Ali Abbasov
Kamil Aida-zade (Eds.)

BOOK OF SELECTED PAPERS

6th International Conference on

Problems of Cybernetics and Informatics

PCI 2025

August 25-28, 2025
Baku, Azerbaijan

Ali Abbasov
Kamil Aida-zade
Editors

BOOK OF SELECTED PAPERS

6th International Conference on Problems of Cybernetics and Informatics

PCI 2025

August 25-28, 2025
Baku, Azerbaijan

Editors

Ali Abbasov
Institute of Control Systems of Ministry of Science and Education of the Republic of Azerbaijan
Baku, Azerbaijan

Kamil Aida-zade
Institute of Control Systems of Ministry of Science and Education of the Republic of Azerbaijan
Baku, Azerbaijan

Technical editors
Ahmadli, Nigar
Alizada, Tahir
Aydazada, Ilham
Azimov, Rustam
Rzayeva, Sona

ISBN: 978-9952-530-36-0

Organizers

Institute of Control Systems (Azerbaijan)

Institute of Information Technology (Azerbaijan)

IEEE Azerbaijan COM/C Joint Chapter

State Examination Center (Azerbaijan)

Ege University (Türkiye)

V.M. Glushkov Institute of Cybernetics (Ukraine)

Institute of Information and Computational Technologies (Kazakhstan)

University of Thessaly (Greece)

Organizing Committee

Chair: Hafiz Pashayev (Azerbaijan)
Co-chair: Ali Abbasov (Azerbaijan)

Members:

- Abdullayev, Vagif (Azerbaijan)
- Agayeva, Cherkez (Türkiye)
- Aliyev, Elchin (Azerbaijan)
- Aliyev, Soltan (Azerbaijan)
- Aliguliyev, Ramiz (Azerbaijan)
- Ashrafova, Yegana (Azerbaijan)
- Eshmamatova, Dilfuza (Uzbekistan)
- Faraslis, Ioannis (Greece)
- Gasimov, Yusif (Azerbaijan)
- Ismayilzada, Fariz (Azerbaijan)
- Kalimoldaev, Maksat (Kazakhstan)
- Melikov, Telman (Azerbaijan)
- Nuriyev, Urfat (Azerbaijan)
- Parasidis, Ioannis (Greece)
- Rahimov, Anar (Azerbaijan)
- Rzayev, Ramin (Azerbaijan)
- Sadigov, Aminaga (Azerbaijan)
- Stetsyuk, Petro (Ukraine)
- Stovba, Viktor (Ukraine)

Program Committee

Chair: Rasim Alguliyev (Azerbaijan)
Co-chair: Kamil Aida-zade (Azerbaijan)
Co-chair: Boris Mordukhovich (USA)
Co-chair: Abzetdin Adamov (IEEE, USA)

Members:

- Aliev, Fikret (Azerbaijan)

- Aliev, Telman (Azerbaijan)
- Aliyev, Rovshan (Azerbaijan)
- Ahmedzade, Perviz (Uzbekistan)
- Akleylek, Sedat (Estonia)
- Artikbayev, Abdulaziz (Uzbekistan)
- Bagirov, Adil (Australia)
- Bashirov, Agamirza (Cyprus)
- Bashirov, Rza (Cyprus)
- Bomba, Andrii (Ukraine)
- Boyd, Stephen (USA)
- Chikriy, Arkadiy (Ukraine)
- Guliyev, Hamlet (Azerbaijan)
- Hajiyev, Asef (Azerbaijan)
- Hajiyev, Chingiz (Türkiye)
- Jaćimović, Milojica (Montenegro)
- Hasanli, Yadulla (Azerbaijan)
- Hasanov, Elman (Türkiye)
- Huseynov, Sharif (Latvia)
- Imanov, Gorkhmaz (Azerbaijan)
- Kasimbeyli, Refail (Türkiye)
- Khaniyev, Tahir (Türkiye)
- Leonov, Sergey (Ukraine)
- Lesnic, Daniel (UK)
- Litman, Amelie (France)
- Lozovanu, Dmitrii (Moldova)
- Mahmudov, Elimhan (Azerbaijan)
- Mahmudov, Nazim (Türkiye)
- Mansimov, Kamil (Azerbaijan)
- Mardanov, Misir (Azerbaijan)
- Melikov, Agassi (Azerbaijan)
- Nakonechniy Alexander (Ukraine)
- Nasiboglu, Efendi (Türkiye)
- Nesterov, Yurii (Belgium)
- Pogoriliy, Sergey (Ukraine)
- Polyak, Roman (USA)
- Providas, Efthimios (Greece)
- Rustamov, Samir (Azerbaijan)
- Serovaisky, Semyon (Kazakhstan)
- Seyidzade, Etibar (IEEE USA)
- Sladkowski, Aleksander (Poland)
- Slodicka, Marian (Belgium)
- Stetsyuk, Petro (Ukraine)
- Tadumadze, Nodar (Georgia)

Preface

The 6th International Conference on Problems of Cybernetics and Informatics (PCI 2025) was held on August 25–28, 2025, in Baku, Azerbaijan. The conference was jointly organized by Institute of Control Systems and Institute of Information Technology, Ege University (Turkey), V.M. Glushkov Institute of Cybernetics (Ukraine), Institute of Information and Computational Technologies (Kazakhstan), and University of Thessaly (Greece).

Since 2006, the PCI conference series has served as a major international platform for discussing current scientific and applied problems in cybernetics, control theory, and informatics. PCI 2025 continued this tradition by bringing together leading researchers and experts from more than 30 countries, including Australia, Austria, Belarus, Belgium, Bulgaria, the United Kingdom, Hungary, Greece, Georgia, Canada, Kazakhstan, Cyprus, China, Kyrgyzstan, Latvia, Moldova, Poland, Russia, Slovakia, the USA, Turkey, Ukraine, Uzbekistan, France, Montenegro, Sweden, Estonia, and Japan. The hybrid format of the event enabled broad participation from the scientific community and contributed to strengthening global cooperation.

The conference was opened by Professor Ali Abbasov, Director General of Institute of Control Systems, who emphasized the relevance of the conference topics in the context of rapid socio-economic and technological changes. He highlighted the growing role of artificial intelligence technologies, networked control systems, and natural language processing in solving modern management challenges. He also noted that the recently approved "Artificial Intelligence Strategy of the Republic of Azerbaijan for 2025–2028" and the "Digital Development Concept" create favorable conditions for the country's digital transformation.

Professor Rasim Aliguliyev, Vice-President of ANAS and Director General of the Institute of Information Technology, welcomed the participants on behalf of the Presidium of ANAS. He emphasized the high scientific significance of the conference and highlighted the long-standing contributions of Azerbaijani scholars — Professor Telman Aliyev, Professor Ali Abbasov, Professor Gorkhmaz Imanov, Professor Kamil Aida-zade, and Professor Ramiz Aliguliyev — to the development of artificial intelligence and cybernetics in the country.

During the opening ceremony, speeches were also delivered by Professor Aminaga Sadygov, Rector of Sumgayit State University; Professor Boris Mordukhovich (Wayne State University, USA); Professor Orken Mamyrbayev, Deputy Director for Research (Institute of Information and Computational Technologies, Kazakhstan); and Professor Abzhetdin Adamov (IEEE representative, ADA University). The speakers emphasized the importance of strengthening international scientific cooperation, supporting young researchers, and promoting interdisciplinary research at the intersection of cybernetics, optimization, artificial intelligence, and computational technologies.

The conference program included plenary talks and several thematic sessions covering optimization and optimal control, mathematical modeling, parametric identification, numerical methods, intelligent systems, pattern recognition, stochastic models, and applications for complex technical systems. The plenary speakers — Professor Piotr Stetsyuk (V.M. Glushkov Institute of Cybernetics, Ukraine), Professor Chingiz Hajiyev (Istanbul Technical University), Professor Orken Mamyrbayev, Professor Semyon Serovaisky (Al-Farabi Kazakh National University), and Dr. Jamaladdin Hasanov (Associate Professor, ADA University) — presented cutting-edge research on nonsmooth optimization, nanosatellite attitude estimation, and AI-based biomedical systems.

PCI 2025 attracted significant international attention and received congratulatory letters from scientific institutions around the world. Among them was a letter from Professor Maktagali Bektemesov, Director General of the Institute of Information and Computational Technologies of Kazakhstan, who highlighted the strengthening of scientific cooperation between Azerbaijan and Kazakhstan.

We express our sincere gratitude to all authors, reviewers, invited speakers, members of the Program and Organizing Committees, and all institutions that supported PCI 2025. We believe that the ideas and results preprofsented at the conference will contribute to the development of new scientific collaborations, support young researchers, and make a significant contribution to the advancement of cybernetics and informatics.

August 2025

Ali Abbasov
Kamil Aida-zade

Table of Contents

Plenary Section

Nanosatellite Attitude Estimation and In-Orbit Sensor Calibration Using Only Kinematic Relations

Chingiz Hajiyev
Dept.of Aeronautical Engineering
Istanbul Technical University
Istanbul, Türkiye
0000-0003-4115-341X

Demet Cilden-Guler
Dept. of Astronautical Engineering
Istanbul Technical University
Istanbul, Türkiye
0000-0002-3924-5422

***Abstract*— In this study, the attitude angles of a nanosatellite are estimated using vector measurements obtained from onboard magnetometers and sun sensors. A nontraditional filtering approach is adopted, relying solely on a kinematic model that is propagated using rate gyroscope and magnetometer data. The estimation process consists of two main stages. In the first stage, attitude measurements are determined using a singular value decomposition (SVD)-based single-frame method. In the second stage, an Extended Kalman Filter (EKF) is implemented, incorporating a linearized attitude model and nonlinear magnetometer measurements. These two stages are integrated into a unified estimation algorithm—referred to as the SVD-Aided EKF—which enables the simultaneous estimation of attitude angles, as well as gyroscope and magnetometer biases, with high accuracy. The rate gyroscope and magnetometer measurements are corrected at each estimation step by subtracting the bias estimates from the corresponding sensor measurements. The bias-free gyroscope measurements are then fed back into the kinematic model, and the bias-free magnetometer measurements are fed back into the SVD algorithm and the measurement update part of Kalman filter, which are also used in the SVD-Aided EKF.**

Keywords—nanosatellite, attitude estimation, kinematics, extended Kalman filter, magnetometer, sun sensor, rate gyroscope

INTRODUCTION

The orientation of a satellite can be estimated using Kalman filtering techniques by combining body-frame measurements with known reference vectors. Conventional approaches for estimating a satellite's attitude angles and angular rates within an Extended Kalman Filter (EKF) framework typically rely on nonlinear vector observations, as the measurement models are inherently nonlinear [1, 2]. Alternatively, linear measurement-based methods use attitude estimates obtained through single-frame attitude determination techniques [3]. These techniques calculate the attitude independently at each time step using vector observations and are then incorporated into the Kalman filter as direct measurement inputs [4, 5]. Filters employing this method are often referred to as single-frame method-aided filters.

To propagate the satellite's orientation within the filter, either a kinematic model alone or a combination of kinematic and dynamic models can be employed. However, dynamics-based models often introduce modeling errors, particularly due to uncertainties in parameters such as the satellite's inertia matrix [2]. This limitation has led to the development of algorithms that rely solely on kinematics, thereby avoiding errors linked to uncertain dynamics.

There are two principal strategies in satellite attitude and angular rate estimation: (1) using only the kinematic equations, or (2) integrating both kinematic and dynamic equations. Replacing a gyro-driven propagation model with one based on rotational dynamics enables the removal of onboard gyroscopes, but this substitution requires an accurate and reliable dynamics model. The study in [6] explores this trade-off by analyzing the threshold of uncertainty in the satellite's mass moment of inertia beyond which the benefits of using a dynamics-based model diminish.

In this context, we present a single-frame method-aided filtering approach that operates with both kinematic-only and combined kinematic-dynamic models. The method is particularly suitable for small satellites, where the accuracy of the inertia matrix may degrade over time. Simulation results demonstrate that when there is no uncertainty in the inertia matrix, the combined model offers superior estimation of gyro biases. However, as uncertainties increase—particularly beyond a 3% deviation in the principal moments of inertia—the purely kinematics-based filter achieves better performance. Therefore, for small satellite missions where inertia uncertainties are expected to evolve over time, kinematics-only attitude estimation provides a more robust and reliable solution by mitigating the impact of uncertain dynamic parameters.

In this study, we propose a single-frame method-aided filter that relies exclusively on the kinematic model, omitting the use of satellite dynamics, making it well-suited for nanosatellite applications. The filter, referred to as the SVD-Aided EKF, utilizes magnetometer and sun sensor measurements to determine the satellite's attitude through a singular value decomposition (SVD)-based attitude determination method. These SVD-derived attitude angles are treated as measurements and incorporated into the Kalman filter. The SVD-Aided EKF concurrently estimates the attitude angles, as well as the biases of the gyroscope and magnetometer. At each estimation step, the raw sensor data are corrected by subtracting the estimated biases, resulting in unbiased gyroscope and magnetometer measurements. The corrected gyroscope data are used to propagate the attitude via the kinematic model, while the corrected magnetometer data are fed back into both the SVD algorithm and the measurement update step of the Kalman filter. This closed-loop structure enhances estimation accuracy by continuously refining both the state and measurement inputs within the filter framework.

MATHEMATICAL MODELS

Model of attitude kinematics

The kinematic equations can be written using Euler angles as follows:

$$\begin{bmatrix}\dot{\phi}\\ \dot{\theta}\\ \dot{\psi}\end{bmatrix}=\begin{bmatrix}1 & s(\phi)t(\theta) & c(\phi)t(\theta)\\ 0 & c(\phi) & -s(\phi)\\ 0 & s(\phi)/c(\theta) & c(\phi)/c(\theta)\end{bmatrix}\begin{bmatrix}p\\ q\\ r\end{bmatrix}, \quad (1)$$

where $c(\cdot)$, $s(\cdot)$ and $t(\cdot)$ are cosine, sine and tangent functions, ϕ,θ,ψ are the yaw, pitch, and roll angles respectively, and $p,\ q,\ r$ are the components of the $\bar{\omega}_{BR}$ vector of the body frame with respect to the reference frame.

The gyro measurement model

The rate gyro measurements can be modelled as,

$$\omega_{BI_m}(k)=\omega_{BI}(k)+\eta_g(k)+b_g(k) \quad (2)$$

where $\omega_{BI}(k)=\begin{bmatrix}\omega_x(k) & \omega_y(k) & \omega_z(k)\end{bmatrix}^T$ is the angular velocity vector of the body frame with respect to the inertial frame, $b_g(k)$ is the gyro bias vector and $\eta_g(k)$ is the zero mean white noise with the characteristic of,

$$E\left[\eta_g(k)\eta_g^T(j)\right]=I_{3x3}\sigma_g^2\delta(kj), \quad (3)$$

where I_{3x3} is the identity matrix with the dimension of 3×3, $\delta(kj)$ is the Kronecker delta function, σ_g is the standard deviation of rate gyro error. The characteristic of gyro biases $b_g(k)=\begin{bmatrix}b_{g_x}(k) & b_{g_y}(k) & b_{g_z}(k)\end{bmatrix}^T$ is given as,

$$b_g(k+1)=b_g(k)+\eta_1(k)\times\Delta t \quad (4)$$

where $\eta_1(k)$ is the zero mean white noise.

The magnetometer measurement model

Three onboard magnetometers of satellite measure the components of the magnetic field vector in the body frame. Overall measurement model may be given as;

$$\begin{bmatrix}B_{mx}(k)\\ B_{my}(k)\\ B_{mz}(k)\end{bmatrix}=\left(A\begin{bmatrix}B_{ox}(k)\\ B_{oy}(k)\\ B_{oz}(k)\end{bmatrix}+b_m(k)+\eta_m(k)\right), \quad (5)$$

where, $B_{ox}(k)$, $B_{oy}(k)$ and $B_{oz}(k)$ are the Earth geomagnetic field direction cosines in the orbit frame, $B_{mx}(k)$, $B_{my}(k)$ and $B_{mz}(k)$ are the measured Earth geomagnetic field direction cosines in body frame, $b_m(k)$ is the magnetometer bias vector as $b_m(k)=\begin{bmatrix}b_{m_x}(k) & b_{m_y}(k) & b_{m_z}(k)\end{bmatrix}^T$ and $\eta_m(k)$ is the zero mean Gaussian white noise with the covariance matrix of

$$E\left[\eta_m(k)\eta_m^T(j)\right]=I_{3x3}\sigma_m^2\delta(kj)=R_m(k). \quad (6)$$

where σ_m is the standard deviation of the magnetometer error. The characteristic of magnetometer biases $b_m(k)=\begin{bmatrix}b_{m_x}(k) & b_{m_y}(k) & b_{m_z}(k)\end{bmatrix}^T$ is given as,

$$b_m(k+1)=b_m(k)+\eta_2(k)\times\Delta t \quad (7)$$

where $\eta_2(k)$ is the zero mean white noise.

The sun sensor measurement model

The unit sun direction vector in the Earth-Centered Inertial (ECI) frame can be computed using the linear model based on the sun's ecliptic longitude [2]. Orbital elements from the satellite's orbit propagation model are required to transform the unit sun direction vector from the ECI frame into the orbital frame. Once this transformation is completed, the model for the sun sensor measurements can be expressed as follows:

$$S_{mes}(k)=S_b(k)=AS_o(k)+\eta_s(k), \quad (8)$$

where, $S_o(k)$ is the sun direction vector in the orbit frame and $S_b(k)$ are the sun sensor measurements in body frame which are corrupted with, $\eta_s(k)$, the zero mean Gaussian white noise with the characteristic of

$$E\left[\eta_s(k)\eta_s^T(j)\right]=I_{3x3}\sigma_s^2\delta(kj). \quad (9)$$

where σ_s is the standard deviation of sun sensor measurement noise.

SINGLE-FRAME ATTITUDE DETERMINATION METHOD

As single-frame attitude determination method the Singular Value Decomposition (SVD) approach was employed [7]. The loss is caused from the divergence of the measurements from the corresponding reference models

$$L(A)=\frac{1}{2}\sum_i a_i\left|\mathrm{b}_i-\mathrm{Ar}_i\right|^2 \quad (10)$$

$$B=\sum a_i b_i r_i^T \quad (11)$$

$$L(A)=\sum a_i-tr\left(AB^T\right) \quad (12)$$

where b_i is measurement vector, r_i is reference vector, a_i is non-negative weight.

$$B = USV^T = U\text{diag}\left(\left[S_{11} \;\; S_{22} \;\; S_{33}\right]\right)V^T \tag{13}$$

$$A_{opt} = Udiag[1 \;\; 1 \;\; \det(U)\det(V)]V^T \tag{14}$$

According to SVD, the U and V matrices are left and right orthogonal matrices. They use main singular values $\left(S_{11},\ S_{22},\ S_{33}\right)$ to express the matrix B. The attitude matrix A_{opt} can be used to determine the Euler angle values. The attitude angle estimation errors' covariance matrix is

$$P_{SVD} = Udiag[(\mathrm{s}_2+\mathrm{s}_3)^{-1} \;\; (\mathrm{s}_3+\mathrm{s}_1)^{-1} \;\; (\mathrm{s}_1+\mathrm{s}_2)^{-1}]U^T \tag{15}$$

where $s_1 = S_{11} \;\; s_2 = S_{22} \;\; s_3 = \det\left(U\right)\det\left(V\right)S_{33}$.

IV. SVD-AIDED EKF FOR ATTITUDE ESTIMATION

Gyroscopes, magnetometers and sun sensors are used as the attitude and rate sensors in this study. The measurements can be presented as,

$$\begin{aligned} Z_\phi(k) &= \phi(k) + v_\phi(k) \\ Z_\theta(k) &= \theta(k) + v_\theta(k) \\ Z_\psi(k) &= \psi(k) + v_\psi(k) \end{aligned} \tag{16}$$

where $Z_\phi(k), \;\; Z_\theta(k), \;\; Z_\psi(k)$ represent the attitude angle measurements determined by SVD method, $v_{(\cdot)}(k)$ is the measurement noises of the attitude angles. We can call the SVD measurements as $Z_1(k) = \left[Z_\phi(k) \;\; Z_\theta(k) \;\; Z_\psi(k)\right]^T$. The gyro measurements described in (3) can be expressed as $Z_2(k) = \omega_{BI_m}(k)$. The magnetometer measurements described in (7) can be written as $Z_3(k) = \left[B_{mx}(k) \;\; B_{my}(k) \;\; B_{mz}(k)\right]^T$.

$Z_1(k)$ represents the measurements from SVD using the magnetometer and sun sensor measurements. The corrected (bias eliminated) rate gyro measurements can be represented from the rate gyroscope measurements as $Z_2^{cor}(k) = Z_2(k) - \hat{b}_g(k-1)$ where $b_{g_{est}}(k-1)$ is the estimated gyro bias vector by the filter.

The augmented state vector in this case is

$$x(k) = \left[\phi(k) \;\; \theta(k) \;\; \psi(k) \;\; b_{g_x}(k) \;\; b_{g_y}(k) \;\; b_{g_z}(k) \;\; b_{m_x}(k) \;\; b_{m_y}(k) \;\; b_{m_z}(k)\right]^T \tag{17}$$

The Kalman filter uses only kinematics and measurements $Z(k) = \left[Z_1(k) \;\; Z_3(k)\right]^T$ for attitude estimation and calibration of attitude sensors. This means that the measurement input vector is composed of the attitude measurements from SVD $Z_1(k)$ and magnetometer measurements $Z_3(k)$. Consequently, the measurement vector $Z(k)$ consists of the output signals of the SVD and the nonlinear measurements of the magnetometers and can be written as follows:

$$Z(k) = \left[Z_\phi(k) \;\; Z_\theta(k) \;\; Z_\psi(k) \;\; B_{mx}(k) \;\; B_{my}(k) \;\; B_{mz}(k)\right]^T \tag{18}$$

The corrected (bias eliminated) rate gyro measurements can be determined from the rate gyroscope measurements as $Z_2^{cor}(k) = Z_2(k) - \hat{b}_g(k-1)$ where $\hat{b}_g(k-1)$ is the estimated gyro bias vector by the filter. In this case, the covariance matrix of the corrected gyroscope measurement noise will be determined as follows:

$$E\left[\eta_g(k)\eta_g^T(j)\right] = \begin{bmatrix} \sigma_g^2 + Var(\hat{b}_{g_x}) & 0 & 0 \\ 0 & \sigma_g^2 + Var(\hat{b}_{g_y}) & 0 \\ 0 & 0 & \sigma_g^2 + Var(\hat{b}_{g_z}) \end{bmatrix} \tag{19}$$

Here $Var(\hat{b}_{g_x})$, $Var(\hat{b}_{g_y})$ and $Var(\hat{b}_{g_z})$ are the variances of gyroscope bias estimation errors. After this procedure, the bias-free rate gyro measurements are fed back into the kinematics model which is also used in the SVD-Aided EKF.

The corrected (bias eliminated) magnetometer measurements can be represented from the magnetometer measurements as $Z_3^{cor}(k) = Z_3(k) - \hat{b}_m(k-1)$ where $\hat{b}_m(k-1)$ is the estimated magnetometer bias vector by the filter. In this case, the corrected covariance matrix of the magnetometer measurement noise is written in the following form:

$$E\left[\eta_m(k)\eta_m^T(j)\right] = R_m^{cor} = \begin{bmatrix} \sigma_m^2 + Var(\hat{b}_{m_x}) & 0 & 0 \\ 0 & \sigma_m^2 + Var(\hat{b}_{m_y}) & 0 \\ 0 & 0 & \sigma_m^2 + Var(\hat{b}_{m_z}) \end{bmatrix} \tag{20}$$

Here $Var(\hat{b}_{m_x})$, $Var(\hat{b}_{m_y})$, $Var(\hat{b}_{m_z})$ are the variances of magnetometer bias estimation errors. The bias-free magnetometer measurements and the corrected covariance matrix of the magnetometer measurement noise are then fed back into the SVD algorithm and measurement update part of Kalman filter, which are also used in the SVD-Aided EKF.

The discrete-time nonlinear state space model is used to formulate the satellite's rotational motion about its mass center

$$x(k+1) = f\left[x(k)\right] + w(k) \tag{21}$$

$$Z(k)=\begin{bmatrix} Z_1(k) \\ Z_3(k) \end{bmatrix}=\begin{bmatrix} H_1 x(k)+v_1 \\ h[x(k)]+b_m(k)+\eta_m(k) \end{bmatrix} \tag{22}$$

where $f[\cdot]$ and $h[\cdot]$ are the nonlinear kinematic and magnetometer measurement functions respectively, $w(k)$ is the zero-mean Gaussian noise with covariance $Q(k)$, $Z(k)$ is the 6-dimensional measurement vector, $v_1(k)$ is the zero-mean Gaussian noise with covariance $P_{SVD}(k)$, $\eta_m(k)$ is the zero-mean Gaussian noise with covariance (20), H_1 is the measurement matrix corresponding to the linear measurements (16) and can be expressed as,

$$H_1=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{23}$$

It is assumed that both noise vectors $w(k)$ and $v(k)=[v_1(k) \quad \eta_m(k)]^T$ are uncorrelated with zero mean.

The filter algorithm is formulated based on the system kinematics described in Equation (1) and the measurement model defined in Equations (16) and (5). The prediction and update steps follow the standard Extended Kalman Filter (EKF) framework, incorporating necessary linearization and approximations. In this unconventional EKF structure, called SVD-Aided EKF, the satellite's rotational motion parameters and the attitude sensor biases are jointly estimated. The estimation process is carried out through a sequence of steps that utilize both the system model and the SVD-based attitude measurements, ensuring accurate and consistent state updates. The estimated value is found as,

$$\hat{x}(k+1)=\hat{x}(k+1/k)+K(k+1)\times \left\{\begin{bmatrix} Z_1(k+1) \\ Z_3(k+1) \end{bmatrix}-\begin{bmatrix} H_1\hat{x}(k+1/k) \\ h[\hat{x}(k+1/k)]+\hat{b}_m(k+1) \end{bmatrix}\right\} \tag{24}$$

The extrapolation value can be determined as,

$$\hat{x}(k+1/k)=f[\hat{x}(k)] \tag{25}$$

Filter-gain of the EKF is,

$$K(k+1)=P(k+1/k)H^T(k+1)\times \left[H(k+1)P(k+1/k)H^T(k+1)+R(k+1)\right]^{-1} \tag{26}$$

where $H(k+1)$ is the 6×9 system measurement matrix

$$H(k+1)=\begin{bmatrix} H_{1(3\times 9)} \\ \left(\dfrac{\partial h[\hat{x}(k+1/k)]}{\partial \hat{x}(k+1/k)}\right)_{(3\times 9)} \end{bmatrix} \tag{27}$$

$\left(\dfrac{\partial h[\hat{x}(k+1/k)]}{\partial \hat{x}(k+1/k)}\right)$ is the 3×9 partial derivatives of the measurement function with respect to the states $\hat{x}(k+1/k)$, R is the covariance matrix of the SVD-Aided EKF measurement noise

$$R=\begin{bmatrix} P_{SVD} & 0_{3\times 3} \\ 0_{3\times 3} & R_m^{cor} \end{bmatrix} \tag{28}$$

The covariance matrix of the extrapolation error is,

$$P(k+1/k)=\frac{\partial f[\hat{x}(k)]}{\partial \hat{x}(k)}P(k/k)\frac{\partial f^T[\hat{x}(k)]}{\partial \hat{x}(k)}+Q(k) \tag{29}$$

The covariance matrix of the estimation error is,

$$P(k+1/k+1)=\left[I-K(k+1)H(k+1)\right]P(k+1/k) \tag{30}$$

The expressions (24)-(30) express the proposed SVD-Aided EKF for estimating the nanosatellite orientation angles and the gyroscope and magnetometer biases.

V. ANALYSIS AND RESULTS

In this study, we considered a nanosatellite with mass moment of inertia $J=\text{diag}\left(\left[2.1\times10^{-3} \quad 2.0\times10^{-3} \quad 1.9\times10^{-3}\right]\right)$ kg.m^2 having a magnetometer, sun sensor and gyroscope as attitude and rate sensors. Fig. 1 represents the estimation results of attitude angles by SVD-Aided EKF. As can be seen, the estimates are very close to the actual values. Figs. 2 and 3 show the errors in estimating the gyroscope bias and the variance of the errors over time. It is concluded that SVD-Aided EKF can accurately estimate the attitude angles and bias of gyroscopes using only kinematics relations.

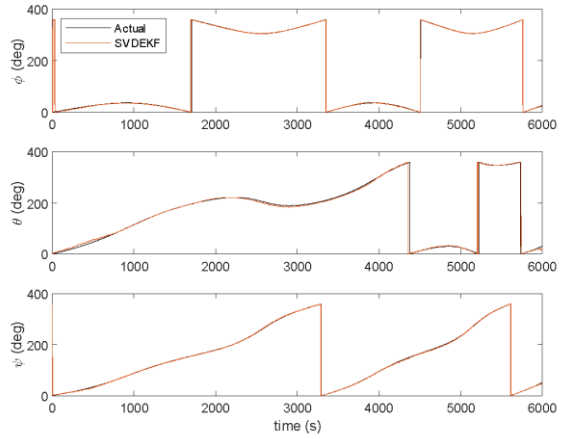


Fig.1. Attitude angles estimation results by SVD-aided EKF

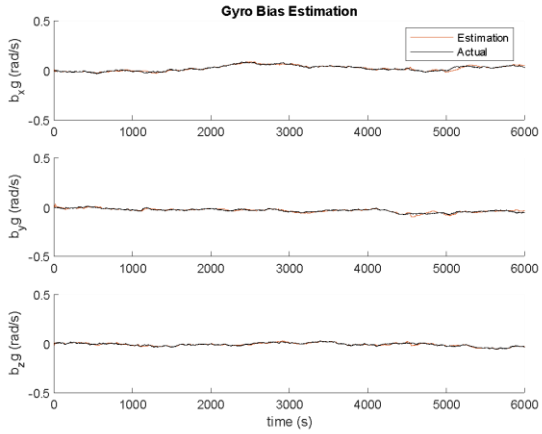


Fig.2. Gyroscope bias estimation results by SVD-aided EKF

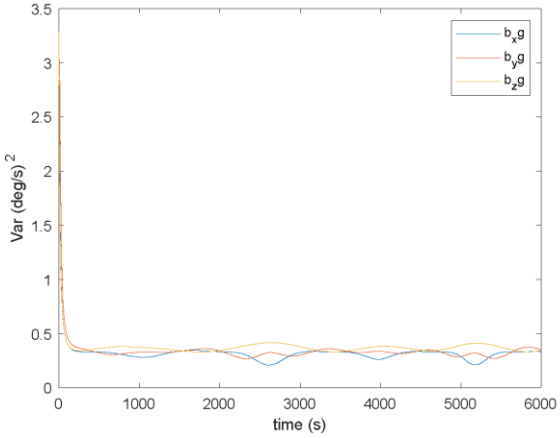


Fig.3. Variances of gyroscope bias estimation errors by SVD-aided EKF

Corrected rate gyroscope measurements can be determined by subtracting the gyroscope bias estimates from the rate gyroscope measurements

$$Z_2^{cor}(k) = Z_2(k) - \hat{b}_g(k-1) = \omega_{BI_m}(k) - \hat{b}_g(k-1) \quad (31)$$

A comparison of the corrected rate gyro measurements with their actual values is shown in the graphs presented in Fig. 4. As can be seen, the corrected rate gyro measurements are very close to the actual values.

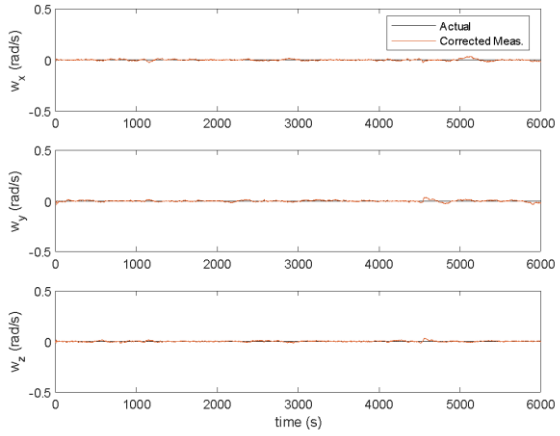


Fig.4. Corrected rate gyroscope measurements

The variances of the corrected rate gyroscope measurement errors, calculated using formula (19), are shown in Fig. 5. As can be seen, the variances are not constant and change depending on the accuracy of the rate gyro bias estimate. Figs. 6 and 7 show the errors in estimating the magnetometer bias and the error variance.

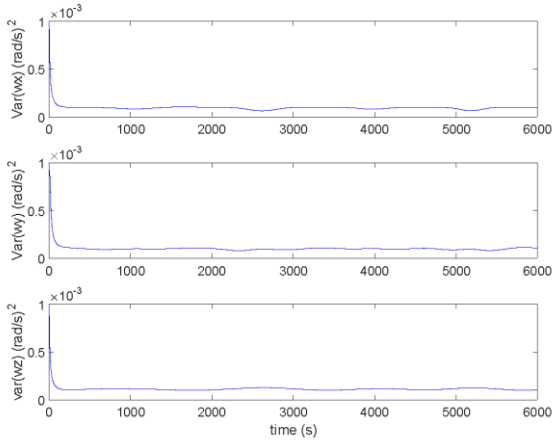


Fig.5. Variances of the corrected rate gyroscope measurements errors

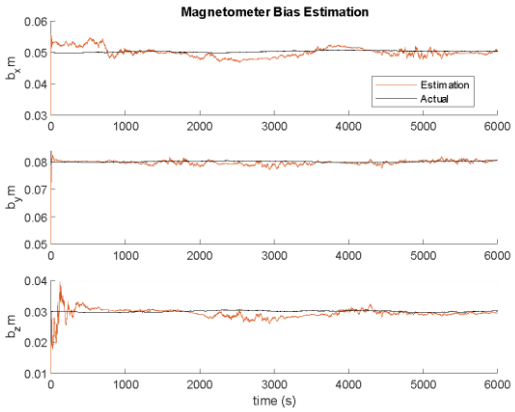


Fig.6. Magnetometer bias estimation results by SVD-aided EKF

As can be seen, the proposed method is able to accurately estimate magnetometer biases using only kinematic relations.

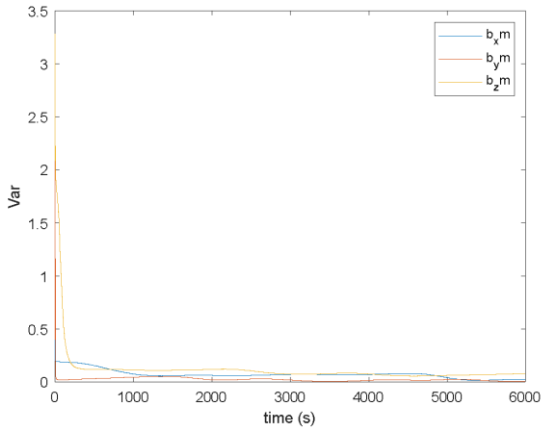


Fig.7. Variances of magnetometer bias estimation errors by SVD-aided EKF

CONCLUSIONS

In this study, we present a single-frame method-aided filter that utilizes only the kinematic equations of a nanosatellite, without relying on a dynamic model. The system incorporates three types of onboard sensors for attitude and rate estimation: magnetometers, sun sensors, and gyroscopes. Attitude angles are computed using a singular value decomposition (SVD)-based method applied to magnetometer and sun sensor measurements. These angles serve as input observations for the proposed filter, referred to as the SVD-Aided EKF. The filter simultaneously estimates the satellite's attitude and the biases in both the gyroscope and magnetometer measurements.

Once the biases are estimated, the corrected (bias-free) gyroscope data are used to propagate the attitude via the kinematic model within the EKF framework. Similarly, the bias-free magnetometer readings, along with an updated magnetometer noise covariance matrix, are fed back into the SVD algorithm and the measurement update step of the Kalman filter, maintaining consistency and accuracy throughout the estimation process.

Future work may focus on comparing the performance of this kinematics-only approach with a filter that integrates both kinematic and dynamic models, particularly under various levels of uncertainty in the satellite's dynamic parameters.

REFERENCES

[1] F.L. Markley, J.L. Crassidis, Y. Cheng, "Nonlinear attitude filtering methods," in: AIAA Guidance, Navigation, and Control Conference and Exhibit. San Francisco, California, 2005.

[2] J.L. Crassidis, F.L. Markley, Y. Cheng, 2007, "Survey of nonlinear attitude estimation methods," J. Guid. Control. Dyn., vol. 30, 2007, pp. 12–28. doi:10.2514/1.22452

[3] D. Cilden-Guler, E.S. Conguroglu, C. Hajiyev, "Single-frame attitude determination methods for nanosatellites," *Metrol. Meas. Syst.,* vol. 24, 2017, pp. 313–324. doi:10.1515/mms-2017-0023

[4] C. Hajiyev, M. Bahar, 2003, "Attitude determination and control system design of the ITU-UUBF LEO1 satellite," Acta Astronaut., vol. 52, pp. 493–499. doi:10.1016/S0094-5765(02)00192-3

[5] C. Hajiyev, D. Cilden, "Nontraditional approach to satellite attitude estimation," *Int. J. Control Syst. Robot.*, 2016, vol.1, pp. 19–28.

[6] D. Cilden-Guler, C. Hajiyev, "SVD-aided EKF for nanosatellite attitude estimation based on kinematic and dynamic relations," Gyroscopy and Navigation, vol. 14, no. 4, 2023, pp. 366–379.

[7] F.L. Markley, D. Mortari, "Quaternion attitude estimation using vector observations," J. Astronaut. Sci., vol. 48, 2000, pp. 359–380.

Section 1: Application of AI in Applied Problems

Clinical Decision Support System for Hepatocellular Carcinoma Treatment Selection

Masuma Mammadova
Department No. 11 Institute of Information Technology of MSERA
Baku, Azerbaijan
mmg51@mail.ru
0000-0002-2205-1023

Nuru Bayramov
I Department of Surgical Diseases of Azerbaijan Medical University
Baku, Azerbaijan
nurubay2006@yahoo.com,
0000-0001-6958-5412

Zarifa Jabrayilova
Department No. 11 Institute of Information Technology of MSERA
Baku, Azerbaijan
djabrailova_z@mail.ru
0000-0002-9661-5805,

Minara Manafli
Department No. 11 Institute of Information Technology of MSERA
Baku, Azerbaijan
minulya@rambler.ru

Mehriban Huseynova
I Department of Surgical Diseases of Azerbaijan Medical University
Baku, Azerbaijan
mehribanhuseynova9@gmail.com

***Abstract*—The article presents an algorithm for creating an intelligent physician decision support system regarding the initial treatment selection for hepatocellular carcinoma (HCC), which is the 5th-6th most common cancer. HCC is determined by a large number of parameters, and according to the values they can receive, resection, transplantation, ablation, embolization, systemic treatment, symptomatic treatments are selected according to the staged classical approach to the HCC treatment based on the clinical situations that arise. The large number, complexity, quantitative and qualitative character of the input parameters symbolizes the choice of treatment for HCC as a difficult formalized issue and makes the development of an intelligent system for its solution relevant. In this regard, this article solves the issue of integrating the doctors' knowledge collected in a multidisciplinary consultation into the knowledge base of the intelligent system for the individual choice of treatment for HCC, and develops the working principle and implementation algorithm of the system's architectural components.**

Keywords—hepatocellular carcinoma; treatment selection; intelligent physician decision support system; descriptive model of knowledge.

INTRODUCTION

Among the most widespread malignant liver tumors, HCC is one of the main causes of cancer-related death in the world [1]. HCC is rated 5-6th in terms of incidence among cancers. HCC is diagnosed in about one million people worldwide every year [2]. HCC, the most widespread among malignant liver tumors, reveals in various clinical conditions and is often mistaken for hepatitis and cirrhosis [3]. The main diagnostic sign of HCC is contrast retention in the arterial phase and contrast washout in the venous phase. It is most common in people aged 60-70 and mainly in males (2.5 times more often than in females). In some risk countries (Far East, Africa), it is more often observed in people aged 30-40. HCC is a progressive and irreversible process and, if not treated, causes complications and lethality within a few months. The following methods are used in the treatment of HCC [4-6]:

- Surgical treatments - liver resection and transplantation;
- Loco-regional treatments – ablations (thermal, electro and chemical destructions), intraarterial embolization (chemoembolization, radioembolization), radiotherapy, arterial infusions;
- Systemic anticancer treatments (targeted, immunotherapy, cytotoxic chemotherapy);
- Symptomatic treatments.

Liver resection, transplantation and ablations of small tumors are considered radical treatments of HCC. Loco-regional and systemic methods aim to reduce tumor progression and aggressiveness. Indications for liver resection in HCC are as follows:

- Tumor is resectable – according to its bio-behavior, it is suitable for surgery, there is no metastasis or it can be removed, the tumor can be completely removed and the blood supply of the remaining liver is preserved, there is no invasion of large vessels and surrounding organs;
- The liver is resectable – it has sufficient reserves (no or minimal parenchymal disease and dysfunction – Child A, MELD<9, ISG<35, Elastography<12), there is no portal hypertension, the remaining liver volume is sufficient (more than 30% in normal liver, more than 50% in cirrhosis);
- The patient is operable – comorbidities and performance allow surgery.

The indications for liver transplantation in HCC are as follows:

- The tumor is transplantable – the tumor can be removed, the tumor has a favorable bio-behavior (low aggressiveness): its size is less than 5 cm, its number is less than 3, there is no invasion or metastasis to large vessels, lymph nodes, or surrounding organs, AFP (α-fetoprotein is a primary liver cancer marker) level is below 200-400.
- Transplantation is possible – there is a donor organ, there are no contraindications to immunosuppression and pulmonary hypertension.
- The patient is operable – concomitant diseases and performance allow surgery and immunosuppression.

In the presence of contraindications to surgery, that is, signs of unresectable, non-transplantable, and inoperability, resection and transplantation are considered contraindicated. In general, the classical approach to the treatment of HCC takes into account the stage of the disease, the tumor's bio-behavior, the liver's condition, the patient's general status, and the effectiveness of the method [7-9].

The first choice in treatment is radical methods (resection, transplantation), the second choice is locoregional treatments (ablation, embolization), the third choice is systemic anticancer therapies, and the last choice is symptomatic treatment. For the general approach algorithm, international classifications (clinical, NCCN, Barcelona, Milan, Malatya, etc.) are recommended, and for the individual treatment selection, discussion in a multidisciplinary council is suggested. The above shows that a large number of parameters refers to in the treatment selection of HCC, which is represented by various clinical symptoms. According to the clinical situations arising from the various combinations of values that these parameters can take, one of the above treatment methods is selected in the multidisciplinary council for stage-based treatment selection.

The large number, complexity, hierarchy, and both quantitative and qualitative nature of the referred parameters justify the fact that the issue of HCC treatment selection is realized in an uncertain environment, and necessitates the development of an intelligent system based on the knowledge of medical specialists.

The creation of such a system requires the development of its architectural and technological principles, the integration of knowledge received from medical specialists into the system, and the algorithm for implementing the system. The presented article is devoted to solving these problems.

Review Of Related Works

Although today, with the application of information technologies and artificial intelligence methods, a traditional methodology has been developed for building medical decision support systems in various fields of medicine [10], there are no instrumental tools in the literature for medical decision support to select treatment for HCC. [11] presents the functioning principles of a medical decision support system for the diagnosis of liver diseases in a local environment. The system includes a knowledge base consisting of 28 rules generated based on the expert physician's knowledge in the diagnosis of liver diseases, a conclusion block, and an interface block. Although the automation of the diagnostic decision-making process related to liver disease is solved in this system, the system doesn't support the decision-making process related to specific diseases such as hepatitis, obesity, cirrhosis, and HCC. [12] describes a complex approach based on principal component analysis and K-nearest neighbors methods in the creation of a liver disease diagnostic system. [13] develops models based on artificial neural networks for the detection of hepatitis disease.

[14] develops a semi-automatic segmentation method for liver tumors, [15] uses fuzzy logic to solve the classification of hepatobiliary diseases. [16] performs the integration of artificial neural networks and fuzzy logic to detect liver diseases, [17] applies the integration of artificial neural networks and fuzzy logic to increase the accuracy of liver disease classification, and [18] uses the integration of artificial neural networks and fuzzy logic to evaluate the accuracy of hepatitis prognosis in addition to the classification of liver diseases. In [19], the integration of artificial neural networks and genetic algorithms is performed to detect liver diseases and stabilize liver fibrosis in chronic hepatitis. In [20], the integration of fuzzy logic and genetic algorithms is used to detect liver diseases, and in [21], the integration of artificial neural networks and genetic algorithms is used to make decisions about liver transplantation. In [22], the possibilities of using machine learning methods for predicting HCC are studied, and the results of the research conducted to select the most effective method are presented. Apparently, these studies generally examined the possibilities of solving various problems with the application of intelligent technologies in liver diseases. The solution of the problem of building intelligent systems based on the knowledge of medical specialists for the diagnosis and selection of treatment methods for HCC, the most widespread among malignant liver tumors, was not considered. In [23-25], it is substantiated that the parameters used for the diagnosis of HCC are multifactorial, manifest themselves with weak and atypical symptoms, information is incomplete and inaccurate, and the decision-making process takes place in terms of uncertainty. In [23], fuzzy rule-based conceptual model of a system based on the decision tree approach for the diagnosis of HCC, namely, the production and logical model of fuzzy logic and knowledge description, is given, and in [24], the mechanisms for transforming knowledge obtained from medical experts into rules based on the production model for the determination of HCC stages, and the formation of the knowledge base of the diagnostic system are developed. The logical continuation of these studies is the development of the functioning principles of an expert system for the diagnosis of HCC [25].

The **purpose of the present work** is to develop a methodology for creating an intelligent physician decision support system using a classical stage-based approach in the selection of HCC treatment.

Result

To achieve the goal set, it is intended to solve the issues of developing and practical implementation of an intelligent support system for the selection of treatment for HCC, that is, the classification of all possible clinical situations of HCC into resection, transplantation, ablation, embolization, systemic and symptomatic treatment.

The intelligent support system for the selection of HCC treatment consists of a database (DB), a knowledge base (KB), inference engine (IE) and an interface block that provides mutual communication between the user and the system. The intended goal is possible to be achieved, when these components interact in a single system. Thus, the solution to the problem is implemented in the following sequence.

Building a database of an intelligent support system for the selection of HCC treatment

The knowledge of the medical specialists of the First Surgical Diseases Clinic of the Azerbaijan Medical University and the classical approach to decision-making are used as a source of knowledge for building an intelligent support system for the selection of HCC treatment.

In the treatment selection of HCC, 44 parameters are referred to, the names of some of which, the values they can take and the conventional designations are presented in Table 1. In the treatment selection of HCC, a database (working memory) of the intelligent system is generated to store the examination data and their current values.

Building a knowledge base of the intelligent support system for the selection of HCC treatment

Consequently, based on the possible values of 44 examination data, shown in Table 1, the knowledge (judgments) obtained from expert physicians regarding the selection of HCC treatment is transformed into rules, which are shown below.

Rule 1. IF $[(x_3=y_{31}) \& (x_6=y_{62}) \& (x_7=y_{72}) \& (x_{12}=y_{122}) \& (x_{13}=y_{132}) \& (x_{16}=y_{162}) \& (x_{17}=y_{172}) \& (x_{21}=y_{212}) \& (x_{22}=y_{222}) \& (x_{23}<3) \& (x_{24}<1{,}7) \& (x_{27}=y_{272}) \& (x_{28}=y_{282}) \& (x_{29}=y_{292}) \& (x_{30}=y_{302}) \& (x_{31}<35) \& (x_{33}=(A \vee B)) \& (x_{34}<9) \& ((x_{35} \geq 30) \vee (x_{36} \geq 40) \vee (x_{37} \geq 35)) \& (x_{40}=y_{402}) \& (x_{41}=y_{412}) \& (x_{42}=y_{422}) \& (x_{43}=(0 \vee 1 \vee 2))]$ **Then "Resection is permissible"**

Rule 2. **IF** $[(x_1 \leq 5) \& (x_2 \leq 3) \& (x_3=(y_{31} \vee y_{32})) \& (x_4=y_{42}) \& (x_5=y_{52}) \& (x_6=y_{62}) \& (x_7=y_{72}) \& (x_8=y_{82}) \& (x_9=y_{92}) \& (x_{10}=y_{102}) \& (x_{11}=y_{112}) \& (x_{12}=y_{122}) \& (x_{13}=y_{132}) \& (x_{14}=y_{142}) \& (x_{15}=y_{152}) \& (x_{16}=y_{162}) \& (x_{17}=y_{172}) \& (x_{19} \leq 200) \& (x_{26} \leq 100) \& (x_{31}<35) \& (x_{38}=y_{381}) \& (x_{39}=y_{392}) \& (x_{40}=y_{402}) \& (x_{41}=y_{412}) \& (x_{42}=y_{422}) \& (x_{43}=(0 \vee 1 \vee 2))]$ **Then "Transplantation is permissible"**

Rule 3. **IF** $[(x_1 \leq 3) \& (x_2 \leq 3) \& (x_3=(y_{31} \vee y_{32})) \& (x_4=y_{42}) \& (x_5=y_{52}) \& (x_6=y_{62}) \& (x_7=y_{72}) \& (x_8=y_{82}) \& (x_9=y_{92}) \& (x_{10}=y_{102}) \& (x_{11}=y_{112}) \& (x_{12}=y_{122}) \& (x_{13}=y_{132}) \& (x_{21}=y_{212}) \& (x_{22}=y_{222}) \& (x_{24} \leq 1.7) \& (x_{28}=y_{282}) \& (x_{29}=y_{292}) \& (x_{33}=(A \vee B)) \& (x_{40}=y_{402}) \& (x_{41}=y_{412}) \& (x_{43}=(0 \vee 1 \vee 2))]$ **Then "Ablation"**

Rule 4. **IF** $[(x_6=y_{62}) \& (x_7=y_{72}) \& (x_{21}=y_{212}) \& (x_{22}=y_{222}) \& (x_{23}<3) \& (x_{24} \leq 1.7) \& (x_{28}=y_{282}) \& (x_{29}=y_{292}) \& (x_{30}=y_{302}) \& ((x_{33}=(A \vee B)) \& (x_{40}=y_{402}) \& (x_{41}=y_{412}) \& (x_{42}=y_{422}) \& (x_{43}=(0 \vee 1 \vee 2))]$ **Then "Embolization".**

Rule 5. **IF** $[(x_{21}=y_{212}) \& (x_{22}=y_{222}) \& (x_{23}>3) \& (x_{24}<1.7) \& (x_{27}=y_{272}) \& (x_{28}=y_{282}) \& (x_{29}=y_{292}) \& (x_{30}=y_{302}) \& (x_{33}=A) \& (x_{40}=y_{402}) \& (x_{42}=y_{422}) \& (x_{43}=(0 \vee 1 \vee 2))]$ **Then "Systemic treatment".**

Forming the logical inference mechanism of the HCC diagnostic expert system

The inference engine of the system consists of 4 submodules: examination data processing module (EDPM), result formation module (RFM), scheme formation module (SFM) according to the HCC treatment method (stage), input data tabular display formation module (DDFM) according to the selected treatment method.

Based on the current input data, i.e., examination data, a new fact is formed in EDPM, and when the fact is checked for compliance with the antecedent (condition) part of the rules in KB and the compliance is confirmed, the rule is activated in RFM. Obtaining the scheme corresponding to the activated rule and transferring the result part of the rule to the interface block is performed by SFM, and transferring the values obtained by the examination data entered into the system to the interface block in the form of a table is performed by DDFM.

TABLE 1. EXAMINATION DATA, THE VALUES THEY CAN RECEIVE AND THEIR CONDITIONAL MARKING

No	Examination data	Marking	Possible values examination data may recieve	Makring the examination data values
1.	Size of tumor	x_1	Pcs	
2.	Number of tumors	x_2	Pcs	
3.	Location of tumor	x_3	In one lobe/ in two lobes/ Central bile duct neighboring, Exophyte grown subcapsular derivative	y_{31}/ y_{32}/ y_{33}
4.	Segmental portal vein invasion	x_4	Yes/no	y_{41}/ y_{42}
5.	Portal vein invasion of one lobe	x_5	Yes/no	y_{51}/ y_{52}
6.	Portal vein invasion of both lobes	x_6	Yes/no	y_{61}/ y_{62}
7.	Main portal vein invasion	x_7	Yes/no	y_{71}/ y_{72}
8.	Right hepatic vein invasion	x_8	Yes/no	y_{81}/ y_{82}
9.	Left hepatic vein invasion	x_9	Yes/no	y_{91}/ y_{92}
10.	Middle hepatic vein invasion	x_{10}	Yes/no	y_{101}/ y_{102}
11	Two hepatic veins invasion	x_{11}	Yes/no	y_{111}/ y_{112}
...	…	…	...	...
33.	Child	x_{33}	A/B/C	*A/B/C*
34.	MELD	x_{34}	Pcs	
35.	Residual volume after resection-normal Qc	x_{35}	Pcs	
36.	Residual volume after resection-cirrhotic Qc	x_{36}	Pcs	
37.	Residual volume after resection-fatty Qc	x_{37}	Pcs	
38.	Donor	x_{38}	Yes/no	y_{381}/ y_{382}
39.	Contraindications to immunosup-pression	x_{39}	Yes/no	y_{391}/ y_{32}

40.	Active infections	x_{40}	Yes/no	y_{401}/ y_{402}
41.	Active malignancies	x_{41}	Yes/no	y_{411}/ y_{412}
42.	Extrahepatic organ failure (uncorrectable)	x_{42}	Yes/no	y_{421}/ y_{422}
43.	Performance (ECOG)	x_{43}	Pcs	
44.	Weakness (minute walk distance - m)	x_{44}	Pcs	

Interface block of the intelligent support system for the HCC treatment selection

The interface block is intended to ensure communication between the user-physician and the expert system. This block performs the functions of controlling the screen, organizing a dialogue with the expert system, entering initial data (examination data values) into the system and delivering the results to the user.

System implementation

The intelligent system for the HCC treatment selection is developed in the object-oriented programming language C#. C# is used to develop a wide range of applications, including desktop and web applications, mobile applications, Unity-based games, cloud applications on Microsoft Azure, and database applications. Figure 1 illustrates the block diagram of the treatment selection decision-making algorithm in the system.

After logging into the system, the window illustrated in figure 2 opens in front of the user-physician. This interface window consists of a large number of sub-windows, each of which corresponds to the examination data of HCC and the values they can receive.

Each of these windows is active and enables the user-physician to enter the current examination data. The current values of the examination data are entered into the system from the opened interface window and the "Initial result" button is pressed. The current values of the examination data entered into the system are recorded in the DB. These data also form the EDPM in the ERM. The latter is compared with the condition part of the rules in the KB, and when an overlap occurs, the result (HCC treatment method) part of that rule is recorded in the ERM of the RFM.

Based on the rule corresponding to the formed result, the selected treatment method of HCC in the ERM (for example, "Resection is permittable") is transferred to the interface window via the SFM. Then, the table formed in the DDFM based on the entered data and their values is transferred to the interface. Consequently, the result on the selected treatment method of HCC registered in the ERM ("Resection", "Transplantation", "Ablation", "Embolization", "Systemic treatment", "Symptomatic treatment") and the entered data and their values are presented to the user via the interface block (fig. 3).

The choice of the treatment method appropriate to the situation shown in Figure 2 corresponds to the decision option "Resection". This decision is made as a result of the activation of Rule 1 in the KB of the intelligent support system for the HCC treatment selection.

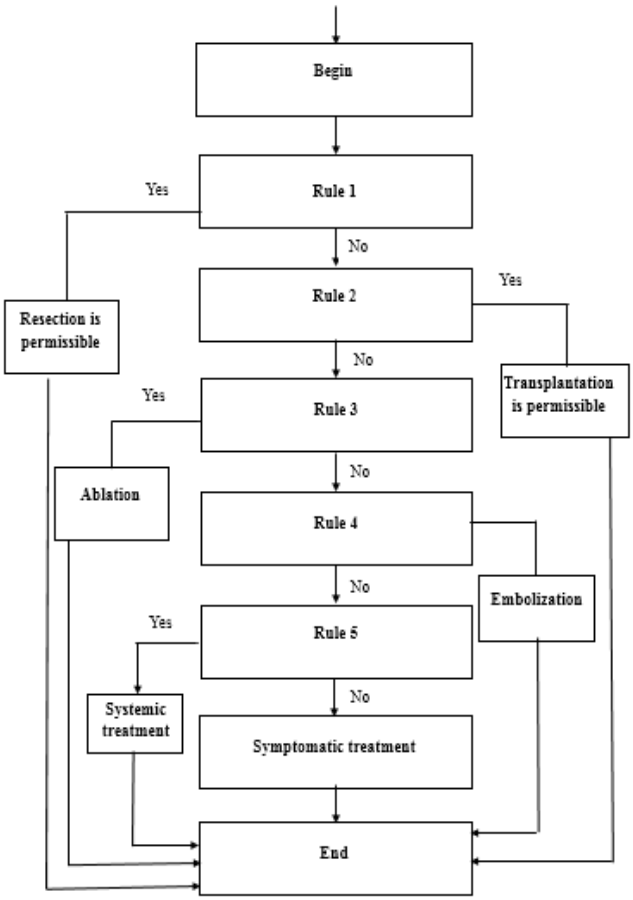


Fig 1. HCC treatment selection decision-making algorithm in the intelligent system

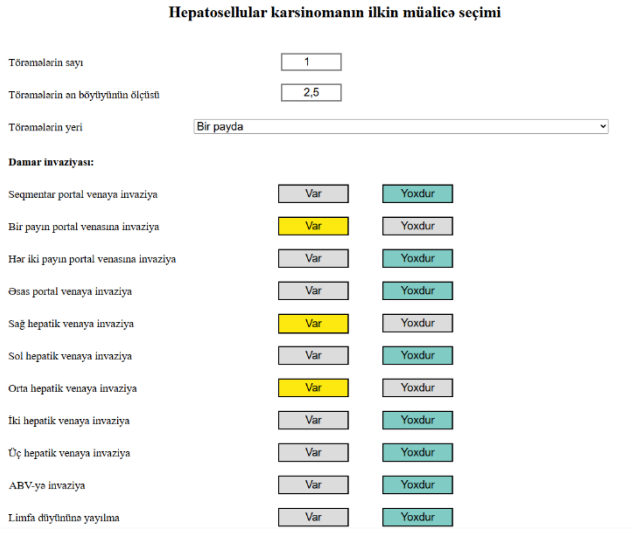


Fig. 2. Entering the examination data values for the HCC treatment selection into the system from the opened interface window

CONCLUSION

The intelligent support system for the HCC treatment selection was developed to support physician decisions. The system, consisting of architectural components, as a database, a knowledge base, an inference mechanism and an interface block, was implemented in the C# programming language on the Visual Study 2019 platform.

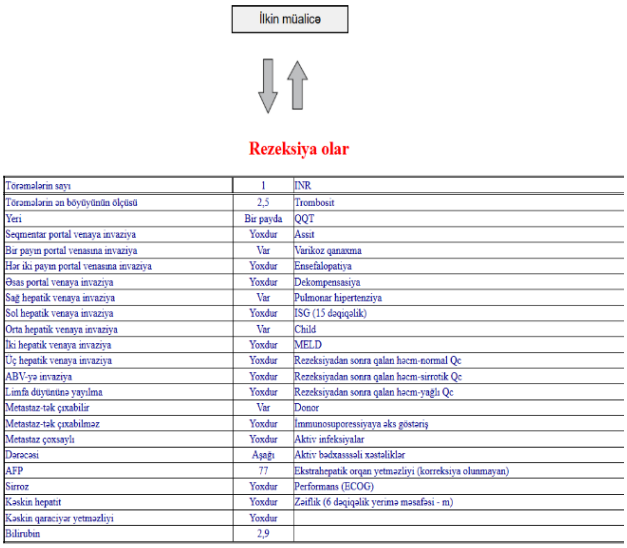


Fig. 3. Interface window for delivering the result to the user in the intelligent system for the HCC treatment selection

During the tests of the intelligent system supporting the treatment selection of HCC with the involvement of physician-specialists of the Department of Surgical Diseases of the Azerbaijan Medical University, it was confirmed that it was suitable for use as an expert physician decision support system in a multidisciplinary concilium for individual treatment selection.

REFERENCES

[1] R. R. Arich, F. Alessandro, T. F David, S. Eizaburo, A. B. Amir, T. H. Matthew, E. T. Mitchell, "Infiltrative hepatocellular carcinoma: what radiologists need to know", Radiographics, 35(2), 371-386 (2015). doi: 10.1148/rg.352140114.

[2] N. Y. Bayramov, "Surgical diseases of the liver", 2012, in press. Baku: Qismet.

[3] D. Q. Huang, A. Tran, E. X. Tan, S. N. Nerurkar, R. Teh, M. L. P Teng, E. J. Yeo, B. Zou, C. Wong, C. O. Esquivel, C. A. Bonham, M. H. Nguyen, "Characteristics and outcomes of hepatocellular carcinoma patients with macrovascular invasion following surgical resection: a meta-analysis of 40 studies and 8,218 patients", Hepatobiliary Surg Nutr, 2022, vol.11, no. 6, pp. 848-860.

[4] S. Ray, N. N. Mehta, A. Golhar, S. Nundy, "Post hepatectomy liver failure – A comprehensive review of current concepts and controversies", Annals of Medicine and Surgery, 2018, no. 34, pp.4-10. doi:10.1016/j.amsu.2018.08.012

[5] F. Calderon, E. Masino, L. Caram, V. Ardiles, "Short-and-Long-Term Outcomes of Hepatic Resection for Hepatocellular Carcinoma in Cirrhotic and Non-Cirrhotic Liver Parenchyma", Journal of Surgery and Clinical Reports, 2021, May.

[6] D. Citterio, A. Facciorusso, C. Sposito, R Rota, S. Bhoori, V. Mazzaferro, "Hierarchic Interaction of Factors Associated With Liver Decompensation After Resection for Hepatocellular Carcinoma", JAMA Surgery, 2016, Vol. 151, no. 9. doi:10.1001/jamasurg.2016.1121

[7] S. Park, A. M. Davis, A. A. Pillai, "Prevention, Diagnosis, and Treatment of Hepatocellular Carcinoma", JAMA, Published online August 29, 2024. doi:10.1001/jama.2024.14101

[8] Z.J. Brown, D. I. Tsilimigras, S. M. Ruff, et al.,"Management of Hepatocellular Carcinoma: A Review", JAMA surgery, 2023, vol. 158, no. 4, pp. 410-420. doi:10.1001/jamasurg.2022.7989

[9] D. J. H. Tan, C. WongC, C. H. Ng, et al., "A Meta-Analysis on the Rate of Hepatocellular Carcinoma Recurrence after Liver Transplant and Associations to Etiology, Alpha-Fetoprotein, Income and Ethnicity", Journal of Clinical Medicine, 2021, vol 10. doi:10.3390/jcm10020238

[10] Российский патент 2021 года по МПК G16H10/60 G16H50/20 «Система для поддержки принятия врачебных решений» RU2752792C1.

[11] M. Mirmozaffari, "Developing an expert system for diagnosing liver diseases", European Journal of Engineering Research and Science, vol. 4, no. 3, March 2019. https://www.ej-eng.org/index.php/ejeng/article/view/1168/462

[12] S. Aman, P. Babita, "An Efficient Diagnosis System for Detection of Liver Disease Using a Novel Integrated Method Based on Principal Component Analysis and K-Nearest Neighbor (PCA-KNN)", International Journal of Healthcare Information Systems and Informatics, 2016, vol. 11, no. 4, pp. 56–61.

[13] J. S. Sartakhti, M. H. Zangooei, K. Mozafari, "Hepatitis disease diagnosis using a novel hybrid method based on support vector machine and simulated annealing (SVM-SA)", Computer Methods and Programs in Biomedicine, 2015, vol. 108, no. 2, pp. 570–579.

[14] D. N. Li, C. K. Chui, S. Chang, S. H. Ong, "A new unified level set method for semi-automatic liver tumor segmentation on contrast-enhanced CT images", Expert Systems with Applications, 2012, vol.39, no.10, pp.9661–9668.

[15] L. K. Ming, L. C. Kiong, L. W. Soong, "Autonomous and deterministic supervised fuzzy clustering with data imputation capabilities. Applied Soft Computing, 2011, vol.11, no.1, pp.1117–1125.

[16] S. Kulluk, L. Ozbakır, A. Baykasoğlu, "Fuzzy DIFACONN-miner: A novel approach for fuzzy rule extraction from neural networks", Expert Systems with Applications, 2013, vol. 40, no.3, pp.938–946.

[17] D. C. Li, C. W. Liu, S. C. Hu, "A learning method for the class imbalance problem with medical data sets", Computers in Biology and Medicine, 2010, vol.40, no.5, pp.509–518.

[18] E. Mezyk, O. Unold, "Mining fuzzy rules using an Artificial Immune System with fuzzy partition learning", Applied Soft Computing, 2011, vol.11, no.2, pp.1965–1974.

[19] F. Gorunescu, S. Belciug., M. Gorunescu, R. Badea, "Intelligent decision-making for liver fibrosis stadialization based on tandem feature selection and evolutionary-driven neural network", Expert Systems with Applications, 2012, vol.39, no.17, pp.12824–12832.

[20] Y. Torun, G. Tohumoglu, "Designing simulated annealing and subtractive clustering based fuzzy classifier", Applied Soft Computing, 2011, vol.11, no.2, pp.2193–2201.

[21] M. Aldape-Perez, C. Yanez-Marquez, O. J. Camacho-Nieto, A. Arguelles-Cruz, "An associative memory approach to medical decision support systems", Computer Methods and Programs in Biomedicine, 2012, vol.106, pp.287–307. doi:10.1016/j.cmpb.2011.05.002 PMID:21703713

[22] M. H. Mammadova, Z. G. Jabrayilova, L. A. Garayeva, A. I. Ahmadova, "Prediction of hepatocellular carcinoma using a machine learning", The 16th IEEE International Conference Application of Information and Communication Technologies (AICT-2022), Washington DC, 12-14 Oct 2022, INSPEC Accession Number: 22541899, DOI: 10.1109/AICT55583.2022.10013575https://ieeexplore.ieee.org/document/10013575

[23] M. H. Mammadova, N. Y. Bayramov, Z. G. Jabrayilova, "Development of the principles of fuzzy rule-based system for hepatocelular carcinoma staging", EUREKA: Physics and Engineering, 2021, no. 3, pp. 3–13, doi:https://doi.org/10.21303/2461-4262.2021.001829

[24] M. H. Mammadova, N. Y. Bayramov, Z. G. Jabrayilova, M. I. Manafli, M. R. Huseynova, "Knowledge transformation in the intelligent system for hepatocellular carcinoma staging" Proceedings of the 8th International Conference on Control and Optimization with Industrial Applications (COIA'2022), 24-26 August 2022, Baku, Azerbaijan, vol.1, pp. 318-321.

[25] M. H. Mammadova, N. Y. Bayramov, Z. G. Jabrayilova, M. I. Manafli, M. R. Huseynova, "Decision support system for the hepatocellular carcinoma diagnosis", Azerbaijan Medical Journal, 2024, No. 4, pp. 163–169.

Enhancing Solar Photovoltaic Power Systems with AI-Based Maximum Power Point Tracking

Omar Imamverdiyev
School of Information Technology and Engineering
ADA University
Baku, Azerbaijan
oimamverdiyev15629@ada.edu.az

Burcu Ramazanli
School of Information Technology and Engineering
ADA University
Baku, Azerbaijan
bramazanli@ada.edu.az

Surkhay Fatullayev
School of Information Technology and Engineering
ADA University
Baku, Azerbaijan
sfatullayev15530@ada.edu.az

***Abstract*— Photovoltaic (PV) power systems are increasingly deployed to meet sustainable energy needs but suffer efficiency losses under variable irradiance and temperature conditions. To address this, Maximum Power Point Tracking (MPPT) algorithms dynamically adjust the converter's duty cycle to maximize power extraction. This study presents a comparative analysis of four advanced MPPT techniques, including Perturb and Observe (P&O), Fuzzy Logic Control (FLC), Particle Swarm Optimization (PSO), and Artificial Neural Networks (ANN), all implemented within a unified MATLAB/Simulink framework. Each algorithm is tested using real-world weather data that captures typical variations in solar irradiance and temperature and accelerated profiles that simulate rapid and abrupt changes in irradiance and photovoltaic cell temperature. Performance metrics include tracking accuracy, dynamic response, steady-state stability, total energy harvested, and overall efficiency. The findings illustrate the trade-offs among these approaches, highlighting their advantages and disadvantages regarding response time, overall efficiency, and energy output. This thorough examination offers comparative findings to identify the suitable MPPT strategy for particular environmental circumstances.**

Keywords—Solar energy, Maximum Power Point, Photovoltaic panels, P&O, FLC, PSO, ANN

INTRODUCTION

The growing demand for sustainable and low-carbon energy solutions has significantly increased the role of renewable energy sources in global electricity generation [1]. Among these, solar energy has demonstrated the fastest growth rate due to rapid advances in photovoltaic (PV) technology and the expansion of solar infrastructure worldwide. However, the power output of PV systems is inherently variable, as it depends directly on external environmental conditions such as solar irradiance and ambient temperature [2]. Fluctuations in sunlight intensity and temperature affect the current–voltage (I–V) characteristics of PV modules, causing continuous shifts in their maximum power point (MPP). Without an appropriate control mechanism, PV systems operating at fixed voltage and current setpoints fail to track these variations effectively, resulting in considerable energy losses over the system's lifetime. For example, sudden increases in irradiance during the day may lead to underutilization of available solar energy, while temperature variations influence cell conversion efficiency if not compensated for in real time. This situation poses a practical engineering challenge to ensure that PV systems continuously operate at or near the true MPP under real-world, dynamic conditions. To address this, Maximum Power Point Tracking (MPPT) algorithms are widely used to adjust system parameters dynamically, maximizing energy extraction and improving overall system efficiency. Conventional methods, such as Perturb and Observe (P&O), provide a simple and low-cost solution but are limited by slow convergence and oscillations under rapidly changing conditions. In response, modern MPPT strategies increasingly incorporate advanced control techniques, including fuzzy logic controllers (FLC), optimization algorithms such as Particle Swarm Optimization (PSO), and artificial neural networks (ANN), to enhance tracking accuracy and system stability. This paper investigates and compares the performance of several MPPT methods, including P&O, FLC, PSO, and ANN-based control, under realistic environmental conditions, with a focus on improving tracking effectiveness and operational reliability.

LITERATURE REVIEW

First, Recent advancements in AI-based Maximum Power Point Tracking (MPPT) algorithms have significantly improved the efficiency and adaptability of photovoltaic (PV) systems under dynamic environmental conditions. Traditional methods such as Perturb and Observe (P&O) and Incremental Conductance (INC) often suffer from slow response and high oscillations, particularly under partial shading or rapidly changing irradiance.

Gundogdu & Celikel [3] explored AI-driven MPPT designs, with particular emphasis on feedforward neural networks (FNN) using voltage, current, and temperature as inputs. Their results demonstrated faster convergence, minimal oscillations, and improved tracking performance compared to conventional methods. Similarly, Yilmaz et al. [4], along with M. A. Sasi [5], highlighted the robustness and effectiveness of Fuzzy Logic Control (FLC) in handling nonlinearities and dynamic conditions without requiring a precise mathematical model, and also noted its slow convergence under low irradiance conditions.

PSO-based MPPT, as discussed by Touti [6], treats the tracking task as an optimization problem, where particles represent candidate solutions. This method proved effective in navigating complex search spaces but is sensitive to parameter tuning and may converge slowly.

Hardware considerations were addressed by Hashim et al. [7], who emphasized the importance of boost converter design and its integration with MPPT algorithms. They showed that tuning inductor and capacitor values can reduce convergence time and improve system responsiveness.

Collectively, the findings indicate that AI-based MPPT methods, including FNN, PSO, and FLC, provide higher efficiency, stability, and adaptability than conventional

approaches, but they still have some limitations and practical challenges.

In order to improve the effectiveness of MPPT algorithms, this study introduces several modifications aimed at enhancing tracking performance and system stability. Specifically, the proposed P&O algorithm incorporates practical operating constraints and a voltage threshold to ensure safe and stable operation. The FLC design employs direct environmental inputs (irradiance and temperature) with clearly defined membership functions and rule bases for better reproducibility and responsiveness. For the PSO approach, an adaptive search space scaling mechanism dynamically adjusts particle limits according to real-time power levels, increasing tracking efficiency under variable irradiance. Finally, the ANN-based MPPT is trained on real NASA POWER data with three inputs, irradiance, temperature and current, and integrates a Ziegler–Nichols-tuned PI controller to generate a stable duty cycle for the boost converter. All proposed algorithms are validated using realistic, real-world environmental datasets to demonstrate robust performance under practical operating conditions.

MPPT ALGORITHMS

Perturb and Observe (P&O)

The P&O algorithm is one of the conventional methods in finding MPP due to its simplicity. The algorithm functions by perturbing the voltage and observing the effect on output power. In this work, a modified version of the classical P&O algorithm is used. The duty cycle is initialized at 0.1 and bounded between 0 and 0.85 to ensure safe operation of the boost converter. At each iteration, the algorithm calculates the change in voltage (ΔV) and change in power (ΔP) between the current and previous states. The core logic is based on the sign of these differences:

- If $\Delta P \neq 0$ and the PV voltage exceeds 30 V, the algorithm decides whether to increase or decrease the duty cycle based on the relationship between ΔP and ΔV.
- If ΔP = 0 or the PV voltage is below 30 V, the duty cycle remains unchanged.
- The output duty cycle is then constrained within user-defined limits to prevent over-driving the converter.

This approach effectively searches for the point where the derivative (ΔP/ΔV) is equal to 0, indicating the MPP.

However, due to its perturbative nature, P&O inherently introduces oscillations around the MPP, particularly under steady-state conditions. It may also exhibit performance degradation rapidly

Particle Swarm Optimization (PSO)

Another common MPPT technique is the PSO algorithm, inspired by the social behavior of birds flocking. This technique uses particles to explore the search space and adjust their positions based on personal experience and the global best solution found by the swarm. In this work, each particle represents a candidate duty cycle value, and the swarm refers to the collection of these particles exploring the solution space to track the MPP.

The PSO algorithm begins by initializing 10 particles with uniformly spaced duty cycle values within the range [0.1,0.5]. This step ensures a diverse starting point, allowing the algorithm to explore different possibilities in the search for the optimal solution.

To minimize ripple and measurement noise, the PV output power is filtered using a low-pass filter with a coefficient (α) of 0.9. The filtered power is then normalized and used as the fitness function to evaluate each particle's effectiveness.

Each particle retains the position that gave it the highest power as its personal best, while the global best is updated to reflect the highest-performing particle in the entire swarm. This ensures that the swarm is guided by both individual experience and collective intelligence. The particle velocities and positions are updated iteratively using (1) and (2) [6].

$$v_i^{k+1} = \omega v_i^k + c_1 r_1 \left(P_{best,i}^k - d_i^k\right) + c_2 r_2 \left(G_{best,i} - d_i^k\right) \#(1)$$

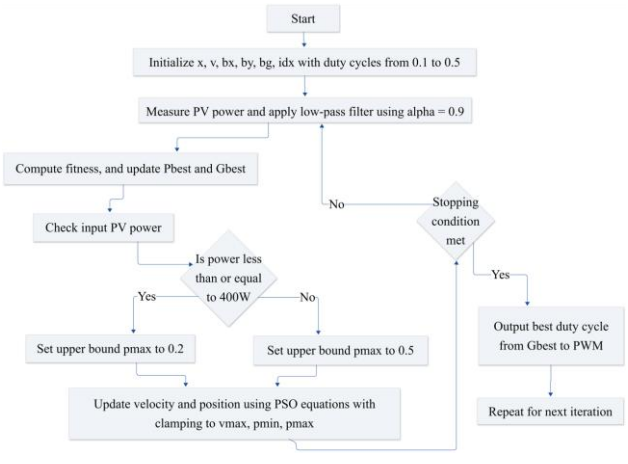
$$x_i^{k+1} = x_i^k + v_i^{k+1} \#(2)$$

Here, ω represents the inertia weight, c_1 and c_2 are acceleration coefficients, and the velocity is limited by v_{max} to ensure stability. Using these equations, particles update their positions based on a balance of past experience and the best-known global solution.

Velocity and position are constrained within set limits to maintain algorithm stability and avoid divergence. The resulting optimal duty cycle is then passed to a PWM generator to control the boost converter, ensuring the PV system operates near the MPP.

In contrast to other studies, in this work adaptive search space scaling based on input conditions was added to enhance mean efficiency of the algorithm. Depending on the power input value, the upper bound of the particle search space (p_{max}) is adaptively set to either 0.2 or 0.5 to reflect low or high power conditions, respectively.

The complete process of the proposed PSO-based MPPT algorithm, including the adaptive search space adjustment, is illustrated in Fig. 1.



- Fig.1 Flowchart of the proposed PSO algorithm

Fuzzy Logic Controller

The FLC has proven to be a highly adaptive approach for MPPT in PV systems [5], [4]. A fuzzy inference system includes three fundamental stages that are acknowledged as fuzzification, rule evaluation, and defuzzification.

In the fuzzification stage, the algorithm processes real-time input signals and converts numerical values into fuzzy sets using predefined membership functions [4]. In this study, the tracking error (E) and its rate of change (ΔE) were selected as the input signals to the fuzzy algorithm, while the output of the FLC is the duty cycle adjustment (ΔD) used to keep the PV system operating near its maximum power point. The tracking error and its rate of change are calculated according to (3) and (4), respectively.

$$E(k) = \frac{P_{in}(k) - P_{in}(k-1)}{V_{in}(k) - V_{in}(k-1)} \#(3)$$

$$\Delta E(k) = E(k) - E(k-1) \#(4)$$

The input space used in this study is defined as (5).

$$X = \{ -8.5 \le E \le 8.5,\ -1.6 \le \Delta E \le 1.6 \} \quad (5)$$

A membership function is used to map each point in the input space to a degree of membership, i.e., real number, between 0 and 1 [5]. The inputs are characterized using five triangular membership functions (MFs), which are represented in Table I.

- THE MF OF E, ΔE, AND ΔD

MF Name	Range Information	
	Range of ΔE and ΔD	***Range of E***
NB	[–2.4, –1.6, –0.8]	[–12.75, –8.5, –4.25]
NS	[–1.6, –0.8, 0]	[–8.5, –4.25, 0]
ZE	[–0.8, 0, 0.8]	[–4.25, 0, 4.25]
PS	[0, 0.8, 1.6]	[0, 4.25, 8.5]
PB	[0.8, 1.6, 2.4]	[4.25, 8.5, 12.75]

Following this, the system proceeds to the inference engine stage, where fuzzy rules are applied to the fuzzified inputs. Each rule is written in the "IF-THEN" structure and forms the decision-making part of the algorithm [5]. A total of 25 rules were assigned in this research, and their relationships are displayed in Table II.

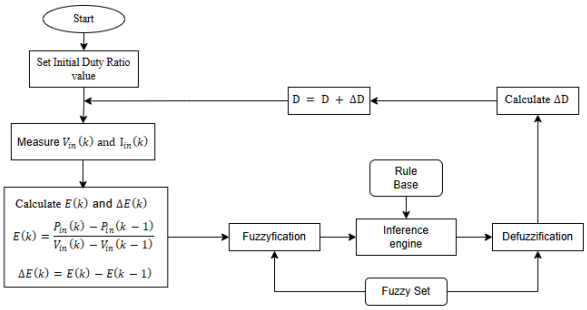
ASSIGNED RULES FOR FLC

Error (E)	Change in Error (ΔE)				
	NB	***NS***	***ZE***	***PS***	***PB***
NB	PS	PB	NB	NB	NS
NS	PS	PS	NS	NS	NS
ZE	ZE	ZE	ZE	ZE	ZE
PS	NS	NS	PS	PS	PS
PB	NS	NB	PB	PB	PS

The final step involves defuzzification, where the resulting fuzzy set is transformed back into numerical output, which is achieved through mapping [4]. The mapping method was chosen to be centroid, due to its simplicity in terms of computation [5].

After defuzzification, the fuzzy logic controller produces ΔD, which represents small adjustments needed to be applied to the duty cycle. A small adjustment ΔD keeps the system smooth and stable [8]. Fig. 2 shows the flowchart of the proposed FLC system.

Another approach involves implementing ANN, which was inspired by biological neural architecture. In this study, the MPPT system, is built around a Feed-Forward Artificial Neural Network (FNN) integrated into a boost converter-based photovoltaic power system.



Flowchart of the FLC system

Artificial Neural Network

To train the ANN, two distinct datasets were used. The first dataset consisted of real-world irradiance and temperature values, collected from the NASA POWER database for various regions of Azerbaijan. These environmental parameters were then fed into a PV array model, which generated the corresponding voltage (V_{pv}) and current (I_{pv}) outputs. The output values were then merged as one dataset with corresponding environmental parameters, yielding 10000 data samples, covering irradiance from 200 to 1000 W/m² and temperature from 10°C to 50°C.

When studies in literature were examined, it has been observed that either voltage and temperature or irradiance values are used as inputs ANN-based MPPT algorithms [3]. In this work, the architecture of the proposed FNN consists of:

Input Layer: 3 neurons (Irradiance, Temperature, Current)

Hidden Layer: 10 neurons

Output Layer: 1 neuron (Voltage)

The network uses Levenberg–Marquardt algorithm for training. The data is normalized using the "mapminmax" function to the range [−1,1]. The dataset is divided into 70% training, 15% validation, and 15% testing sets to ensure generalization and avoid overfitting.

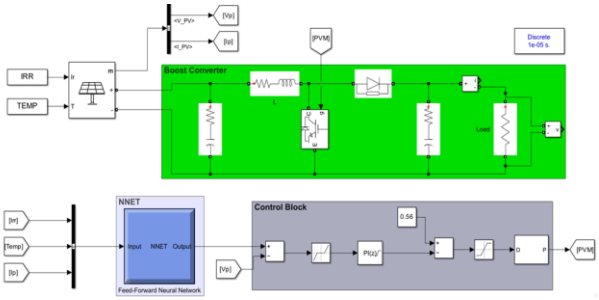
The network produces the reference voltage (V_{ref}), which corresponds to the predicted voltage at the MPP. This reference voltage is compared to the actual PV voltage (V_{pv}), computing error value, as shown in (6).

$$ERROR = V_{ref} - V_{pv} \quad (6)$$

The resulting error signal is then passed to the discrete PI controller with proportional gain of 0.1 and integral gain of 0.0025. The given parameters for the PI controller were found by the Ziegler–Nichols Method [9]. The output of the PI controller is passed on to the saturation block, ensuring it stays within valid bounds, i.e., between 0.1 and 0.9. The defined duty cycle is fed into the PWM generator, operating at 20 kHz with 50 μs sample time.

The output of the PWM generator is sent to the gate terminal of the power switch (IGBT), to determine how long

the switch stays ON during each cycle. Fig. 3 displays simulation of the FNN in PV array system.



PV array system with ANN

SIMULATION MODEL

To evaluate the performance of the MPPT algorithms (P&O, PSO, FLC, and ANN), a complete photovoltaic system was modeled and simulated in MATLAB/Simulink. The system includes a PV array, a DC-DC boost converter, and the MPPT controller, forming a closed-loop control architecture.

PV Array

The PV array is modeled using the single-diode equivalent circuit, also known as the five-parameter model [10]. The array consists of two modules connected in series and two strings connected in parallel ($N_s = 2$, $N_p = 2$), with each module rated at $V_{mp} = 29.94$ V and $I_{mp} = 7.18$ A. The model generates V_{pv} and I_{pv} under varying irradiance levels (200–1000 W/m²) and ambient temperatures (10–50 °C).

Boost Converter

A DC-DC boost converter is used to raise the PV voltage and ensure MPP tracking. The converter is controlled via a Pulse Width Modulation (PWM) signal generated from the output of the MPPT algorithms. The switching frequency is set to 20 kHz. The converter consists of a 4 mH inductor, a 470 µF input and output capacitor, and a 20 Ω load resistor.

Test Conditions

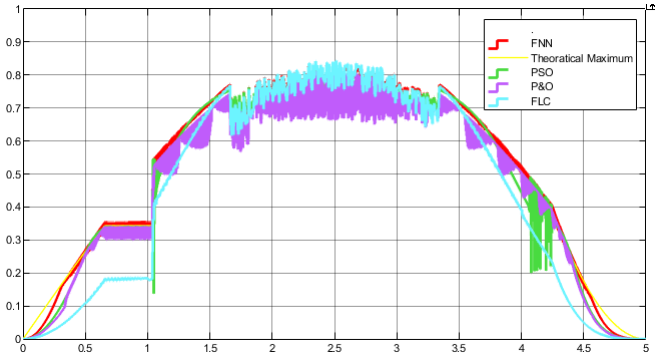
To evaluate the performance and robustness of the MPPT algorithms, two types of irradiance and temperature profiles were used during simulation:

1. Synthetic Random Profile: Irradiance values were randomly generated within the range of 400 to 1000 W/m², and temperature values were varied from 15°C to 50°C.
2. Real-World Inspired Profile: The second dataset was derived from NASA POWER irradiance and temperature data for regions in Azerbaijan [11]. Daily profiles were collected and analyzed, then merged to create a realistic irradiance-temperature curve. The time axis was normalized to a 5-second simulation window, representing a full 24-hour day. The final dataset is publicly accessible for reference [12].

RESULTS

Comparing Algorithms

Firstly, the performance of all four MPPT algorithms was evaluated using a real-world irradiance and temperature profile derived from NASA POWER data [11], [12]. The results are illustrated in Fig. 4.



FNN, PSO, FLC, P&O on real-world inspired dataset

Among the tested methods, the FNN-based MPPT demonstrated the best overall performance in terms of both accuracy and stability. It achieved convergence to the MPP by t=0.2s and maintained a consistently smooth power output for the rest of the simulation. The second-best performance was achieved by the PSO algorithm. Although PSO converged more slowly than FNN, t=0.5s, it exhibited fewer oscillations, particularly in the interval between t=0.5s and t=1.0s. However, the difference in oscillation behavior outside this interval was negligible. The FLC ranked third. It converged to MPP more slowly, around t=1.3s, but once stabilized, the output remained steady with minimal oscillations. The P&O method exhibited the weakest performance, despite reaching the MPP as early as PSO at t=0.5s. It suffered from high-frequency oscillations, particularly near the peak power region, which compromises its effectiveness in real-world applications with fluctuating irradiance. The average tracking efficiency observed for each method was as follows:

FNN: 99.1%

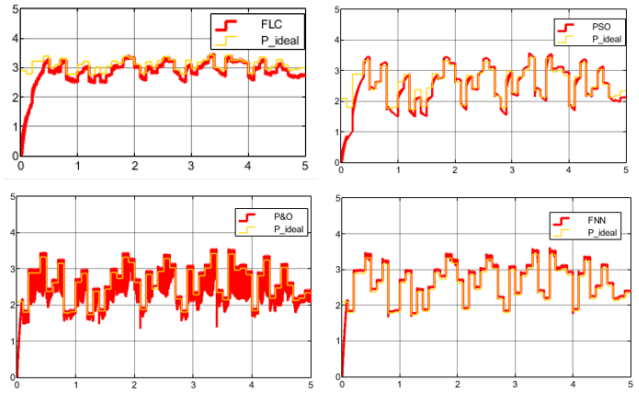
PSO: 98.1%

FLC: 93.1%

P&O: 89.6%

Afterwards, the four MPPT algorithms were also evaluated under randomly generated irradiance (400–1000 W/m²) and temperature (15–50 °C) conditions to evaluate their dynamic response and robustness. The results of this test are presented on Fig. 5.

The FNN-based MPPT again demonstrated superior performance, achieving fast convergence to the MPP at t=0.1s and maintaining a low oscillation frequency throughout the simulation. It also responded swiftly to abrupt changes in environmental conditions, confirming its adaptability and stability. The PSO algorithm reached the MPP at approximately t=0.5s and exhibited low oscillation, but showed slower adaptation to sudden changes in irradiance, indicating some vulnerability in highly dynamic conditions. The FLC method also converged to the MPP around t=0.5s, similar to PSO, but its reaction to environmental variations was less effective, showing delayed and less stable behavior during transitions. The P&O algorithm showed the poorest overall performance due to its high oscillation amplitude, especially during steady-state periods. However, it did achieve a relatively fast convergence at t=0.2s and a moderate response to sudden changes, making it slightly more reactive but less stable.



FLC, PSO, P&O, FNN on random environmental parameters

The average tracking efficiencies recorded during this test were as follows:

- FNN: 99.5%
- PSO: 96.9%
- FLC: 95.1%
- P&O: 90.1%

Validation

Multiple sources have been analyzed to compare the efficiencies of the proposed algorithms. According to several studies, the P&O algorithm achieved efficiency between 76.5% and 85.8% under varying operating conditions [13], [14]. The FLC technique has demonstrated efficiencies ranging from 94% to 95% in some researches [4], [8]. Compared to the conventional P&O algorithm, PSO achieved a higher efficiency range of 96-97.6%, as shown in the study [14]. According to several literature, the FNN showed the highest efficiency results, ranging from 97% to 98.5% [3], [15], [16]. Table III shows the efficiency results obtained from review papers.

EFFICIENCY DATA FROM REVIEW PAPERS

Algorithms	**Efficiency (%)**	
	Other studies	***In this work***
P&O	76.5-85.8%	89.6-90.1%
FLC	94-95%	93.1-95.1%
PSO	96-97.6%	96.9-98.1%
FNN	97-98.5%	99.1%-99.5%

CONCLUSION

This study presented a comparative analysis of four MPPT algorithms, P&O, FLC, PSO, and FNN with the goal of enhancing their performance relative to existing literature. All algorithms were implemented under the same simulation framework and tested using both randomly generated and real-world environmental data. The results demonstrated that the FNN-based MPPT approach achieved the highest performance, with an efficiency range of 99.1% to 99.5%, exceeding values reported in previous studies. PSO and FLC exhibited stable and consistent behavior, achieving efficiency levels in line with the literature, between 96.9%–98.1% and 93.1%–95.1% respectively. The P&O algorithm, while offering fast convergence, showed significant oscillations and lower overall efficiency, ranging from 89.6% to 90.1%, as expected. In conclusion, while PSO and FLC maintained their established performance, the FNN model showed a notable improvement over prior implementation. This highlights the potential of neural network–based MPPT strategies for achieving high tracking efficiency and robust performance under dynamic conditions.

REFERENCES

[1] M. Wiatros-Motyka, N. Fulghum, & D. Jones, "Global Electricity Review," Ember, 2024.

[2] Z. Ishrat, A. K. Gupta, S. Nayak, "A comprehensive review of MPPT techniques based on ML applicable for maximum power in solar power systems," jree.ir, Apr. 14. 2023. [Online]. Available: https://www.jree.ir/article_171383_e47e94299dfe49b3aa6a9c794d0a58d8.pdf

[3] Gundogdu, and R. Celikel, "ANN-Based MPPT Algorithm for Photovoltaic Systems," dergipark.org.tr, Aug. 23, 2020, [Online]. Available: https://dergipark.org.tr/en/download/article-file/1192226

[4] U. Yilmaz, A. Kirchay, and S. Borekchi, "PV system fuzzy logic MPPT method and PI control as a charge controller," sciencedirect.com,Aug.31,2017.[Online].Available: https://doi.org/10.1016/j.rser.2017.08.048

[5] M. A. Sasi, "Fuzzy Logic Control of MPPT Controller for PV Systems," core.ac.uk,May.2017.[Online].Available:https://core.ac.uk/download/pdf/154914752.pdf.

[6] E. Touti, "Reactive power analysis and frequency control of autonomous wind induction generator using particle swarm optimization and fuzzy logic," Energy Exploration & Exploitation, vol. 38, no. 3, pp. 755–782, 2020. [Online]. Available: https://www.jstor.org/stable/26966856

[7] H. M. H. Farh, and A. M. Eltamaly, "Maximum Power Extraction from the Photovoltaic System Under Partial Shading Conditions," link.springer.com, Jul. 31, 2019. [Online]. Available: https://doi.org/10.1007/978-3-030-05578-3_4

[8] F. A. Omar, "Comparative Performance Analysis Of A Feed-Forward Neural Network-Based Mppt For Rapidly Changing Climatic Conditions" dergipark.org.tr, Nov. 11, 2022. [Online]. Available: https://doi.org/10.36306/konjes.1179030.

[9] G. Ellis, "Chapter 6 - Four Types of Controllers," in Control System Design Guide (Fourth Edition), sciencedirect.com, May. 11, 2012. [Online].Available:https://doi.org/10.1016/B978-0-12-385920-4.00006-0.

[10] E. Tayyan, "A simple method to extract the parameters of the single-diode model of a PV system," researchgate.net, Mar. 2013. [Online].Available: https://doi.org/10.3906/fiz-1206-4

[11] NASA POWER Project, "Daily solar irradiance and temperature data for Azerbaijan," NASA Langley Research Center, Hampton, VA, USA. [Online]. Available: https://power.larc.nasa.gov/

[12] O. Imamverdiyev, Time, Irradiance, Temperature parameter dataset for MPPT, GitHub, 2024. [Online]. Available: https://github.com/OmarImamverdiyev/Dataset-for-MPPT-/blob/main/Time%2C%20Irradiance%2CTemperature%20parameter%20.xlsx

[13] Ch. Sai Babu and J. Surya Kumari, "Design and investigation of improved perturb & observe MPPT algorithm for photovoltaic systems," *J. Electr. Eng.*, vol. 14, no. 2, pp. 1–8, Jun. 2014. [Online]. Available: https://www.researchgate.net/publication/271199594

[14] M. S. Endiz, "Performance Analysis Of P&O And PSO Mppt Algorithms For PV Systems Under Partial Shading," researchgate.net, Mar.2024.[Online].Available:https://doi.org/10.36306/konjes.1359177.

[15] R. Salviati, Particle Swarm Optimization in MPPT Algorithms for Photovoltaic Systems, thesis, Univ. of Padova, Italy, 2020. [Online]. Available: https://thesis.unipd.it/retrieve/9ddf2c62-1d9c-4648-9630-ed3edcc08c1c/Salviati_Riccardo_1183982.pdf

[16] M. T. Makhloufi and Y. Abdessemed, "A feed-forward neural network MPPT control strategy applied to a modified Cuk converter," *Int. J. Electr. Comput. Eng.*, vol. 6, no. 4, pp. 1421–1433, Aug. 2016. doi: 10.11591/ijece.v6i4.9704.

Clustering of Countries Based on Multi-Domain Indicators for Sustainable Development

Nigar Ismayilova
Department of General and Applied Mathematics
Azerbaijan State Oil and Indsutry University
Baku, Azerbaijan
nigar.ismailova@asoiu.edu.az

Ibrahim Muradov
Department of General and Applied Mathematics
Azerbaijan State Oil and Indsutry University
Baku, Azerbaijan
ibrahim.muradv@gmail.com

Nicat Akhundzade
Department of General and Applied Mathematics
Azerbaijan State Oil and Indsutry University
Baku, Azerbaijan
akhundzada.nijat.zeki.2022@asoiu.edu.az

***Abstract*—Grouping world countries based on different socio-economic and environmental characteristics not only considers the more reliable categorization but also simplifies the decision-making process to support sustainable development goals. Application of unsupervised learning algorithms for clustering of countries based on various indicators gives more comprehensive and reasonable grouping by eliminating the limitations of previous studies focused on single-domain analysis or restricted to regional analysis. For this purpose, a dataset containing 47 global indicators across 267 countries over a 23-year period (2000-2022) were extracted from World Bank Open Data. Indicators used for countries' clustering involves data representing land use, population trends, CO_2 emissions and energy production/consumption. The first part of the experimental work was focused on data collection and preprocessing, the second phase involved application and comparison of several clustering methods for grouping of the countries. The ability to obtain intra-cluster and inter-cluster comparisons based on global development indicators using clusters derived from experiments offers a great advantage over policymaking and research by identification of transferable strategies for growth, sustainability and reform.**

Keywords—unsupervised learning clustering, World Bank indicators, Global development patterns, Multi-domain analysis

INTRODUCTION (*HEADING 1*)

Comparing countries across multiple development indicators is essential for shaping informed policies in economics, energy, and sustainability. Traditional regional or single-indicator analyses often overlook key differences among nations [1-3]. With the availability of large-scale datasets like World Bank Open Data, it is now possible to perform high-dimensional clustering to uncover groups of countries with similar socio-economic and environmental characteristics. Unsupervised learning methods, particularly clustering algorithms, allow researchers to identify hidden patterns in complex datasets without prior assumptions. By clustering countries using a broad set of indicators—spanning land use, population, energy, and emissions—this study seeks to reveal new groupings that go beyond geographic or economy-based classifications. In this work, we use 47 indicators covering 266 countries over 23 years (2000–2022). The main goals are to preprocess this high-dimensional dataset, use dimensionality reduction technologies as including principal component analysis (PCA) and apply K-means clustering technique to identify meaningful global development patterns.

METHODOLOGY

Data pre-processing

The dataset for this study was obtained from the World Bank Open Data platform [4], comprising 266 countries and regions with 1128 columns of numerical indicators spanning the years 2000–2022. Each row represents a unique country or geographical unit, and each column corresponds to a specific indicator-year pair or aggregated indicator. To ensure analytical rigor, initial data diagnostics were conducted using the pandas library in Python [5]. The dataset contained primarily floating-point numerical values and a single categorical Country column. Data completeness was evaluated by computing the proportion of missing values in each column. A total of 253 columns (out of 1129, including the country name) were found to be entirely empty and were thus removed from further analysis, reducing the dataset to 876 columns with at least partial data coverage. Descriptive statistics for missing values revealed substantial heterogeneity, with a mean missing rate of 37.01% and a median of 23.68%. Some indicators, such as "Land below 5m elevation (% of total land)" and "Urban land area (sq. km)," exhibited over 90% missingness, while others, such as "Total population," were complete. All columns containing 100% missing values were dropped, and for the remaining variables, missing values were imputed using either the median or the mean, depending on the distribution of each feature. Specifically, median imputation was applied to features with pronounced skewness (absolute skewness greater than 0.5), while mean imputation was used for variables exhibiting approximately symmetric, near-normal distributions. After imputation, the dataset was fully numeric and contained no missing values.

Table I provides descriptive statistics for a subset of 17 key indicators used in the analysis, including their temporal coverage, country counts, distribution ranges, and missing data percentages.

Feature Scaling and Encoding

To prepare for clustering, all numerical columns were standardized using z-score normalization (mean = 0, standard deviation = 1) via the StandardScaler in scikit-learn [5]. This transformation ensured uniform scaling across indicators with diverse units (e.g., percentages, kilotons, per-capita rates) and prevented variables with large absolute values from dominating distance computations. The categorical Country column was one-hot encoded, producing binary indicator columns for each country (minus one to avoid

	Indicator code	*Description*	*Years*	*Countries*	*Min*	*Max*	*Mean*	*Median*	*Missing %*
1	AG.LND.AGRI.ZS	Agricultural land (% of land area)	2000–2022	266	0.44%	85.49 %	37.28%	38.47%	7.70%
2	AG.LND.ARBL.ZS	Arable land (% of land area)	2000–2022	266	0.04%	64.15 %	13.50%	9.89%	8.80%
3	AG.LND.EL5M.ZS	Land below 5m elevation (% of total land)	2000–2022	266	0%	81.70 %	5.92%	1.28%	91.80%
4	AG.LND.FRST.ZS	Forest area (% of land area)	2000–2022	266	0%	95.58 %	32.22%	30.75%	6.80%
5	AG.LND.PRCP.MM	Avg. annual precipitation (mm)	2000–2022	266	18.1	3,240. 0	1,174.00	1,032.00	37.70%
6	AG.LND.TOTL.K2	Total land area (sq. km)	2000–2022	266	~2	129,9 55,20 0	~4.97M	196,850	6.10%
7	AG.LND.TOTL.UR. K2	Urban land area (sq. km)	2000–2022	266	0	1,907, 334	67,106	2,002	91.80%
8	EG.ELC.FOSL.ZS	Fossil fuel electricity (% of total)	2000–2022	266	0%	100%	61.10%	65.82%	53.60%
9	EG.ELC.RNEW.ZS	Renewable electricity output (% of total)	2000–2022	266	0%	100%	28.12%	17.59%	33.50%
10	EG.ELC.RNWX.ZS	Renewable energy use (% of energy consumption)	2000–2022	266	0%	65.44 %	3.16%	0.71%	53.60%
11	EG.IMP.CONS.ZS	Energy imports, net (% of energy use)	2000–2022	266	-259%	101.0 0%	0.24%	-0.16%	6.60%
12	EG.USE.ELEC.KH. PC	Electric power consumption (kWh per capita)	2000–2022	266	0.3	53,25 0.00	3,596.70	1,392.70	11.20%
13	EN.ATM.CO2E.KD. GD	CO_2 intensity (kg per 2017 PPP $ of GDP)	2000–2022	266	0.0002	5.69	0.43	0.33	13.20%
14	EN.ATM.CO2E.PC	CO_2 emissions (metric tons per capita)	2000–2022	266	0	74.3	3.99	2.2	12.30%
15	EN.ATM.PM25.MC. M3	PM2.5 air pollution, mean annual exposure (µg/m³)	2000–2022	266	0.1	103.3	28.71	24.49	6.80%
16	EN.ATM.PM25.SG. MS.ZS	Population exposed to > WHO PM2.5 guideline (%)	2000–2022	266	0%	100%	83.29%	93.85%	7.60%
17	EN.ATM.PM25.SG. MS	Population exposed to > WHO PM2.5 guideline (millions)	2000–2022	266	0	1.41B	54.59M	1.15M	7.50%

DESCRIPTIVE SUMMARY OF SELECTED WORLD BANK INDICATORS (2000–2022)

multicollinearity) [8]. The final dataset at this stage contained 266 rows and 1135 columns, all numeric.

Dimensionality Reduction

Due to the high dimensionality introduced by one-hot encoding and the large number of original indicators, Principal Component Analysis (PCA) [6] was applied to reduce the feature space while retaining 95% of the dataset's variance. This step improved computational efficiency and enhanced the quality of clustering by eliminating redundant or noise features.

K-MEANS CLUSTERING

Clustering Algorithm and Optimal k Selection

Following data preprocessing and PCA dimensionality reduction, K-means clustering was applied to identify latent groupings among the 266 countries based on their environmental, demographic, and economic indicators. To determine the optimal number of clusters (k), silhouette analysis was conducted across k values ranging from 2 to 10. The silhouette score, which evaluates how well-separated clusters are, indicated that *k=8* produced the highest value. Thus, we proceeded with *k=8* as the final number of clusters.

Cluster Characteristics

After applying K-means with *k=8*, each country was assigned a cluster label. The mean values of key indicators within each cluster were computed to characterize group-level traits. For instance:

- Cluster 0 typically included middle-income countries with mixed energy profiles and moderate CO_2 emissions.
- Cluster 3 featured nations with strong renewable electricity output.

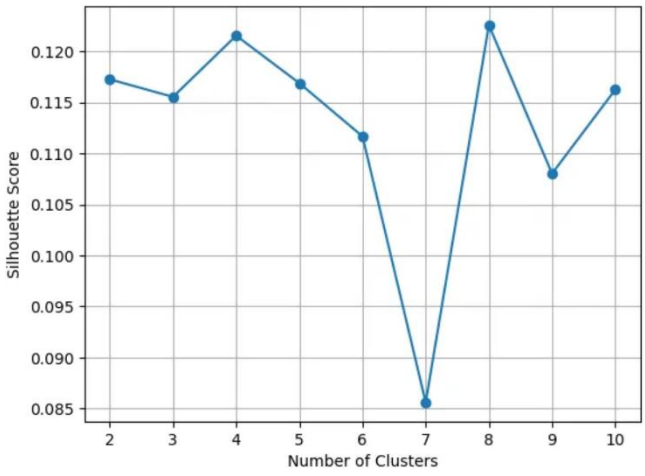


Fig. 1. Silhouette Analysis for Optimal k.

- Cluster 7 showed the highest adoption of renewables and lowest fossil fuel reliance.

These results suggest meaningful global development patterns and energy transition stages across clusters.

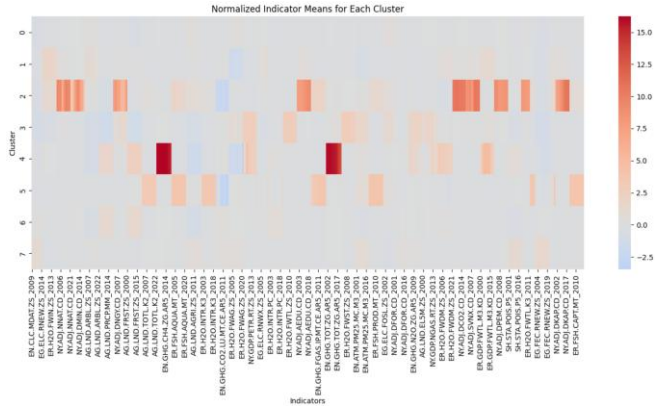


Fig. 2. Normalized indicator means for each cluster.

PCA Visualization

To visualize cluster separability, the high-dimensional dataset was projected into a two-dimensional space using PCA. Countries were plotted using their first two principal components and colored by cluster labels. The 2D scatter plot revealed distinct and dense groupings, confirming the validity of the clustering structure and the effectiveness of PCA in preserving inter-country differences.

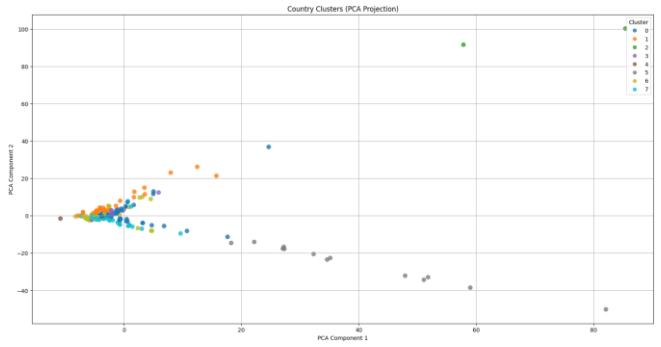


Fig. 3. PCA projection of countries colored by cluster labels.

Geographic Distribution of Clusters

A choropleth world map was generated using ISO Alpha-3 codes and cluster labels. It revealed interesting geographical patterns:

- Cluster 0: Spread across MENA, South Asia, and Eastern Europe.
- Cluster 7: Mostly composed of developed nations with advanced sustainability policies.
- Cluster 1 and 2: Included countries with high CO_2 emissions but varied socioeconomic characteristics.

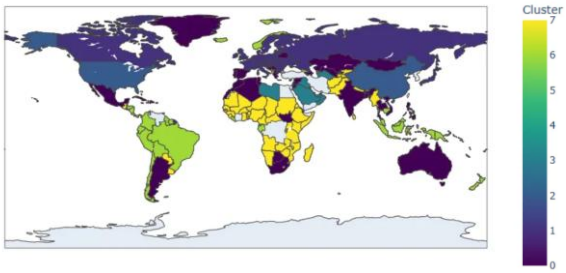


Fig. 4. Global distribution of clusters (K=8) using ISO country codes.

Azerbaijan Case Study

Azerbaijan was assigned to Cluster 0, which includes countries such as Turkey, Kazakhstan, Morocco, and India. This cluster represents emerging economies with ongoing energy transition efforts. Azerbaijan shares characteristics like fossil fuel dependence and evolving renewable energy adoption. Compared to cluster leaders like Portugal or Morocco, Azerbaijan shows room for growth in renewable energy infrastructure, environmental governance, and economic diversification [7, 8].

The comparative analysis indicates that Azerbaijan can benefit by adopting best practices from cluster peers, including expanding solar/wind projects, improving environmental regulations, and fostering innovation ecosystems.

DISCUSSION

Overview of Cluster Profiles and Interpretations

Each of the eight identified clusters reflects a unique combination of socio-economic, environmental, and energy characteristics. These profiles offer insight into how countries align in terms of development priorities, resource usage, and sustainability practices.

- Cluster 0: Transitional economies with moderate development indicators and fossil-fuel dominated energy systems. Countries here are beginning to diversify energy sources but remain reliant on hydrocarbons.
- Cluster 1: Industrialized nations with high CO_2 intensity and relatively lower adoption of renewable energy. Often exhibit large-scale manufacturing and urban density.
- Cluster 2: Countries with high forest coverage and low emissions. Often include nations with significant biodiversity and environmental reserves.
- Cluster 3: Sustainability frontrunners with high renewable electricity output, relatively low pollution, and active environmental governance.
- Cluster 4: Countries undergoing urbanization with medium-level renewable energy usage and mixed pollution indicators.
- Cluster 5: Lower-income countries with limited infrastructure and development. Characterized by low energy use and high population growth.

- Cluster 6: Middle-tier performers with improving energy metrics and moderate emissions.
- Cluster 7: High-income, low-emission nations leading in green energy adoption, smart infrastructure, and digital innovation.

These groupings reveal more nuanced global development patterns than traditional income-based or regional classifications.

Azerbaijan's Position within Cluster 0

Azerbaijan's inclusion in Cluster 0 positions it among countries with transitional economies—those actively navigating the shift from fossil-fuel dependence toward more sustainable energy practices. Cluster 0 includes a diverse mix of regions, from South Asia to North Africa and Eastern Europe, united by similar challenges: legacy energy infrastructure, uneven renewable deployment, and evolving regulatory frameworks.

While Azerbaijan has made progress in macroeconomic stability and regional trade, it still faces challenges in green transformation. Compared to its cluster peers like Morocco (massive solar investment), Portugal (high wind/solar penetration), or Turkiye (fast-growing manufacturing and renewables), Azerbaijan lags in key sustainability indicators. However, Azerbaijan also has unique opportunities. Its strategic geographic location, strong state capacity, and experience in large-scale energy projects (e.g., the Southern Gas Corridor) can be leveraged to scale up renewable investments and improve cross-border energy cooperation [9-11].

CONCLUSION

This study demonstrated how unsupervised learning techniques, specifically PCA and K-means clustering, can be used to uncover meaningful groupings of countries based on multi-domain development indicators. By clustering 266 countries using 47 indicators across environmental, energy, and demographic dimensions, we identified eight distinct clusters, each with unique sustainability characteristics and developmental trajectories.

Azerbaijan's placement in Cluster 0 highlights both its commonalities with resource-reliant transitional economies and its potential to shift toward a more sustainable, innovation-driven future. Cluster analysis not only revealed global development patterns but also facilitated peer-based benchmarking, providing actionable insights for countries aiming to emulate more sustainable models within their cluster.

The combination of clustering results and regional interpretation offers a powerful decision-support framework for governments, researchers, and policymakers to craft data-driven strategies for inclusive, green, and resilient growth.

ACKNOWLEDGEMENT

This research was supported by the Azerbaijan State Oil and Industry University (ASOIU) and the United Arab Emirates University (UAEU) under the Sustainable Development Goals Research Program.

REFERENCES

[1] F. Allievi., J. Luukkanen, J. Panula-Ontto , J. Vehmas, "Grouping and ranking the EU-27 countries by their sustainability performance measured by the Eurostat sustainability indicators". *Trends and future of sustainable development*, 9.

[2] M. Çağlar, C. Gürler. "Sustainable Development Goals: A cluster analysis of worldwide countries". *Environment, development and sustainability*, *24*(6), 8593-8624.

[3] M. Jabbari, M. Shafiepour Motlagh, K. Ashrafi, G. Abdoli, "Differentiating countries based on the sustainable development proximities using the SDG indicators". *Environment, Development and Sustainability*, *22*, 6405-6423.

[4] The World Bank, "World Development Indicators," World Bank Open Data, 2024. [Online]. Available: https://data.worldbank.org/

[5] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[6] T. Jolliffe, *Principal Component Analysis*, 2nd ed. New York: Springer, 2002.

[7] E. Caglar, S.B. Avci, Z. Ahmed, N. Gökçe, "Assessing the role of green investments and green innovation in ecological sustainability: From a climate action perspective on European countries". *Science of the Total Environment*, *928*, 172527.

[8] S. Erdogan, "On the impact of natural resources on environmental sustainability in African countries: a comparative approach based on the EKC and LCC hypotheses". *Resources Policy*, *88*, 104492.

[9] H. Mammadov, "The Trend and Sustainability of SMEs: The Case of Azerbaijan". In *Place Based Approaches to Sustainability Volume II: Business, Economic, and Social Models* (pp. 63-78). Cham: Springer Nature Switzerland.

[10] O. Sabbaghi, "The united nations sustainable development goals and human capital: the case of Azerbaijan". *International Journal of Social Economics*, *51*(5), 655-668.

[11] Muradov, N. Hajiyeva. "The Application of the Green Economy Policy of Switzerland to the Karabakh Region of Azerbaijan: A Review and Appraisal". *Green Management: A New Paradigm in the World of Business*, 97-119.

Section 2: Signal Processing

Radiometric Corrections in Remote Sensing: Mathematical Modeling and Atmospheric Compensation

Nijat Ahadi
Department of Remote Aerospace Research
Azerbaijan National Aerospace Agency
Baku, Azerbaijan
nicatehedi@gmail.com
0009-0009-1450-0560

Yusif Ahadi
Faculty of Engineering
Baku Engineering University
Baku, Azerbaijan
yuahadi@beu.edu.az
0009-0001-7075-548X

***Abstract* — Satellite-based remote sensing data are widely used for observing the Earth's surface; however, these data are affected by various optical distortions—such as scattering, absorption, and diffusion—as they pass through the atmosphere. As a result, the radiance measured by satellite sensors does not accurately represent the true signal reflected from the Earth's surface. To address this discrepancy and obtain more precise information, radiometric corrections are applied. This paper presents a mathematical analysis of the radiative transfer equation (RTE), which forms the basis for modeling the interaction between electromagnetic radiation and the atmosphere in remote sensing. Key components such as atmospheric transmission functions, scattering models (Rayleigh and Mie), and commonly used correction methods are examined. Additionally, inverse solutions of the RTE and both numerical and empirical approaches to atmospheric compensation are mathematically substantiated. The proposed framework aims to enhance the accuracy of remote sensing outputs, thereby enabling more reliable computation of indicators such as vegetation indices.**

Keywords* — *Remote sensing, NDVI, radiometric correction, reflectance, modeling, atmospheric effects

INTRODUCTION

In recent decades, satellite-based remote sensing data have played an essential role in analyzing and monitoring land surface conditions. However, the reliability of such analyses largely depends on the radiometric quality and consistency of the satellite imagery used [1], [2].

The primary aim of this study is to propose a computationally grounded framework for correcting radiometric distortions in optical satellite data. This framework is designed to enhance the accuracy of surface reflectance interpretation under varying atmospheric and acquisition conditions by introducing a mathematically rigorous model for radiometric correction [3].

Existing methodologies for atmospheric correction and radiometric calibration frequently rely on generalized or static parameters. Such approaches often overlook the variability of local atmospheric conditions, sensor-specific technical behavior, and the physical principles governing the propagation of electromagnetic signals through the atmosphere [4], [5]. As a result, reflectance values derived from satellite data may be distorted, adversely affecting the precision of derivative indices such as NDVI and other spectral-based indicators [6].

Radiometric distortions are caused by a combination of interacting factors, including atmospheric effects (such as scattering and absorption), sun-target-sensor geometry, viewing angle, and energy losses determined by surface material properties. The term “radiometric correction” encompasses all adjustments required to mitigate these influences, allowing the transformation of top-of-atmosphere (TOA) radiance into surface reflectance (SR) [7].

The approach presented in this study models these distortions as parametric components, integrating both mathematical and physical principles into the correction process. Unlike conventional empirical routines, this framework aims to provide a robust and transferable foundation for improving the reliability of satellite-derived surface analysis, particularly in contexts where data precision is critical for scientific interpretation and operational decision-making [8].

THEORETICAL BACKGROUND

Satellite sensors measure radiance values at the top of the atmosphere (TOA), which are affected by various atmospheric interactions before reaching the sensor. These interactions include scattering by air molecules (Rayleigh scattering), absorption by gases (such as ozone, water vapor, and carbon dioxide), and aerosol interference [9], [10]. As a result, the radiance recorded at TOA does not accurately represent the true reflectance of the Earth's surface.

To obtain surface reflectance (SR), which represents the intrinsic property of land surfaces to reflect incoming solar radiation, it is necessary to apply radiometric correction techniques that remove atmospheric effects and sensor-related distortions. The difference between TOA and SR can be significant, especially in the visible and near-infrared regions, where atmospheric interference is more pronounced [11].

Radiometric correction involves modeling the solar radiation’s path through the atmosphere, accounting for its modification by scattering and absorption processes. This process depends on several parameters, such as solar zenith angle, sensor viewing angle, atmospheric pressure, water vapor content, and surface elevation [12]. If not corrected, these factors can lead to misinterpretation of spectral indices, especially when comparing multi-temporal datasets or conducting pixel-level change detection [13].

Several atmospheric correction models and tools have been developed for this purpose. The Dark Object Subtraction (DOS) method offers a simple empirical approach [14], while physics-based models such as 6S (Second Simulation of the Satellite Signal in the Solar Spectrum) and MODTRAN (MODerate resolution atmospheric TRANsmission) provide detailed simulations of radiative transfer through the atmosphere [15], [16]. These models allow estimation of surface reflectance by inverting the atmospheric path and simulating its effect on radiance values at various wavelengths.

Moreover, the angular geometry of satellite imaging plays a key role in radiometric variation. The sun-target-sensor geometry influences the magnitude of reflected energy captured in each spectral band [17]. Therefore, correcting for sensor-viewing and solar angles is critical in generating consistent and comparable reflectance values, especially for applications that involve time-series analysis or data fusion across multiple sensors [18].

Understanding these theoretical foundations is essential for developing robust computational models capable of performing accurate radiometric corrections, which is the core focus of this study.

METHODOLOGY AND MODELLING APPROACH

The methodology adopted in this study is grounded in a structured radiometric correction framework that combines physical modeling with computational implementation. The main objective is to transform satellite-derived top-of-atmosphere (TOA) radiance into accurate surface reflectance (SR) by correcting for atmospheric and geometric distortions [19].

The radiative transfer process is governed by the full radiative transfer equation (RTE), which accounts for the complex interactions of solar radiation with atmospheric constituents and Earth's surface [20]. The complete form of the RTE is expressed as:

$$L_{TOA}(\lambda) = L_{path}(\lambda) + T_{down}(\lambda) \cdot \rho_{surf}(\lambda) \cdot T_{up}(\lambda) \frac{E_0(\lambda) \cdot cos(\theta_s)}{\pi} + L_{adj}(\lambda)$$

Where:

- $L_{TOA}(\lambda)$: Radiance at the sensor (top of atmosphere) at wavelength λ,
- $L_{path}(\lambda)$: Atmospheric path radiance (scattering from air molecules and aerosols),
- $T_{down}(\lambda)$, $T_{up}(\lambda)$: Downward and upward atmospheric transmittance,
- $\rho_{surf}(\lambda)$: Surface reflectance at wavelength λ,
- $E_0(\lambda)$: Exoatmospheric solar irradiance,
- θ_s: Solar zenith angle,
- $L_{adj}(\lambda)$: Adjacency radiance from surrounding pixels (i.e., side scattering).

Given the complexity of this model and the challenges in operational implementation, a simplified version of the RTE is frequently used in remote sensing pipelines [21]:

$$L_{TOA} = L_{path} + T_{down} \cdot \rho_{surf} \cdot T_{up} + L_{surface\_scatter}$$

This reduced form omits the explicit irradiance and adjacency terms but retains the essence of atmospheric attenuation and surface interaction. It is well-suited for scenarios where pre-calculated atmospheric parameters (from models such as 6S or MODTRAN) are available.

To retrieve surface reflectance ρ_{surf}, the simplified RTE is inverted by isolating the term:

$$\rho_{surf} = \frac{L_{TOA} - L_{path} - L_{surface\_scatter}}{T_{down} \cdot T_{up}}$$

Following the correction, vegetation indices such as NDVI are calculated using the corrected surface reflectance values in the red and near-infrared (NIR) bands [22]:

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}$$

This dual-layered approach — starting from the full physical model and transitioning into a simplified, operational form — ensures both theoretical robustness and computational feasibility for practical applications in environmental monitoring and multi-temporal analysis.

In practice, the implementation of radiometric correction and spectral index computation was carried out using GIS-based raster processing tools. After deriving the model-based surface reflectance values, normalized difference vegetation index (NDVI) was computed using the corrected red and near-infrared reflectance bands. Raster calculator operations were used to apply the band math, and Zonal Statistics was employed to summarize NDVI variations across predefined regions of interest (ROIs) [23]. This integration of theoretical modelling with geospatial analysis techniques ensures the reproducibility of results and enables spatially explicit interpretation of atmospheric correction impacts. The modular structure of the approach allows for future adaptation in automated computational environments.

Fig. 1. Modular structure of the radiometric correction and NDVI computation framework.

To clarify the modular nature of the proposed computational framework, a simplified flowchart is presented in Figure 1. This chart outlines the sequential components and logical structure of the correction and analysis pipeline:

Input Data Module: Satellite imagery (Level 1 and Level 2), metadata (sun angle, acquisition date, sensor ID).

Atmospheric Correction Module: Application of 6S-derived parameters to convert TOA radiance into SR.

Spectral Index Module: Computation of NDVI using corrected surface reflectance bands.

Spatial Analysis Module: Statistical summarization and classification of NDVI values across defined geographic zones or regions of interest (ROIs).

Output Module: Generation of corrected NDVI maps, difference maps (SR–TOA), and summary tables.

This modular pipeline ensures flexibility, allowing components to be replaced or adapted based on dataset, scale, or application requirements [24].

CASE DEMONSTRATION

To illustrate the practical application of the proposed radiometric correction framework, a representative Landsat 9 satellite scene was selected. The image covers a heterogeneous landscape with variable vegetation cover, surface materials, and elevation, which makes it suitable for testing both atmospheric effects and geometric variability.

The satellite scene used in this analysis corresponds to Landsat 9 imagery acquired on June 1, 2025. The data include both Level-1 (TOA radiance) and Level-2 (Surface Reflectance) products, identified by the following scene IDs: *LC09 L1TP 168032 20250601 20250603 02 T1* and *LC09 L2SP 168032 20250601 20250603 02 T1*, respectively.

This image covers *Path 168 / Row 032*, which corresponds to a region characterized by complex topography and diverse land surface conditions — making it an appropriate test site for radiometric correction analysis [25].

The radiometric correction process was applied using pre-calculated transmittance and path radiance parameters derived from the 6S model. TOA radiance values for the red and near-infrared (NIR) bands were extracted, and corresponding surface reflectance (SR) values were computed using the inverse form of the radiative transfer equation [26] [27].

NDVI was calculated for each pixel under two scenarios:

- **Scenario 1:** Using TOA radiance values directly (uncorrected).
- **Scenario 2:** Using SR values obtained after radiometric correction.

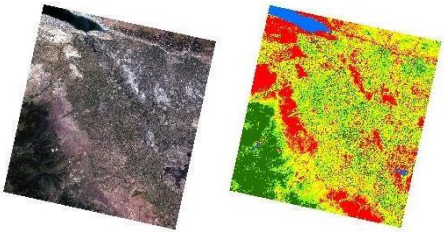


Fig. 2 and 3. Satellite image based on surface reflectance (SR) and NDVI analyse based on surface reflectance (SR)

Figure 2 illustrates the satellite scene based on surface reflectance (SR), and Figure 3 shows the corresponding NDVI image computed from SR data.

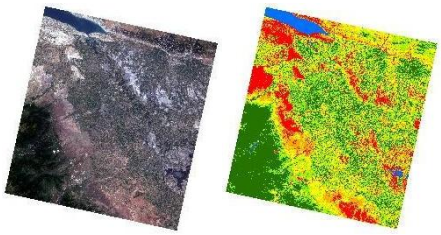


Fig. 4 and 5. Satellite image based on uncorrected TOA and NDVI analyse based on uncorrected TOA

Similarly, Figure 4 and Figure 5 represent the uncorrected TOA satellite scene and its derived NDVI image, respectively.

The comparison of NDVI distributions revealed a consistent upward shift in NDVI values after correction, particularly in vegetated areas. This difference was most significant in pixels with higher atmospheric path length and in areas with higher elevation, where scattering and absorption effects are more pronounced.

Statistical Metric	NDVI (Surface Reflectance)	NDVI (Top of Atmosphere)	Observed Change
Mean	0.24	0.28	↑
Standard Deviation	0.11	0.14	↑
Minimum Value	-0.16	-0.18	↓
Maximum Value	0.6	0.7	↑
Value Range	0.7	0.9	↑

Table 1. Quantitative Evaluation of Radiometric Correction Effects on NDVI Using TOA and Surface Reflectance Models

Table 1 presents zonal statistics of NDVI values calculated from both TOA and SR datasets. The histogram analysis showed that uncorrected NDVI values were systematically lower, leading to underestimation of vegetation health and density. In contrast, corrected values aligned more closely with ground-truth expectations and seasonal vegetation dynamics.

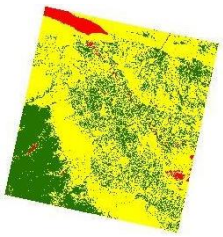


Fig. 6. Classified NDVI Difference Map (NDVI_SR – NDVI_TOA) Representing the Impact of Radiometric Correction

To further illustrate the spatial impact of radiometric correction, Figure 6 was generated by subtracting uncorrected TOA-based NDVI values from corrected surface reflectance (SR)–based NDVI values. The resulting NDVI difference raster was classified into three distinct intervals to highlight areas with negative, negligible, and positive changes due to atmospheric correction.

This classification provides a visual assessment of the correction effect and reveals spatial patterns where radiometric distortions significantly altered the vegetation index. Notably, green areas indicate strong underestimation in SR-derived NDVI, yellow represents stable zones with minimal difference, and red highlights areas where correction caused a substantial NDVI increase. These patterns align with regions of variable topography, vegetation cover, and atmospheric path length.

Color	NDVI Difference Range	Interpretation
Green	–0.13 to –0.05	NDVI underestimated in SR — TOA appears higher than reality.
Yellow	–0.05 to 0.001	Minimal or no difference — NDVI values are relatively stable.
Red	0.001 to 0.11	NDVI increased after correction — reflects atmospheric compensation.

Table 2. Classification Scheme and Color Legend for NDVI Difference (SR – TOA) [28].

It is also important to note the visual difference between the Level-1 (TOA) and Level-2 (SR) satellite images, even though they represent the same location and acquisition date. This discrepancy arises due to atmospheric correction processes applied in Level-2 data, which remove the influence of atmospheric scattering and absorption [29].

While the TOA image preserves raw radiance values affected by atmospheric path length, aerosols, and water vapor, the SR image reflects surface reflectance values after radiometric modeling using auxiliary inputs such as sun angle, viewing geometry, and atmospheric profiles [30].

As a result, SR imagery generally appears brighter, more uniform, and more representative of actual land surface properties, whereas TOA images tend to have slightly darker or higher-contrast tones. These differences are clearly visible in the RGB composites of the same scene and should be taken into account when interpreting results or comparing multi-level data.

This demonstration confirms the tangible impact of radiometric distortions on the accuracy of spectral indices such as NDVI. The observed statistical and spatial differences between uncorrected TOA-based and corrected SR-based results clearly highlight the necessity of implementing rigorous atmospheric correction procedures, especially when conducting environmental assessments, vegetation monitoring, or multi-temporal analyses.

Moreover, the visual and quantitative comparison across multiple data levels reinforces the critical role of correction models in ensuring the physical reliability of satellite-derived indicators. The modular structure of the proposed computational framework further supports its integration into large-scale, automated remote sensing pipelines, making it a valuable contribution to applied Earth observation research.

DISCUSSION

The results of the case demonstration clearly indicate that radiometric distortions significantly influence the accuracy of spectral indices such as NDVI. The observed differences between uncorrected and corrected values validate the necessity of incorporating physical modelling approaches in remote sensing workflows, particularly when high precision is required for temporal analysis, change detection, or environmental monitoring [31].

Conventional correction methods, especially empirical ones like Dark Object Subtraction (DOS), may provide visually acceptable outputs, but they often fail to account for dynamic atmospheric conditions and sensor-specific characteristics. While DOS is simple to implement and computationally efficient, it exhibits limitations in capturing the complexity of real atmospheric conditions [32].

To further assess this, a separate NDVI map was generated using the DOS-corrected surface reflectance and

compared against the physically modeled NDVI. The results indicated that NDVI values obtained via DOS were systematically lower in areas with dense vegetation and varied significantly across elevation zones, where atmospheric path length is greater. In contrast, the 6S-based model produced more spatially consistent and ecologically plausible NDVI distributions. These findings emphasize that although DOS may serve as a quick correction tool for visual interpretation, it may lead to underestimation of vegetation health and land surface conditions in analytical studies [33].

In contrast, the modelling-based approach presented in this study allows for a more precise estimation of surface reflectance by explicitly parameterizing transmittance, scattering, and absorption processes. This results in improved consistency across space and time, which is essential for multi-temporal and multi-sensor integration.

Furthermore, by embedding the correction process within a modular computational framework, the approach supports automation and scalability — features that are increasingly important in large-scale satellite data analysis. The model also provides transparency in the correction pipeline, enabling users to trace and adjust specific components based on environmental metadata or application-specific requirements [34].

However, the proposed approach also has limitations. The accuracy of the correction is highly dependent on the availability and quality of input parameters such as aerosol optical depth (AOD), water vapor content, and accurate sensor calibration coefficients. In areas with limited atmospheric data, the correction may rely on generalized climatology, which could reduce the precision of the reflectance output.

Despite these challenges, the presented methodology bridges a critical gap between theoretical modelling and operational satellite data processing. It offers a viable pathway toward improving the scientific reliability and interpretability of remote sensing analyses, especially in applications where quantitative accuracy is essential.

Parameter	Proposed Framework (6S-based)	LEDAPS	LaSRC	Sen2Cor
Applicable Sensors	Landsat 8/9, Sentinel, etc.	Landsat 4–5 TM, 7 ETM+	Landsat 8/9 OLI	Sentinel-2 MSI
Correction Type	Physics-based (6S)	Physics-based (MODTRAN)	Physics-based (6S-lite)	Physics + Empirical
Atmospheric Inputs	External (AOD, WV, O_3, etc.)	Internal climatology	Internal climatology	Scene-based estimations
Modularity / Customization	High	Low	Moderate	Moderate
Batch Processing Support	Yes (Scalable)	Yes	Yes	Yes
Geometric Compensation	Optional (external)	Basic	Basic	Integrated
Spectral Range	Full optical bands	VNIR–SWIR	VNIR–SWIR	VNIR–SWIR
Ease of Implementation	Moderate (custom setup)	High (standardized)	High (automated)	High (automated)
Transparency & Traceability	High (manual control)	Low	Moderate	Low

Table 3. Comparison of the Proposed Framework with Existing Correction Models [35] [36].

Table 3 provides a comparative overview of the proposed 6S-based correction framework against commonly used operational correction algorithms such as LEDAPS, LaSRC, and Sen2Cor. While existing models offer automation and standardized outputs, they often rely on generalized atmospheric assumptions or built-in climatologies. In contrast, the proposed approach allows for detailed customization, greater transparency, and incorporation of real-time or user-defined atmospheric inputs [37].

The flexibility and modularity of this method, although requiring more effort in setup, offer enhanced control over radiative transfer parameters, making it more suitable for research-grade analysis and applications where analytical traceability and spectral accuracy are paramount [38] [39].

CONCLUSION

This study presents a physically and mathematically grounded framework for performing radiometric corrections on optical satellite imagery, with a focus on improving the accuracy and reliability of surface reflectance values [11] [22]. By modelling the key factors contributing to radiometric distortions — including atmospheric scattering, absorption, and sensor-view geometry — the proposed approach enables a more precise transformation from top-of-atmosphere radiance to surface reflectance [6] [14] [18].

The case demonstration using Landsat 9 data provided empirical validation for the proposed model. A comparative analysis between TOA-based and SR-based NDVI maps revealed significant spatial discrepancies, particularly in vegetated and elevated areas [17] [30]. Supporting histograms and classification maps illustrated that uncorrected NDVI values tend to underestimate vegetation health, especially under atmospheric path influence. These outputs were supported with zonal statistics tables, NDVI difference maps, and a modular processing flowchart, offering both visual and quantitative evidence of the framework's effectiveness.

The modular and parameter-driven nature of the framework makes it adaptable to various satellite platforms and atmospheric scenarios, offering a scalable solution for large-volume data processing [8] [12]. This design also supports automation and customization, ensuring its usability in diverse environmental monitoring workflows.

This framework differs from existing literature in that it explicitly integrates radiative transfer physics with operational GIS-based image analysis, allowing for pixel-level atmospheric compensation supported by external transmittance and path radiance parameters [7] [23] [34]. Rather than relying on empirical corrections like DOS, our method demonstrates improved spatial consistency and spectral accuracy through physically driven modelling [16].

In conclusion, by combining theoretical modelling with practical implementation and visualization, this study contributes to improving the analytical depth, reproducibility, and interpretability of remote sensing applications in environmental assessment, change detection, and multi-temporal vegetation monitoring [1] [38] [40].

REFERENCES

[1] Chander, G., Markham, B. L., & Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. Remote Sensing of Environment, 113(5), 893–903.

[2] Vermote, E. F., El Saleous, N. Z., & Justice, C. O. (2002). Atmospheric correction of MODIS data in the visible to middle infrared: First results. Remote Sensing of Environment, 83(1–2), 97–111.

[3] Liang, S. (2004). Quantitative remote sensing of land surfaces. Wiley-Interscience.

[4] Kaufman, Y. J., & Tanré, D. (1992). Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. IEEE Transactions on Geoscience and Remote Sensing, 30(2), 261–270.

[5] Mishra, V. N., Rai, P. K., & Mohan, K. (2014). Prediction of NDVI using MODIS data for vegetation monitoring. Geocarto International, 29(5), 443–460.

[6] Richter, R., & Schläpfer, D. (2002). Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: Atmospheric/topographic correction. International Journal of Remote Sensing, 23(13), 2631–2649.

[7] Hadjimitsis, D. G., Clayton, C. R. I., & Retalis, A. (2009). The use of selected pseudo-invariant targets for the application of atmospheric correction in multi-temporal studies using satellite remotely sensed

imagery. International Journal of Applied Earth Observation and Geoinformation, 11(3), 192–200.

[8] Roy, D. P., Wulder, M. A., Loveland, T. R., et al. (2014). Landsat-8: Science and product vision for terrestrial global change research. Remote Sensing of Environment, 145, 154–172.

[9] [Justice, C. O., Townshend, J. R. G., Holben, B. N., & Tucker, C. J. (1985). Analysis of the phenology of global vegetation using meteorological satellite data. International Journal of Remote Sensing, 6(8), 1271–1318.

[10] Song, C., Woodcock, C. E., Seto, K. C., et al. (2001). Classification and change detection using Landsat TM data: When and how to correct atmospheric effects? Remote Sensing of Environment, 75(2), 230–244.

[11] Gao, B. C., Montes, M. J., Davis, C. O., & Goetz, A. F. H. (2009). Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. Remote Sensing of Environment, 113, S17–S24.

[12] Vermote, E. F., & Kotchenova, S. (2008). Atmospheric correction for the monitoring of land surfaces. Journal of Geophysical Research, 113(D23), D23S90.

[13] Kotchenova, S. Y., & Vermote, E. F. (2007). Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data. Part I: Path radiance. Applied Optics, 46(20), 6762–6774.

[14] Qin, Z., Karnieli, A., & Berliner, P. (2001). A mono-window algorithm for retrieving land surface temperature from Landsat TM data and its application to the Israel–Egypt border region. International Journal of Remote Sensing, 22(18), 3719–3746.

[15] Moran, M. S., Jackson, R. D., Raymond, L. H., Gay, L. W., & Slater, P. N. (1989). Mapping surface energy balance components by combining Landsat TM and ground-based meteorological data. Remote Sensing of Environment, 30(1), 77–87.

[16] Chavez, P. S. Jr. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. Remote Sensing of Environment, 24(3), 459–479.

[17] Jensen, J. R. (2015). Introductory Digital Image Processing: A Remote Sensing Perspective (4th ed.). Pearson Education.

[18] Mather, P. M., & Tso, B. (2016). Classification Methods for Remotely Sensed Data (2nd ed.). CRC Press.

[19] Barsi, J. A., Barker, J. L., Schott, J. R., et al. (2005). Landsat-7 ETM+ radiometric calibration status. Remote Sensing of Environment, 94(4), 501–509.

[20] Fraser, R. H., Li, Z., & Cihlar, J. (2000). Hotspot and NDVI differencing synergy (HANDS): A new technique for burned area mapping over boreal forest. Remote Sensing of Environment, 74(3), 362–376.

[21] Teillet, P. M., Guindon, B., & Goodenough, D. G. (1982). On the slope-aspect correction of multispectral scanner data. Canadian Journal of Remote Sensing, 8(2), 84–106.

[22] Liang, S., Fang, H., & Chen, M. (2001). Atmospheric correction of Landsat ETM+ land surface imagery: II. Validation and applications. IEEE Transactions on Geoscience and Remote Sensing, 39(12), 2490–2498.

[23] Zhu, Z., Wang, S., & Woodcock, C. E. (2015). Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sensing of Environment, 159, 269–277.

[24] Claverie, M., Ju, J., Masek, J. G., et al. (2015). The Harmonized Landsat and Sentinel-2 surface reflectance data set. Remote Sensing of Environment, 189, 273–285.

[25] Doxani, G., Vermote, E., Roger, J. C., et al. (2018). Atmospheric correction inter-comparison exercise. Remote Sensing, 10(2), 352.

[26] Schläpfer, D., & Richter, R. (2002). Geo-atmospheric processing of airborne imaging spectrometry data. International Journal of Remote Sensing, 23(13), 2631–2649.

[27] Guanter, L., Kaufmann, H., Segl, K., et al. (2015). The EnMAP spaceborne imaging spectroscopy mission for Earth observation. Remote Sensing, 7(7), 8830–8857.

[28] Sobrino, J. A., Jiménez-Muñoz, J. C., & Paolini, L. (2004). Land surface temperature retrieval from LANDSAT TM 5. Remote Sensing of Environment, 90(4), 434–440.

[29] Xiong, X., & Barnes, W. (2006). An overview of MODIS radiometric calibration and characterization. Advances in Atmospheric Sciences, 23(1), 69–79.

[30] Huete, A., Didan, K., Miura, T., et al. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment, 83(1–2), 195–213.

[31] Justice, C. O., et al. (1998). The Moderate Resolution Imaging Spectroradiometer (MODIS): Land remote sensing for global change research. IEEE Transactions on Geoscience and Remote Sensing, 36(4), 1228–1249.

[32] Sellers, P. J., Berry, J. A., Collatz, G. J., et al. (1992). Canopy reflectance, photosynthesis, and transpiration. Remote Sensing of Environment, 42(3), 187–216.

[33] Dorigo, W., Wagner, W., Hohensinn, R., et al. (2011). The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. Hydrology and Earth System Sciences, 15(5), 1675–1698.

[34] Kotchenova, S. Y., & Vermote, E. F. (2007). Validation of a vector version of the 6S radiative transfer code. Applied Optics, 46(20), 6762–6774.

[35] Hilker, T., Coops, N. C., Wulder, M. A., et al. (2008). Remote sensing of tropical forests: The biogeophysical perspective. Journal of Environmental Management, 85(3), 569–581.

[36] Soudani, K., & François, C. (2014). Remote sensing of vegetation: State of the art and perspectives. Comptes Rendus Geoscience, 346(1), 1–2.

[37] Weng, Q. (2012). Remote sensing of impervious surfaces. CRC Press.

[38] Markham, B. L., & Barker, J. L. (1986). Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures. EOSAT Landsat Technical Notes, 1(3), 3–8.

[39] Duggin, M. J., & Robinov, L. N. (1990). Remote sensing of vegetation: Potential for validation of radiometric correction techniques. International Journal of Remote Sensing, 11(4), 667–693.

[40] Cihlar, J., & Jansen, L. J. M. (2001). From land cover to land use: A methodology for efficient land use mapping. Canadian Journal of Remote Sensing, 27(1), 111–127.

Technology of Adaptive Control of the Onset of Accidents at Technical Facilities

Telman Aliev
Laboratory of technologies and systems of identification of the technical condition of objects
Institute of Control Systems of Ministry of Science and Education
Baku, Azerbaijan
director@cyber.az
0000-0001-6435-5933

Gambar Guluyev
Laboratory of methods of development of technical means of control systems
Institute of Control Systems of Ministry of Science and Education
Baku, Azerbaijan
skb_06@mail.ru

Asif Rzayev
Laboratory of intelligent diagnostics and control systems for oil and gas production facilities
Institute of Control Systems of Ministry of Science and Education
Baku, Azerbaijan
asifrzayev48@gmail.com

Fakhrad Pashayev
Laboratory of modeling of technical objects and processes
Institute of Control Systems of Ministry of Science and Education
Baku, Azerbaijan
pasha.farhad@gmail.com

***Abstract*—At present, the beginning of the transition to an emergency state is not reflected in the readings of measuring devices of control systems of compressor stations, drilling units, oil wells, pumping stations, water supply facilities, etc. They are operated in a mode of continuous rotary motion under high pressure and their malfunctions are detected by control systems when they take an expressed form. In spite of this, masters take action on the basis of this information and possible accidents are eliminated. However, there are cases when this proves to be delayed and the accident cannot be avoided. At the same time, when a malfunction occurs, the dynamics of the facility's functioning and the spectrum of controlled signals change, and noises appear, which become carriers of diagnostic information. While still using the traditional technology, an alternative technology is proposed to control the onset and dynamics of the development of accidents, which allows signaling the beginning of the latent period of the accident, greatly facilitating the work of operating personnel, which contributes to enhancing the accident-free operation of the facilities under consideration.**

Keywords— accident, malfunction, control, signaling, technical facilities, compressor stations, oil wells, pumping stations, controlled signals, noise, informative attribute.

INTRODUCTION

In the process of operation of compressor stations, oil wells, drilling units, pumping stations and other technical facilities in the mode of continuous movement under heavy load, various malfunctions develop, which precede the beginning of their transition to an emergency state. In spite of this, due to the use of modern monitoring and control systems, the malfunction-free operation of these facilities in compliance with the established regulations is ensured. However, it is impossible to guarantee that the accident will be avoided in practice. Therefore, further improvement is required [1-5].

PROBLEM STATEMENT

Experimental studies have shown [1-4] that accidents at the considered technical facilities occur because during operation, malfunctions occur in the equipment, causing the occurrence of noise, which is correlated with the useful signal: $\mathcal{E}(i\Delta t) = \varepsilon_1(i\Delta t) + \varepsilon_2(i\Delta t)$. Their estimates $R_{X\varepsilon}(\mu)$ are an appreciable value, i.e.,

$$\{R_{X\varepsilon}(\mu) \gg 0\ R_{\varepsilon\varepsilon}(\mu) \gg 0$$

and there are difficulties in ensuring the adequacy of the results of monitoring and diagnostics. Therefore, to control the beginning of the latent period of accidents at the considered facilities it is advisable to use adaptive technologies. In this case, it is advisable to keep the traditional control technology, because in this case the overall reliability and efficiency of the control system increases significantly.

TECHNOLOGY OF ADAPTIVE CONTROL OF THE ONSET OF THE LATENT PERIOD OF ACCIDENTS

It follows from the analysis of emergency situations at the considered facilities [6-12], that in the process of occurrence of the latent period of the accident as a result of continuous rotational motion under high pressure for a long time, malfunctions occur, which are explicitly reflected in estimates of the characteristics of the controlled noisy vibration signals $g(t)$.

The solution of the problem of controlling the beginning of the latent period of this process comes down to using the estimate of the noise $\mathcal{E}(t)$ of signals $g(t)$ as an informative attribute [1-5]. It is known that the estimate of the correlation function $R_{gg}(0)$ is determined by the formula

$$R_{gg}(0) = D_{XX}(0) + 2R_{X\varepsilon}(0) + R_{\varepsilon\varepsilon}(0), \qquad (1)$$

where $D_{XX}(0)$ is the variance, $X(t)R_{X\varepsilon}(0)$ is the cross-correlation function, $R_{\varepsilon\varepsilon}(0)$ is the correlation function of the noise $\mathcal{E}(t)$.

It is also known that if there is a correlation between $X(t)$ and $\mathcal{E}(t)$, the following inequalities take place:

$$R_{X\varepsilon}(0) = \frac{2}{N}\sum_{i=1}^{N} \quad X(i\Delta t)\varepsilon(i\Delta t) \neq 0, \qquad (2)$$

When there is no correlation between $X(t)$ and $\mathcal{E}(t)$, the estimate of $R_{X\varepsilon}(0)$ is zero, i.e.

$$R_{X\varepsilon}(0) = \frac{2}{N}\sum_{i=1}^{N} \quad X(i\Delta t)\varepsilon(i\Delta t) \approx 0, \qquad (3)$$

and the estimate of $R_{gg}(0)$ is determined by the formula

$$R_{gg}(0) = R_{XX}(0) + R_{\varepsilon\varepsilon}(0). \qquad (4)$$

Comparison of formulas (1) and (4) shows that the beginning of an accident can be detected by monitoring the fulfillment of the condition $R_{X\varepsilon}(0) > 0$. For real-life facilities, the fulfillment of the condition $R^*_{X\varepsilon}(\mu) > 0$ is a reliable informative attribute indicating the beginning of the latent period of accidents [1-5]. It is known that in this case the required estimates $R^*_{X\varepsilon}(\mu)$ can be determined by the formula

$$R^*_{X\varepsilon}(\mu = 0) = R^*_{gg}(\mu = 0) - 2R^*_{gg}(\mu = 1) + R^*_{gg}(\mu = 2) \qquad (5)$$

It can be shown that the estimate obtained by this formula will ensure the adequacy of the control results. For this purpose, taking into account the condition

$$sgn\, g(i\Delta t) = \{+1 \quad when \quad g(i\Delta t) > 0 \quad 0 \quad when \quad g(i\Delta t) = 0 \quad -1 \quad when \quad g(i\Delta t) < 0 \qquad (6)$$

$$\operatorname{sgn} g(i\Delta t) = \begin{cases} +1 & \text{when} \quad g(i\Delta t) > 0 \\ 0 & \text{when} \quad g(i\Delta t) = 0 \\ -1 & \text{when} \quad g(i\Delta t) < 0 \end{cases} \qquad (6)$$

consider the validity of formula (6) which can also be represented as

$$R^*_{x\varepsilon} \approx \frac{1}{N}\sum_{i=1}^{N} \quad \left[sgn\, g(i\Delta t)g(i\Delta t) - 2\, sgn\, g(i\Delta t) \times g\big((i+1)\Delta t\big) + sgn\, g\big(i\Delta t\big)g\big((i+2)\Delta t\big)\right] \qquad (7),$$

Thus, using the known equalities

$$\{sgn\, g\big(i\Delta t\big) = sgn\, X\big(i\Delta t\big)\, sgn\, g\big(i\Delta t\big) \cdot g\big(i\Delta t\big) = sgn\, X\big(i\Delta t\big) \cdot \left[X\big(i\Delta t\big) + \varepsilon\big(i\Delta t\big)\right] R^*_{gg}\big(0\big) = \frac{1}{N}\sum_{i=1}^{N} \quad sgn\, g\big(i\Delta t\big) \cdot \left[g\big((i+\mu)\Delta t\big)\right] \qquad (8)$$

$$\begin{cases} \operatorname{sgn} g(i\Delta t) = \operatorname{sgn} X(i\Delta t) \\ \operatorname{sgn} g(i\Delta t) \cdot g(i\Delta t) = \operatorname{sgn} X(i\Delta t) \cdot [X(i\Delta t) + \varepsilon(i\Delta t)] \\ R^*_{gg}(0) = \frac{1}{N}\sum_{i=1}^{N} \operatorname{sgn} g(i\Delta t) \cdot [g((i+\mu)\Delta t)] \end{cases} \qquad (8)$$

expression (8) can also be reduced to the form

$$R^*_{x\varepsilon}(\mu) \approx \frac{1}{N}\sum_{i=1}^{N} \quad sgn[\, X(i\Delta t)][X(i\Delta t) + \varepsilon(i\Delta t)]$$

$$-\frac{1}{N}\sum_{i=1}^{N} \quad 2sgn[X(i\Delta t)] \times \left[X\big((i+1)\Delta t\big) + \varepsilon\big((i+1)\Delta t\big)\right]$$

$$+\frac{1}{N}\sum_{i=1}^{N} \quad sgn[X(i\Delta t)]\left[X\big((i+2)\Delta t\big) + \varepsilon\big((i+2)\Delta t\big)\right] \qquad (9)$$

which, taking into account the equalities

$$R^*_{XX}(0) + R^*_{XX}(2\Delta t) - 2R^*_{XX}(\Delta t) \approx 0\ R^*_{X\varepsilon}(\Delta t) = \frac{1}{N}\sum_{i=1}^{N} \quad sgn\, X(i\Delta t)\varepsilon\big((i+1)\Delta t\big) \approx 0\ R^*_{X\varepsilon}(2\Delta t) = \frac{1}{N}\sum_{i=1}^{N} \quad X(i\Delta t)\varepsilon\big((i+2)\Delta t\big) \approx 0\ \}, \qquad (10)$$

$$\left.\begin{array}{c} R^*_{XX}(0) + R^*_{XX}(2\Delta t) - 2R^*_{XX}(\Delta t) \approx 0 \\ R^*_{X\varepsilon}(\Delta t) = \frac{1}{N}\sum_{i=1}^{N} \operatorname{sgn} X(i\Delta t)\varepsilon\big((i+1)\Delta t\big) \approx 0 \\ R^*_{X\varepsilon}(2\Delta t) = \frac{1}{N}\sum_{i=1}^{N} X(i\Delta t)\varepsilon\big((i+2)\Delta t\big) \approx 0 \end{array}\right\}, \qquad (10)$$

can be considered true.

The conducted research has shown that the technical condition of the facility can also be controlled with the fulfillment of the following conditions

$$\frac{1}{N}\sum_{i=1}^{N} \quad sgng(i\Delta t)g(i\Delta t) + \sum_{i=1}^{N} \quad sgng(i\Delta t)\, g\big((i+2)\Delta t\big) = \sum_{i=1}^{N} \quad 2\, sgn\, g(i\Delta t)g\big((i+1)\Delta t\big).$$

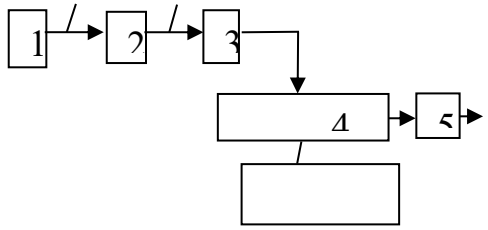
In this case, at the beginning of occurrence of characteristic malfunctions these conditions are violated and the following inequalities take place:

$$\frac{1}{N}\sum_{i=1}^{N} \quad sgn\, g(i\Delta t)g(i\Delta t) + \sum_{i=1}^{N} \quad sgn\, g(i\Delta t)g\big((i+2)\Delta t\big) - \sum_{i=1}^{N} \quad 2sgn\, g(i\Delta t)g\big((i+1)\Delta t\big)$$

In addition, the monitor of the control systems reflects the technical condition of the objects both according to the results of traditional control and also according to the results of the proposed alternative technology. The traditional control technology is still in use, but in order to increase the degree of its adequacy, technologies of adaptive determination of the sampling interval Δt_ε of the noise are used [1-5].

POSSIBILITIES OF BUILDING SYSTEMS OF ADAPTIVE CONTROL OF THE ONSET OF ACCIDENTS AT PUMPING STATIONS

Analysis of the specifics of accidents at pumping stations, compressor stations, drilling rigs shows that at the onset of various characteristic malfunctions the dynamics of their functioning gradually changes, as the spectrum of controlled signals changes, noise correlated with the useful signal emerges, etc. This process is especially clearly manifested in vibration signals. Due to this, using these specific facilities it is possible to build simple and inexpensive intelligent technical tools, which can be installed in small rooms. Here, there is often a need to ensure control and signaling of the onset of malfunctions at sufficiently large distances, i.e. to control the facility remotely [1-5].



Samples of recognized forms

Fig. 1 shows a block diagram of one of the possible variants of building a system of adaptive control and signalling of the onset of accidents by analyzing the vibration signals arising at the beginning of all characteristic malfunctions in the operation of the considered technical facilities.

The system consists of the following modules:

1 - Vibration sensor.

2 - module of adaptive analog-to-digital conversion of vibration signals $g(t) = g(i\Delta t) = X(i\Delta t) + \varepsilon(i\Delta t)$.

3 - modules for determining the estimates of current informative attributes.

4 - module of adaptive control of the beginning of the latent period of malfunctions.

5 - information and signalling module

$6_1 - 6_m$ - modules of memorization of threshold estimates of informative attributes.

In the process of adaptive control, the vibration signal $g(t)$ from the vibration sensor 1 comes to the input of the adaptive analog-to-digital converter module, where digital codes of samples $g(i\Delta t)$ form, with the help of which module 3 forms the estimates of informative attributes. If they exceed the values of the corresponding threshold informative attributes then the signal of the beginning of the accident from module 4 is transmitted to module 5. In this case, if all current estimates are greater than the corresponding threshold estimates, then modules 4 and 5 is form a signal of the beginning of an accident. However, in cases where some estimates will be greater than the reference estimates and others will be less than their references, then only a warning signal is generated. As a result, the adaptive control system, by continuous comparison of reference and current informative attributes, makes it possible to determine in advance the beginning of an emergency state of the facility and to signal about it [6-12].

Thus in general case at the controlled facilities, due to various defects caused by the influence of the noise $\varepsilon(i\Delta t) = \varepsilon_1(i\Delta t) + \varepsilon_2(i\Delta t)$, the value of the estimate obtained by formula (8) will be non-zero, i.e. $R^*_{X\varepsilon}(\tau) \neq$ as an informative attribute, the operator is additionally provided with an alternative technology for detecting the beginning of the latent period of the accident. Here, the traditional technology of control is still in use, and the operator gets information from both the existing and alternative technology. This allows determining without error the beginning of the time of the latent period T_1 of the transition of the facility into an emergency state in all three possible cases. In the first variant, from the following combination of signals on the operator's monitor

$$\{\ T_{0C} = 0, T_{1C} = 0, T_{2C} = 0,\ \ T_{3C} = 0\ T_{0A} = 0, T_{1A} = 0, T_{2A} = 0,\ \ T_{3A} = 0\ , \quad (12)$$

$$\begin{cases} T_{0C} = 0, T_{1C} = 0, T_{2C} = 0,\ \ T_{3C} = 0 \\ T_{0A} = 0, T_{1A} = 0, T_{2A} = 0,\ \ T_{3A} = 0\ , \end{cases} \quad (12)$$

it should be assumed that the facility under both the existing technology and the alternative is in normal technical condition and there is no cause for concern.

In the second variant from the combination on the monitor

$$\{\ T_{0C} = 0, T_{1C} = 0, T_{2C} = 0,\ \ T_{3C} = 0\ T_{0A} = 1, T_{1T} = 0, T_{2T} = 0,\ \ T_{3T} = 0\ , \quad (13)$$

$$\begin{cases} T_{0C} = 0, T_{1C} = 0, T_{2C} = 0,\ \ T_{3C} = 0 \\ T_{0A} = 1, T_{1T} = 0, T_{2T} = 0,\ \ T_{3T} = 0\ , \end{cases} \quad (13)$$

it follows that the facility according to the existing technology is in normal condition, and according to alternative technology in an emergency state; therefore, it is necessary to take measures.

In the third case from the combination on the monitor

$$\{\ T_{0C} = 1, T_{1c} = 0, T_{2C} = 0,\ \ T_{3C} = 0\ T_{0A} = 1, T_{1A} = 1, T_{2A} = 0,\ \ T_{3A} = 0 \quad (14)$$

$$\begin{cases} T_{0C} = 1, T_{1c} = 0, T_{2C} = 0,\ \ T_{3C} = 0 \\ T_{0A} = 1, T_{1A} = 1, T_{2A} = 0,\ \ T_{3A} = 0 \end{cases} \quad (14)$$

it follows that according to both technologies, the facility has gone into the latent period of an emergency state, so it is necessary to take measures.

where $T_{0C}, T_{1C}, T_{2C}, T_{3C}$ are indicators of technical condition of the facility according to the existing technology in time periods T_0, T_1, T_2, T_3;

where $T_{0A}, T_{1A}, T_{2A}, T_{3A}$ are indicators of technical condition of the facility according to alternative technology in time periods T_0, T_1, T_2, T_3.

CONCLUSION

Control systems of technical facilities do not provide monitoring and control of the beginning of the latent period of accidents. During this period the dynamics of operation of technological processes and the spectrum of analyzed signals change. In these cases, to ensure accident-free operation of technical facilities, it is necessary to control the beginning and dynamics of accident development. In vibration diagnostics the adequacy of the control result depends to a great extent on fulfillment of this condition. In this period noise has a significant diagnostic potential. To ensure the adequacy of vibration control results it is advisable to use it. Ignoring the noise as a carrier of diagnostic information makes it difficult to ensure the effectiveness of vibration control in practice. If this condition is fulfilled, it is possible to determine reliably enough the initial stages of malfunctions, which usually precede catastrophic accidents, based on the estimates of several informative attributes obtained from the analysis of various vibration signals.

In the proposed variant of control at the considered facilities, at the beginning of the latent period of accidents, the operator's monitor will reflect information about the beginning of the latent period of the emergency state of the facility at the beginning of the latent period of accidents. Normally, when the facility is in normal technical condition both lines of the operator's monitor will reflect the results of both the conventional technology and the proposed alternative technology. In case of suspicious results of the control only on the first line on the monitor will show a warning about the possibility of an emergency situation. In case of emergency state of the object both lines on the monitor will have information about it. In this way, the operating personnel will receive in advance information from both the existing and alternative technology and will be able to enhance the accident-free operation of the facility.

REFERENCES

[1] T. Aliev, Noise control of the Beginning and Development Dynamics of Accidents, Springer, 2019, 201 p.

[2] Aliev TA, Alizada TA, Rzayeva NE et al (2017) Noise technologies and systems for monitoring the beginning of the latent period of accidents on fixed platforms. Mechanical Systems and Signal Processing 87:111-123. doi:10.1016/j.ymssp.2016.10.014

[3] T.A. Aliyev, T.A. Alizada, N.E. Rzayeva. Noise technologies and systems for monitorinf the beginning of the latent period of accidents on fixed platforms Mechanical Systems and Signal Processing, 2017, vol.87, pp.111-123.

[4] T. Aliyev, A.H.Rzayev, G.A.Guluyev, T.A.Alizada, N.E.Rzayeva. Robust technology and system for management of sucker rod pumping units in oil wells. Mechanical Systems and Signal Processing, 2018, Vol. 99 (15), pp.47-56 (WoS - 4,116).

[5] Aliev T.A., A Rzayev.A.H., Guluyev G.A. et al. Robust technology and system for management of sucker rod pumping units in oil wells, Mechanical Systems and Signal Processing (2018) 99:47-56. https://doi.org/10.1016/j.ymssp.2017.06.010

[6] Intelligent Seismic-Acoustic System for identifying the Area of the Focus of an Expected Earthquake. Earthquakes tectonics. Hazard and risk mitigation, Edited by Taher Zouaghi., Published by in Tech, Janeza Trdine 9, 51000 Rijeka, Croatia, The Editor(s) and the Author(s), İngilis-dilində.2017, pp. 293-315. T.A. Aliev

[7] Bendat J.S., Piersol A.G. Random Data, Analysis & Measurement Procedures, Wiley, New York, 2000

[8] R.A. Collacott. Mechanical Fault Diagnosis and condition monitoring. 1977. 506 p.

[9] Popkovich G.S., Kuzmin A.A. Automation of water supply and sewerage systems. — Moscow: Stroyizdat, 1983. 151 p. (in Russian)

[10] Yakovlev S.V., Karelin Y.A., Laskov Yu.M., Kalitsun V.I. Water drainage and wastewater treatment. Moscow: Stroyizdat, 1996. — 591 p. (in Russian)

[11] Muzaffer Metin and Rahmi Guclu. Rail Vehicle Vibrations Control Using Parameters Adaptive PID Controller. Mathematical Problems in Engineering, Hindawi, 2014, pp 1-10.

[12] C.C. Lin, J.F. Wang and B.L. Chen. Train-Induced Vibration Control of High-Speed Railway Bridges Equipped with Multiple Tuned Mass Dampers. Journal of Bridge Engineering, 2005. 10(4), pp. 398-414.

Studying Propagation of Nonstationary Dynamical Waves in Inhomogeneous Layered Media

Nabi Kurbanov
Professor of the Department of Mathematical Analysis and Differential Equations
Sumgait State University,
Sumgait, Azerbaijan
nabi.qurbanov@sdu.edu.az
qurbanov53@mail.ru

Vusala Babajanova
Associate Professor of Mathematical Analysis and Differential Equations
Sumgait State University,
Sumgait, Azerbaijan
vusala.babacanova@sdu.edu.az

Kamala Agamaliyeva
Teacher of Mathematical Analysis and Differential Equations
Sumgait State University,
Sumgait, Azerbaijan
kamala.aqamaliyeva@sdu.edu.az

Abstract: Analytical methodology for solving boundary value problems for two-layered inhomogeneous viscoelastic semi-spaces is developed and exposure of external in influences on medium stability on the constructed mathematical models, is studied. The form of nonstationary surface phenomena corresponding to force boundary conditions is determined. It is obtained that the rate of attenuation of wave that are not attenuated in an elastic layered medium, turns out to be less than the rate of attenuation of waves in areas of non-transmission depending on viscoelastic properties of materials.

Keywords: inhomogeneity, lamination, viscoelastic, Laplace operator, original, image, iterated kernels.

INTRODUCTION

Recently, considerable attention was paid to the study of the regularities of interaction of wave processes of different medium. This is related with the fact that many problems of nature relate to the solution of the problems of seismology and seismic stability of engineering systems, that leads to studding wave processes allowing for inhomogeneity, laminarity and rheological properties of the medium. There are many works devoted to this problem [1, 2, 3].

PROBLEM STATEMENT

Let us assume that in a rectangular coordinate system $OXYZ$, a two-layer heterogeneous viscoelastic medium is located on the surface of a half-space. The coordinate axes X and Z lie on the surface of the first layer, while the y-axis is directed into the depth of the half-space. The first layer occupies the region $0 \le y \le l_1$, $-\infty < x, z < +\infty$, the second layer occupies the region $l_1 \le y \le l_2$, $-\infty < x, y < +\infty$; and the half-space occupies the region $y > l_2$, $-\infty < x, y < +\infty$.

Suppose that a dynamic load $\sigma_1 = g(t)$ applied at time $t = 0$ to the surface $y = 0$ of the first layer (for $y \ge 0$), which was initially at rest. We aim to determine the wave state of the medium at subsequent moments in time.

Mathematically, the problem reduces to solving the equation of motion [1, 2, 3].

$$\frac{\partial \sigma_i(y,t)}{\partial x} = \rho_i \frac{\partial^2 v_i(x,t)}{\partial t^2} \qquad (1)$$

under the following boundary conditions:

$$\sigma_1 = g(t) \; for \; y = 0 \qquad (2)$$

$$V_2(y,t) = 0 \; for \; y = l_2 \qquad (3)$$

The boundary conditions are zero

$$V_i(y,t) = \frac{\partial v_i(y,t)}{\partial t} = 0 \; for \; t = 0 \qquad (4)$$

The contact conditions are accepted in the form:

$$\sigma_1(y,t) = \sigma_2(y,t) \; for \; y = l_1$$

$$V_1(y,t) = V_2(y,t) \; for \; y = l_1 \qquad (5)$$

We determine the defining relations and inhomogeneity of the medium by the following formulas:

$$\sigma_1(y,t) = \int_0^t [R^{(1)}(t-\tau, y) + \frac{2}{3} R^{(i)}(t-\tau, y)] d\left(\frac{\partial v_i(y,t)}{\partial y}\right) \qquad (6)$$

$$\rho_i(y) = \rho_i^0 (1+ay)^m \quad R_i(y,t) = R^i(t)(1+ay)^m \qquad (7)$$

where $i = 1,2; j = 1,2;$ $\sigma_i(y,t)$ is a stress, $\rho_i^0 - const$, is medium density, $V(y,t)$ is displacement, l_1 and l_2 are layer thicknesses, a and m are constants, $R_i^{(1)}(y,t)$ and $R_i^{(2)}(y,t)$ are volume and shear relaxation functions, $g(t)$ is the given external load.

Note that for $i = 1$ all the relations belong to the first layer, for $i = 2$,to the second layer. As can be seen the solution of the problem is led to the solution of equation (2) under the conditions (2)-(7).

Applying the Laplace transform in time t to the equation (1) allowing for (4), (6) and (7) we obtain:

$$\frac{d^2 \underline{V_i}(z,p)}{\partial z^2} + \frac{k}{z} \frac{d\underline{V_i}(z,p)}{\partial z} - \frac{\beta_i^2}{a^2} \underline{V}(z,p) = 0 \qquad (8)$$

where:

$$z = 1 + ay,$$

$$\beta_i^2(p) = \frac{p^2}{C_i^2(p)};$$

$$C_i^2(p) = \frac{pR_i^{(1)}(p) + \frac{2}{3} p\underline{R_i^{(1)}}(p)}{\rho_i^0}$$

p is a parameter of Laplace , $\underline{V_i}(z,p)$ is the image of the function $\underline{V_i}(y,t)$, $\underline{R_i^{(1)}}(p)$ and $R_i^2(p)$ are the images of the functions $R_i^{(1)}(t)$, and $R_i^{(2)}(t)$ respectively. The solution of the equation (8), satisfying conditions (2), (3) and (5) has the form:

$$\underline{V_1}(z,p)=\frac{z^y}{\Delta}\left[\Delta_1 K_\nu\left(\frac{\beta_1 z}{a}\right)+\Delta_2 I_\nu\left(\frac{\beta_1 z}{a}\right)\right]$$

$$\underline{V_2}(z,p)=\frac{z^y}{\Delta}\left[\Delta_3 K_\nu\left(\frac{\beta_2 z}{a}\right)+\Delta_4 I_\nu\left(\frac{\beta_2 z}{a}\right)\right] \quad (9)$$

where

$$\Delta=\left[I_{\nu-1}\left(\frac{\beta_1}{a}\right)K_{\nu-1}\left(\frac{\beta_1 z_1}{a}\right)-K_{\nu-1}\left(\frac{\beta_1}{a}\right)I_{\nu-1}\left(\frac{\beta_1 z_1}{a}\right)\right]\times$$

$$\times\left[I_\nu\left(\frac{\beta_2 z_2}{a}\right)K_\nu\left(\frac{\beta_2 z_1}{a}\right)-K_\nu\left(\frac{\beta_2 z_2}{a}\right)I_\nu\left(\frac{\beta_2 z_1}{a}\right)\right]-$$

$$-\alpha\left[K_{\nu-1}\left(\frac{\beta_1}{a}\right)I_n\left(\frac{\beta_1 z_1}{a}\right)+I_{\nu-1}\left(\frac{\beta_1}{a}\right)K_\nu\left(\frac{\beta_1 z_1}{a}\right)\right]\times$$

$$\times\left[I_\nu\left(\frac{\beta_2 z_2}{a}\right)K_{\nu-1}\left(\frac{\beta_2 z_1}{a}\right)+K_\nu\left(\frac{\beta_2 z_2}{a}\right)I_{\nu-1}\left(\frac{\beta_2 z_1}{a}\right)\right];$$

$$\Delta_1=\frac{\underline{g}(p)}{p\underline{R_1}^{(1)}+\frac{2}{3}p\underline{R_2}^{(1)}}\cdot\frac{1}{\beta_1}\{\alpha I_\nu\left(\frac{\beta_1 z_1}{a}\right)\times$$

$$\times\left[K_{\nu-1}\left(\frac{\beta_2 z_1}{a}\right)I_\nu\left(\frac{\beta_2 z_2}{a}\right)+K_\nu\left(\frac{\beta_2 z_2}{a}\right)I_{\nu-1}\left(\frac{\beta_2 z_1}{a}\right)\right]+$$

$$+I_{\nu-1}\left(\frac{\beta_1 z_1}{a}\right)[I_\nu\left(\frac{\beta_2 z_2}{a}\right)K_\nu\left(\frac{\beta_2 z_1}{a}\right)-$$

$$-K_\nu\left(\frac{\beta_2 z_2}{a}\right)I_\nu\left(\frac{\beta_2 z_1}{a}\right)]\}$$

$$\Delta_2=\frac{\underline{g}(p)}{p\underline{R_1}^{(1)}+\frac{2}{3}p\underline{R_2}^{(1)}}\cdot\frac{1}{\beta_1}\{K_{\nu-1}\left(\frac{\beta_1 z_1}{a}\right)\times$$

$$\times\left[I_\nu\left(\frac{\beta_2 z_2}{a}\right)K_\nu\left(\frac{\beta_2 z_1}{a}\right)-I_\nu\left(\frac{\beta_2 z_1}{a}\right)K_{\nu-1}\left(\frac{\beta_2 z_2}{a}\right)\right]-$$

$$-\alpha K_\nu\left(\frac{\beta_1 z_1}{a}\right)[K_\nu\left(\frac{\beta_2 z_2}{a}\right)I_{\nu-1}\left(\frac{\beta_2 z_1}{a}\right)+$$

$$+I_\nu\left(\frac{\beta_2 z_1}{a}\right)K_{\nu-1}\left(\frac{\beta_2 z_1}{a}\right)]\}$$

$$\Delta_3=\frac{\underline{g}(p)}{p\underline{R_1}^{(1)}+\frac{2}{3}p\underline{R_2}^{(1)}}\cdot\frac{1}{\beta_1}I_\nu\left(\frac{\beta_2 z_2}{a}\right)[K_\nu\left(\frac{\beta_1 z_1}{a}\right)\times$$

$$\times I_{\nu-1}\left(\frac{\beta_1 z_1}{a}\right)+I_\nu\left(\frac{\beta_1 z_1}{a}\right)K_{\nu-1}\left(\frac{\beta_1 z_1}{a}\right)]$$

$$\Delta_4=\frac{\underline{g}(p)}{p\underline{R_1}^{(1)}+\frac{2}{3}p\underline{R_2}^{(1)}}\cdot\frac{1}{\beta_1}K_\nu\left(\frac{\beta_2 z_2}{a}\right)[I_\nu\left(\frac{\beta_1 z_1}{a}\right)\times$$

$$\times K_{\nu-1}\left(\frac{\beta_1 z_1}{a}\right)+K_\nu\left(\frac{\beta_1 z_1}{a}\right)I_{\nu-1}\left(\frac{\beta_1 z_1}{a}\right)]$$

where $I_\nu(z)$ and $K_\nu(z)$ are cylindrical Bessel functions [1, 4].

$$\gamma=\frac{1-m}{2}$$

$$\alpha=\frac{p\underline{R_1}^{(2)}+\frac{2}{3}p\underline{R_2}^{(2)}}{p\underline{R_1}^{(1)}+\frac{2}{3}p\underline{R_2}^{(1)}}\cdot\frac{\underline{\beta_2}}{\beta_1};$$

$$z_1=1+al_1;$$

$$z_2=1+al_2$$

By means of expansion of cylindrical functions, for large values of arguments we obtain: [1]

$$\underline{V_1}(z,p)=\frac{\underline{g}(p)z^{\gamma-\frac{1}{2}}}{p\rho_1^0 C_1}\sqrt{\frac{\rho_1^{(1)}(0)}{pR_1^{(1)}+\frac{2}{3}pR_2^{(1)}}}\times$$

$$\times\sum_{n=0}^{\infty}\theta^n\{exp\left[-\frac{p(z+2nl_1)}{a}\sqrt{\frac{\rho_1^{(1)}(0)}{pR_1^{(1)}+\frac{2}{3}p\underline{R_2}^{(1)}}}\right]+$$

$$+\theta exp\left[-\frac{p\big((2n+1)l_1-z\big)}{a}\sqrt{\frac{\rho_1^{(1)}(0)}{p\underline{R_1}^{(1)}+\frac{2}{3}pR_2^{(1)}}}\right]\}$$

$$\underline{V_2}(z,p)=\frac{aU_1(l_1,p)}{2\gamma\sqrt{z_1 z_2}}\frac{1}{p}\sqrt{\frac{pR_1^{(2)}+\frac{2}{3}p\underline{R_2}^{(2)}}{\rho_0^{(2)}}}\times$$

$$\times\{exp\left[-\frac{p(z-z_2)}{a}\sqrt{\frac{\rho_0^{(2)}}{p\underline{R_1}^{(2)}+\frac{2}{3}p\underline{R_2}^{(2)}}}\right]-$$

$$-exp\left[-\frac{p(z_2-z)}{a}\sqrt{\frac{\rho_0^{(2)}}{p\underline{R_1}^{(2)}+\frac{2}{3}pR_2^{(2)}}}\right]\}$$

To compute the original solution, let us assume that the Poisson's γ is constant. Then the functions $R_1^{(i)}(t)$ and $R_2^{(i)}(t)$ are proportional:

$$R_1^{(i)}(t)=\mu^{(i)}R^{(i)}(t) \text{ where } \mu^{(i)}=\frac{1+\gamma_i}{3(1-2\gamma_i)}$$

In this case, the solution takes the form:

$$\underline{V_1}(z,p)=\frac{\underline{g}(p)}{p\rho_1^0 C_1}\sqrt{\frac{R_1^{(1)}(0)}{p\underline{R_1}^{(1)}}}\times$$

$$\times\sum_{n=0}^{\infty}\theta^n[exp\left(-\frac{p}{C_1}\frac{z+2nl_1}{a}\sqrt{\frac{R_1^{(1)}(0)}{p\underline{R_1}^{(1)}}}\right)+$$

$$+\theta exp\left(-\frac{p}{C_1}\frac{2(n+1)-z}{a}\sqrt{\frac{R_1^{(1)}(0)}{p\underline{R_1}^{(1)}}}\right)]$$

$$\underline{V_2}(z,p)=\frac{a\underline{V_1}(l_1,p)}{2\gamma p\sqrt{z_1 z_2}}\sqrt{\frac{R^{(2)}(0)}{pR^{(2)}(p)}}\times$$

$$\times \left[exp\left(-\frac{p}{C_2}\frac{z-z_2}{a}\sqrt{\frac{R^{(2)}(0)}{p\underline{R}^{(2)}}}\right)-\right.$$

$$\left.-exp\left(-\frac{p}{C_2}\frac{z_2-z}{a}\sqrt{\frac{R^{(2)}(0)}{pR^{(2)}(p)}}\right)\right] \qquad (10)$$

where

$$C_i=\sqrt{\frac{R_1^{(i)}(0)+\frac{2}{3}R^{(i)}(0)}{\rho_i^0}};$$

$$\theta=\frac{\beta_2\gamma z_1+(\gamma y-1)a\mu}{\beta_2\gamma z_1-(\gamma y-1)a\mu};$$

$$\gamma=\frac{z_2-z_1}{z_1};$$

$$\frac{p\underline{R}_1^{(1)}}{pR^{(i)}}=\mu=const.$$

Hence it is seen that the solution of the stated problem is reduced to calculating the inverse Laplace transformation of the function of the form:

$$\underline{\varphi}(z,p)=-\frac{\underline{g}(p)}{p}\sqrt{\frac{R(0)}{p\underline{R}(p)}}exp\left(-\frac{p}{C}\frac{r}{a}\sqrt{\frac{R(0)}{p\underline{R}}}\right)$$

It is known that $p\underline{R}=R_0\left(i-\varepsilon\underline{\Gamma}\right)$. Taking this into account, representing the exponential function in the form of a Fourier integral, we obtain:

$$\underline{\varphi}(z,p)=-\frac{2\underline{g}(p)}{p\pi}\int_0^\infty \frac{cos\left(\frac{\lambda z}{C}\right)d\lambda}{p^2+\lambda^2-\varepsilon\lambda^2\underline{\Gamma}(p)}$$

Taking into account $\left|\frac{\varepsilon\lambda^2\underline{\Gamma}(p)}{p^2+\lambda^2}\right|<1$ we expand the integrand function in series:

$$\varphi(z,p)=-\frac{2\underline{g}(p)}{\pi p}\int_0^\infty \left[\frac{1}{p^2+\lambda^2}+\frac{\varepsilon\lambda^2\underline{\Gamma}(p)}{(p^2+\lambda^2)^2}+\cdots+\right.$$

$$\left.+\frac{\left(\varepsilon\lambda^2\underline{\Gamma}(p)\right)^m}{(p^2+\lambda^2)^{m+1}}\right]cos\left(\frac{\lambda z}{C}\right)d\lambda$$

Calculating the integrals, we obtain:

$$\varphi(z,p)=-\frac{\underline{g}(p)}{p}e^{-\frac{pz}{C}}\left[\frac{1}{p}+\frac{1}{2^2}\left(\frac{z}{C}\right)\frac{\varepsilon\underline{\Gamma}}{p^2}+\cdots+\right.$$

$$\left.+\frac{p^{-(m+1)}}{2^m m!}\left(\frac{z}{C}\right)^m\left(\varepsilon\underline{\Gamma}(p)\right)^m+\cdots\right]$$

Passing to the space of originals, we find:

$$\varphi(z,p)=-\int_0^t g(\tau)d\tau*\left[H\left(t-\frac{z}{C}\right)+\frac{1}{2^2}\left(\frac{z}{C}\right)\varepsilon\Gamma(t)+\right.$$

$$\left.\cdots+\right.$$

$$\left.+\frac{1}{2^m m!}\varepsilon^m\Gamma^{(m)}(t)+\cdots\right]$$

where

$$\Gamma_1(t)=\Gamma(t),$$

$$\Gamma^{(2)}(t)=\int_0^t \Gamma(t-\tau)\Gamma_1(\tau)d\tau,\ ...,$$

$$\Gamma^{(m)}(t)=\int_0^t \Gamma(t-\tau)\Gamma^{m-1}(\tau)d\tau$$

then the solution is determined by the formulas:

$$V_1(z,t)=-\frac{1}{\rho_1^0 C_1}\int_0^t g(\tau)d\tau$$

$$*\sum_{n=1}^{\infty}\theta^n[\varphi(z+2nl_1,t)+$$

$$+\varphi[2(n+1)-z;t]]$$

$$V_2(z,t)=\frac{a}{2\gamma\sqrt{z_1z_2}}\int_0^t V_1(l,t)[\varphi(z-z_2,t)-\varphi(z_2$$

$$-z,t)]$$

CONCLUSION

1. A problem on propagation of nonstationary waves in inhomogeneous viscoelastic laminated semi-spaces for arbitrary hereditary functions at low viscosity was solved by the method of Laplace integral transforms.

2. The influence of viscosity and homogeneity of materials on wave propagation when the density and relaxation function depend on coordinates directed into the subspace was studied.

REFERENCES

[1] Ilyasov M.Kh. nonstationary viscoelastic waves // - Baku "Elm", -2011.

[2] Kurbanov N.T. , Babajanova V.G. Studying the reaction of inhomogenous viscoelastic bodies to nonstationary extrenal effects. // Proc.of International confrence of mathetaticans.Russia-Alshuta- -2009.

[3] Fillipov I.G., Bahramov B.M. Waves in homogenous and inhomogenous media. // Tashkent: FAN, 1978.

[4] Kurbanov, N.T., Babajanova, V.G. An investigation of the longit udinal fluctuation of viscoelastic cores // -New York: USA, Life Sience Journal, -2014. 11(9). –p. 557-561

Section 3: Problems and Methods of Decision Making

Aspects of Improving Decision-Making Mechanisms for Technological Economic Development Processes on the Industry 4.0 Platform

Alovsat Aliyev
Institute of Information Technology, Baku, Azerbaijan
alovsat_qaraca@mail.ru
orcid.org/0000-0002-1174-8036

Roza Shahverdiyeva
Institute of Information Technology, Baku, Azerbaijan
Azerbaijan Technical University
shahverdiyevar@gmail.com
orcid.org/0000-0003-0842-7300

***Abstract*—Improving decision-making mechanisms for technological economic development processes on the Industry 4.0 platform is one of the pressing issues of the modern era. The article dedicated to this topic notes the importance of such main goals as technological development, digitalization, cyber security, and the application of Artificial Intelligence in the country. A comprehensive analysis of scientific and research works in the field of improving decision-making mechanisms by applying digital technologies to increase the efficiency of the new generation technological economy sectors was conducted. The main directions of the transformation of the technological economic development process on the Industry 4.0 platform were identified. Some tools for processing management information for decision-making in the digital economy were provided, and some approaches to the decision-making process were analyzed. The features of the application of decision-making methods in managing technological economic development processes were investigated. The structural scheme of multi-criteria decision-making methods was developed. A conceptual model for improving decision-making mechanisms in the technological economic development process was proposed, and some recommendations were given.**

Keywords—digital technologies and transformations, Industry 4.0 platform, technological economic development process, decision-making mechanisms, multi-criteria decision-making methods

INTRODUCTION

The Industry 4.0 platform has formed a new modern stage of the technological development process in the economy. Thus, as a result of digitalization and the application of innovative technologies, traditional economic models are transformed, new business processes are automated, and more effective management mechanisms are created. The application of digital technologies has led to increased efficiency in the economy, reduced costs, and improved product quality. This stage is characterized by the integration of advanced technologies such as digitalization, automation, and Artificial Intelligence into industrial and manufacturing sectors.

Through technologies such as Artificial Intelligence, Internet of Things (IoT), Big Data, Blockchain, Cloud, 5G, cyber-physical systems, Machine Learning, etc., production processes have become more intelligent and adaptive, and effective decision-making opportunities have emerged in real time.

The technological economic development process formed on the Industry 4.0 platform requires the management of numerous and complex information flows, rapid decision-making, and flexible adaptation to changing market conditions. In these circumstances, the issue of improving decision-making mechanisms is of particular importance. Modern technologies and expert systems, by increasing the intellectual level of management processes, allow for risk-free, accurate and strategic decision-making. This restructuring of decision-making mechanisms increases the competitiveness of enterprises, increases the efficiency of resource use and facilitates the achievement of strategic goals.

Industry 4.0 technologies create great opportunities in this area. It is possible to make more accurate and quick decisions based on data collected in real time. Thus, predictive analytics based on Artificial Intelligence allows for the early detection of equipment failures and the taking of appropriate measures. This also reduces costs and prevents production stops. Such issues indicate that improving decision-making mechanisms is very important. It should be noted that the formation and technological development of new generation technological economy sectors are also important tasks for Azerbaijan. These issues are reflected in many State Programs and documents of the country. These program documents include 1)Increasing the competitiveness of the country's economy; 2)Creating favorable conditions for the application of Artificial Intelligence; 3)Preparing a qualified workforce to work in the Artificial Intelligence industry; 4)It is intended to achieve goals such as widespread promotion of the benefits of Artificial Intelligence to all members of society (https://president.az/az/articles/view/68364; https://president.az/az/articles/view/67938; https://president.az/az/articles/view/60949; https://president.az/az/articles/view/63979; https://president.az/az/articles/view/50474).

In the indicated State documents, taking into account the great importance of technological development in the field of both industrial development and the civil and military industrial complex, its main goals such as technological development, digitalization, cybersecurity, and the application of Artificial Intelligence have been set as a task for state institutions and society, becoming signs of everyday life in the country.

Issues of sustainability and sustainability in the formation of a new generation of technological economic development form the basis of economic strategies. Governments are trying to integrate the development of new generation technologies with the principles of green, innovative, digital, cognitive technologies and circular economy, ultimately balancing economic growth with environmental control [1].

Therefore, the formation of appropriate methodological apparatus, approaches, and implementation mechanisms for improving decision-making mechanisms on technological economic development processes on the Industry 4.0 platform is an urgent problem. Special attention to solving such a problem and conducting relevant research are of high relevance and great importance both from a scientific and practical point of view.

Problem Statement And Analysis Of Related Scientific Research Work

It is known that in the current era of rapid development of digitalization, innovations, science, and high technologies, the economy is developing rapidly in unity with other areas. Various sectors of the new generation national digital economy are also rapidly forming and developing. In such a situation, determining the aspects of improving decision-making mechanisms for technological economic development processes on the Industry 4.0 platform, as well as developing a conceptual methodological approach to the formation of a decision-making model for the technological economic development process with the application of digital technologies and providing relevant recommendations is a very relevant and important issue.

The Industry 4.0 platform has created new opportunities and conditions for improving decision-making mechanisms for technological economic development processes. Many scientists and researchers have also conducted certain studies on this issue [2-8]. The requirements for solving other relevant problems in those publications should be taken into account in the improvement of decision-making mechanisms in the process of technological economic development in a comprehensive manner. Therefore, although the analysis of the above-mentioned problems has been studied to some extent in the scientific works of many scientists and specialists, there is still a need for a deeper and more comprehensive study of that process, as well as the development of various approaches to solving the problems that have arisen. Therefore, in relation to the development of the problem under consideration, it should be noted that [2] assessed the impact of using big data analytics on the quality of decision-making in organizations, [4] analyzed multi-criteria decision-making methods for managing business process data, [6] comprehensively analyzed the review of scientific literature on the application of multi-criteria decision-making methods in business analytics, [7] fuzzy large-scale group decision-making based on big data in a circular economy (Large-Scale Group Decision Making-LSGDM), [8] database modeling in Machine Learning theories of economic decision-making, [9] opportunities and problems of intelligent systems in the perspective of cognitive decision-making, [10] comprehensively analyzed behavioral characteristics in financial decision-making and from classical finance to behavioral finance perspectives, [11] hybrid fuzzy COPRAS-based criterion method for multi-criteria decision-making, [12] analyzed behavioral economics for decision support system researchers, [13] analyzed the application of the Bayesian approach to decision-making in modern economics, [14] analyzed the application of managerial economics in decision-making in enterprise management, etc. The requirements and features specified in the analyzed sources of this type play an important role in improving the relevant decision-making mechanisms. Therefore, it is necessary to study the process in more depth and develop different approaches to solving the problems that have arisen.

Transformation Of The Technological Economic Development Process In The Industry 4.0 Platform

Industry 4.0 is a new industrial platform formed by the application of digital technologies such as Artificial Intelligence, Big Data, Cloud technologies, IoT, etc., in the industrial sector. This approach involves not only the application of technologies but also a change in economic development models, that is, their dynamic transformation. The application of Industry 4.0 technologies ensures the transition of economic development to a new stage. This transformation requires both technological and structural and institutional changes.

The main directions of the transformation of the technological economic development process on the Industry 4.0 platform can be attributed to: 1)Automation and digitalization of production, 2)Creation of new economic models and digital ecosystems, 3)Formation of smart factories and enterprises, 4)Transformation of the labor market, 5)Integration of real-time decision-making with cloud technologies, 6)Rapid application of digital innovative technologies, 7)More effective use of resources, 8)Improvement of legislative mechanisms, 9)Reduction of cyber security risks and information protection, 10)Increasing the role of human capital, Increasing the need for personnel with digital and analytical skills, 11)Increasing human and machine cooperation, 12)Creation of global value chain models, etc.

Regarding the characteristics of the impact of this type of new generation digital technologies and transformation on economic development, it can be noted that these technologies have a very serious impact on the structure, dynamics, and quality of economic development. New generation technologies such as Artificial Intelligence, Internet of Things (IoT), Big Data, Blockchain, 5G, and cloud technologies not only increase the efficiency of the economy but also drive the formation of new economic models. Economic value is created through digital platforms due to changes in consumer behavior. The demand for digital skills is constantly increasing. The process of strengthening regional and global integration is taking place.

New generation digital technologies and transformation are changing the speed, quality, and inclusiveness of economic development, making it more agile, digitalized, and innovative. For the effective application of these technologies, the development of digital transformation strategies and the development of human capital in that direction are of crucial importance.

Processing Management Information For Decision Making In The Digital Economy

Management information is analyzed and purposefully structured information used by management at various

levels to make strategic and operational decisions. The transformation of theoretical provisions on management decision-making in the digital economy is carried out through the integration of new digital technologies and data analysis methods into decision-making processes. The main tools for improving the information-analytical support of management in the new environment include: 1)Big Data processing technologies; 2)business analytics and machine learning; 3)digital platforms, 4)programs and services; 5)digital risk management, etc.

The adequacy of a specific model or technology for preparing a management decision in accordance with the conditions of the digital environment should be assessed, considering the specific management situation and existing resource provision. Some models allow to increase efficiency in solving structured problems and to make decisions based on precise criteria. Other models are applied to decision-making in conditions of more complex and unstructured problems. It should be noted that each model has its own advantages and limitations. The choice of model should be based on the characteristics of a specific management situation [3]. The main tools for processing management information for decision-making in the digital economy can be presented as in Table 1.

TABLE 1. Some Tools For Processing Management İnformation For Decision-Making İn The Digital Economy

Tools	Importance in the Digital Economy
Big Data analysis	The digital economy generates vast amounts of data, the analysis of which is becoming an important tool for making management decisions.
Machine learning	Artificial intelligence allows us to process large amounts of data and predict outcomes, creating additional opportunities for making more informed and accurate decisions.
Artificial intelligence-based support systems in decision-making	Predictive analytics. Scenario modeling through machine learning
Cloud technologies and online systems	Ensuring uninterrupted access to data. Real-time capabilities of distributed databases
ERP and CRM systems	Integration of internal data. Analysis of customer data and its impact on strategic decisions
Digital platforms	Building smart connections. Opportunities for using smart management mechanisms. Promptly filling information gaps
Visualization. Tools for visually presenting and managing data	Visual decision support systems (real-time indicators, dashboards), clear and interactive presentation of information

The effectiveness of management decisions in the digital economy directly depends on the level of application of information technologies and the functionality of the applied analytical tools. Digital information tools for modern management have become a necessity, not an option. The advantages of these tools include: Fast decision-making, Real-time analysis and response, Reasoned decisions, Support with statistical and artificial intelligence-based models, Strategic relevance, and the key role of the obtained data in planning and development.

An important aspect in the development of decision-making tools in the digital economy is the application of "smart management mechanisms" based on effective methods of supporting decision-making. In the classical digital technology version of decision-making, a person makes the decision. Digital products provide appropriate recommendations. Then the results obtained are compared and analyzed with the proposals of the computer system.

This version also has disadvantages, such as 1)difficulties in building a management object model, 2)low level of protection against corruption, and 3)possible disruption of feedback mechanisms. In the case of "reverse decision-making", the computer system is at the center of the technology. The person acts as an observer and intervenes in the process only in emergency cases.

The effectiveness of the "reverse decision-making" technology depends on the competence and experience of the team of specialists who perform the monitoring function in the decision-making process.

In the conditions of digital transformation, the role of the manager in making management decisions changes. He acts not only as an organizer of production and economic and financial activities in the enterprise, but also as a coordinator who provides coordination and control over the activities of the staff. These changes in the management system of the enterprise in the conditions of the digital economy have led to the formation of a number of features in the application of methods in making management decisions [3].

The features of the methods of making management decisions in the conditions of the digital economy can be presented as follows: 1)The role of the manager is changing from an organizational leader to a communicative leader. 2)Organization of online interaction in conditions of full-time remote availability of the staff. 3)Formation of working groups without taking into account geographical restrictions. 4)Expansion of the possibilities of preliminary assessment of risks based on the processing of Big Data. 5)Multifunctionality, lack of rigid hierarchy, and expansion of horizontal relations. 6)Real-time management decision-making. 7)Increasing the role of digital skills of personnel, etc.

The application of digital technologies has a serious impact on management decision-making and the operating conditions of companies in general. The conducted analyses confirm the need for complex changes in the management system in order to achieve a positive effect from the widespread application of ICT and increase operational efficiency.

These features play a decisive role in clarifying the main features of decision-making mechanisms in the processes of new generation technological economic development. In accordance with them, the main directions of improving decision-making mechanisms can be attributed to: 1)Data-driven decision-making, 2)Real-time decision-making, 3)Artificial intelligence and automated decision-making mechanisms, 4)Use of cloud technologies and distributed systems, 5) Improvement of Multi-Criteria Decision Making (MCDM) methods, 6)Decision-making in integrated systems, 7)Decision-making under risk and uncertainty, 8)Decision-making through simulation and digital twins, 9)Solving cybersecurity and information protection issues in decision-making, etc.

Some Approaches To The Decision-Making Process

In general, some approaches to the decision-making process are a set of various theoretical and practical methods used to make correct and justified decisions in organizations and at the individual level. Decision-making approaches are based on understanding the problem, evaluating alternatives, and making the best choice. Decision-making approaches can be classified into several groups. First of all, the mainly uncertain Multi-Attribute Decision-Making (MADM) models and methods are noted. The Multi-Objective Decision-Making (MODM) approach can also be noted [4]. Apart from these, approaches such as 1) Multiple-Criteria Decision-Making, 2) Information Aggregation Operators (Fuzzy Sets Theory, Neutrosophic Theory), 3) Rough and Fuzzy-Rough Sets, 4) Neuro-fuzzy approach, 5)Fuzzy MADM dynamic approach, 6)Deterministic finite automatabased model, 7)Interval multiplicative preference relations, etc., can also be mentioned. Each group itself can be expressed in the form of various subgroups and methods [6].

Multi-attribute decision-making (MADM) methods consist of a large part of Multi-Criteria Decision-Making (MCDM) methods. The main characteristics of multi-attribute decision-making methods are that there is a limited number of alternatives in the decision-making problem. Here, a new structure is proposed by adding method classes based on interaction and utility. A brief description of that structure is given in Figure 1.

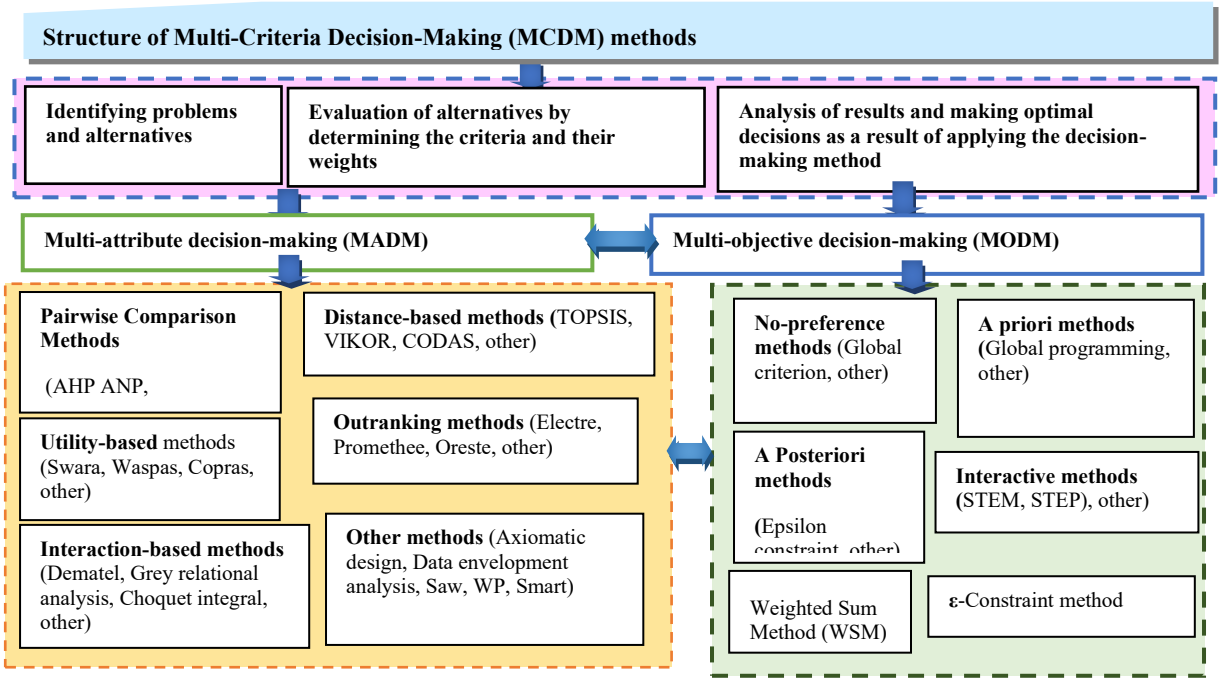


Fig. 1. Structure of Multi-Criteria Decision-Making (MCDM) methods *(compiled by the authors based on the analysis and systematization of scientific literature)*

Features Of Application Of Decision-Making Methods In Management Of Technological Economy Development Processes

The application of decision-making methods in technological digital economic sectors is based on a more information-based, flexible, technological, and intellectual basis than traditional management. This creates broad opportunities for reducing risks in management, efficient use of resources, and faster achievement of strategic goals.

With Big Data and real-time analytics, a transition to information-based decision-making is taking place in the conditions of direct integration of data into management. Traditional instinctive and experience-based approaches are replaced by analytical and statistically based decisions. Decisions can be made more objectively, quickly, and dynamically. These take place based on the application of Artificial Intelligence and Machine Learning methods. The application of scenario-based decision-making models is becoming more relevant. The scope of multi-criteria decision-making methods continues to expand. In this direction, intuitive and intellectual decision support systems are being created. Real-time decision-making capabilities are increasing. Special attention is paid to it due to the need for institutional and organizational coherence.

The level of application of predictive and prescriptive analytics is increasing and expanding.

Analysis of decision-making models and their characteristics shows that a decision-making model is a systematic approach based on information to make a choice between alternatives available in an organization or economic system. These models create additional opportunities for rational, analytical, and systematic decision-making in the management of technological and economic processes. In the application of decision-making methods, systematization based on data analysis and synthesis is used [3].

Conceptual Model Of Improving Decision-Making Mechanisms In The Technological Economic Development Process

In the modern era, the acceleration of technological progress and the penetration of digital transformation into economic systems necessitate a transition to a new stage.

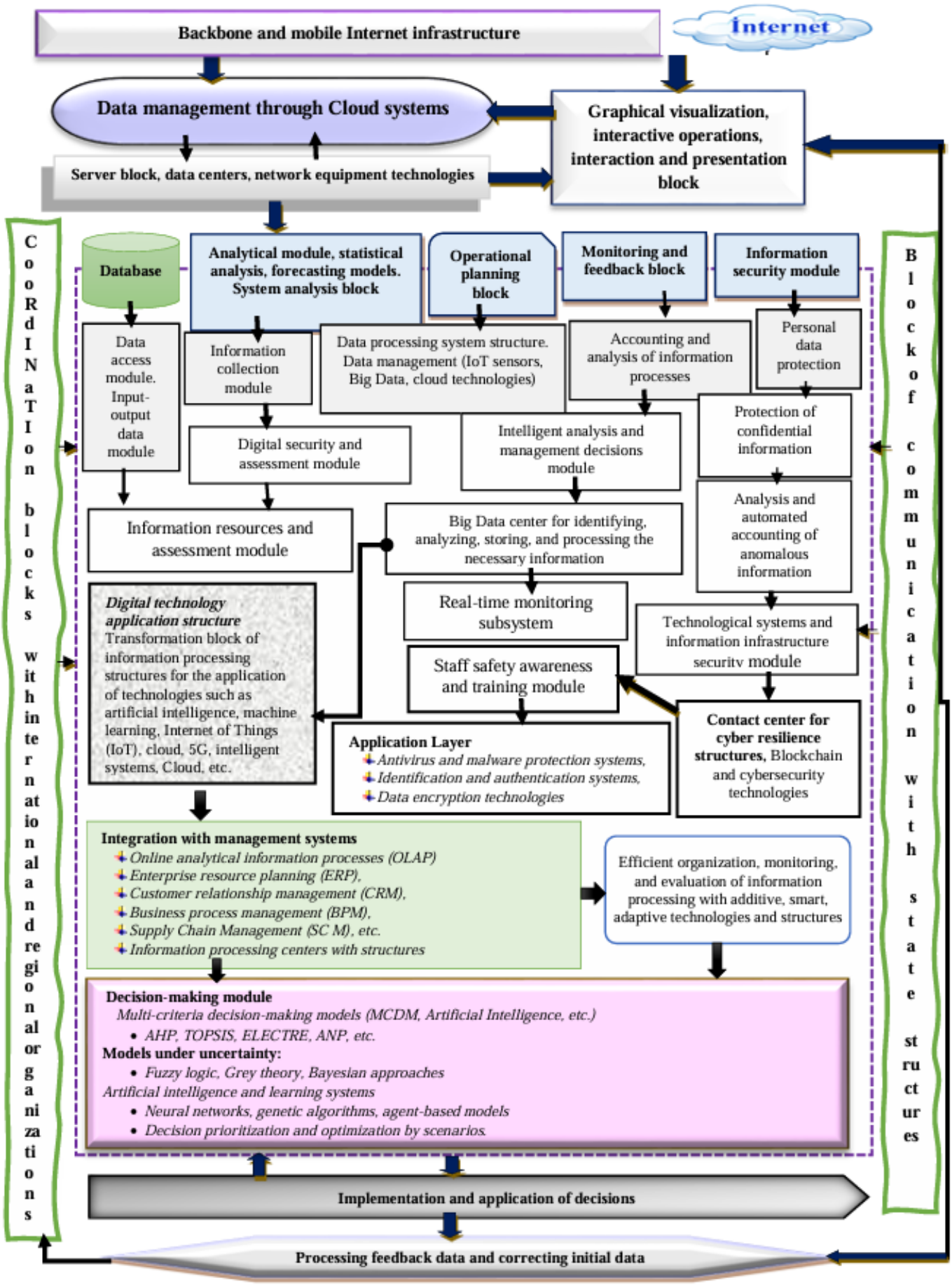


Fig.2. Conceptual model of improving decision-making mechanisms of the technological economic development process on the Industry 4.0 platform *(compiled by the authors)*

This stage is mainly based on an innovative technological platform called Industry 4.0. Industry 4.0 is a new generation production and management concept formed as a result of the integration of cyber-physical systems, Artificial Intelligence,

Big Data, IoT, and automated management technologies. This transformation necessitates the improvement of decision-making mechanisms not only in industrial sectors but also in all areas of economic development.

The decision-making process is carried out in a particularly complex and rapidly changing technological environment, in the presence of uncertainties, multiple alternatives, and various criteria. Since traditional decision-making approaches cannot adequately respond to the changing and digitized information flow, the application of new, more flexible and intelligent models in this area is becoming relevant. In this sense, the formation of intelligent decision-making systems that ensure effective management of technological economic development using the capabilities of the Industry 4.0 platform is of great scientific and practical importance. Therefore, we have developed an appropriate conceptual model to improve the decision-making mechanisms of the technological economic development process using the capabilities of the Industry 4.0 platform. Through this model, it is possible to ensure effective management (Figure 2).

The interaction of the conceptual coordination blocks of the proposed structural model can have a positive impact on improving the decision-making mechanisms of the technological economic development process on the Industry 4.0 platform in the conditions of digital transformation, increasing its level of application and the efficiency of decision-making processes. Since such a model has an open system nature, it reflects the progressive technological change proposed in state programs, digital development strategies, and perspective concepts, and creates conditions for the inclusion of other structural components in the model.

CONCLUSION

In the era of modern globalization and integration, the interaction of technological and economic development has led to fundamental changes in the activities of industrial enterprises. As a result of the widespread application of advanced technologies such as artificial intelligence, the Internet of Things, Big Data, cyber-physical systems, and automation, traditional decision-making mechanisms no longer meet modern realities. New technological realities require the establishment of more flexible, data-based, predictive, and multi-level decision-making systems for enterprises. The study found that the Industry 4.0 platform requires not only the improvement of production and technical operations, but also the transformation of management and decision-making systems. In this regard, enterprises should improve their decision-making structures in the relevant areas. Thus, the integration of rational and artificial intelligence-based models should be implemented. As a result of the application of intelligent algorithms in decision-making, a more flexible and risk-resistant management system is formed. The management of big data in technological economic processes, their rapid analysis, significantly increases the quality and speed of decision-making. The phased establishment of management systems ensures more systematic decision-making in line with the strategic goals of the enterprise. In terms of cybersecurity and information security, decision-making systems should be protected against digital risks, and critical information security should be ensured. Effective organization of human-computer cooperation in decision-making processes should be accompanied by personnel training that keeps pace with technological development.

Analyses show that improving decision-making mechanisms in technological economic systems requires restructuring the organizational structure of management and strategic planning. Enterprises should pay special attention to management flexibility, efficiency of information flows, and quality of decisions.

The creation of intelligent and multifunctional decision-making mechanisms for successful technological and economic development on the Industry 4.0 platform is one of the main priorities. Systematic steps to be taken in this direction will not only increase the competitiveness, flexibility, and sustainability of enterprises and structures but will also make a significant contribution to strengthening the innovation-based economic development of the digital economy and technological economic sectors.

REFERENCES

[1] A.G.Aliyev, R.O.Shahverdiyeva. Organizational problems of innovation activities and their solution mechanisms. Monograph, Baku, "Information Technologies", 2023, 532 p.

[2] Li L. et al. Evaluating the impact of Big Data analytics usage on the decision-making quality of organizations. Technological Forecasting & Social Change 2022, (175) 121355. https://doi.org/10.1016/j.techfore.2021.121355

[3] A.A.Vereshchaka. Teoreticheskiye i prakticheskiye aspekty prinyatiya upravlencheskikh resheniy v tsifrovoy ekonomike. Vestnik UU, 2024, T.34, vyp.3, str.422-427. DOI: 10.35634/2412-9593-2024-34-3-422-427

[4] K.Z.Edmundas. Multiple-Criteria Decision-Making techniques for business processes information management. Information 2019, 10, 4. doi:10.3390/info10010004

[5] A.G.Aliyev, R.O.Shahverdiyeva. Some problems of the formation of the new generation digital economy based on artificial intelligence technologies Informatica Economica journal, 2024, vol. 28, no. 3, pp.49-64.

[6] A.S.Yalcin et al. The use of multi-criteria decision-making methods in business analytics: A comprehensive literature review. Technological Forecasting & Social Change, 2022, (174), 121193. https://doi.org/10.1016/j.techfore.2021.121193

[7] L.Xuan. Big Data-driven fuzzy large-scale group decision making (LSGDM) in circular economy environment. Technological Forecasting & Social Change 2022, (175), 121285. https://doi.org/10.1016/j.techfore.2021.121285

[8] T.Tobias. Modelling dataset bias in machine-learned theories of economic decision-making. Nature Human Behaviour, 2024, volume 8, pp.679–691

[9] Z.Xu, X.Wang et al. Intelligent systems from a perspective of cognitive decision-making: Opportunities and challenges. Journal of Industrial Engineering and Engineering Management, 2025

[10] R.Kanapickienė. A comprehensive review of behavioral biases in financial decision-making: From classical finance to behavioral finance perspectives. Journal of Business Economics and Management, 2024, volume 25, Issue 5, pp. 1006–1029. https://doi.org/10.3846/jbem.2024.22314

[11] M.Narang et al. A hybrid fuzzy COPRAS-base-criterion method for multi-criteria decision making. Soft Computing, 2021, 25(2), pp.1-9

[12] D.Arnott, S.Gao. Behavioral economics for decision support systems researchers. Decision Support Systems, 2019, volume 122, 113063. https://doi.org/10.1016/j.dss.2019.05.003

[13] S.V.Zaytsev. Voprosy primeneniya Bayyesovskogo podkhoda v sovremennoy ekonomike dlya prinyatiya resheniy. Ekonomika i upravleniye: problemy, resheniya, 2024, №6, T.1, s.20–27. https://doi.org/10.36871/ek.up.p.r.2024.06.01.003

[14] V.SH.Rasumov, Ye.V.Novikova, M.S.Yusupova. Primeneniye upravlencheskoy ekonomiki v prinyatii resheniy po upravleniyu predpriyatiyem. Ekonomika i upravleniye: problemy, resheniya, 2025, №2, T.10, s.2-57. https://doi.org 10.36871/ ek.up-.p.r.2025.02.010.007

Section 4: Optimization and Optimal Control

On Determining Higher Coefficient of a Second Order Hyperbolic Equation by the Variational Method

Hamlet Guliyev
Doctor of Physical and Mathematical Sciences, Professor
Institute of Mathematics and Mechanics, Baku, Azerbaijan
hamletquliyev51@gmail.com

Vusala Nasibzadeh
Senior Teacher of Mathematical Analysis and Differential Equations
Sumgayit State University, Sumgayit
vusala.nasibzada@sdu.edu.az

***Abstract* -The paper deals with an inverse problem of determining a higher coefficient of a second order hyperbolic equation. This problem is reduced to an optimal control problem and the new problem is studied by the methods of optimal control theory. Is proved existence theorem for optimal control and obtained necessary condition of optimality in the form integral inequality.**

Keywords: inverse problem, higher coefficient, optimal control, existence theorem, necessary condition for optimality.

I. INTRODUCTION

In direct problems of theory of partial differential equations or in mathematical physics problems the functions that describe various physical phenomena as propagation of heat, sound, various vibrations, electromagnetic waves, etc. are sought. This time, the features of the medium under consideration or coefficients of equations are assumed to be known. However, just the features of medium in great majority of cases are unknown. Then there arise inverse problems in which on the information on the solution of the direct problem it is required to determine the coefficients of equations. As is known, these problems in many cases are ill-posed. But, at the same time, the desired coefficients of the equations characterize the medium under consideration. Therefore, solving inverse problems is very important both from a practical and theoretical point of view [1-3, 8-13].

II. PROBLEM STATEMENT

Let $\Omega-$ be a bounded domain in the space R^n with a smooth boundary $\Gamma,$ $T>0-$ be a given number, $Q=\{(x,t): x\in\Omega, t\in(0,T)\}-$ be a cylinder in R^{n+1}, $S=\{(x,t): x\in\Gamma, t\in(0,T)\}-$ be a lateral surface of the cylinder $Q.$

It is reguired to determine a pair of functions $(u(x,t),\upsilon(x))$ from the conditions

$$\frac{\partial^2 u}{\partial t^2}-\sum_{i=1}^{n}\frac{\partial}{\partial x_i}\left(\upsilon(x)\frac{\partial u}{\partial x_i}\right)+a_0(x)u=f(x,t),(x,t)\in Q, \tag{1}$$

$$u(x,0)=u_0(x),\frac{\partial u(x,0)}{\partial t}=u_1(x),x\in\Omega\quad u|_S=0, \tag{2}$$

$$\int_0^T K(x,t)u(x,t)dt=\varphi(x), \tag{3}$$

$\upsilon=\upsilon(x)\in V,$ where

$$V=\left\{\upsilon(x)\in W_2^1(\Omega): \nu_0\le\upsilon(x)\le\mu_0,\left|\frac{\partial\upsilon}{\partial x_i}\right|\le\mu_i,i=1,...,\right\}- \tag{4}$$

n *a.e. on* Ω

is a given set, $\nu_0,\mu_0,\mu_1,...,\mu_n-$ are the given positive numbers, $a_0(x)\ge 0,$ $a_0\in L_\infty(\Omega), f\in L_2(Q), u_0\in \overset{0}{W}{}_2^1(\Omega), u_1\in L_2(\Omega), K\in L_\infty(Q), \varphi\in L_2(\Omega)-$ are the given functions.

For the given function $\upsilon(x)$ the problem (1), (2) is a direct problem in the domain $Q,$ for the unknow function $\upsilon(x)$ the problem (1)-(4) is said to be an inverse problem to the problem (1), (2). Note that for each fixed function $\upsilon(x)\in V$ the solution of the boundary value problem (1), (2) understood as a generalized solution from the space $W_{2,0}^1(Q)$ [4].

Under the solution from $W_{2,0}^1(Q)$ of the boundary value problem (1), (2) for the given function $\upsilon\in V$ we will understand the function $u=u(x,t),$ equal to $u_0(x)$ for $t=0$ and satisfying the integral identity

$$\int_Q\left[-\frac{\partial u}{\partial t}\frac{\partial\eta}{\partial t}+\sum_{i=1}^{n}\upsilon(x)\frac{\partial u}{\partial x_i}\frac{\partial\eta}{\partial x_i}+a_0(x)u\eta\right]dxdt-\int_\Omega u_1(x)\eta(x,0)dx=\int_Q f\eta dxdt \tag{5}$$

for all $\eta=\eta(x,t)$ from $W_{2,0}^1(Q)$, equal to zero for $t=T.$

From the results of [4, p.209-215] follows that under the above assumptions the boundary value problem (1), (2) for each fixed function $\upsilon\in V$ has a unique generalized solution from $W_{2,0}^1(Q)$ and the estimation

$$\|u\|_{W_2^1(Q)} \le c\left[\|u_0\|_{W_2^1(\Omega)} + \|u_1\|_{L_2(\Omega)} + \|f\|_{L_2(Q)}\right] \quad (6)$$

is valid. Here and in the sequel, by c we will denote various constants independent of the estimated quantities and admissible controls.

To the problem (1)-(4) we associate the following optimal control problem: it is reguired to minimize the functional

$$J(\upsilon) = \frac{1}{2}\int_\Omega \left[\int_0^T K(x,t)u(x,t;\upsilon)dt - \varphi(x)\right]^2 dx \quad (7)$$

under the conditions (1), (2), (4), where $u = u(x,t) = u(x,t;\upsilon)-$ is the solution of the boundary value problem (1), (2) corresponding to the function $\upsilon = \upsilon(x) \in V$.

We call the function $\upsilon(x)$ a control, the class $V-$ a set of admissible controls. There is close connection between the problems (1)-(4) and (1), (2), (4), (7) if in the problem (1), (2), (4), (7) $\min\limits_{\upsilon\in V} J(\upsilon) = 0,$ then additional integral condition (3) is fulfilled.

In futher, in order to avoid possible degeneration in the obtained we consider the following functionol condition for optimality:

$$J_\alpha(\upsilon) = J(\upsilon) + \frac{\alpha}{2}\int_\Omega \left[\upsilon^2(x) + \sum_{i=1}^n \left(\frac{\partial \upsilon}{\partial x_i}\right)^2\right]dx =$$

$$= J(\upsilon) + \frac{\alpha}{2}\|\upsilon\|^2_{W_2^1(\Omega)}, \quad (8)$$

where $\alpha > 0-$ is a given number.

III. ON THE EXISTENCE OF THE SOLUTION TO PROBLEM (1), (2), (4), (8).

Theorem 1. Let the conditions accepted in the statement of problem (1)-(4) be fulfilled. Then the set of optimal controls in the problem (1), (2), (4), (8) $V_* = \left\{\upsilon_* \in V : J(\upsilon_*) = J_* = \inf\limits_{\upsilon\in V} J(\upsilon)\right\}$ is non-empty, weakly compact in $W_2^1(\Omega)$ and any minimizing sequence $\{\upsilon^{(m)}\}$ weakly in $W_2^1(\Omega)$ converges to the set V_*. This theorem was prowed by some results in [5, 6].

Differentiability of the functional (8) and necessary condition for optimality.

Let $\psi = \psi(x,t;\upsilon)-$ be a generalized solution from $W_{2,0}^1(Q)$ of the adjoint problem

$$\frac{\partial^2 \psi}{\partial t^2} - \sum_{i=1}^n \frac{\partial}{\partial x_i}\left(\upsilon(x)\frac{\partial \psi}{\partial x_i}\right) + a_0\psi =$$

$$= -K(x,t)\left[\int_0^T K(x,\tau)u(x,\tau;\upsilon)d\tau - \varphi(x)\right], (x,t) \in Q, \quad (9)$$

$$\psi(x,T) = 0,\ \frac{\partial \psi(x,T)}{\partial t} = 0, x \in \Omega, \psi|_S = 0. \quad (10)$$

Under the generalized solution of the boundary value problem (9), (10) for each fixed control $\upsilon \in V$ we will understand the function $\psi = \psi(x,t;\upsilon)$ from $W_{2,0}^1(Q)$, that equal to zero for $t = T$ and satisfying the integral identity

$$\int_Q \left[-\frac{\partial \psi}{\partial t}\frac{\partial g}{\partial t} + \sum_{i=1}^n \upsilon(x)\frac{\partial \psi}{\partial x_i}\frac{\partial g}{\partial x_i} + a_0\psi g\right]dxdt =$$

$$= -\int_Q K(x,t)\left[\int_0^T K(x,\tau)u(x,\tau;\upsilon)d\tau - \varphi(x)\right]g(x,t)dxdt \quad (11)$$

for all $g = g(x,t)$ from $W_{2,0}^1(Q)$, that equal to zero for $t = 0.$

From the results of the [4, p.209-215] it follows that the boundary value problem (9), (10) for each fixed control $\upsilon(x) \in V$ has a unique generalized solution from $W_{2,0}^1(Q)$ and the estimation

$$\|\psi\|_{W_2^1(Q)} \le c\left[\|u\|_{L_2(Q)} + \|\varphi\|_{L_2(\Omega)}\right]$$

is valid. Taking into account estimation (6), hence we have

$$\|\psi\|_{W_2^1(Q)} \le c\left[\|u_0\|_{W_2^1(\Omega)} + \|u_1\|_{L_2(\Omega)} + \|f\|_{L_2(Q)} + \|\varphi\|_{L_2(\Omega)}\right] \quad (12)$$

Let the generalized solutions $u = u(x,t;\upsilon)$ and $\psi = \psi(x,t;\upsilon)$ from $W_2^1(Q)$ of problem (1), (2) and (9), (10), respectively, have the derivatives

$$\frac{\partial^2 u}{\partial x_i^2}, \frac{\partial^2 \psi}{\partial x_i^2},\ i = 1,...,n, \quad \text{that belong to the space } L_2(Q). \quad (13)$$

Theorem 2. Let the conditions of theorem 1 and condition (13) be fulfilled. Then the functional (8) is continuously Frechet differentiable on V and its differential at the point $\upsilon \in V$ for the increment $\delta\upsilon \in W_\infty^1(Q)$ is determined by the expression

$$\langle J'_\alpha(\upsilon), \delta\upsilon\rangle = \int_\Omega \left[\int_0^T \sum_{i=1}^n \frac{\partial u}{\partial x_i}\frac{\partial \psi}{\partial x_i}dt\right]\delta\upsilon(x)dx +$$

$$+ \alpha\int_\Omega \left[\upsilon\delta\upsilon + \sum_{i=1}^n \frac{\partial \upsilon}{\partial x_i}\frac{\partial \delta\upsilon}{\partial x_i}\right]dx. \quad (14)$$

Theorem 3. Let the conditions of theorem 2 be fulfilled. Then for the optimality of the control $\upsilon_* = \upsilon_*(x) \in V$ in the problem (1), (2), (4), (8) it is necessary for the inequality

$$\int_{\Omega}\left[\int_0^T\sum_{i=1}^n\frac{\partial u_*}{\partial x_i}\frac{\partial\psi_*}{\partial x_i}dt\right](\upsilon(x)-\upsilon_*(x))dx+$$
$$+\alpha\int_{\Omega}\left[\upsilon_*(x)(\upsilon(x)-\upsilon_*(x))+\sum_{i=1}^n\frac{\partial\upsilon_*}{\partial x_i}\left(\frac{\partial\upsilon(x)}{\partial x_i}-\frac{\partial\upsilon_*(x)}{\partial x_i}\right)\right]dx\geq 0 \qquad (15)$$

to hold for any $\upsilon=\upsilon(x)\in V$, where $u_*=u(x,t;\upsilon_*)$ and $\psi_*=\psi(x,t;\upsilon_*)-$ are the solutions of the problem (1), (2) and (9), (10) respectively for $\upsilon=\upsilon_*(x)$.

This theorem was proved by the results in [7].

IV. CONCLUSION

In this work considered one inverse problem at defining of higher coefficient for hyperbolic equation. The stated problem is reduced to finding the minimum of some functional structured by means of additional information.

Then, the optimal control problem is studiet by means of methods of optimal control theory, the existence of the control, differentiability of the functional is shown, a necessary condition for optimality is derived.

REFERENCES

[1] Tikhonov A.N., Arsenin V.Yu. Methods for solving ill-posed problems. M.: Nauka, 1974, 224 p.

[2] Kabanikhin S.I., Iskakov K.T. Optimizational methods for solving coefficient inverse problem. Novosibirsk: NSU, 2001.

[3] Kabanikhin S.I. Inverse and ill-posed problems. Novosibisk: Sib. Sci. publ., 2009, 457 p.

[4] Ladyzhenskaya O.A. Boundary value problems of mathematical physics. M.: Nauka, 1973, 408 p.

[5] Mikhailov V.P. Partial differential equations. II edition, M.: Nauka, 1983, 424 p.

[6] Sobolev S.L. Some applications of functional analysis in mathematical physics. M.: Nauka, 1988, 334 p.

[7] Vasil'ev F.P. Methods for solving extremal problems. M.: Nauka, 1981, 400 p.

[8] Vladimir G. R, Masahiro Y. (2019). Recovering two coefficients in an elliptic equation via phaseless information. Inverse Problems and Imaging, Volume 13, No. 1, 81–91

[9] Hasanov K.K., Gasumov T.M. A minimal energy control problem for second-order linear hyperbolic systems with two independent variables, Optim. Control Appl. Meth., 33(1), 2012, pp.51-60.

[10] Guliyev H.F. Seyfullayeva Kh.I., Optimal control problem with coefficients for the equation of vibrations of a elastic plate with discontinuous solution, International Journal of Applied Mathematics, Volume 36 No. 5, 2023, pp. 699-713

[11] Guliyev H.F., Nasibzadeh V.N. On optimization method in the Neumann problem for wave equation. International Journal of Applied Mathematics, Volume 30, 2017, № 6, pp.515-526.

[12] Guliyev H.F., Nasibzadeh V.N. On determining higher coefficient of a second order hyperbolic equation by the variational method // International Journal of Applied Mathematics, 2025, Volume 38, No 3, pp.323-334.

[13] Guliyev H.F., Safarova Z.R. On a determination of the initial functions from the observed values of the boundary functions for the second-order hyperbolic equation, Advanced Mathematical Models and applications, Vol.3, 2018, pp.215-222.

Synthesis of Optimal Control for the Heat Propagation Process in Rarefied Media with Conductive Viscosity

Rashad Mamedov
Department of General and Applied Mathematics,Faculty of Information Technology and Control, Azerbaijan State Oil and Industry University
Baku, Azerbaijan
rasadmammedov@gmail.com

Sardar Gasimov
Department of General and Applied Mathematics,Faculty of Information Technology and Control, Azerbaijan State Oil and Industry University
Baku, Azerbaijan
sardarkasumov1955@mail.ru

***Abstract*— The problem of building an optimal regulator in the process of heat exchange in viscous heat-conducting gaseous medium is considered. In this case, not only the heat transfer density, but also the velocity of heat transfer changes depends on temperature gradient. Moreover, as we consider a medium with temperature viscosity, we have to take into account the velocity of changes in temperature gradient. Therefore, the mathematical model of the considered process is a third order partial differential equation. The minimized functional is quadratic.**

The solution of the corresponding boundary value problem is sought in the form of Fourier series, therefore we restate our problem in infinite-dimensional phase space. To solve the newly obtained problem, we use the dynamic programming method, and we derive the optimal control in the form of synthesis.

Keywords*— *optimal regulator, gaseous medium, dynamic programming method.

Problem Statement

In gases, the heat is conducted according to generalized Fourier's law [1]. Also, the heat-conduction viscosity causes the changes in temperature gradient [2]. In this case, the function $u(x,t)$, which describes this process inside the domain $\underline{Q} = \{0 \le x \le 1, 0 \le t \le T\}$, satisfies the equation [3]

$$\beta \frac{\partial^2 u}{\partial t^2} + \frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} + \xi \frac{\partial^3 u}{\partial t \partial x^2} + p(x,t) \quad (1)$$

and the initial and boundary conditions

$$u(x,0) = u^0(x), \frac{\partial u(x,0)}{\partial t} = u^1(x), \quad (2)$$

$$\frac{\partial u(0,t)}{\partial x} = 0, \frac{\partial u(1,t)}{\partial x} + \alpha u(1,t) = 0, \quad (3)$$

where $a, \alpha > 0$ are the given numbers, β is a relaxation coefficient, $\xi = const > 0$ is a heat-conduction viscosity coefficient, $u^0(x) \in L_2(0,1)$ is an initial distribution and $u^1(x) \in L_2(0,1)$ is an initial velocity of temperature change, and $p(x,t) \in L_2(Q)$ is a control parameter.

Optimal control problem is to find a control parameter in the form of feedback $p[x,t,u(x,t)]$ such that the functional

$$I[p] = \gamma_1 \int_0^1 [u(x,T) - \psi_1(x)]^2 dx + \gamma_2 \int_0^1 [u(x,T) - \psi_2(x)]^2 dx + \gamma \iint_Q p^2(x,t) dx dt \quad (4)$$

takes the smallest possible value.

In [4], the optimal control problem with minimum energy for this process has been considered.

Stating Problem In Infinite-Dimensional Phase Space

Consider in $L_2(0,1)$ the orthonormal system of functions $X_n(x) = \frac{coscos\ \lambda_n x}{\sqrt{\omega_n}}, n = 1,2,\ldots,$ where λ_n's are the eigenvalues of the boundary value problem

$$X''(x) + \lambda^2 X(x) = 0, 0 < x < 1;$$
$$X'(0) = 0, X'(1) + \alpha X(1) = 0; \quad (5)$$

which are at the same time the positive roots of the equation $\lambda tg\lambda = \alpha$, and $\omega_n = \frac{\alpha + \alpha^2 + \lambda_n^2}{2(\alpha^2 + \lambda_n^2)}$ is a normalizing factor. With these in mind, we will seek for the approximate solution of the problem (1)-(3) in the form of Fourier series

$$u(x,t) = \sum_{n=1}^{\infty} u_n(t) X_n(x),$$
$$u_n(t) = \int_0^1 u(x,t) X_n(x) dx \quad (6)$$

Let's multiply both sides of the equation (1) by $X_n(x)$ and integrate with respect to x from 0 to 1. Then, in view of boundary conditions (3) and (6), the coefficients $u_n(t)$ satisfy the following system of ordinary differential equations [3]:

$$\beta \frac{d^2u_n(t)}{dt^2} + (1 + \xi\lambda_n^2)\frac{du_n(t)}{dt} + a^2\lambda_n^2 u_n(t) = p_n(t),$$

$$n = 1,2,\ldots \quad (7)$$

As the function (7) must satisfy the condition (2), we have

$$u_n(0) = u_n^0, \frac{du_n(0)}{dt} = u_n^1, n = 1,2,\ldots, \quad (8)$$

where $p_n(t)$ and u_n^0 are the Fourier coefficients of the functions $p(t,x)$ and $u_n^0(x)$, respectively.

Then the functional (4) becomes

$$I[\bar{p}] = \sum_{n=1}^{\infty} I_n[p_n], \quad I_n[p_n] = \gamma_1[u_n(T) - \psi_{1n}]^2 + .$$

$$+\gamma_2[u_n(T) - \psi_{2n}]^2 + \gamma\int_0^T p_n^2(t)dt \quad (9)$$

Obviously, minimization now can only be considered for the functional $I_n[p_n]$.

So we obtain an optimal control problem of finding a control in the form of feedback $p_n[t, u_n(t)]$ such that the functional (10) takes the smallest possible value.

Introduce new notations:

$$u_n(t) = u_{1n}(t), \dot{u}_n(t) = u_{2n}(t),$$
$$y(t) = (u_{1n}(t)\ u_{2n}(t)), \quad y_0 = (u_n^0\ u_n^1),$$
$$\underline{p}(t) = \left(0\ \frac{1}{\beta}p_n(t)\right),$$
$$A = \begin{pmatrix} 0 & 1 \\ -\frac{a^2}{\beta}\lambda_n^2 & -(1 + \xi\lambda_n^2) \end{pmatrix},$$
$$Q = \begin{pmatrix} \gamma_1 0 & 0\gamma_2 \end{pmatrix}, \quad \underline{\psi} = \left(\psi_{1n}\ \psi_{2n}\right)$$

Then the problem (7)-(9) can be rewritten in the following matrix form:

$$\dot{y}\ (t) = Ay(t) + p(t), \quad y(0) = y_0. \quad (10)$$

$$I[p] = [y(T) - \underline{\psi}]'Q[y(T) - \underline{\psi}] + \gamma\int_0^T p^2(t)dt \quad (11)$$

The solution of the problem (11) has the following form:

$$u_n(t) = \Phi_n(t,0) + \int_0^t \Phi_n(t-\tau)p_n(\tau)d\tau, n = 1,2,\ldots,N \quad (12)$$

where $\boldsymbol{\Phi}_n(t) = \frac{k_2(n)u_n^0 - u_n^1}{k_2(n) - k_1(n)}e^{k_1(t)} + \frac{u_n^1 - k_1(n)u_n^0}{k_2(n) - k_1(n)}e^{k_2(t)}$ is a solution of homogeneous equation and

$k_{1,2}(n) = -\frac{1}{2\beta}(1 + \xi\lambda_n^2) \pm \frac{1}{2\beta}\sqrt{(1 + \xi\lambda_n^2)^2 - 4a^2\beta\lambda_n^2}$.

In general, the function (6) with the Fourier coefficients (12) may not have the second order derivatives. Therefore, this function should be understood as a generalized solution, i.e. under the conditions of the problem, every admissible control defines a unique generalized solution $u(t,x) \in W_2^{1,2}(Q)$ which satisfies the integral identity [4]

$$\int_0^T \int_0^1 [-\beta\frac{\partial u(t,x)}{\partial t}\cdot\frac{\partial\Phi}{\partial t} + u(t,x)\Phi + a^2\frac{\partial u(t,x)}{\partial x}\cdot\frac{\partial\Phi}{\partial x} + \xi\frac{\partial^2 u(t,x)}{\partial x^2}\Phi]dxdt -$$

$$-\int_0^1 \phi_1(x)\Phi(0,x)dx =$$

$$\int_0^T \int_0^1 p(t,x)\Phi(t,x)dxdt$$

for every function $\Phi(t,x) \in W_2^{1,2}(Q)$ such that $\Phi(T,x) = 0$, $\Phi_x(t,0) = 0$, $\Phi_x(t,1) = 0$, and the satisfaction of conditions (2) is understood in the sense of $L_2(Q)$.

APPLYING DYNAMIC PROGRAMMING METHOD

Introduce the functional

$$S[t,y] = \min_{\bar{p}}\{[y(T) - \bar{\psi}]'Q[y(T) - \bar{\psi}] + \gamma\int_t^T \bar{p}^2(t)dt\} \quad (13)$$

Applying the algorithm of dynamic programming method, we obtain the following Bellman equation:

$$-\frac{\partial S[t,y]}{\partial t} = \min_{\bar{p}}\{\overleftrightarrow{}\ \gamma\bar{p}^2(t) + gradS[t,y]\cdot(Ay(t) + \bar{p}(t))\} \quad (14)$$

Assume that the admissible controls are the arbitrary functions from $L_2(Q)$. Then, from the right-hand side of the above Bellman equation we find

$$\bar{p}(t) = -\frac{1}{2\gamma}gradS[t,y]. \quad (15)$$

Excluding $\bar{p}$ from (14), we obtain

$$-\frac{\partial S}{\partial t} = -\frac{1}{4\gamma} grad^2 S[t,y] + gradS[t,y] \cdot Ay. \quad (16)$$

From the definition (13) of the functional S it directly follows that $S \geq 0$ and

$$S[T,y] = [y(T) - \bar{\psi}]' Q[y(T) - \bar{\psi}]. \quad (17)$$

The solution of the functional equation (16) with the additional condition (17) has the form of

$$S[t,y] = [y(t) - \bar{\psi}]' K(t)[y(t) - \bar{\psi}], \quad (18)$$

where $K(t)$ is a solution of the Riccati-type matrix differential equation

$$\dot{K}(t) = -A'K(t) - K(t) + \frac{1}{\gamma}K(t), K(T) = Q \quad (19)$$

CONSTRUCTING AN OPTIMAL CONTROL IN THE FORM OF SYNTHESIS

As the Bellman equation has the form (18), using (15), we define the optimal control as follows:

$$\bar{p}(t) = -\frac{1}{2\gamma}K(t)[y - \bar{\psi}],$$

or, in coordinates,

$$p_n(t) = -\frac{1}{\gamma}K_{1n}(t)u_n(t) - \frac{1}{\gamma}K_{2n}(t)\dot{u}_n(t) + \frac{1}{\gamma}(K_{2n}(t)\psi_1 + +K_{3n}(t)\psi_2).$$

Then, introducing the functions

$$K_\nu(x,s,t) = -\frac{1}{\gamma}\sum_{n=1}^{\infty} K_{in}(t)X_n(x)X_n(s),$$

$$\nu = 1,2;$$

$$q(t,x) = \frac{1}{\gamma}\sum_{n=1}^{\infty} (K_{2n}(t)\psi_{1n} + +K_{3n}(t)\psi_{2n})X_n(x),$$

we obtain the following synthesis control for our initial problem:

$$p(x,t) = \int_0^1 K_1(x,s,t)u(s,t)ds + . + \int_0^1 K_2(x,s,t)u_t(s,t)ds + q(x,t)$$

Thus, the optimal control is constructed in the form of feedback, i.e. in the form of a functional defined as a state function. This provides the optimal regulation for the heat exchange process in gaseous medium.

REFERENCES

[1] Kerimova M.E. Asymptotic of the solution of a boundary value problem stated in an infinite half strip for a non-classical type singularly perturbed differential equation. // Journal of Contemporary Applied Mathematics, 2021, v.11, No 2, p.61-70.

[2] Mammadov R.S., Gasimov S.Y. Approximate solution of optimal problem for cooling process with minimum energy // Baku Mathematical Journal, 2022, Vol. 1, No.1, p. 106-110.

[3] Yagubova M.M. Constructing an optimal control by the method of moments for the third order linear equation with quadratic quality criterion // Problemi upravleniya i informatiki. 2016, №4, p.47-53. (in Russian)

[4] Mammadov R.S., Gasimov S.Y. Optimal control problem with minimum energy for process of heat transfer in rarefied media with heat-conduction viscosity // 5th International Conference on Problems of Cybernetics and Informatics (PCI 2023), August 28-30, Baku, Azerbaijan.

[5] Mammadov R.S., Gasimov S.Y. Solution of optimal control problem for process of heat transfer with minimum energy in media with temperature viscosity // Proceedings of International Conference on Actual Problems of Applied Mathematics, May 22-26, 2018, Nalchik, Russia. (in Russian)

[6] Yegorov A.I. The fundamentals of control theory. M.: FIZMATLIT, 2004. – 504 p. (in Russian)

A Solution Method for the Integer Knapsack Problem with Fuzzy Number Coefficients

Knyaz Mammadov
Baku State University
Institute of Control Systems [2]
Baku, Azerbaijan
mamedov_knyaz@yahoo.com

Saqif Huseynov
Institute of Control Systems
Baku, Azerbaijan
saqif.huseynov@gmail.com

***Abstract*—The paper considers the integer knapsack problem with fuzzy coefficients. Using known properties of fuzzy numbers, an equivalent representative is selected of each fuzzy number. It has been proven in the literature that the optimal solutions of the obtained problem and the problem with fuzzy coefficients are the same. Therefore, an exact solution method of the dynamic programming type has been developed for the problem with fuzzy coefficients.**

Keywords— integer knapsack problem, fuzzy coefficients, precise representative, dynamic programming

INTRODUCTION

The classical knapsack problem in operations research is one of the frequently encountered optimization problems indecision-making processes. This problem is widely applied in building practicalmodels aimed at achieving maximum benefit under limited resources. Howeverin real-life situations, it is not always possible to determine thecoefficients (such as profit,cost, weight, or amount of resources) precisely. In such cases, the application of fuzzy logic becomes necessary.

In this study, an integer knapsack problem with coefficients given as triangular fuzzy numbers is considered. The objective is to develop an appropriate modeling and solution method for finding optimal decisions under uncertainty.

Let us consider the following integer knapsack problem with fuzzy coefficients.

$$\sum_{j=1}^{n} \tilde{c}_j x_j \to \ , \qquad (1)$$

$$\sum_{j=1}^{n} \tilde{a}_j x_j \le \tilde{b} \quad , \qquad (2)$$

$$0 \le x_j \le d_j \ , \quad (j = \underline{1,n}), \quad (3)$$

$$x_j - integer \ , \quad (j = 1,n), \quad (4)$$

Here $\tilde{c}_j = (c_j, \alpha, \alpha)$, $\tilde{a}_j = (a_j, \beta, \beta)$, $\tilde{b} = (b, \lambda, \lambda)$ are the given fuzzy numbers; the pairs (α, α), (β, β), (λ, λ) are the right and left circumcircles of the given numbers $c_j, a_j, (j = 1,n)$ and b.

It should be noted that in a special case, these numbers can be graphically represented as follows.

The numbers c_j, a_j and b are the chosen fuzzy representatives of the fuzzy numbers $\tilde{c}_j, \tilde{a}_j, \tilde{b}$ respectively.

Here the geometric representation of the given fuzzy numbers $\tilde{c}_j, \tilde{a}_j$, and b as follows (figure 1 – 3).

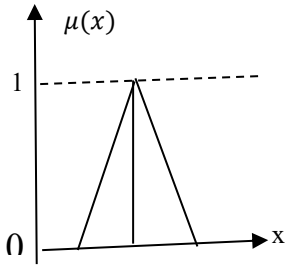


Figure 1

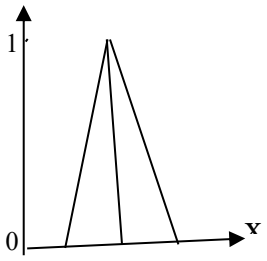


Figure 2

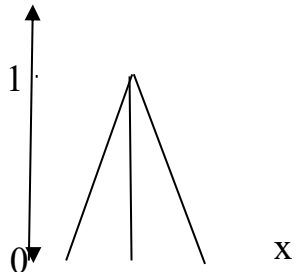


Figure 3

It should be noted that when the coefficients in problems (1) – (3) are given as integers, the problem is referred to as an integer knapsack problem, and various exact and

approximate methods have been developed for its solution [1,3], among others.

In general, models in the form of integer programming problems arise in various fields of economics and engineering. Therefore, studying this problem and developing new solution methods is of significant importance. However, since these problems belong to the class of NP-complete problems, i.e., the class of "hard-to-solve" problems, numerous approximate solution methods have been developed.

In this study, we have developed a dynamic programming-based method for the exact solution of problem (1) – (4). In this case, the coefficients of the problem to be solved are chosen as the representatives of the corresponding fuzzy numbers.

It has been proven in [4] that the optimal solution of this problem coincides with the optimal solution of the fuzzy-coefficient problem.

On the other hand, that work also provides the following rules for determining the crisp representative of an arbitrary fuzzy number A.

$$I(A) = C_L I_L(A) + C_R I_R(A),\ C_L \geq 0,\ C_R \geq 0, C_L + C_R = 1. \quad (5)$$

$$I_L(A) = \int_0^1 L_A(\xi)P(\xi)d\xi\ ,\ \ I_R(A) = \int_0^1 R_A(\xi)P(\xi)d\xi\ . \quad (6)$$

Here

$$\int_0^1 P(\xi)d\xi = 1\ , \forall \xi\ \in (0,1]\ ;\ P(\xi) \geq 0\ , L_A(\xi) = \mu_{\uparrow}^{-1}(\xi), R_A(\xi) = \mu_{\downarrow}^{-1}(\xi)\ .$$

Here $P: [0;1] \to E_+ = [0;\ +\infty)$ is a function with $E_+ = [0;\ +\infty)$, the inverse functions of the increasing and decreasing parts of the function $\mu_A(x)$ are $\mu_{\uparrow}^{-1}(\xi)$ and $\mu_{\downarrow}^{-1}$ respectively.

Specifically:

$$\mu_{\tilde{a}}(x) = \{0, if\ x < a(1-p)\ or\ x > a(1+p)\ \frac{x-(1-p)a}{p\cdot a},\ or\ (1-p)a \leq x \leq a\ \frac{(1+p)a-x}{p\cdot a},\ or\ a < x \leq (1+p)a$$

$$\mu_{\tilde{a}}(x) = \begin{cases} 0, if\ x < a(1-p)\ or\ x > a(1+p) \\ \dfrac{x-(1-p)a}{p\cdot a},\ or\ (1-p)a \leq x \leq a \\ \dfrac{(1+p)a-x}{p\cdot a},\ or\ a < x \leq (1+p)a \end{cases}$$

Here, p is the spread percentage; for example, it can be taken as $p = 0.1$

We assume P(A)=1, and based on formulas (5) and (6), we will select the crisp representatives of the fuzzy numbers $\widetilde{c_j}$, $\widetilde{a_j}$ and $\tilde{b}$ as follows:

$$I(\widetilde{c_j}) = C_L I_L(\widetilde{c_j}) + C_R I_R(\widetilde{c_j})$$

$$C_L I_L(\widetilde{c_j}) = \int_0^1 \frac{1}{2}(c_j - \alpha + \alpha\xi^*)d\xi^*\ , C_R I_R(\widetilde{c_j}) == \int_0^1 \frac{1}{2}(c_j + \alpha - \alpha\xi^*)d\xi^*$$

$$I(\widetilde{a_j}) = C_L I_L(\widetilde{a_j}) + C_R I_R(\widetilde{a_j})$$

$$C_L I_L(\widetilde{a_j}) = \int_0^1 \frac{1}{2}(a_j - \beta + \beta\xi^*)d\xi^*\ , C_R I_R(\widetilde{a_j}) = = \int_0^1 \frac{1}{2}(a_j + \beta - \beta\xi^*)d\xi^*$$

$$I(\tilde{b}) = C_L I_L(\tilde{b}) + C_R I_R(\tilde{b})$$

$$C_L I_L(\tilde{b}) = \int_0^1 \frac{1}{2}(b - \gamma + \gamma\xi^*)d\xi^*\ , C_R I_R(\tilde{b}) = \int_0^1 \frac{1}{2}(b + \gamma - \gamma\xi^*)d\xi^*.$$

As a result, is obtain the following problem with fluent coefficient.

$$\sum_{j=1}^{n} I(\widetilde{c_j})x_j \to \quad (7)$$

$$\sum_{j=1}^{n} I(\widetilde{a_j})x_j \leq I(\tilde{b}) \quad (8)$$

$$0 \leq x_j \leq d_j, (j = \underline{1,n})\ and\ xj\ is\ an\ integer \quad (9)$$

Here, $I(\widetilde{c_j})$, $I(\widetilde{a_j}), I(\tilde{b})$ are integers. Let us apply the principle of dynamic programming to solve the obtained problem (7) – (9). To derive the recurrence formula of this principle, let $f_k(z)$ denote the maximum value of the function (7) when $n = k; I(\tilde{b}) = z$. Here, *k=1,2,...,n;* $z = 0,1,2,\dots, I(\tilde{b})$.

To derive the recursive formula, let us perform the following operations.

$$f_k(z) = \max_{x_j=0,1,\dots,d_k} \left\{ \sum_{j=1}^{k} I(\widetilde{c_j})x_j \Big| \sum_{j=1}^{k} I(\widetilde{a_j})x_j \leq z \right\}$$

$$= \max_{x_j=0,1,\dots,d_{k-1}} \left(\sum_{j=1}^{k-1} I(\widetilde{c_j})x_j + I(\widetilde{c_k}) \Big| \sum_{j=1}^{k-1} I(\widetilde{a_j})x_j + I(\widetilde{a_k}) \leq z \right);$$

$$= \max_{x_j=0,1,\dots,d_{k-1}} \left(\sum_{j=1}^{k-1} I(\widetilde{c_j})x_j + I(\widetilde{c_k}) \cdot 2 \Big| \sum_{j=1}^{k-1} I(\widetilde{a_j})x_j + I(\widetilde{a_k}) \cdot 2 \leq z \right)$$

$$\max_{x_j=0,1,\dots,d_{k-1}} \left(\sum_{j=1}^{k-1} I(\widetilde{c_j})x_j + I(\widetilde{c_k}) \cdot d_k \middle| \sum_{j=1}^{k-1} I(\widetilde{a_j})x_j + I(\widetilde{a_k}) \cdot d_k \le z\right) =$$
$$\max\{f_{k-1}(z);\ f_{k-1}(z - I(\widetilde{a_k})) + I(\widetilde{c_k}); f_{k-1}(z - 2I(\widetilde{a_k})) + 2I(\widetilde{c_k}) \dots; f_{k-1}(z - d_k I(\widetilde{a_k})) + d_k I(\widetilde{c_k})\} =$$
$$\max_{h=0,1,\dots,d_k} \{ f_{k-1}(z - hI(\widetilde{a_k})) + hI(\widetilde{c_k})\}$$

Thus, for solving the problem (7) – (9), we obtain the following recursive formula based on the dynamic programming principle:

$$\max_{h=0,1,\dots,d_k} \{ f_{k-1}(z - hI(\widetilde{a_k})) + hI(\widetilde{c_k})\} \quad (10)$$

Here *k=1,2,...,n;* *z= 0,1,2,...,* $I(\tilde{b})$. However, initially üçün $f_o(z) = 0$, for $\forall z \ge 0$, $f_k(0) = 0$ for $\forall k$ and $f_k(z)$ is not defined when z<0.

It can be seen from formula (10) that during successive calculations, when *k=n* and $z = I(\tilde{b})$ the number $f_n(I(\tilde{b}))$ is the maximum value of the function (7) in problem (7) – (9). The optimal solution $X = (x_1, x_2, \dots, x_n)$ that gives this value is determined by the following formula:

$$x_j = H\left[j, \quad I(\tilde{b}) - \sum_{i=j+1}^{n} I(\widetilde{a_j})x_j\right]$$

Here, *j* takes values *n, n-1*, ..., *2, 1*. That is *j = n, n-1, ..., 2, 1.*

It should be noted that the matrix **H** used in equation (11) is $(n+1) \times (I(\tilde{b}) + 2)$. The elements of this matrix are the maximum values of the variable **h** obtained using equation (10).

Conclusion: A method has been developed to find the optimal solution of the integer knapsack problem with fuzzy coefficients. For this purpose, a new recursive formula based on the principle of dynamic programming has been derived for the problem. Using this formula, the optimal value of the objective function is obtained. The solution that yields this value is then determined through a special matrix constructed during the computational process. Thus, the solutions of the practical integer knapsack problem with fuzzy number cofficients were given.

REFERENCES

[1] Kellerer H., Pferschy U., Pisinger D. *Knapsack Problems*. Springer-Verlag, Berlin, Heidelberg, New York, 2004, 546 p.

[2] Мамедов К.Ш. Один подход к решению целочисленной задачи о ранце // Журнал «Автоматика и Вычислительная Техника», 2005, №5, pp. 56–63.

[3] Chen, S. J., & Klein, C. M. (1997). An efficient heuristic for solving the fuzzy knapsack problem. *European Journal of Operational Research*, 100(3), 593–603.

[4] Насибов Э.Н. Методы обработки нечеткой информации в задачах принятия решений. Баку: Элм, 2000, 260 с.

[5] Məmmədov K. Ş., Hüseynov S. Y. Qeyri–səlis çanta məsələsinin bir həll üsulu / *"İnformasiyalaşdırma, Kibernetika və İnformasiya texnologiyalarının müasir problemləri"* Respublika elmi konfransının əsərləri. Bakı, 2003, III cild, səh. 10–13.

Some Necessary Conditions for Optimality in Fractional Caputo Systems with Control Delay

Shakir Yusubov
Department of Mechanics and Mathematics,
Baku State University
Baku, Azerbaijan
yusubovshsh@gmail.com
0000-0001-5330-5519

Elimhan Mahmudov
Azerbaijan National Aviation Academy, Baku, Azerbaijan;
Azerbaijan University of Architecture and Construction
Baku, Azerbaijan
elimhan22@yahoo.com
0000-0003-2879-6154

Shikhi Yusubov
Department of Mathematics, Shanghai University,
Shanghai, China
yusubovsixi@gmail.com

***Abstract*— This paper considers an optimal control problem involving a dynamic system described by a nonlinear fractional differential equation of Caputo type of order $0 < \alpha < 1$ with control delay, associated with a Bolza-type cost functional expressed as the sum of a Mayer cost and a Lagrangian cost given by a Riemann-Liouville fractional integral of order $\beta > 0$ with control delay. For this problem, an analogue of the Euler equation and the Legendre-Clebsch condition has been obtained. In deriving the analogue of the Legendre-Clebsch condition, a higher-order necessary optimality condition was obtained.**

Keywords—fractional Caputo derivative, fractional optimal control, necessary optimality condition

INTRODUCTION

It is known that the theory of optimal control, reflecting the modern stage of development of the calculus of variations, arose in connection with the need to solve a number of problems posed by practice in various areas of development of new technology. In modern mathematics, the theory of optimal control occupies an important place. Its field of application is constantly expanding due to the emergence of new applied problems. The ability to study systems described by equations with fractional derivatives has revealed a new layer of unexplored control problems for such systems. Note that models with fractional derivatives are more adequate than integer-order models, especially when describing memory and heredity effects in real systems [1-3].

In the last thirty years, problems of optimal control of dynamic systems with fractional derivatives have been actively studied and various necessary optimality conditions have been obtained for them [4-11].

To the best of our knowledge, there is only one result on fractional dynamics problems with control delay for singular controls [11]. In work [11] a necessary condition for first-order optimality is obtained, expressed in the form of Pontryagin's maximum principle. In addition, a necessary condition for high-order optimality is obtained for controls that are singular in the Pontryagin sense. Note that if Pontryagin's maximum principle is valid with respect to some control but is not degenerate, then the results [11] leave this control among the candidates for optimality. Therefore, for further study of such controls it is natural to have new necessary optimality conditions.

In this paper, we consider the problem of optimal control of a dynamic system whose motion is described by a nonlinear differential equation with a fractional Caputo derivative of order $0 < \alpha < 1$ and a control delay. The duration of process control is defined and limited. The goal of management is to minimize a given Bolza-type cost functional, consisting of two components. One of them estimates the state vector of the system, and the second is the Riemann-Liouville integral of order $\beta > 0$ with an integrand that also depends on the delay control. The set of values of control functions is non-empty and open. The stated optimal control problem is investigated using the first and second variations of the functional. The cases $0 < \alpha < 1$, $\beta \geq \alpha$ and $0 < \beta \leq \alpha < 1$ are considered separately. In each of these cases, the first and second variations of the functional are derived, as well as the corresponding conjugate integral equation. Note that the solutions of these integral equations have no singularities. The existence and uniqueness of continuous solutions of these integral equations is proven. The Euler equations, the Legendre conditions and, when the Legendre condition degenerates, a necessary condition for high-order optimality are obtained.

STATEMENT OF THE PROBLEM

Consider a dynamical system whose motion is described by a differential equation with a fractional Caputo derivative of order $\alpha \in (0,1)$, coupled with a control delay

$$\left({}^{C}D^{\alpha}_{t_0+}x\right)(t) = f(t, x(t), u(t), v(t)), a.e.\ t \in [t_0, t_1], (1)$$

with the initial condition

$$x(t_0) = x_0. \quad (2)$$

Here $x(t)$ is n-dimensional vector of the state variable, $u(t)$ is r-dimensional vector of the control variable, $v(t) = u(t-h)$, t represents the time, positive constant h-delay effect, and $t_0, t_1 \in R, x_0 \in R^n$ are fixed points, and $t_1 > t_0 + h$, function $f(t,x,u,v)$ is continuous on totality of arguments on $[t_0, t_1] \times R^n \times R^r \times R^r$ alongside partial derivatives with respect to x, u, v up to the second order, inclusively;

$$({}^{C}D^{\alpha}_{t_0+}x)(t) = \frac{d}{dt}\left(I^{1-\alpha}\left(x(\cdot) - x(t_0)\right)\right)(t) = \frac{1}{\Gamma(1-\alpha)}\frac{d}{dt}$$

$$\times \int_{t_0}^{t} (t-\tau)^{-\alpha}\big(x(\tau)-x(t_0)\big)d\tau$$

the Caputo derivative of order α. We define the set of admissible controls as the collection of measurable, bounded r-dimensional vector functions $u(t), t \in [t_0 - h, t_1]$, and these functions assume values from a given non-empty open set $V \subset R^r$.

$$u(t) \in V, t \in [t_0 - h, t_1]. \quad (3)$$

As a solution of the problem (1), (2) corresponding to the fixed control function $u(t), t \in [t_0 - h, t_1]$, we consider the function $x(\cdot) \in AC_\infty^\alpha\big([t_0, t_1], R^n\big)$ satisfies differential equations (1) for almost every $t \in [t_0, t_1]$ and the initial condition (2), where $AC_\infty^\alpha([t_0, t_1], R^n)$ is the set of functions $x(\cdot)$ representable in the form

$$x(t) = b_0 + \big(I_{t_0+}b\big)(t) = b_0 + \frac{1}{\Gamma(\alpha)}\int_{t_0}^{t} (t-\tau)^{\alpha-1}b(\tau)d\tau,$$

$b_0 \in R^n, b(\cdot) \in L_\infty([t_0, t_1], R^n)$.

The goal of the optimal control problem is the minimization of the functional

$$J(u) = \varphi\big(x(t_1)\big) + \tfrac{1}{\Gamma(\beta)}\int_{t_0}^{t_1} (t_1 - t)^{\beta-1} f_0(t, x(t), u(t), v(t))dt \quad (4)$$

established within the solutions of the problem (1), (2) for admissible controls that meet the condition (3). In this context, we assume that $\varphi(x)$ is a given, twice continuously differentiable scalar function, and the scalar function $f_0(t, x, u, v)$ remains continuous in the totality of variables on $[t_0, t_1] \times R^n \times R^n \times R^r$ and has continuous partial derivatives with respect to x, u, v up to the second order inclusive. It is assumed that $0 < \alpha < 1$ and $\beta > 0$.

Let assume problem (1), (2) for each admissible control $u(\cdot)$ have a unique solution $x(\cdot) \in AC_\infty^\alpha([t_0, t_1], R^n)$. Let $x(\cdot)$ be a solution of the problem (1), (2) corresponding to $u(\cdot)$,then we call pair$\big(u(\cdot), x(\cdot)\big)$ an admissible process. We call an admissible control $u(\cdot)$ that is the solution of problem (1)-(4) an optimal control, and its corresponding trajectory $x(\cdot)$, an optimal trajectory. Then the pair $(u(\cdot), x(\cdot))$ is said to be an optimal process.

NECESSARY OPTIMALITY CONDITIONS

We will consider the cases $0 < \alpha \leqslant \beta$ and $0 < \beta \leqslant \alpha < 1$ separately. According to these cases, let us consider the Hamilton- Pontryagin functions $H(t, x, u, v, \psi) = \psi' f(t, x, u, v) - \frac{\Gamma(\alpha)}{\Gamma(\beta)}(t_1 - t)^{\beta-\alpha} f_0(t, x, u, v)$ and $H_1(t, x, u, v, \psi_1) = \psi_1' f(t, x, u, v) - f_0(t, x, u, v)$, as well as the functions ψ and ψ_1, which are solutions of the adjoint integral equations

$$\psi(t) = -\varphi_x\big(x(t_1)\big) + \frac{(t_1 - t)^{1-\alpha}}{\Gamma(\alpha)} \times \int_t^{t_1} (t_1 - \tau)^{\alpha-1}(\tau - t)^{\alpha-1} H_x(\tau)d\tau, t \in [t_0, t_1], \quad (5)$$

and

$$\psi_1(t) = -\frac{\Gamma(\beta)}{\Gamma(\alpha)}(t_1 - t)^{\alpha-\beta}\varphi_x\big(x(t_1)\big) + \frac{(t_1 - t)^{1-\beta}}{\Gamma(\alpha)} \times \int_t^{t_1} (t_1 - \tau)^{\beta-1}(\tau - t)^{\alpha-1} H_{1x}(\tau)d\tau, t \in [t_0, t_1], \quad (6)$$

respectively, where

$$H_x(t) = H_x(t, x(t), u(t), v(t), \psi(t)), H_{1x}(t) = H_{1x}(t, x(t), u(t), v(t), \psi_1(t)).$$

The expressions of the first and second variations of the functional have the following form:

$$\delta J(u) = -\frac{1}{\Gamma(\alpha)}\int_{t_0}^{t_1} (t_1 - t)^{\alpha-1}[H_u'(t)\delta u(t) + H_v'(t)\delta v(t)]dt,$$

$$\delta^2 J(u) = -\frac{1}{\Gamma(\alpha)}\int_{t_0}^{t_1} (t_1 - t)^{\alpha-1}[\delta x'(t)H_{xx}(t)\delta x(t) + \delta u'(t)H_{uu}(t)\delta u(t) + \delta v'(t)H_{vv}(t)\delta v(t) + 2\delta x'(t)H_{xu}(t)\delta u(t) + 2\delta x'(t)H_{xv}(t)\delta v(t) + 2\delta u'(t)H_{uv}(t)\delta v(t)]dt, \delta u(t) \in L_\infty([t_0, t_1], R^r)$$

$$(\delta_1 J(u) = -\frac{1}{\Gamma(\beta)}\int_{t_0}^{t_1} (t_1 - t)^{\beta-1}[H_{1u}'(t)\delta u(t) + H_{1v}'(t)\delta v(t)]dt,$$

$$\delta_1^2 J(u) = -\frac{1}{\Gamma(\beta)}\int_{t_0}^{t_1} (t_1 - t)^{\beta-1}[\delta x'(t)H_{1xx}(t)\delta x(t) + \delta u'(t)H_{1uu}(t)\delta u(t) + \delta v'(t)H_{1vv}(t)\delta v(t) + 2\delta x'(t)H_{1xu}(t)\delta u(t) + 2\delta x'(t)H_{1xv}(t)\delta v(t) + 2\delta u'(t)H_{1uv}(t)\delta v(t)]dt).$$

Using the first variation of the functional, the following theorem can be proven.

Theorem 1. Let the admissible process $(u(\cdot), x(\cdot))$ be optimal in problem (1)-(4) and let $\psi(\cdot)\big(\psi_1(\cdot)\big)$ be a solution of adjoint problem (5) ((6)) calculated on optimal process. Then following equality is fulfilled

$$\chi(t)(t_1 - t)^{\alpha-1}H_u(t) + \chi(t + h)\big(t_1 - (t + h)\big)^{\alpha-1} \cdot \cdot H_v(t + h) = 0, a.e.\ t \in [t_0 - h, t_1)$$

$$(\chi(t)(t_1 - t)^{\beta-1}H_{1u}(t) + \chi(t + h)\big(t_1 - (t + h)\big)^{\beta-1} \times H_{1v}\big(t + h\big) = 0, a.e.\ t \in [t_0 - h, t_1)), \quad (7)$$

where $\chi(t), t \in R$, is the characteristic function of the set $\big[t_0, t_1\big)$.

Theorem 2. If an admissible control $u(t)$ satisfies the Euler condition (7), then for its optimality in problem (1)-(4) it is necessary that the inequality

$$v'[\chi(t)(t_1-t)^{\alpha-1}H_{uu}(t)$$
$$+\chi(t+h)(t_1-(t+h))^{\alpha-1}H_{vv}(t+h)]v \leqslant 0$$
$$(v'[\chi(t)(t_1-t)^{\beta-1}H_{1uu}(t) \quad (8)$$
$$+\chi(t+h)(t_1-(t+h))^{\beta-1}H_{1vv}(t+h)]v \leqslant 0)$$

hold for all $v \in R^r$ and for a.e. $t \in [t_0-h, t_1)$. The possiblity of degeneration of the necessary optimality condition (8) is not ruled out.

Definition. An admissible control is said to be singular in the classical sense if for all $v \in R^r$ and for a.e. $t \in [t_0-h, t_0)$

$$\chi(t)(t_1-t)^{\alpha-1}H_u(t) + \chi(t+h)\big(t_1-(t+h)\big)^{\alpha-1} \times H_v(t+h) = 0,$$
$$v'[\chi(t)(t_1-t)^{\alpha-1}H_{uu}(t)$$
$$+\chi(t+h)(t_1-(t+h))^{\alpha-1}H_{vv}(t+h)]v = 0$$
$$(\chi(t)(t_1-t)^{\beta-1}H_{1u}(t) + \chi(t+h)\big(t_1-(t+h)\big)^{\beta-1} \times H_{1v}(t+h) = 0, \quad (9)$$
$$v'[\chi(t)(t_1-t)^{\beta-1}H_{1uu}(t)$$
$$+\chi(t+h)(t_1-(t+h))^{\beta-1}H_{1vv}(t+h)]\, v = 0).$$

It is clear that when conditions (9) are satisfied the analogue of the Legendre-Clebsch condition (8) degenerates and does not provide any information about the optimality of the control under study. In this case, additional research is required.

Let $u(t)$ be a singular control in the classical sense. Then the following theorem is true.

Theorem 3. For optimality of control $u(t)$ singular in the classical sense in problem (1)-(4) it is necessary that the inequality

$$v'[\chi(t)(t_1-t)^{\alpha-1}H_{ux}(t)f_u(t) + \chi(t+h)\big(t_1-(t+h)\big)^{\alpha-1}H_{vx}(t+h)f_v(t+h)]v \leq 0$$
$$(v'[\chi(t)(t_1-t)^{\beta-1}H_{1ux}(t)f_u(t) + \chi(t+h)\big(t_1-(t+h)\big)^{\beta-1} \times H_{1vx}(t+h)f_v(t+h)]v \leq 0)).$$

hold for all $v \in R^r$ and for a.e. $t \in [t_0-h, t_1)$.

CONCLUSION

In this paper, we considered a general optimal control problem involving a dynamic system described by a nonlinear fractional differential equation of order $0 < \alpha < 1$ with a control delay associated with a common Bolza cost written as the sum of a standard Mayer cost and a Lagrange cost given by a Riemann-Liouville fractional integral of order $\beta > 0$ with a control delay. The posed problem of optimal control is investigated using a new version of the increment method, in which the concept of a conjugate equations of an integral form is essentially used. The cases $0 < \alpha < 1, \beta \geqslant \alpha$ and $0 < \beta \leqslant \alpha < 1$ are considered separately. In each of these cases, the first and second variations of the functional are derived, as well as the corresponding conjugate integral equations. Using the first and second variations of the functional, we obtain the Euler equations, the Legendre conditions, and when the Legendre condition degenerates we obtain the necessary high-order optimality condition. The approach presented here can be applied to the derivation of necessary optimality conditions in the form of a Euler equation, the Legendre-Clebsch condition and a high order for an optimal control problem in which the system is controlled by a nonlinear fractional Caputo partial differential equation.

REFERENCES

[1] S.G. Samko, A.A. Kilbas, O.I. Marichev, Fractional integrals and derivatives: theory and applications, Gordon and Breach Science Publishers, Switzerland: Yverdon, (1993).

[2] A.A. Kilbas, H.M. Srivastava, and J.J. Trujillo, Theory and applications of fractional differential equations, Volume 204 of North-Holland Mathatics Studies, Amsterdam:Elsevier Science B.V. (2006).

[3] E.N. Mahmudov, and Sh.Sh. Yusubov, Nonlocal boundary value problems for hyperbolic equations with a Caputo fractional derivative, J. Comput. Appl. Math. 398 (2021), 1-15.

[4] M. Bergounioux and L. Bourdin, Pontryagin maximum principle for general Caputo fractional optimal control problems with Bolza cost and terminal constaints, ESAIM Control Optim.Calc. Var., 26 (2020), Paper No. 35, 38 pp.

[5] R. Kamocki and M. Majewski, Fractional linear control systems with Caputo derivative and their optimization, Optim. Control Appl. Math., 36 (2014), 953-967.

[6] Sh.Sh. Yusubov, E.N. Mahmudov, Optimality conditions of singular controls for systems with Caputo fractional derivatives, J. Indust. Manag. Optim. 19(1)(2023), 246-264.

[7] Sh.Sh. Yusubov, E.N. Mahmudov, Necessary and sufficient optimality conditions for fractional Fornasini-Marchesini model, J. Indust. Manag. Optim. 19(10)(2023), 7221-7244.

[8] Sh.Sh. Yusubov, E.N. Mahmudov, Some necessary optimality conditions for systems with fractional Caputo derivatives, J. Indust. Manag. Optim. 19(12)(2023), 8831-8850.

[9] Sh.Sh. Yusubov, E.N. Mahmudov, Necessary optimality conditions for quasi-singular controls for systems with Caputo fractional derivatives, Arch. Contr. Sci. 33(3)(2023), 463-496.

[10] Sh.Sh. Yusubov, E.N. Mahmudov, Pontryagin's maximum principle for the Roesser model with a fractional Caputo derivative, Arch. Contr. Sci. 34(2)(2024), 271-300.

[11] Sh.Sh. Yusubov, E.N. Mahmudov, Necessary optimality conditions for singular controls of Caputo fractional systems with delay in control, Qual. Theory Dynam. Syst. 24(61), (2025) https://doi.org/10.1007/s 12346-025-01225-x

Obtaining Estimates for the Rate of Convergence of the Gradient Descent via Machine Learning

Yujia Feng
Department of Applied mathematics and Control Process
Saint-Petersburg State University
Saint Petersburg, Russia
fx2000fyj@gmail.com

Nikita Izmailov
Department of Applied mathematics and Control Process
Saint-Petersburg State University
Saint Petersburg, Russia
nikitaizmaylovv@yandex.ru

***Abstract*—The aim of this study is to propose a novel approach for estimating the convergence rate of optimization algorithms using gradient descent as an example, Specifically, we aim to establish the decay law of the function value depending on the parameters of the optimized function, the starting point of the iteration process and the number of iterations. The proposed technique allows one to avoid complicated theoretical calculations that usually arise in convergence analysis and to use a trained ML model instead of a theoretical estimate for the convergence rate of a studied algorithm.**

Keywords—convergence rate, machine learning, CatBoost model, optimization, gradient descent

INTRODUCTION

It is well-known that with the improvement of computing power, the application of big data and the continuous innovation of algorithms, the performance and application scenarios of machine learning models are developing rapidly. Machine learning models are powerful in solving practical problems [1] [2] [3] [4].

Research on convergence rate has received more and more attention. In recent years, many research results on convergence rate have been published. Yang and Feng [5] derived the optimal convergence rate of CNN estimator in nonparametric regression and binary classification problems. Vojnovic and Yun [6] analyzes the convergence of gradient descent and expectation maximization algorithms for maximum likelihood estimation and maximum a posteriori probability estimation in generalized Bradley-Terry models. Xiao [7] proposes the theory of weak gradient mapping dominance and uses it to improve the sublinear convergence rate of the projected policy gradient method. However, existing research has not yet explored in depth the application of machine learning models in obtaining the convergence and convergence rate of theoretical estimates. In theory, convergence refers to the ability of an algorithm to gradually approach the optimal solution as training iterations proceed, while the convergence rate measures the time or number of iterations required for the algorithm to reach the optimal solution. Research on these aspects is of great significance because they are related to the efficiency and reliability of the algorithm in practical applications. So far, Azar and Munos [8] propose the SQL algorithm to optimize the convergence rate of standard Q-learning. Rankovic [9] proposed and evaluated the low-iteration convergence rate of four simple artificial neural network architectures.

However, there is still a large gap in the research of estimating the convergence rate of machine learning models. This paper will explore the convergence rate of the gradient descent algorithm on the optimal solution neighborhood of a more general smooth function based on the CatBoost model. As an efficient gradient boosting tree model, CatBoost has become an important technical tool in many fields and has demonstrated its powerful predictive ability in many practical problems. For example, it has achieved remarkable success in tasks such as epidemic prediction, water resources management, geological exploration and energy development. [10] [11].

DESCRIPTION OF THE DATA GENERATION PROCESS

To begin with, the convergence behavior of gradient descent algorithms exhibits significant dependence on the functional characteristics. Therefore, to systematically capture this variability, our study employs different types of two-dimensional convex quadratic functions to construct training data to cover the function variants as widely as possible. Furthermore, to assess the efficacy and generalization ability of the trained model, we will evaluate its performance on four classic benchmark optimization functions. Now we detail the methodology for generating both the training dataset and the benchmark optimization functions.

Generation of training dataset.

We define a convex quadratic function $f(x_1, x_2) = \frac{1}{2}(ax_1^2 + bx_2^2)$, where (x_1, x_2) is a two-dimensional vector, a and b are randomly generated constants. The maximum eigenvalue of the Hessian matrix of the function f at the minimum point $(0,0)$ is denoted by L, and the minimum eigenvalue is denoted by l. For this function, the Hessian matrix is a diagonal matrix with eigenvalues a and b. That is, $L = max(a, b)$ and $l = min(a, b)$. Then we randomly generate the parameters a and b, the initial point $x_0 = (x_0^1, x_0^2)$ within the range [0, 0.5], and the step size h that satisfies $h \leq \frac{1}{2L}$ where L is the maximum eigenvalue of the Hessian matrix of currently generated function($L = max(a, b)$). We follow by optimizing the generated function f for 1000 iterations using gradient descent starting from the generated initial point. During each optimization iteration, the gradient vector ∇f is computed and the parameter coordinates are updated following the gradient descent rule with step size h. For comprehensive convergence analysis, we systematically record the following quantities: (i) the current iteration count k, (ii) difference (defined as $1.1 \cdot (f(x_k) - f^*)$, where f^*represents the value of f at its global minimum), (iii) step size h, the eigenvalue pair of the Hessian matrix L and l, (iv) the initial point x_0.

And the optimization process is performed under the four different settings. Each setting perform the above

optimization process 1000, 1000, 1000, 3000 times respectively.

a is in the range of (0, 10], b is in the range of (100, 400];

a is in the range of (0, 400], $b = a$;

a and b are both in the range of (0, 5];

a and b are both in the range of (0, 500].

Generation of optimization functions.

In this section, we employ four classic benchmark optimization functions to evaluate the proposed methodology, including: Sphere, Rastrigin, Matyas and Rosenbrock. The data generation process are as follows.

The mathematical expression of Sphere function in two-dimensional space is

$$f(x_1, x_2) = x_1^2 + x_2^2.$$

Its gradient is given by

$$\nabla f(x_1, x_2) = [2x_1, 2x_2].$$

The Hessian matrix at the minimum point (0,0) is

$$H = [2\ 0\ 0\ 2\ ].$$

Consequently, all eigenvalues of the Hessian are

$$\lambda_1 = 2, \lambda_2 = 2.$$

So $L = (\lambda_1, \lambda_2) = 2, l = min(\lambda_1, \lambda_2) = 2.$

Then randomly choose h within the range [0,0.25], the initial point x_0 to be in the range [0,0.5], use the gradient descent method to iterate 1000 times, and the update rule is as follows:

$$x^{(k+1)} = x^{(k)} - h\nabla f(x^{(k)})$$

where $x^{(k)}$ represents the k-th iteration point, $\nabla f(x^{(k)})$ represents the gradient of f at $x^{(k)}$.

The mathematical expression of Rastrigin function in two-dimensional space is

$$f(x_1, x_2) = 20 + \sum_{i=1}^{2} (x_i^2 - 10\ cos\ cos\ (2\pi x_i))$$

Its gradient is given by

$\nabla f(x) = 2x + 20\pi sin(2\pi x)$, x is a vector

Consequently, all eigenvalues of the Hessian at the minimum point (0,0) are

$$\lambda_1 = \lambda_2 = 396.7842$$

So $L = (\lambda_1, \lambda_2) = 396.7842, l = min(\lambda_1, \lambda_2) = 396.7842.$

Then randomly choose h within the range [0,0.00126], the initial coordinates x_0 to be in the range [0, 0.5], use the gradient descent method to iterate 1000 times.

The mathematical expression of Matyas function in two-dimensional space is

$$f(x_1, x_2) = 0.26(x_1^2 + x_2^2) - 0.48x_1x_2$$

Its gradient is given by

$$\nabla f(x_1, x_2) = [0.52x_1 - 0.48x_2, 0.52x_2 - 0.48x_1].$$

The Hessian matrix at the minimum point (0,0) is

$$H = [0.52\ -0.48\ -0.48\ 0.52\ ].$$

Consequently, all eigenvalues of the Hessian are

$$\lambda_1 = 1, \lambda_2 = 0.04.$$

So $L = (\lambda_1, \lambda_2) = 1, l = min(\lambda_1, \lambda_2) = 0.04.$Then randomly choose h within the range [0,0.5], the initial point x_0 to be in the range [0,0.5], use the gradient descent method to iterate 1000 times

The mathematical expression of Rosenbrock function in two-dimensional space is

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + x_1^2$$

Its gradient is given by

$$\nabla f(x_1, x_2) = [2x_1 - 400(x_2 - {x_1}^2)x_1, 200(x_2 - {x_1}^2)].$$

The Hessian matrix at the minimum point (0,0) is

$$H = [2\ 0\ 0\ 200\ ].$$

Consequently, all eigenvalues of the Hessian are

$$\lambda_1 = 2, \lambda_2 = 200.$$

So $L = (\lambda_1, \lambda_2) = 200, l = min(\lambda_1, \lambda_2) = 2.$Then randomly choose h within the range [0,0.5], the initial coordinates x_0 to be in the range [0,0.5], use the gradient descent method to iterate 1000 times.

For each optimization function, after calculating L and l, at each iteration k we record the theoretical upper bound of the improvement value $f(x_k) - f^*$ of the gradient descent (the formula for the upper bound is presented below). Finally we record the iteration count k, improvement value $f(x_k) - f^*$, step size h, eigenvalues of the Hessian matrix L and l, and initial point x_0.

$$upper\ boundary = \frac{2 \cdot (f(x_0) - f^*)\|x_0 - x^*\|^2}{2\|x_0 - x^*\|^2 + k \cdot h(2 - Lh) \cdot (f(x_0) - f^*)}$$

NUMERICAL SIMULATION EXPERIMENT

The CatBoost model was selected for this study which performs well in handling large-scale, high-dimensional numerical datasets while maintaining computational efficiency. The dataset was divided into training (60%), validation (20%), and test sets (20%) . The model's effectiveness was quantified using three performance metrics and the calculation formulas for each metrics are as follows.

1. R^2.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

2. Mean Absolute Error. (MAE)

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

3. Mean Square Error. (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

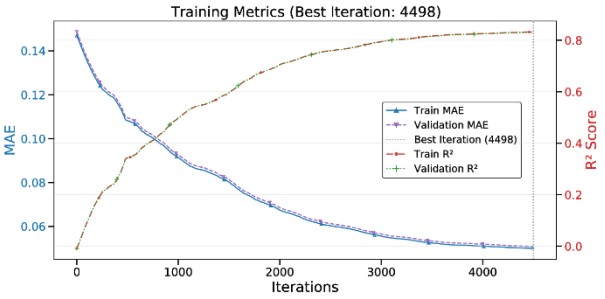
4. Root Mean Square Error. (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value.

To train the CatBoost model, we employed the learning rate η = 0.002 combined with iteration count 4500, the tree depth 13 and applying L2 regularization $\lambda = 3$. For optimization, we adopted the Mean Absolute Error (MAE) as the loss function, supplemented by R^2as the primary performance indicator.

From the learning curve (fig. 1), we can see that R^2 increases, Mean Absolute Error decreases. After that, the curve does not change much, so we conclude that the model converges well and is well trained.



Learning curve

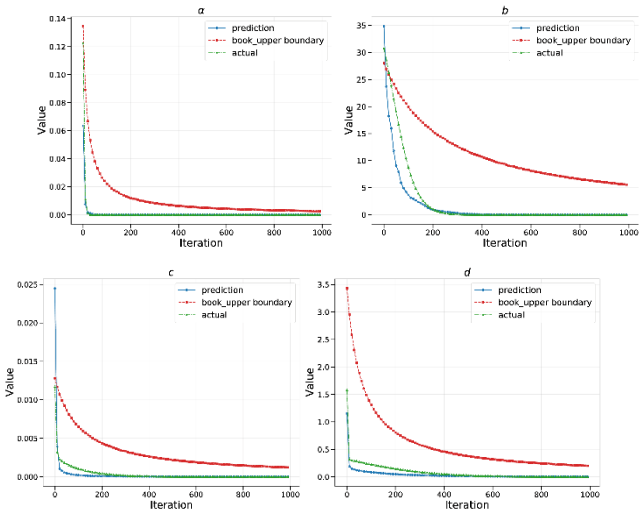
As we focus on the asymptotic properties of the estimates, that is, the behavior of the algorithm when it is close to the optimal solution, they should remain accurate over a large number of iterations, like 400. Therefore, starting from the 400th iteration, we calculate a series of evaluation metrics for the four optimized functions, including MAE, MSE, and RMSE to test the generalization ability of the model under different optimization environments.

EVALUATION METRICS

Optimization Function	Evaluation Metrics		
	Mean Absolute Error	*Mean Square Error*	*Root Mean Square Error*
Matyas	0.00001	0	0.00001
Sphere	0	0	0
Rastrigin	0.01128	0.00034	0.01843
Rosenbrock	0.00719	0.00012	0.01073

Using the trained model we can plot the predicted improvement values, actual improvement values, and the theoretical estimate $\frac{2\cdot(f(x_0)-f^*)\|x_0-x^*\|^2}{2\|x_0-x^*\|^2+k\cdot h(2-Lh)\cdot(f(x_0)-f^*)}$from book [1] for the Sphere, Rastrigin, Matyas, and Rosenbrock functions.

From fig. 2, we find that starting from a large iteration number, the predicted improvement value of the function is close to the actual value and converges to the actual improvement value asymptotically.



Prediction image of the optimization functions: (a) Sphere, (b) Rastrigin, (c) Matyas, (d) Rosenbrock

CONCLUSION

Our experimental results demonstrate that when the gradient descent step size h satisfies $h \leq \frac{1}{2L}$ where L is the maximum eigenvalue of the current Hessian matrix, the prediction effect of the model we trained on the optimization functions meets our expectations. The performance on the optimization function confirms that our model is feasible. These findings establish that our approach provides reliable convergence rate estimates for the gradient descent method.

REFERENCES

[1] Nesterov Y. Introductory lectures on convex optimization: A basic course[M]. Springer Science & Business Media, 2013.

[2] Kourou K, Exarchos T P, Exarchos K P, et al. Machine learning applications in cancer prognosis and prediction[J]. Computational and structural biotechnology journal, 2015, 13: 8-17.

[3] Kruppa J, Ziegler A, König I R. Risk estimation and risk prediction using machine-learning methods[J]. Human genetics, 2012, 131: 1639-1654.

[4] McCoy J T, Auret L. Machine learning applications in minerals processing: A review[J]. Minerals Engineering, 2019, 132: 95-109.

[5] Yang Y, Feng H, Zhou D X. On the rates of convergence for learning with convolutional neural networks[J]. arXiv preprint arXiv:2403. 16459, 2024.

[6] Vojnovic M, Yun S Y, Zhou K. Convergence rates of gradient descent and MM algorithms for Bradley-Terry models[C]//International Conference on Artificial Intelligence and Statistics. PMLR, 2020: 1254-1264.

[7] Xiao L. On the convergence rates of policy gradient methods[J]. Journal of Machine Learning Research, 2022, 23(282): 1-36.

[8] Chen M, Poor H V, Saad W, et al. Convergence time minimization of federated learning over wireless networks[C]//ICC 2020-2020 IEEE International Conference on Communications (ICC). IEEE, 2020: 1-6.

[9] Rankovic D, Rankovic N, Ivanovic M, et al. Convergence rate of Artificial Neural Networks for estimation in software development projects[J]. Information and Software Technology, 2021, 138: 106627.

[10] Kim J S . COVID-19 Prediction and Detection Using Machine Learning Algorithms: Catboost and Linear Regression[J]. Science Publishing Group, 2021(5).

[11] Lu C, Zhang S, Xue D, et al. Improved estimation of coalbed methane content using the revised estimate of depth and CatBoost algorithm: A case study from southern Sichuan Basin, China[J]. Computers & Geosciences, 2022, 158: 104973.

Radial Epiderivative Calculus: Explicit Formulas for Structured Nonsmooth Functions

Mohamed Muhumed Hassan
Department of Mathematics
Eskisehir Technical University
Eskisehir, Turkey
harirdigly@gmail.com

Refail Kasimbeyli
Department of Industrial Engineering
Eskisehir Technical University
Eskisehir, Turkey
rkasimbeyli@eskisehir.edu.tr

***Abstract*—This paper develops the concept of the radial epiderivative for general nonsmooth, nonconvex functions, extending classical directional derivatives to broader settings. We derive explicit formulas for radial epiderivatives of classes of structured nonsmooth functions. These constructive results deepen theoretical understanding of nonsmooth analysis and provide a foundation for algorithmic development in nonsmooth and nonconvex optimization.**

Keywords—radial epiderivative, structured nonsmooth functions, nonconvex optimization, generalized differentiation.

INTRODUCTION

Classical differential calculus often fails to describe the local behavior of functions that are nondifferentiable or nonconvex. Many optimization problems involve such functions, where standard directional derivatives are inadequate. The radial epiderivative overcomes these limitations by exploiting the geometric properties of the function's epigraph.

Over the past decades, several generalized derivative frameworks have been developed to address nonsmoothness in optimization. Notable among these are the subdifferential calculus by Rockafellar, the generalized gradient by Clarke, the quasidifferential constructions by Dem'yanov and Rubinov, and the bundle method foundations laid by Shor and Pshenichnyi. These tools have played a fundamental role in nonsmooth analysis and have enabled the formulation of necessary and sufficient optimality conditions for broad classes of functions.

While this paper builds on that tradition, it introduces the concept of the radial epiderivative as a distinct and constructive alternative. Unlike classical subdifferentials that often yield set-valued or implicit characterizations, the radial epiderivative is formulated using epigraph geometry and directional limits. This approach leads to explicit analytical expressions for a wide range of structured nonsmooth functions, thereby enhancing both theoretical transparency and computational applicability.

The concept of a radial epiderivative for set-valued mappings was initially introduced by Flores-Bazán [2,3], whose formulation operates within a different theoretical framework. In subsequent work, a separate notion of the radial epiderivative has been developed, distinct from Flores-Bazán's approach in both methodology and application. This later definition shares structural similarities with the contingent epiderivative for set-valued maps proposed by Jahn and Rauh [1], particularly in its geometric interpretation and analytical mechanics. However, the proposed radial epiderivative introduces independent definitions, conceptual foundations, and analytical tools tailored to address specific challenges in set-valued optimization. While inspired by prior frameworks, its formulation and theoretical underpinnings reflect a self-contained approach, differing in both scope and technique from existing derivatives.

This paper presents a comprehensive exposition of the radial epiderivative. We introduce the necessary mathematical background, formalize the core definitions, and prove key results with full generality. In particular, we derive explicit formulas for the radial epiderivatives of structured nonsmooth functions, including affine-norm compositions, pointwise extrema, convex combinations, piecewise definitions, and composite mappings. These developments provide a foundation for advancing both theory and computation in nonsmooth, nonconvex optimization.

The explicit formulas for the radial abrasive derivative developed in this paper form the analytical backbone for new classes of optimization algorithms designed for nonsmooth and nonconvex problems. In particular, two such algorithms were proposed in [9] , where the radial derivative was used in place of classical directional derivatives. These algorithms highlight the practical power of the radial calculus framework, offering an alternative pathway for descent and feasibility in problems where traditional methods fall short.

PRELIMINARIES AND DEFINITIONS

Definition 1. [5, Definition 1.2] *Let U be a nonempty subset of a real normed space* $(Z, ||\cdot||_Z)$ *and let* $z \in cl(U)$*. The closed radial cone at z, denoted* $R(U,z)$*, is defined as the set of all vectors* $w \in Z$ *for which there exists a sequence of positive scalars* $\{\lambda_n\}$ *and a sequence* $\{z_n\} \subset U$ *such that*

$$\lambda_n (z_n - z) = w.$$

Equivalently, $R(U,z) = cone(U - z)$ *is the minimal closed conical set containing* $U - z$.

Definition 2. [5, Definition 1.3] *Let* $(X, ||\cdot||_X)$ *and* $(Y, ||\cdot||_Y)$ *be real normed spaces and* $S \subset X$ *be nonempty. For a set-valued mapping* $f: S \to Y$:

The graph *of* f *is*

$$graph\,(f) = \{(x,y) \in X \times Y \mid x \in S, y \in f(x)\}.$$

The domain *of* f *is*

$$domain(f) = \{x \in X \mid f(x) \neq \emptyset\}.$$

If Y is partially ordered by a convex cone $C \subset Y$*, the* epigraph *of* f *is defined as*

$$epi(f) = \{(x,y) \in X \times Y \mid x \in S, y \in f(x) + C\}$$

Definition 3. [4, Definition 1.6] *Let* $(X, ||\cdot||_X)$ *and* $(Y, ||\cdot||_Y)$ *be normed spaces, with Y partially ordered by a convex cone C. For a nonempty set* $S \subset X$ *and a set-valued mapping* $f: S \to Y$*, the* radial epiderivative *of* f *at a point* $(x,y) \in graph(f)$ *is defined as the unique mapping* $f^r(x,y) : X \to Y$ *satisfying*

$$epi\big(f^r(x;y)\big) = R\big(epi(f),(x,y)\big).$$

In recent advancement, R. Kasimbeyli and G. D. Yalcin [7], established an equivalent definition for the radial epiderivative

$$f^r(\underline{x};x) = \frac{f(\underline{x}+tu) - f(\underline{x})}{t}$$

Where $\underline{x}, x \in R^n$. This new definition offers significant computational advantages, providing a practical and straightforward method to calculate the radial epiderivative for any function f. If the radial epiderivative exists and is finite for every direction x, the function f is said to be radially epidifferentiable at $\underline{x}$.

MAIN THEORETICAL AND COMPUTATIONAL RESULTS

Throughout this section, the radial epiderivative of a function f at $\underline{x}$ in the direction of x is denoted $f^r(\underline{x};x)$.

Theorem 1 (Norm-linear function)**.** *Let*

$$f(x) = \langle v, x\rangle - c||x||,$$

where $v \in R^n$ and $c \in R_+$. Then the radial epiderivative of f at $\underline{x}$ in the direction of x is given by

$$f^r(\underline{x};x) = \langle v, x\rangle - c||x|| \ = f(x).$$

Proof. For all $\underline{x}, x \in R^n$, we have

$$f^r(\underline{x};x) = \frac{f(\underline{x}+tu) - f(\underline{x})}{t}$$

$$= \frac{\langle v, \underline{x}+tu\rangle - c\big|\underline{x}+tu\big|| - \big[\langle v,\underline{x}\rangle - c||\underline{x}||\big]}{t}$$

Since $t > 0$, and using the triangle inequality $||\underline{x}+tu|| \le ||\underline{x}\,|| + t||u||$ we have

$$\ge \frac{\langle v, \underline{x}+tu\rangle - c||\underline{x}|| - ct||u|| - \langle v,\underline{x}\rangle + c||\underline{x}||}{t}$$

$$= \langle v, x\rangle - c||x||.$$

Similarly, we can show that $f^r(\underline{x};x) \le f(x)$, thus

$$f^r(\underline{x};x) = f(x). \qquad \square$$

Theorem 2 (Norm-affine function). *Let* $f(x) = \langle v, x \pm a\rangle - c||x \pm b|| + \alpha$. *Then* $f^r(\underline{x};x) = \langle v, x\rangle - c||x||$.

Proof. A translation does not alter the local behavior in the limit. The difference quotient cancels the constant shift, and by an argument analogous to Theorem 1 yields the result. $\square$

Corollary 1. *Let* $f(x) = \langle v, x \pm a\rangle + c||x \pm b|| + \beta$, *where* $v \in R^n$, $c \in R_+$ *and* $\beta \in R$.*Then the radial epiderivative at b in the direction of x is given by*

$$f^r(b;x) = \langle v, x\rangle + c||x||.$$

Proof. The proof follows directly from Theorem 1. $\square$

Pointwise Minima and Maxima

Theorem 3 (Radial Epiderivative of a Minimum in R)**.** *Let, for* $i = 1,2,$

$$f_i(x) = \langle v_i, x - a_i\rangle - c_i|x - b_i| + \alpha_i$$

and define $g(x) = \{f_1(x), f_2(x)\}$.

Then $g^r(\underline{x};x) = ax - b|x|,$

where the constants a and b are determined by

$$a + b = \{v_1 + c_1, v_2 + c_2\},$$

$$a - b = \{v_1 - c_1, v_2 - c_2\}.$$

Proof. Analyze the difference quotient for $g(x)$ by considering the two cases $f_1(x) \le f_2(x)$ and $f_2(x) \le f_1(x)$. In each case, the limit yields a linear function, and the coefficients must be chosen to match the largest and smallest slopes among the functions. The resulting system yields formulas for $a + b$ and $a - b$.

Example 1. *Let*

$$f_1(x) = 2(x-2) - 3|x-1|,$$

$$f_2(x) = 5(x-3) - 2|x-3|.$$

For $g(x) = \{f_1(x), f_2(x)\}$, *compute:*

$$a + b = \{2 + 3{,}5 + 2\} = 7\ ,$$

$$a - b = \{2 - 3{,}5 - 2\} = -1.$$

Thus $a = 3$ *and* $b = 4$, *so*

$$g^r(\underline{x};x) = 3x - 4|x|.$$

Corollary 2 (Extension to m functions)**.** *If*

$$g(x) = \big\{\langle v_i, x - a_i\rangle - c_i|x - b_i| + \alpha_i\ |, i = 1,2,\dots,m\big\},$$

then

$$g^r(\underline{x};x) = ax - b|x|,$$

With $a + b = \{v_1 + c_1, v_2 + c_2\}$, $a - b = \{v_1 - c_1, v_2 - c_2\}$, *where* $a, b \in R$.

Proof. The proof of this corollary follows immediately from Theorem 3. $\square$

Theorem 4 (Linearity of the radial epiderivative for certain functions)**.** *Let* $g_i: R^n \to R$ *such that* $g_i(x_i) = v_i x_i - c|x_i|$ *and suppose that*

$$f(x) = \sum_{i=1}^{n} g_i(x_i),$$

then the radial epiderivative of $f(x)$ is

$$f^r(\underline{x};x) = \left(\sum_{i=1}^{n} g_i(\underline{x}_i;x_i)\right)^r = \sum_{i=1}^{n} g_i^r(\underline{x}_i;x_i).$$

In other words, for these functions, the radial epiderivative of a sum is equal to the sum of the radial epiderivatives.

Proof. For all $\underline{x}, x \in R^n$, we have

$$f^r(\underline{x};x) = \frac{f(\underline{x}+tx) - f(\underline{x})}{t}$$

$$= \left(\sum_{i=1}^{n} \ [v_i(\underline{x}_i + tx_i) - c|\underline{x}_i + tx_i|] - [v_i\underline{x}_i - c|\underline{x}_i|\,]\right)/t$$

Using the triangle inequality, $\big|\underline{x}+tx\big| \ge t|x| - \big|\underline{x}\big|$, we have

$$\le \sum_{i=1}^{n} \ v_i x_i - c|x_i| + \frac{2c\big|\underline{x}_i\big|}{t}$$

$$f^r(\underline{x};x) \le \sum_{i=1}^{n} v_i x_i - c|x_i| = g_1 + g_2 + g_3 + \cdots + g_m$$

We know from *theorem 1* that $g_i = g_i^r$, thus we have

$$f^r(\underline{x};x) \le \sum_{i=1}^{n} g_i^r(\underline{x}_i;x_i) = \sum_{i=1}^{n} g_i(\underline{x}_i;x_i) = f(x).$$

Similarly, we can show that

$$f^r(\underline{x};x) \ge \sum_{i=1}^{n} g_i(\underline{x}_i;x_i) = f(x)$$

Thus we have

$$f^r(\underline{x};x) = \sum_{i=1}^{n} g_i(\underline{x}_i;x_i) = f(x). \qquad \square$$

Theorem 5. *Let* $g: R^n \to R$ *be defined by*

$$g(x) = \{f_1(x), f_2(x), \dots, f_m(x)\},$$

where each $f_i(x) = \langle v_i, x - a_i\rangle - c_i||x - b_i||_1 + \alpha_i$, *with* $v_i, a_i, b_i \in R^n$, $c_i \ge 0$, *and* $||\cdot||_1 = \sum |x_i|$. *Then* g *is radially epidifferentiable at any* $\underline{x} \in R^n$, *and its radial epiderivative is*

$$g^r(\underline{x};x) = \sum_{j=1}^{m} (a_j x_j - b_j|x_j|),$$

where for each $j \in \{1,2,\dots,m\}$,

$$a_j + b_j = \{v_{ij} + c_i\} \text{ and } a_j - b_j = \{v_{ij} - c_i\}$$

Proof. The proof follows from coordinate-wise analysis for each x_i and applying Theorem 4.

Example 2. *Let* $f_1(x) = 2x_1 + 3x_2 - x_3 - 4||x||_1$, *and* $f_2(x) = 5x_1 - 2x_2 + 4x_3 - 3||x||_1$. *Define* $g(x) = \{f_1(x), f_2(x)\}$. *Then:*

For x_1:

$$a_1 + b_1 = \{6,8\} = 8$$

$$a_1 - b_1 = \{-2,2\} = -2$$

Solving for a_1 *and* b_1, *we have*

$$a_1 = 3, b_1 = 5$$

For x_2:

$$a_2 + b_2 = \{7,1\} = 7$$

$$a_2 - b_2 = \{-1,-5\} = -5$$

determining a_2 *and* b_2, *we obtain*

$$a_2 = 1, \qquad b_2 = 6$$

For x_3:

$$a_3 + b_3 = \{3,7\} = 7$$

$$a_3 - b_3 = \{-5,1\} = -5$$

$$\therefore\ a_3 = 1, b_3 = 6$$

Thus, the radial epiderivative of g *is*

$$g^r(\underline{x};x) = 3x_1 - 5|x_1| + x_2 - 6|x_2| + x_3 - 6|x_3|.$$

Building on our earlier analysis of the pointwise minimum, we now derive analogous properties and results for the pointwise maximum. These findings mirror the structure and implications of those established for the minimum, illustrating the symmetry between the two operations in theoretical and applied contexts.

Theorem 6 (Radial Epiderivative of max of two functions)**.** *Let, for* $i = 1,2$,

$$f_i(x) = \langle v_i, x - a_i\rangle - c_i|x - b_i| + \alpha_i$$

and define

$$g(x) = \{f_1(x), f_2(x)\}.$$

Then the radial epiderivative of f at $\underline{x}$ *in the direction of* x *is*

$$g^r(\underline{x};x) = ax - b|x|$$

where

$$a + b = \{v_1 + c_1, v_2 + c_2\},$$

$$a - b = \{v_1 - c_1, v_2 - c_2\}.$$

Proof. A similar analysis to Theorem 3 can be applied with the roles of the extreme values reversed due to the maximum operation. $\square$

Example 3. *Let* $f_1(x) = 2(x-2) - 3|x-1|$, $f_2(x) = 5(x-3) - 2|x-3|$. *For* $g(x) = \{f_1(x), f_2(x)\}$, *we have*

$$a + b = \{2 + 3{,}5 + 2\} = 5,$$

$$a - b = \{2 - 3{,}5 - 2\} = 3.$$

Thus, $a = 4$ *and* $b = 1$, *so that*

$$g^r(\underline{x};x) = 4x - |x|.$$

Corollary 3 generalizes Theorem 6 to any number of functions.

Corollary 3. *If* $g(x) = \{\langle v_i, x - a_i\rangle - c_i|x - b_i| + \alpha_i\}$ *where* $i = 1,2,3,\dots,m$, *then*

$$g^r(\underline{x};x) = ax - b|x|$$

with

$$a + b = \{v_1 + c_1, \dots, v_m + c_m\},$$

$$a - b = \{v_1 - c_1, \dots, v_m - c_m\}.$$

Proof. It follows immediately from Theorem 6. $\square$

Theorem 7 extends Theorem 6 to R^n.

Theorem 7. *Let* $g: R^n \to R$ *be defined by*

$$g(x) = \{f_1(x), f_2(x), \dots, f_m(x)\},$$

where each $f_i(x) = \langle v_i, x - a_i\rangle - c_i||x - b_i||_1 + \alpha_i$, *with* $v_i, a_i, b_i \in R^n$, $c_i \ge 0$. *Then* g *is radially epidifferentiable at any* $\underline{x} \in R^n$, *and*

$$g^r(\underline{x};x) = \sum_{j=1}^{m} (a_j x_j - b_j|x_j|),$$

where for each j,

$$a_j + b_j = \{v_{ij} + c_i\}, \text{ and } a_j - b_j = \{v_{ij} - c_i\}.$$

Example 4. *Let*

$$f_1(x) = 2x_1 + 3x_2 - x_3 - 4||x||_1,$$

$$f_2(x) = 5x_1 - 2x_2 + 4x_3 - 3||x||_1,$$

and $g(x) = \{f_1(x), f_2(x)\}$ *. Then:*

For x_1:

$$a_1 + b_1 = 6$$

$$a_1 - b_1 = 2$$

$$a_1 = 4, \qquad b_1 = 2.$$

For x_2:

$$a_2 + b_2 = 1$$

$$a_2 - b_2 = -1$$

$$a_2 = 0, \qquad b_2 = 1$$

For x_3:

$$a_3 + b_3 = 3$$

$$a_3 - b_3 = 1$$

$$a_3 = 2, \qquad b_3 = 1$$

Thus, the radial epiderivative of g *is*

$$g^r(\underline{x}; x) = 4x_1 - 2|x_1| - |x_2| + 2x_3 - |x_3|.$$

Remark 1 (Geometric Interpretation)**.** *In the formulas for the radial epiderivative of pointwise minima and maxima, the expressions for* $a + b$ *and* $a - b$ *are determined by selecting the extreme values among the affine components. This reflects the geometry of the epigraph, capturing the extreme directional slopes.*

Remark 2. (Pointwise Extrema for Finite Collections)**.** Let $f_1, f_2, \dots, f_k : R^n \to R$ be a finite collection of real-valued functions. For any $x \in R^n$, then

$$f_i(x) = inf\ inf\ f_i(x)\ , f_i(x) = sup\ sup\ f_i(x) \quad \forall x \in R^n$$

This holds unconditionally due to the finiteness of $f(x)$ at each x which ensures extrema are attained.

The results in Remark 2 are independent of dimensions (n), the structure of f (eg., convexity), and the choice of norms in thef_i. Remark 2 also implies that the results established in Theorems 3-7 and Corollaries 2-3 extend naturally to the infimum and supremum operators.

Piecewise-Defined Functions

Theorem 8 (Piecewise Function in *R*)**.** *Consider*

$$f(x) = \begin{cases} v_1 x - c_1|x|, & x \le 0, \\ v_2 x - c_2|x|, & x > 0. \end{cases}$$

Then $f^r(\underline{x}; x) = ax - b|x|,$

with $a + b = v_1 + c_1$ *and* $a - b = v_2 - c_2.$

Proof. For $x \le 0$ the one-sided limit is computed using $f(x) = v_1 x - c_1|x|$, and for $x > 0$ using $f(x) = v_2 x - c_2|x|$. The continuity at $x = 0$ and the linear behavior imply the existence of a and b satisfying the given relations. □

Example 5. *Suppose*

$$f(x) = \begin{cases} 2x - 3|x| & x \le 0 \\ 3x - |x| & x > 0 \end{cases}$$

Then $a + b = 2 + 3 = 5$ and $a - b = 3 - 1 = 2$. Solving this system of linear equations, we have

$$a = \frac{7}{2}, \quad and \quad b = \frac{3}{2}$$

so that

$$f^r(\underline{x}; x) = \frac{7}{2}x - \frac{3}{2}|x|.$$

Theorem 9 (Extension to R^n)**.** *Let*

$$f(x) = \begin{cases} -c_1||x + a|| + \alpha, & x_i \le 0 \\ -c_2||x + b|| + \beta, & otherwise. \end{cases}$$

where $x = (x_1, x_2, \dots, x_n)$*. Then*

$$f^r(\underline{x}; x) = \begin{cases} -c_1||x||, & x_i \le 0, \\ -c_2||x||, & otherwise. \end{cases}$$

Proof. Using the triangle inequality, translations a or b do not affect the linear rate for $t > 0$, so the radial epiderivative is determined solely by c_1 or c_2.

Linearity and Convex Combinations

Theorem 10 (Linearity of Sum and Difference)**.** *Suppose*

$$f_1(x) = v_1 x \pm c_1|x| \text{ and } f_2(x) = v_2 x \pm c_2|x|.$$

Then $(f_1 \pm f_2)^r(\underline{x}; x) = f_1^r(x) \pm f_2^r(x).$

Proof. The difference quotient for $f_1 \pm f_2$ is the sum (or difference) of the difference quotients of f_1 and f_2. Since each satisfies $f_i^r(\underline{x}; h) = f_i(x)$, the result follows. □

Theorem 11 (Convex Combination)**.** *Let*

$$f_1(x) = v_1 x - c_1|x|, \text{ and } f_2(x) = v_2 x - c_2|x|.$$

Define $f(x) = \lambda f_1(x) + (1 - \lambda) f_2(x), \quad \lambda \in [0,1].$

Then $f^r(\underline{x}; x) = ax - b|x|$

Where $a = \lambda v_1 + (1 - \lambda) v_2$ *and* $b = \lambda c_1 + (1 - \lambda) c_2.$

Proof. Since the radial epiderivative is linear over convex combinations, and $f_i^r(\underline{x}; x) = f_i(x)$ for $i = 1,2$, the convex combination preserves the property.

Example 6. *Let*

$$f_1(x) = 2x - 3|x|, \text{ and } f_2(x) = x - 3|x|.$$

Then

$$f(x) = \lambda(f_1(x)) + (1 - \lambda)(f_2(x)) = (1 + \lambda)x - 3|x|.$$

The radial epiderivative of $f(x)$ *is*

$$f^r(\underline{x}; x) = (1 + \lambda)x - 3|x|.$$

Theorem 12 (Extension to R^n)**.** *For* $i = 1,2$*, let*

$$f_i(x) = \sum_{j=1}^{n} v_{ij} x_j - c_i \sum_{j=1}^{n} |x_j|,$$

and define $f(x) = \lambda f_1(x) + (1 - \lambda) f_2(x), \ \lambda \in [0,1].$

Then

$$f^r(\underline{x}; x) = \sum_{j=1}^{n} \left(a_j x_j - b_j |x_j|\right),$$

where for each j,

$$a_j = \lambda v_{1j} + (1 - \lambda) v_{2j}, \qquad b_j = \lambda c_1 + (1 - \lambda) c_2.$$

Proof. The result follows by applying the one-dimensional convex combination result coordinate-wise. □

Example 7 (Convex Combination in R^n)**.** *Let*

$$f_1(x) = 2x_1 - x_2 - 3(|x_1| + |x_2|),$$

$$f_2(x) = x_1 - 3x_2 - 2(|x_1| + |x_2|).$$

For $f(x) = \lambda f_1(x) + (1-\lambda) f_2(x)$ *with* $\lambda = 1/4$, *one obtains*

$$f^r(\underline{x}; x) = \frac{5}{4}x_1 - \frac{5}{2}x_2 - \frac{9}{2}(|x_1| + |x_2|).$$

Composition of Functions

Theorem 13 (Composition of Functions)**.** *Let*

$$f_1(x) = v_1 x - c_1|x|, \text{ and } f_2(x) = v_2 x - c_2|x|.$$

Then the radial epiderivative of the composition $(f_1 \circ f_2)$ *is given by*

$$(f_1 \circ f_2)^r(\underline{x}; x) = v_1(v_2 x - c_2|x|) - c_1(ax + b|x|),$$

where

$$a = \frac{|v_2 - c_2| - |v_2 + c_2|}{2},$$

$$b = \frac{|v_2 - c_2| + |v_2 + c_2|}{2}.$$

Proof. Write the difference quotient for the composition of f_1 and f_2 $(f_1 \circ f_2)$ and linearize the inner function f_2. Approximating $\left|v_2 x - c_2|x|\right|$ by a linear function $ax + b|x|$ yields the desired expression. □

Example 8 (Composition)**.** *For* $f_1(x) = x - 3|x|$ *and* $f_2(x) = 2x - 5|x|$, *we obtain*

$$(f_1 \circ f_2)^r(\underline{x}; x) = 8x - 20|x|,$$

$$(f_2 \circ f_1)^r(\underline{x}; x) = 7x - 21|x|.$$

REFERENCES

[1] J. Jahn and R. Rauh, "Contingent epiderivatives and set-valued optimization," Math. Methods Oper. Res., vol. 46, pp. 193–211, 1997.

[2] F. Flores-Bazan, "Optimality conditions in nonconvex set-valued optimization," Math. Methods Oper. Res., vol. 53, pp. 403–417, 2001.

[3] F. Flores-Bazan, "Radial epiderivatives and asymptotic functions in nonconvex vector optimization," SIAM J. Optim., vol. 14, no. 1, pp. 284–305, 2003.

[4] R. Kasimbeyli, "Radial epiderivatives and set-valued optimization," Optimization, vol. 58, no. 5, pp. 521–534, 2009.

[5] R. Kasimbeyli and M. Mammadov, "On weak subdifferentials, directional derivatives, and radial epiderivatives for nonconvex functions," SIAM J. Optim., vol. 20, no. 2, pp. 841–855, 2009.

[6] G. D. Yalcin and R. Kasimbeyli, "On radially epidifferentiable functions and regularity conditions in nonconvex optimization," arXiv preprint arXiv:2209.02118, 2022. [Online]. Available: https://arxiv.org/abs/2209.02118

[7] G. D. Yalcin and R. Kasimbeyli, "Generalized derivatives and optimality conditions in nonconvex optimization," Bull. Malays. Math. Sci. Soc., vol. 47, no. 3, p. 81, 2024.

[8] R. Kasimbeyli, G. D. Yalcin, G. B. Yildiz, and E. Ozcetin, "Radial epiderivative-based line search methods in nonconvex and nonsmooth box-constrained optimization," arXiv preprint arXiv:2504.05090, 2025. [Online]. Available: https://arxiv.org/abs/2504.05090

[9] M.M.Hassan. "Optimality Conditions and Solution Methods for Some Classes of Nonconvex and Nondifferentiable Optimization Problems," Ph.D. Dissertation, Eskisehir Technical University, Eskisehir (2025)

A New Variant of the Brown Bear Optimizer Driven with Chebyshev Chaotic Map Approximation

Iclal GOR
Department of Mathematics
Faculty of Science
Aydin Adnan Menderes University
Aydin, Turkiye
iclal@adu.edu.tr, 0000-0002-1999-8283

***Abstract*: This work proposes a novel variation of the Brown Bear Optimization Algorithm (BOA) by integrating Chebyshev chaotic map into its design. The Brown Bear Optimization Algorithm, while capable of maintaining a balance of exploration and exploitation, may suffer from low population diversity and premature convergence, particularly in multimodal search spaces. These issues can lead to suboptimal solutions and reduced global search capability. To evaluate the effectiveness of this modification, Chebyshev chaotic map is used. The performance of the improved algorithm is evaluated using some different types of standard benchmark functions and compared to the original BOA. The experimental results suggest that the integrating Chebyshev chaotic map may offer improvements in the solution quality of the algorithm. The results highlight the potential of chaos-based strategies to improve the exploration capacity of the metaheuristic algorithms.**

Keywords: global optimization; metaheuristics; nature-inspired algorithm; brown bear optimization; chaotic map; benchmark test function.

INTRODUCTION

Optimization algorithms are designed to get the minimum or maximum value of the objective function with its constraints. Nature-inspired optimization algorithms are the variety of the optimization methods, which focus on the natural systems of the nature for the solving engineering problems. One of these algorithms is the Brown Bear Optimization Algorithm (BOA), which is inspired by the pedal scent-marking behavior of large-sized bears known as Ursus arctos, observerd in North America and the northern regions of Eurasia [1]. This optimization algorithm is developed based on two fundamental phases such as the pedal scent marking and the sniffing of the brown bear's proposed by Prakash et al. [2]. BOA utilizes the approach, which the bears actively perform using the pedal scent marking, also communicate via the trace that includes the scent.

In the literature, chaotic approaches are proposed to improve the performance of various bio-inspired optimization algorithms. Some examples include Grey Wolf Optimizer (GWO) [3], Local Search-Based Differential Evolution Algorithms (LSDE) [4], Social Group Optimization (SGO) [5], Coyote Optimization Algorithm (COA) [6], Jaya Optimization Algorithm (JOA) [7], Slime Mould Algorithm (SMA) [8], Multi-Swarm Particle Swarm Optimization (MSPSO) [9], Wind-Driven Optimization (WDO) [10], White Shark Optimizer (WSO) [11], Aquila Optimization Algorithm (AOA) [12], Harris Hawks Optimization (HHO) [13], Butterfly Optimization Algorithm (BOA) [14], Whale Optimization Algorithm (WOA) [15], Seahorse Optimization Algorithm (SHO) [16], Bat Swarm Optimization (BSO) [17], Fruit Fly Optimization Algorithm (FOA) [18], and Biogeography-Based Optimization (BBO) [19].

The paper has the following structure: The following part of the paper provides the detailed explanation of the BOA. Next, the Chebyshev chaotic map approach is explained indetail. Following that, the proposed novel optimization approach is explained called Chaotic Brown Bear Optimization Algorithm (cBOA). The other section of the paper is about the performance analysis of the classical BOA and chaotic BOA with solving traditional benchmark functions. Finally in the last section, the results obtained from optimization algorithms is discussed.

BROWN BEAR OPTIMIZATION ALGORITHM

Inspritaion of BOA

The brown bear optimization algorithm (BOA) is inspired by some behaviors of the brown bears such as pedal scent marking, sniffing and movement strategies. These animals has the hunting methods for the effective exploration and exploitation process in the search space. In this optimization process, for the dynamic balance of BOA's, the algorithm has some aggressive and conservative behaviors.

The next subsection has the mathematical explanation of BOA.

Mathematical Models of BOA

The working principle of the Brown Bear Optimization Algorithm (BOA) can be examined under three main headings. The first stage is initial population generation, the second one is main phases and behaviors, and the third stage is algorithm output and convergence curve. BOA starts by determining an initial population randomly generated within the search space. Within this population, every candidate solution is represented by a bear. There are a total of N bears, and each bear is represented as a D-dimensional vector in the search space. The candidate solution, X_i is the position vector of the i-th bear, is shown in (1). For $i = 1,2,\cdots,N$

$$X_i = [x_{i1}, x_{i2}, \cdots, x_{iD}] \quad (1)$$

In (1), x_{ij} is the j-th component of the i-th bear's position. Each x_{ij} value is randomly initialized within the specified lower bound Min_j and upper bound Max_j of the search space using (2).

$$x_{ij} = Min_j + rand(0,1).\left(Max_j - Min_j\right) \quad (2)$$

Where $rand(0,1)$ is a uniformly distributed random number between 0 and 1. Min_j and Max_j is the minimum and maximum allowable values for the j-th variable, respectively.

The optimization algorithm randomly generated initial population forms for the foundation for the algorithm to evolve towards better solutions in the following steps. The first phase called initialization population generation. The second stage is the main stages and behaviors. The Brown Bear Optimization Algorithm uses the natural foraging behavior of brown bears in two main periods: pedal scent marking behavior (exploitation phase) and sniffing behavior (exploration phase). These two phases enable the algorithm to get the balance between the local search (exploitation) and the global search (exploration) in the optimization process.

a. Pedal Scent Marking Behavior (Exploitation)

The Brown Bear Optimization Algorithm has this stage for the emulation the way brown bears leave pedal marks while they act, which includes three different movements: gait walking, careful stepping, and foot twisting. These movements provide the algorithm to focus on promising areas in the search space and improve the solutions at the local level. $P = \frac{t}{T}$ is the ratio of the current iteration t to the total number of iterations T. Depending on the value of P, different movement strategies are applied in the following.

In (3), the gait walking are calculated for $0 < P \leq \frac{T}{3}$.

$$X_i^{new} = X_i + (-P.R_1 \circ X_i) \quad (3)$$

In (4), the careful stepping are obtained for $\frac{T}{3} < P \leq \frac{2T}{3}$.

$$X_i^{new} = X_i + Q \circ (X_{best} - S.X_{worst}) \quad (4)$$

The values of Q and S are calculated in (5) and (6), respectively.

$$Q = P.R_2 \quad (5)$$

$$S = round\left(1 + \beta_{2,k}\right) \quad (6)$$

In (6), S is the step length parameter taking values 1 or 2 and $\beta_{2,k}$ is a random number uniformly generated in the range [0,1].

In (7) the foot twisting are estimated for $\frac{2T}{3} < P \leq T$.

$$X_i^{new} = X_i + [W \circ (X_{best} - |X_i|)] - [W \circ (X_{worst} - |X_i|)] \quad (7)$$

The values of W is evaluated in (8).

$$W = 2.P.\pi.R_3 \quad (8)$$

In this movements, the values of R_1, R_2 and R_3 are the random vectors with the values between 0 and 1, which has the properties uniformly distributed, $\circ$ is the the Hadamard product, which has the element-wise multiplication, X_{best} is the best solution found so far, X_{worst} is the worst solution in the current population and S is the step size used to control exploration intensity. When the expressions such as $P.R_1$, $P.R_2$ and $P.\pi.R_3$ are utilized, the operation denotes the scalar vector multification, where each component of random values (R_1, R_2 and R_3) is multiplied by the scalar P. After the generating new positions using one of the above equations, a greedy selection mechanism is used. The greedy selection occurs, while the new position gets a better objective function value, leading to the replacement of the current position.

b. Sniffing Behavior (Exploration)

The Brown Bear Optimization Algorithm has this phase, for the exploration of the brown bears their environment by identifying other pedal marks left by other bears. This behavior helps the algorithm to discover potentially superior solutions within the search space.

For each bear X_i, the other bear X_k is randomly selected such that $k \neq i$. Next, based on the fitness evaluation, the update rule is applied in (9) and in (10).

If $f(X_i) < f(X_k)$

$$X_i^{new} = X_i + R \circ (X_i - X_k) \quad (9)$$

Else

$$X_i^{new} = X_i + R \circ (X_k - X_i) \quad (10)$$

In (9) and (10), R is a random vector between 0 and 1, which is the uniformly distributed. This mechanism promotes the movement toward better solutions while preserving diversity in the population. These two main stages (the pedal marking and the sniffing) are repeated iteratively until the stopping criteria is reached. Such stopping criteria could involve parameters like the maximum allowable number of iterations.

At the end of the optimization process, the Brown Bear Optimization Algorithm (BOA) returns the best solution in the search process. This includes three different definitions, such as the best solution, best position and the convergence curve.

Best Solution ($Best\_sol$): The highest (or the lowest, based on the problem) objective function value achieved in all iterations.

Best Position ($Best\_X$): The best objective function value of the highest (or the lowest, based on the problem) objective function value achieved in all iterations .

Convergence Curve: A vector that contains the best objective function value achieved in each iteration, allowing for the performance of the convergence analysis and the comparative studies with other algorithms.

In (11), the best solution is updated iteratively.

$$Best\_sol^t = min\left(f\left(X_1^{(t)}\right), f\left(X_2^{(t)}\right), \cdots, f\left(X_N^{(t)}\right)\right) \quad (11)$$

In (11), t represents the current iteration number, $f\left(X_i^{(t)}\right)$ denotes the best objective function value in the iteration t. The convergence curve is formulated in (12), as a vector containing the best solution at each iteration.

$$Convergence = \left[Best\_sol^{(1)}, Best\_sol^{(2)}, \cdots, Best\_sol^{(T)}\right] \quad (12)$$

This curve illustrates how quickly and effectively the algorithm converges to the optimal solution.

CHEBYSHEV CHAOTIC MAP

In general, chaos is a deterministic process that displays random behavior within nonlinear dynamical systems. It is characterized by three key properties: non-periodicity, boundedness, and extreme sensitivity to initial conditions—generally known as the 'butterfly effect'. Despite the deterministic nature of the chaotic sequences, they are highly unpredictable and uniformly distributed, making them more effective than pseudo-random sequences in preserving population diversity. Furthermore, the chaotic maps require just a few parameters such as initial value and control parameter to generate long, non-repeating sequences. A minor change in the initial condition produces a completely different sequence, enabling extensive exploration of the search space. Another advantage is their reproducible, ensuring the experimental repeatability of the algorithm. In this context, the chaotic maps act as effective sources of randomness, replacing traditional random number generators to enhance exploration capabilities and avoid premature convergence.

Chaotic maps are the discrete-time dynamical systems. Similar to the metaheuristic optimization methods, long period random number sequences play a crucial role in this algorithm. In general, the chaotic maps have successful applications in a variety of areas because of their properties. Utilizing chaotic maps can help avoid local optimum points or reduce the tendency to the likelihood of converging to the suboptimal solutions. Various chaotic maps have been produced effective results in improving the constrained stochastic nature of some original optimization algorithms [20].

In this study, Chebyshev chaotic map is used to improve the performance of classical BOA for the aiming of reducing the probability of getting stuck in the local optimum and also to increase the global convergence speed. In Chebyshev chaotic map, the generated numbers are not the same.

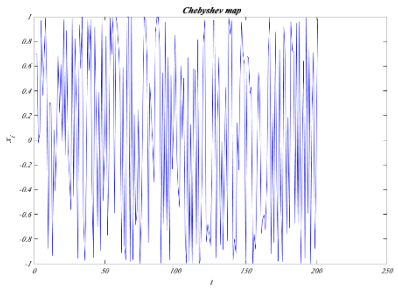
Fig. 1 demonstrates the graphs of the Chebyshev chaotic map in 200 iterations that is employed.

The Chebyshev map is given in (13).

$$x_{n+1} = cos\ cos\ (kx_n) \quad (13)$$

CHAOTIC BROWN BEAR OPTIMIZATION ALGORITHM (CBOA)

In this study, unlike the classical BOA algorithm, some random values are calculated with Chebyshev chaotic approach instead of random computation, in order to improve the exploration of the solution space. The proposed new method has been named the Chaotic Brown Bear Optimization Algorithm (cBOA). The results obtained for solving some classical benchmark functions are compared between both the classical BOA and cBOA.



Chebyshev Chaotic Map [20].

In this new method, there is a 75% probability of using chaotic Chebyshev numbers and a 25% probability of using classical uniform random numbers. This approach improves the exploration capability with using a balance of structured unpredictability and stochastic variation.

According to this novel method, the Chebyshev approach is first applied in the gait walking phase. In this manner, (14) uses Chebyshev. Other parameters used in the cBOA algorithm have been explained in detail in the previous section in classical BOA and will therefore not be discussed in this section.

$$X_i^{new} = X_i + (-P.R_{Chebshev} \circ X_i) \quad (14)$$

In the careful stepping phase of the cBOA method, the Chebyshev approach is applied with a 75% probability using (15) and (16) for the calculations.

$$Q = P.R_{Chebshev} \quad (15)$$

$$S = round(1 + R_{Chebshev}) \quad (16)$$

The W value used in the foot twisting phase is similarly calculated with a 75% probability using (17).

$$W = 2.P.\pi.R_{Chebshev} \quad (17)$$

A mutation mechanism is integrated into the cBOA during the final phase of pedal scent marking. If the current iteration ratio $P > \frac{3}{4}$, the current solution is replaced with a

newly generated random solution. In the next iteration, the optimization process continues with the updated population, potentially including newly generated random solutions that improve exploration.

EXPERIMENTAL RESULTS

In order to evaluate the performance of the proposed cBOA, four different benchmark functions is selected. These functions include Rosenbrock and Schwefel 2.22 as unimodal type and Step and Penalized2 functions as multimodal type. Table I presents the informations about these benchmark functions.

In this study, both the classical BOA and cBOA is implemented with a population size of 30 and 500 maximum iterations and with 30 independent runs conducted for each.

The benchmark functions Rosenbrock, Step, Schwefel 2.22, and Penalized2 are represented by (18), (19), (20) and (21), respectively.

$$f_1(x) = \sum_{j=1}^{D-1} \left(100\left(x_{j+1} - x_j^2\right)^2 + \left(x_j - 1\right)^2\right) \quad (18)$$

$$f_2(x) = \sum_{j=1}^{D} \left(x_j + 0.5\right)^2 \quad (19)$$

$$f_3(x) = \sum_{j=1}^{D} -x_j \, sin \, sin \left(\sqrt{|x_j|}\right) \quad (20)$$

$$f_4(x) = 0.1\Big((3\pi x_1) + \sum_{j=1}^{D} \left(x_j - 1\right)^2 (1 + (3\pi x_1 + 1)) + (x_D - 1)^2 (1 + (2\pi x_D))\Big) + \sum_{j=1}^{D} p\left(x_j, 5, 100, 4\right) \quad (21)$$

BENCHMARK FUNCTIONS

No	The properties of benchmark functions			
	Name	***Search Space***	***Dimension***	f_{min}
1	Rosenbrock	$(-30,30)$	30	0
2	Step	$(-100,100)$	30	0
3	Schwefel 2.22	$(-500,500)$	30	0
4	Penalized2	$(-50,50)$	30	0

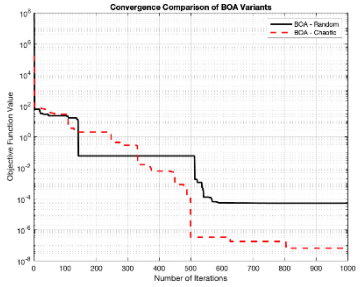
The obtained results are presented in Table II. As seen in Table II, the cBOA algorithm demonstrates superior performance compared to classical BOA in terms of best solution, mean score, and standard deviation for solving the benchmark functions.

EXPERIMENTAL RESULTS

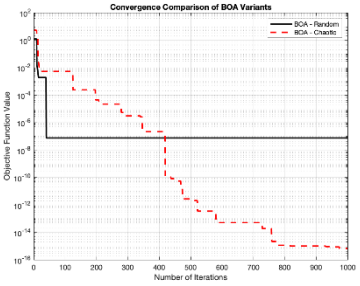
No		The results of benchmark functions solutions					
	Function Name	***Best Solution***		***Mean Score***		***Standard Deviation***	
	Method Name	***BOA***	***cBOA***	***BOA***	***cBOA***	***BOA***	***cBOA***
1	Rosenbrock	0.000051	0.000000	0.909344	0.107536	1.753354	0.273712
2	Step	0.000000	0.000000	0.007477	0.000133	0.007298	0.000425
3	Schwefel 2.22	-12569.4874962	-12569.4866587	-12561.436665	-12569.108921	7.316988	0.528261
4	Penalized 2	0.000166	0.000000	0.001095	0.000018	0.000698	0.000070

The convergence comparisons of the BOA variants has been obtained for each benchmark function, convergence comparisons of the BOA variants is obtained and is presented in Figure 2, Figure 3, Figure 4, and Figure 5 for the Rosenbrock, Step, Schwefel 2.22, and Penalized2 functions, respectively. The figures show that the applying of the Chebyshev chaotic approach to the BOA algorithm leads to more efficient solutions for these functions.

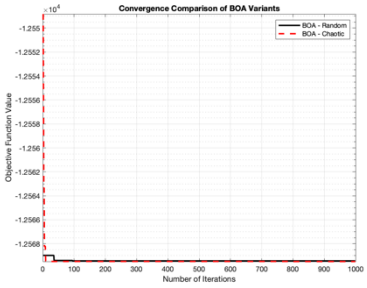
In this work, a computational efficiency analysis is included to evaluate the runtime performance of the Brown-bear Optimization Algorithm (BOA) and Chaotic Brown-bear Optimization Algorithm (cBOA). For each of the 30 independent trials, the execution time was measured using MATLAB's tic and toc. The results are added in Table III for each benchmark functions. The computational efficiency analysis indicates that the Chaotic Brown Bear Optimization Algorithm (cBOA) generally achieves superior runtime performance compared to the standard BOA. Specifically, cBOA gets faster convergence on Rosenbrock, Step, and Schwefel 2.22 functions, with lower standard deviations. Although cBOA exhibits a longer average execution time on the Penalized2 function, it gets a more stable runtime. These results show that the adding of chaotic maps, in generally, improves both efficiency and stability.



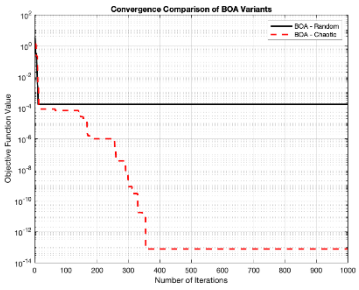
Convergence Results of BOA (BAO Random) and cBOA (BAO Chaotic) for Rosenbrock.



Convergence Results of BOA (BAO Random) and cBOA (BAO Chaotic) for Step.



Convergence Results of BOA (BAO Random) and cBOA (BAO Chaotic) for Schwefel 2.22.



Convergence Results of BOA (BAO Random) and cBOA (BAO Chaotic) for Penalized2.

COMPUTATIONAL ANALYSIS OF EXPERIMENTAL RESULTS

No		**The computational analysis of benchmark functions solutions**			
	Function Name	***Average Time (seconds)***		***Standard Deviation***	
	Method Name	***BOA***	***cBOA***	***BOA***	***cBOA***
1	Rosenbrock	0.166208	0.161041	0.020707	0.00764 7
2	Step	0.163184	0.168174	0.001149 5	0.00630 9
3	Schwefel 2.22	0.379609	0.347626	0.014009	0.00744 2
4	Penalized2	1.755071	2.128807	0.074223	0.03245 6

CONCLUSION

In this work, a novel approach has been proposed for the Brown Bear Optimization Algorithm (BOA), which is one of the nature-inspired algorithms. The original BOA is the effective method, that achieves good balance of exploration and explotaiton. However, BOA may have some problems about the convergence and the population diversity, particularly in complex search spaces. These limitations can cause the algorithm to lead to suboptimal solutions and reduced global search capability. In order to solve these issues, in the proposed algorithm, Chaotic Brown Bear Optimization Algorithm (cBOA), some random values in the classical BOA algorithm have been replaced with Chebyshev chaotic map. The aim of this modification is to increase population diversity and accelerate convergence, enhancing the algorithm's capability to avoid local optima and get the global optimal solutions. To evaluate the performance of the proposed cBOA, four different benchmark functions are solved including Rosenbrock and Step (unimodal), and Schwefel and Penalized2 (multimodal). In this study, both the classical BOA and the proposed cBOA is set with a population size of 30 and a maximum iteration count of 500, and each has been run 30 times. As a future work, it would be possible to apply the chaotic approach to other optimization methods mentioned in the literature along with the proposed approach. The codes used in this study can be accessed at the following address: https://github.com/iclalgor/Chaotic-Brown-Bear-Optimization-Algorithm-cBOA-.

REFERENCES

[1] Sergiel, J. Naves, P. Kujawski, R.Maslak, E. Serwa, D. Ramos, A. Fernández-Gil, E. Revilla, T. Zwijacz-Kozica, F. Zieba, J. Painer, and N. Selva, "Histological, chemical and behavioural evidence of pedal communication in brown bears," Scientific reports, vol. 7, 1, pp.1–10, 2017.

[2] T. Prakash, P. P. Singh, V. P. Singh, S. N. Singh, "A Novel Brown-Bear Optimization Algorithm for Solving Economic Dispatch Problem," In Advanced Control and Optimization Paradigms for Energy System Operation and Management; River Publishers: Gistrup, Denmark, 2022.

[3] M. Kohli and S. Arora, "Chaotic grey wolf optimization algorithm for constrained optimization problems, " Journal of Computational Design and Engineering, Vol. 5, pp. 458–472, 2018. https://doi.org/10.1016/j.jcde.2017.02.005.

[4] S. Gao, Y. Yu, Y. Wang, J. Wang, J. Cheng and M. Zhou, "Chaotic Local Search-Based Differential Evolution Algorithms for Optimization," in IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 51, no. 6, pp. 3954-3967, June 2021, doi: 10.1109/TSMC.2019.2956121.

[5] Naik, "Chaotic Social Group Optimization for Structural Engineering Design Problems," Journal of Bionic Engineering, vol. 20, pp. 1852–1877, 2023. https://doi.org/10.1007/s42235-023-00340-2.

[6] H. Tong, Y. Zhu, J. Pierezan, Y. Xu and L. dos Santos Coelho, "Chaotic Coyote Optimization Algorithm, " Journal of Ambient Intelligence and Humanized Computing, vol. 13, pp. 2807–2827, 2022. https://doi.org/10.1007/s12652-021-03234-5.

[7] W. Lei, Z. Zhang, J. Zhu, Y. Lin, J. Hou and Y. Sun, "An Improved Jaya Optimization Algorithm with Hybrid Logistic-Sine-Cosine Chaotic Map, " 2022 14th International Conference on Advanced Computational Intelligence (ICACI), Wuhan, China, 2022, pp. 176-181, doi: 10.1109/ICACI55529.2022.9837758.

[8] O. Altay, "Chaotic slime mould optimization algorithm for global optimization, " Artificial Intelligence Review, vol. 55, pp. 3979–4040, 2022. https://doi.org/10.1007/s10462-021-10100-5.

[9] K. Akyol and S. O. F. Feneaker, "Chaotic multi-swarm particle swarm optimization for the welded beam design engineering problem," Journal of Polytechnic, vol. 25, no. 4, pp. 1645–1660, 2022, doi: 10.2339/politeknik.880994.

[10] Z. Tang, S. Tao, K. Wang, B. Lu, Y. Todo and S. Gao, "Chaotic Wind Driven Optimization with Fitness Distance Balance Strategy," International Journal of Computational Intelligence Systems, vol. 15, no. 46, 2022. https://doi.org/10.1007/s44196-022-00099-0.

[11] F. Lepe-Silva, B. Crawford, F. Cisternas-Caneo, J. Barrera-Garcia and R. Soto, "A Binary Chaotic White Shark Optimizer," Mathematics, vol. 12, no. 20, 2024, pp. 3171. https://doi.org/10.3390/math12203171.

[12] O. E. Turgut, M. S. Turgut, M.S. and E. Kırtepe, "Chaotic Aquila Optimization Algorithm for Solving Phase Equilibrium Problems and Parameter Estimation of Semi-empirical Models, " Journal of Bionic Engineering, vol. 21, pp. 486–526, 2024. https://doi.org/10.1007/s42235-023-00438-7.

[13] H. Gezici and H. Livatyalı, "Chaotic Harris hawks optimization algorithm," Journal of Computational Design and Engineering, vol. 9, no. 1, February 2022, pp. 216–245. https://doi.org/10.1093/jcde/qwab082

[14] W. Li, J. Yang and P. Shao, "Circle Chaotic Search-Based Butterfly Optimization Algorithm,"In: Tan, Y., Shi, Y. (eds) Advances in Swarm Intelligence. ICSI 2024. Lecture Notes in Computer Science, vol 14788. 2024, Springer, Singapore. https://doi.org/10.1007/978-981-97-7181-3_10.

[15] G. I. Sayed, A. Darwish, and A. E. Hassanien, "A New Chaotic Whale Optimization Algorithm for Features Selection, " Journal of Classification, vol. 35, pp. 300-344, 2018. https://doi.org/10.1007/s00357-018-9261-2.

[16] F. A. Ozbay, " A modified seahorse optimization algorithm based on chaotic maps for solving global optimization and

engineering problems ” Engineering Science and Technology, an International Journal, vol. 41, pp. 101408, 2023. https://doi.org/10.1016/j.jestch.2023.101408.
[17] R. Jordehi, “Chaotic bat swarm optimisation (CBSO), ”Applied Soft Computing, vol. 26, pp 523-530, 2015, https://doi.org/10.1016/j.asoc.2014.10.010.
[18] M. Mitić, N. Vuković, M. Petrović and Z. Miljković, “Chaotic fruit fly optimization algorithm, ” Knowledge-Based Systems, vol. 89, pp 446-458, 2015, https://doi.org/10.1016/j.knosys.2015.08.010.
[19] W. Guo, W. Li, Q. Kang, L. Wang, and Q. Wu, “Chaotic biogeography-based optimisation, ” International Journal of Computing Science and Mathematics, vol. 5, no. 2, pp. 127-136, 2014, doi: 10.1504/IJCSM.2014.064057.
[20] S. Saremi, S. Mirjalili and A. Lewis, “Biogeography-based optimisation with chaos, ” Neural Computing and Applications, vol. 25, pp. 1077-1097, 2014. https://doi.org/10.1007/s00521-014-1597-x.

An Optimal Control Problem for Equations in Special Cases of Thin Plates with Boundary Control

Hamlet Guliyev
Department of Optimal Control
Institute of Mathematics and Mechanics
Baku, Azerbaijan
hamletquliyev51@gmail.com

Khayala Seyfullayeva
Department of Mathematical Analysis
and Differential Equations
Sumgait State University
Sumgait, Azerbaijan
0000-0002-5421-7809

***Abstract*—In the work the boundary control problem is considered for the linear equation of thin plate vibrations. The theorem on the existence and uniqueness of the optimal control is proved, differential of the functional is calculated, necessary and sufficient condition of optimality in the form of integral inequality is derived.**

***Keywords*—thin plate, optimal control, theorem of existence and uniqueness, necessary and sufficient condition.**

INTRODUCTION

It is known that a couple of physical and mechanical processes, as well as vibration of the bar, tuning-fork, elastic and thin plate, are described by the fourth order partial differential equations [1]-[3]. Therefore, investigation of the optimal control problems in the processes, described by such equations is of theoretical and practical interests. When the controls are the boundary functions, study of these problems become more complicated. But it should be noted that the boundary control problems are more natural as compare with the problems with distributed parameters. In the works [4]-[10] some close control problems are considered.

PROBLEM STATEMENT

Let the controlling process be described by the equation of the thin plate vibrations

$$\frac{\partial^2 u}{\partial t^2}+a^2\Delta^2 u=0 \quad \text{in } Q_T=\Omega\times(0,T),\ \Omega=(0,l_1)\times(0,l_2) \tag{1}$$

with initial

$$u(x_1,x_2,0)=\varphi_0(x_1,x_2),\ \frac{\partial u(x_1,x_2,0)}{\partial t}=\varphi_1(x_1,x_2),\ (x_1,x_2)\in\Omega \tag{2}$$

and boundary conditions

$$\begin{aligned}&u(0,x_2,t)=\upsilon(x_2,t),\ u(l_1,x_2,t)=0,\\ &\frac{\partial u(0,x_2,t)}{\partial x_1}=0,\ \frac{\partial u(l_1,x_2,t)}{\partial x_1}=0,\ (x_2,t)\in(0,l_2)\times(0,T),\\ &u(x_1,0,t)=0,\ u(x_1,l_2,t)=0,\\ &\frac{\partial u(x_1,0,t)}{\partial x_2}=0,\ \frac{\partial u(x_1,l_2,t)}{\partial x_2}=0,\ (x_1,t)\in(0,l_1)\times(0,T),\end{aligned} \tag{3}$$

where a^2,l_1,l_2,T are given positive numbers, $\upsilon(x_2,t)$ is a boundary control function, $\varphi_0(x_1,x_2)\in W_2^2(\Omega)$, $\varphi_1(x_1,x_2)\in L_2(\Omega)$ are given functions, Δ is Laplace operator with respect to x_1,x_2.

Consider the space of controls $H=W_2^{4,2}\left((0,l_2)\times(0,T)\right)$.

As a class of admissible controls U_{ad} we consider the set of functions $\upsilon(x_2,t)$ from H, for which hold: $\upsilon(0,t)=\upsilon(l_2,t)=0,\ \dfrac{\partial\upsilon(0,t)}{\partial x_2}=\dfrac{\partial\upsilon(l_2,t)}{\partial x_2}=0,$ $\upsilon(x_2,0)=\dfrac{\partial\upsilon(x_2,0)}{\partial t}=0$, moreover $\left\|\dfrac{\partial^4\upsilon}{\partial x_2^4}\right\|_{L_2((0,l_2)\times(0,T))}\le M_1$, $\left\|\dfrac{\partial^2\upsilon}{\partial t^2}\right\|_{L_2((0,l_2)\times(0,T))}\le M_1$, where M_1 is a given number.

Here and further on by M_i we denote different constants not depending on the admissible controls and estimating quantities.

It is supposed that the functions $\varphi_0(x_1,x_2)$ and $\upsilon(x_2,t)$ satisfy the proper conditions.

The problem is: On the set U_{ad} to find a function that gives minimum to the functional

$$J(\upsilon)=\frac{1}{2}\int_\Omega\left[u(x_1,x_2,T)\right]^2dx_1dx_2+\frac{\alpha}{2}\int_0^{l_2}\int_0^T\upsilon^2(x_2,t)dx_2dt, \tag{4}$$

Subject to (1)-(3), where $\alpha>0$ is a positive number.

As a solution of the problem (1)-(3) for each fixed admissible control $\upsilon(x_2,t)$ we take the function $u(x_1,x_2,t)\in W_2^{2,1}(Q_T)$, which for any function $\eta\in W_2^{2,1}(Q_T)$, $\eta(x_1,x_2,T)=0,$

$$\eta(0,x_2,t)=0,\ \eta(l_1,x_2,t)=0,\ \frac{\partial\eta(0,x_2,t)}{\partial x_1}=0,\ \frac{\partial\eta(l_1,x_2,t)}{\partial x_1}=0,$$

$$\eta(x_1,0,t)=0,\ \eta(x_1,l_2,t)=0,\ \frac{\partial\eta(x_1,0,t)}{\partial x_2}=0,\ \frac{\partial\eta(x_1,l_2,t)}{\partial x_2}=0$$

Satisfies to the integral identity

$$\int_{Q_T}\left[-\frac{\partial u}{\partial t}\cdot\frac{\partial \eta}{\partial t}+a^2\Delta u\Delta\eta\right]dx_1dx_2dt-$$
$$\int_{\Omega}\varphi_1(x_1,x_2)\eta(x_1,x_2,0)dx_1dx_2=0 \quad (5)$$

and conditions

$$u(x_1,x_2,0)=\varphi_0(x_1,x_2),\ u(0,x_2,t)=\upsilon(x_2,t) \quad (6)$$

in the usual sense.

Note that the theorem may be proved on the existence and uniqueness of the solution $u(x_1,x_2,t)$ of the boundary problem (1)-(3) by each fixed admissible control $\upsilon(x_2,t)$ and this solution has a property $u\in C\left([0,T];W_2^2(\Omega)\right)$, $u\in C\left([0,T];L_2(\Omega)\right)$ [11].

Main Results

Existence of the optimal control in the problem (1)-(4).

Theorem 1. Let the conditions above on the date of the problem (1)-(4) satisfied. Then there exists the unique optimal control in the problem (1)-(4).

Proof. Let $\{\upsilon_n\}\in U_{ad}$ be a minimizing sequence, i.e.

$$\lim_{n\to\infty}J(\upsilon_n)=\inf_{\upsilon\in U_{ad}}J(\upsilon). \quad (7)$$

Then from the definition of the class U_{ad} follows that

$$\|\upsilon_n\|_{W_2^{4,2}((0,l_2)\times(0,T))}\le M_2.$$

Let us denote by $u_n(x_1,x_2,t)$ the solution of the problem (1)-(3), corresponding to $\upsilon_n(x_2,t)$.

As in the first part of the work [12] for each function $u_n(x_1,x_2,t)$ the Galerkin's approach $u_n^N(x_1,x_2,t)$ to the solution $u_n(x_1,x_2,t)$ may be constructed and the following estimation may be obtained

$$\left\|u_n^N\right\|_{W_2^{2,1}(Q_T)}\le M_3.$$

From this considering weak lower semicontinuity of the norm in the Hilbert spaces we get the same estimation for $u_n(x_1,x_2,t)$, i.e.

$$\|u_n\|_{W_2^{2,1}(Q_T)}\le M_3.$$

Then taking into account the weak compactness in the Hilbert spaces we can regard that

$$\upsilon_n\to\upsilon_0 \text{ weakly in } W_2^{4,2}\left((0,l_2)\times(0,T)\right) \quad (8)$$

by $n\to\infty$ and

$$u_n\to u_0,\ \frac{\partial u_n}{\partial x_1}\to\frac{\partial u_0}{\partial x_1},\ \frac{\partial u_n}{\partial x_2}\to\frac{\partial u_0}{\partial x_2},\ \frac{\partial u_n}{\partial t}\to\frac{\partial u_0}{\partial t}$$
$$\frac{\partial^2 u_n}{\partial x_1^2}\to\frac{\partial^2 u_0}{\partial x_1^2},\ \frac{\partial^2 u_n}{\partial x_1\partial x_2}\to\frac{\partial^2 u_0}{\partial x_1\partial x_2},\ \frac{\partial^2 u_n}{\partial x_2^2}\to\frac{\partial^2 u_0}{\partial x_2^2}$$
$$\text{weakly in } L_2(Q_T). \quad (9)$$

From the relation (8) in accordance with the embedding theorems [13] follows that

$$\upsilon_n(x_2,t) \text{ converges uniformly to } \upsilon_0(x_2,t) \quad (10)$$

on $[0,l_2]\times[0,T]$ by $n\to\infty$.

From the relation (9) in accordance with the embedding theorems [14] follows that by $n\to\infty$

$$u_n(0,x_2,t)\to u_0(0,x_2,t) \text{ strongly in } L_2\left((0,l_2)\times(0,T)\right),$$
$$u_n(x_1,x_2,t)\to u_0(x_1,x_2,t) \text{ strongly in } L_2(Q_T). \quad (11)$$

In the definition of the generalized solution of the problem (1)-(3) we take $\upsilon=\upsilon_n$, $u=u_n$:

$$\int_{Q_T}\left[-\frac{\partial u_n}{\partial t}\cdot\frac{\partial \eta}{\partial t}+a^2\Delta u_n\Delta\eta\right]dx_1dx_2dt-$$
$$\int_{\Omega}\varphi_1(x_1,x_2)\eta(x_1,x_2,0)dx_1dx_2=0$$

and $u_n(x_1,x_2,0)=\varphi_0(x_1,x_2)$, $u_n(0,x_2,t)=\upsilon_n(x_2,t)$.

If to pass to the limit by $n\to\infty$ in this equalities considering (9), (10) and (11) we obtain

$$\int_{Q_T}\left[-\frac{\partial u_0}{\partial t}\cdot\frac{\partial \eta}{\partial t}+a^2\Delta u_0\Delta\eta\right]dx_1dx_2dt-$$
$$\int_{\Omega}\varphi_1(x_1,x_2)\eta(x_1,x_2,0)dx_1dx_2=0$$
$$u_0(x_1,x_2,0)=\varphi_0(x_1,x_2),\ u_0(0,x_2,t)=\upsilon_0(x_2,t).$$

These relations show that the function $u_0(x_1,x_2,t)$ is a solution of the problem (1)-(3), which corresponds to the admissible control $\upsilon_0(x_2,t)$.

Since the boundary problem (1)-(3) is linear and the functional (4) is quadratic, then (4) weak lower on U_{ad}. Therefore

$$\underline{\lim_{n\to\infty}}J(\upsilon_n)\ge J(\upsilon_0). \quad (12)$$

Then as follows from (7) and (12)

$$\inf_{\upsilon\in U_{ad}}J(\upsilon)\ge J(\upsilon_0).$$

From the last

$$\inf_{\upsilon\in U_{ad}}J(\upsilon)=J(\upsilon_0),$$

which shows that $\upsilon_0(x_2,t)$ is an optimal control in the considered problem.

From the form of the functional (4) is clear that it is strongly convex in $L_2\left((0,l_2)\times(0,T)\right)$. Therefore the optimal control is unique.

Theorem 1 is proved.

Differentialbilty of the functional (4) and necessary conditions of optimality.

Let us introduce the adjoint problem for the control

$$\frac{\partial^2\psi}{\partial t^2}+a^2\Delta^2\psi=0 \quad \text{in } Q_T, \quad (13)$$

$$\psi(x_1,x_2,T)=0,\quad \frac{\partial\psi(x_1,x_2,T)}{\partial t}=-u(x_1,x_2,T),$$
$$(x_1,x_2)\in\Omega, \tag{14}$$
$$\psi(0,x_2,t)=\psi(l_1,x_2,t)=0,$$
$$\frac{\partial\psi(0,x_2,t)}{\partial x_1}=\frac{\partial\psi(l_1,x_2,t)}{\partial x_1}=0,$$
$$(x_2,t)\in[0,l_2]\times[0,T],$$
$$\psi(x_1,0,t)=\psi(x_1,l_2,t)=0,$$
$$\frac{\partial\psi(x_1,0,t)}{\partial x_2}=0,\ \frac{\partial\psi(x_1,l_2,t)}{\partial x_2}=0,$$
$$(x_1,t)\in[0,l_1]\times[0,T], \tag{15}$$

where $u(x_1,x_2,t)$ is a solution of the problem (1)-(3) for the given control $\upsilon(x_2,t)$.

Since $u(x_1,x_2,T)\in W_2^2(\Omega)$, the adjont problem (13)-(15) has the only solution in the space $W_2^{4,2}(Q_T)$ [11].

Theorem 2. Let the condition above on the data of the problem (1)-(4) be satisfied. Then the functional (4) is continuously differentiable in the Freshet sense on H and its differential in the point $\upsilon(x_2,t)\in U_{ad}$ with increment $\delta\upsilon(x_2,t)\in H$, $\upsilon(x_2,t)+\delta\upsilon(x_2,t)\in U_{ad}$ is defined by the following relation

$$\langle J'(\upsilon),\delta\upsilon\rangle_H =$$
$$=\int_0^T\int_0^{l_2}\left[\alpha\upsilon(x_2,t)-a^2\left(\frac{\partial^3\psi(0,x_2,t)}{\partial x_1^3}+\frac{\partial^3\psi(0,x_2,t)}{\partial x_1\partial x_2^2}\right)\right]\delta\upsilon(x_2,t)dx_2dt. \tag{16}$$

Theorem 3. Let the conditions above on the data of the problem (1)-(4) be satisfied. Then for the optimality of the control $\upsilon_*(x_2,t)\in U_{ad}$ in the problem (1)-(4) it is necessary and sufficient fulfillment of the inequality

$$\int_0^T\int_0^{l_2}\left[\alpha\upsilon(x_2,t)-a^2\left(\frac{\partial^3\psi_*(0,x_2,t)}{\partial x_1^3}+\frac{\partial^3\psi_*(0,x_2,t)}{\partial x_1\partial x_2^2}\right)\right]\times$$
$$\left(\upsilon(x_2,t)-\upsilon_*(x_2,t)\right)dx_2dt\ge 0\quad\forall\upsilon\in U_{ad} \tag{41}$$

where $\psi_*(x_1,x_2,t)$ is a solution of the problem (13)-(15) by $u=u_*(x_1,x_2,t)$, and $u_*(x_1,x_2,t)$ is a solution of the problem (1)-(3) by $\upsilon=\upsilon_*(x_2,t)$.

CONCLUSION

In the proposed work, the problem of boundary optimal control for a linear equation of oscillations of a thin plate is considered. In the work, the theorem of existence and uniqueness of optimal control is proved, the differential of the functional is calculated and a necessary and sufficient condition for optimality in the form of an integral inequality is derived.

REFERENCES

[1] Tikhonov A.N., Samarskiy A.A. Equations of Mathematical Physics, Moscow, Nauka, 1972, 736 p.

[2] Komkov V. Optimal Control Theory in the Damped Vibrations of the Simple Elestic Systems, Moscow, Mir, 1975, 160 p.

[3] Arman J.-L.H. Application of the Optimal Control Theory with the Distributed Parameters Systems to the Optimal Design Problems, Moscow, Mir, 1977, 144 p.

[4] Zhang Xinong, Zhang Jinghui. The hybrid control of vibration of thin plate with active constrained damping layer, Applied Mathematics and Mechanics, (English Edition, Vol. 19, No.12, Dec., 1998).

[5] Sadek S., Adali S., Sloss J. M., Bruchjr J.C. Vibration damping of a thin plate by optimal open-and closed-loop control forces, Journal of the Franklin Institute, Pergamon Presa ple. (vol. 329, No. 2, pp. 207-214, 1992. Printed in Great Britain).

[6] Daqun Tong and Robert L. Williams II, Sunil K. Agrawal. Optimal shape control of composite thin plates with piezoelectric actuators. Journal of Intelligent Material Systems and Structures. Vol. 9, pp. 458-467, June, 1998.

[7] V.S. Deineka. Optimal control of the dynamic state of a thin compound plate. Cybernetics and Systems Analysis. Vol. 42, pp. 151-175, No 4, 2006.

[8] Maria L. Blanton & Ibrahim S. Sadek. Optimal active pointwise of tuhin plates via state-control parametrization. International Journal of Systems Science. Vol. 25. Issue 11, 1994.

[9] Guliyev, H.F., Nasibzadeh, V.N. On determining higher coefficient of a second order hyperbolic equation by the variational method // International Journal of Applied Mathematics, 2025, Volume 38, No 3, pp.323-334.

[10] Hasanov K.K., Gasumov T.M. A minimal energy control problem for second-order linear hyperbolic systems with two independent variables // Optimal Control Applications and Methods 33(1), 2012, pp.51-60.

[11] Ломовцев Ф.Е., Юрчук Н.И. Задача Коши для гиперболических дифференциально-операторных уравнений второго порядка. // Дифференциальные уравнения, 1976, Т. 12, №12, 2242-2250 с.

[12] Quliyev H.F., Seyfullayeva X.İ. Nazik lövhənin rəqsləri tənliyi üçün optimal idarəetmə məsələsi. Bakı Universitetinin Xəbərləri, Fizika-riyaziyyat elmləri seriyası. №3, 2013, səh. 64-73.

[13] Бесов О.В., Ильин В.П., Никольский С.М. Интегральные представления функций и теоремы вложения, М.: «Наука», 1975, 482 с.

[14] Лионс Ж.-Л. Управление сингулярными распределенными системами. М.: «Наука», 1987, 367 с.

[15] Ладыженская О.А. Краевые задачи математической физики, М.: «Наука», 1973, 408 с.

[16] 14. Васильев Ф. П. Методы решения экстремальных задач, М.: «Наука», 1981, 399 с.

Section 5: Methods of Image Processing and Recognition

Adaptive Traffic Light Optimization for Dynamic Urban Intersections

Zumrud Isgandarli
School of IT and Engineering
ADA University
Baku, Azerbaijan
zisgandarli15645@ada.edu.az

Ilkin Sharafkhanov
School of IT and Engineering
ADA University
Baku, Azerbaijan
isharafkhanov15578@ada.edu.az

Rufat Ismayilov
School of IT and Engineering
ADA University
Baku, Azerbaijan
rismayilov15680@ada.edu.az

Musa Afandiyev
School of IT and Engineering
ADA University
Baku, Azerbaijan
mafandiyev14056@ada.edu.az

Jamaladdin Hasanov
School of IT and Engineering
ADA University
Baku, Azerbaijan
jhasanov@ada.edu.az

***Abstract*—Traffic congestion is a major urban challenge, and smart cities increasingly adopt Adaptive Traffic Light Systems (ATLS) to address it. In Azerbaijan, traffic lights follow static schedules, requiring manual adjustments. This paper presents an Adaptive Traffic Flow Optimization System using Reinforcement Learning (RL) and Deep Q-Network (DQN) to introduce dynamic responsiveness. The Simulation Urban Mobility (SUMO) simulator is utilized to model real-world traffic sce- narios from Baku. Initially, a single-agent RL model adjusted signal durations based on vehicle-to-capacity ratios, optimizing decisions through Q-learning. Later, the DQN algorithm was implemented to conduct further experiments. The system modernizes Azerbaijan's traffic infrastructure, aligning it with advanced global solutions.**

***Index Terms*—Adaptive Traffic Light System, Rein- forcement Learning, Deep Q-Network, SUMO, Traffic Congestion, Q-learning, Smart Cities.**

I. Introduction

A. Purpose

The traffic congestion in densely populated areas, such as the capital city of Azerbaijan - Baku - continues to pose a significant problem. Smart cities all around the world try to use various solutions for tackling the mentioned problem, and one of the solutions that is being commonly implemented in developed countries is the ATLS [1]. This system uses AI-driven technology to enable the dynamic management of traffic light systems. Other methods are manual or automated management, which are not effective enough since they do not adapt to real-time traffic.

B. Project Objectives, Significance, Novelty

Communication with the Center for Intelligent Trans- port Management of Azerbaijan demonstrated that, in Azerbaijan, traffic lights are only adjusted according to the time of day. This means, it does not react to traffic flow and fluctuations happening in the traffic. It operates with predefined rules that change at certain times of the day, specifically during peak hours. Additionally, in some cases, police officers are also involved in adjusting the traffic flow. Using ML, ATLS aims to tackle this problem and seeks to bring dynamism to the static system that is currently being applied in the country.

The significance of this project includes improvement of traffic management, reduction in emissions and fuel consumption, and scalable, dynamic, real-time adaptive, and cost-effective solutions.

In terms of novelty, even though RL and DQN are well-known approaches to tackle the relevant issue and develop ATLS, they may initiate the first steps in introducing this system locally, in Azerbaijan. The first approach, with pure RL, which is minimizing the longest queue in the intersection, is unique and has been developed from scratch. Moreover, a custom sim- ulation environment has been designed, and dynamic simulation techniques has been adopted for having more realistic test results.

C. Problem statement

According to the State Statistics Committee [2], statistics show that there are more than 1 million 700 hundred registered personal vehicles in the country, and the number is growing exponentially every day. This not only creates discomfort for drivers but also harms the environment. For example, according to Urban Mobility in 2019, 8.8 billion hours and 3.3 gallons of fuel were lost by drivers because of traffic congestion [3]. Furthermore, research by the Spanish Institute for Energy Diversification and Saving traffic congestion causes an 80% rise in tailpipe emissions [4]. According to observations, traffic lights are one of the reasons for the increasing rate of traffic in Baku. According to traffic expert Ferhad Eyyubov, sometimes the red light is set for 70-90 seconds on the roads in one direction [5]. This creates discomfort and a waste of time for drivers. But it would be more effective if traffic light phases were set according to the intensity of traffic on the road. Therefore, there are some errors in the arrangement.

II. Related Work

Most of the ATLS use RL [6, 7, 8] and NN [6, 8] as Machine Learning algorithms, which are similar to the implemented approach. There are some different algorithms as well, which are utilized in other research, including Fuzzy Logic (FL) , Metaheuristic (MH),Genetic Algorithm (GA) , Model

Predictive Con- trol (MPC), and Dynamic Programming (DP) [6, 8, 9]. The results of experimental models are compared with those of other methods such as PressLight, MPLight, NeighborRL, Fixed-time Control (FT), Self-Organizing Traffic Light Control (SOTL), Deep Reinforcement Learning for Traffic Light Control (DRL), and many others [7, 8, 9]. Multiple success evaluation metrics aredefined: based on travel time, road congestion, different environment adaptation, dataset adaptation, and others. After conducting experiments and analyzing results, the best model is chosen for the corresponding dataset and environment.

Other approaches for data collection include both synthetic and real-world data [6, 7, 8]. They collect real-world data from CCTV cameras and sensors on the roads. Number of vehicles, speed, travel time, types of vehicles, flow of traffic, and congestion are retrieved from these technologies. Although it is more realistic to collect real-world data, it is very difficult to analyze the video and extract these parameters. To collect synthetic data, different traffic flow simulators were used, including VISSIM, AIMSUM, MATSIM, CORSIM, and Paramics [10]. In some research, the grid concept is used because different grid structures can affect the balance of safety, efficiency, and complexity in autonomous traffic light control systems [7].

Other research set different values for the parameters. Most of them defined action as the phase change, instead of the phase duration change. For example, if it is phase 0, the action changes it to phase 1 if necessary and vice versa [8]. The reward functions are more complex in their case, as their environments are complex as well. Different techniques include Memory Palace (MP) theory, Phase gate (PG), and different computations of distance between neighboring agents (in the multi-agent approach) [8]. DQN model of some projects does not only have two layers – train and target networks – but also inner layers called fully- connected (FC) layers. Those layers are used in PG theory. The DQN models are trained for approximately 1000 episodes [7, 8].

III. Research Methodology

A. Architecture, Model, Diagram Description

System Architecture:

The main components of the system architecture and the interaction between them are depicted in Figure 1: SUMO simulator - for collecting data and ML algorithms: Q-Learning and DQN. These components always work together throughout the project. Data is generated and retrieved from the SUMO simulator. RL and DQN algorithms are applied to train the data and achieve the results of more experience. The output of training is generated in the same Python code and written to the Excel file. After getting the Excel files, diagram representation tools are used to analyze the data and see the optimized results.

As mentioned above, two machine learning algo- rithms are applied: Q-Learning and DQN. The values of hyperparameters control the learning process:

Epsilon: The probability of agent exploring various action.

Learning rate: the rate at which an algorithm updates its estimates, in another word, the change of neural network weights.

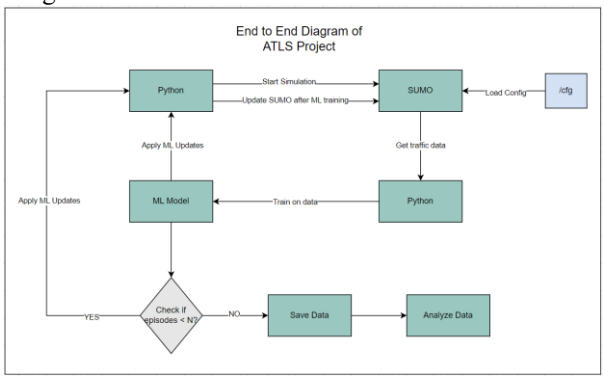


Figure 1: System Architecture Model

Description:

Discount factor: it shows whether the agent pays attention to the rewards in the distant future or the immediate future.

1) Reinforcement Learning - Q-Learning

For the first half of the project, the Q-Learning algo- rithm was trained. The values of predefined parameters for this methodology are listed below:

- ***Agent***: makes decisions, takes actions - traffic light
- ***State***: current situation of the agent - ratio (number of vehicle/capacity).
- ***Action***: move that agent can take - change duration of phases.
- ***Policy***: strategy or rule the agent follows to take actions - decrease max ratio.
- ***Reward function***: rule that gives feedback (reward or penalty – for each action) - increase the Q-value if the max ratio decreases; decrease the Q-value if the max ratio increases.

Hyperparameters whose values control the learning process:

- ***Epsilon***: 0.9
- ***Learning rate:*** 0.9
- ***Discount factor***: 0.1

Here is how the working algorithm:

It calculates the ratio by dividing the number of vehicles by the capacity of the lane, takes the maximum ratio among those, and tries to minimize it.

The q-value is calculated and updated on the q-table in each iteration.

Appropriate action is chosen based on the maximum q-value.

If the new maximum ratio is less than the previous one, the reward is increased.

If the new maximum ratio is more than the previous one, the reward is decreased.Here is the formula of Q-value [11]:

$$Q(S, A) \leftarrow Q(S, A) + \alpha\,(R + \gamma Q(S', A') - Q(S, A))$$

- **S** - current state.
- **A** - action taken by the agent.
- **S'** - next state the agent moves to.
- **A'** - the best next action in state S'.
- **R** - reward received for action A in state S.
- γ (Gamma) - discount factor.

- α (Alpha) - learning rate.

This model worked well for a single-agent approach. After conducting experiments with a single-agent approach, it was decided to continue with a multi-agent approach. Facing the challenge of not being able to adjust to neighbors' actions, agents share some attributes and communicate with each other to work together to find optimal phase durations. Multiple Agent - Reinforcement learning serves exactly for the following purpose. There were two parts of the goal: Goals for Local Agent and Global Network of Agents. Local Agent goals included the reduction of queue size inside each traffic light system and the waiting time of cars. The Global Agent's goal was to reduce the number of cars on the road. The parameters remained the same with the single agent. Shared attributes parameters were added to the system, which are the attributes shared between the neighbors of traffic light systems. They included parameters such as traffic light phases (Red/Green light durations) and traffic density (number of cars in TLS). For the reward function, instead of taking only the maximum ratio of one traffic light, the neighbors of that agent are added as well. In other words, reward function = union max ratio (local agent), max ratio (neighbor agent). Rewards and punishments were applied based on the union of both local and neighbor agents as well. Although for some agents the congestion was decreased, this approach did not optimize all of the agents.

2) Deep Q-Learning (DQN)

To further experiment with different models and algorithms, the DQN algorithm is implemented in the second half of the project, which is an improvement over Q-learning that uses neural networks instead of Q-tables to estimate Q-values [11]. This algorithm is better for complex environments and huge dataset. The different parameters for this model are listed below:

State: *{queue_length, total_vehicle, road_length, phase_dur, phase1, phase0, sim_time}*

Reward function: two options were used throughout the experi- ment.

The first one is simple reward: R = queue_length (number of halting vehicles)

The second one is a complex reward with defined weights:

$R = w_1 \cdot$ total_queue $+ w_2 \cdot$ total_delay $+ w_3 \cdot$ step_waiting $+ w_4 \cdot C$(action) $+ w_5 \cdot$ passed_count $+ w_6 \cdot T$ (passed_travel_time)

Table I: Reward Function Weights

Weight	Value
$w_1 = w_2 = w_3$	−0.25
w_4	−5
$w_5 = w_6$	1

Hyperparameters whose values control the learning process:

- **Epsilon:** 1.0 started with ha igh value and decayed gradually by multiplying 0.9978 (these values are adjusted based on the episode).
- **Learning rate:** 0.001
- **Discount factor:** 0.8

Experience replay: There is a replay buffer that saves transitions (s, a, r, s'). Afterward, these transitions are sampled and used for training.

Action Selection: The agent either chooses a random action (its rate corresponds with the value of epsilon) or takes the best action (according to Q-value) using the Q-network.

Train Network: This network is used to select the best action. Mini-batch of experiences is sampled here. The amount of the sample is decided based on the value specified in the code. For example, if the batch size is 256, this is the size of the training data. The Bellman equation is used to compute the target Q-values [11]:

$$Q(s, a) = r + \gamma \cdot \max_{a'} Q'(s', a')$$

- **Q (s, a)** - The Q-value for state s and action a (train).
- **r** - immediate reward.
- γ (Gamma) - discount factor.
- $\max_{a'} Q'(s', a')$**:** maximum future Q-value.
- **Q' (s', a')** - The Q-value for state s' and action a' (target).

This step updates the neural network weights at the end.

Target Network: This network is generated after the learn process and is used as benchmark for Q-values. This network is updated periodically by copying the weights from the train Q-network. After the agent takes action, it observes the reward r and the next state s'. Then, the train Q-network estimates the Q-value. The target network calculates the target values using the Bellman equation formula. The train network changes weights to reduce the difference between Q-values [11].

Figure 2 depicts the general architecture of DQN, which includes input and output parameters of the layer.

To prevent DQN from confusing the natural reduction of traffic intensity due to the time of the day, all agent functions are being called, and parameters are being adjusted during peak-hour periods:

- 3500 < sim_time < 4500
- 6500 < sim_time < 7000
- 9000 < sim_time < 10000

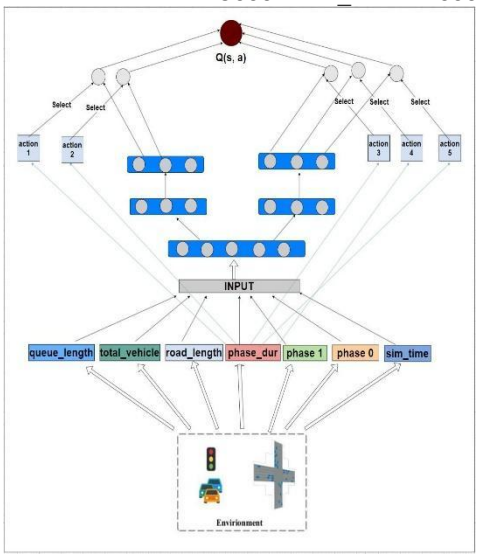


Figure 2: General architecture of DQN.

B. Experimental Results

1) Reinforcement Learning with Q-table

Following road intersection was chosen to conduct experiment for assessing RL's effectiveness:

Scenario 1: Uniform Traffic Distribution, Figure 3

Figure 3: Road intersection in SUMO

Description: Traffic flows across all lanes.

Without RL:

- Equal green phase durations across all lanes.
- Moderate congestion as idle green phases are wasted on low-

traffic lanes.

With RL: The RL agent detects uniformity in traffic and optimizes signal timings to minimize idle phases.

- By balancing green phase durations across lanes, throughput is maximized.

Outcome: RL reduces the average waiting time by 20% and ensures smoother flow, depicted in Figure 4:

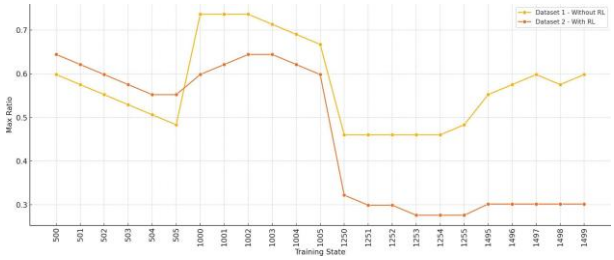


Figure 4: All roads open RL performance compari- son

2) Reinforcement Learning with DQN

For testing the first theory, it was decided to check the code with a stable flow. This could give a clear picture of whether agents can optimize the traffic flow when there is no fluctuation due to the time of day. Results of the simulation was as follows, shown in Figure 5:

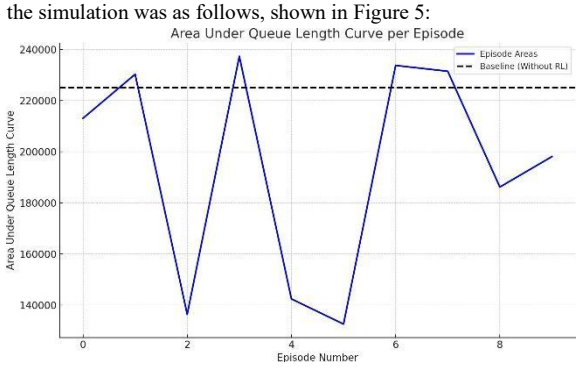


Figure 5: Simple reward DQN with Static flow

As can be seen from the result, in 8 out of 10 episodes, DQN agents performed better than statically set TLS phases, achieving 5% - 37% reductions in queue length.

IV. Conclusion

Several experiments were conducted with both Q-learning and DQN algorithms. On some roads, the congestion is reduced, but on others, it is not optimized enough. In the intersection of R. Behbudov and Bakhikhanov streets, in the case of uniform traffic distribution, RL reduced the average waiting time by 20%. In the DQN experiments, static flow was easily optimized since the vehicle flow was not randomly generated, and the reward did not fluctuate in peak times during the simulation. However, when dynamic flow was introduced, the variability in traffic patterns led to fluctuations in the reward function and introduced uncertainty in the learning process. As a result, although optimized results were achieved in some episodes, this was not consistent. Therefore, it was challenging to fully optimize the roads when there was dynamic flow.

A. Future work

As mentioned previously, the main limitation and challenge throughout the project was the lack of resources. This issue not only blocked the efficient acquisition of test results, but also prevented the experiment from being conducted in a multitude of possible scenarios. Therefore, in the future, having more computational power and the ability to process experiments on multiple machines would speed up the R&D and would give the researcher freedom to test every possible scenario, enriching knowledge about agent behavior. However, for the best effectiveness, ATLS should be tested in real life. It is feasible to implement ATLS in regions of Azerbaijan where traffic is not as heavy as in Baku to observe how it will prove
itself in practice.

Acknowledgment

We thank the Intelligent Transportation Management Center (NIIM) for their support and help during the project.

References

1. Intelligent Transportation Management Center (ITMC), “Internal communications and information provision,” 2025, data gained by ITMC during the meeting with them.
2. A. Samadov, “Azerbaycanın neqliyyatı,” https://www.stat.gov.az/source/transport/, azerbaycan Respublikasının Dövlet Statis- tika Komitesi. Accessed: Mar. 30, 2025.
3. M. Shaygan, C. Meese, W. Li, X. G. Zhao, and M. Nejad, “Traffic prediction using artificial intelligence: Review of recent advances and emerging opportunities,” *Transportation Research Part C: Emerging Technologies*, vol. 145, p. 103921, 2022.
4. S. Xu, C. Sun, and N. Liu, “Road congestion and air pollution analysis of spatial and temporal congestion effects,” *Science of The Total Environment*, vol. 945, p. 173896, 2024.
5. F. Eyyubov, “Tıxaclı Bakı: Çıxıs¸ yolu ndir?” https://www.yeniazerbaycan.com/Sosial_e72238_az.html, 2022.
6. H. Wei, N. Xu, H. Zhang, G. Zheng, H. Zang, and Z. Li, “Colight: Learning network-level cooperation for traffic signal control,” in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM)*, 2019.
7. C. Chen, H. Wei, H. Wu, N. Xu, Z. Zhang, J. Zhu, and B. Zhang, “Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, Apr. 2020, pp. 3414–3421.
8. H. Wei, G. Zheng, H. Yao, and Z. Li, “Intellilight: A reinforcement learning approach for intelligent traffic light control,” in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018.
9. K.-H. N. Bui, O.-J. Lee, J. J. Jung, and D. Camacho, “Dynamic traffic light control system based on process synchronization among connected vehicles,” in *Advances in Intelligent Systems and Computing*. Springer, May 2016, pp. 77–85. [Online]. Available: https://doi.org/10.1007/978-3-319-40114-0_9
10. M. R. Ullah, K. S. Khattak, Z. H. Khan, M. A. Khan, N. Minallah, and A. N. Khan, “Vehicular traffic simulation software: A systematic comparative analysis,” *Pakistan Journal of Engineering and Technology*, vol. 4, no. 1, pp. 66–78, Mar. 2021.
11. H. Sutton, “Peter morgan sutton,” *BMJ*, vol. 348, no. mar31 11, pp. g2466–g2466, Mar. 2014. [Online]. Available: https://doi.org/10.1136/bmj.g2466

Development of a Mobile Price Tag Recognition System for Visually Impaired Customers

Sevda Aliyeva, Ibrahim Aliyev,
Fidan Yusifova, Jamaladdin Hasanov
School of Information Technology and Engineering
ADA University, Baku, Azerbaijan
{saliyeva14249, ialiyev14159, fyusifova14091, jhasanov}@ada.edu.az

Abstract—Shopping for goods in markets sounds like an easy task, but people with disabilities, including visually impaired individuals, struggle to complete it. To address this problem, a mobile application for reading price information is developed, and while introducing it in this paper, traditional and machine learning models for OCR and TTS are compared. Classical computer vision algorithms are integrated with pre-trained machine learning models by the system to accurately detect and recognize price tags in various real-world conditions. Following image recognition, a dynamic text generation module extracts and structures relevant information, which is then converted into speech for the user. This application provides a user-friendly and relatively cheaper solution while giving a chance for future improvements.

Index Terms—Price tag recognition, visually impaired users, mobile application, computer vision, pre-trained models, image processing, text-to-speech, voice synthesis, accessibility, assistive technology, traditional models

INTRODUCTION

In recent years, technological advancements have made remarkable progress in improving quality of life and addressing previously unsolvable problems. As digital devices have become more integrated into our lives, people rely on smart devices more often for personal and professional needs. The combination of artificial intelligence with smartphone sensors and cameras has allowed us to have and consider new opportunities in the field of assistive technology [1]. These tools are remarkably promising for addressing accessibility problems, yet far fewer solutions exist for people who are visually impaired.

The World Health Organization indicates that more than 2.2 billion individuals worldwide have some kind of vision impairment [2]. This group is neither small nor isolated members of society, and visually impaired individuals struggle to complete simple daily activities; for example, students studying in higher education institutions mention their struggle with access to information and reference materials [3] [4], which is their basic right. Insufficient support in reading, writing or communication keeps people from doing everyday tasks independently. Although many assistive technologies such as screen readers, OCR tools and wearable devices, do exist, most do not work as expected when applied in real-world environments, like supermarkets where light, the style of labels and the layout can significantly affect the performance.

The psychological results of such situations can be devastating for the mental well-being of these people. The need for continuous help causes people to question their freedom and privacy, which deepens their feelings of needing someone else's help. This condition causes people to experience anxiety [5] and triggers their hesitation to shop by themselves since it destroys their confidence in public environments. Dependence on others for everyday tasks creates negative impacts on selfesteem and social participation [6] because these elements form essential parts of mental well-being.

This paper discusses a mobile app that helps tackle one of these problems: reading and understanding price tags in market environments. The main motivation behind this product is the fact that traditional price tags rely heavily on the visual information, which makes comfortable shopping inaccessible for the visually impaired people. Thus, our system enables visually impaired users with an accessible and efficient service that recognizes the price tag information directly and converts it to speech out loud. The system helps users become more independent and feel more included while shopping.

While providing practical functionality is significant, another purpose needs to be mentioned is educating people about the everyday difficulties experienced by people with visual impairments. The project targets to motivate developers, businesses, and society as a whole to consider accessibility as an essential element in innovation by promoting inclusive technology within public and commercial domains. The system functions to support consumer autonomy and equality combined with dignity for visually impaired users in the everyday consumer context.

RELATED WORK

A. Assistive Tools and Mechanisms for the Visually Impaired

With recent progress in technology, assistive devices help visually impaired people participate in regular activities more easily. [7]. The analysis of this interaction and tools that are used are especially important for visually impaired people since there is an increase in the assistive technologies that are integrated to the mobile devices. Diverse solutions and mechanisms have been developed to target this group, and they use different approaches to make life easier. According to a survey on assistive technologies for visually impaired, although some of the solutions employ sensor-based, image processing and AI-based, deep learning-based, and app-based approaches, all of them share a common point, which is having a trade off between the features and resources that are used. Currently, some of the app-based solutions are Seeing AI by Microsoft, LookTel by IPPLEX, TetraMail, and all of them serve visually impaired people in different areas of life. The survey also discusses the need for the cost-effective, lightweight, and high-accuracy solutions with advanced features [8]. Another overview paper of assistive technologies

suggests various categories in daily life, such as personal care, money, timekeeping, alarms and alerting, finance, and shopping - last one being the direction of this paper. The main conclusion is based on the assistive technologies being over complex and unreachable. Moreover, the primary challenge has been indicated the integration of different assistive solutions in a single assistive system [9].

B. Advances and Challenges in Optical Character Recognition and Price Tag Recognition

The emergence of the sophisticated software and hardware designs caused a transform in OCR technology beginning from 1980s and 90s. The integration of machine learning into the OCR systems allowed the models to learn from the given data, and it improved the accuracy of general OCR technology by being a foundation for more advanced OCR systems. Later, in the 2000s, the introduction of deep learning, Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) took OCR to the next level and enabled text recognition even in noisy, hard-to-identify, and complex environments [10]. Years of research has been conducted to improve the accuracy of Optical Character Recognition (OCR) models while also gaining more abilities in recognizing the text in very diverse and natural environments, such as supermarkets, hospitals, universities, and transportation. The real world applications of text recognition in image search, industrial automation, and instant translation [11] have increased the research interest even more although challenges to address still exist, especially in the field of price tag recognition and verification. A neural network-based price tag verification solution offers a unique system to address the irrelevant and incorrect information issue on price tags that decreases customer loyalty. The mobile app compares the information on the price tag with the database to automatically check whether the price tag reflects accurate information [12]. Another similar study offers an even better deep learning-based solution that utilizes a dataset with 50000 images that is adjusted according to diverse features and difficulty levels, such as lighting and different camera angles. The system detects the price tag area using the EfficientNet-Lite-0 models, which is a more efficient version of the original EfficientNet. Compared to the various methods used in Google Tensorflow Object Detection API, the highest efficiency, speed, and accuracy results were obtained with the Efficientnet-Lite-0 model. The average accuracy rate that the tests confirmed is more than 98% [13]. Considering the accuracy, which is highly significant to avoid mistakes and incorrect information, a proposed approach uses Niblack binarization algorithm for the task of search and recognition of price tag zone by targeting minimal energy usage. The algorithm has been tested on 708 images dataset, which demonstrated a remarkable accuracy result with 95.8% [14]. In addition to this, a price tag recognition system uses the HSV color model to distinguish text more easily, isolating the Region of Interest (ROI), and eventually improving the OCR performance [15].

THE DESIGN OF THE SYSTEM

There are two goals in this project, one is to introduce models for the mobile application and the other is to compare those models. Both of the strategies introduced follow the same structure; first, OCR, then formatting the output text, and voicing it. Therefore, the only thing left to compare is the OCR techniques used.

Machine Learning Models: Llama Vision 3.2

Although two viable solutions were proposed, the pipeline using the Llama Vision 3.2 model was implemented as it showed high accuracy in real-world scenarios, had a simpler pipeline, and was more scalable compared to the pipeline using Paddle OCR. In this way, steps like barcode detection, cropping, grayscale conversion and noise filtering are eliminated. Misread barcodes or low-quality images do not affect the result and do not need to be manually re-read, thus creating a better user experience. Reducing these types of stages affecting the process creates fewer points of failure and faster response times. This pipeline uses Llama 3.2 Vision for information extraction, FastAPI for text formatting and backend routing, and an open-source Text-to-Speech engine for audio feedback.This solution provides scale, consistency, and high-level contextual understanding of information, such as price tags in complex visual environments. Llama Vision 3.2 is able to semantically know what constitutes a “price”—whether or not it’s not called out in some straightforward manner. The system follows the following steps. 1. The user takes a picture using the mobile app. 2. The image is stored on the server. 3. Llama Vision processes the image to obtain relevant pricing data. 4. This data is received by FastAPI, the output is formatted, and it’s passed to the TTS engine. 5. eSpeak NG produces an audio file, and the file is returned to the app and played back to the user.

The Android mobile application was developed in (Java/Kotlin) using Android Studio. The interface has been made as user-friendly as possible to ensure accessibility for users. Users will use a single, high-contrast, large action button to take pictures.

Instead of using traditional OCR (such as PaddleOCR), using Llama 3.2 Vision, a cutting-edge vision-language model created by Meta AI that enables image-based context reasoning. It recognizes image relationships (such as product → barcode → price) without the need for specific spatial directions. Unlike classical OCR requiring preprocessing (i.e., segmentation), the Llama Vision may analyze real-world images directly regardless of the lighting and background conditions. Also, it could extract and pick out semantically relevant data like prices from noisy and blended images. This alleviates the preprocessing step, so better results are achieved.

FastAPI (a modern, asynchronous web framework) handles the string extracted by the Llama model. After normalizing the text, such as standardizing currency labels, the string is passed to eSpeak NG for processing in the TTS model.The API also handles the response, including converting the speech to .mp3 and returning it via HTTP. This modular backend enables each of the functions to run in isolation so it is possible to swap out the components (such as eSpeak with Google TTS) without having to rewrite the fundamental logic.

For the last step, as mentioned above, the Speak engine was selected for the extracted and formatted price text into speech. Its support for many languages, including Azerbaijani, was one of the main factors in choosing this model. In addition, the fact that it is open source and lightweight, and can be used for academic purposes without license restrictions, made this model suitable for us. One of its advantages is that it converts text to audio with minimal delay compared to other TTS models has been tested. Finally, having cross-platform

compatibility makes this model appropriate for any future scaling.

The model's execution begins with a system call in the FastAPI backend. The input text is converted into a .mp3 file, and then streamed back to the mobile device to automatically play it to the user.

Traditional Models: PaddleOCR

Next model proposed in this paper represents a classic computer vision pipeline that combines barcode detection, image preprocessing, traditional OCR (PaddleOCR), and textto-speech. It is implemented and tested as an alternative solution for the product. It was noted that Paddle OCR lost this competition against the use of Llama model due to its dependence on image quality and multi-step pipeline. In addition, the lack of context-awareness during detection was the factor that caused it to produce noisy and precise output.Despite these, the advantages of using Paddle OCR over traditional OCR are also worth noting.The most important of them is that it requires less computational power.It is quite accessible as it can run even on mid-range CPUs or mobile devices. Because it does not depend on a large visual language model or web-based APIs, it has the potential for full offline deployment (e.g. on embedded systems) and no GPU or cloud computing is required, significantly reducing operational costs.The fact that Paddle OCR is more affordable than the Llama model makes it a better candidate for budget-constrained environments or lightweight deployments. Additionally, from the perspective of transparency of the operating principles of artificial intelligence models, which has been relevant recently, the Paddle OCR model can also be preferred. Because each step (barcode detection, clipping, OCR) is explainable, observable, and adjustable — suitable for debugging and monitoring.

The initial stage of this pipeline also begins with the user taking a photo. This image is uploaded to the server either via Google Drive (test) or directly via HTTP. In the next step, the image is gray-scaled by OpenCV to reduce noise and make it easier to input for text and barcode detection. After that comes the barcode detection stage with the pyzbar Python library. The reason for finding the barcode is the assumption that the price is close to it. It extracts the bounding box coordinates (x, y, width, height) of the barcodes detected in the image, thereby determining the location of the price tag. Based on the measured coordinates, the system cuts the image to fit the Region of Interest. The problem with this step is that if the barcode is off-center or not clear in any way, or if it is not placed regularly on the barcode, this step can cause incorrect or incomplete regions to be cut. As a result, the OCR model later fails to produce the desired results. Good performance of Paddle OCR is highly dependent on this stage. Assuming that the cropping was done correctly, Paddle OCR has no trouble extracting the information. However, the model is not smart enough to distinguish price from detected content. The text needs to be formatted accordingly. So, regex filters or numeric checks are applied to extract the most likely value from all detected text elements (e.g., look for numbers like "money" or with "AZN", or manat sign). If there are multiple numbers, priority rules are applied based on location or value range (e.g., ignore values below 0.10 or above 100). The TTS model used to read the detected price is the same as Option A - eSpeak NG. The audio is streamed or downloaded to the mobile app and is automatically played.

Data Collection

The dataset used in this project was collected from three different local markets to experiment with various OCR models. Therefore, the dataset has three types of images which mainly varies because of the price tags; different markets use different barcode styles and placements for product information on price tags. Some of the examples can be seen in Fig. 1.

Market A Market B Market C

Fig. 1: Comparison of Market A, Market B, and Market C

Experimental Results

One of the proposed solutions is PaddleO, which has been extensively tested for extracting prices from price tags. The price tag image dataset used for this purpose consisted of pictures taken in a real-life market environment including variations in angle, lighting, background clutter, and label design. Initially, PaddleOCR was applied directly to these raw images. In the initial tests, the model had difficulty detecting the price on many price tags due to multiple product labels, shelf lines, and background distractions. The model failed to identify correct location containing price which caused either system to detect no price or produce incorrect results. The observations showed that a proper preprocessing step must be added because it serves more than one purpose: it enhances OCR accuracy while guaranteeing PaddleOCR maintains its attention on a specific targeted area. To solve this, a preprocessing pipeline was implemented, consisting of barcode detection defining the area of interest and expanding the bounding box to reach the actual price on the price tag, followed by grayscale conversion and binary thresholding on the image. These steps boosted PaddleOcr's core recognition capability. Despite increased localization in inputs and improved performance, it was unable to correctly classify numeric values of the price, specifically when the digit appeared in different front shifts (refer to image). Inspection revealed the most frequent price tag layout as the cause for incorrect detections because the top placement of some digits confuses the model at times. Therefore, the digit recognition of PaddleOCR resulted in either missing or misidentifying numbers that caused the extraction of process like "1.80" as ".80" .This formatting inconsistency, inherent in the physical design of many supermarket price tags, created a persistent obstacle.The fact that some price tags have non-standard or differently sized barcodes also created a problem during the testing phase. Differences between barcodes affected the region of interest in preprocessing, resulting in ambiguous image cropping results. Image cropping resulted in two potential problems: the price area was sometimes cut off, or distracting elements from nearby labels or shelf material entered the frame. Such cases accounted for the majority of failed detections, even after initial processing.

To validate the model's performance, we collected 25 price tag images from each of 4 different supermarkets to ensure they were representative and comprehensive, and formed a total dataset of 100 images in various conditions.PaddleOCR achieved about 65 percent accuracy across all samples. In the remaining 35 percent, the model returned either partial values or no values at all—usually due to misaligned digits and inconsistency between barcode types.

The proposed solution, option A, uses the Llama 3.2 Vision model of Meta, and the results are highly promising. Firstly, we used Roboflow environment, which offers a graphical interface and built-in version of the model shown in Fig. ?? to avoid setting up the model on local and testing it on the first stage. As for all other models, Llama 3.2 Vision asked for the api key, and we obtained it from OpenRouter API. Among various task types, we utilized text recognition (OCR) and open prompt, which allows us to give a prompt to the model, such as "extract only the price" or "extract the name of the product" for the purpose of text recognition and getting the required part of the text on the image. Although the dataset contains 100 images, which are highly noisy and complex, the model worked successfully and outperformed all the other models tested by having near 100% accuracy. In addition, the LLaMA 3.2 Vision model clearly demonstrated how large language models can be extended to vision-based tasks with label-based instructions and minimal pre-processing to achieve the desired output. An example scenario can be considered to better understand the testing process with the LLama 3.2 Vision model. After Fig. ?? was given as an input to the model, we chose open prompt as the task type and gave it a prompt like this: "Give me only product price in price tag, in X.XX format". Later, the model processed the text on the image and gave us the final price information shown in Fig. ??. The successful results we obtained in Roboflow allowed us to continue with this model, and we decided to install and test it on our local server. The local deployment gave us more control over the environment and allowed us to simulate the real-world scenarios. In local testing, after the images are taken, they are sent to the AWS ECS2 instance server and are uploaded to the "Images" folder there. Later, Llama 3.2 Vision model starts to process the data based on the given prompt and extracts the price information. Although the testing results are still highly accurate on the local server, the response time is relatively high since the model is loaded for each request.

CONCLUSION

In this paper, all the research and work done to create this mobile application for visually impaired people to read the price tag have been discussed. Firstly, different approaches were explored and the drawbacks of existing solutions were considered, and as an essential part of the project, the team investigated the major problems faced by people with visual impairment while shopping at the markets. After gathering all the necessary knowledge, two main models for the mobile application were proposed. Both utilize the same strategy for voicing the acquired information, yet they use different models for OCR; PaddleOCR and Llama Vision 3.2. The project concludes that traditional models, such as PaddleOCR, are a cheaper and faster solution, yet machine learning models, such as Llama, are more accurate and able to provide more detailed information without requiring additional preprocessing.

A. Future Work

For this stage of the project, the main goal was to sound the price information of the products on the markets for visually impaired people. Currently, the application gives exactly the desired result. However, there is always room for growth. Firstly, due to high computation power, the current response time differs between 6-10 seconds. With better server and GPU, this can be reduced. After these improvements, increasing the capabilities of the project is the primary objective. In the next stage, the application should be able to voice not only the price but also the name and quantity of the products. These data can also be read from the price tags. A more complex task for this project would be reading expiration dates, since they are only mentioned on the packaging of the products. One more important issue that individuals with visual impairments face is not knowing which product is located where. In general, to make finding goods faster, the markets divide them into sections and have boards near the sections for customers. Obviously, this is a struggle for people with lower vision, and by adding a necessary functionality, the proposed application would be able to address this issue in the future. While increasing the number of features, keeping the application as simple as possible is important for users to easily utilize it. To ensure this, voice-based command control can be enabled in later stages. One more main point that needs to be considered is that, to make sure everyone can use this application, it should also be adapted to iOS phones. Currently, the version of the application is available only on Android phones.

Acknowledgments

The authors are grateful to Tural Gasimov, who contributed to this project as a mobile developer.

Disclosure statement:

The authors do not report potential conflicts of interest.

REFERENCES

[1] P. Naayini, P. K. Myakala, C. Bura, A. K. Jonnalagadda, and S. Kamatala, "Ai-powered assistive technologies for visual impairment," 2025. [Online]. Available: https://arxiv.org/abs/2503.15494

[2] World Health Organization, "World report on vision," 2019. [Online]. Available: https://www.who.int/publications/i/item/worldreport-on-vision

[3] A. Amin, N. Sar, N. Md Akhir, S. Marziah, and R. Badri, "Main challenges of students with visual impairment at higher education institutions," *International Journal of Academic Research in Progressive Education and Development*, vol. 10, 03 2021.

[4] E. Croft, "Experiences of visually impaired and blind students in uk higher education: An exploration of access and participation," *Scandinavian Journal of Disability Research*, vol. 22, no. 1, pp. 382–392, 2020.

[5] H. Winterflood and E. A. Climie, "Learned helplessness," *The Wiley Encyclopedia of Personality and Individual Differences: Personality Processes and Individual Differences*, pp. 269–274, 2020.

[6] T. Govier, "Self-trust, autonomy, and self-esteem," *Hypatia*, vol. 8, no. 1, pp. 99–120, 1993. [Online]. Available: http://www.jstor.org/stable/3810303

[7] J. M. Carroll, "Human–computer interaction: Psychology as a science of design," *International journal of human-computer studies*, vol. 46, no. 4, pp. 501–522, 1997.

[8] K. Manjari, M. Verma, and G. Singal, "A survey on assistive technology for visually impaired," *Internet of Things*, vol. 11, p. 100188, 2020.

[9] S. T. Brassai, L. Bako, and L. Losonczi, "Assistive technologies for visually impaired people." *Acta Universitatis Sapientiae-Electrical & Mechanical Engineering*, vol. 3, 2011.

[10] O. Timilehin, “Enhancing industrial efficiency: The evolution and applications of robust optical character recognition systems,” 2024.

[11] S. Long, X. He, and C. Yao, “Scene text detection and recognition: The deep learning era,” *International Journal of Computer Vision*, vol. 129, no. 1, pp. 161–184, 2021.

[12] P. Laptev, S. Litovkin, S. Davydenko, A. Konev, E. Kostyuchenko, and A. Shelupanov, “Neural network-based price tag data analysis,” *Future Internet*, vol. 14, no. 3, p. 88, 2022.

[13] M. Turan, M. Peker, H. Ozkan, C. Balaban, N. Kocakır, and¨ O. Ka-¨ rademir, “Development of a price tag detection system on mobile devices using deep learning,” *Orclever Proceedings of Research and Development*, vol. 1, no. 1, pp. 178–187, 2022.

[14] M. A. Aliev, D. A. Nikolaev, I. Kunina, and D. Bocharov, “A low computational approach for price tag recognition,” in *Twelfth International Conference on Machine Vision (ICMV 2019)*, vol. 11433. SPIE, 2020, pp. 301–309.

[15] M. Hussin, A. Ahmad, and M. A. Razak, “Price tag recognition using hsv color space,” *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, vol. 9, no. 3-9, pp. 77–84, 2017.

A Satellite-Based Wildfire Dataset for Azerbaijan

Gunel Aliyeva, Turgay Farhadli, Ahmad Huseynli, Sahil Guluzade
School of Information Technology and Engineering ADA University, Baku, Azerbaijan
{galiyeva15577, tferhadli14021, ahuseynli7563, sguluzada14101}@ada.edu.az

***Abstract*—Wildfires are a growing global concern. However, the availability of high-quality, processed satellite imagery for developing detection systems remains limited, especially in under- represented areas. This study presents a novel dataset developed by integrating NASA FIRMS fire alerts with Sentinel-2 satellite imagery, covering the Azerbaijan region. The dataset is openly accessible through a dedicated website that offers tools for visual exploration, filtering, and downloading, addressing the challenge of limited wildfire imagery data. A comprehensive data extraction and processing pipeline was implemented, including cloud and noise filtering, spatial and temporal alignment with fire alerts, contrast enhancement, and manual verification. The overall approach is designed to be scalable and easily extend- able to other regions. Special emphasis was placed on smoke plume visibility and geographic representation. The dataset contributes to the ongoing efforts in wildfire detection, fire risk mitigation, and data-driven environmental policy, in response to the increasing frequency and impact of human-caused fires influenced by changing climate conditions. Images are accessible through the project website (https://wildfireaze.github.io), and the complete source code is available in the GitHub repos- itory (https://github.com/gunelaliyevaa/wildfire-detection-using-satellite-imagery).**

***Keywords*—Wildfire dataset, Satellite imagery, Remote sens- ing, Sentinel-2**

I. Introduction

Wildfires damage millions of hectares around the world each year, causing extensive destruction of forests and wildlife, along with severe social, economic and environmental con- sequences [1]. Detecting fires at an early stage is important to prevent catastrophic outcomes, protect lives, and preserve property [2]. As wildfires commonly occur in remote loca- tions, their early detection through on-site monitoring becomes difficult [3]. Conventional wildfire detection methods, such as manual patrolling, fire lookout towers, and ground-based sensors, are limited by human error, slow response times, and poor coverage. They depend on site visits for data collection, which is insufficient for proactive wildfire management given how quickly these events unfold and change [4]. Aerial solu- tions offer improvements but are not widely adopted due to operational and cost constraints. On the other hand, satellite- based remote sensing offers a more suitable way of detection, as the necessary infrastructure is already in place. Despite advancements on a global scale, wildfire detection is still a valid problem in areas with limited access to data, including Azerbaijan. To address this, our study presents a publicly avail- able wildfire dataset covering Azerbaijan, constructed from Fire Information for Resource Management System (FIRMS) fire alert data offered by the National Aeronautics and Space

Administration (NASA) and Sentinel-2 satellite imagery. Our dataset focuses exclusively on fire events within Azerbaijan's national boundaries and includes rigorous preprocessing, la- beling, and enhancement pipeline to make it directly usable by researchers, particularly for developing and evaluating Machine Learning (ML) models and supporting studies on the application of remote sensing in the region. Additionally, to address a larger community, we developed an interactive web platform that visualizes our wildfire imagery data through customizable filters. The website allows users to query fire occurrences by date range, geographic region, sensor type, satellite, and view the location of the fire through an interactive map.

The rest of this paper is organized as follows. Section 2 reviews prior work on wildfire detection, focusing on the role of satellite imagery in wildfire science, datasets, and research gaps. Section 3 describes the methodology, covering data acquisition, preprocessing, and labeling. Finally, Sections
4 and 5 summarize the study and outline future research directions, respectively.

II. Related Work

Satellite-based observation has advanced wildfire monitor- ing by enabling scalable, timely, and reliable coverage of remote areas. It became even more effective with the emer- gence of advanced remote sensing technologies and ML techniques. To picture the Earth from space, satellites are equipped with sensors that capture electromagnetic radiation across various spectral bands, including visible light, infrared, and thermal wavelengths [5]. These are then processed and converted into image products for analysis. Moderate Resolution Imaging Spectroradiometer (MODIS) and Visible Infrared Imaging Radiometer Suite (VIIRS) are two of the widely available satellite-based sensors used for fire detection from satellite imagery [6]. These sensors pro- vide a broad, synoptic view in large areas, providing daily observations essential for monitoring active fires and burned areas.

Wildfire detection systems rely on a diverse range of datasets that are important for the timely detection, mapping, and prediction of wildfire events. Although various countries such as Portugal and Italy have conducted research in this field, their datasets mainly consist of wildfire images cap- tured by regular cameras, rather than satellite imagery. These datasets (e.g., a dataset from Barcelona, Spain, "The Wildfire Dataset Enhancing Deep Learning-Based Forest Fire Detec- tion with the Wildfire Dataset" [7]) may provide solutions to detect smoke/fire via ML techniques, but these are not sufficient sources for us due to the absence of satellite imagery. In contrast, satellite datasets provide significant advantages. MODIS and VIIRS offer daily observations over broad areas, supporting the monitoring of active fires and burned regions. For higher resolution needs, Sentinel-2 has become essential for analyzing fire damage due to its multispectral imaging and frequent revisit capabilities [8]. Beyond imagery alone, an integrated multi-modal dataset approach has emerged, fusing remote sensing data with

environmental and meteorological variables such as temperature, humidity, wind speed, vegeta- tion indices, and topography. An example of such advancement is the "Next Day Wildfire Spread" dataset, which spans nearly a decade and includes various explanatory variables to forecast fire behavior and support emergency management efforts [9]. These evolving datasets reflect the increasing sophistication and utility of wildfire-related data, especially when enhanced by satellite-derived information.

Despite recent advancements, many model-driven ap- proaches used today for wildfire detection are still primarily reaction-oriented, and there exists a significant gap in proactive measures. This gap is particularly relevant in underrepresented regions, where wildfire research and data infrastructure are still developing. Even after considering our research, there is still a noticeable gap in predictive modeling that takes meteorological data into account. Key environmental factors such as temperature, wind speed and direction, and humidity play a crucial role in assessing both the severity and spread of ongoing wildfires and the likelihood of new fire ignitions. The absence of early prediction models and warning systems causes delays in response efforts, making it harder for fire-fighting organizations to take preventive action.

III. Methodology

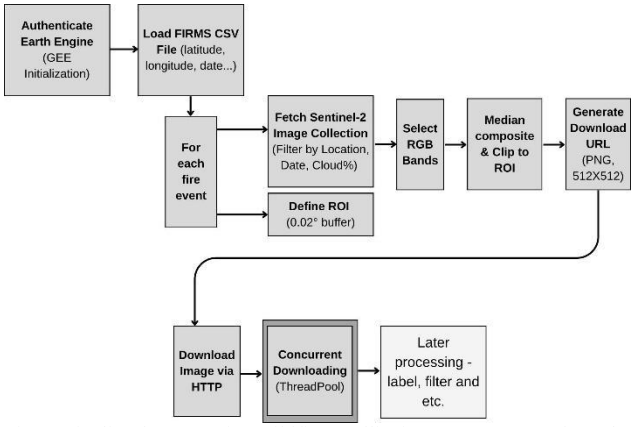


Fig. 1. Pipeline for extracting wildfire satellite imagery. It covers from the first phase of the process, data fetching until downloading.

To solve the problem of having limited availability of pro- cessed satellite datasets for wildfire detection and to increase the geographical diversity of existing data, we developed a

pipeline for generating wildfire imagery from satellite sources. The dataset is intended for use in ML and Deep Learning (DL) applications of remote sensing and is therefore processed and labeled accordingly. A labeled subset containing confirmed wildfire instances is publicly shared through an interactive website to support further exploration and development.

A. Dataset Description

Dataset consists of 178 true wildfire images collected across the regions of Azerbaijan using multiple satellite instruments. Each image is the product of merging NASA FIRMS fire alerts data and Sentinel-2 imagery. The images contain various metadata such as a description, geographic coordinates, city, district, brightness, scan and track values, acquisition date and time, satellite and instrument details, confidence score, version information, brightness temperature (T31), fire radiative power (FRP), and a day/night indicator.

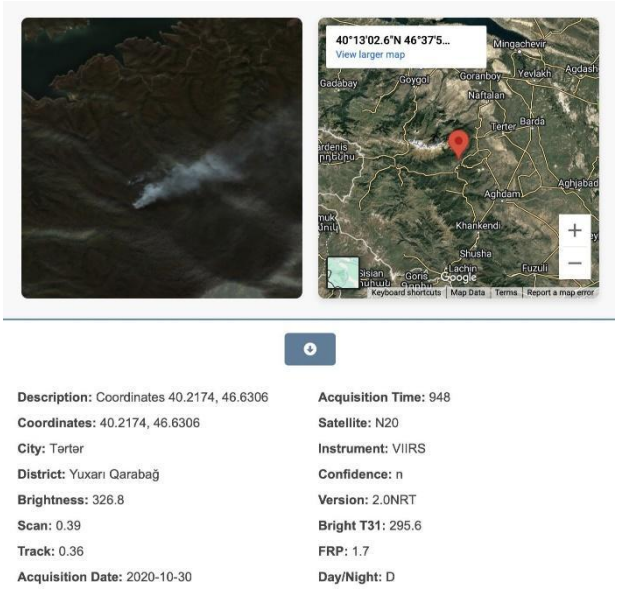


Fig. 2. Sample satellite image with wildfire detection metadata.

B. Data Choices and Preprocessing

Sentinel-2 was selected, with bands B2, B3, and B4 (RGB) chosen, due to its high-resolution imagery and moderate revisit time. The surface reflectance product was used, because it provides visually interpretable, atmospherically corrected data, ensuring accurate surface reflectance values that are not influenced by atmospheric scattering, distortion, or absorption. Sentinel-2 data was accessed freely through Google Earth Engine (GEE), which allowed prototyping to be carried out without the need for large storage capacity or a fast internet connection, making it ideal for the initial stages of the project. Sentinel-2 also offers 13 spectral bands, which will be useful when the project is expanded further to use multispectral data. For the initial prototype, RGB data was chosen exclusively due to several practical constraints, including GPU memory limitations, large data volumes, and ease of presentation. Addi- tionally, since the dataset initially contained over 3000 images that required rigorous filtering, using RGB-only data sped-up this process. Overall, it balances detail and coverage making it a suitable source for our purposes and keeping the system lightweight for initial stages. The image extraction pipeline starts with fetching the Sentinel-2 Level-2A (COPERNI- CUS/S2_SR_HARMONIZED) image collection product [10], provided publicly by European Space Agency, using GEE's Python API. Data collection was limited to the national boundaries of Azerbaijan. The date range of images covers the years 2017-2024 as the collection was available only for these years. The images captured from the collection were downloaded in PNG format and used in preprocessing. Before fetching the view corresponding to the fire alert, the Sentinel-2 collection is filtered based on the percentage of cloud cover, using the CLOUDY_PIXEL_PERCENTAGE

metadata value derived from the Scene Classification Layer (SCL) generated by the Sen2Cor processor, where images with more than 10% of cloud cover were not fetched. Despite applying this filter, some noise in the images remained due to thin clouds or shadows, which were later manually screened during labeling.

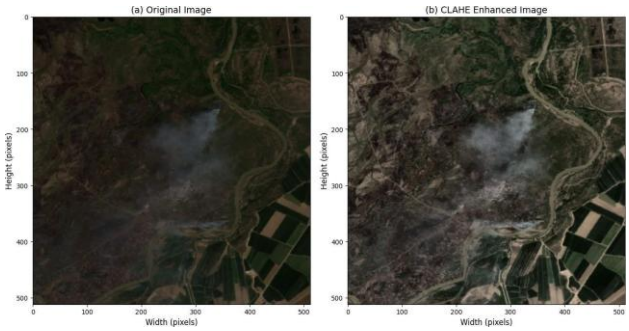


Fig. 3. Preprocessing of a wildfire image captured from Sentinel-2 L2A collection.

The raw Sentinel-2 images were already visually inter- pretable to some extent. However, smaller smoke indicators were difficult to detect due to the overall darkness and the gray color of the smoke blending with the dark terrain background. Therefore, Contrast Limited Adaptive Histogram Equalization (CLAHE), a contrast enhancement method, was applied, improving lighting and visibility, as shown in Fig.
3. Default parameters were retained, as the imagery was primarily captured during daylight and under typical wildfire season conditions.
We used NASA FIRMS, a Near Real-Time (NRT) fire detection system, as the main source to pinpoint fire locations. NASA FIRMS uses the MODIS aboard the Aqua and Terra satellites, and VIIRS aboard S-NPP, NOAA 20, and NOAA 21 [11]. The FIRMS dataset has 15 attributes, 7 of which were longitude, latitude, acquisition date, acquisition time, satellite, instrument, and confidence that were used in the dataset creation process. It is also important to note that these alerts are not preprocessed, so they contain false alarms,

repetitive alerts, etc. The acquisition date and time were used as the primary indicator of fire events. MODIS has revisit time of approximately 1-2 days globally and VIIRS provides daily coverage per satellite, around 12-hour revisit when combining all satellites [12], [13]. Sentinel-2, on the other hand, offers a revisit time of 10 days per satellite, which reduces to approximately 5 days with both satellites, Sentinel- 2A and Sentinel-2B, combined [14]. The lower temporal resolution compared to MODIS and VIIRS may result in missed detections of short-duration fire events. Considering the temporal resolution mismatch between the products and the rapid evolution of wildfire events, we implemented a 1- day non-aggressive buffer window before and after each fire alert to capture the closest available Sentinel-2 imagery. We used longitude and latitude as the locator of the fire incident. The coordinates in MODIS represent the center of a 1 km pixel and for VIIRS these coordinates represent the center of a 375 m pixel [11]. In contrast to the coarse resolution of MODIS and VIIRS sensors, Sentinel-2 Level 2A products are delivered in three spatial resolutions: 10 m, 20 m, and 60 m, depending on the spectral bands used [10]. Since we are using RGB composites, the spatial resolution is 10 m. To address the spatial resolution mismatch between the coarser resolution fire alert data from MODIS and VIIRS and the higher resolution Sentinel-2 imagery, spatial buffers of 0.02 degrees were applied around each fire point. This approach increased the likelihood of capturing relevant fire indicators, such as smoke or plumes, within the buffer area, even when the alert coordinate did not represent the exact hotspot location.
We experimented with different image resolutions dur- ing development. The resulting images have a resolution of 512x512 pixels and contain true-color composites, represent- ing the terrain and smoke features in a balanced way. Due to limited ground truth and resolution mismatches between Sentinel-2 and FIRMS data, manual filtering was applied. Only images showing visible smoke with confirmed FIRMS fire alerts were kept. This resulted in 178 high-quality wildfire samples for public release.

C. *Web Interface and Dataset Access*

After processing wildfire images, the remaining task was to accurately link them with corresponding wildfire records. Due to spatio-temporal mismatches between NASA FIRMS data and Sentinel-2, and a significant amount of unprocessed data, the number of usable images was limited. Multiple CSV files containing over 60,000 entries were cleaned and consolidated into a final dataset of 583 entries, each precisely matched to an image.
From these, 178 representative images were selected and displayed in an interactive online gallery. Users can sort and filter the images based on wildfire details, view each image with its location on Google Maps, and download individual or grouped results as .zip files (containing a .jpg and a
.txt), all after agreeing to the MIT License. The source code worked on the coordinates in the file, then calculated and compared the Azerbaijani regions' coordinates to find the prevention and strategic resource allocation during extreme conditions accurate longitude/latitude values. Using these coordinates, we embedded a Google Maps frame into the image details modal, displaying a pin at the center of the buffer zone, whose size varies based on the instrument's spatial resolution.

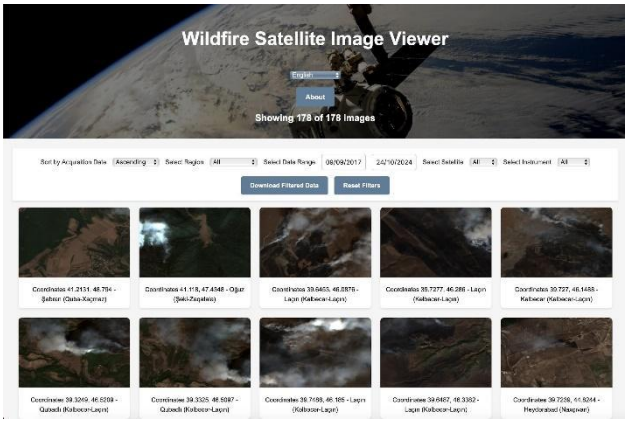


Fig. 4. Website, main page showing images, navigation sector, data filtering buttons etc.

Additionally, to support the user experience, the website supports both English and Azerbaijani. A JSON-based key-value translation system was implemented to switch languages dynamically, ensuring consistent user experience. The entire dataset and interface are publicly available to support both academic research and broader public understanding.

IV. Conclusion

Using Sentinel-2 and NASA FIRMS data, we collected satellite-based wildfire imagery that now contributes to the growing body of resources available for wildfire research and remote sensing applications, taking a step toward unlocking the potential of advanced DL techniques for near real-time fire detection in underrepresented regions like Azerbaijan. In order to visualize the results in a user-friendly and accessible format, we also developed a web-based interface, presenting the imagery in an intuitive format, making the results of extensive effort publicly available and easily usable.

With careful processing and optimization, we made the most of a relatively small dataset and created a valuable tool for environmental monitoring.

V. Future Work

Looking ahead, this project opens several promising directions for future development. Combining Sentinel-2 and NASA FIRMS data produced useful results for this study. However, it remains insufficient for broader, long-term appli- cations and real-world adaptability. A key step is expanding the dataset and integrating temporal forecasting to develop a predictive system capable of generating early warnings and risk assessments. While the current focus is on dataset creation and preprocessing, future work will explore DL methods using satellite imagery to detect wildfire indicators more efficiently. Such systems could support proactive measures, including fire

References

[1] J. R. Martinez-de Dios, B. C. Arrue, A. Ollero, L. Merino, and F. Go´mez-Rodr´iguez, "Computer vision techniques for forest fire per- ception," *Image and vision computing*, vol. 26, no. 4, pp. 550–562, 2008.

[2] Q. Zhang, J. Xu, L. Xu, and H. Guo, "Deep convolutional neural networks for forest fire detection," in *2016 International forum on man- agement, Education and information technology application*, pp. 568– 575, Atlantis Press, 2016.

[3] D. Rashkovetsky, F. Mauracher, M. Langer, and M. Schmitt, "Wild- fire detection from multisensor satellite imagery using deep semantic segmentation," *IEEE Journal of Selected Topics in Applied Earth Ob- servations and Remote Sensing*, vol. 14, pp. 7001–7016, 2021.

[4] T. Sung, Y. Kang, and J. Im, "Enhancing satellite-based wildfire mon- itoring: Advanced contextual model using environmental and structural information," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–16, 2024.

[5] C. Corradino, E. Amato, F. Torrisi, and C. Del Negro, "Data-driven random forest models for detecting volcanic hot spots in sentinel-2 msi images," *Remote Sensing*, vol. 14, no. 17, 2022.

[6] Y. Chen, J. Hall, D. van Wees, N. Andela, S. Hantson, L. Giglio, G. R. van der Werf, D. C. Morton, and J. T. Randerson, "Multi-decadal trends and variability in burned area from the fifth version of the global fire emissions database (gfed5)," *Earth System Science Data*, vol. 15, no. 11, pp. 5227–5259, 2023.

[7] I. El-Madafri, M. Pen˜a, and N. Olmedo-Torre, "The wildfire dataset: Enhancing deep learning-based forest fire detection with a diverse evolving open-source dataset focused on data representativeness and a novel multi-task learning approach," *Forests*, vol. 14, no. 9, p. 1697, 2023.

[8] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage, "Sentinel-2 data for land cover/use mapping: A review," *Remote Sensing*, vol. 12, no. 14, 2020.

[9] F. Huot, R. L. Hu, N. Goyal, T. Sankar, M. Ihme, and Y.-F. Chen, "Next day wildfire spread: A machine learning dataset to predict wildfire spreading from remote-sensing data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, p. 1–13, 2022.

[10] European Commission, Directorate-General for Internal Market, Indus- try, Entrepreneurship and SMEs, Space Policy, Copernicus and Defence, "Legal notice on the use of copernicus sentinel data and service information." https://sentinels.copernicus.eu/documents/247904/690755/ Sentinel Data Legal Notice, 2024. Accessed: Oct. 15, 2024.

[11] NASA Fire Information for Resource Management System (FIRMS), "Fire Information for Resource Management System (FIRMS)." https://firms.modaps.eosdis.nasa.gov. Accessed: Apr. 4, 2025.

[12] NASA, "MODIS web." https://modis.gsfc.nasa.gov/about/. Accessed: April 20, 2025.

[13] USGS, "VIIRS overview." https://lpdaac.usgs.gov/data/get-started-data/ collection-overview/missions/s-npp-nasa-viirs-overview/. Accessed: April 20, 2025.

[14] European Space Agency, "Sentinel-2 operations." https://www.esa.int/ Enabling Support/Operations/Sentinel-2 operations, 2023. Accessed: Apr. 20, 2025.

[15]

Information Search System of World Geoglyphs

Aydin Kazim-zada [1][0000-0003-0714-8481] and [2]Hayat Huseynova [2][0000-0001-6062-6599]

[1,2] AZERBAIJAN UNIVERSITY OF ARCHITECTURE AND CONSTRUCTION
Ayna Sultanova str.,5, Baku, AZ1073, Azerbaijan
aydin91@gmail.com, ibrahimlihayat@yahoo.com.tr

Abstract. The article examines informative features of geoglyphs of the world, found in many countries of the continents of the globe. The most famous of the geoglyphs are found in South America - the lines and drawings of the Nazca plateau, located in Peru, as well as on the neighboring to Nazca slightly less known Palpa plateau. Of the individual images, the most famous, for example, are the "Giant from the Atacama Desert" and the "candelabrum" on the Paracas Peninsula. The systematization and classification of geoglyphs is carried out and informative features are determined. A block diagram of the information retrieval system of geoglyphs is developed and presented. The geographic coordinates of the geoglyphs are given, which can be viewed on the Internet using the appropriate search programs.

Keywords: informative features, geoglyphs, classification, systematization, information retrieval system, geometric dimensions, period of discovery, geographic coordinates, cultural heritage.

1. INTRODUCTION

The ancient inhabitants of our land created not only petroglyphs on rocks, but also geoglyphs on the surface of the earth. If petroglyphs drawn on rocks, stone slabs, in caves have sizes from several centimeters to several meters, then geoglyphs from several tens, hundreds of meters, and sometimes even several kilometers. If petroglyphs can be seen close up at a short distance, then most geoglyphs can only be seen from a certain height above the earth's surface. Therefore, many geoglyphs were discovered after man gained the ability to travel by air. A number of geoglyphs were drawn on inclined terrain, so they are visible from a certain distance on the earth's surface. Their purposes remain unsolved to this day.

2. STATEMENT OF THE PROBLEM

Scientists around the world still cannot clearly answer the questions of who, how, with the help of what technology, for what purpose and for whom such huge drawings on the ground were intended. Natural and climatic conditions contribute to the fact that the lines of geoglyphs are erased over time. Tourist groups wishing to see geoglyphs with their own eyes and visiting these places can also damage the drawings. These drawings are of interest to specialists in various fields of science, technology, art, culture, etc. There is a need to preserve geoglyph images for future generations both in their natural form and in all possible media (books, magazines, optical CDs, DVDs, Blu-ray discs, semiconductor flash drives, SSD disks, etc.). All this work should ensure the preservation of the cultural heritage of past centuries. Geoglyphs are found on all continents of the globe in the territories of various countries. The developed information retrieval system will allow the systematization, classification, and search for geoglyphs by their images and informative features.

3. SOLUTION

Let's consider a number of the most famous geoglyphs in the world and their characteristics. In the southern part of Peru, in the Nazca desert, there are the most famous geoglyphs in the world, called the “Nazca Lines” (Fig. 1) [1,2].

a) b) c)

Fig. 1 Nazca Lines Scientists believe that the lines were created between about 200 BC and 500 AD.

There are several hundred figures in this area. Among the figures are images of animals such as spiders (Fig.1a), monkeys (Fig.1b), hummingbirds (Fig.1c), as well as sharks, dolphins, lizards, etc. There is even an image of a supposed “astronaut”. As with all geoglyphs in the world, the purpose of their creation and their meaning remain unknown. A number of theories have been put forward about the purpose of the geoglyph. Due to their enormous size, the most realistic reasons for their creation are considered to be either astronomical purposes or communication with extraterrestrial civilizations. But none of them finds its full confirmation. The largest geoglyph in the world is the image of an aborigine (the geoglyph “Marree Man” (Fig.2a) found in a sparsely populated region of Australia) hunting birds or kangaroos with a bat. Marree Man (English Marree Man) - Geoglyph, was discovered from the air on June 26, 1998. It is located on the Finneys Springs plateau, 60 km from the village of Marree (29 ° 31′45.75 ″ S 137 ° 27′56.02 ″ E (G) (O)). It is 4.2 km long. The lines of the figure are 20 - 30 cm deep, and up to 35 m wide. The origin of the Marree giant still remains a mystery, but a hypothesis is put forward that the image was created by the aborigines of the Pitjantjatjara tribe Scientists have determined that the drawing was created in our era and its origin, meaning and purpose also remain a mystery. In the area where this giant drawing is located, no human footprints, car tires, or traces of any construction equipment have been found. On a hill in the English county of Oxfordshire is an image called the “Uffington White Horse” (Fig.2b). It was discovered that the drawing was created by deep trenches filled with tiny white chalk. It was created between 1200 and 800 BC [1]. The Giant of the Atacama Desert is a large anthropomorphic geoglyph (Fig.2c). The geoglyph is the largest prehistoric anthropomorphic drawing in the world, 86 meters long. The age of the drawing is estimated at 9000 years. The image of the giant is located 1370 km

from the geoglyphs of the Nazca Desert, on the lonely mountain Cerro Unica, in the Atacama Desert (Chile). See this in full the geoglyph can only be seen from an airplane. The Andean Candelabra geoglyph (Fig.2d) can be seen on the northern side of the Paracas Peninsula in Peru. It is believed to have been created around 200 BC. The geoglyph was dug into a coastal slope made of compacted sand. The walls of the canals are lined with stone, which has ensured their long preservation. The depth of the canals is up to 60 cm. The size of the drawing: length - 128 m, width - 74 m (according to other sources - a little more than 70 m), and the thickness of the lines is from 1.5 to 4 m. One of the assumptions about the purpose of this geoglyph is that it was created as a sign for sea vessels [1].

a) b) c) d) e)

Fig.2 a) Marree Man, b) Uffington White Horse, c) Atacama Giant, d) Andean Candelabra, e) Cerne Abbas Giant

Geoglyphs have also been discovered in the Californian town of Blythe in the Colorado Desert. In the Amazon basin, deforestation has revealed a vast network of mysterious ancient geoglyphs, believed to have been created by a pre-Columbian civilization. The Cerne Abbas Giant (Fig.2e) was found near the village of Cerne Abbas in England, and many others. Measuring 55 meters high and 51 meters wide, the giant figure of a naked man is carved into the side of a steep hill, and is best seen from across the valley or from the air. The figure is drawn on the hill by a trench 30 centimeters wide and about the same depth, which was made through the grass and earth, as a result of which the chalk was exposed. Along with the geoglyphs presented in the article, the most famous ones include the geoglyphs stretching across the territories of Saudi Arabia, Syria and Jordan and discovered by aerial photography called "Works of the Old Men". Hundreds of thousands of geoglyphs are scattered in these territories with the image of many "wheels". It is believed that some of them, called "Works of the Old Men", are several thousand years old, which makes them much older than the more famous Nazca Lines. It should be noted that along with photographs of geoglyphs in scientific articles, in popular printed publications, modern information technologies and the Internet allow you to view their drawings on a computer screen. To do this, you need to enter the geographic coordinates of the geoglyphs into the maps of such search engines as Google, Yandex, OpenStreetMap, etc. [5]. The geographic coordinates of some geoglyphs are given in Table 1. Based on the considered examples of geoglyphs and their characteristics, located on different continents and countries of the world, the following informative features were defined for the systematization and classification of geoglyphs:

- Continent
- Country
- Geoglyph name
- Discovery date
- Creation period
- Type of lines
- Geometric dimensions of lines
- Geometric dimensions of geoglyph
- Geographic coordinates.

These informative features are the basis of the structure of the information retrieval system of geoglyphs [4]. The block diagram of this information retrieval system with an example of the “Marree Man” geoglyph is shown in Fig. 3.

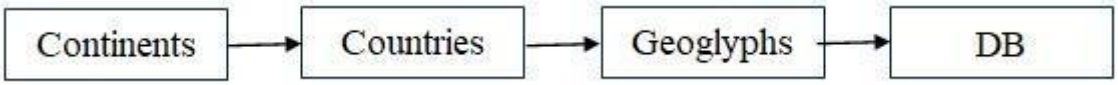


Fig.3 Block diagram of the information retrieval system

Table 1. Table of some data of the information retrieval system of geoglyphs of the world

		Date and place of discovery	Period of creation	Geometric dimensions	Type and size of lines	Geographic coordinates
1	Marree Man	Discovered from the air on June 26, 1998. Located on the Finneys Springs plateau, 60 km from the village of Murray	The image was created by the Pitjantjatjara tribe (hypothesis).	4.2 km length	20 - 30 cm deep and up to 35 cm wide	(29°31′45.75″ S 137°27′56.02″ E (G) (O))
2	"The Uffington White Horse"	The White Horse Hill first appeared in records in the 1070s, and the image itself was mentioned in the records of Abingdon Abbey in 1190.	The mysterious image is dated back to at least the Iron Age (approximately 1200 and 800 BC).	Length 112 meters Width 30 meters	filling deep trenches with broken chalk	51°34′39″ N 1°34′00″ W / 51.5775° N 1.566667° W 51.5775, -1.566667

3	Andean Candelabrum		created around 200 BC	Size of the drawing: length - 128 m, width - 74 m,	The depth of the ditches is up to 60 cm the thickness of the lines is from 1.5 to 4 m.	3°47′42″ S w. 76°18′18″ W long / 13.795° south w. 76.305° W d.-13.795, -76.305

4. CONCLUSION

Examples of the most famous geoglyphs in the world are considered and given. Based on the data and characteristics known about them, informative features are defined. An information retrieval system has been developed for systematization and classification of geoglyphs based on their informative features, the block diagram of which is given. This information retrieval system can be used by a wide range of scientists, specialists in various fields, and those simply interested in mysterious geoglyphs.

References

[1] Sklyarov A.Yu. - Nazca: giant drawings on the fields. "Nazca: giant drawings on the fields": Veche; Moscow; 2013, 169 p. ISBN 978-5-4444-0192-7 (In Russian)

[2] Erich von Daniken. Mysterious Nazca Drawings. Kharkov, Belgorod, 2010, 192 p. (In Russian)

[3] One of the earliest references to the «Rude Man» is in: Simon Welfare, John Fairley, «Arthur C. Clarke's mysterious world», A & W Publishers, 1981, ISBN 0-89104-268-7, ISBN 978-0-89104-268-6. 217 pages.

[4] Abdullaeva G.G., Kazim-zada A.K., Kurbanova N.G. "Information system of petroglyphs of the countries of the world". 12th International Interdisciplinary Scientific and Practical School-Conference "Modern Problems of Science and Education", Evpatoria, April 27 - May 09, 2012. pp. 213-214

[5] https://ru.wikipedia.org/wiki/Геоглифы_Наски (In Russian)

Section 6: Applied Problems

Solution of an Optimal Control Problem Related to Technological Operating Modes of Gas Wells

Kamil Mamtiyev
Department of Digital Technologies and Applied İnformatics
Azrerbaijan State University of Economics (UNEC)
Baku, Azerbaijan
kamil.mamtiyev@unec.edu.az

Rena Mikayilova
Department of Digital Technologies and Applied İnformatics
Azrerbaijan State University of Economics (UNEC)
Baku, Azerbaijan
rana.mikayilova@unec.edu.az

Hafiz Bayramov
Department of Digital Technologies and Applied İnformatics
Azrerbaijan State University of Economics (UNEC)
Baku, Azerbaijan
hafiz.bayramov@unec.edu.az

Abstract— Selecting a well operating mode is a critical step in field development planning and well operation. For gas wells, the operating mode includes regulating the flow rate or bottom-hole pressure to ensure compliance with subsoil protection regulations and safe operation. Mathematical formulas, such as the maximum permissible depression mode, allow us to describe these operating modes. The use of optimization methods, including the Pontryagin maximum principle, has advanced the modeling and solution of oil and gas field development problems. However, there are still unresolved issues, such as the formulation of optimal control problems, the definition of special controls, and numerical solution methods. In this paper, the gradient projection method is used to determine the optimal operating modes of gas wells, paying attention to minimizing reservoir depletion and refining control strategies.

Keywords— gas wells, field development planning, nonlinear differential equations, oil and gas industry, gradient projection method

INTRODUCTION

Well operating mode is a key element of field development and ensuring safe operation. For gas wells, these modes play a decisive role, allowing to regulate bottom-hole pressure and flow rate by the rules of subsoil protection and accident-free operation requirements. Mathematical models, such as the maximum permissible depression mode, are widely used in practice, ensuring the reliability and efficiency of operational processes. Modern optimization methods, in particular, the Pontryagin maximum principle, have opened up new opportunities for solving complex problems associated with field development. Despite the progress, many unresolved issues remain in the field of field development management, such as identifying special controls and refining optimal strategies. This article is devoted to the study of these problems and the application of numerical methods to develop solutions that can minimize reservoir depletion and improve management efficiency.

Selecting the technological operating mode of wells is one of the most critical decisions made during the field development planning and the well operation process. The technological operating mode of gas wells generally refers to maintaining specific conditions at the wellbore by regulating the production rate or bottom-hole pressure to ensure compliance with subsurface conservation regulations and the safe, accident-free operation of the wells [1].

Some technological regimes can be expressed using mathematical formulas [2]. The maximum allowable reservoir drawdown mode is one of the simplest technological operating modes for gas wells. This widely used mode in practice is mathematically represented as follows:

$$p_{layer}(t) - p_c(t) = \delta,$$

where $p_{layer}(t)$ is a reservoir pressure of a certain well at a given time t; $p_c(t)$ is a bottom-hole pressure in the same well at the moment of time t; δ is a permissible depression on the layer.

It should be noted that after the emergence of Pontryagin's maximum principle, optimization methods have been widely applied in modeling and solving various problems related to the development of oil and gas fields. Despite the existing successes, there are still no sufficiently convincing formulations for field development control problems. This is despite the fact that, by their nature, these problems should be based on optimal control theory methods. Undoubtedly, optimization methods have not yet been adequately utilized when assessing the future development of individual fields or the long-term growth of the oil and gas industry.

PROBLEM STATEMENT

In this note, the technological operating mode of gas wells is determined under the condition that the gas reservoir is depleted by a specified deadline. In terms of dimensionless variables, this problem leads to defining a piecewise-continuous function $q(t)$, which does not exceed an absolute value of one over the interval $[0, T]$, based on the criterion of minimizing a given functional

$$S(q) = \int_0^1 [p(x,T) - p^0(x)]^2 dx\,, \qquad (1)$$

that characterizes reservoir depletion. Here $p^0(x)$ is a given function based on technological conditions, $p(x,t)$ describes the distribution of gas pressure, which in a rectangular region $\{0 < x < 1, 0 < t \le T\}$ satisfies the equation

$$\frac{\partial p}{\partial t} = \frac{1}{2} \cdot \frac{\partial^2 p^2}{\partial x^2}, \qquad (2)$$

and at the border Q satisfies the conditions

$$p(x,0) = const = 1, 0 \le x \le 1 \qquad (3)$$

$$\frac{\partial p^2(0,t)}{\partial x} = q(t), \frac{\partial p^2(1,t)}{\partial x} = 0, 0 \le t \le T. \qquad (4)$$

Equation (2) is a nonlinear differential equation of the parabolic type, describing the process of unsteady filtration of an ideal gas in a homogeneous porous environment. In the literature, it is referred to as the L.S. Leibenson equation [2].

Condition (3) implies that at the initial moment, the reservoir is undisturbed with an initial uniform pressure. The first boundary condition in (4) indicates that the well, located at the point $x = 0$, operates with a flow rate $q(t)$. The task is to select $q(t)$ under the constraint $|q(t) \leq 1|$ in such a way that by the end of the process, the deviation of the pressure distribution $p(x,T)$ from the given function $p^0(x)$ is minimized. The second boundary condition in (4) specifies the impermeability of the reservoir boundary $x = 1$.

NUMERICAL SOLUTION OF PROBLEM (1)-(4)

It can be shown that the variation of the functional (1), corresponding to a variation $\Delta q(t)$, takes the following form:

$$\Delta S = -\int_0^T \psi(0,t)\Delta q(t)dt + \eta, \quad (5)$$

where

$$\eta = \int_0^1 \Delta p^2(x,T)dx + \frac{1}{2}\int_0^1 \int_0^T \frac{\partial \Delta p^2(x,t)}{\partial x}\frac{\partial \psi(x,t)}{\partial x}dxdt \quad (6)$$

is a remainder term, $\psi(x,t)$ is a solution of the adjoint boundary value problem

$$\frac{\partial \psi}{\partial t} = -p \cdot \frac{\partial^2 \psi}{\partial x^2}, 0 < x < 1, 0 < t \leq T \quad (7)$$

$$\psi(x,T) = 2[p(x,T) - p^0(x)], 0 \leq x \leq 1, \quad (8)$$

$$p(0,t)\frac{\partial \psi(0,t)}{\partial x} = p(1,t)\frac{\partial \psi(1,t)}{\partial x} = 0, 0 \leq t \leq T, \quad (9)$$

and $\Delta p(x,t)$ is a solution to the boundary value problem (2) – (4) in increments.

From (5), it follows that the Hamiltonian function *H* in the problem (1)–(4) is given by $H = -\psi(0,t)q(t)$. According to Pontryagin's maximum principle, if the control $q^*(t)$ minimizes the functional (1), then it must maximize the function *H*, that is:

$$q^*(t) = -sign(\psi(0,t)) \quad (10)$$

Equation (10) provides insight into the structure of the optimal control and its general form without the need to explicitly solve the optimal control problem (1) – (4). Thus, in the considered problem, the optimal control formally takes the form of a piecewise constant function that alternates between the boundary values *1* and *-1*. At first glance, solving the problem reduces to selecting the optimal sequence of control intervals and their switching points.

However, it is important to note that the formal nature of equation (10) also implies that the function $\psi(0,t)$ under the $sign$ function may vanish not only at isolated points within the interval $[0,T]$ but also over entire sub-intervals. In such cases, Pontryagin's maximum principle alone is insufficient to determine the optimal control, and additional analysis is required to identify so-called singular controls [3].

Another challenge arises in obtaining estimates for the remainder term in equation (6) within formula (5). Furthermore, an open question remains regarding the possible number of switching points and their precise locations within the time interval $[0,T]$.

Nevertheless, according to the recommendations of the paper [4], in some *a priori-specified* classes, the solution to problem (1)–(4) can be obtained by numerical methods, in particular, by the gradient projection method.

The essence of the gradient projection method can be described as follows: based on some considerations, some initial control $|q^k(t)| \leq 1$ is specified, and the boundary value problem (2) – (4) is being solved. Then, from the obtained solution at the time $t = T$, the value (8) is determined, and the equations for the conjugate variables are integrated to clarify the new control $q^{k+1}(t)$ according to the scheme

$$q^{k+1}(t) = \{q^k(t) + \delta q^k(t), if|q^k(t) + \delta q^k(t)| \leq 1, 1, if q^k(t) + \delta q^k(t) > 1, -1, if q^k(t) + \delta q^k(t) < -1, \quad (11)$$

Where

$$\delta q^k(t) = \alpha \psi^k(0,t), k = 0,1,2,\ldots \quad (12)$$

Here, k is an iteration number, $\psi^k(0,t)$ is a solution of problem (7) – (9) with iterations, and α is a step size, which is selected by one of the methods specified in [5].

In the numerical solution of problem (1) - (4), equations (2) and (7) are integrated using the method of straight lines, taking into account the corresponding boundary conditions. To check the optimality of the control found using scheme (11)-(12), the solution of problem (2)-(4) is taken as $p^0(x)$ for a given control $q^*(t)$. In this case, the value of the functional is zero, and the approximately found optimal control does not have a character significantly different from $q^*(t)$ in the time interval [0, T], and with an increase in the number of iterations, approaches it.

The convergence of the process in terms of the functional is as follows:

k	$S(q^k)$	$max\psi^k(0,t)$	k	$S(q^k)$	$max\psi^k(0,t)$
0	$1.6026 \cdot 10^{-2}$	0.2812	20	$6.1205 \cdot 10^{-3}$	$1.7562 \cdot 10^{-1}$
1	$1.5429 \cdot 10^{-2}$	0.2760	39	$9.0553 \cdot 10^{-4}$	$7.0936 \cdot 10^{-2}$
2	$1.4843 \cdot 10^{-2}$	0.2708	47	$8.8225 \cdot 10^{-5}$	$2.3010 \cdot 10^{-2}$
3	$1.3702 \cdot 10^{-2}$	0.2656	52	$3.6741 \cdot 10^{-5}$	$-1.6203 \cdot 10^{-3}$
10	$1.0534 \cdot 10^{-2}$	0.2341	53	$5.8798 \cdot 10^{-6}$	$5.2725 \cdot 10^{-3}$

CONCLUSION

Thus, the calculations performed enable us to draw the following conclusions:

1. The presented gradient method provides more accurate results in a relatively short time, although it requires a comparatively large number of iterations.
2. For any initial feasible control, convergence by functional is observed.
3. For ill-posed problems, the gradient projection method with a specially chosen step size did not show a tendency to "diverge" and provided a convergent sequence of controls.

REFERENCES

[1] K. Mamtiyev, T. Aliyeva, U. Rzayeva (2021). Analysis of one class of optimal control problems distributed-parameter systems, Eastern-Eurepan Jurnal of Enterprise Technologies, 5(4(113)), pp.26-33, https:// doi.org/10.15587/1729-4061.2021.241232.

[2] S. N. Zakirov, B. B. Lapuk. Design and development of gas fields. Moscow, "Nedra", 1974, 374 p.

[3] R. Gabasov. Special optimal controls R. Gabasov, F. M. Kirillova. - Moscow: Nauka, 1973.-256 p.

[4] G. Butkovsky. Theory of mobile control of systems with distributed parameters. Moscow: Nauka, 1980.-384 p.

[5] F. P. Vasiliev. Methods for solving extremal problems. Study guide, Moscow: Nauka, 1981, 400 p.

On the Method of Forecasting Abnormal Pressure in Oil and Gas Wells and the Technical Implementation of an Intelligent Forecasting System

Bikas Aghayev
Institute of Information Technologies
of the Ministry of Science and Education
Baku, Azerbaijan
bikies418@gmail.com
[0000-0002-5258-7718]

Maleyka Pashayeva
Institute of Information Technologies
of the Ministry of Science and Education
Baku, Azerbaijan
maleykapashayeva15@gmail.com
[0000-0002-9360-7041]

***Abstract*— The article examines the processes of forecasting abnormal reservoir pressure in oil and gas wells and identifying abnormal zones. Some accidents and complications arising as a result of abnormal reservoir pressure are considered. A method for predicting abnormally high reservoir pressure is proposed. The method is based on the use of the relationship between the technical parameters of the drilling process. A prototype of a forecasting system operating on the basis of the method has been developed. For this purpose, a schematic technical solution of the system, an operating algorithm and a control program in the C^{++} language and other engineering solutions have been developed. In order to ensure the intelligence of the forecasting process, software blocks have been developed that compare current drilling parameters with the design documentation for well construction and take into account the drilling results of previously drilled wells located nearby. The functionality of the prototype was verified through computer modeling experiments in laboratory conditions. The experiments used a data block of real mechanical drilling parameters obtained from a borehole during the implementation of the SOCAR grant project. A brief description of the system's operation is provided.**

Keywords—drilling of oil and gas wells, oil and gas production, abnormal reservoir pressure, accidents and complications, forecast, method of prompt detection of high-pressure zones, forecasting system, test experiments

I. INTRODUCTION

Macroeconomic indicators of many countries depend significantly on the oil and gas industry. Thus, last year, 92.5 percent of our country's total exports were oil, gas and oil products. It is known that during drilling and operation of wells of any type, a number of accidents and complications occur due to abnormally high reservoir pressure (AHRP), which emits harmful gases and other compounds into the atmosphere, creating a large amount of greenhouse gases and ozone-depleting substances, the effects of which cause environmental pollution.

The SOCAR company is the largest enterprise in our republic and the enterprise that pollutes the environment the most. According to statistics, every fourth drilled well has fluid manifestations of varying strength, open fountains, griffins, etc. cases. At this time, a significant amount of oil, gas, mineralized water, etc. is emitted into the environment (atmosphere, soil, water basins, underground waters) [1, 2]. During the normal drilling process, substances such as drilling mud, drilling mud, formation water, condensate, etc. are emitted from the bottom of the well into the atmosphere. These ingredients contain many substances and compounds that cause global warming, have a destructive effect on the ozone layer and are harmful to the human body.

Therefore, along with a number of other methods, preventing the formation of oil and gas fountains and other complications is relevant and very important from the point of view of environmental protection and ecological safety. One of the important methods used for this purpose is preventing the occurrence of fountains through forecasting. In most cases, abnormally high reservoir pressure is taken as a forecasting parameter.

The article proposes a method for forecasting abnormally high reservoir pressure during the drilling of oil and gas wells, and also considers a system created on the basis of this method.

II. ON THE CAUSE-AND-EFFECT RELATIONSHIPS OF COMPLICATIONS AND ACCIDENTS

The history of oil drilling shows that all oil and gas provinces of the Earth have zones of abnormally high pressure in the rocks. Starting with the fountain that appeared in the world's first industrially drilled oil well (Bibi-Heybat on July 14, 1848), many wells drilled and operated to this day have experienced various accidents.

In technical literature, accidents that occur during well construction mainly mean the following:

- accidents with the drill string and protective column (or their elements);
- seizure of the drill string and protective column;
- accidents with the drilling tool;
- poor cementing of the well walls, etc.;

Complications mainly relate to situations arising due to abnormal reservoir pressure:

- fluid manifestations;
- violation of the integrity of the well walls;
- absorption of drilling mud;
- collapse and spillage of the well walls, etc.

The main reasons for complications are:

- non-scientific organization of the drilling process;
- failure to use advanced equipment and technologies (especially blowout prevention devices, control and measuring devices, etc.);
- use of equipment with an expired shelf life;

- errors made in the design documentation related to well construction;
- human factor, etc.

The harmful impact on the environment and people (primarily on the operating personnel) during drilling and operation of oil and gas wells, especially flowing ones, is mainly associated with the following factors:

- emissions into the atmosphere (methane, volatile organic compounds, etc.) during normal operation, especially during open and uncontrolled fountains;
- flaring of associated gas;
- removal of drilling products (drilling mud, oil sand, highly mineralized formation water, etc.) to the surface of the well;
- well flushing, well equipment, etc. items;
- storage/burial of drilling mud, formation water, etc. drilling products in accumulation dumps, open ground barns).

The above factors affect all components of the ecosystem, including:

- into the atmospheric air;
- to surface and underground water bodies;
- soil and flora;
- to the animal world; - has a harmful effect on drilling personnel and the population living in certain areas.

As an example of environmental pollution as a result of AHFP, we can cite a powerful open gas fountain that formed during the drilling of gas well No. 90 at the Bulla-Deniz field (Horizon VIII, design depth 5800 m, Girmeki Sandy Bunch). The fountain existed for 68 days, daily ejecting 8 million cubic meters of gas condensate and 704 tons of oil into the atmosphere. As a result, the environment was seriously polluted. Over the past 10 years, more than 88.3 thousand hours have been spent on eliminating accidents that occurred at wells in offshore fields in Azerbaijan alone, and 54 wells have been liquidated (mothballed) [3].

Conducted studies prove that oil and oil product discharges polluting the environment are also a source of serious danger to human health [2]. For example, nitrogen oxide has an irritating effect on the mucous membrane and respiratory organs. Its concentration above the permissible limit causes irreversible changes in the lungs. Sulfur oxide causes general cardiovascular and respiratory diseases. High concentrations can lead to death. Sulfuric acid, formed as a result of a reaction with water vapor in the air, causes pathological conditions in the respiratory organs.

It is clear from the above that the assessment of pressure anomalies in order to prevent oil and gas fountains, as well as the forecast of abnormal pressure zones, is important from the point of view of environmental protection. Most of the methods used for this purpose are based on the use of geophysical well survey data, but collection, storage, processing and decision-making are time-consuming. That is, it is impossible to make operational decisions. On the other hand, the process of geophysical surveys is expensive and requires suspension of drilling (deepening of wells).

In order to forecast AHRP, it is technologically and economically more expedient to use mechanical drilling parameters. Because these methods do not require stopping the drilling process, but at the same time allow using forecast data for decision-making in real time. It becomes possible to fully automate the forecasting process.

III. ABOUT THE PROPOSED METHOD OF FORECASTING AHRP

Research shows that in countries (companies) that do not use modern achievements of geology and geophysics, modern methods and technologies, including methods of prompt detection of abnormal pressure zones, accidents occur more often during drilling of oil and gas wells and these cases cause greater damage.

As mentioned above, systems created on the basis of the dependence of mechanical parameters of drilling (MPD) are more widely used for forecasting AHRP zones due to a number of advantages [3,4]. The operating principle of these systems is based on the use of formulas created on the basis of mathematical formalization of the patterns of change in dependencies between MPD. These systems also allow for the full automation of the forecasting process. The disadvantage of these methods is the fact that the level of forecasting depends on the lithological stability of rocks. In practice, to improve the accuracy of detection of these systems, one or more simple forecasting methods are used in parallel [5,6].

As a result of monitoring conducted within the framework of the SOCAR grant project (21 LR AMEA), we have established that the specified systems are not used in the company's drilling practice for the purpose of forecasting. For this purpose, a method was developed and a test sample of the system created on the basis of this method was tested using computer modeling experiments. The mathematical expression of the method is as follows.

$$\alpha = \frac{lg\frac{V_{mex}}{60V_{fs}}}{lg\frac{P_y}{\rho_{qm}D_{qa}^2}} \qquad (1)$$

where

V_{mex} – denotes - meter/hour;
V_{fs} - drilling tool rotation speed-cycle/min;
ρ_{qm} – drilling fluid density-g/cm^3;
P_y – drilling tool load-kg;
D_{qa}– drilling tool diameter-cm.

To facilitate practical use, the formula includes the drilling mud density parameter, and the parameter values are converted to the international system of units. This formula can also be used to calculate the density of the weighted drilling mud required to restore the disturbed pressure balance $(-\Delta P)$ between the hydrostatic and formation pressure in the high-pressure zone (P_{hidst}= (1÷1,3 $P_{f.p}$):

$$\rho_{qm} = \frac{P_y}{10^n D_{qa}^2} \qquad (2)$$

where,

$$n = \frac{lg\frac{V_{mex}}{60V_{fs}}}{\alpha_{max}} \qquad (3)$$

Analysis of this formula shows that α function allows for tracking the degree of compaction/porosity and differential pressure. In other words, there is a strong correlation between reservoir pressure and V_{mex} and α parameters. In normal mode with depth increase, reservoir pressure increases along the "normal seal line", whereas MDS decreases linearly inversely. Correspondingly, the value of α function increases.

From the moment of opening of porous layers (while entering AHRP zone), reservoir pressure P_f increases abruptly, and rocks are drilled faster, i.e., $V_{mex.}$ increases suddenly. In accordance with the change in these parameters, the value of α function increases monotonically over the whole AHRP zone. Hence, the pattern among MDP can be used for the evaluation of high reservoir pressure and AHRP zone forecasting.

The principle of identifying high pressure zones using this method is more clearly illustrated in the graph below (fig. 1).

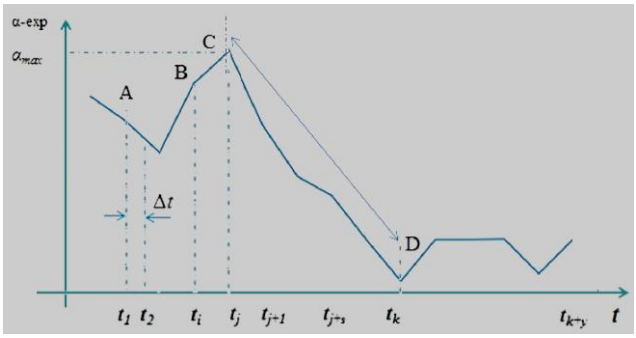


Fig.1. AHRP zone identification chart

If the drilling process continues under normal reservoir pressure conditions, the average value of α increases monotonically with depth (small amplitude deviations are ignored). From the moment when the drilling tool (DT) opens the porous zone, the mechanical drilling speed (MDS) increases sharply and, accordingly, the value of α begins to decrease monotonically, starting from the moment $N\Delta t$ (Δt is the parameter inverse to the frequency of sensor polling). On the graph, for the period t_1-t_j, the decrease in the value of α occurred only twice (in the period $2\Delta t$, N=2), and since it is less than some m, it is not considered as a case of detecting an anomalous zone.

Since the value of α is successively reduced by $N = m$ times in the CD segment (the formation interval t_j-t_k, the drilled during this period is defined as the AHRP zone. From this moment on, the weighted density of the DT for neutralizing the abnormal pressure is calculated according to one of the known rules. The value of m is selected experimentally depending on the formation thickness, drilling mode and other parameters specified in the well design documentation.

The general architectural scheme of the forecasting system is as follows (fig. 2):

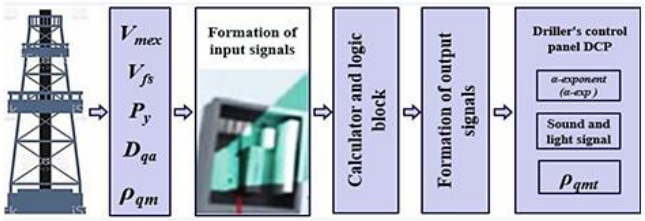


Fig.2. Forecasting principle of AHRP zones

A brief explanation of how the system works is as follows. The MPD is received by the input unit of the system from digital sensors of the equipment installed on the well platform, where it undergoes primary processing (noise removal, normalization of the amplitude-time parameters of the input signals, formatting according to the requirements of the input-output protocols of the computer, etc.). These signals are received by the computer unit from the input unit and the system is started: based on the current MPD values, the values of the function α are calculated and output to the driller's control panel (DCP). The calculation cycle is repeated until the value of the function α decreases consistently and monotonically $m=m_t$ times. In this case, sound and light warning signals are transmitted to the DCP. From this point on, the system calculates the required density of the DM using the sequence of α values and the α_{max}, value, and outputs the result to the DCP.

The current parameters of the AHRP zone (depth, formation pressure, formation temperature, etc.) are compared with the parameters: a) with the corresponding values of previously drilled wells nearby; b) well data specified in the design documentation. Information about significant differences that have occurred is displayed on the driller's screen. Drilling crew makes data-driven decisions.

Based on the above-mentioned general functionality of the system, an algorithm and a control program in C^{++} were developed, demonstrating the full work cycle. The control of the system's operation was verified by experiments conducted using computer modeling.

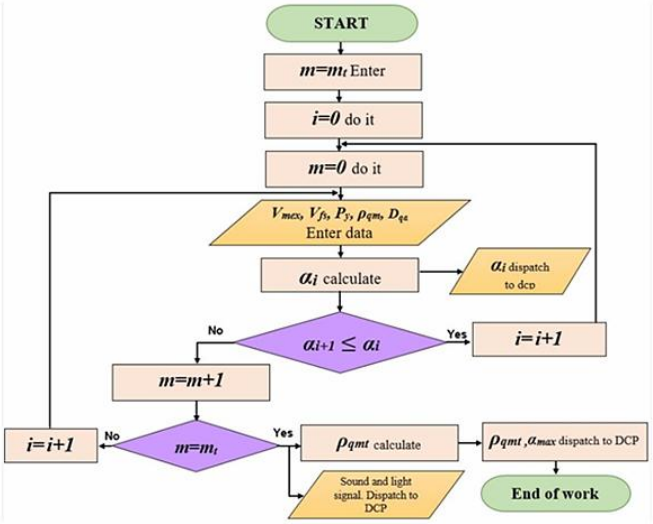


Fig. 3. Algorithm for the forecasting system

Experiments, an array formed on the basis of real MDP values obtained for wells drilled during the implementation of the grant project was used as input data. These data are entered into the system in two ways: sequentially for each computing session manually, or as a data block placed in the program code for automatic input. It should be noted that the forecasting accuracy of this group of systems mainly depends on the accuracy class of the devices measuring the density of the DM. Currently, a wide variety of digital devices with different functionality (and price) measuring the density of drilling mud are produced in a number of foreign countries. Information on some of these devices is given in [7].

IV. CONCLUSION

Article considers a number of accidents and complications caused by abnormal reservoir pressure in oil and gas wells. It is shown that during fountaining caused by abnormally high reservoir pressure, oil, gas condensate, drilling mud, mineralized formation water, oil sand and other

substances are emitted into the atmosphere. The causes of fountaining, the potential danger they pose to the environment and human health are noted. The importance of predicting abnormal reservoir pressures and identifying abnormal zones to prevent harmful discharges is substantiated. A method for predicting abnormally high pressure zones is proposed. Based on this method, a test sample of the system was made, an operating algorithm was developed and a control program was written in C^{++}. A brief description of the structural diagram and the operating principle of the system based on the algorithm is given. The operability of the system was tested by computer modeling on a test sample.

REFERENCES

[1] S.L. Davydova and V.I. Tagasov, "Oil and oil products in the environment," M., RUDN University Press, 2004, 163 p.

[2] A.A. Zhakisheva, "Environmental consequences of oil and gas extraction, "Bulletin of Chelyabinsk State University. Economics., Issue 33, No. 31(246), 2011, pp. 137–141.

[3] I.H. Dadashov and C.H. Abyshov, "The main indicators of drilling operations in Azerbaijan and the possibility of their improvement", Azerbaijan oil industry, 10, 2012, pp. 14-18.

[4] A.A. Mammadov, Q.N. Qakhramanov and G.A. Mammadova, "The role of abnormally high formation pressure in the forecast of oil and gas distribution in the South Caspian depression," Azerbaijan oil industry, 2021. 2, pp. 4–9.

[5] B.A. Korotayev, B.M. Basekha and A.M. Onufrik, " A method for assessing reservoir pressure during exploratory drilling," Bulletin of MSTU, 20 (1), 2017, pp. 104–110. DOI: 10.21443/1560-9278-2017-20-1/1-104-110.

[6] R.M. Alguliyev and B.S. Aghayev, "On the feasibility of using technologies for operational forecasting of abnormal reservoir pressures in oil and gas wells of SOCAR, "SOCAR Proceedings No.2, 2024, pp. 095-100. DOI: 10.5510/OGP20240200972.

[7] M.T. Biletskiy, B.T. Ratov and D. Delikesheva, "Automatic mud density measurement device," MIAB. Mining Inf. Anal. Bull, 7, 2019, pp. 140-148.: 10.25018/0236-1493-2019-07-0-140-148.

Reducing Heat Density and Improving Energy Efficiency in Cooling Systems of Data Centers

Rashid Alakbarov
Institute of Information Technolgy Baku, Azerbaijan
rashid.alakberov@gmail.com

Ogtay Alakbarov
Institute of Information Technolgy
Baku, Azerbaijan
oqtayalakbarov@yahoo.com

***Abstract*—This paper underscores the significance of implementing advanced cooling technologies to improve energy efficiency in data center operations. Data centers (DCs) play a fundamental role in processing, modeling, and simulating large-scale datasets, yet they require substantial energy resources. Optimizing cooling systems is essential to maintain energy efficiency and system reliability. Inadequate cooling performance can lead to elevated operational costs, hardware malfunctions, and service disruptions. This study explores key energy-related challenges in DC environments and investigates strategic approaches to minimize power consumption.**

Keywords— data centers, cooling systems, carbon emissions, energy consumption, heat density, energy efficiency.

INTRODUCTION

In today's rapidly evolving technological landscape, data centers (DCs) have become foundational components of both the digital ecosystem and the global economy. These facilities are responsible for processing, storing, and distributing massive volumes of information, positioning them among the highest consumers of electrical power. A significant share of this energy supports the continuous function of computational infrastructure, including servers, storage systems, and networking hardware. Equally important is the energy dedicated to thermal regulation, as managing the heat generated by such high-performance systems is essential. Thermal inefficiencies can severely impact both operational effectiveness and long-term sustainability. Prolonged exposure to elevated temperatures may lead to recurrent overheating, increased risk of component degradation, and even complete hardware failure. As a result, optimizing cooling mechanisms within DCs remains a key priority in ensuring reliability and energy-conscious operations.

COOLING SYSTEMS IN DATA CENTERS

A substantial portion of the energy consumed by data centers (DCs) is allocated to their cooling systems. On average, around 52% of the electrical energy used in a DC is consumed by IT equipment, while 38% is dedicated to cooling systems, and the remaining 10% powers other infrastructure components such as power distribution units and UPS systems [1]. Most of the energy supplied to IT equipment is eventually transformed into heat, which must be dissipated to maintain system functionality. The energy distribution across devices varies based on factors such as data center size, geographical location, cooling technology employed, and the overall energy efficiency of the center. Traditional cooling systems, such as CRAC/CRAH units and raised-floor systems, tend to lead to higher energy consumption. In contrast, more modern methods such as liquid cooling and natural (free) cooling are typically more energy-efficient. The application of effective cooling technologies and the optimization of cooling system energy efficiency can result in significant reductions in overall energy costs in DCs [2,3]. As these systems are primarily designed to manage the heat generated by servers, their efficient operation is crucial for minimizing energy expenditure [4]. This section provides a comparative analysis of various cooling technologies used in DCs, including traditional systems, liquid cooling, natural cooling, in-row cooling, and evaporative cooling.

2.1. Traditional Cooling System (CRAC/CRAH + Raised Floor)

The traditional cooling system, commonly used in many data centers, consists of CRAC (Computer Room Air Conditioning) or CRAH (Computer Room Air Handler) units. These systems deliver cooled air to servers via an underfloor plenum created by raised floor systems [5]. Despite their widespread use, traditional systems are often characterized by high energy consumption, as cooling the entire room leads to unnecessary energy expenditure. The Power Usage Effectiveness (PUE) coefficient for traditional systems typically ranges from 1.7 to 2.0, indicating relatively low energy efficiency.

2.2. Liquid Cooling

Liquid cooling systems (LCS) offer a more energy-efficient solution for cooling high-performance processors and other components. This system operates by circulating a coolant through cooling blocks, which absorb and dissipate heat more effectively than air. Given that liquid is a superior heat conductor compared to air, LCS provides substantial energy savings and high cooling efficiency [6,7]. The PUE coefficient for liquid cooling systems ranges from 1.2 to 1.4, significantly reducing energy consumption in comparison to traditional methods.

2.3. Free Cooling

Free cooling leverages the external environment's low temperature to cool air or water inside the data center. This method minimizes the reliance on mechanical air conditioning, thereby reducing energy consumption to a bare minimum [8]. The PUE coefficient for free cooling systems

typically ranges from 1.3 to 1.5, providing substantial savings in energy usage.

2.4. Evaporative Cooling Systems

Evaporative cooling systems utilize the natural process of water evaporation to cool the air. These systems are particularly effective in dry, hot climates and are designed to minimize energy consumption while ensuring optimal operating conditions for equipment [9]. The PUE coefficient for evaporative cooling systems typically ranges from 1.2 to 1.4, making them a highly efficient cooling solution.

2.5. In-Row Cooling

In-row cooling is a technology specifically designed to optimize heat management in high-density server environments. Cooling units are strategically placed between server racks, allowing for precise control of airflow and cooling in only the server rows that require it [10]. This method enhances energy efficiency by focusing cooling efforts where they are most needed. The PUE coefficient for in-row cooling systems typically ranges from 1.3 to 1.5, offering a significant reduction in energy consumption and operating costs.

A MODEL FOR REDUCING HEAT DENSITY AND IMPROVING ENERGY EFFICIENCY IN DATA CENTERS

Cooling systems play a crucial role in regulating the temperature, air quality, and humidity within the indoor environments of buildings, data centers (DCs), and various other facilities. Data centers, which house computational infrastructure including servers, storage systems, network equipment racks, and dedicated cooling units, generate substantial amounts of heat. Efficient management of this heat is essential to ensure the proper functioning and longevity of the equipment. In data centers, two critical factors — thermal density and energy efficiency — serve as key performance indicators for evaluating the effectiveness of cooling systems [4]. These metrics are vital for optimizing energy consumption and enhancing operational performance. Thermal Density refers to the heat load generated in a specific region of a data center, typically quantified in watts per square meter (W/m²) or watts per rack (W/rack). As thermal density increases, more powerful cooling solutions are required to prevent overheating and ensure system stability. In high-density environments, such as server rooms, managing thermal density is paramount to achieving operational reliability. The Power Usage Effectiveness (PUE) is a metric used to gauge the efficiency of energy consumption in cooling systems. It is calculated as the ratio of the total energy consumed by the entire data center to the energy consumed by IT equipment alone. A lower PUE value signifies greater energy efficiency, indicating that less energy is being used for cooling relative to the energy used by IT equipment. Thermal density is a widely recognized metric in data centers, server rooms, and high-density computing environments. It quantifies the amount of heat generated in a given space, such as within a server rack. This value is typically expressed in W/m², representing the energy consumption per square meter of the area where the equipment is housed. Managing thermal density effectively is critical for improving cooling efficiency and reducing the overall energy usage of data centers.

This metric is measured based on the energy consumption per square meter of the area where the equipment is located and is generally expressed in W/m². Thermal density is calculated using the following formula:

$$Q = P_{total}/A$$

Here:

- Q: Heat density (W/m²)
- Ptotal : Total amount of heat generated in the data center (Watts)
- A: Area of the data center (m²)

The total amount of heat P_{total} consists of the following components:

- Power consumption of servers ($P_{servers}$)
- Power consumption of the cooling systems ($P_{cooling}$)
- Power consumption of other equipment (P_{other})

$$P_{servers} = P_{servers} + P_{cooling} + P_{other}$$

In data centers, the heat density indicator — that is, the amount of heat generated in a specific area — is usually measured in BTU (British Thermal Unit) or W/m² (watts per square meter). This indicator mainly depends on the heat dissipation capability of servers, network devices, and other equipment. To determine heat management in data centers, the equation for total heat amount is used.

İn data centers, the total amount of heat is used to determine heat management. The equation is:

$$P_{total} = h \cdot A \cdot \Delta T$$

where:

- h: Heat transfer coefficient (W/m²)
- ΔT: The difference between the internal temperature of the center (Tin) and the external temperature (T_{out}), where $\Delta T = T_{in} - T_{out}$

With the help of this equation, it is possible to manage how the heat load (P_{total}) generated in the processing center is transferred to the environment (A) and how the temperature is balanced. If the heat load in a system is high, meaning a lot of heat is generated, and we want to remove this heat more efficiently, it can be done using one or more of the following methods:

- increase the surface area (A) – a larger surface area will dissipate heat better (e.g., large radiators),
- increase the temperature difference (ΔT) – if the cooling air or fluid is colder, heat will transfer faster,
- increase the heat transfer coefficient (h) – using better conductive materials or fluids will spread heat more quickly (e.g., using liquid cooling systems).

To ensure that heat density is evenly distributed between server racks, the PUE (Power Usage Effectiveness) indicator is used. PUE is the main metric for energy efficiency in data centers. It is used to evaluate how efficiently the cooling system and other auxiliary systems operate. Let's look at the model for optimizing the cooling requirements of IT equipment used in a processing center. The main goal of improving the PUE indicator in a data center is to reduce the energy consumption of cooling systems and other auxiliary systems while increasing the ratio of energy consumed by IT equipment. In other words, by maximizing the share of IT equipment's energy consumption within the overall energy consumption of the center, energy efficiency is improved. To achieve this:

- increase the efficiency of cooling systems,
- reduce the energy consumption of non-IT systems, such as lighting and ventilation.
- apply high efficiency equipment.

With this method, the PUE value becomes lower, meaning more efficient energy use. By optimizing PUE, the overall energy efficiency of the data center increases, operational costs decrease, and the environmental impact is minimized. To ensure optimal distribution of heat density, a model has been proposed that takes into account the server rack area, the total area, the heat density of each rack, and cooling capabilities. The goal is to ensure that the heat density is distributed as evenly as possible across the areas where the servers are located and to optimize cooling requirements.

Let's adopt the following notations:

- A – total area of the processing center (m^2),
- n – number of server racks,
- A_i – area of the i-th rack (m^2),
- A_{total} – area occupied by servers (m^2),
- $P_{IT,i}$ – energy consumption of IT equipment in the i-th rack (W),
- P_{IT} – total energy consumption (W),
- q_i – heat density in the i-th rack (W/m^2),
- q – heat density in the local area occupied by servers (W/m^2),
- q_{max} – maximum allowable heat density (W/m^2),
- $P_{cooling,i}$ – energy consumed for cooling in the i-th rack (W),
- $P_{cooling}$ – total energy consumed for cooling (W),
- q_{avg} – average heat density of the processing center (W/m^2).

The overall heat density (q) of the area where the servers are located is determined by the following formula:

$$q = P_{IT} / A_{total}$$

where,

$$P_{IT} = \sum_{i=1}^{n} P_{IT,i}\ ,\ A_{total} = \sum_{i=1}^{n} Ai$$

Here, a higher q value requires more energy for cooling.

- the heat density of each server is calculated using the following formula:

$$q_i = P_{IT,i} / A_i$$

- the total heat density of the servers is calculated by the formula:

$$q_{total} = \sum_{i=1}^{n} P_{IT,i} / A_{total}$$

- the average heat density of the processing center system is calculated by the formula:

$$q_{avg} = P_{cooling} / A$$

To maintain the energy balance of the cooling system, the cooling power of each local area is adjusted based on the ratio of the area's heat density to the average heat density:

$$P_{cooling,\ i} = q_i / q_{avg} * P_{cooling}$$

This formula is used to ensure the balanced and efficient distribution of cooling power across regions in the data center. That is:

- If the heat density in a region is higher than the average heat density ($q_i > q_{avg}$.), more cooling power is required in that region.
- If the heat density in a region is lower than the average ($q_i < q_{avg}$.), then less cooling power is needed there.

This approach ensures that cooling resources are directed where needed, meaning regions with higher heat density are cooled more effectively. Thus, both energy losses are reduced, and the cooling system operates more efficiently.

This model provides balanced cooling and optimizes the effective distribution of energy. To ensure the optimal distribution of heat density on the racks, the energy load of the racks must be assigned in such a way that the difference in heat density is minimized. In the data center, heat should be evenly distributed across different racks (zones). This can be expressed with the following condition:

$$\max(|q_i - q_{avg}|) \leq \epsilon$$

here, ϵ is the allowed heat density difference (for sensitive systems, ϵ is typically in the range of 10-25 W/m^2, and this value is determined by the engineer or system designer).

This formula ensures the most even distribution of the heat load across local areas. The optimal distribution of the heat density indicator across the server racks in cooling systems is crucial for increasing the efficiency of data processing centers (DPC) and reducing energy losses. Heat density expresses the heat load per unit area, and proper placement of servers can help optimize this indicator. If the condition $\max(|q_i - q_{avg}|) \leq \epsilon$ is not met, then the following actions must be taken:

- increase the cooling power in racks with high heat density,
- reduce the cooling power in racks with low heat density,
- regulate the heat density by properly distributing the servers across the racks.

The optimal distribution of the heat density indicator across the server racks in cooling systems is crucial for reducing energy losses. Heat density expresses the heat load per unit area, and proper placement of servers can help optimize this indicator. Minimizing the difference in heat density between racks ensures that the heat load is as evenly distributed as possible across the racks of the data center (DPC). This approach increases the efficiency of cooling systems and reduces energy consumption. The heat density difference between the racks should be minimized ($\min(\max(q_i) - \min(q_i))$.

Here:

- qi: Heat density of the i-th rack
- $\max(q_i)$: The highest heat density among the racks in the data center
- $\min(q_i)$: The lowest heat density among the racks in the data center

The difference in heat density between racks can be expressed as:

$$\Delta q = \min(\max(q_i) - \min(q_i))$$

To minimize this difference, it is necessary to regulate the energy consumption or areas of the shelves. Equalizing the heat densities of the shelves leads to more efficient operation of the cooling system and a reduction in energy costs. By calculating the heat densities of all the shelves, we can determine the highest (q_{max}) and lowest (q_{min}) values:

- Maximum heat density:

$$q_{max} = \max(q_1, q_2, \ldots, q_n)$$

- Minimum heat density:

$$q_{min} = \min(q_1, q_2, \ldots, q_n)$$

In this case, the difference in heat density is calculated as:

$$\Delta q = \min(q_{max} - q_{min})$$

To minimize the difference in heat density between shelves (i.e., to make the heat densities of the shelves as close to each other as possible), the following mechanisms can be used to regulate heat densities:

• regulating energy consumption: The energy consumption of shelves with higher heat densities can be reduced, or the energy consumption of shelves with lower heat densities can be increased,

• optimizing cooling power: By adjusting the power and distribution of the cooling system, more uniform distribution of heat densities can be ensured.

As a result of these measures, the difference in heat density between shelves will be minimized, and the efficiency of the cooling system will increase. Minimizing the difference means making the heat densities of the shelves closer to each other (equal), which will ensure optimal distribution of the cooling power within the system. Considering the above, let's look at the equal distribution of heat density among the servers located in the processing center:

If $q_i \geq q_{max}$:

• in this case, the i-th shelf has the highest heat load, and there is already IT equipment in this zone. To reduce the heat density, it would be appropriate to move some of the equipment from this shelf to other, less loaded shelves.

If $q_i \leq q_{max}$:

• in this case, the i-th shelf is considered underloaded, and the heat load is low. To balance the heat density in this zone, it is possible to place additional IT equipment.

The proposed model ensures the equal distribution of heat density across shelves, thereby increasing the energy efficiency of the cooling system. The difference should be minimized under the following conditions:

● the total energy consumption of the system's IT must be equal to the sum of the energy consumption of the IT equipment on the shelves:

$$P_{IT} = \sum_{i=1}^{n} P_{IT,i}$$

● the total area of the shelves must not exceed the total available area:

$$\sum_{i=1}^{n} Ai \leq A$$

● the heat density must not exceed a certain threshold:

$$q_i \leq q_{max}, \forall i$$

The proposed model ensures that some equipment from high heat load shelves is moved to less loaded shelves, thus providing a more balanced distribution of the heat load. This approach also reduces the load on the cooling system, increasing energy efficiency.

IV Conclusion

The article analyzes various cooling systems that can be utilized in data centers. It emphasizes the importance of addressing the challenges associated with cooling systems in order to reduce energy consumption and ensure the long-term reliability of server equipment. The significance of reducing heat density and enhancing energy efficiency within data centers is highlighted, and a model is proposed to improve these aspects. The suggested model involves relocating certain IT equipment from high heat-load shelves to those with lower heat load, facilitating a more balanced heat distribution across the data center. This strategy not only optimizes the heat management process but also reduces the strain on cooling systems, thereby contributing to better overall energy efficiency.

References

[1] Nadjahi, H. Louahlia, S. Lemasson, (2018), "A review of thermal management and innovative cooling strategies for data center", Sustain. Comput. Informatics Syst. 19, pp. 14–28.

[2] A. Nada, M.A. Said, M.A. Rady, (2016), "Numerical investigation and parametric study for thermal and energy management enhancements in data centers buildings", Appl. Therm. Eng., 98, pp. 110–128.

[3] Data Center Cooling Systems: Challenges and Solutions, https://www.sangfor.com/blog/cloud-and-infrastructure/data-center-cooling-systems-challengesand-solutions.

[4] J. Ni, X. Bai, (2017), "A review of air conditioning energy performance in data centers", Renewable and Sustainable Energy Reviews 67, pp. 625–640.

[5] D.D. Gerijih, K.Z. Hazrati, R. M. Qibtiah, N. Baharudin, N. Razali, K. F Ali, (2022), "A Comprehensive Review of a Data Centre for a Cooling System", International Journal of Mechanical Engineering, vol.7, no.1, pp.1–7.

[6] R. Kong, H. Zhang, M. Tang, H. Zou, C. Tian, T. Ding, (2024) "Enhancing data center cooling efficiency and ability: A comprehensive review of direct liquid cooling technologies", Energy 308, 32846, pp. 1–18.

[7] N. A. Pambudi, A. Sarifudin, R. A. Firdaus, D. K. Ulfa, I. M. Gandidi, R. Romadhon, (2022), "The immersion cooling technology: Current and future development in energy saving", Alexandria Engineering Journal, 61, pp. 9509–9527

[8] A.H. Khalaj, S.K. Halgamuge, (2017), "A review on efficient thermal management of airand liquid-cooled data centers: From chip to the cooling system", Appl. Energy 205, pp. 1165–1188.

[9] C. Nadjahi, H. Louahlia, S. Lemasson, (2018), "A review of thermal management and innovative cooling strategies for data center", Sustainable Computing: Informatics and Systems, 19, pp. 14–28

[10] The Evolution Of Data Center Cooling Technologies: Innovations For Sustainability. https://powerandcoolingme.com/the-evolution-of-data-center-cooling-technologies-innovations-for-sustainability/

Investigation of the Boundary Functional of the Random Walk Process with a Special Barrier

Elshan Ibayev
Laboratory of applied probabilistic statistical methods
Institute of Control Systems
Azerbaijan Technical University
Baku, Azerbaijan
0000-0002-5455-8865

***Abstract*— In this paper we consider a complex stochastic process that incorporates random walk behavior with variable durations in each state, including negative drift and positive jumps. We derive an integral equation for the Laplace transform of the conditional distribution of the boundary functional. In this work, we define the residence time of a system using a mathematical model based on absolutely continuous distribution with varying parameters. Each parameter could represent a different characteristic of the system. The main objective of the paper is to reduce the integral equation for the Laplace transform to a fractional order differential equation with constant coefficients.**

Keywords— Inverse Laplace transform, semi-Markov process, Riemann-Liouville fractional derivative.

INTRODUCTION

A semi-Markov processes (SMPs) are investigated in different directions. There are many research papers and books about theoretical aspects and applications of SMPs. In recent years, a semi-Markov random walk process (SMRWP) with one or two barriers are being used to solve a number of very interesting problems in a variety of fields, including queueing theory, reliability analysis, inventory control, insurance, and more. In a Markov chain, the memoryless property of the exponential distribution is a fundamental assumption. This property implies that the future evolution of the system depends only on its current state, and the time until the next transition is exponentially distributed, independent of the past history. However, many real-world systems exhibit behaviors where the memoryless property may not hold, and the waiting times between transitions may follow distributions other than the exponential. But in particular case, the exact formulas for the moments of the SMRWP can be obtained. In the works of [7, 8, 9], Laplace transform (LT) of the distribution boundary functional of the SMRWPs have found by using the ordinary differential equations. In the works [3, 6] the random variable was introduced . This random variable is called boundary functional of the process. Fractional differential equation in the Weyl sense for LT of the boundary functional of a SMRWP with screen was obtained. LTs are often used to find moments and other statistical characteristics of random processes. This is a powerful tool in understanding the statistical properties of the SMRWPs. We recall that, the study in [2] asymptotic formulas were found for the mathematical expectation and variance of the boundary functional of the SMP. But in this paper, a mathematical model is considered, for the SMP with specific characteristics, including negative drift and positive jumps, and applicable to a general class of probability distributions. The fractional order differential equation with constant coefficients has been derived for the Laplace transform within the class of absolutely continuous distributions. Using fractional differential equation with fractional Riemann-Liouville derivative we can find the Laplace transform of the conditional distribution of the boundary functional.

PROBLEM STATEMENT

Let's assume that sequences of independent and identically distributed pairs of random variables $\{\xi_k, \zeta_k\}_{k=1}^{\infty}$,be given on the any probability space (Ω, F, P), where ξ_k and ζ_k take only positive values. Suppose that ξ_1 and ζ_1 are independent random variables.

Let us denote the cumulative distribution functions of the random variables ξ_1 and ζ_1

$$F(t) = P\{\xi_1 < t\} \ and \quad G(t) = P\{\zeta_1 < t\}.\ t \in R^+$$

respectively.

Using this assumption the stochastic process $X_1(t)$ is constructed as follows

$$X_1(t) = z - t + \sum_{i=0}^{k-1} \zeta_i, \ if \quad \sum_{i=0}^{k-1} \xi_i \le t < \sum_{i=0}^{k} \xi_i,$$

$$k = \underline{1, \infty},$$

where $\xi_0 = \zeta_0 = 0$. This stochastic model is called "the SMRWP with negative drift, positive jumps". Let this model is delayed by a barrier zero (see, [1])

$$X(t) = X_1(t) - inf_{0 \le s \le t}\{0, X_1(s)\}.$$

Let the random variable τ_0 defined by

$$\tau_0 = inf\{t:\ X(t) = 0\}.$$

We set $\tau_0 = \infty$ if $X(t) > 0$ for every t.τ_0 is called boundary functional of the process $X(t)$. The aim of this study is to find the LT of the conditional distribution of the τ_0, which represents the time of the first crossing of the process into the delaying barrier at the zero level. The LT of the conditional distribution of the τ_0 by

$$L(\lambda|z) = E\left[e^{-\lambda\tau_0}\middle|X(0) = z\right], \quad \lambda > 0, \qquad z \ge 0,$$

where λ is the LT parameter.

Let us denote the conditional distribution of the τ_0, as

$$N(t|z) = P[\tau_0 > t|X(0) = z],$$

Similarly, we denote the LT of the conditional distribution as

$$\tilde{N}(\lambda|z) = \int_0^\infty e^{-\lambda t} N(t|z)dt.$$

We obtain from the last formula that

$$\tilde{N}(\lambda|z) = \frac{1-L(\lambda|z)}{\lambda}.$$

Obviously,

$$L(\lambda|z) = 1 - \lambda\tilde{N}(\lambda|z).$$

OBTAINING THE FRACTIONAL ORDER DIFFERENTIAL EQUATION FOR $\tilde{N}(\lambda|z)$

Using total probability formula, we can get

$$\tilde{N}(\lambda \mid z) = \int_0^z e^{-\lambda t} P\{\xi_1 > t\}\, dt + \quad (1)$$

$$+ \int_z^\infty \tilde{N}(\lambda \mid y) \int_0^z e^{-\lambda t}\, d_t P\{\xi_1 < t\}\, d_y P\{\zeta_1 < y - z + t\} + \int_0^z \tilde{N}(\lambda \mid y) \int_{z-y}^z e^{-\lambda t}\, d_t P\{\xi_1 < t\}\, d_y P\{\zeta_1 < y - z + t\}.$$

Suppose that the random variable ξ_1 has a probability density function $p_{\xi_1}(u)$, $u > 0$ and the random variable ζ_1 has a probability density function $p_{\zeta_1}(u)$, $u > 0$. Then, equation (1) has the form

$$\tilde{N}(\lambda \mid z) = \frac{1}{\lambda}\left[1 - e^{-\lambda z}\right] + \frac{1}{\lambda} e^{-\lambda z} P\{\xi_1 < z\} - \quad (2)$$

$$- \frac{1}{\lambda} \int_0^z e^{-\lambda t} p_{\xi_1}(t) d\, t +$$

$$+ \int_z^\infty \tilde{N}(\lambda \mid y) \int_0^z e^{-\lambda t}\, p_{\xi_1}(t) p_{\zeta_1}(y - z + t)\, dtdy +$$

$$+ \int_0^z \tilde{N}(\lambda \mid y) \int_{z-y}^z e^{-\lambda t}\, p_{\xi_1}(t) p_{\zeta_1}(y - z + t)\, dtdy.$$

Suppose that distribution functions of ξ_1 and ζ_1 are known. Let density functions of the distributions for ξ_1 and ζ_1 as follows

$$p_{\xi_1}(u) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot u^{\alpha-1} e^{-\beta u}, & u > 0, \\ 0, & u \le 0, \end{cases}$$

$$\rho_{\zeta_1}(u) = \begin{cases} \mu\, e^{-\mu u}, & u \ge 0, \\ 0, & u < 0, \end{cases}$$

where

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du$$

is Euler's gamma function.

In the class of these probability distributions the integral equation (2) can be written as follows

$$\tilde{N}(\lambda \mid z) = \frac{1}{\lambda}\left[1 - e^{-\lambda z}\right] + \frac{\beta^\alpha e^{-\lambda z}}{\lambda\Gamma(\alpha)} \int_0^z e^{-\beta y} y^{\alpha-1} dy -$$

$$- \frac{\beta^\alpha}{\lambda\Gamma(\alpha)} \int_0^z e^{-(\lambda+\beta)t} t^{\alpha-1} d\, t +$$

$$+ \frac{\mu\beta^\alpha e^{\mu z}}{\Gamma(\alpha)} \int_z^\infty e^{-\mu y} \tilde{N}(\lambda \mid y) \int_0^z e^{-(\mu+\lambda+\beta)\, t}\, t^{\alpha-1}\, dtdy +$$

$$+ \frac{\mu\beta^\alpha e^{\mu z}}{\Gamma(\alpha)} \int_0^z e^{-\mu y} \tilde{N}(\lambda \mid y) \int_{z-y}^z e^{-(\mu+\lambda+\beta)\, t}\, t^{\alpha-1}\, dtdy. \quad (3)$$

After some transformations, the (3) integral equation takes the form

$$e^{(\lambda+\beta)z} \tilde{N}'(\lambda \mid z) - \mu e^{(\lambda+\beta)z} \tilde{N}(\lambda \mid z) =$$

$$= \frac{1}{\lambda}\left[(\mu + \lambda) e^{\beta z} - \mu e^{(\lambda+\beta)z}\right] -$$

$$- \frac{\beta^\alpha(\mu + \lambda)}{\lambda\Gamma(\alpha)} z^\alpha \sum_{n=0}^\infty \frac{\beta^n z^n}{\alpha(\alpha+1)\ldots(\alpha+n)} +$$

$$+ \frac{\mu\beta^\alpha}{\lambda\Gamma(\alpha)} z^\alpha \sum_{n=0}^\infty \frac{(\lambda+\beta)^n z^n}{\alpha(\alpha+1)\ldots(\alpha+n)} +$$

$$+ \frac{\mu\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} \int_0^\infty e^{-\mu y} \tilde{N}(\lambda \mid y) d\, y -$$

$$- \frac{\mu\beta^\alpha}{\Gamma(\alpha)} \int_0^z e^{(\lambda+\beta)y} \tilde{N}(\lambda \mid y)(z - y)^{\alpha-1} dy. \quad (4)$$

We denote

$$K(\lambda|z) = e^{(\lambda+\beta)z} \tilde{N}(\lambda \mid z) \quad (5)$$

Then equation (4) can be rewritten as follows

$$K'(\lambda \mid z) - (\mu + \lambda + \beta) K(\lambda \mid z) =$$

$$= \frac{1}{\lambda}\left[(\mu + \lambda) e^{\beta z} - \mu e^{(\lambda+\beta)z}\right] -$$

$$- \frac{\beta^\alpha(\mu + \lambda)}{\lambda\Gamma(\alpha)} z^\alpha \sum_{n=0}^\infty \frac{\beta^n z^n}{\alpha(\alpha+1)\ldots(\alpha+n)} +$$

$$+ \frac{\mu\beta^\alpha}{\lambda\Gamma(\alpha)} z^\alpha \sum_{n=0}^\infty \frac{(\lambda+\beta)^n z^n}{\alpha(\alpha+1)\ldots(\alpha+n)} +$$

$$+ \frac{\mu\beta^\alpha}{\Gamma(\alpha)} z^{\alpha-1} \int_0^\infty e^{-\mu y} \tilde{N}(\lambda \mid y) d\, y -$$

$$- \frac{\mu\beta^\alpha}{\Gamma(\alpha)} \int_0^z K(\lambda \mid y)(z - y)^{\alpha-1} dy. \quad (6)$$

It is known that the Riemann-Liouville integral can be expressed by (see, [4] and [5])

$$D_z^{-\alpha}(K(\lambda\,|z)) = \frac{1}{\Gamma(\alpha)}\int_0^z K(\lambda\,|y)(z-y)^{\alpha-1}dy, \quad 0 < \alpha \le 1.$$

where $\alpha(\alpha > 0)$ is the fractional order, Γ is the gamma function.

Using this equality, we can write the equation (6) in the following form

$$K'(\lambda \mid z) - (\mu+\lambda+\beta)K(\lambda \mid z) =$$

$$= \frac{1}{\lambda}\left[(\mu+\lambda)e^{\beta z} - \mu e^{(\lambda+\beta)z}\right] -$$

$$-\frac{\beta^\alpha(\mu+\lambda)}{\lambda\Gamma(\alpha)}\sum_{n=0}^{\infty}\frac{\beta^n z^{n+\alpha}}{\alpha(\alpha+1)\dots(\alpha+n)} +$$

$$+\frac{\mu\beta^\alpha}{\lambda\Gamma(\alpha)}\sum_{n=0}^{\infty}\frac{(\lambda+\beta)^n z^{n+\alpha}}{\alpha(\alpha+1)\dots(\alpha+n)} +$$

$$+\frac{\mu\beta^\alpha}{\Gamma(\alpha)}z^{\alpha-1}\int_0^\infty e^{-\mu y}\widetilde{N}(\lambda \mid y)d\,y - \mu\beta^\alpha D_z^{-\alpha}K(\lambda \mid z). \quad (7)$$

Using Riemann-Liouville fractional derivative of order α, (see, [4] and [5]), we can write the equation (7) in the following form

$$D_z^{\alpha+1}K(\lambda \mid z) - (\mu+\lambda+\beta)D_z^\alpha K(\lambda \mid z) +$$

$$+\mu\beta^\alpha K(\lambda \mid z) = \frac{1}{\lambda}(\mu+\lambda)z^{-\alpha}E_{1,1-\alpha}(\beta z) -$$

$$-\frac{1}{\lambda}\mu z^{-\alpha}E_{1,1-\alpha}((\lambda+\beta)z) - \frac{\beta^\alpha(\mu+\lambda)}{\lambda}E_{1,1}(\beta z) +$$

$$+\frac{\mu\beta^\alpha}{\lambda}E_{1,1}\big((\lambda+\beta)z\big)$$

where $E_{\alpha,\beta}(z)$ is the Mittag-Leffler function.

CONCLUSION

We investigated a specific type of the SMRWP with delaying barrier, specifically focus on the LT of the conditional distribution of the random variable within the class of absolutely continuous distributions. Furthermore, we constructed an integral equation for the LT and obtained a fractional order differential equation, in terms of fractional derivatives in the Riemann-Liouville sense. Finally, we can find a solution to this fractional order differential equation..

REFERENCES

[1] A. Borovkov, Probability theory. Amsterdam, Gordon and Breach Science Publishers, 1998.

[2] A.S. Shukurov. Asymptotic results for a semi-Markov process describing the behavior of some stochastic systems Informatics and Control Problems 41 Issue 1, 2021, pp. 45-50.

[3] E.A. Ibayev, Obtaining the fractional order differential equation for the Laplace transform of the distribution of the boundary functional. Journal of Baku Engineering University Mathematics And Computer Science. 2022, Volume 6, Number 1, pp. 43-48.

[4] F. Mainardi Fractional calculus and waves in linear viscoelasticity: an introduction to mathematical models. World Scientific, 2010.

[5] K.S. Miller and B. Ross, An Introduction to the fractional calculus and fractional differential equations. Wiley, New York, 1993.

[6] R.A. Bandaliyev, E.A. Ibayev and K.K. Omarova. Investigation of fractional order differential equation for boundary functional of a semi-Markov random walk process with negative drift and positive jumps. Chaos, Solitons & Fractals, 152, 111394, 2021 https://doi.org/10.1016/j.chaos.2021.111394

[7] S.Maden, B.G. Shamilova the Laplace transform of a boundary functional of the semi-Markovian random walk process with two delaying barriers. Ordu Univ. J. Sci. Tech., Vol:6, No:1, 2016, pp. 43-53.

[8] T. H. Nasirova and U.Y. Kerimova, Definition of Laplace transform of the first passage of zero level of the semi-Markov random process with positive tendency and negative jump, Appl. Math. 2(7), 2011, pp. 908-911.

[9] T.I. Nasirova and K.K. Omarova, Distribution of the lower boundary functional of the step process of semi-Markov random walk with delaying screen at zero. Ukr. Math.J. 59(7), 2007, pp. 1010-1018.

Section 7: Information and Control Systems

Evaluation of Term Creation Processes within the Framework of the National Terminology Information System

Rasim Alguliyev
Institute of Information Technology
Baku, Azerbaijan
r.alguliev@gmail.com

Afruz Gurbanova
Institute of Information Technology
Baku, Azerbaijan
afruz1961@gmail.com

***Abstract*—The research work examines the issues of analysis and evaluation of terminological processes within the framework of the National Terminology Information System. Analysis, evaluation and monitoring of terminological processes are the main elements of effective terminology management in any terminological information systems. Comparison of these processes between different fields of knowledge allows for a better understanding of the characteristics of each field and the development of more efficient approaches to the organization of terminological work.**

An indicator was proposed to characterize the dynamics of the term creation process in various fields of knowledge. With the help of this indicator, it is possible to determine the progress or decline of the term creation process in fields of knowledge. The intersections of two or more fields of knowledge based on terms were considered. By making comparisons between different fields of knowledge, it is possible to identify similarities and differences between these fields, that is, common and unique terms, concepts, and relationships between them in different fields. Experiments were conducted and the results were analyzed to implement the proposed statistical indicator. The proposed indicator, based on structured metadata and statistical analysis, significantly improves the efficiency and transparency of terminology management in the National Terminology Information System.

Keywords—terminological processes, national terminological information system, dynamic processes, term creation

INTRODUCTION

Terms are the most important element of scientific and practical activity, and their systematization is closely connected with scientific and technological progress, language features, historical and cultural factors and processes of international integration [1].

The main issue of terminological activity is the correct construction of the term creation process [2]. Term creation is the process of creating, systematizing, standardizing and using terms in various fields of knowledge. It includes the following main aspects:

- Creation of new terms: Borrowing the equivalent of a term from other languages, forming new terms based on existing words, combining two or more words into one term, etc.
- Changing the meaning of a term. The meaning of a term can narrow, expand, or change completely over time. This is due to the development of science and technology, the emergence of new knowledge and technologies.
- Standardization of terms: Standardization of terms is carried out to ensure uniform meaning and form [3].
- Systematization of terms: Organizing terms in the form of dictionaries, glossaries, or terminology databases. This makes it easier to find and use terms.
- Obsolescence of terms: With the development of science and technology, some terms become obsolete and fall out of use. This is due to the emergence of new, more precise and modern terms [4].

Terminological processes are dynamic changes in terminology, i.e. in the system of specific words and expressions used in a certain field of knowledge or activity. These processes are associated with the development of science, technology, culture and society as a whole and reflect the emergence of new concepts, clarification of old terms, as well as changes in their meanings [5]. The main terminological processes are:

- Terminologization. It is the process of creating special words used to express concepts, objects and phenomena related to a particular field of knowledge in science, technology and other fields [6].
- Determinologization. This is a process of exchange between specialized and general vocabulary, leading to the term going beyond the framework of the terminological system and turning it into a common word [7].
- Semantic changes of terms. This is a change in the lexical meaning of a term.
- Harmonization of terms. Harmonization of terminology implies interlingual regulation of terms at the level of comparable terminological systems [8].
- Synonymy of terms. This is a special type of semantic connection, in which different terminological units can express the same concept, differing only in a certain style and emotional coloring.

In [9] a conceptual model of the National Terminology Information System (NTIS) is presented, its functional subsystems are analyzed and the capabilities of the system are defined. NTIS is formed as an integral part of the Azerbaijani language ecosystem on the digital state platform and provides users with electronic terminology services. The integration of terminological dictionaries into a single information system increases the efficiency of terminological research.

Let us consider the analysis and evaluation of terminological processes within the framework of the NTİS.

RELEATED WORK

The presented analysis of relevant works focuses on various approaches and methods for effective terminology management in terminology information systems, with special attention paid to dynamic processes in terminology systems, interdisciplinary connections and assessment of research directions.

In [10], the processes of terminologization and determinologization are analyzed, dynamic processes in the National Terminology Information System are formulated on the basis of mathematical laws, and a method based on a population model is proposed for the general analysis of the system.

In [11], a method based on machine learning was proposed for the analysis and evaluation of dynamic processes in the NTIS. A method based on the k-means clustering algorithm and the Naive Bayes classification model was proposed to automate the distribution of new terms into appropriate disciplines.

In [12] the theoretical and practical aspects of interdisciplinary connections are scientifically substantiated. Researchers present scientific results on interdisciplinary connections in primary education and ways of their practical implementation. The emergence of interdisciplinary connections based on the integration of sciences is defined as a scientific pattern, and the need to create educational and methodological materials that take into account interdisciplinary connections is noted.

Interdisciplinary communication is the result of dynamics on two levels. First, research questions are addressed using data from different fields. Second, the results of interdisciplinary research feed back into different research fields. This may contribute to the further development of fields or the creation of a new interdisciplinary field. This approach, proposed in [13], is a useful method for studying the dynamics of interdisciplinary research and, in general, the subject area.

In [14], the similarity of research areas is assessed using the Jaccard index. The dynamics of the Jaccard index distribution does not confirm the hypothesis of an increase in the share of interdisciplinary research over time. It turns out that the most cooperative research area for the time intervals 2000-2009 and 2010-2019 is mechanical engineering. This research area has the highest Jaccard indices with all other research areas. It was found that the most closely related triad of research areas for both time intervals are commercial services, marketing and tourism.

In [15] analyses were conducted based on a OLAP model in the NTISystem for various parameters (required information (space, time and area of knowledge) characterizing terms introduced into the language over certain periods). As a result of the analyses, it is possible to determine trends occurring in the national terminology environment, the ecology and safety of the language, what area of activity is developing in the country, etc., and make decisions based on the information obtained.

Thus, the analyzed works represent different methodologies, such as mathematical modeling, machine learning, OLAP, etc. All these works aim to understand, evaluate and manage the complex and dynamic landscape of terminology, especially in the context of evolving knowledge and interdisciplinary interactions.

METHODOLOGY

The main objective of the research work is to analyze and evaluate the term creation processes in the National Terminology Information System, as well as to conduct comparisons between different fields of knowledge. For this purpose, an indicator has been developed. The main goal and task of the indicator is to determine the dynamics of the development of term creation processes in different fields of knowledge, the relationships between fields of knowledge, the definition of common and unique terms and concepts in these fields, and to make decisions based on the results obtained.

In general, there are statistical, semantic and linguistic types of indicators. The indicator proposed in the research work is based on statistical analysis and its main task is to determine the number of common and unique terms, the frequency of use of terms in different fields and the measures of terminological similarity and to make decisions based on the results obtained. With the application of the indicator, it is possible to carry out the analysis of scientific and technical information, the development of educational programs, the creation of multilingual terminological resources, and support decision-making in various fields. As a result, with the help of the proposed indicator, scientifically based, strategic and applied decisions are made aimed at the development, adaptation and application of terms in various fields of activity.

EVALUATION OF TERMINOLOGICAL PROCESSES IN THE NATIONAL TERMINOLOGICAL INFORMATION SYSTEM

Let us provide each term (accepted for discussion and approved) in the Terminological Register (TR) of the national terminology information system with accompanying metadata. This metadata includes the name and address of the organization and authors that submitted the term, the time the term was submitted for discussion, the time the term was approved, an indication of the scientific field to which the term belongs, information about the language from which the term is derived, the name (spelling) of the term in that language, the etymology of the term, the individual identification code given to the approved term, etc. Table 1 provides an example of a detailed description of metadata.

TABLE 1. DETAILED DESCRIPTION OF METADATA

Metadata	**Description**
Term	The word or phrase itself
Synonyms	Other words or phrases with the same meaning
Antonyms	Words or phrases with the opposite meaning
Definition	A clear and concise explanation of the meaning of the term
Usage Examples	Examples of the use of the term (in context)
Field of Knowledge	The field of knowledge to which the term belongs
Author	The person or organization that proposed the term
Date of Creation	The date the term was created
Date of Approval	The date the term was approved by the expert committee
License	The license under which the term was distributed

By analyzing the necessary information (space, time, field of knowledge, etc.) characterizing the terms that have entered our language over certain periods, it is possible to identify

trends, threats, etc., occurring in the national terminological environment, and make the necessary decisions.

Formally, we denote the metadata related to the i-th term from the s-th knowledge area by the following vector:

$$T_i^{(s)} = \langle t_{ij}^{(s)} | j = \underline{1, m_s} \rangle, s = \underline{1, S}, i = \underline{1, n_s} \quad (1)$$

Here $t_{ij}(s)$ is the j-th attribute (metadata) of the i-th term belonging to the s-th knowledge area. Thus, TR consists of terms t_i and their accompanying metadata bases:

$$TR = \bigcup_{s=1}^{S} \bigcup_{i=1}^{n_s} (t_i, T_i^{(s)}) = \bigcup_{s=1}^{S} TR_s \quad (2)$$

$$TR_s = \bigcup_{i=1}^{n_s} (t_i^{(s)}, T_i^{(s)}), s = \underline{1, S} \quad (3)$$

Here the variable ns is the number of terms related to the s-th area of knowledge, and ms is the amount of metadata characterizing these terms.

Let us introduce the indicator ns(t) for the purposes of analysis, evaluation, monitoring of terminological processes in the national terminological information system, conducting comparisons between different areas of knowledge, etc. Here $n_s(t)$ is the number of terms in the Terminological Register for the *t*-th period in the *s*-th area of knowledge. Naturally, the total number of terms in the Terminological Register for the *t*-th period $N_s(t)$ will be as follows:

$$N_s(t) = \sum_{s=1}^{S} \quad n_s(t) \quad (4)$$

To characterize the dynamics of the development of the term formation process in the *s*-th sphere, we introduce the indicator $\Delta n_s(t)$:

$$\Delta n_s(t) = n_s(t) - n_s(t-1), t=1,2,\ldots,m \quad (5)$$

With this indicator, we can track the growth or decline of terminology creation in each area over time, allowing us to determine the progress (decline) of the terminology creation process by knowledge area. It also allows us to identify active and stagnant areas in terms of terminology development. This also allows us to identify active and stagnant areas in terms of term development, thereby identifying trends or threats. It also includes term clarification.

A comprehensive analysis of several such indicators allows us to gain a more complete understanding of the dynamics of terminology development in a specific field of knowledge: the frequency of emergence of new terms; the frequency of use of terms, etc. By making comparisons between different fields of knowledge, it is possible to identify similarities and differences between those fields, that is, common and unique terms, concepts and relationships between them in different fields. Let's look at the dynamics of the intersection of two fields ($U_s(t)$ and $U_{s'}(t), s \neq s'$):

$$|U_s(t) \cap U_{s'}(t)| = \Delta U_{ss'}(t) \quad (6)$$

if the condition $\Delta U_{ss'}(t) > \Delta U_{ss'}(t-1)$ is satisfied, the integration of the 2 fields has occurred over time;

If the condition $\Delta U_{ss'}(t) < \Delta U_{ss'}(t-1)$ is satisfied, the differentiation of the 2 fields has occurred over time.

Thus, based on a comparison of the terminology of the two subject areas over time, the following can be determined:

- identifying common terms in both domains;
- identifying unique terms in each domain;
- identifying the frequency of use of common terms in each domain;
- identifying semantic relationships between common terms in different domains;
- identifying terms with different meanings in each domain.

This information can be used to analyze the relationship between two fields, develop interdisciplinary educational programs, or create a multilingual dictionary. Such comparisons can also be made between 3 fields, 4 fields, etc. This is one of the main functions performed by the National Terminological Information System. The proposed indicator, based on structured metadata and statistical analysis, significantly enhances the efficiency and transparency of managing terminology within the National Terminology Information System.

The experiment was conducted using data on the number of terms accepted into the National Terminology Information System (NTIS) in three knowledge areas in 2022–2024:

- s = 1: Computer science
- s = 2: Biomedicine
- s = 3: Ecology

TABLE 2. NUMBER OF TERMS ACCEPTED INTO NTIS IN 2022–2024

year (t)	Computer science ($n_1(t)$)	Biomedicine ($n_2(t)$)	Ecology ($n_3(t)$)
2022	120	80	95
2023	150	105	110
2024	170	115	130

The growth rate of terms in each area for 2023-2024 is calculated based on formula (5) as follows:

2023:

$$\Delta n_1(2023) = 150 - 120 = +\mathbf{30}$$

$$\Delta n_2(2023) = 105 - 80 = +\mathbf{25}$$

$$\Delta n_3(2023) = 110 - 95 = +\mathbf{15}$$

2024:

$$\Delta n_1(2024) = 170 - 150 = +\mathbf{20}$$

$$\Delta n_2(2024) = 115 - 105 = +\mathbf{10}$$

$$\Delta n_3(2024) = 130 - 110 = +\mathbf{20}$$

As can be seen from the calculations, computer science developed most actively in 2023, but the growth rate slowed down in 2024. In 2024, the growth rate of ecology accelerated and equaled that of computer science. Biomedicine showed steady but slower growth.

Let us consider a comparison of terms between knowledge areas, the purpose of which is to identify common and unique terms, as well as to identify the processes of

integration or differentiation occurring between knowledge areas:For the comparison of terms between two subject areas, five terms from both areas were selected.

Let:

In 2024, the general set of computer science terms is: $U_1(2024)=\{data, algorithm, network, intelligence, model\}$

In 2024, the general set of ecology terms is: $U_3(2024)=\{model, environment, biodiversity, sustainability\}$

The intersection of two subject areas will be as follows:

$U_1 \cap U_3 = \{model\}$

$|U_1 \cap U_3| = 1$

In 2023 there were 0 such terms.

$\Delta U_{13}(2024) = 1 - 0 = +1$

The experiments conducted show that there is a tendency towards the integration of informatics and ecology, since the term “model”, previously characteristic of informatics, has become an interdisciplinary term in ecology.

Let's look at the frequency of use of terms in computer science and ecology. Suppose the database records the use of the term "model":

- In computer science: 250 occurrences
- In ecology: 35 occurrences

Thus, the proposed indicator allows for formalization and quantitative assessment:

- intensity of creation of new terms;
- integration and differentiation between areas of knowledge,
- the need for lexical and semantic standardization,
- strategic directions for the development of science, education and translation.

CONCLUSION

The meaning of a term can narrow, expand or change completely over time. This is due to the development of science and technology, the emergence of new knowledge and technologies.

Terminological processes are an integral part of the development of language and culture. They reflect the dynamics of knowledge and technology and ensure effective communication in various fields of activity.

Analysis, evaluation and monitoring of terminological processes are key elements of effective terminology management in any terminology information systems. Comparison of these processes between different fields of knowledge allows for a better understanding of the characteristics of each field and the development of more effective approaches to the organization of terminological work. This is a continuous process that requires attention to user needs, technology development and the dynamics of subject areas.

The indicator proposed for comparing different fields of knowledge in the national terminology information system is a tool that allows for the analysis and comparison of the terminology of these fields.

The study showed that, based on the presented indicator, it is possible to identify unique terms, concepts and relationships between them in various fields, assess the intersection and degree of interaction between knowledge fields based on their terminology, and provide information for the analysis and synthesis of knowledge from various fields. Thus, using the proposed statistical indicator, it is possible to make decisions in several directions based on the results of the analysis of terminological processes:

- Determining which terms should be standardized, unified or updated.
- Formation and development of national terminology databases.
- Analysis of the terminological composition of scientific and technical documents and identification of terminological gaps.
- Development and adaptation of educational programs taking into account the terminological content of various disciplines.
- Creation of methodological materials for teaching terminology in various fields.
- Creation of multilingual terminology resources based on the analysis of similarities and differences of terms.
- Support for the harmonization of terms between different fields of knowledge, which is important for interdisciplinary research and communication, etc.

REFERENCES

[1] Khudaybergenova D. M., Rakhimova F. S. “Development and Standardization of Terminology Across Scientific Disciplines”, European International Journal of Multidisciplinary Research and Management Studies, 2025. v. 5, No 4, pp. 41-43.

[2] Zhukova Yu.V., Antonova E.I. “Borrowing as an active mechanism for the formation of scientific and technical terminology”, The main issues of linguistics, linguodidactics and intercultural communication, Collection of works of the XI international scientific and practical conference, Publishing house "Astrakhan University", 2020, pp. 39-42. DOI: 10.21672/978-5-9926-1237-0-039-042.

[3] Galinski C., Weissinger R. “Terminology standardization and translation standards”, Magyar Terminológia (3), 2010, v. 11. pp. 8-20.

[4] Jakob Halskov. “Probing the Properties of Determinologization - the DiaSketch”, 2005, Conference: Nordic Graduate School of Language Technology (NGSLT) Seminar, 2005, Stockholm, Sweden, pp. 39-63.

[5] Faber, P., León Araúz, P., & Prieto Velasco, J. A. “Semantic relations, dynamicity, and terminological knowledge bases”. Current Issues in Language Studies, 2009, 1(1), pp. 1-23.

[6] Ohaike, N. Terminological Creation: A Practical Approach. Journal of Nigeria Languages’ Studies (NILAS), 2015, 2(3), 77-82.

[7] Meyer, Ingrid, Kristen Mackintosh. "When terms move into our everyday lives: An overview of determinologization" In: Terminology, 2000, vol. 6:1, pp. 111-138, John Benjamins.

[8] ISO 860:2007, Terminology work – Harmonization of concepts and terms. URL: https://www.iso.org/obp/ui/#iso:std:iso:860:ed-3:v1:en.

[9] Alguliyev R., Gurbanova A. “The Conceptual Foundations of National Terminological Information System”, International Journal of Education and Management Engineering, MECS Press, 2018, 4, p. 19-30.

[10] Aliguliyev R.M., Gurbanova A.M. “Mathematical modeling of dynamic processes within the framework of the terminological information system”, Information Technologies, 2022, no. 5, pp. 226-233.

[11] Gurbanova A., Alakbarova I., Mammadzade S. “Development of a method for intellectual analysis of terminological dynamic processes”, Proceedings of the 2nd Information Technologies and Its Applications Conference, Part 3, Baku, Azerbaijan: 23-25 april 2024, pp. 1–15.
[12] Saurykova Z. M., Ybyraimzhanov K., Mailybaeva G. “Implementation of interdisciplinary relationships in education on the basis of science integration”, Opción: Revista de Ciencias Humanas y Sociales, 2018, № 85, pp. 353-385. DOI: 10.26577/JES.2023.v74.i1.018
[13] Trnova E. “Development of interdisciplinary instruction using inquiry-based science education”, International Conference on Education and New Developments, 2019, pp. 232-236. DOI: 10.36315/2019v1end049
[14] Shtovba S., Petrychko M. “Jaccard index-based assessing the similarity of research fields in dimensions”, 1st International Workshop on Digital Content & Smart Multimedia, 2020, pp. 1-12.
[15] Rasim M. Alguliyev, Gulnara Ch. Nabibayova, Afruz M. Gurbanova. “Development of a Decision Support System with the use of OLAP-Technologies in the National Terminological Information Environment”, International Journal of Modern Education and Computer Science(IJMECS), 2019, Vol.11, No.6, pp. 43-52.

Development of IELTS Essay Evaluation System with Deep Learning

Fuad Mammadov
Faculty of Mechanics and Informattion Tehnology
Department of Information Tecnology and Systems
University of Architecture and Construction
Baku, Azerbaijan
fuad.mammadov.ilham@gmail.com

Huseyn Sultanli
Faculty of Mechanics and Informattion Tehnology
Department of Information Tecnology and Systems
University of Architecture and Construction
Baku, Azerbaijan
huseynsultanli426@gmail.com

***Abstract*—The IELTS Writing Task 2 consists of well-written essays by non-native English speakers to be graded on four categories: Task Response, Coherence and Cohesion, Lexical Resource, and Grammatical Range and Accuracy. Manual grading of these essays is time-consuming and heterogeneous in nature, thus solutions through automation are required. This project gives an example of how a Long Short-Term Memory (LSTM) model, a sequence-trained recurrent neural network, could be employed to mark IELTS Writing Task 2 essays. The model, having been trained on a labeled corpora of essays, returns a total band score and in-depth feedback per criterion. Exploiting LSTM's ability to process text contextual dependence, the system is extremely human-like and accurate marking as possible. Metrics of performance such as prediction accuracy and processing time indicate its potential usability in real-time applications. It enables actionable, real-time feedback for student self-learning and aids teachers in low-resource settings. The project exemplifies automated essay marking, which proves the effectiveness of LSTM-based systems for edtech.**

Keywords—LSTM, IELTS Writing Task 2, automated scoring, natural language processing, educational technology

INTRODUCTION

The International English Language Testing System (IELTS) is an English language proficiency test, taken by millions of test takers every year. Writing Task 2, as the most important aspect of the IELTS test, asks candidates to compose an essay on a given topic. Candidate essays are marked by human markers on four aspects: Task Response, Coherence and Cohesion, Grammar, and Lexical Resource. Even though strict in nature, human assessment is time-consuming and subject to errors because from human subjectivity. Automated essay scoring (AES) systems provide an answer of high promise by furnishing rapid, objective, and trustworthy assessments.

Natural language processing (NLP) and deep learning have revolutionized the quality of AES systems. Architectures like recurrent neural networks (RNNs) and transformers have proven to have an ability to identify complex linguistic structures and structural nuances in written language. In this paper, we present an innovative LSTM-based classification model constructed to predict IELTS Writing Task 2 essays on all four official criteria. The method employs two input texts: the essay of the candidate and a reference text, for example, high-scoring exemplar or model response. By comparing the inputs, the model assesses the candidate's performance in managing the prompt, building their argument, and using effective language.

The dataset used in this research was downloaded from Hugging Face's "chillies/IELTS-writing-task-2-evaluation" repository. The dataset offered essay texts with scores in one text column. Utilizing the processing of regular expressions, we obtained the individual scores for every criterion and reorganized them into different columns to enable model training.

This paper is structured as follows: Section II briefly reviews current research on AES and applications of deep learning in NLP. Section III establishes a theoretical basis, with LSTMs, word embeddings, and multi-task learning. Section IV explains the methodology, e.g., data preprocessing, model architecture, and training methods. Section V provides the experimental design and possible evaluation criteria. Section VI addresses implications, constraints, and future work of this research. Section VII concludes the paper with contributions and their importance.

LITERATURE REVIEW

The area of computer-based essay scoring has evolved considerably in the past several decades. The initial AES systems based on hand-designed features such as essay length and vocabulary density were used to provide score estimates [1]. Although adequate for simple assessments, these approaches did not have the capacity to assess superior linguistic and structural features. The emergence of machine learning and thus deep learning has changed the paradigm towards more complex methods that can bypass these limitations.

Within deep learning methods, recurrent neural networks, and specifically Long Short-Term Memory (LSTM) networks, have become favored due to their effectiveness in sequence modeling tasks like text classification and sentiment analysis [2]. LSTM's sequential processing ability is particularly suited to grading essays where paragraph, sentence, and word sequence is imperative. LSTMs have been used extremely effectively for AES research, where performance has been proven to be superior to that obtained through conventional statistical methods [3], [4].

One of the limitations with the majority of AES systems is that they are capable of producing a single general score without taking into consideration the multi-dimensional character of writing assessment. For IELTS Writing Task 2, where four different criteria must be graded, a multi-output model is required. Multitask learning was discovered to be a logical approach in this regard, allowing a single model to produce several scores simultaneously by sharing representations across tasks [5]. The approach can be potentially enhanced by finding interdependencies between criteria.

Reference text usage is another important advance in AES. Human graders usually compare a test-taker's essay against a perfect answer in order to estimate its quality. Likewise, on computer-based platforms, a reference text may serve as a comparative benchmark to allow the model to estimate compliance with the anticipated in terms of content and structure. While such an idea has been investigated in content-based AES approaches [6], its extension to IELTS-specific multi-criteria grading is unknown.

Our contribution leans on these developments by introducing an LSTM-based framework that takes in both candidate's essay and reference text as inputs to generate scores on the four IELTS dimensions. This two-input system is intended to give a more precise assessment of Task Response and Coherence & Cohesion, highly reliant on content appropriateness and organisational coherence.

THEORETICAL BACKGROUND

Long Short-Term Memory Networks

Long Short-Term Memory (LSTM) networks are a dedicated type of recurrent neural network specially designed to avoid the vanishing gradient problem in typical RNNs. They are especially used to map long-term dependencies in sequential data, which is an important aspect in natural language processing tasks.

An LSTM unit has a cell state and three control gates: an input gate, forget gate, and output gate. The input gate controls what new info is put into the cell state, the forget gate controls what to forget, and the output gate controls what to pass on to the next layer. This structure allows LSTMs to maintain contextual information pertinent for extended sequences, and they are suited to handle essays where thematic development and syntactic coherence are conducted over paragraphs or sentences.

Model Architecture

Word embeddings are low-dimensional dense vector spaces of words that capture semantic and syntactic relationships. In contrast to sparse one-hot encodings, embeddings capture context similarities based on word usage in large corpora and represent a better and more informative representation for NLP applications.

Among the most widely used embedding methods are Word2Vec, GloVe, and FastText. Pre-computed embeddings learned over a large corpus offer a suitable initial point by leveraging external linguistic knowledge, which could potentially make the model wiser about word subtleties in essays.

Multi-Task Learning

Multi-task learning (MTL) is a learning paradigm where one model is trained to accomplish several related tasks at the same time. By making use of the shared representations learned across tasks, MTL could potentially enhance generalizability in addition to the prevention of overfitting compared to training models separately for each task.

In IELTS Writing Task 2, the four assessment criteria—Task Response, Coherence & Cohesion, Grammar, and Lexical Resource—are interrelated components of quality writing. A multi-task learning approach allows the model to leverage these correlations, potentially leading to more precise and consistent predictions on each criterion.

METHODOLOGY

Data Preprocessing

The used dataset in the present research is from the Hugging Face "chillies/IELTS-writing-task-2-evaluation" repository containing IELTS Writing Task 2 essays and scores. Four criteria scores were initially in a text field with the essay text. In order to facilitate systematic analysis, we drew upon regular expressions to scrape these numerical ratings—Task Response, Coherence & Cohesion, Grammar, and Lexical Resource and re-mapped them into independent columns of the dataset.

Text preprocessing started with stripping special characters from essay texts and lowercasing all to provide uniformity. Keras Tokenizer was then employed to split the text into words, generating the vocabulary from frequency of words. For the handling of computational resources, we limited the vocabulary size to 10,000 distinct words. We then encoded each essay as a sequence of integers representing this vocabulary. Since essay lengths varied, we normalized all sequence to a maximum of 500 words by shortening lengthy essays to size and adding zeros to short essays.

Reference texts used as comparative baselines were taken from the dataset as the best-scoring essays for the same prompt as the candidate essay. Such alignment ensures content and structure coherence. The reference texts were preprocessed with the identical preprocessing that was used in the experimental setup for conformity.

Raw scores for the criterion between 0 and 9 were scaled to [0,1] using division by 9. This scaling is the same as that provided by sigmoid activation used in the model output layer to allow efficient learning. Synonym replacement and sentence shuffling data augmentation techniques were also employed to enhance model diversity. During training, the data was divided into 80% train and 20% validation sets, data shuffled per epoch, and processed with batch-size 32.

Model Architecture

The proposed model utilizes the LSTM architecture, which is renowned for handling sequential data. The architecture is two-input with distinct branches for candidate essay and reference text. Both the branches have an embedding layer, which accepts tokenized words and converts them into 100-dimensional dense vectors, and an LSTM layer with 128 units. The two inputs share a common embedding layer in order to obtain typical word representations, but the LSTM layer handles the sequential dependence within each piece of text independently.

The concatenated output from each of the LSTM layers is used to obtain an aggregate representation that captures the interaction between the candidate's essay and the reference. This is then passed through a dense layer of 64 units and ReLU activation, which finalizes the features before the final prediction. The model has four output nodes, each representing one of the IELTS criteria, using sigmoid activation to yield normalized scores between 0 and 1.

The structure can be summarized thus:

- Input 1: Candidate's essay (integer sequence)
- Input 2: Reference text (integer sequence)
- Embedding Layer: 100-dimensional, shared by both inputs

- LSTM Layer: 128 units, return_sequences=False
- Concatenation Layer: Concatenates LSTM outputs
- Dense Layer: 64 units, ReLU activation
- Output Layers: Four 1-unit layers, sigmoid activation

This configuration allows the model to compare the candidate's essay with an ideal response and thus makes it more suitable to determine content-based criteria such as Task Response and Coherence & Cohesion.

Training

The model was adjusted with Adam optimizer and the learning rate of 0.001, a common choice due to its being adaptive learner in nature. Mean squared error (MSE) was used as a loss function to all four outputs, and overall loss was calculated as a sum of per output MSEs. This optimizes for all parameters equally.

To prevent overfitting, we had early stopping whereby we stopped training when the validation loss failed to improve. Training was carried out for a maximum of 50 epochs, giving us enough time for convergence while utilizing the batch size of 32 to keep computation efficient and consistent predictions on each criterion.

EXPRETIMENTAL SETUP

We had proposed the performance evaluation of our model on various criteria such as mean squared error (MSE) for prediction, Pearson correlation coefficient for linear concordance, and quadratic weighted kappa (QWK) for inter-rater concordance—a practice in AES literature.

Instead of presenting results, the section defines the study protocol for evaluation in the entire study. The model would be trained on the provided training set and evaluated on the validation set. Hyperparameter search would entail a grid search over important parameters like the hyperparameters of the number of LSTM units (e.g., 64, 128, 256), embedding size (e.g., 50, 100, 200), and learning rate (e.g., 0.001, 0.0001).

To place our strategy in perspective, we created comparisons against baseline models like a bag-of-words model that loses sequence information and a one-input LSTM model that doesn't see the reference text. These comparisons would highlight the power of our dual-input, multi-output strategy.

DISCUSSION

The double-input LSTM model provides unique benefits to grading IELTS Writing Task 2. It is able to more effectively judge the candidate's alignment with the prompt and structural coherence essential elements of Task Response and Coherence & Cohesion when provided with a reference text. The multi-output design is also suitable for in-depth feedback, enabling candidates to know particular strengths and weaknesses within the four criteria.

There are, nevertheless, some limitations that must be remembered. The success of the model depends on the representativeness of the reference text; an exemplar reference would corrupt predictions. The model will similarly fail essays that lie outside the training distribution, i.e., those with unusual structures or on insufficiently covered subjects in the dataset.

Such automatic generation or choice of the best reference texts can be achieved using, e.g., clustering or similarity metrics in future work. Another possible direction is the inclusion of attention mechanisms, which enable the model to concentrate on significant parts of the text and could improve accuracy and interpretability.

The model's interpretability is a particularly important issue in educational applications. Deep learning models can also be themselves opaque and thus difficult to supply useful feedback. By adding attention mechanisms, we hope to make the model's decision-making process more transparent. For example, attention weights might highlight focus on grammatical mistakes in the instance of the Grammar criterion or notable argument points in the instance of Task Response, providing test takers with explicit guidance for improvement.

Extending the model's generalizability to other types of essays or languages might make it more universally applicable, but implementation in learning environments would require overcoming scalability and user interface barriers. These guidelines emphasize the strength of our approach to enable the creation of AES systems.

CONCLUSION

We introduce an LSTM-based classification model with two inputs here for the automatic evaluation of IELTS Writing Task 2 essays. From both the candidate essay and a reference text, the model gives in-depth evaluations on four criteria, setting the stage for future advancement in AES. While empirical results are not provided, the detailed methodology and experimental setup presented here provide a good starting point for further exploration.

As AES technology further develops, it has the potential to transform education evaluation through timely and unbiased feedback to students worldwide. Our own contribution to this vision is leveraging deep learning in order to overcome the challenges of multi-criteria essay grading.

REFERENCES

[1] S. Dikli, ”An overview of automated scoring of essays,” Journal of Technology, Learning, and Assessment, vol. 5, no. 1, 2006.

[2] Y. Kim, ”Convolutional neural networks for sentence classification,” in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1746–1751.

[3] K. Taghipour and H. T. Ng, ”A neural approach to automated essay scoring,” in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 1882–1891.

[4] Alikaniotis, H. Yannakoudakis, and M. Rei, ”Automatic text scoring using neural networks,” in Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016, pp. 715–725.

[5] J. C. S. Wu, C. Chang, and H. Chang, ”Multitask learning for automated essay scoring with sentiment consistency,” in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLPIJCNLP), 2019, pp. 5885–5890.

[6] P. Chen, Z. Sun, L. Bing, and W. Yang, ”Recurrent attention network on memory for aspect sentiment analysis,” in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017, pp. 452–461.

Section 8: Natural Language Processing

Primary and Secondary Language Identification in Multilingual Texts Using a Multi-Layer Perceptron and Kolmogorov-Arnold Network

Nurzhan Amantay
Department of Computer Engineering, Karabuk University
Karabuk, Turkey
nurzhan.amantay@aurorabilisim.com

Yasin Ortakcı
Department of Computer Engineering, Karabuk Universitya
Karabuk. Turkey
yasinortakci@karabuk.edu.tr

Oğuz Fındık
Department of Computer Engineering, Karabuk University
Karabuk. Turkey
oguzfindik@karabuk.edu.tr

***Abstract*—Language identification in multilingual texts is a critical task for Natural Language Processing (NLP) applications, particularly when texts combine primary and secondary languages. This study introduces a novel approach to identifying both primary and secondary languages using a dataset of 11,000 multilingual sentences, encompassing 27 primary and 58 secondary languages, including Indian languages such as Hindi, Tamil, and Telugu. We utilize Multi-Layer Perceptron (MLP) and Kolmogorov-Arnold Networks (KAN) classifiers, capitalizing on their pattern recognition capabilities to detect linguistic features. The MLP model achieves an accuracy of 99% for primary language identification and 85% for secondary language identification. In comparison, the KAN model attains a similar primary language accuracy of 99% while improving secondary language accuracy to 87%. Both models were tested on a subset of the dataset reserved for evaluation, ensuring robust performance assessment. This work underscores the potential of KAN-based models for robust language identification and provides a foundation for future research in multilingual NLP applications.**

Keywords—Language Identification, Multilingual text, Multi-Layer Perceptron, Kolmogorov-Arnold Network.

INTRODUCTION

With the rise of social media, human communication has become limitless. Social networking platforms are now essential in daily life, enabling global interaction and information sharing. For instance, social media enhances customer engagement, boosts brand visibility, generates leads, drives sales, and increases business revenue.

On social media, users frequently write posts that don't follow standard language rules [1], often mixing several languages within a single sentence. This is common among multilingual people who switch between languages [2]. People often express themselves on social media by mixing languages, like their native tongue and English [3]. This practice, known as code-mixing, involves blending elements such as phrases, vocabulary, and morphemes from different languages [4]. Code-mixing is frequently observed in both spoken and written interactions within multilingual societies [5], such as Hindi-English [6], English-Bengali [7], Persian-English [8], Indonesian-English [9], and Malay-English [10]. Code-mixing is classified into intra-sentential (within a sentence), intra-word (within a word), and inter-sentential (between sentences) types. Differences in spelling and grammar make processing code-mixed social media content challenging for natural language processing (NLP) [6]. Therefore, code-mixed text demands more extensive preprocessing compared to monolingual data [11].

Language identification (LID) is a common preprocessing step that automatically detects languages in a text [12]. It is crucial for downstream NLP tasks like sentiment analysis and machine translation, significantly impacting their overall performance [9],[13].

Most LID research focuses on detecting a single language at the document or sentence level, making LID in code-mixed text still challenging. Sentence-level LID often misses important details and can misclassify languages due to mixing within sentences [13],[14]. This has led researchers to focus on token-level LID. A major gap in current research is the scarcity of code-mixed datasets for low-resource languages, which are less studied due to limited resources [15]. While South Asian languages like Hindi and Bengali mixed with English are common [2], exploring other low-resource language pairs is needed. Creating new datasets is crucial to tackle code-mixed LID challenges in these under-resourced languages. Additionally, dictionary-based methods struggle with borrowed words and code-mixing [16], and capturing context is difficult due to ambiguity and inconsistent phonetic spellings in code-mixed text [17].

This study addressed identifying primary and secondary languages in multilingual, code-mixed texts by comparing two neural network models: Multi-Layer Perceptron (MLP) and Kolmogorov-Arnold Network (KAN). While both performed well on primary language detection, KAN showed greater robustness for secondary and low-resource languages. Challenges remain with extremely low-resource languages due to data sparsity. This work highlights KAN's potential as a superior approach and encourages future research on data imbalance to enhance rare language detection.

The rest of this paper is organized as follows. Section 2 presents a review of related work. In Section 3, the proposed methods are described in detail. Section 4 provides experimental results to evaluate the performance of the proposed approaches. Finally, the conclusion is presented in Section 5.

RELATED WORKS

Support Vector Machine (SVM) is widely used in LID due to its effective classification and strong performance [18]. Kalita and Saharia achieved 89.51% accuracy detecting Assamese-English code-mixed text using linear kernel SVM with N-gram and dictionary features [19]. Shanmugalingam et al. also found linear kernel SVM performed best with 89.46% accuracy for Tamil-English code-mixed LID [20]. Kazi et al. reported that SVM with RBF kernel and N-gram features reached the highest accuracy of 92% [21].

Conditional Random Fields (CRF) is a statistical method used in sequence labeling tasks like named entity recognition

and LID, leveraging context for better accuracy [22]. Yirmibeşoğlu and Eryiğit achieved 95.6% micro-F1 with CRF using character-level N-grams [23]. Mave et al. found CRF outperformed LSTM and BLSTM, scoring 98% and 96% F1 on Hindi-English data [24]. Mishra and Sharma showed CRF accurately detected multilingual code-mixed text with 97.77% accuracy and 95% F1 [25].

Naïve Bayes is commonly used as a baseline due to its simplicity, applying Bayes' theorem with an assumption of feature independence [26]. Gupta et al. [27] showed that combining edit distance with Naïve Bayes on an N-gram Markov model effectively identified languages from misspelled words. Lakshmi and Shambhavi [28] found Bernoulli Naïve Bayes with TF-IDF and a dictionary module outperformed SVM, Random Forest, and Logistic Regression, achieving 94.8% accuracy, 96.3% precision, and 95.2% recall. Kalita et al. [29] reported Naïve Bayes outperformed Decision Tree and MLP, with an F1 of 65.9%, precision of 76.2%, and recall of 69.3%.

Bansal et al. used Logistic Regression for English-Punjabi code-mixed LID, outperforming Decision Tree and Gaussian Naïve Bayes with 86.63% accuracy and 88% F1 score [30]. Shanmugalingam and Sumathipala found Random Forest achieved the best results for Sinhala-English code-mixed LID, with 90.5% accuracy and F-measures of 94.9% for Sinhala and 75.8% for English [31].

Neural network approaches in showed SegRNN outperforming Naïve Bayes and BLSTM, with F1 scores of 94.84% for intra-word labeling and 99.17% for segmentation [26]. LSTM proved effective for Hindi-English and Bengali-English code-mixed texts [32]. Samih et al. combined char-word LSTM with CRF, achieving a highest accuracy of 96.3% [33].

In text processing, CNN is often used to extract features before machine learning. The convolutional layer applies filters to text segments. Sarma et al. found CNN outperformed other methods with an F1 score of 91.03% [34]. Jaech et al. combined CNN and BLSTM for Spanish-English word-level LID, where CNN created word embeddings from characters and BLSTM captured sequence context. Their model achieved F1 scores of 95.1% for English and 94.1% for Spanish [35].

Since the introduction of the transformer model by Vaswani et al. [36], transformers have quickly become a reliable NLP method, outperforming earlier networks like CNNs and RNNs [37]. Thara and Poornachandran proposed a transformer-based word-level code-mixed LID system for Malayalam-English, evaluating five BERT variants: BERT, DistilBERT, ELECTRA, XLM-RoBERTa, and CamemBERT. ELECTRA achieved the best results, with an F1 score of 99.33% and accuracy of 99.41% [38].

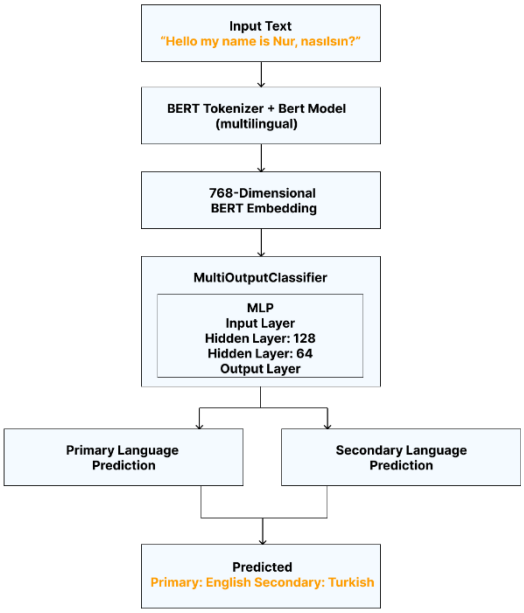
Proposed Methods

Multilayer Perceptron

Environmental modeling and complex classification tasks often require techniques capable of handling nonlinear relationships without depending on strict assumptions about data distribution. The MLP, a feedforward neural network structure, has proven effective in these contexts due to its ability to approximate nearly any smooth function [39]. Composed of interconnected neurons organized into input, hidden, and output layers, the MLP uses nonlinear activation functions to model highly intricate input-output mappings. It is trained using supervised learning, progressively adjusting connection weights through backpropagation to minimize prediction errors, thereby achieving strong generalization to new data [40].

In this study, we utilize an MLP as a basic classifier for multilabel LID. Rather than using raw text, we leverage contextual embeddings derived from the pretrained *bert-base-multilingual-cased* model, specifically the [CLS] token representation from BERT's final hidden layer, as input features. These embeddings encode extensive semantic and syntactic information from multilingual datasets, offering a robust, language-sensitive representation (fig. 1). The MLP takes these BERT-derived feature vectors as input and generates predictions for both primary and secondary language labels concurrently within a multi-output classification setup (table 1). The network's structure includes several hidden layers with nonlinear activation functions, allowing it to capture intricate decision boundaries essential for precise multilabel classification. The model is trained in a supervised fashion using annotated data, fine-tuning the parameters to generalize well across varied language inputs.

This combined method capitalizes on the advantages of transfer learning from BERT's contextual embeddings along with the adaptability and capability of MLPs to capture nonlinear relationships, making it a strong and efficient solution for multilabel LID.



MLP architecture for Language identification

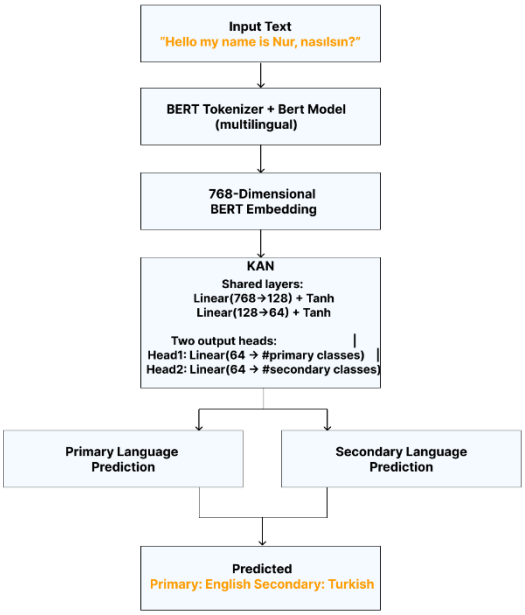
Kolmogorov-Arnold Network

KAN is an innovative neural architecture inspired by the Kolmogorov–Arnold representation theorem, which asserts that any multivariate continuous function can be expressed as a composition of univariate continuous functions. Unlike traditional neural networks such as MLPs, which utilize fixed nonlinear activation functions at each neuron, KANs use trainable nonlinear functions along the connections between nodes. Each edge is parameterized by a spline or a small neural network, while the nodes themselves perform simple

summation without additional nonlinear activation. This edge-centric design allows KANs to capture complex feature transformations while maintaining interpretability and potentially reducing the overall network complexity [41].

In our multilingual LID task, we utilize KAN as an alternative classifier operating on contextual embeddings generated by the bert-base-multilingual-cased model (fig. 2). Specifically, we extract the [CLS] token embedding from BERT as a 768-dimensional feature vector and input it into a compact KAN consisting of two KAN layers. Each layer performs trainable spline-based transformations on the connections, allowing precise control over feature mapping. The architecture features two output branches—each a softmax classifier—responsible for predicting primary and secondary languages. During supervised training, a dual cross-entropy loss function is employed, and the model is optimized with the Adam optimizer (table 2).

KAN's capacity to represent high-dimensional nonlinear relationships via adjustable edge functions makes it especially suitable for structured prediction tasks such as multilingual LID. It provides a promising and interpretable alternative to conventional MLPs, particularly when model transparency and flexibility are important.



KANs architecture for Language identification

EXPERIMENTAL RESULTS

Dataset

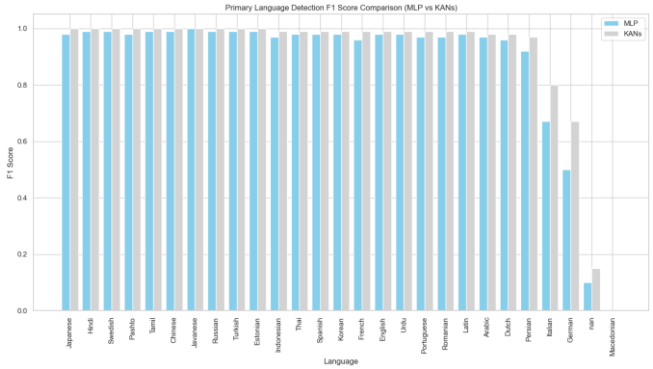
The experiments in this study were conducted using the *WiLI-2018 dataset* (Wikipedia Language Identification), a widely recognized benchmark for multilingual LID tasks [42]. For the purposes of this study, the original dataset [43] was preprocessed to include an additional column labeled *secondary_language*, representing potential secondary languages that may appear in the text [44]. This secondary label was generated using the *Gemini 2.5* model to simulate multilingual usage within the data, thereby enabling the development and evaluation of models capable of dual-language detection. After preprocessing, the final version of the dataset comprised *11,000* sentence rows.

Experimental Results of MLP

EXPERIMENTAL SETUP FOR MLP

Parameter	Value
Classifier	MultiOutputClassifier with MLP
Text Embedding	[CLS] token from BERT output (768-dim)
Max Token Length	64
MLP Hidden Layers	(128, 64)
Max Iterations	100
Early Stopping	Yes

The evaluation of the multilingual language detection system reveals strong performance in primary language detection, achieving an impressive overall accuracy of 99%. Most primary languages including Chinese, English, Hindi, Turkish, and Spanish – were identified with near-perfect precision, recall, and F1-scores. In contrast, secondary language detection posed greater challenges, with a lower macro F1-score of 0.47 and weighted F1-score of 0.84, though the accuracy remained high at 85%. While languages such as English, French, Japanese, and Latin performed well as secondary languages, many others – especially those with very few instances like Azerbaijani, Bengali, and Kannada had zero performance, highlighting the difficulty of accurate secondary LID in imbalanced or sparse data scenarios (fig. 3).

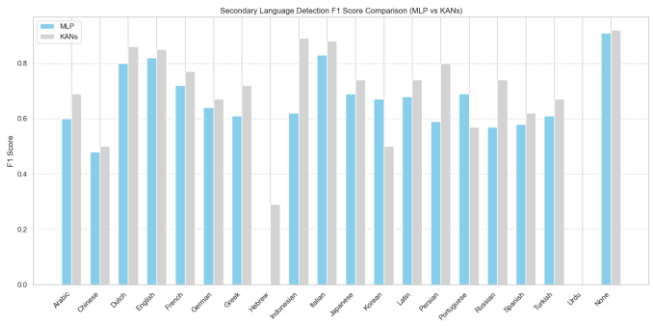


Primary language F1 score comparison (MLP and KANs)

Experimental Results of KAN

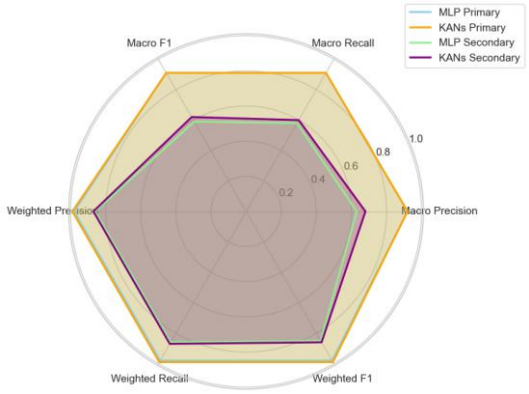
EXPERIMENTAL SETUP FOR KAN

Parameter	Value
Classifier	Shared FFN: Linear(768->128) ->Tanh ->Linear(128 -> 64) ->Tanh Head 1: Linear(64 -> #primary_classes) Head 2: Linear(64 -> #secondary_classes)
Text Embedding	[CLS] token from BERT output (768-dim)
Max Token Length	64
MLP Hidden Layers	(128, 64)
Max Iterations	100
Early Stopping	Yes
Activation Function	Tanh
Loss Function	CrossEntropyLoss (one for each head, summed)
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Epoch	30



Secondary language F1 score comparison (MLP and KANs)

The language detection model demonstrates excellent performance in identifying primary languages, achieving an overall accuracy of 99% with high precision, recall, and F1-scores across nearly all major languages. Particularly strong results are observed for widely spoken languages such as English, Chinese, Russian, Turkish, and Hindi, all with F1-scores near or at 1.00. For secondary languages, the model achieves a respectable 87% accuracy, but the performance is more variable due to class imbalance and fewer examples per language. While commonly occurring secondary languages like English, French, German, and Latin show good results, rare or low-resource languages (e.g., Kurdish, Urdu, Tagalog) are often misclassified or missed entirely (fig. 4).



Macro and Weighted scores (MLP and KANs)

The macro and weighted precision, recall, and F1-scores for both MLP and KAN models are compared. The turtle bar visualization shows that the MLP model consistently achieves higher scores across all metrics, especially in the weighted averages, indicating better overall performance when considering class imbalance (fig. 5).

CONCLUSION

This study addressed the critical challenge of identifying both primary and secondary languages within multilingual, code-mixed texts. By leveraging a comprehensive dataset of 11,000 sentences, we evaluated and compared the performance of two distinct neural network architectures: a MLP and KAN.

Our findings indicate that both models are highly effective for primary LID, each achieving an exceptional accuracy of 99%. However, for the more complex task of secondary language detection, the KAN model demonstrated a clear advantage, reaching 87% accuracy compared to the MLP's 85%. The superiority of the KAN architecture was particularly evident in its handling of low-support languages, where it achieved a significantly better F1-score, underscoring its robustness. While overall performance was strong, both models faced limitations with extremely low-resource secondary languages, where data sparsity and class imbalance hindered effective identification.

This work successfully highlights the substantial potential of KANs as a powerful and more effective alternative to traditional models for nuanced LID tasks. It provides a solid foundation for future research, which should focus on addressing data imbalance through techniques like data augmentation to further enhance the detection of rare languages in multilingual contexts.

REFERENCES

[1] F. Hidayatullah, ‘‘Language tweet characteristics of Indonesian citi- zens,’’ in *Proc. Int. Conf. Sci. Technol. (TICST)*, Pathum Thani, Thailand, Nov. 2015, pp. 397–401, doi: 10.1109/TICST.2015.7369393.

[2] N.Jose,B.R.Chakravarthi,S.Suryawanshi,E.Sherly,andJ.P.McCra e, ‘‘A survey of current datasets for code-switching research,’’ in *Proc. 6th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Coimbatore, India, Mar. 2020, pp. 136–141, doi: 10.1109/ICACCS48705.2020.9074205.

[3] K. Shanmugalingam and S. Sumathipala, ‘‘Language identification at word level in Sinhala-English code-mixed social media text,’’ in *Proc. Int. Res. Conf. Smart Comput. Syst. Eng. (SCSE)*, Colombo, Sri Lanka, Mar. 2019, pp. 113–118, doi: 10.23919/SCSE.2019.8842795.

[4] Myers-Scotton, ‘‘Common and uncommon ground: Social and struc- tural factors in codeswitching,’’ *Lang. Soc.*, vol. 22, no. 4, pp. 475–503, Dec. 1993.

[5] Gautam, P. Kodali, K. Gupta, A. Goel, M. Shrivastava, and P. Kumaraguru, ‘‘CoMeT: Towards code-mixed translation using parallel monolingual sentences,’’ in *Proc. 5th Workshop Comput. Approaches Lin- guistic Code-Switching*, 2021, pp. 47–55.

[6] S.MandalandA.K.Singh,‘‘Languageidentificationincode-mixeddata using multichannel neural networks and context capture,’’ in *Proc. EMNLP Workshop W-NUT: 4th Workshop Noisy User-Generated Text*, Brussels, Belgium, 2018, pp. 116–120, doi: 10.18653/v1/W18-6116.

[7] A.Jamatia,S.D.Swamy,B.Gambäck,A.Das,andS.Debbarma,‘‘De ep learning based sentiment analysis in a code-mixed English-Hindi and English-Bengali social media corpus,’’ *Int. J. Artif. Intell. Tools*, vol. 29, no. 5, Aug. 2020, Art. no. 2050014, doi: 10.1142/s0218213020500141.

[8] N. Sabri, A. Edalat, and B. Bahrak, ‘‘Sentiment analysis of Persian- English code-mixed texts,’’ in *Proc. 26th Int. Comput. Conf., Com- put. Soc. Iran (CSICC)*, Tehran, Iran, Mar. 2021, pp.1–4, doi: 10.1109/CSICC52343.2021.9420605.

[9] M. Barik, R. Mahendra, and M. Adriani, ‘‘Normalization of Indonesian-English code-mixed Twitter data,’’ in *Proc. 5th Workshop Noisy User-Generated Text (W-NUT)*, Hong Kong, 2019, pp. 417–424, doi: 10.18653/v1/d19-5554.

[10] M. J. Fuadvy and R. Ibrahim, ‘‘Multilingual sentiment analysis on social media disaster data,’’ in *Proc. Int. Conf. Electr., Electron. Inf. Eng. (ICEEIE)*, Bali, Indonesia, Oct. 2019, pp.269–272, doi: 10.1109/ICEEIE47180.2019.8981479.

[11] R. Srinivasan and C. N. Subalalitha, ‘‘Sentimental analysis from imbal- anced code-mixed data using machine learning approaches,’’ *Distrib. Par- allel Databases*, pp. 1–16, Mar. 2021, doi: 10.1007/s10619-021-07331-4.

[12] T. Jauhiainen, M. Lui, M. Zampieri, T. Baldwin, and K. Lindén, ‘‘Auto- matic language identification in texts: A survey,’’ *J. Artif. Intell. Res.*, vol. 65, pp. 675–782, Aug. 2019, doi: 10.1613/jair.1.11675.

[13] N.Sarma,S.R.Singh,andD.Goswami,‘‘Wordlevellanguageidentifica- tion in Assamese-Bengali-Hindi-English code-mixed social media text,’’ in *Proc. Int. Conf. Asian Lang. Process. (IALP)*, Bandung, Indonesia, Nov. 2018, pp. 261–266.

[14] Das and B. Gambäck, ‘‘Code-mixing in social media text: The last language identification frontier?’’ *Traitement Automatique des Langues*, vol. 54, no. 3, pp. 41–64, 2013.

[15] P. King, ‘‘Practical natural language processing for low-resource lan- guages,’’ Ph.D. dissertation, Dept. Comput. Sci. Eng., Univ. Michigan, Ann Arbor, MI, USA, 2015.

[16] S.KentandD.Claeser,‘‘Incorporatingcode-switchingandborrowingin Dutch-English automatic language detection on Twitter,’’ in *Proc. Future Technol. Conf. (FTC)* (Advances in Intelligent Systems and Computing), vol. 880.

Cham: Springer, 2018, pp. 418–434, doi: 10.1007/978-3-030-02686-8_32.
[17] S. D. Das, S. Mandal, and D. Das, ''Language identification of Bengali- English code-mixed data using character & phonetic based LSTM mod- els,'' in *Proc. 11th Forum Inf. Retr. Eval.*, Kolkata, India, Dec. 2019, pp. 60–64, doi: 10.1145/3368567.3368578.
[18] V. Veena, M. A. Kumar, and K. P. Soman, ''An effective way of word-level language identification for code-mixed Facebook comments using word-embedding via character-embedding,'' in *Proc. Int. Conf. Adv. Comput., Commun. Informat. (ICACCI)*, Udupi, India, Sep. 2017, pp. 1552–1556.
[19] N.J.KalitaandN.Saharia,''Languageidentificationoncode-mixsocial text,'' in *Proc. Int. Conf. Comput. Commun. Syst.* (Lecture Notes in Networks and Systems). Singapore: Springer, 2018, pp. 433–440, doi: 10.1007/978-981-10-6890-4_42.
[20] K.Shanmugalingam,S.Sumathipala,andC.Premachandra,''Wordl evel language identification of code mixing text in social media using NLP,'' in *Proc. 3rd Int. Conf. Inf. Technol. Res. (ICITR)*, Moratuwa, Sri Lanka, Dec. 2018, pp. 1–5, doi: 10.1109/ICITR.2018.8736127.
[21] M.Kazi,H.Mehta,andS.Bharti,''Sentencelevellanguageidentification in Gujarati-Hindi code-mixed scripts,'' in *Proc. IEEE Int. Symp. Sustain. Energy, Signal Process. Cyber Secur. (iSSSC)*, Gunupur Odisha, India, Dec. 2020, pp. 1–6, doi: 10.1109/iSSSC50941.2020.9358837.
[22] P.LamabamandK.Chakma,''Alanguageidentificationsystemforc ode- mixed English-Manipuri social media text,'' in *Proc. IEEE Int. Conf. Eng. Technol. (ICETECH)*, Coimbatore, India, Mar. 2016, pp. 79–83.
[23] Z. Yirmibeşoğlu and G. Eryiğit, ''Detecting code-switching between Turkish-english language pair,'' in *Proc. EMNLP Workshop W-NUT: 4th Workshop Noisy User-Generated Text*, Brussels, Belgium, 2018, pp. 110–115, doi: 10.18653/v1/W18-6115.
[24] D.Mave,S.Maharjan,andT.Solorio,''Languageidentificationanda nal- ysis of code-switched social media text,'' in *Proc. 3rd Workshop Comput. Approaches Linguistic Code-Switching*, Melbourne, VIC, Australia, 2018, pp. 51–61.
[25] Mishra and Y. Sharma, ''Language identification and context-based analysis of code-switching behaviors in social media discussions,'' in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Los Angeles, CA, USA, Dec. 2019, pp. 5951–5956, doi: 10.1109/BigData47090.2019.9006032.
[26] Sabty, I. Mohamed, Ö. Çetinoglu, and S. Abdennadher, ''Language identification of intra-word code-switching for Arabic–English,'' *Array*, vol. 12, Dec. 2021, Art. no. 100104, doi: 10.1016/j.array.2021.100104.
[27] Gupta, G. Bhatt, and A. Mittal, ''Language Identification and dis-ambiguation in Indian mixed-script,'' in *Proc. Int. Conf. Distrib. Com- put. Internet Technol. (ICDCIT)*, vol. 9581. Cham: Springer, 2016, pp. 113–121, doi: 10.1007/978-3-319-28034-9_14.
[28] S. Sowmya Lakshmi and B. R. Shambhavi, ''An automatic language identification system for code-mixed English-Kannada social media text,'' in *Proc. 2nd Int. Conf. Comput. Syst. Inf. Technol. Sus- tain. Solution (CSITSS)*, Bengaluru, India, Dec. 2017, pp. 1–5, doi: 10.1109/CSITSS.2017.8447784.
[29] N. J. Kalita, A. G. Agarwala, and J. Das, ''Word level language identifi- cation on code-mixed English-bodo text,'' in *Proc. 6th Int. Conf. Comput. Manage. Math. Sci. (ICCM)*, Nirjuli, India, vol. 1020, 2021, pp. 1–6, doi: 10.1088/1757-899X/1020/1/012027.
[30] N. Bansal, V. Goyal, and S. Rani, ''Experimenting language identifica- tion for sentiment analysis of English Punjabi code mixed social media text,'' *Int. J. E-Adoption*, vol. 12, no. 1, pp. 52–62, Jan. 2020, doi: 10.4018/ijea.2020010105.
[31] K. Shanmugalingam and S. Sumathipala, ''Language identification at word level in Sinhala-English code-mixed social media text,'' in *Proc. Int. Res. Conf. Smart Comput. Syst. Eng. (SCSE)*, Colombo, Sri Lanka, Mar. 2019, pp. 113–118, doi: 10.23919/SCSE.2019.8842795.
[32] S. D. Das, S. Mandal, and D. Das, ''Language identification of Bengali- English code-mixed data using character & phonetic based LSTM mod- els,'' in *Proc. 11th Forum Inf. Retr. Eval.*, Kolkata, India, Dec. 2019, pp. 60–64, doi: 10.1145/3368567.3368578.
[33] Y.Samih,S.Maharjan,M.Attia,L.Kallmeyer,andT.Solorio,''Multi lin- gual code-switching identification via LSTM recurrent neural networks,'' in *Proc. 2nd Workshop Comput. Approaches Code Switching*, Austin, TX, USA, 2016, pp. 50–59.
[34] N. Sarma, R. S. Singh, and D. Goswami, ''SwitchNet: Learning to switch for word-level language identification in code-mixed social media text,'' *Natural Lang. Eng.*, vol. 28, no. 3, pp. 337–359, 2022, doi: 10.1017/s1351324921000115.
[35] A.Jaech,G.Mulcaire,M.Ostendorf,andN.A.Smith,''Aneuralmodel for language identification in code-switched tweets,'' in *Proc. 2nd Workshop Comput. Approaches Code Switching*, Austin, TX, USA, 2016, pp. 60–64.
[36] A.Vaswani,N.Shazeer,N.Parmar,J.Uszkoreit,L.Jones,A.N.Gomez, Ł. Kaiser, and I. Polosukhin, ''Attention is all you need,'' in *Proc. 31st Conf. Neural Inf. Process. Syst. (NIPS)*, Long Beach, CA, USA, 2017, pp. 5998–6008.
[37] T.Wolf,J.Chaumond,L.Debut,V.Sanh,C.Delangue,A.Moi,P.Cistac, M. Funtowicz, J. Davison, and S. Shleifer, ''Transformers: State-of-the-art natural language processing,'' in *Proc. Conf. Empirical Methods Natural Lang. Process., Syst. Demonstrations*, 2020, pp. 38–45.
[38] S. Thara and P. Poornachandran, ''Transformer based language identi- fication for Malayalam-English code-mixed text,'' *IEEE Access*, vol. 9, pp. 118837–118850, 2021, doi: 10.1109/access.2021.3104106.
[39] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359–366, 1989.
[40] M. W. Gardner and S. R. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric Environment*, vol. 32, no. 14-15, pp. 2627-2636, 1998, doi: 10.1016/S1352-2310(97)00447-0.
[41] Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, "KAN: Kolmogorov–Arnold Networks," Massachusetts Institute of Technology, California Institute of Technology, Northeastern University, and NSF Institute for Artificial Intelligence and Fundamental Interactions.
[42] M. Thoma, "The WiLI benchmark dataset for written language identification," *Data in Brief*, vol. 19, pp. 2167–2175, Aug. 2018, doi: 10.1016/j.dib.2018.06.091.
[43] https://www.kaggle.com/code/martinkk5575/language-detection/input
[44] https://github.com/IT-special/LID/blob/main/LID_dataset.csv

Section 9: Mathematical Modeling Methods and Their Applications

Assessing Migration Patterns of Researchers in Azerbaijan Using Digital Trace Data

Makrufa Hajirahimova
Department 13
Institute of Information Technology
Baku, Azerbaijan
0000-0003-0786-5974

Aybeniz Aliyeva
Department 13
Institute of Information Technology
Baku, Azerbaijan
0000-0002-1739-1808

***Abstract*— The process of globalization that is happening all over the world has given rise to international mobility of a skilled workforce with innovative thinking and high knowledge. Understanding the mechanisms behind the mobility of skilled workers is of great importance for the development of science, education and immigration policies. The lack of access to official data stimulates the use of digital trace data obtained from ORCID public records. ORCID profiles contain standardized information education and employment, which allows tracking individuals over time. This study used ORCID data to analyze the mobility of Azerbaijani researchers during 1990-2024. A multivariate nonlinear regression model was proposed for the evaluation of researcher's migration. The findings of the study may have serious practical implications for the making of migration policies to attract and retain skilled talent.**

Keywords—high-skilled migration, scientific migration, ORCID, digital trace data, regression analysis

INTRODUCTION

The globalization of knowledge has promoted to international mobility by helping to transform a skilled workforce and researchers with innovative thinking and high knowledge into key drivers of economic and social development. This has increased performance of researches by facilitating the recombination of knowledge and created the basis for scientific discoveries [1]. As the role of highly skilled researchers in economic development is recognized and countries look to close skills gaps global competition for talent are intensified. Policymakers are actively seeking ways to improve research activity and skills attraction capabilities in their regions in order to achieve economic growth and productivity. However, despite the large economic impact of the scientific achievements of researchers and scientists, the relevance of intellectual migration and the need for a deep understanding of migration processes in order to make sensible policy interventions, the possibilities to measure and track the mobility of scientific researchers remain limited [2]. One of the main reasons for this is the lack of high-quality and accessible data over an extended period of time or the factor that measuring the mobility of highly skilled workers using traditional methods (including surveys, census data and CV analysis) is costly and time-consuming [2, 3].

Understanding the flows of scientists – where they studied, where they moved, how often they moved, how this varies by field, and the type of research university they worked at – is crucial for science and immigration policy.

Recently, the use of non-traditional data sources, in particular bibliometric data [2, 3] and digital trace data [4-6], has become an increasingly researched subject and interest in their application is constantly growing [2, 7-10].

In this study, we monitor the international mobility of Azerbaijani researchers and scholars by proposing the use of ORCID profile data. In particular, we compile the most comprehensive dataset of 1400 international migration of researchers for the period 1990 to 2024 and use it to analyze and estimate the mobility of migration flows of researchers.

PRE-PROCESSING DATA

This section describes the steps taken to retrieve data from ORCİD, preparing it for further processing and analysis.

Migration Dataset

Since official data for the analysis of the phenomenon of researchers migration are not available, we extract the necessary migration patterns from the ORCID public data. ORCID (Open Researcher and Contributor ID) is a non-profit organization that provides a unique identifier to each registered user member engaged in research. It contains information on researcher's works, publications and affiliations. This data is a valuable source of education and employment histories of PhD students, academics, engineers, medical professionals, researchers. The majority of ORCID users have or are in the process of obtaining a PhD. It means that ORCID data captures a subset of high-skilled workers [3, 11]. From ORCID public database we extract migration patterns of registered users, across Azerbaijan in the time horizon spanning from 1990 to 2024 and create a dataset . The data was obtained by analyzing the educational and employment histories of over 13,000 users registered in ORCID. Comparison with independent data sources supports the technical validity and representativeness of this data. Incomplete information on years of employment or education was completed using diaspora and workplace websites, social networking sites such as Linkedin and Facebook. Upon graduation all students, especially PhD students, have the option to register themselves in the ORCID database. The dataset provides new dimensions of the geography of a subset of highly skilled labor and opens up opportunities to explore hypotheses about intellectual migration. The key variables included in this database and their description are given in Table 1.

TABLE I. VARIABLES AND THEIR DESCRIPTION

Variables	***Description***
orcid_id	Unique open digital identifier
mig_year	Year in which the researcher left the country
dest_country	Country where the researcher first migrated
has_phd	Does researcher have a PhD?
phd_year	Year in which the PhD researcher received the degree
phd_country	Country where the PhD researcher received the degree

science_field_group	The main field of science in which the researcher is involved
gender	Gender of researcher (female or male)
earliest_degree	Degree when the researcher leaves the country
science_degre_2024	Current degrees of researchers
country_2024	The country where the researcher is currently active
current_position_2024	Current affiliation roles of researchers

Figure 1, displayed Azerbaijani researchers (with ORCID profile) migration over time, from 1990 to 2024. The overall number of researchers shows an increasing trend, with a peak in the year 2018. The total number of migrated researchers also experienced an increasing trend over time from 1990 to 2018, the number of researchers has dropped in 2019 and 2024.

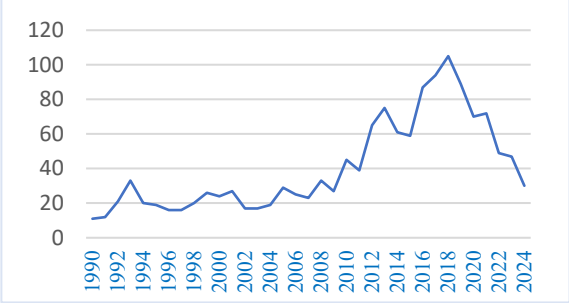


Fig.1. Intellectual migration in Azerbaijan (1990-2024) [6, 12].

The distribution of Azerbaijani researchers and scientists registered in the Orcid database by country is depicted in Figure 2. As can be seen from the graph, the devoleped countries , like Turkey, Germany, USA and Russia take the leading position in the process of "brain drain" from Azerbaijan.

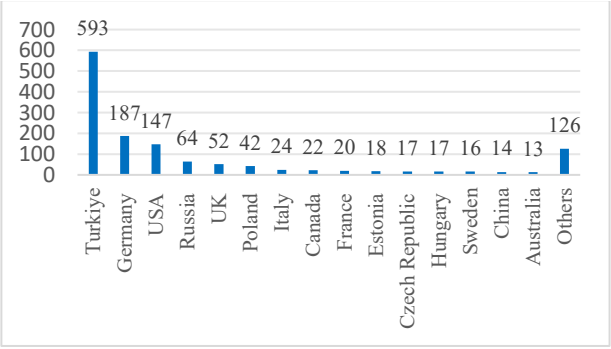


Fig. 2. Distribution Azerbaijani researchers and scientists registered in the Orcid database by countries [6, 12].

In terms of degree, the highest number of migrated researchers regestred in ORCİD is in the PhD degree (42%) and Master of Science degree (27%) (Fig. 3). The other degrees of Doctor of Science (15%) and PHD Student (16%) have fewer than the number of PhD student and MSc.

Also number of migrated non-PhD researchers regestred in ORCİD (58%) have more than the number of PhD researchers (42%) (Fig. 4).

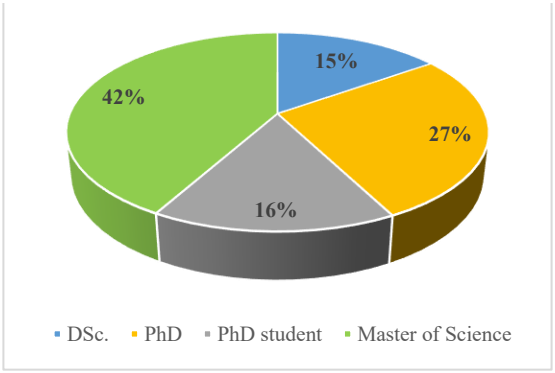


Fig. 3. Number of researchers regestred in ORCİD in terms of degree (%).

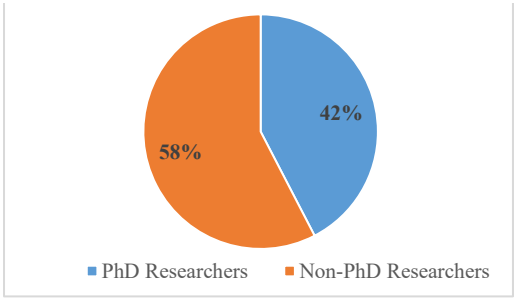


Fig.4. Number of migrated PhD researchers and non-PhD researchers regestred in ORCİD (%).

METHODOLOGY

This section presents the methodology for analyzing the migration patterns of researchers in Azerbaijan. The methodology includes several stages. The first step is to create dataset for study.

Selection Variables and Creating Dataset

Considering previous research works on migration, we collect a selection of variables we deem to be tied with the phenomenon and interesting from a policy standpoint.

1*) GDP per Capita.* We select GDP per capita as it is one of the most universally adopted and straightforward indicators of the overall economic wealth of a region. This metric is often used to compare economic well-being and living standards between countries. GDP per capita is the total value of a country's finished goods and services (gross domestic product) divided by its total population (per capita) [13].

2**)** *Education Index.* As an indicator of the overall schooling provision of a region we select the Education Index. This indicator is a component of the Human Development Index and it is defined as the geometric average of mean years of schooling and expected years of schooling [39]. The basis for the data is taken from Global Data Lab [14].

3) *Average monthly wages ($).* Averag monthly wages represent the typical earnings of workers in a specific area, influencing various economic indicators.They also represent a standard of living This data is taken from UNECE Statistical Database, compiled from national and international (OECD, EUROSTAT, CIS) official sources. Gross average monthly wages cover total wages and salaries, before any tax deduction and before social security contributions **[**15].

We create a dataset taking into account migration data and parameters. The dataset covers intellectual migration data and selected parameter values for the period 1995-2023. An example record from a dataset is shown in Table 2.

TABLE 2. DESCRİPTİON OF THE DATASET

	Year	Number_mig	GDP_per_capita	Average_ monthly_wages	Education_Index
0	1995	19	1.098	14.2	0.649
1	1996	16	1.096	20.8	0.654
2	1997	16	1.144	35.5	0.658
3	1998	20	1.245	43.5	0.662
4	1999	26	1.324	44.8	0.667

A. Method

There are several statistical approaches for exploring relationships between two or more parameters. Regression analysis model is one of the statistical approaches which recommended for this kind of analysis. In the present study, to make the regressions more flexible, Multiple NonLinear Regression (MNLR) models were applied . The general form of a MNLR function can be represent (explained) by the following equation [16]:

$$Y = \alpha + \beta_1 X_i + \beta_2 X_j + \beta_3 X_i^2 + \beta_4 X_j^2 \ldots + \beta_n X_i X_j$$

Where, Y is the dependent variable (e.g. migration flow for our study) and $X_1, X_2, \ldots X_n$ are the independent variables. Also, α is the intercept and $\beta_1, \beta_2 \ldots, \beta_n$ are the coefficients of the independent variables while n is the number of observations.

One of the major problems of statistical analysis was to establish the appropriate relationship between the dependent variable and a set of independent variables. Different functions including the exponential, power, cubic, logarithmic, quadratic and linear functions are used to identify the best relation.

The MNLR analysis was conducted in a Python programming environment using the Scikit-Learn library. The performance of the developed MNLR models were assessed by statistical performance measures, Pearson correlation value (r) and the Root Mean Square Error (RMSE), the Mean Absolute Error(MAE) and Mean Squared Error (MSE). The lower values of the RMSE, MAE, MSE will indicate the better performance of the model.

RESULTS

One key condition for justifying final results In multifactor correlation models is having a requirement of dependencies between factors. We check the Pearson correlation (Fig 7). The Pearson correlation is a statistical measure that assesses the linear relationship between two variables. A correlation matrix which created using Pearson correlation is shown in Table 3. Analysis of the pairwise correlation coefficient matrix shows that the number of migrants (the "Number of migration" dependent variable) has a strong relationship with the factors GDP per capita (0.770659) and Education Index (0.802589). At the same time, there is a strong relationship between GDP per capita (0.918665) and Education Index.

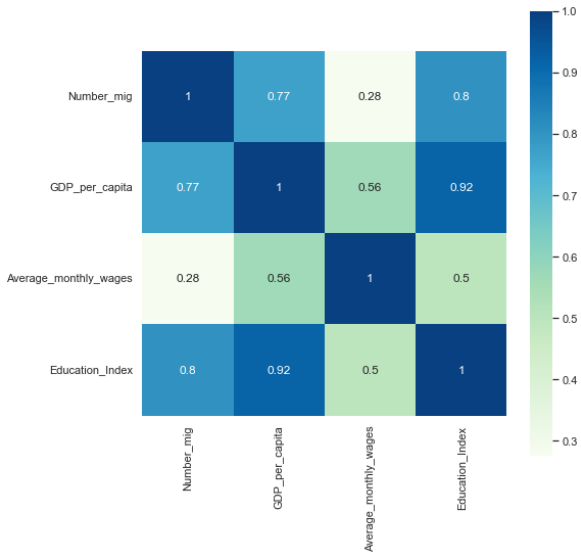


Fig. 7. Pearson correlation.

TABLE III. MATRİX OF CORRELATİON COEFFİCİENTS BETWEEN VARİABLE

	Number of migration	**GDP per capita**	**Average monthly wages**	**Education Index**
Number of migrations	1			
GDP per capita	0.770659	1		
Average monthly wages	0.276535	0.563194	1	
Education Index	0.802589	0.918665	0.498175	1

The ability of the best MNLR model to predict future migration flows has been demonstrated flows in Figure 8.

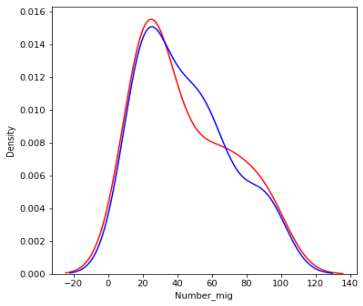


Fig. 8. Comparison between the observed (blue line) and simulated (red line) migration flows.

Some differences can be identified between the observed and predicted (simulated) migration flow. The reason can be that some other indices might have been more influential at that time rather than the selected indices in this study.

The results showed that, our model exhibited the best performance, with an R2 value of 0.893, Adj.R2-Squared value of 0.833, RMSE value of 8.544, a MAE value of 6.644, and MSE value of 73.005, indicating a very close match between predicted and observed values (Table 4).

TABLE IV. THE PERFORMANCE METRICS OF THE MNLR MODEL

Metrics	*Values*
R2-Squared	0.893
Adj.R2-Squared	0.833
RMSE	8.544
MSE	73.005
MAE	6.644

The formation of the multiple non-linear equation can be explained by the following equation where cubic functions are found to be more suitable for the indices based on the correlation values.

$$M = \alpha + \beta_1 AVR^3 + \beta_2 AVR^2 + \beta_3 AVR + \beta_4 GDP^3 + \beta_5 GDP^2 + \beta_6 GDP + \beta_7 EI + \beta_8 AVR * GDP + \beta_9 AVR * EI + \beta_{10} GDP * EI$$

Where, M denotes migration, AVR denotes Average monthly wage, GDP denotes GDP per capita, and EI denotes Education index.

The regression coefficients of the model are obtained using optimization methods such as the least squares method. The least squares method consists in finding such values of the coefficients that minimize the sum of the squares of the forecast errors. Considering the values of the regression coefficients of the model, the predictor model is as follows:

$$M = 383.417 - 0.001 AVR^3 + 0.003 AVR^2 - 2.326 AVR + 4.048 GDP^3 - 1.272 GDP^2 + 11.244 GDP - 558.110 EI - 0.659 AVR * GDP + 5.339 AVR * EI - 38.826 GDP * EI$$

By comparing the coefficients obtained for individual factors in this model, it can be seen that the impact of "GDP per capita" on intellectual migration is higher.

CONCLUSION

In this study, non-linear regression model were developed with the combination of different indices (GDP, Average, Edication index) to evalutaion intellectual migration flows in Azerbaijan . The performance of the developed MNLR model were assessed by statistical performance measures. The results showed that, our model exhibited the best performance, with an R2 value of 0.893, Adj.R2-Squared value of 0.833, RMSE value of 8.544, a MAE value of 6.644, and MSE value of 73.005. The use of digital footprint data allows us to overcome some of the limitations of traditional approaches. The development of this area can contribute to the improvement of existing migration statistics, as well as the creation of new models and the development of new concepts related to population migration. To explore the better predictive skills for migration forecast more sophisticated models will be included in the extension of this research study.

REFERENCES

[1] V Bosetti, C Cattaneo, E Verdolini, "Migration of skilled workers and innovation: A European perspective," Journal of International Economics, vol. 96 (2), pp. 311-322, 2015.

[2] Sanliturk, E. Zagheni, M.J. Danko, T. Theile, and A. Akbaritabar, "Global patterns of migration of scholars with economic development, PNAS, 2023, vol. 120, no. 4. https://doi.org/10.1073/pnas.2217937120

[3] J. El-Ouahi, N. Robinson-García and R. Costas, "Analyzing scientific mobility and collaboration in the Middle East and North Africa." Quantitative Science Studies, vol.2(3), pp. 1023–1047, 2021. https://doi.org/10.1162/qss_a_00149

[4] S. Orazbayev, "International stocks and flows of students and researchers reconstructed from ORCID biographies," MPRA Paper No. 79242, April 6, 2017. [Online]. Available: https://mpra.ub.uni-muenchen.de/79242/

[5] S. J. Porter, "Measuring ORCID Research Information Citizenship, Frontiers in Research Metrics and Analytics," March 2022, vol. 7:779097, www.frontiersin.org, doi: 10.3389/frma.2022.779097

[6] J. Bohannon, K. Doran, "Introducing ORCID," Science, 2017, vol. 356(6339), pp. 691-692. http://dx.doi.org/10.1126/science.356.6339.691

[7] H. Kim, "Global mobility of the recent STEM postdoctoral workforce registered in ORCID," Quantitative Science Studies, December 2024, vol. 6, no. 2., pp.1-23. DOI:10.1162/qss_a_00341

[8] J. Gomez, A. C. Herman and P. Parigi, "Moving more, but closer: Mapping the growing regionalization of global scientific mobility using ORCID," Journal of Informetrics, 2020, vol. 14(3), 101044. https://doi.org/10.1016/j.joi.2020.101044

[9] Subbotin, S. Aref, "Brain Drain and Brain Gain in Russia: analyzing International Migration of Researchers by Discipline using Scopus Bibliometric Data 1996-2020, "Scientometrics, vol.126, pp.875–7900, 2021. https://doi.org/10.1007/s11192-021-04091-x

[10] J. Ghirri, M. Mastropietro, S. Vantini, F. Ieva, M. Fontana, "An Evaluation of Researchers' Migration Patterns in Europe using Digital Trace Data," arXiv:2302.07764v1 [stat.AP] 15 Feb 2023, pp.1-34

[11] Raghupathi, J. Ren, W. Raghupathi, "Exploring the Nature and Dimensions of Scientific Mobility: Insights from ORCID Database - A Visualization Approach," International Journal of Technology Diffusion, vol. 14, iss. 1, pp. 1-31. January 2023. DOI:10.4018/IJTD.331090

[12] ORCID: Connecting research and researchers, https://orcid.org/orcid-search/search?searchQuery=Azerbaijan

[13] GDP per capita (currentUS$)-Azerbaijan, https://data.worldbank.org/indicator/NY.GDP.PCAP.CD?locations

[14] Educational Index, https://globaldatalab.org/shdi/table/edindex/

[15] Gross Average Monthly Wages by Indicator, Country and Year, UNECE Statistical Database, https://w3.unece.org/PXWeb2015/web/

[16] R.I. Eshaa. and M.A. Imteaza, "Non-linear multiple regression analysis for predicting seasonal streamflow using large scale climate mode," 22nd International Congress on Modelling and Simulation, Hobart, Tasmania, Australia, 3 to 8 December 2017, pp. 1586-1589. mssanz.org.au/modsim

Reducing Delays and Optimizing Model in Fog Computing-Based SCADA Systems

Rashid Alakbarov
Institute of Information Technology,
Ministry of Science and Education
Republic of Azerbaijan
Baku, Azerbaijan
t.direktor_muavini@iit.science.az

Mammad Hashimov
Institute of Information Technology,
Ministry of Science and Education
Republic of Azerbaijan
Baku, Azerbaijan
mamedhashimov@gmail.com

***Abstract*—SCADA systems are of great importance in monitoring and safely managing operations in the oil and gas industry. However, these systems face network challenges due to delays in communication channels. Fog Computing technology provides an effective solution to handle these challenges. Processing data at the network edge, explicitly, at points close to the source, increases the response speed of SCADA systems, reduces network load, and enhances security. This article examines the architecture of Fog Computing-based SCADA systems and their implementation possibilities in the oil and gas industry. The proposed mathematical model provides optimal deployment of sensor data in fog devices in Fog Computing-based SCADA systems, upsurges the response speed of the system, lessens network load and balances computing resources.**

Keywords—SCADA systems, FOG computing, Fog Computing-Based SCADA System, Objective Function

INTRODUCTION

SCADA (Supervisory Control and Data Acquisition) systems collect data from various sensors and devices, send this data to a central server or cloud, and ultimately provide system management. Traditional SCADA systems are built on a local infrastructure and rely on physical components such as servers, network equipment, and software. In these systems, data is stored on local servers and SCADA operators can access this data only through an internal network. Setting up such systems usually requires a large initial investment and ongoing technical support. Security is mainly provided by local firewalls and physical security measures. Traditional SCADA systems are ideal for local and stable environments.

Cloud-based SCADA systems operate over the internet and data is stored on cloud servers such as AWS (Amazon Web Services), Microsoft Azure, Google Cloud. These systems can be accessed from anywhere and on any device, which facilitates remote monitoring and management. The cloud-based approach makes it easier and more flexible to scale and update systems. Additionally, it offers a more accessible and cost-effective solution with a monthly service fee model. Cloud-based SCADA is more suitable for modern enterprises with agility, mobile monitoring and analytics needs [1].

However, SCADA systems should operate in real time and analyze data instantly. If the data is first sent to the central server, this can cause delays, decreasing the speed of operations. Fog computing systems have been recently used to overcome the mentioned problems [2]. Fog computing systems refer to a computing model that involves processing data not entirely in the cloud, but on the network edge, on local devices and network devices (i.e., adjacent to sensors). Systems used in the oil and gas industry include thousands of sensors. Each sensor collects data, such as temperature, pressure, gas leakage, etc. Sending all this data directly to the cloud would produce a lot of latency and traffic. Instead, they use nodes located in fog computing systems. These devices are a server or smart device close to the sensors that first processes the data locally, reacts immediately if an abnormality is detected, and only transmits important data to the cloud [3].

Fog computing systems are based on the principle of processing and storing data close to the devices and users. In this approach, data is processed and stored on local devices and networks before being sent to cloud servers. This approach creates opportunities for real-time data analysis and processing. By storing data on local networks before being sent to the cloud, the latency and bandwidth required for data transmission are minimized. Moreover, storing and processing data locally also provides high security, as data is sent to the cloud less often and is better protected on the local network [4].

FOG COMPUTING-BASED SCADA SYSTEMS

SCADA systems used in the automation and monitoring of industrial processes have played an indispensable role in critical areas such as manufacturing, energy, transportation, water supply, and particularly the oil and gas industry for many years. However, since traditional SCADA systems are based on a centralized architecture, they rely on central servers or cloud platforms to process, store, and respond to large amounts of data. This can lead to system delays, data security risks, and overload of network resources. Lately, Fog Computing technology has begun to be integrated into SCADA systems in order to overcome these shortcomings.

Fog Computing is a distributed computing model that provides processing close to the point of data collection, i.e., sensors and control devices, that is, at the "edge of the network". Through this model, data is processed in real time, preliminary analysis is performed, and only important and valuable data is transmitted to the cloud environment. Accordingly, both the system response speed increases and data security and network performance are improved. The advantages of Fog Computing-based SCADA systems are presented below [5, 6].

- *Low latency:* Due to the data processing on local networks, latency is significantly reduced. Data can be analyzed immediately and decisions can be made that have an instantaneous impact on operations. This option

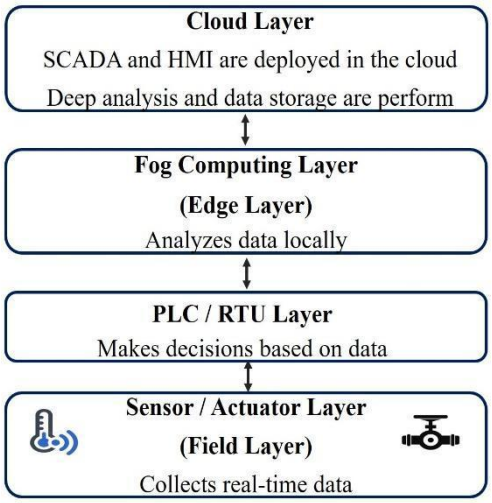
is specifically imperative for systems that are managed in real time.

- *Bandwidth utilization:* The data collected in SCADA systems can be very large, and sending this data to the cloud overloads the bandwidth. Fog computing processes the data locally, sending only the important data to the cloud. This allows for more efficient use of bandwidth.
- *Security and data protection*: While sending data to the cloud poses additional security risks, fog computing ensures data to be processed locally, which intensifies data protection. Since data sharing between users and devices occurs on local networks, the risks of data leakage are reduced.
- *Resource optimization:* Fog computing ensures resource optimization, as the data is processed locally and only the important results are sent to the central system. This ensures more efficient use of server and other network resources.
- *Seamless integration:* Fog computing can be easily integrated into existing SCADA systems. This allows companies to implement new technologies without replacing their existing infrastructure. This approach can be the only and effective way to improve SCADA system performance.

FOG COMPUTING-BASED SCADA SYSTEM ARCHITECTURE

Traditional SCADA systems have been used for decades to automate and monitor industrial processes. However, since the classic structure of these systems is based on centralized data processing and remote control, they face numerous shortcomings. Issues such as delays, network overload, limitations in real-time decision-making and data security come to the fore. These problems are especially important in critical areas such as energy, transportation, oil and gas, and water supply. Along with the development of new technologies, the Fog Computing approach has begun to be integrated into SCADA systems to address these difficulties. The Fog Computing architecture ensures data to be processed locally, close to the sensors, and only the important results are transmitted to central or cloud systems. This approach brings significant advantages to SCADA systems in terms of low latency, local decision-making, efficient use of bandwidth, and security. Fog Computing-based SCADA systems are built on four main functional layers. This architecture creates both a flexible and reliable industrial control environment by regulating the data flow between different technological layers. The main layers of this architecture and their functions are presented below (fig. 1) [7]:

- *Sensor / Actuator Layer (Field Layer)* - collects real-time data (temperature, pressure, vibration, humidity, level, etc.) through physical sensors and executes commands via actuators. It initially transmits data to Programmable Logic Controllers (PLCs) or Remote Terminal Unit (RTUs).
- *PLC / RTU Layer* - Here, PLC devices make decisions according to programmed logic based on the data collected, and the RTU collects data remotely and sends it to the upper layers. This layer ensures fast and local response of the system.
- *Fog Computing Layer (Edge Layer)* - Data is analyzed at the local level through Fog nodes, edge gateways, and micro-servers. Real-time analysis, data filtering, anomaly detection, and local alarms are performed at this stage. The system can operate effectively even without going to the cloud.
- *Cloud Layer (including SCADA/HMI)* - At this layer, SCADA systems control blocks and HMI (Human-Machine Interface) interfaces are deployed in the cloud. Here, operators control the system in real time through web-based panels or mobile applications, send control signals, and all visual monitoring functions are performed.



Fog-based SCADA system architecture

When integrating fog computing into SCADA systems, two components are mainly used: local network devices (edge devices) and fog computing nodes. These components are combined to collect and analyze data, and establish communication between different parts of SCADA systems [8]:

- *Local network devices (Edge Devices)* - perform the initial collection and pre-processing of data. In SCADA systems, this can consist of sensors, actuators, PLCs (Programmable Logic Controllers), and other devices. These devices collect data, perform initial analysis, and send only the important information to local fog devices. This stage is very important as it is the first stage in the data processing, and it reduces the volume of data and transmission time.
- *Fog level (Fog Nodes)* - these devices are more powerful and perform more extensive data analysis. Fog nodes are involved in the control and analysis stages of the SCADA system, process data from various devices in a more extensive way, and then send important data to central servers or the cloud. In this way, operations can be managed faster and more effectively.

OPTIMIZATION MODEL OF DATA DISTRIBUTION FOR FOG COMPUTING-BASED SCADA SYSTEMS

In fog-based SCADA systems, optimal data processing and effective distribution of data (tasks) to devices are of great importance. In these systems, computing power, memory capacity, and response time of each fog device are limited. Therefore, mathematical optimization models are applied to plan tasks in accordance with these resources. The proposed model aims to assign tasks entering the system to fog devices in such a way that the total response time (i.e., the sum of loading and execution times) is minimized. Moreover, restrictions such as the execution of each task on only one device, the memory capacity of fog devices not being exceeded, and the execution of only one task on one device at a time are taken into account. The mathematical model presented in this section formally expresses the optimization objective taking these conditions into account. Along with the model, the definition of variables, the construction of the objective function, and the applied constraints are explained in sequence.

Objective Function Z: It aims to minimize the sum of the times of transferring (loading) to fog devices and executing tasks:

$$min\ Z = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(L_{ij} + C_{ij}\right) \cdot x_{ij}$$

here:

- L_{ij}— loading time for sending the i-th task to the j-th fog device,
- C_{ij} — execution time for processing the i-th task on the j-th fog device,
- x_{ij}— binary variable:
- Here x_{ij} is a binary variable:
- $x_{ij} = \{1, if\ the\ i-th\ task\ is\ sent\ to\ the\ j-th\ fog\ device.\ 0,\ otherwise$

Within the following conditions:

Memory Limit. Since the memory capacity of each fog device is limited, the total memory usage of the tasks placed in the memory of each device should not be greater than a certain limit (M_j).

$$\sum x_{ij} \cdot S_{ij} < M_{j,} \quad \forall i \in \{1, 2, \ldots, n\}, \quad \forall j = \in \{1, 2, \ldots, m\}$$

where:
S_{ij} — the amount of memory used by the i-th task on the j-th fog device,
M_j— the total amount of memory on the j-th fog device.
Task placement: Each task can be executed on only one device. This condition is represented by $\sum x_{ij} = 1$

$$\sum x_{ij} = 1, \quad \forall i \in \{1, 2, \ldots, n\} \quad , \quad \forall j = \in \{1, 2, \ldots, m\}$$

$$x_{ij} \in \{0, 1\}, \quad \forall i \in \{1, 2, \ldots, n\}, \quad \forall j = \in \{1, 2, \ldots, m\}$$

The proposed model creates the following possibilities in the SCADA system by minimizing the sum of the loading and execution times of tasks (data):

- *Optimization in terms of time.* Time losses in both the transmission (loading) and processing (execution) stages of tasks are taken into account and the total delay is minimized. This is important for real-time systems.
- *Optimal distribution of tasks.* The model ensures balancing of the computational load by directing each task to the most suitable fog device. Consequently, situations such as overloading of some devices and leaving others vacant are eliminated.
- *Efficient use of fog device resources.* The overall performance of the system increases as the resource requirements (e.g., memory and computing power) of the tasks to be executed are distributed according to the capabilities of each device.
- *Reduced network load.* Due to the loading time component, the model encourages sending data to fog devices as close as possible, which saves overall traffic on the network.

Since the decision variable in the proposed model is $x_{ij} \in \{0, 1\}$, this is a Binary Integer Linear Programming (BILP) task. Algorithms (Branch and Bound, Genetic Algorithm, Heuristics) are used to solve such problems.

Due to these capabilities, the presented model plays an important role in building intelligent, efficient and flexible fog-based control systems.

EXPERIMENT

This problem is a 0-1 binary integer programming model for the optimal distribution of tasks to fog devices. The goal is to minimize the sum of the task transfer (loading) and execution times. Let's explain the problem step by step below. Let's set up an experiment on the model for 4 tasks (data from sensors) and 4 fog devices (nodes). The goal is characterized as follows:

- Assign tasks to devices,
- Each task must go to only one node,
- Each node receives only one task.
- There are 4 tasks (data) and 4 nodes, and each can only be used once, which means 4! = 24 (permutations)

The objective function must be minimized.

$$min\ Z = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(L_{ij} + C_{ij}\right) \cdot x_{ij}$$

TABLE I. DATA LOADING TIME (L_IJ)

Task/ Node	Node 1	Node 2	Node 3	Node 4
Task 1	3	2	4	6
Task 2	5	3	2	4
Task 3	6	4	3	2
Task 3	4	5	3	1

TABLE II. EXECUTION TIME OF THE DATA (C_IJ)

Task/ Node	Node 1	Node 2	Node 3	Node 4
Task 1	7	5	6	8
Task 2	4	6	3	7
Task 3	5	7	6	4
Task 3	6	4	7	5

TABLE III. SUM OF LOADING AND EXECUTION TIMES (L + C)

Task/ Node	Node 1	Node 2	Node 3	Node 4
Task 1	10	7	10	14
Task 2	9	9	5	11
Task 3	11	11	9	6
Task 3	10	9	10	6

The task (L + C)-node correspondence and the Z-value of the objective function of the full 24 variants are shown in Table 4.

TABLE IV. COST OF THE OBJECTIVE FUNCTION ACROSS OPTIONS

Variant	Task 1	Task 2	Task 3	Task 4	Z-value
1	N1	N2	N3	N4	34
2	N1	N2	N4	N3	35
3	N1	N3	N2	N4	32
4	N1	N3	N4	N2	30
5	N1	N4	N2	N3	42
6	N1	N4	N3	N2	39
7	N2	N1	N3	N4	31
8	N2	N1	N4	N3	32
9	N2	N3	N1	N4	29
10	**N2**	**N3**	**N4**	**N1**	**28**
11	N2	N4	N1	N3	39
12	N2	N4	N3	N1	37
13	N3	N1	N2	N4	36
14	N3	N1	N4	N2	34
15	N3	N2	N1	N4	36
16	N3	N2	N4	N1	35
17	N3	N4	N1	N2	41
18	N3	N4	N2	N1	42
19	N4	N1	N2	N3	44
20	N4	N1	N3	N2	41
21	N4	N2	N1	N3	44
22	N4	N2	N3	N1	42
23	N4	N3	N1	N2	39
24	N4	N3	N2	N1	40

The lowest value is obtained in option 10 (Z=28). Therefore, we need to place the data from the sensors in the nodes corresponding to option 10 so that the value of the objective function is minimal.

Calculation of the objective function (Z) for option 10:

$$Z = (L_{1,2} + C_{1,2}) + (L_{2,3} + C_{2,3}) + (L_{3,4} + C_{3,4}) + (L_{4,1} + C_{4,1}) = 7 + 5 + 6 + 10 = 28$$

CONCLUSION

The article proposed a model that ensures optimal placement of sensor data to fog devices in Fog Computing-based SCADA systems. The model aimed to minimize the overall loading and execution time and effectively distribute tasks (data) taking into account resource constraints. This approach increases the system's response speed, reduces network load, and balances computing resources, ultimately creating a more reliable and efficient control system.

ACKNOWLEDGMENT

This work is supported by the Science Foundation of the State Oil Company of Azerbaijan Republic (SOCAR) (Contract No. 01LR-EF/2024).

REFERENCES

[1] R. G. Alakbarov and M.A. Hashimov. "Migration İssues Of Scada Systems To The Cloud Computing Environment (Review)," SOCAR Proceedings No.3, 2020, pp. 155-164.

[2] R. J. Tom and S. Sankaranarayanan. "IoT based SCADA Integrated with Fog for Power Distribution Automation," 12th Iberian Conference on Information Systems and Technologies (CISTI), 21-24 June 2017, pp. 1-4.

[3] M. A. Hashimov. "Issues of the use of fog technologies in the IoT environment," Problems of Information Technology, No.2, 2020, pp. 80-90.

[4] R. G. Alakbarov and M.A. Hashimov. "Fog Computing Application In Oil And Gas Industry And Analysis Of Cybersecurity Problems," The 8th International Conference on Control and Optimization with Industrial Applications, COIA 2022, pp. 63-65.

[5] R. Das and M. M. Inuwa. "A review on fog computing: Issues, characteristics, challenges, and potential applications," Telematics and Informatics Reports, Vol. 10, June 2023, 100049.

[6] Hazra, P. Rana, M. Adhikari and T. Amgoth. "Fog computing for next-generation Internet of Things: Fundamental, state-of-the-art and research challenges," Computer Science Review, Vol. 48, May 2023, 100549.

[7] M. D. Stojanović, S. V. Boštjančič Rakas and J. D. Marković-PetroviC, "Scada systems ın the cloud and fog envıronments: mıgratıon scenarıos and securıty ıssues," Electronics and Energetics, 32(3), 2019, pp. 345–358.

[8] V. Hurbungs, V. Bassoo and T. P. Fowdur. "Fog and edge computing: concepts, tools and focus areas," International Journal of Information Technology, Vol. 13, 2021, pp. 511-522.

Issues of Constructing an Integrated Fire Hazard Assessment System

Huseyn Bayramov
National Aerospace Agency
Baku, Azerbaijan
huseyn.bayramov0228@gmail.com
0009-0008-4084-5450

Amida Aliyeva
National Aerospace Agency
Baku, Azerbaijan
amida.aliyeva@nasa.az
0000-0001-8858-4917

Sevda Akhmedova
Azerbaijan State University of Economics
Baku, Azerbaijan
axmedovasevda111@gmail.com
0000-0003-3383-643X

Rena Huseynova
Azerbaijan University of Architecture and Construction
Baku, Azerbaijan
renahuseynova55@gmail.com
0000-0002-7686-7503

Hikmat Asadov
National Aerospace Agency, Scientific Research Institute of Aerospace Informatics
Baku, Azerbaijan
asadzade@rambler.ru
0000-0003-1180-1535

Abstract **– The article discusses issues related to the development of an integrated fire hazard assessment system for forest areas based on improvements to the existing FMA and FMA$^+$ indices. The authors analyze the methodological basis of the indices used in Brazil and other countries and propose modifications that allow for more accurate consideration of the influence of meteorological factors, in particular precipitation intensity, relative humidity, and wind speed. The proposed methodology is based on a probabilistic correction of index values depending on discrete precipitation values, which allows for a more accurate assessment of forest fire risk. The task of minimizing the integral fire hazard index using variational calculus methods is formulated. An analytical expression has been obtained that describes the optimal ratio between wind speed and relative humidity at which the fire hazard index is minimal.The developed model allows for the improvement of the forest fire warning system and can serve as a basis for the further development of intelligent systems for monitoring and managing natural risks.**

Keywords - Forest fires, fire hazard, FMA index, FMA$^+$ index, integrated control system, meteorological parameters, variational optimization.

INTRODUCTION

Forest fires, as noted in [1], are a significant factor leading to climate change. Forest fires lead to the emission of about 25-35% of the total volume of CO 2 emitted into the atmosphere, as well as the generation of a significant amount of aerosol. Forest fires cause significant damage to the economy and ecology and can lead to human casualties [2], [3]. Forest fires are characterized by their scale, and most often occur in such countries and regions as Siberia, Mongolia, China, Thailand, Indonesia, Canada, the Mediterranean, Brazil.

In order to prevent and provide early warning of forest fires, integrated forest fire monitoring systems are currently used to collect and process all the necessary information related to the occurrence and course of this natural process. At the same time, the multifactorial nature of forest fires dictates the need to form some generalized indicators, the values of which allow us to assess the degree of risk or probability of their occurrence. Such generalized indicators, in principle, are the methodological basis for building integrated systems for monitoring the occurrence of fires in forest areas. One of such indicators is the *FMA* and *FMA* $^+$ indices described in [4]. The specified fire hazard indices were the methodological basis for building a system for collecting and processing the monitoring of the fire hazard situation in forest areas. As noted in [5-8], such indices include meteorological and climatic indicators such as temperature, relative humidity, wind speed, precipitation.

Index *FMA* (Monte Alegre Formula) was proposed in [9] and is widely used in Brazil. This index includes such indicators as relative humidity and precipitation.

FMA$^+$ index was proposed in [10] and includes, in addition to the above, also wind speed. The purpose of this article is: (1) A brief description of the known methodological base for constructing a control system and the system itself, described in [11]; (2) Improvement of the known methodological base and the fire hazard control system itself, developed on the basis of the improved methodological base.

MATERIALS AND METHODS

As noted in the work [11], the FMA index is calculated using the following formula:

$$FMA = \sum_{i=1}^{n} \left(\frac{100}{H_i}\right), \quad (1)$$

where is H - the relative humidity in percent, measured at 13:00; nis the number of days in which the amount of precipitation did not exceed 13.0 mm.

In this case, the index FMA is accumulated by days, taking into account some restrictive rules for their summation.

The index FMA^+ is defined as follows:

$$FMA^+ = \sum_{i=1}^{n} \left(\frac{100}{H_i}\right) \cdot exp\ exp\ (0{,}04V) \quad (2)$$

where is V – the wind speed.

Depending on the precipitation limits on $i$the day of calculation, formulas (1) and (2) are transformed into the following expressions:

For precipitation amounts (γ) from 2.5 mm to 4.9 mm:

$$FMA_i = 0{,}7FMA_{i-1} + \left(\frac{100}{H_i}\right), \quad (3)$$

where is i - the day of index calculation.

For precipitation amounts from 5.0 mm to 9.9 mm:

$$FMA_i = 0{,}4FMA_{i-1} + \left(\frac{100}{H_i}\right) \quad (4)$$

With precipitation ranging from 10.0 mm to 12.9 mm

$$FMA_i = 0{,}2FMA_{i-1} + \left(\frac{100}{H_i}\right) \quad (5)$$

According to the above, the index FMA^+ in the i - th day is calculated using the following formula:

$$FMA_i^+ = 0{,}7FMA_{i-1}^+ + \left(\frac{100}{H_i}\right)$$

$$(with\ precipitation\ 2{,}5\ \text{мм} < \gamma \leq 4{,}9\ \text{мм}) \quad (6)$$

$$FMA_i^+ = 0{,}4FMA_{i-1}^+ + \left(\frac{100}{H_i}\right)(5{,}0\ mm < \gamma \leq 9{,}9\ mm) \quad (7)$$

$$FMA_i^+ = 0{,}2FMA_{i-1}^+ + \left(\frac{100}{H_i}\right)(10{,}0\ mm < \gamma \leq 12{,}9\ mm) \quad (8)$$

Figure 1 shows a general representation of the fire hazard calculation system using indices $FMA$and FMA^+.

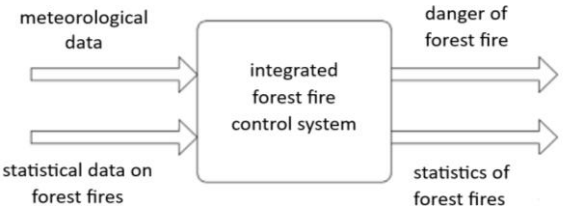


Fig. 1. Conceptual representation of the fire hazard calculation system

Figure 2 shows the flow chart of data collection and processing in the forest fire hazard assessment system.

As can be seen from formulas (3)-(8), in the general case, the calculation formulas FMAcan FMA^+be represented as

$$FMA = \sum_{i=1}^{n} \alpha_i \left(\frac{100}{H_i}\right) + \frac{100}{H_n} \quad (9)$$

$$FMA^+ = \sum_{i=1}^{n} \alpha_i \left(\frac{100}{H_i}\right) exp\ exp\ (0{,}04V_i) + \frac{100}{H_n} exp\ exp\ (0{,}04V_n) \quad (10)$$

where is α_ia random variable that depends on the precipitation level but does not depend on H_i.

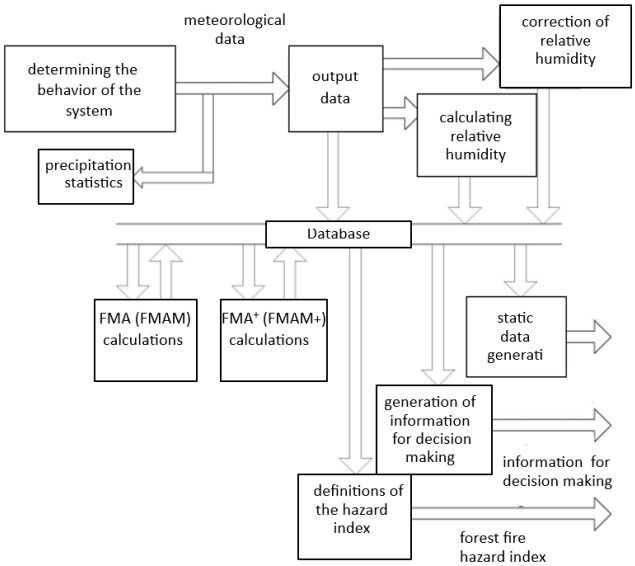


Fig. 2. Flow diagram of data collection and processing in the fire hazard assessment system. The dotted line shows the necessary addition to the diagram, in the form of entering static data on the probability of occurrence of rain intensity within the specified limits

The modification of indices $FMA$is FMA^+carried out on the basis of the following assumptions:

For $n \gg 1$ true

$$\sum_{i=1}^{n} \alpha_i \left(\frac{100}{H_i}\right) \gg \frac{100}{H_n}; \quad (11)$$

$$\sum_{i=1}^{n} \alpha_i \left(\frac{100}{H_i}\right) exp\ exp\ (0{,}04V_i) \gg \frac{100}{H_n} exp\ exp\ (0{,}04V_n)\ . \quad (12)$$

Having designated the primary modified indices as $FMAM$and $FMAM^+$we write

$$FMAM = \sum_{i=1}^{n-1} \alpha_i \left(\frac{100}{H_i}\right) \quad (13)$$

$$FMAM^+ = \sum_{i=1}^{n-1} \alpha_i \left(\frac{100}{H_i}\right) exp\ exp\ (0{,}04V_i)\ . \quad (14)$$

For further modification, taking into account the discreteness of the values α_i, we rewrite formulas (13) and (14) as

$$FMAM = \sum_{j=1}^{n_1} \alpha_1 \left(\frac{100}{H_j}\right) + \sum_{j=1}^{n_2} \alpha_2 \left(\frac{100}{H_j}\right) + \sum_{j=1}^{n_3} \alpha_3 \left(\frac{100}{H_j}\right), \quad (15)$$

where

$$n_1 + n_2 + n_3 = n - 1;\ \alpha_1 = 0{,}7;\ \alpha_2 = 0{,}4;\ \alpha_3 = 0{,}2.$$

n_1- the number of days when condition (6) is true; n_2- the number of days when condition (7) is true; n_3- the number of days when condition (8) is true.

Accordingly (15) $FMAM^+$ we write

$$FMAM^{+} = \sum_{j=1}^{n_1} \alpha_1 \left(\frac{100}{H_j}\right) exp\ exp\ (0{,}04V_j) + + \sum_{j=1}^{n_2} \alpha_2 \left(\frac{100}{H_j}\right) exp\ exp\ (0{,}04V_j) + \sum_{j=1}^{n_3} \alpha_3 \left(\frac{100}{H_j}\right) exp\ exp\ (0{,}04V_j). \quad (16)$$

Having designated the probabilities of implementation $n_1, n_2\ and\ n_3$ accordingly, as $P(n_1); P(n_2)$ и $P(n_3)$ we write

$$FMAM = \sum_{j=1}^{n_1} \alpha_1 P(n_1) \left(\frac{100}{H_j}\right) + \sum_{j=1}^{n_2} \alpha_2 P(n_2) \left(\frac{100}{H_j}\right) + + \sum_{j=1}^{n_3} \alpha_3 P(n_3) \left(\frac{100}{H_j}\right), \quad (17)$$

where $P(n_1), P(n_2), P(n_3)$ the probabilities of occurrence of conditions (6), (7), (8) are known.

Similarly to (17) $FMAM^{+}$ we write

$$FMAM^{+} = \sum_{j=1}^{n_1} \alpha_1 P(n_1) \left(\frac{100}{H_j}\right) exp\ \exp(0{,}04V_j) + + \sum_{j=1}^{n_2} \alpha_2 P(n_2) \left(\frac{100}{H_j}\right) exp\ \exp(0{,}04V_j) + \sum_{j=1}^{n_3} \alpha_3 P(n_3) \left(\frac{100}{H_j}\right) exp\ exp\ (0{,}04V_j) \quad (18)$$

Taking into account that

$$P(n_1); P(n_2); P(n_3) = const \quad (19)$$

We can formulate a minimization problem $FMAM^{+}$. Let us formulate a single minimization problem for all three components of formula (18).

It is required to determine the minimum of the function F, where

$$F = \sum_{j=1}^{n_0} \left(\frac{100}{H_j}\right) exp\ exp\ (0{,}04V_j) \quad (20)$$

with the newly introduced connection function f, where

$$V_j = f(H_j) \quad (21)$$

At the same time, according to statistics

$$\sum_{j=1}^{n_0} f(H_j) = C; C = const \quad (22)$$

To solve this problem, we assume that when $j = \underline{1, n_0}$, the indicator $H_j$makes decisions

$$H_j = H_{j-1} + \Delta H;\ \Delta H = const;\ H_0 = 0, \quad (23)$$

i.e. the set $\{H_j\}$ is ordered.

Based on the above, conditionally assuming a continuous model, we will form the following objective functional $F_0$of unconditional variational optimization:

$$F_0 = \int_{H_{min}}^{H_{max}} P(\alpha) \left(\frac{100}{H}\right) exp\ exp\ [0{,}04 f(H)]\ dH - \lambda \left[\int_{H_{min}}^{H_{max}} f(H) dH - C\right], \quad (24)$$

where is λ - the Lagrange multiplier.

The solution of problem (24) using Euler's method [11] gave expression

$$f(H) = \frac{1}{a_2} ln\ ln\ \frac{\lambda H}{a_1 a_2}, \quad (25)$$

where $a_1 = 100;\ a_2 = 0{,}04$.

Calculation λ using the well-known method [11] gave

$$\lambda = exp\ exp \left[\frac{a_2 \left[C^{-\frac{1}{a_2}} \int_{H_{min}}^{H_{max}} lnln \left(\frac{H}{a_1 a_2}\right) dH\right]}{H_{max} - H_{min}}\right] \quad (26)$$

Taking into account (25) and (26) we obtained:

$$f(H) = \frac{1}{a_2} ln\ ln\ \frac{H}{a_1 a_2} + a_3, \quad (27)$$

where $a_3 = \frac{1}{a_2} \frac{a_2 (C - C_1)}{(H_{max} - H_{min})}$.

$$C_1 = \frac{1}{a_2} \int_{H_{min}}^{H_{max}} ln\ ln \left(\frac{H}{a_1 a_2}\right) dH.$$

The Lagrange test showed that when solving (17) the functional F_0 reaches a minimum.

DISCUSSION

Thus, the known fire hazard indices have been analyzed $FMA$and $FMA^{+}$the properties of these indices are such that rain intensity measurements should be taken daily and the integrated indicator of these indices should be adjusted.

A modification of the specified indices is proposed, where the correction of the indices by the value of rain intensity is carried out taking into account the probability of occurrence of specified discrete values of rain intensity. The current summation of indices in all three groups is carried out with the current sum multiplied by the corresponding correction coefficients, as well as by the known probabilities of occurrence of rains of a specified intensity during the corresponding days. The advantage of such a modification is the possibility of determining such a functional dependence between the wind speed and the current relative humidity, at which the fire danger indices could reach a minimum. It is determined that such a minimum is achieved with a direct logarithmic dependence between the wind speed and relative humidity at a specified limitation on the average annual value of the wind speed.

CONCLUSION

As a result of the research, an improved methodology for assessing fire danger in forests has been developed, based on the modification of the existing FMA and FMA^{+} indices.

The main difference between the proposed approach is the introduction of probabilistic correction coefficients that depend on discrete values of precipitation intensity, which makes it possible to more realistically model and evaluate the behavior of natural factors influencing the occurrence and development of forest fires. Through mathematical modeling and the use of methods of calculus of variations, the optimal functional relationship between wind speed and relative humidity has been determined, at which the integral indicator of fire danger reaches a minimum value. This ratio can be used as a basis for adaptive configuration of fire monitoring and forecasting systems in real time. It is also shown that the

proposed technique is highly flexible and can be easily adapted to different climatic zones and weather scenarios. By integrating statistical data on the probability of precipitation of varying intensity, a more accurate and reliable assessment of the current fire situation is provided. Thus, the results obtained have both theoretical and practical significance: on the one hand, they expand the scientific understanding of the interaction of climatic factors in the context of fire danger, and on the other, they contribute to improving the effectiveness of early warning systems and disaster risk management. The application of the developed model can significantly reduce the economic and environmental damage from forest fires and increase the level of public safety in vulnerable regions.

REFERENCES

[1] https://www.fao.org/4/i3028e/i3028e.pdf
[2] Dimitrakopoulos A., Mitsopoulos I. Thematic report on forest fires in the Mediterranean Region// Forest Fire management working paper 8. FAO. In global forest resources assessment. 43 p.
[3] Koutsias N., Xanthopoulos G., Founda D., et al. 2013. On the relationships between forest fires and weather conditions in Greece from long-term national observations// (1894-2010). International journal of wildland fire 22. Pp. 493-507.
[4] Nunes J.R., Soares R.V., Batista A.C. Prometheus – an integrated wildfire control system// WIT Transactions on ecology and the environment. Vol. 119. 2008.
[5] Soares R.V., Batista A.C. Incendios florestais: controle, efeitos e uso do fogo. Curitiba. 250 p. 2007.
[6] Cheney N.P. Predicting fire behavior with fire danger tables// Australian Forestry. V. 32. n. 2. P. 71-79. 1968.
[7] Schroeder M.J., Buck C.C. Fire weather// USDA forest service. Agriculture handbook 360. Pp. 229. 1970.
[8] Heikkila T.V., Gronovist R., Jurvelius M. Handbook on forest fire control// Helsinki, forestry training programe publication 21. P. 239. 1993.
[9] Soares R.V. Determinaçao de um indice de perigo de incendio para a regiao centro paranaense// Brasil, Turrialba. Costa Rica. CATIE/IICA. P. 72. 1972.
[10] Nunes J.R.S. FMA+ um novo indice de perigo de incendios Florestais para o estado do Parana – Brasil// Curitibaş 2005. 150 p. Setor de ciencias agrarias, universidade federal do parana.
[11] Эльсгольц Л.Э. Дифференциальные уравнения и вариационное исчисление// М. Наука. 1974. Стр. 432. (Elsgolts L.E. Differential equations and calculus of variations// M. Nauka. 1974. P. 432.)

Section 10: Cybersecurity Methods

Comparative Analysis of Global Cybersecurity Indices in the Context of the Formation of Cyber Sovereign States

Rasim Alguliyev
Instıtute of Informatıon Technology
Baku, Azerbaijan
r.alguliev@gmail.com

Babek Nabiyev
Instıtute of Informatıon Technology
Baku, Azerbaijan
Babek.nabiyev@gmail.com

Konul Dashdamirova
Instıtute of Informatıon Technology
Baku, Azerbaijan
konulahmed@gmail.com

***Abstract* — In the modern era, the deep penetration of digital technologies into all spheres of human activity has made the cybersecurity an integral component of the security policy of states. States use global cybersecurity indices and assessment methods to evaluate their cybersecurity status, identify existing gaps, and define strategic priorities. The thesis examines the mechanisms for assessing the cybersecurity posture of countries, and emphasizes the importance of international indices and assessment methods in the field of cybersecurity. Key international assessment frameworks, including the Global Cybersecurity Index, National Cybersecurity Index, Cybersecurity Capacity Maturity Model, National Cyber Sovereignty Index is analyzed. The impact of cyber sovereignty on national security, information independence, and digital governance are analyzed. It is recommended to create national cybersecurity indices in accordance with international standards based on strategic documents in the field of digital development and cybersecurity in Azerbaijan.**

Keywords—cybersecurity, national security, digital sovereignty, global cybersecurity indices, national cybersecurity indices, assessment methods

I. INTRODUCTION

Today, due to rapid digitalization, cybersecurity issues have become one of the priority areas in the security strategies of states. The widespread application of information technologies and the integration of various spheres of state and public life into digital platforms necessitate the strengthening of the defense capabilities of global and national information infrastructures against cyber threats [1]. There is a need to implement comprehensive measures to prevent cyber threats at both national and global levels and ensure the sustainability of information infrastructure.

Ensuring the cyber sovereignty of countries is of particular significance for maintaining both national and global cybersecurity. Cyber sovereignty refers to the ability of a state to independently manage information flows, digital resources and technological infrastructure within its cyberspace as well as to protect it from external influences. This is not limited to technical security measures, but also includes legal, institutional and strategic approaches. Cyber sovereignty, in addition to ensuring the state's full control over the digital domain, functions as a fundamental component of national security and information independence [2].

Under such circumstances, various international and national cybersecurity indices have been developed to assess the current state of cybersecurity in countries. These indices measure the legal frameworks, institutional capacities, technical capabilities, and levels of cooperation of states, identify existing gaps, and assess strategic development directions.

Currently used international and national cybersecurity indices apply a variety of criteria such as technical, legal, organizational, international cooperation, scientific research. These criteria include components such as the existence of a national cybersecurity policy, regulatory and legal framework, technological defense infrastructure, intelligent cyber threat analysis and monitoring systems, awareness and training programs, institutional response mechanisms, and the level of international cooperation [3]. Such a comprehensive and structured assessment approach enables an objective determination of the current cybersecurity status in countries, facilitates comparative analysis, and guides the identification of development priorities.

II. EXISTING GLOBAL CYBERSECURITY INDICES AND THEIR COMPARATIVE ANALYSIS

Cybersecurity indices are metrics used to assess a country's cybersecurity preparedness, strategic capabilities, resilience against risks, and protection of digital infrastructure. Currently, several indices and models are used to assess the current state of cybersecurity in countries at the global and national levels. The main assessment indices are as follows [4]:

- Global Cybersecurity Index (GCI);
- National Cybersecurity Index (NCSI);
- Cybersecurity Capacity Maturity Model (CMM);
- EU Cybersecurity Index;
- National Cyber Sovereignty Index.

A. Global Cybersecurity Index (GCI);

GCI is a global index developed by the International Telecommunication Union (ITU) and used to assess countries' preparedness, commitments, and capabilities in the field of cybersecurity. The GCI provides an opportunity for states to improve their national cybersecurity policies. It promotes the global exchange of knowledge and experience, and also serves to increase resilience to cyber threats and strengthen the level of digital security. The assessment process is conducted at five primary levels:

- Legal measures: The presence of a national strategy, laws, regulatory and legal framework in the field of

cybersecurity, as well as the degree of compliance with international cybersecurity standards;

- Technical measures: The level of protection of critical infrastructure (energy, transport, finance, etc.), state information systems, the presence of technical infrastructure such as Computer Emergency Response Team/Computer Security Incident Response Team (CERT/CSIRT), Security Operations Center (SOC) and Security Information and Event Management (SIEM);
- Organizational measures: The presence of state bodies and organizations responsible for cybersecurity, the level of cooperation between the state and the private sector, the effectiveness of the national CERT group in responding to cybersecurity incidents;
- Human resources: The presence of educational programs for training cybersecurity specialists cybersecurity specialists, the level of awareness of the population in the field of cybersecurity, and the implementation of cybersecurity training sessions. Conducting scientific research in the field of cybersecurity, allocating investments for the development of cybersecurity technologies, deployment of innovative cybersecurity solutions;
- International cooperation: The level of cooperation with international cybersecurity organizations, participation in global cybersecurity trainings and events, the presence of international treaties and agreements in the field of cybersecurity.

The assessment is conducted based on a survey. Each survey contains 83 questions covering 20 indicators that allow for a comprehensive assessment of the country's cybersecurity landscape. The questions are developed by ITU and distributed to the national representative bodies of member states. These bodies are the official structures responsible for cybersecurity within the country. The survey is completed through an online platform and countries are required to provide evidence with relevant documents for each answer.

A maximum of 20 points can be awarded for each level, with a maximum of 100 points across all 5 levels. This approach allows countries to identify their strengths and weaknesses in cybersecurity and to make comparisons at a global level [5].

The country's total score for each criterion is calculated using the following formula:

$$\mathrm{GCI}_p = \left(\frac{\sum_{i=1}^{n}(q_i \times w_{i)}}{\sum_{i=1}^{n} w_{i)}}\right) \times 20 \qquad (1)$$

Here:

$\mathrm{GCI_p}$ – GCI score for each criterion, q_i – the score assigned to the response for each question related to the criterion (can be 0, 0.5 or 1), w_i – the weight coefficient determined for that question, n – the number of questions under the corresponding criterion.

Countries' cybersecurity levels are categorized according to the following GCI score ranges:

95–100 points: countries with high-level legislation, developed technical infrastructure, international cooperation and regular training programs. For instance, the USA, Great Britain, Singapore, France, Canada;

85–95 points: countries with a well-established legal framework, technical defense systems and international cooperation, but with gaps in some areas. For instance, Spain, UAE, South Korea, India;

55–85 points: countries with progress in the field of cybersecurity, but legislative and technical measures are not fully finalized. For instance, Brazil, South Africa, Vietnam;

20–55 points: countries with poorly developed legal and technical infrastructure, limited awareness-raising and international cooperation. For instance, Nigeria, Kyrgyzstan, Bolivia;

0–20 points: countries with virtually no cybersecurity policies, legislation, or technical measures. For instance, Somalia, Afghanistan, Haiti.

B. National Cybersecurity Index (NCSI)

NCSI was developed by the Estonian e-Governance Academy (eGA) in 2018 [6]. It is one of the most comprehensive mechanisms for assessing countries' cybersecurity policies, the security level of digital services, and their ability to respond to threats. The assessment process is based on 49 indicators across 12 areas, which are grouped into three main categories (Table 1):

TABLE 1. NCSI AREAS

Strategic Opportunities	Cyber Threat Prevention Opportunities	Cyber Threat Response Opportunities
1. Cybersecurity policy development 2. Global collaboration 3. Education and professional development 4. Scientific research	5. Protection of essential services 6. Ensuring digital technological capabilities 7. Threat analysis and information 8. Protection of personal data	9. Cyber Incident Response 10. Cyber Crisis Management 11. Cyber Crime Prevention 12. Military Cyber Defense

NCSI is calculated as follows:

$$\text{NCSI Score} = \left(\frac{\text{Country Points} \times 100}{\text{Maximum Points}}\right) \qquad (2)$$

NCSI Score is the overall indicator of the country's national cybersecurity index. Country Points represent the total points a country has collected for the existing areas of cybersecurity, while Maximum Points refer to maximum possible number of points provided for in the index.

$$\text{Country Points} = \sum_{i=1}^{n}(V_i \times W_i) \qquad (3)$$

Vi — the value of the *i*-th indicator (0 – not relevant, 0.5 – partially relevant, 1 – fully relevant), *Wi* — the weight coefficient of the *i*-th indicator (indicates the importance of that indicator in the index), *n* — the total number of indicators (in the NCSI this is 49

$$\text{Maximum Points} = \sum_{i=1}^{n} W_i \qquad (4)$$

Wi — weight coefficient of the i-th indicator, *n* — total number of indicators.

The data used in the NCSI is primarily collected from the following sources:

- Data provided by country government officials;
- Data provided by various organizations or individuals;
- Data that is publicly available on the Internet and collected by the NCSI team.

The accuracy of the information provided for the calculation of each indicator must be verified. This evidence includes laws, decisions and other legal acts covering regulatory and legal documents, official documents approved by the state, including strategies and action plans, as well as reliable information posted on the websites of the government or relevant official bodies. Only evidence based on official and legal grounds is accepted as the primary source for assessing the indicators.

C. Cybersecurity Capacity Maturity Model (CMM)

CMM – first developed by the Global Cyber Security Capacity Centre (GCSCC) at the University of Oxford in 2014 and later refined in collaboration with several international partners (UN, World Bank, ITU, CTO, etc.) [7].

The CMM model serves to identify countries' current capabilities and weaknesses in the field of cybersecurity, support the development of national strategies and determine long-term development directions. The model assesses cybersecurity according to five main criteria. These criteria encompass all aspects necessary for a country to operate effectively in the field of cybersecurity:

- Development of cybersecurity policy and strategy;
- Promotion of a responsible cybersecurity culture in society;
- Development of cybersecurity knowledge and skills;
- Establishment of effective legal and regulatory legislation;
- Management of risks through standards and technologies.

The structure of the CMM is as follows (Figure 1):

1) Dimension: Five main criteria collectively reflect a country's national cybersecurity potential, which is assessed based on the CMM model. Each criterion consists of certain factors.

2) Factor: The factors included within the criteria define the key components of a country's cybersecurity potential.

3) Aspect: When factors consist of multiple components, they are further divided into aspects. The number of aspects depends on the diversity and complexity of the topics included in the factor.

4) Stage: Levels indicate the degree of development achieved by a country in the field of cybersecurity in a particular factor or aspect. Each aspect or factor is assessed based on indicators in five stages:

- Start-up;
- Formative;
- Established;
- Strategic;
- Dynamic

5) Indicator: Indicators are the most basic structural units of the CMM and describe the activities and processes that fulfill the requirements of a specific level A country must provide evidence for all indicators to achieve a certain level. Most indicators are binary in nature - a country either meets the criteria or does not meet them. Achieving a higher level of cybersecurity is not possible without meeting all indicators.

D. EU Cybersecurity Index

The EU Cybersecurity Index is a set of indicators designed to assess the level of cybersecurity preparedness and capabilities of member states. The index is based on various indicators. Its main objectives are to identify the cybersecurity strengths and weaknesses of EU member states in the field of cybersecurity, assess the adequacy of national and cross-border security measures, strengthen the protection of digital infrastructures, and formulate cybersecurity policies based on the recommendations of ENISA (the European Union Agency for Cybersecurity) and other organizations [8].

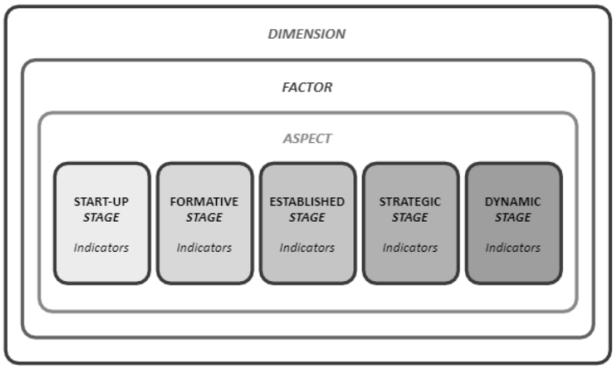


Fig.1. Structure of CMM

The EU Cybersecurity Index is a comprehensive, hierarchical index comprising 84 qualitative and quantitative indicators, structured in 4 main areas and 16 sub-areas. These areas are as follows:

- Operations;
- Market / Industry;
- Policy;
- Capacity.

Of the 84 indicators:

- 60 indicators are collected from the relevant authorities of the Member States through ENISA and aggregated at the European Union level.
- 24 indicators are aggregated across the EU.

The indicators for each sub-area are calculated as follows:

$$SubArea_i = \frac{\sum_{j=1}^{n} w_j \cdot I_j}{\sum_{j=1}^{n} w_j} \quad (5)$$

Where:

I_j— the value of the jth indicator (in the range 0-100), w_j— the weight of that indicator, n — the number of indicators included in the subfield.

For each main area (e.g. Policy, Operations, etc.), the sub-areas are calculated as follows:

$$Area_k = \frac{\sum_{i=1}^{m} w_i \cdot SubArea_i}{\sum_{i=1}^{m} w_i} \quad (6)$$

where:

$SubArea_i$— the value of the ith subarea, w_i— the weight of the subarea, m— the number of subareas in the area.

The index is found by calculating the average price across all areas:

$$Overall\ Index = \frac{1}{4}\sum_{k=1}^{4} Area_k \quad (7)$$

E. National Cyber Sovereignty Index

In today's borderless and complex cyber threat environment, states around the world are striving to maintain the initiative in cyberspace. Accordingly, states should assess their cyber sovereignty to determine how effectively they control the security level of cyberspace used in both their domestic and foreign activities [9].

A nation's cyber sovereignty is determined by various factors. These factors include the degree of dependence on foreign technologies, the availability of national cyber resilience capabilities, the political context, the level of institutional maturity of stakeholders, the structural complexity of the cyber governance ecosystem, and other key indicators. All of these factors directly influence the dynamics of cyber sovereignty development and its potential for improvement.

The National Cyber Sovereignty Index (NCSI), developed by Thales Group, is an assessment tool designed to measure the level of cyber sovereignty of states. The NCSI offers a more specific and targeted approach, allowing countries to assess their current state of cyber sovereignty and identify priority areas for development in this field.

While the exact formula used by Thales to calculate the index remains undisclosed, the assessment is known to be based on six core areas:

- National Cybersecurity Strategy - the existence and implementation of a national strategy in the field of cybersecurity.
- Education and training strategy - the development of educational programs and professional training in the field of cybersecurity.
- National Research and Technology Strategy - support for research and development in the field of cybersecurity.
- Strengthening corporate cyber culture - increasing business awareness and responsibility in cybersecurity issues.
- Sovereign solutions and technologies - the development and implementation of national technological solutions, reducing dependence on foreign suppliers.
- State technical capabilities - the availability of the state's necessary technical resources and infrastructure to ensure cybersecurity.

Each of these areas is assessed against several criteria, such as the availability of regulations, specialized institutions, training programs, and technical solutions. The assessment is carried out by Thales experts based on existing data and information provided by the countries.

Although international and national assessment methods are widely used, each has advantages and challenges.

Table 2 presents a comparative analysis of the Global Cybersecurity Index, National Cybersecurity Index, Cybersecurity Capacity Maturity Model EU Cybersecurity Index, and National Cyber Sovereignty Index.

TABLE 2. COMPARATIVE ANALYSIS OF GLOBAL CYBERSECURITY INDEX, NATIONAL CYBERSECURITY INDEX, CYBERSECURITY CAPACITY MATURITY MODEL, EU CYBERSECURITY INDEX AND NATIONAL CYBER SOVEREIGNTY INDEX

Features	GCI (ITU)	NCSI (eGA, Estonia)	CMM (Oxford, GCSCC)	EU CSI (ENISA)	NCSI (Thales Group)
Preparing organization	ITU	Estonian e-Governance Academy (eGA)	Oxford University	ENISA	Thales Group
Scope	Global 180 countries	160 countries globally	Global practice-based (50 countries)	EU countries (27 member states)	Global
Methodology	Survey, public information	Open source data	In-depth quality assessment	Weight-based, numerical average, indicator system	Undisclosed
Assessment model	Score system in 5 key areas (0–100)	Open scores in 12 areas (0–100)	5 levels	4 domains, 16 subdomains, 84 indicators (0–100)	6 areas
Data sources	Surveys documents, open source	Government data	Field meetings and reports	Member state surveys, statistical documents	Government documents, public data, analytical reports
Frequency of updates	Every 2–3 years	Constantly updated open platform	Project-based	Phase-by-phase implementation and periodic updates	Not clearly stated
Social implications	Global ranking	Country profiles and rankings	Full report only on request	EU country indicators and average score	Global ranking
Purpose of usage	Policy recommendations, comparison	Transparency, national security development	Political	EU coordination	Measuring and comparing the level of cyber sovereignty of states

The results of the analysis show that the GCI is one of the most widely used indices worldwide for measuring the cybersecurity preparedness of states. By covering five key areas, it allows for a structured analysis of the state's

cybersecurity ecosystem. Its support by the UN and its reliance on official government data increase the legitimacy of the results. However, the main part of the assessment is based on self-assessment data provided by countries, which makes the results subjective. In-depth analysis of technical dimensions and real operational capabilities is not sufficiently conducted, and technological development indicators are poorly covered. Additionally, the frequency and update interval of reports may be insufficient for operational implementation.

NCSI is a platform based on official documents, as well as open source, and is updated in real time. It is ideal for transparency and rapid comparison. The rationale for each indicator used is publicly accessible. The results are open and subject to intense public scrutiny. However, technical measures and operational security aspects are not covered in sufficient depth. This creates a gap in the analysis of technological risks. Legal indicators are difficult to adapt to the national legal systems of some countries, and diversity can create inaccuracies in the context. Indicators are mainly assessed using a "yes/no" approach, which does not consider the degree of quality and effectiveness.

The CMM model is based on in-depth qualitative analysis and a structured assessment methodology. The phased approach provides a structured approach for developing a long-term development strategy. The complex methodology based on interviews, focus groups and document analysis provides more realistic and contextual results. However, the assessment process requires high expertise and time, which limits its widespread application. Subjective approaches between experts can weaken the reliability and comparability of the results. The data are mainly based on on-site assessments, making it challenging to apply in resource-poor countries.

The EU CSI specializes in monitoring legal compliance and strategic measures within the EU, particularly monitoring the implementation of directives such as NIS2 and DORA. The quarterly updated indicators are suitable for rapid feedback. It includes separate indicators measuring the integration of industry and the public sector. However, since it only covers EU and EEA member states, it is not universally applicable. While some indicators reflect regulatory compliance, they do not assess operational security and technological aspects. Since the data is often based on self-reports by countries, the objectivity of the analysis can be questioned.

The Thales Group National Cyber Sovereignty Index is a unique and strategically important indicator for assessing digital sovereignty. This index is important for assessing geopolitically significant areas such as digital autonomy, dependence on foreign technologies and domestic technological capabilities. It is one of the few models that assess the security capabilities of the private sector. However, the concept may conflict with Western democracies regarding ethical and normative compliance, and carries risks of suppressing freedom of information. The number of indicators and the methodology are not fully disclosed to the public, which limits scientific and practical application. An approach based on a certain degree of state control may raise issues of transparency and human rights.

CONCLUSION

The conducted research shows that the global cybersecurity environment is undergoing dynamic development, and international indices based on different methodologies are widely used to objectively assess the state of countries in this area. The main assessment tools including the Global Cybersecurity Index (GCI), National Cybersecurity Index (NCSI), Cybersecurity Capacity Maturity Model (CMM), EU Cybersecurity Index and National Cyber Sovereignty Index allow assessing the technological, legal, institutional and strategic capabilities of the cyber environment of countries from various aspects.

While each of these indices relies on a specific methodological framework, information sources, and measurement criteria, the general goal is to systematically analyze the cybersecurity ecosystem of countries, identify weak points, and form a roadmap for development directions. Contemporary approaches show that it is impossible to be satisfied with just technical measures. It is necessary to form a comprehensive approach to this issue, considering aspects such as science, law, education, international cooperation, and national security.

The main innovative contribution of the article is that, based on a comprehensive comparison of international indices, the idea of developing a local cybersecurity index that considers the national characteristics and regional challenges of Azerbaijan is put forward. Such an index could serve as an important tool for assessing the country's level of digital sovereignty, information security, and technological resilience.

Consequently, a national cybersecurity index developed through the scientific analysis of international practices and its adaptation to local realities will enhance the effectiveness of Azerbaijan's information security, strengthen its position in international rankings, and support its transition toward a cyber-sovereign state model.

REFERENCES

[1] R. Alguliyev, B. R. Nabiyev, K. Dashdamirova, "Cyber threats and their intellectual analysis issues in the context of technological challenges of the IV Industrial Revolution," 2023 IEEE 17th International Conference on Application of Information and Communication Technologies (AICT), IEEE, 2023, p. 1-6.

[2] Voronenko, M. Nehrey, A. Laptieva, et al. "National cybersecurity: assessment, risks and trends," International Journal of Embedded Systems, 2022, T. 15, №. 3, pp. 226-238.

[3] K. Farahbod, C. Shayo, J. Varzandeh, "Cybersecurity indices and cybercrime annual loss and economic impacts," Journal of Business and Behavioral Sciences, 2020, T. 32, №. 1, pp. 63-71.

[4] R. Bruggemann, P. Koppatz, M. Scholl et al. "Global cybersecurity index (GCI) and the role of its 5 pillars," Social Indicators Research., 2022, pp. 1-19.

[5] Global Cybersecurity Index. https://www.itu.int/en/ITU-D/Cybersecurity/pages/global-cybersecurity-index.aspx (accessed May 24, 2025)

[6] NCSI, https://ncsi.ega.ee/indicators/ (accessed May 21, 2025)

[7] Cybersecurity Capacity Maturity Model for Nations. (CMM) https://www.nrdcs.eu/cybersecurity-capacity-maturity-model-for-nations-cmm/ (accessed June 04, 2025)

[8] EU Cybersecurity Index. https://www.enisa.europa.eu/topics/state-of-cybersecurity-in-the-e (accessed June 01, 2025)

[9] Cyber Sovereignty. https://www.thalesgroup.com/en/markets/defence-and-security/cyberdefence-solutions/cyber-sovereignty (accessed June 01, 2025)

Cloud Cyber Attack Images Classification using GAN and ViT+ML Algorithms

Fargana Abdullayeva
Department of Information Security
Institute of Information Technology
Baku, Azerbaijan
a_farqana@mail.ru
https://orcid.org/0000-0003-2288-6255

Gurban Farajov
Department of Information Security
Institute of Information Technology,
Baku, Azerbaijan
farajovqurban@gmail.com

***Abstract*—The emergence of the Industry 4.0 concept and the development of modern technologies have made the detection of cyber attacks in cloud systems an important issue. In the article, a hybrid model based on the combination of machine learning algorithms with Generative Adversarial Networks (GANs) was developed to identify various attack categories targeting cloud systems. In the model, the integration of functions that enhance image quality within the GAN algorithm significantly improved classification performance by increasing the quality of cyber attack images. Here, the damage in the images is repaired, and their appearance is restored and generated to resemble the original as closely as possible. To enhance the model's robustness against various changes in input images, during the data augmentation phase, the process of rotating images and generating them in different variations was also carried out using GAN. The proposed method classified various cyber attacks on cloud systems more effectively than existing methods, achieving a classification accuracy of 0.9451.**

Keywords—Deep learning, GANs, CatBoost, Cyber Attacks, vision transformer (ViT)

INTRODUCTION

Cloud systems are one of the main technological pillars enabling the digitalization and automation of industrial enterprises within the framework of Industry 4.0. These systems play a crucial role in the processing of big data, the application of artificial intelligence, and the optimization of production processes [1]. Accessibility from anywhere, scalability, multi-tenancy, and the ability to process complex tasks in parallel are among the advantages of cloud systems [2]. Cloud systems are widely used in the finance, military, healthcare, education, energy, transportation, and e-government sectors [3]. The widespread use of cloud systems across various fields has exposed them to different types of cyber attacks. In cloud systems, the main cyber attacks occur in components such as the network, infrastructure, software, access and account management, and data protection. The realization of these cyber attacks can result in the destruction of target assets and cause significant economic damage. Timely detection of these cyber attacks is considered a critical issue.

There have been real cyber attack incidents targeting cloud systems. On February 21, 2018, hackers gained access to Tesla's Amazon Web Services (AWS) cloud account and used the service (a computer operating on the internet) to mine cryptocurrency [4]. This incident occurred due to configuration errors in the cloud systems, lack of monitoring, and absence of mechanisms for detecting anomalies. In 2014, the account information of the company Code Spaces was compromised on AWS [5]. The attackers who took control were able to delete all customer data belonging to Code Spaces. This incident occurred due to the absence of a multi-factor authentication mechanism. In 2025, the SaaS service hosted on the Microsoft Azure platform by Commvault, a company operating in the field of data protection, was affected by cyber attacks exploiting a zero-day vulnerability [6]. The attackers gained unauthorized access to customers' confidential data and took control of their SaaS services. In 2023, as a result of a cyber attack on Ukraine's huge mobile network Kyivstar, customer access to telephone station and the internet network was disrupted, and the cloud storage and backup systems were destroyed [7].

There are various types of DoS cyber attacks that target the services and resources of cloud systems [8]: DOS Golden, DOS Hulk, DOS Slow, DOS slowloris. These attacks primarily target the availability, service level agreement (SLA), and performance of the cloud. In addition, cyber attacks such as FTP Patator, SSH Patator, and web attack brute force, which are aimed at capturing account login passwords on cloud platforms, are also common [9]. Another type of attack is port scanning. Hackers carry out this attack to identify potential vulnerabilities in the cloud and gather information about open ports. Cyber attacks such as Heartbleed, infiltration, SQL injection, and XSS are also threats that can be executed in cloud systems.

Numerous approaches have been developed for detecting cloud cyber attacks. Compared to traditional methods based on the analysis of conventional features, image-based methods are more effective in detecting cyber attacks [10]. These methods can detect spatial dependencies and contextual relationships with high accuracy, which are often overlooked by traditional feature engineering techniques when identifying cyber attacks. Spatial dependencies refer to how closely pixels are positioned relative to each other. Contextual relationships refer to how different parts of an image are positioned in relation to one another. Existing approaches first segment the images to detect cyber attacks. However, during the segmentation process, useful parts of the image may be cut out and excluded from the analysis. This reduces the confidence in the effectiveness of cyber attack detection systems. In smart systems like the cloud, methods for detecting cyber attacks based on image analysis should include components such as self-attention mechanisms, filtering, and restoration of blurred images.

Related Works

This section analysis works related to the topic. In [10] a Gaussian mixture model (GMM) based probabilistic model was proposed for detecting malware based on image analysis. In this approach, malware is detected by evaluating the similarity of images. Primarily, changes made in the images are identified. In [11], the DDoSViT approach was proposed for detecting cyber attacks targeting Internet of Things (IoT) devices. The proposed multi-vector DDoS and DoS attack detection approach based on ViT converts attack flows into images and trains ViT on the attack image dataset. The CICIoT2023 and CICIoMT2024 datasets were used for conducting experiments. The effectiveness of the model was evaluated based on accuracy, precision, recall, and F1-score metrics. In [12], gray-scale images were analyzed for malware detection in GCS (Grid Computing Systems). The proposed hybrid model consists of ResNet-50 and Support Vector Machine (SVM). ResNet-50 is used to extract relevant features from the images, while SVM classifies the malware based on the extracted features. In [13], a hybrid method based on multi-objective optimization was proposed, allowing the use of the most efficient features of lightweight deep learning models for cyber attack detection. First, QR code images are generated from large volumes of data with many classes. Then, QR code images are trained on neural networks using CNN models such as MobileNetV2 and ShuffleNet. Features are extracted from the trained images, and the Harris Hawk Optimization (HHO) algorithm is used to select the most effective features for classification purposes. In [14], a supervised machine learning approach was proposed for cyber attack detection. Random Forest, K-Nearest Neighbors (KNN), and Logistic Regression algorithms were used to classify the data into normal and cyber attack categories. In [15], a method for detecting malware based on resource consumption metrics using transformers was proposed. Here, the input data is encoded as a sequence of processes, and each process is described by its resource consumption metrics (CPU, memory, and disk usage). In [16], the application of ViT for detecting cyber threats related to malware and network intrusions was examined. To address this problem, a deep neural network called VINCENT (ViT-based distillation for Cyber-Threat detection) was proposed, which converts cyber threat data into vectors of color images. The ViT block of this model is trained on images to extract visual features from the data for each class. In [17], a ResDNViT approach was proposed by combining ViT and ResNet models for detecting NetFlow-based cyber attacks. In the ResDNViT model, the ViT-based architecture analyzes network traffic by representing NetFlow features as 2D matrices and dividing them into equal-sized submatrices, which serve as input fragments for the encoder component. In [18], a deep learning model was developed for detecting various types of cyber attacks. The algorithm emphasizes the importance of feature selection, and the significance of attention mechanisms for improving feature evaluation within the same model is analyzed. In [19], the application potential of transformers and large language models (LLMs) in cyber threat detection systems was investigated. The fundamentals of transformers, various types of cyber attacks, and the datasets used in this field were discussed. The study explored the use of attention-based models, BERT-type LLMs, GPT, CNN/LSTM-transformer hybrids, and hybrid models such as ViT in intrusion detection systems.

Proposed Method

The architecture of the proposed approach for detecting cyber attacks in cloud systems is illustrated in Figure 1.

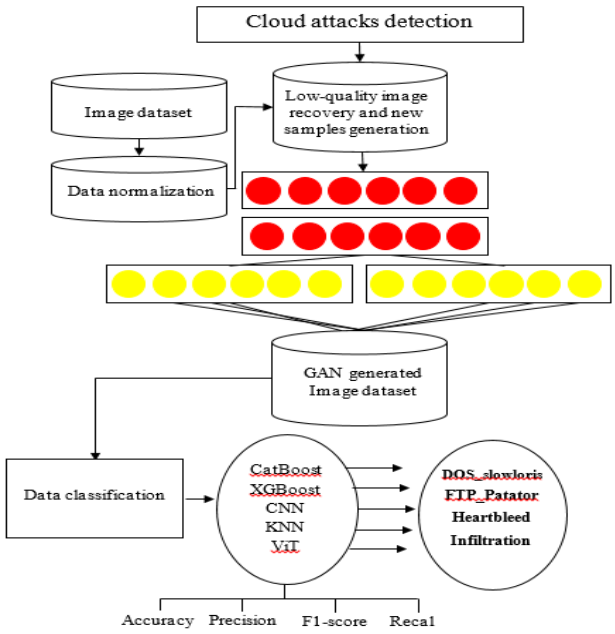


Fig 1. Cloud attacks detection approach

As seen in Fig. 1, the graphical abstract illustrates a hybrid approach that combines GANs with machine learning techniques. The approach consists of the following stages:

In the first stage, GANs are used to restore blurred images and generate new samples based on high-quality images. Subsequently, GANs are also applied to address the class imbalance problem.

In the second stage, feature extraction was performed on the dataset. Each image was first resized to (32, 32, 1), and then converted into a 1024-dimensional vector consisting of pixel intensity values.

In the third stage, the feature vectors composed of pixel values were fed into classifiers to categorize the cyber attack data into different classes.

The operations performed on the data significantly improved the detection accuracy of the proposed approach for cyber attacks. To evaluate its effectiveness, GAN+CatBoost, GAN+XGBoost, and GAN+KNN models were developed for detecting cyber attacks in cloud environments. Compared to traditional methods, the proposed approach demonstrated superior performance.

Experiments

For the experiments the Cloud Attack Dataset was used [16]. The Cloud Attack Dataset was created based on the CIC-IDS 2017 dataset from the Canadian Institute for Cyber security. A total of 100,541 traffic samples were extracted from the dataset. These samples belong to 14 traffic classes: one class represents benign traffic, while the other 13 represent attack classes. The network samples were converted into 9x9 sized traffic images. The

experiments were carried out on the Linux operating system using the Python 3.9 programming package. Several image samples from the dataset are shown in Figure 2.

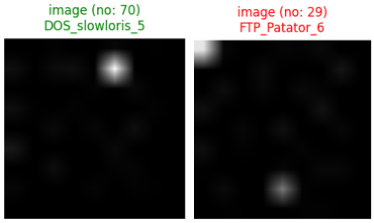


Fig. 2. Cloud Attack Dataset image samples

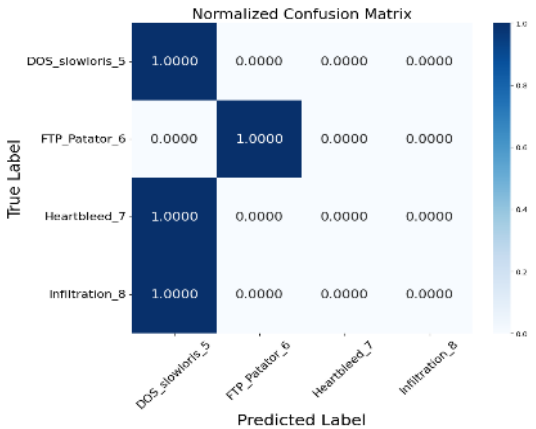
The effectiveness of the methods was evaluated based on accuracy, precision, recall, and F1-score metrics, and the obtained results are presented in Table 1.

Table 1. Comparative analysis of the methods based on the Recall metric

Method / Class	CatBoost	GAN+ CatBoost	XGBoost	GAN+ XGBoost	KNN	GAN+ KNN
DOS Slowloris	1.0000	1.0000	0.9815	0.9929	0.9940	0.9965
FTP Patator	1.0000	1.0000	1.0000	1.0000	0.4444	0.9648
Heartbleed	0.0000	0.9852	0.0000	0.9925	0.0000	0.9852
Infiltration	0.0000	0.9754	0.5000	0.9520	0.0000	0.9836

It should be noted that there is a significant class imbalance in the Cloud Attack Dataset. For instance, the DoS Slowloris class contains 635 samples, the FTP Patator class has 59, the Heartbleed class has 11, and the Infiltration class has 24 samples. Due to the small number of samples in some classes, classification using existing machine learning methods resulted in very low performance across all metrics. For example, when applying the CatBoost algorithm to the dataset, classes 1 and 2 were recognized with 100% accuracy according to the Recall metric, whereas samples from the Heartbleed and Infiltration classes were not detected at all by the algorithm, resulting in a recall of 0%. However, the proposed GAN+CatBoost algorithm demonstrated high accuracy in detecting each class. Specifically, this method was able to recognize samples from the DoS Slowloris and FTP Patator classes with 100% accuracy, Heartbleed samples with 0.9852 accuracy, and Infiltration samples with 0.9754 accuracy.

The confusion matrices of the CatBoost and GAN+CatBoost algorithms are illustrated in Figure 3.



CatBoost simple

GAN+CatBoost

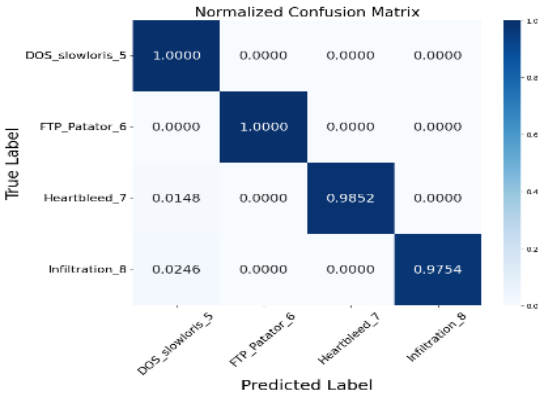
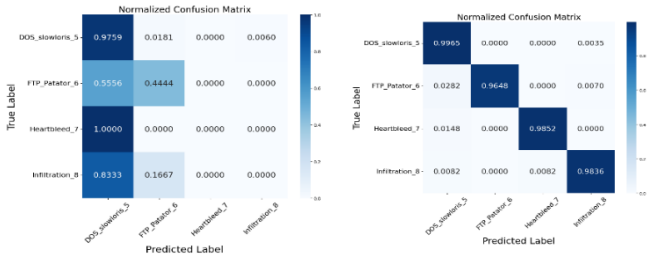


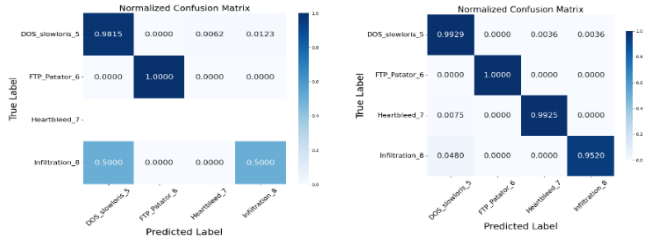
Fig. 3. Confusion matrix of the CatBoost and GAN+CatBoost

As seen in the figure 3, in the first matrix, the samples from the first two classes were recognized with high accuracy by the CatBoost algorithm. However, the algorithm misclassified samples from the other two classes, assigning them to the first class. In this case, the elements of the matrix could not be accurately aligned along the diagonal. In contrast, the second matrix shows the opposite scenario. Here, the proposed GAN+CatBoost algorithm was able to accurately recognize samples from each class, and the matrix elements were precisely aligned along the diagonal.

The confusion matrices of the XGBoost, GAN+XGBoost, KNN, and GAN+KNN algorithms are presented in Figure 4.



KNN simple GAN+ KNN



XGBoost simple simple GAN+ XGBoost

Fig. 4. Confusion matrices of the KNN, GAN+KNN, XGBoost, and GAN+XGBoost algorithms

Comparative analysis of the methods based on the Precision metric is presented in Table 2.

Table 2. Comparative analysis of the methods based on the Precision metric

Method / Class	Catboost	GAN+ Catboost	XGBoost	GAN+ XGBoost	KNN	GAN+ KNN
DOS Slowloris	0.9540	0.9827	0.9815	0.9754	0.9429	0.9759
FTP Patator	1.0000	1.0000	1.0000	1.0000	0.6667	1.0000
Heartbleed	0.0000	1.0000	0.0000	0.9925	0.0000	0.9925
Infiltration	0.0000	1.0000	0.6000	0.9917	0.0000	0.9836

Comparative analysis of the methods based on the F1-score metric is presented in Table 3.

Table 3. Comparative analysis of the methods based on the F1-score metric

Method / Class	Catboost	GAN+ Catboost	XGBoost	GAN+ XGBoost	KNN	GAN+ KNN
DOS Slowloris	0.9765	0.9913	0.9815	0.9841	0.9677	0.9965
FTP Patator	1.0000	1.0000	1.0000	1.0000	0.5333	0.9648
Heartbleed	0.0000	0.9925	0.0000	0.9925	0.0000	0.9852
Infiltration	0.0000	0.9876	0.5455	0.9714	0.0000	0.9836

As seen from Table 2 and Table 3, the proposed approach achieved higher Precision and F1-score values across all classes compared to the existing methods.

When image transformations were applied to the dataset and feature reduction techniques were used on the resulting tabular data, better results were achieved across all classes in the dataset. For this purpose, the ViT model was used for feature extraction, and PCA with 45 components was applied for feature reduction. The results obtained on the dataset using the proposed GAN+ViT+PCA+CatBoost model are presented in Table 4.

Table 4. Classification results of the GAN+ViT +PCA+CatBoost model

Method / Class	Precision	Recall	F1-score	Overall accuracy of the model
DOS Slowloris	0.9187	0.9912	0.9536	0.9451
FTP Patator	0.9608	0.8750	0.9159	
Heartbleed	0.9412	0.9412	0.9412	
Infiltration	1.0000	0.9231	0.9600	

As seen from the table 4, the proposed GAN+ ViT+PCA+CatBoost model was able to classify samples from the DoS Slowloris, FTP Patator, Heartbleed, and Infiltration classes with high accuracy. The model achieved an accuracy score of 0.9451.

However, the classification results of the simple ViT+CatBoost model, which was built without applying GAN and PCA techniques, were lower compared to the GAN+ ViT+PCA+CatBoost model. The classification results of the ViT+CatBoost model are presented in Table 5.

Table 5. Classification results of the ViT+CatBoost model

Method / Class	Precision	Recall	F1-score	Overall accuracy of the model
DOS Slowloris	0.9014	1.0000	0.9481	0.9041
FTP Patator	1.0000	0.2857	0.4444	
Heartbleed	0.0000	0.0000	0.0000	
Infiltration	0.0000	0.0000	0.0000	

As seen from the table 5, while the ViT+CatBoost model was able to classify samples from the DoS Slowloris class with high accuracy, it failed to recognize samples from the other three classes.

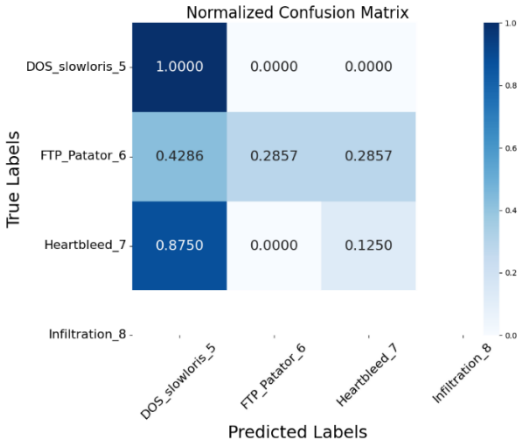
The comparative analysis of the ViT+CatBoost and ViT+GAN+PCA+CatBoost models based on the precision metric is presented in Table 6.

Table 6. Analysis of the precision metrics of the ViT+CatBoost and GAN+ ViT+PCA+CatBoost models

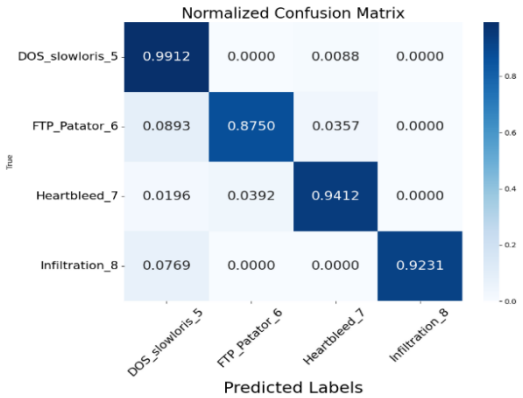
Method / Class	ViT+CatBoost	GAN+ ViT+PCA+CatBoost
DOS Slowloris	0.9014	0.9187
FTP Patator	1.0000	0.9608
Heartbleed	0.0000	0.9412
Infiltration	0.0000	1.0000

It is clearly evident from Table 6 that the GAN+ ViT+PCA+CatBoost model accurately recognized the Cloud Attack Dataset images with high precision.

The confusion matrices of the ViT+CatBoost and GAN+ ViT+PCA+CatBoost algorithms are presented in Figure 5.



ViT+CatBoost



GAN+ViT+PCA+CatBoost

Fig. 5. Confusion matrix of the ViT+CatBoost and GAN+ ViT+PCA+CatBoost models

As seen from the confusion matrix, constructing the more complex GAN+ViT+PCA+CatBoost model significantly improved the classification results. This is an expected outcome.

CONCLUSION

In this paper, images from the Cloud Attack Dataset were used to detect cyber attacks on cloud systems. The proposed approach is based on the combination of GAN and machine learning models, which enabled the achievement of high performance in cyber attack detection. The workflow of the approach begins with a preprocessing stage. In this stage, a sharpening operation within the GAN is applied to enhance important image details, filtering is used to remove noise, and rotation is employed to generate image variations.

In the next stage, normalization and data augmentation are performed to improve robustness and eliminate class imbalance in the dataset. During the experiments, 80% of the dataset samples were used for training and 20% for testing. Accuracy, Precision, Recall, and F1-score metrics were used to evaluate the model's effectiveness. The proposed model achieved 0.9451% accuracy in detecting cloud cyber attacks.

However, the extremely small number of samples in some classes of the Cloud Attack Dataset required complex preprocessing operations. In future work, we plan to propose more advanced preprocessing techniques to further improve detection accuracy for cloud-based cyber attacks.

Acknowledgments

This work was supported by the Science Foundation of the State Oil Company of Azerbaijan Republic (SOCAR) (Contract No 01LR-EF/2024)

REFERENCES

[1] M.H. Onik, C.S. Kim, J. Yang, "Personal data privacy challenges of the fourth industrial revolution," Proc. of the IEEE International conference on advanced communications technology, PyeongChang, Korea (South), pp. 635–638, 17-20 February 2019.

[2] R.M. Alguliev, F.C. Abdullayeva, "An investigation and analysis of security problems of the cloud computing," Problems of Information Technology, no. 1(7), pp. 3–14, 2013.

[3] "Hökumət buludu"nun (G-cloud) yaradılması və "bulud" xidmətlərinin göstərilməsi sahəsində tədbirlər haqqında Azərbaycan Respublikası Prezidentinin Fərmanı, № 718, 3 iyun 2019-cu il

[4] R. Browne, Hackers hijack Tesla's cloud system to mine cryptocurrency, Feb 21 2018

[5] J. Goldman, Code Spaces Destroyed by Cyber Attack, June 23, 2014

[6] Commvault Discloses Zero-Day Exploit Breach in Azure Cloud Environment, https://cyberpress.org/commvault-discloses-zero-day-exploit/

[7] Ukraine mobile network Kyivstar hit by 'cyber-attack', 2023, https://www.bbc.com/news/world-europe-67691222

[8] M. Masdari, M. Jalali, "A survey and taxonomy of DoS attacks in cloud computing: DoS attacks in cloud computing," Security and Communication Networks, vol. 9, no. 16, pp. 28. 2016.

[9] T. Hussain, C. Nugent, J. Liu, A. Beard, L. Chen, A. Moore, "An Attack Impact and Host Importance based Approach to Intrusion Response Action Selection," Proc. of the 4th International Conference on Information Technology and Computer Communications, Guangzhou, China, pp. 84-91 June 23 - 25, 2022.

[10] F.J. Abdullayeva, "Malware detection in cloud computing using an image visualization technique," Proc. of the IEEE 13th International Conference on Application of Information and Communication Technologies, Baku, Azerbaijan, pp. 1-5. 23-25 October 2019.

[11] M. Ali, Y. Saleem, S. Hina, G.A. Shah, "DDoSViT: IoT DDoS attack detection for fortifying firmware Over-The-Air (OTA) updates using vision transformer," Internet of Things, Vol. 30, 101527, 2025.

[12] O. Valikhanli, Detection of malware in ground control stations of unmanned aerial vehicles based on image processing, International Journal of Information and Computer Security, Vol. 26, No. 1-2, pp. 147-158, 2025.

[13] Y. Alaca, Y. Çelik, "Cyber attack detection with QR code images using lightweight deep learning models," Computers & Security, Vol. 126, 103065, 2023.

[14] Fathima, G.S. Devi, M. Faizaanuddin, "Improving distributed denial of service attack detection using supervised machine learning," Measurement: Sensors, Vol. 30, 100911, 2023.

[15] Natsos, A.L. Symeonidis, "Transformer-based malware detection using process resource utilization metrics," Results in Engineering, Vol. 25, pp. 1-19, 2025.

[16] L.D. Rose, G. Andresini, A. Appice, D. Malerba, "VINCENT: Cyber-threat detection through vision transformers and knowledge distillation," Computers & Security, Vol. 144, pp. 1-13, 2024.

[17] H. Wasswa, H.A. Abbass, T. Lynar, "ResDNViT: A hybrid architecture for Netflow-based attack detection using a residual dense network and Vision Transformer," Expert Systems with Applications, Vol. 282, pp. 1-14, 2025.

[18] F.J. Rendón, J.A. Álvarez, A.J. Varela, "Paying attention to cyber-attacks: A multi-layer perceptron with self-attention mechanism," Computers & Security, Vol. 132, 103318, 2023.

[19] H. Kheddar, "Transformers and large language models for efficient intrusion detection systems: A comprehensive survey," Information Fusion, Vol. 124, pp. 1-32, 2025.

[20] Cloud Attack Dataset, IEEE Dataport, https://ieee-dataport.org/documents/cloud-attack-dataset. (Accessed on 09 June 2025)

Machine Learning-Based Credit Card Fraud Detection: Balanced and Imbalanced Cases Evaluation

Abdullah Güney IŞIK
Electrical and Electronics Engineering
Bakırçay University
Izmir, Turkiye
200603012@bakircay.edu.tr

Şeyma GİRGİN
Electrical and Electronics Engineering
Bakırçay University
Izmir, Turkiye
210603048@bakircay.edu.tr

Hazal Su BIÇAKCI YEŞİLKAYA
Electrical and Electronics Engineering
Bakırçay University
Izmir, Turkiye
hazalsu.bicakci@bakircay.edu.tr

***Abstract*— Credit card fraud has become an increasing threat with the development of the digital economy.brown In this study, different machine learning methods are examined by addressing the class imbalance problem in order to detect credit card frauds. The dataset is public and its name is a credit card transaction provided by Kaggle and represents real-world data. The training data is balanced by SMOTE (Synthetic Minority Over-sampling Technique) method, and the model performance is increased by extracting new features. Logistic Regression, Random Forest and XGBoost algorithms are applied and the performance of each is compared in terms of accuracy, precision, recall, F1 score and ROC/AUC values. In addition, the test data is evaluated both as imbalanced and balanced sample sets. The results show that the XGBoost model shows high performance especially in balanced test data.**

Keywords—credit card fraud detection, online transactions, machine learning, SMOTE, classification

INTRODUCTION

With the acceleration of digitalization, credit card usage has increased worldwide. However, this increase has also brought with it a serious increase in credit card fraud cases. In 2020, the total losses due to credit card fraud worldwide were $28.58 billion [1].

In the United States, losses due to credit card fraud reached $12.5 billion in 2024 [2]. In 2020, 393.207 people in the United States were victims of identity theft, with a significant number of these cases being attributed to credit card fraud [3].

A similar trend is observed in Türkiye. There has been a significant increase in credit card fraud cases in Türkiye compared to previous years. This situation poses a serious threat to both individual users and financial institutions.

Credit card fraud not only causes financial losses but also undermines user trust and damages the reliability of financial systems. Therefore, preventing and detecting fraud cases is of great importance in ensuring financial security. In this study, a model was developed using machine learning techniques for the detection of credit card fraud.

In this study, a model is developed using machine learning techniques (Logistic Regression, Random Forest and XGBoost) for the detection of credit card fraud. The dataset is a set of credit card transactions provided via Kaggle and represents real-world data [4]. In this study, a SMOTE method was used to balanced the training data set [5]. The performance of the model is evaluated using various data preprocessing methods and algorithms. The results show the effectiveness of machine learning techniques in the detection of credit card fraud.

Traditional rule-based systems have proven insufficient for detecting fraud cases, as they operate only based on predefined criteria and struggle to identify emerging fraud patterns [6]. However, machine learning algorithms have the capacity to dynamically predict fraud cases by learning patterns on data.

Machine learning makes it possible to distinguish normal and abnormal behaviours by analysing past transactions. These technologies are especially suitable for the structural characteristics of fraud data, such as imbalanced classes, high-dimensional data, and the need for fast processing. For this reason, they are widely used in fraud detection in many areas, especially in the financial sector. Machine learning-based solutions become distinctive not only with their high success rates, but also with the possibility of real-time intervention. In this study, a model was developed using machine learning techniques for the detection of credit card fraud. The performance of the model was evaluated using various data preprocessing methods and algorithms. The results obtained reveal the effectiveness of machine learning techniques in the detection of credit card fraud.

RELATED WORKS

Credit card fraud dataset has been used in many studies in the literature and evaluated with different machine learning algorithms. In this section, selected studies utilising the same dataset are presented and their results are summarized comparatively.

Niu et al. [7] compared supervised and unsupervised learning methods; in particular, XGBoost and Random Forest algorithms achieved the highest success with area under the ROC curve (AUC) scores of 98.9% and 98.8%, respectively. In study [8], the effect of sampling methods such as SMOTE on model performance has been investigated; it has been observed that applying SMOTE before training/test separation causes data leakage and the results increase artificially. For this reason, in our study, SMOTE technique was applied to the training set after the separation.

Btoush et al. [9] achieved results with F1 scores above 90% by combining deep learning with classical machine learning methods. It was observed that these models can generalize strongly even with a limited number of fraud samples. Thimonier et al. [10] compared anomaly detection and supervised learning methods. In this study, the

LightGBM model achieved high success. However, it was stated that it was sensitive to shifts in the data distribution.

MATERIALS AND METHODS

The main problem of fraud data is the imbalance between classes. This situation may cause prediction algorithms to ignore fraudulent transactions. In this study, SMOTE is used to reduce the effect of imbalanced classes; new data-specific features are extracted and three different machine learning algorithms are tested. Our aim is to objectively evaluate the performance of the models in both imbalanced and balanced conditions and to determine the most effective approach in fraud detection.

Dataset

The dataset used in the study consists of 284,807 credit card transactions obtained from a bank in Europe and shared on Kaggle [4]. Only 492 of these transactions were labeled as fraudulent, meaning the fraud rate in the data is approximately 0.17% [10].

Pre-processing

When the dataset was examined, no missing or outlier data was found. All variables were numeric and were subjected to Principal Component Analysis (PCA) transformation for anonymization purposes. This process aimed to minimize confidentiality risks and reduce dimensionality by transforming the original variables into linearly independent components. However, the 'Time' and 'Amount' columns were retained in their original form within the dataset.

Feature Extractions

The following new features have been added to the dataset to improve the performance of the model:

• **Hour**: A variable derived from the Time column that corresponds to the transaction hour.

• **Log_Amount**: The log transformation of the amount variable.

• **Amount_per_time**: Calculated by taking the ratio of the amount to the transaction time.

• **Amount_zscore**: The normalized version of the amount value with the standard deviation.

These operations contribute to the determination of transaction patterns.

The 10 most effective features were investigated utilising the Random Forests model. The results are as shown in Figure 1.

These features have the highest importance scores in training the RFC model. However, the features that best discriminate fraudulent behaviour are observed to be V14, V17, V12, Amount_zscore and V10.

Data Scaling

Prior to model training, all variables were scaled using StandardScaler to have a mean of 0 and a standard deviation of 1. This shortened the training time and increased stability in some algorithms.

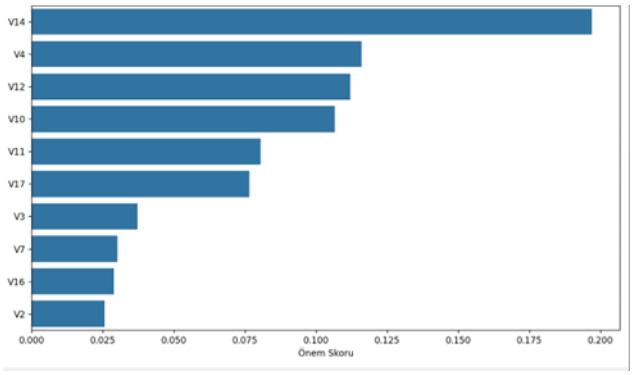


Figure 1 The most important features for RFC training

Classification Model

The dataset was randomly split into 80% for training and 20% for testing, ensuring the class ratio remained intact. This approach maintained the presence of fraudulent examples in both sets, supporting balanced model evaluation.

The imbalance in the training data is addressed by the SMOTE algorithm [11]. This technique balances the majority class by generating new synthetic examples using existing fraud data. This significantly improves the model's minority class learning [8].

In order to fairly evaluate the model performance, the test set is also balanced. In addition to the unbalanced test set, a second test set is created by taking equal numbers (98 samples each) from class 0 and class 1. This test set provides evaluation based on hypothetical scenarios where the classes are equally distributed, rather than the rare fraud scenarios in reality.

Three different methods were chosen to compare machine learning algorithms:

Logistic Regression (LR)

It is a basic and easy to interpret algorithm based on linear classification. It performs well on weighted classes but may have difficulty capturing complex patterns. In our experiments, the number of maximum iterations was 1000.

Random Forests (RFC)

It is a collection of decision trees created with random subsamples. It offers high performance, especially in multi-dimensional data sets. It is also suitable for feature importance analysis.

The hyperparameter search for the RFC classifier was conducted using GridSearchCV, and the optimal combination of hyperparameters was determined. These parameters are:

- n_estimators: 200
- max_depth: None
- min_samples_split: 5

XGBoost (XGB)

It is an advanced classification method based on boosting, which updates the weights of the model by targeting error-prone examples. It has high performance and is resistant to over-fitting. In our experiments, ***logloss*** was utilized as an evaluation metric. The hyperparameters of the XGB classifier were used as default. The scale positive weight (scale-pos_weight) used in imbalanced dataset was set to 1 because the SMOTE technique was used in the training set.

Performance Metrics

The performance of the models is measured by the following metrics:

• Accuracy
• Precision: Accuracy of positive predictions
• Recall: True positive detection rate
• F1 Score: Measures the balance between precision and recall.
• ROC Curve and AUC: Shows the classification ability of the model.
• Confusion Matrix: Visualizes the correct and incorrect classification rates.

RESULTS

As a result of the evaluations, it was observed that the XGBoost algorithm showed the highest performance in both unbalanced and balanced test sets.

The XGBoost model produced approximately 96% accuracy, 81% recall and 86% F1 score in the unbalanced test set. In the evaluations made with the balanced test set, these accuracy rates increased even more, and AUC value exceeded 98%.

Table I shows precision, recall, F1 score and accuracy for credit card fraud classification results. Figure 2 shows the ROC curve for the imbalanced test set, while Figure 3 visualizes the performance comparisons. Tables II, III and IV show the confusion matrices obtained from the LR, RFC and XGB classifiers, respectively.

CREDIT CARD FRAUD CLASSIFICATION RESULTS FOR IMBALANCED TEST SET.

Models	Metrics	0	1	Accuracy
LR	*Precision*	1.00	0.05	0.97
	Recall	0.97	0.91	
	F1 Score	0.99	0.10	
RFC	*Precision*	1.00	0.86	1.00
	Recall	1.00	0.83	
	F1 Score	1.00	0.84	
XGB	*Precision*	1.00	0.71	1.00
	Recall	1.00	0.87	
	F1 Score	1.00	0.78	

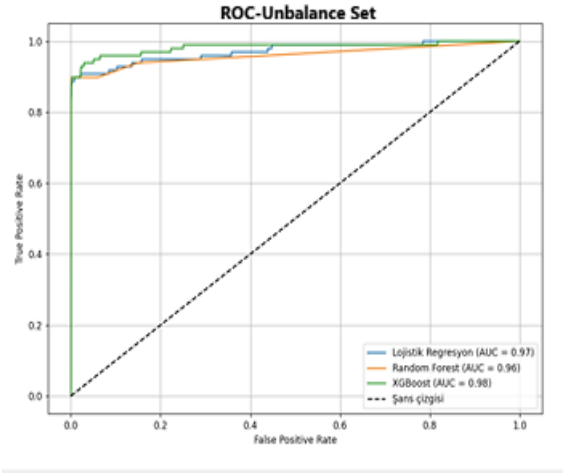


Figure 2 The ROC curve of imbalance test set

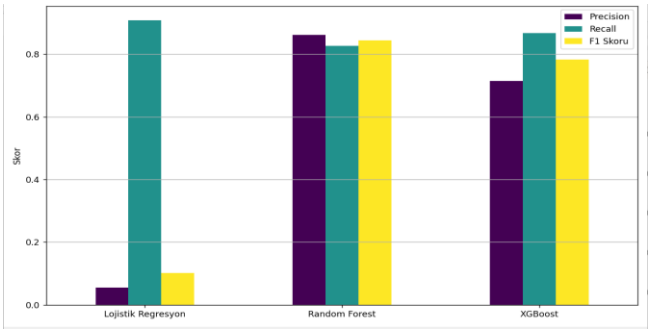


Figure 3 Performance comparisons of different models in terms of precision, recall and F1 score for imbalance test set

LR CONFUSION MATRIX

True Label	**0**	**55306**	**1558**
	1	**9**	**89**
		0	**1**
		Predicted Label	

RFC CONFUSION MATRIX

True Label	**0**	**56851**	**13**
	1	**17**	**81**
		0	**1**
		Predicted Label	

XGB CONFUSION MATRIX

True Label	**0**	**56830**	**34**
	1	**13**	**85**
		0	**1**
		Predicted Label	

When the classes in the test set were equalised, 98 samples were randomly selected for both classes. This step provided the opportunity to classify the training set balanced with SMOTE with the test set having a balanced sample number.

Table V shows precision, recall, F1 score and accuracy for credit card fraud classification results. Figure 4 shows the ROC curve for the balanced test set, while Figure 5 visualizes

the performance comparisons. Tables VI, VII and VIII show the confusion matrices obtained from the LR, RFC and XGB classifiers, respectively.

CREDIT CARD FRAUD CLASSIFICATION RESULTS FOR BALANCED TEST SET

Models	Metrics	0	1	Accuracy
LR	*Precision*	0.91	0.98	0.94
	Recall	0.98	0.91	
	F1 Score	0.95	0.94	
RFC	*Precision*	0.85	1.00	0.91
	Recall	1.00	0.83	
	F1 Score	0.92	0.91	
XGB	*Precision*	0.88	1.00	0.93
	Recall	1.00	0.87	
	F1 Score	0.94	0.93	

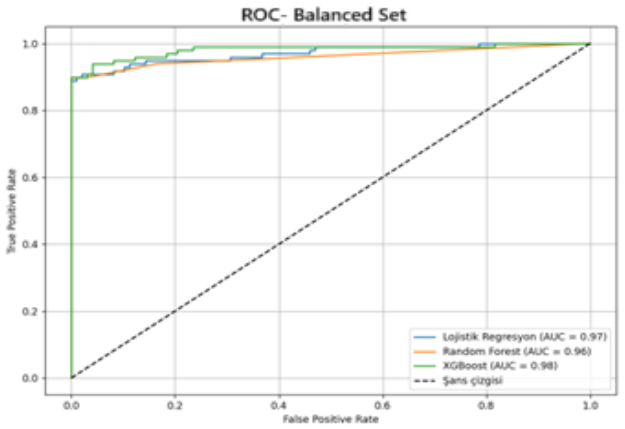


Figure 4 The ROC curve of balanced test set

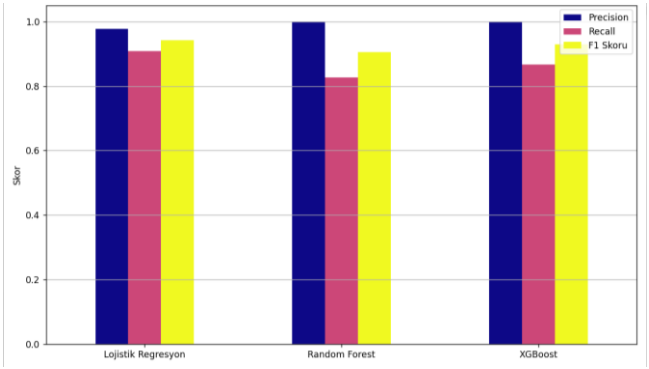


Figure 5 Performance comparisons of different models in terms of precision, recall and F1 score for balanced test set

LR CONFUSION MATRIX

True Label	**0**	**96**	**2**
	1	**9**	**89**
		0	**1**
		Predicted Label	

RFC CONFUSION MATRIX

True Label	**0**	**98**	**0**
	1	**17**	**81**
		0	**1**
		Predicted Label	

XGB CONFUSION MATRIX

True Label	**0**	**98**	**0**
	1	**13**	**85**
		0	**1**
		Predicted Label	

A comparative analysis of the balanced and unbalanced test set scenarios reveals that class 0, representing real transactions, is classified with high accuracy in the unbalanced dataset, whereas class 1, corresponding to fraudulent transactions, exhibits lower classification performance. However, in the case of the balanced test set, the model demonstrates equal effectiveness in distinguishing both real and fraudulent transactions.

DISCUSSION

This study investigates the application of various machine learning algorithms to a real-world dataset obtained from Kaggle for the detection of credit card fraud [4]. To address the imbalance in the training data, the SMOTE was employed, additional features were extracted for better classification performances, and the dataset was standardized. Evaluation results indicate that the XGBoost algorithm demonstrated the highest performance across both imbalanced and balanced test sets.

Considering the computational costs, LR was the classifier with the least computational cost, while XGB was the classifier with the highest computational cost. Since the number of test data is less in the balanced dataset case, the computational cost is reduced compared to the imbalanced case.

A comparison with existing studies reveals that the results obtained in this study outperform many findings in the literature. For instance, Kabane [8] reported that applying the SMOTE technique prior to the train-test split led to artificial increases in model accuracy and data leakage. To mitigate this issue, SMOTE was applied exclusively to the training dataset in this study, ensuring more reliable outcomes. Similarly, while the study by Niu et al. [7] achieved a commendable AUC score of 98.9% using the XGBoost algorithm, it did not incorporate feature engineering. In contrast, this study introduced additional features such as "Hour" and "Amount_zscore," which contributed to improved model performance.

Moreover, Zhu et al. [4] proposed a neural network–SMOTE-based approach that achieved an F1 score of approximately 90%. While this result is comparable to the performance of the model developed in this study, the latter exhibited greater computational efficiency. Conversely, Thimonier et al. [10] attained higher accuracy rates using the LightGBM algorithm. However, their study highlighted that

the model was highly sensitive to variations in data distribution, raising concerns regarding its stability in real-world applications.

In this context, the present study offers a more balanced approach relative to existing literature while simultaneously ensuring reliability by mitigating common issues such as data leakage. Specifically, the selective application of SMOTE to the training data, the evaluation of the test set in both balanced and imbalanced conditions, and the enhancement of model performance through feature extraction underscore the study's contributions to the field.

CONCLUSIONS AND FUTURE WORKS

This study incorporated additional features into the PCA data and utilised them in machine learning models. The training dataset was balanced using the SMOTE method, while the test dataset was evaluated under both balanced and unbalanced conditions. Among the machine learning models tested, the XGBoost algorithm demonstrated the highest performance, with superior classification results observed in the balanced test dataset.

Future research could explore the impact of various features on classical machine learning models. Additionally, the influence of different deep learning models on credit card fraud detection could be examined. Furthermore, alternative data balancing techniques may be investigated to enhance the performance of models applied to imbalanced datasets.

REFERENCES

[1] Nilson Report, "Card fraud losses dip to $28.58 billion in 2020," *Nilson Report*, 2025. [Online]. Available: Nilson Report. [Accessed: May 24, 2025].

[2] Federal Trade Commission, "New FTC data show a big jump in reported losses to fraud to $12.5 billion in 2024," Federal Trade Commission, Mar. 10, 2025. [Online]. Available: FTC. [Accessed: May 24, 2025].

[3] K. Alarfaj, I. Malik, H. U. Khan, N. Almusallam, M. Ramzan, and M. Ahmed, "Credit Card Fraud Detection Using State-of-the-art Machine Learning and Deep Learning Algorithms," IEEE Access, vol. PP, pp. 1-1, 2023.

[4] MLG-ULB, "Kaggle Credit Card Fraud Dataset," Kaggle, 2025. [Online]. Available: Kaggle. [Accessed: May 24, 2025].

[5] Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating probability with undersampling for unbalanced classification," in 2015 IEEE Symposium Series on Computational Intelligence, Dec. 2015, pp. 159-166.

[6] O. A. Bello, A. Folorunso, O. E. Ejiofor, F. Z. Budale, K. Adebayo, and O. A. Babatunde, "Machine learning approaches for enhancing fraud prevention in financial transactions," International Journal of Management Technology, vol. 10, no. 1, pp. 85-108, 2023.

[7] X. Niu, L. Wang, and X. Yang, "A comparison study of credit card fraud detection: Supervised versus unsupervised," arXiv preprint, arXiv:1904.10604, Apr. 2019.

[8] S. Kabane, "Impact of Sampling Techniques and Data Leakage on XGBoost Performance in Credit Card Fraud Detection," arXiv preprint, arXiv:2412.07437, 2024.

[9] Btoush, X. Zhou, R. Gururajan, K. C. Chan, and O. Alsodi, "Achieving Excellence in Cyber Fraud Detection: A Hybrid ML+DL Ensemble Approach for Credit Cards," Applied Sciences, vol. 15, no. 3, p. 1081, 2025.

[10] Thimonier, F. Popineau, A. Rimmel, B. L. Doan, and F. Daniel, "Comparative evaluation of anomaly detection methods for fraud detection in online credit card payments," in *International Congress on Information and Communication Technology*, Singapore: Springer Nature Singapore, Feb. 2024, pp. 37-50.

[11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, pp. 321-357, 2002.