Bulletin of Electrical Engineering and Informatics
Vol. 15, No. 2, April 2026, pp. 1866~1878
ISSN: 2302-9285, DOI: 10.11591/eei.v15i2.11380

Hybrid analytical framework for evaluating socio-economic factors in regional development

Ayman Akynbekova[1], Raikhan Muratkhan[2], Zhanar Lamasheva[1], Ayagoz Mukhanova[1], Gulbakhar Yussupova[3], Serik Eslyamov[4], Saya Santeyeva[5], Alfiya Abdrakhmanova[1]
[1]Department of Information Systems, Faculty of Information Technology, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
[2]Department of Applied Mathematics and Computer Science, Faculty of Mathematics and Information Technology, Karaganda Buketov University, Karaganda, Kazakhstan
[3]Department of Radio Engineering and Telecommunications, ALT University, Almaty, Kazakhstan
[4]Department of Radio Engineering, Electronics, and Telecommunications, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan
[5]Department of Information Security, Faculty of Information Technology, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

Article Info

Article history:

Received Sep 11, 2025
Revised Feb 23, 2026
Accepted Mar 10, 2026

Keywords:

Correlation analysis
Hybrid model
Machine learning
Principal component analysis
Socio-economic factors

ABSTRACT

This study aims to develop and validate a hybrid analytical framework for evaluating the influence of socio-economic factors on regional development. The framework combines correlation analysis, principal component analysis (PCA), and fuzzy inference modeling into a unified approach, applied to 2023 data from the city of Taraz, Kazakhstan, covering 16 socio-economic indicators across demographic, economic, social, and industrial domains. The findings reveal that investments in fixed assets (r=0.8963 and q=0.000010), average monthly salary (r=0.8907 and q=0.000010), and retail trade (r=0.8885 and q=0.000010) exert the strongest positive influence, while migration balance and manufacturing show weak or negative effects. The results demonstrate that the hybrid model offers more comprehensive insights compared to single-method approaches, validating its effectiveness in capturing complex and uncertain dependencies. Practically, the model provides policymakers with a robust decision-support tool for identifying priority areas, designing targeted strategies, and ensuring sustainable regional growth, with adaptability to other regions and datasets.

Corresponding Author:

Raikhan Muratkhan
Department of Applied Mathematics and Computer Science
Faculty of Mathematics and Information Technology, Karaganda Buketov University
Karaganda, Kazakhstan
Email: Muratkhan_Raikhan@buketov.edu.kz

1. INTRODUCTION

Hybrid data analysis models provide a powerful tool for studying and forecasting complex processes in various fields, especially in socio-economic research [1]-[3]. They combine several analysis methods, such as correlation analysis, principal component analysis (PCA) and fuzzy logic. This combination allows taking into account many factors and their interactions. In modern conditions, when the processes under study are becoming increasingly multifaceted and dynamic, hybrid models [4]-[6] are becoming indispensable for obtaining accurate and reliable conclusions. The objective of the study is to assess the influence of various socio-economic factors on key processes occurring in the region. For example, such indicators as population, employment, investment and crime rates [7], [8] have different impacts on economic and social processes.

However, their impact is not always direct and is subject to other factors. Traditional analysis methods, as a matter of fact, cannot fully take into account such complex dependencies, which implies that a mixed approach allows for the estimation of such interactions with greater precision based on quantitative and qualitative data processing. One of the significant advantages of hybrid models is that they have the ability to address nonlinear interdependencies among variables, which is essential in studying social and economic systems. In most cases, standard techniques such as linear regression cannot adequately address complex and multivariate interdependencies. In such situations, fuzzy logic methods become indispensable, as they model the uncertainties and ambiguities [9]-[11] that are often found in real data.

The aim of this study is to create a hybrid model that can take into account all these aspects and allow for the analysis of factors influencing social and economic processes. Based on the stated objectives, the research seeks to answer the following questions: Which socio-economic factors have the most significant positive or negative impact on regional development processes? How effectively can the proposed hybrid model—combining correlation analysis, PCA, and fuzzy logic—identify and quantify these impacts compared to traditional analysis methods? Can the hybrid model be adapted for different regions to provide accurate predictions for decision-making in socio-economic planning?

In line with these research questions, the study tests the following hypotheses: H1: the hybrid model allows for more accurate identification of significant socio-economic factors influencing regional development than single-method approaches. H2: investments in fixed assets, population size, and education-related indicators have a stronger positive influence on economic and social processes compared to other factors. H3: certain factors, such as high birth rates or specific industry outputs, may exert a negative influence on socio-economic processes despite their apparent importance.

In this work, we used correlation analysis to determine the dependencies between factors [12], [13] and processes, PCA to reduce the dimensionality of data [14], [15], and fuzzy logic [16]-[18] to model uncertain and complex relationships [19]. These methods were combined into a single model, which provided a complete picture of the influence of factors on socio-economic processes. Research by Tian *et al.* [20] presents a hybrid gray artificial neural network model for forecasting the volume of waste electrical and electronic equipment (WEEE) in 31 regions of China, and also conducted a socio-economic analysis of seven types of WEEE. The model showed the smallest errors compared with other methods and was the most suitable for forecasting some types of equipment in urban areas. The forecast indicates an increase in WEEE volumes by 2025, which requires the creation of additional recycling centers. The article highlights the importance of effective monitoring and collection systems for WEEE management.

Research by Kumar [21] discusses the application of hybrid machine learning models to predict CO_2 emissions based on energy and socioeconomic data from 1960 to 2018. The proposed model, which combines PCA and machine learning, showed higher accuracy and efficiency compared to other methods, such as linear regression, random forest, RNN, and LSTM. The model achieved minimal error values and computation time. The study can help in developing measures to reduce and manage carbon emissions for decision-making at the policy and government levels. Research by Qian *et al.* [22] presents a systematic review of studies examining the relationship between area-level socioeconomic factors and suicide using spatial analysis methods. Of the 58 included studies, more than half used methods that take into account spatial effects, such as Bayesian and regression models. The results showed that areas with high unemployment, low income, and low education levels have a higher risk of suicide. The study confirms that area-level socio-economic factors are associated with suicide risk, regardless of whether spatial methods were used. Based Kumar *et al.* [23], an improvement of the classical JAYA algorithm using the fractional order concept is proposed to improve the efficiency of renewable technologies. An intelligent system for assessing the impact of solar unpredictability, biomass planning, and optimizing techno-economic decisions is presented using a novel metaheuristic algorithm FO-JAYA. Solutions for village electrification in Eastern India using hybrid renewable energy sources are investigated. The results show that the proposed algorithm reduces the energy cost compared to other methods, demonstrating better performance.

The research is applicable in the sense that modern social and economic processes are characterized by a complex interdependence of their determinants which can change under the influence of internal and external factors. For example, demographic processes can be caused by migration, birth and death rates, economic status, and even access to employment and social services. At the same time, the impact of these factors is not linear and direct in numerous instances, which creates difficulties for quantitative modeling by traditional means. The use of fuzzy logic allows [24], [25] to model uncertainties and fuzzy boundaries, especially in situations where the data is incomplete or accurate, or when the influence of factors is multifaceted. The use of a hybrid model, including correlation analysis, the PCA and fuzzy models, provides a more complete and accurate picture of the analyzed processes based on a variety of data from various sources. This, in turn, allows us to develop more effective strategies for regional development, taking into account the influence of social, economic and production factors.

2. METHOD

For deep data analysis and identification of relationships between socio-economic factors and processes occurring in the region, both traditional statistical methods and modern computational approaches can be used [26], [27]. Among them are machine learning methods and deep data analysis [28]. In this study, various methods were employed, each playing a distinct role in the overall model. These methods enabled quantitative assessment, data dimensionality reduction, and accounting for uncertainties in the information. In Figure 1, the algorithm is a hybrid model for analyzing the influence of factors on processes in the economy, demography, and other areas. It includes several stages of data processing, from their standardization to the construction of a fuzzy logic model. The model combines dimensionality reduction methods (PCA) with fuzzy logic rules to assess the influence of each factor on specific processes.

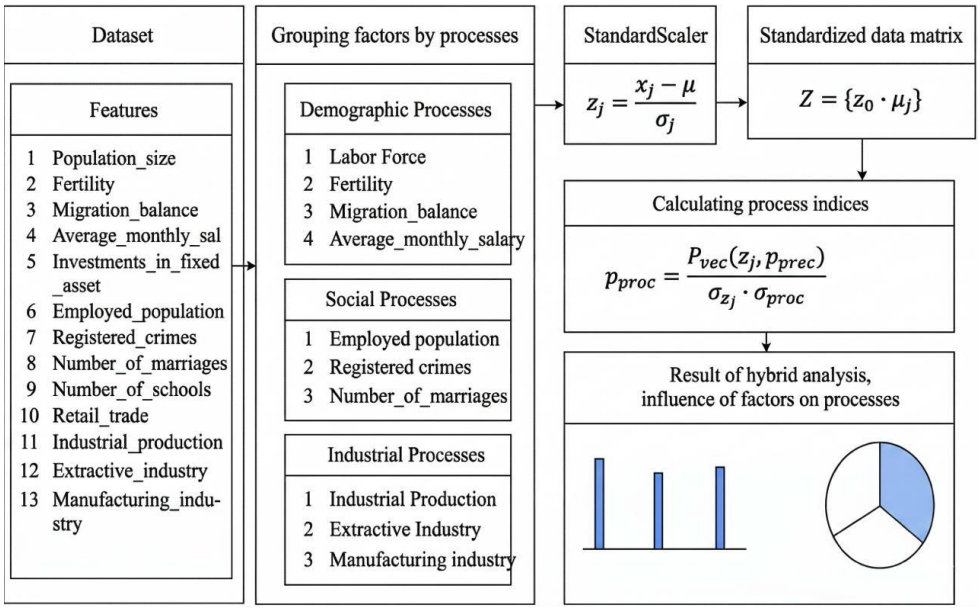

Figure 1. Hybrid model for analyzing the influence of factors on processes

a. Step 1. Loading data. The loaded data represents a variety of factors reflecting various aspects of the processes. Included for analysis are demographic processes, such as population size, birth rate and migration balance; economic processes, including labor force, average monthly wage and investment in fixed capital; social processes, such as employment, registered crimes, number of marriages and cost of living; business and trade processes, including the number of schools, legal entities and retail trade; and industrial processes, such as industrial production, mining and manufacturing.
b. Step 2. Data standardization. The data is normalized using the StandardScaler algorithm so that each factor has a zero mean and a unit standard deviation (1):

$$z_{ij} = \frac{x_{ij} - \mu_{ij}}{\sigma_{ij}} \quad (1)$$

where x_{ij} is value of the factor, μ_{ij} is average value of the factor, and σ_{ij} is standard deviation of the factor.
c. Step 3. Calculating process indices. For each process, an average index is calculated based on the corresponding factors. This simplifies the analysis, allowing you to work with processes as separate entities (2):

$$P_{proc} = \frac{1}{n_{proc}} \sum_{i \in proc} z_{ij} \quad (2)$$

where P_{proc} is process index for year j, n_{proc} is number of factors in the process, and z_{ij} is standardized value of factor i in year j.
d. Step 4. Correlation analysis. Correlation coefficients are calculated for each factor and process index. Correlation helps to understand how each factor is related to the process. The correlation coefficient is calculated using (3):

$$\rho_{i,proc} = \frac{Cov\ (z_i, P_{proc})}{\sigma_{z_i} \sigma P_{proc}} \quad (3)$$

where $\rho_{i,proc}$ is correlation coefficient of factor i with the process, $Cov\ (z_i, P_{proc})$ is covariance between factor and process index, and σ_{z_i} and σP_{proc} are standard deviations for the factor and process index, respectively.

e. Step 5. PCA. PCA is used to reduce the dimensionality of the data. For each process, the first principal component is calculated, which is the combination of factors that explain the most significant portion of the variance in the data. The principal component PC_1 is calculated as a linear combination of factors (4):

$$PC_1 = WZ \quad (4)$$

where W is the vector of weights (loadings) that determines the importance of each factor and Z is the standardized data.

f. Step 6. Building a fuzzy logic model. Based on the results of the previous steps, a fuzzy logic model has been developed, serving as a key tool for assessing the impact of various factors on processes. The model uses three input variables: Z-score (standardized factor value), correlation (factor impact on the process), and principal component loading (PCA). In this step, a three-dimensional input vector is formed for each factor k, composed of the standardized value, correlation coefficient, and PCA loading (5):

$$x_k = [z_k, r_k, p_k] \quad (5)$$

where z_k is the standardized value of factor k, r_k is the correlation coefficient between factor k and the process index, and p_k is the loading of the first principal component associated with factor k.

g. Step 7. Fuzzy logic rules. Membership functions are created for each input variable, and fuzzy rules are constructed to determine the final impact value. For example, if the Z-score is positive, correlation is positive, and PCA loading is high, the influence is assessed as positive. Using fuzzy logic rules, the influence score for each factor k is estimated through an aggregation of fuzzy rule evaluations (6):

$$S_k = \sum_{r=1}^{R} \omega_r \cdot \mu_r(z_k, r_k, p_k) \quad (6)$$

where S_k is the final influence score of factor k, μ_r is the result of the fuzzy membership function for rule r, ω_r is the weight assigned to rule r, and R is the total number of fuzzy rules in the rule base.

h. Step 8. Factor impact assessment and visualization. The results are shown in heat maps and graphs. This enables the ranking of factors by their influence and the identification of the most significant drivers in regional socio-economic development. To facilitate comparison across processes, the final influence scores are linearly normalized to the range [0,1] (7):

$$S_k^* = \frac{S_k - \min(S)}{\max(S) - \min(S)} \quad (7)$$

where S_k^* is the normalized influence score for factor k, $\min(S)$ and $\max(S)$ are the minimum and maximum values among all raw influence scores S_k.

This hybrid analysis enables an integrated and multidimensional understanding of factor interactions, facilitating simplified interpretation, informed decision-making, and the formulation of targeted regional development strategies. The complete workflow of the proposed model is summarized in Algorithm 1, which combines preprocessing, dimensionality reduction, and fuzzy reasoning into a reproducible, step-by-step procedure.

Algorithm 1. Hybrid fuzzy-PCA model for factor impact evaluation

```
Input: Dataset with socio-economic indicators for a region
Output: Ranked influence scores of each factor on selected processes
1: Load raw data (demographic, economic, social, trade indicators)
2: Standardize data using Z-score normalization
3: For each process (e.g., economy, demography):
4: Identify relevant subset of indicators
5: Compute average process index
6: Calculate correlation between each factor and process index
7: Apply PCA to reduce dimensionality
8: Extract first principal component loading for each factor
9: For each factor:
10: Compute fuzzy input values:
11: - Z-score value
12: - Correlation coefficient
```

```
13: - PCA loading
14: Apply fuzzy inference rules to estimate influence score
15: Visualize results (heatmaps, bar charts, tables)
```

In summary, the proposed hybrid method combines correlation analysis, PCA, and fuzzy inference modeling. This approach offers a solid framework for assessing how socio-economic factors affect regional processes. By integrating statistical normalization, reducing dimensionality, and using expert reasoning in sequence, the model captures both the numerical relationships and the qualitative aspects found in complex regional systems. The clear, step-by-step implementation guarantees reproducibility. Also, using easy-to-understand metrics like influence scores improves the model's effectiveness in real-world policy situations. This thorough approach lays the groundwork for evaluating regional data, which is discussed in the next section.

Before normalization, exploratory analysis of distributions (skewness, kurtosis, normality test) was performed, after which logarithmic stabilization (log1p) was applied for strictly positive and strongly skewed features (rule |skew|>1|). Then, standardization by Z-score is applied to all features (8):

$$z_{ij} = \frac{x_{(ij)} - \mu_j}{\sigma_j} \tag{8}$$

This ensures comparability of scales and matches the z-score input with the membership functions of the fuzzy system. To analyze the factor-process index relationship, an adaptive correlation selection is used: Pearson if both variables are close to normal (D'Agostino test with p>0.05), otherwise Spearman. P-values are adjusted by FDR (Benjamini-Hochberg) (Table 1).

Table 1. EDA summary and applied transformations (excerpt)

Factor	Skew	Kurtosis	p(normal)	Transformation
population_chislennost	1.905	3.245	0.0008	log1p
gornodob_primyshlennost	1.757	3.289	0.0013	log1p
obrabotka_proizvodstvo	1.441	2.328	0.0090	log1p
prom_proizvodstvo	1.293	1.928	0.0209	log1p
saldo_migratsii	1.151	0.661	0.0875	log1p
investitsii_v_osnovnoi_kapital	1.144	0.922	0.0750	log1p
srednemesyachnaya_zarplata	1.060	0.845	0.1024	log1p
kol_vo_shkol	0.897	0.127	0.2434	—
prozhitochnyi_minimum	0.891	0.239	0.2368	—
zanyatoe_naselenie	0.605	−0.185	0.5223	—

Before presenting the results of the correlation analysis, an assessment of the relationship between socio-economic indicators and target indicators of regional development was carried out. To identify significant factors, the Pearson correlation method was used, which allows one to assess the degree of linear dependence between variables. The main attention was paid to those indicators that potentially have the greatest impact on economic sustainability and the standard of living of the population. The most significant factors with high correlation coefficients and statistically significant q(FDR) values are listed (Table 2):

Table 2. Significant socio-economic factors identified by Pearson correlation analysis

Factor	r	q(FDR)	Method
investitsii_v_osnovnoi_kapital	0.8963	0.000010	Pearson
srednemesyachnaya_zarplata	0.8907	0.000010	Pearson
roznichnaya_torgovlya	0.8885	0.000010	Pearson
prozhitochnyi_minimum	0.8647	0.000030	Pearson
kol_vo_shkol	0.8411	0.000065	Pearson
obrabotka_proizvodstvo	0.8401	0.000065	Pearson

These results indicate a strong positive correlation between fixed capital investment, wage levels, retail volume and other key indicators. This allows us to conclude that these factors play a significant role in shaping the socio-economic climate and can be used as indicators for assessing the effectiveness of management decisions. In order to ensure the reliability of the analysis and interpretation of input data in the fuzzy system, a sequence of stages was implemented: exploratory data analysis (EDA), logarithmic stabilization of distributions, standardization of variables, and calculation of adaptive correlations with control of the false discovery rate (FDR). This sequence eliminates the risks associated with skewed distributions and

heteroscedasticity, thereby ensuring the correct interpretation of normalized ("z-transformed") inputs in subsequent fuzzy modeling.

The choice of methods in the proposed hybrid model is based not only on their technical efficiency but also on their ability to adequately reflect the specific characteristics of socio-economic data. Correlation analysis is employed as the first stage because it enables the rapid identification of the strength and direction of linear relationships between indicators. In socio-economic systems, where many variables may be interrelated, this step helps select the most relevant factors and exclude those with minimal contribution, thereby increasing the informativeness of the subsequent analysis.

PCA is applied to reduce dimensionality and eliminate multicollinearity among features, which is particularly important when many socio-economic indicators exhibit similar trends. PCA aggregates the original variables into new components while retaining most of the original information, reducing noise and improving the robustness of subsequent computations. Fuzzy logic is a key element of the model because socio-economic data are often accompanied by uncertainty, incompleteness, and blurred boundaries between categories. For example, the influence of a factor cannot always be strictly classified as “high” or “low” — in practice, there are intermediate states. Fuzzy logic allows such vague concepts to be formalized and integrates both quantitative and qualitative indicators, enabling a more flexible and realistic assessment of factor influence compared to traditional crisp methods.

3. RESULTS

The study conducted a hybrid analysis of the impact of socio-economic factors on key processes occurring in the region. The data used for the analysis included demographic, economic, social and business indicators of the region, such as population, average monthly wage, investment in fixed assets and many other factors. The analysis was aimed at determining the contribution of each factor to the processes of regional development, as well as identifying both positive and negative impacts. The structure of the present columns of factors in the data set reflects a variety of socio-economic and demographic indicators. The population at year-end is one of the prominent factors, which presents up-to-date information on the demographic pattern in the region. Data on the overall population, and also separately for women and men, are also taken into account, allowing for more detailed analysis of demographic change and social processes.

Demographic indicators are supplemented by such important data as birth and mortality rates, which allows for analysis of natural population growth. This data allows for assessment of long-term demographic trends and identification of possible changes in the population structure. Apart from that, signs of marriage are also taken into consideration, like the incidence of marriages and divorces and the marriage rate, which represents social change in society. Migration processes are presented through the migration balance, and signs of arrivals and departures, which allow us to find out how attractive the area is for habitation and employment. Infrastructure and educational characteristics are reflected by the number of schools, colleges, universities and students, reflecting the quality and level of accessibility of education within the region.

Economic characteristics are reflected by cost of living, labor force, employment, and employed workers, self-employed and unemployed data. These figures add to assessing the state of the labor market and standard of living of the populace. There are statistics as well related to the average nominal monthly salary and minimum salary, which is an important indicator assessing the prosperity of the region economy. Investment indicators are characterized by a scale of investment volume in fixed capital and the physical scale of investment volume, which makes it possible to assess regional economic activity. The amount of registered and functioning legal entities is the situation of business activity, and R&D spending data make it possible to approximate the level of innovative.

Production measures comprise levels of industrial production, and the share of the region in total production. Manufacturing, mining, food production, beverages, chemical and light industries are all covered by the data, and this allows us to establish the contribution of various industries to the regional economy. Agriculture is represented by statistics on gross output, crop production, livestock rearing, crop yields and the livestock and poultry numbers. These indicators are part of the analysis of the regional economy's agricultural sector. The construction sector is also covered by indicators of the volume of completed construction and the total area of residential buildings commissioned. Financial and trade aspects are covered by data on fixed assets in the economy and the volume of retail trade. All these indicators provide for the thorough examination of economic and social processes in the region.

To check the hybrid analysis, data for the city of Taraz for 2023 were used, including various socio-economic indicators seen in Table 3. The data included such parameters as number of population (431,192), birth rate (8,826), migration balance (2,375), number of the labor force (202,900 people), average monthly wage (2,658,658 tenge), investment in fixed capital (1,646,258 tenge), number of employed individuals (192,900 people), number of committed crimes (2,476), number of marriages (2,762), subsistence minimum (43,344 tenge), number of schools (106), number of active legal entities (6,341), volume of retail trade

(362,050.98 tenge), volume of industrial production (462,704.68 tenge), mining (683.3 tenge), and production in manufacturing (371,229.9 tenge).

Table 3. Test data

Year	2023
Population size	431192
Birth rate	8826
Migration balance	2375
Labor force	202.9
Average monthly salary	2 658 658
Investments in fixed capital	1 646 258
Employed population	192.9
Registered crimes	2476
Number of marriages	2762
Living wage	43344
Number of schools	106
Number of legal entities	6341
Retail trade	362050.98
Industrial production	462704.68
Mining industry	683,3
Processing industry	371229.9

The Figure 2 illustrates a comparative analysis of how different factors influence four major processes: demographic, economic, social, and industrial. Each line represents one process, allowing us to track the dynamics of factor influence across multiple domains simultaneously. The x-axis contains the full list of factors, while the y-axis shows their corresponding influence scores. This visualization enables a holistic view, making it easier to identify both common trends and process-specific differences.

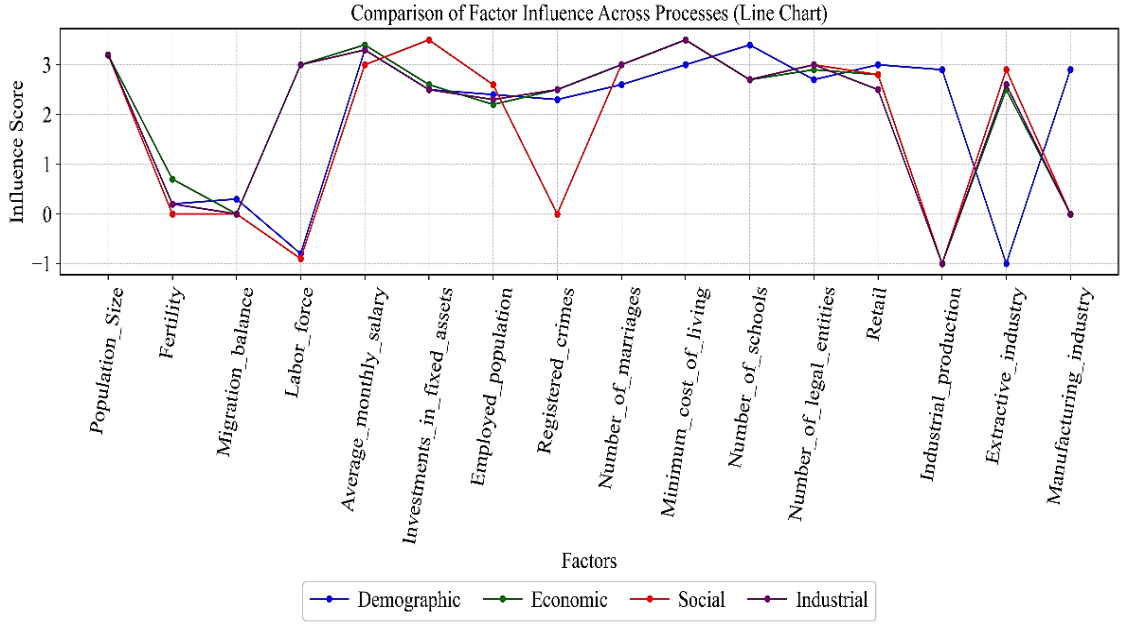

Figure 2. Comparison of factor influence across processes

As the chart demonstrates, certain factors, such as population size, average monthly salary, and investments in fixed assets, have consistently strong influence across all processes, confirming their role as universal drivers of systemic development. In contrast, factors like migration balance and registered crimes exhibit relatively low or unstable impact, indicating their more localized or short-term effects. From a comparative perspective, the demographic and economic processes show the most stable and aligned patterns, while the social process demonstrates higher variability across factors. The industrial process stands out due to the presence of both strongly positive and negative influences, suggesting structural imbalances within production-related indicators.

The joint representation on a single graph highlights three main insights:

- Cross-process consistency: several core factors maintain high influence across all domains.
- Process-specific sensitivity: certain indicators have impact only within individual processes.
- Systemic imbalances: negative or fluctuating values in Social and Industrial domains require deeper examination.

This comparative analysis provides a clearer understanding of which factors serve as universal growth drivers and which act as weak links, helping to identify priority areas for targeted policy and management decisions.

The Figure 3 visualizes the quantitative influence of 16 socio-economic factors on four processes: demographic, economic, social, and industrial. The values range from -1.04 to 3.44, where positive scores reflect strengthening effects and negative scores indicate weakening or destabilizing impacts. The graphical representation allows us to compare not only the overall level of influence but also its distribution across different domains. Particular attention was paid to the highest and lowest influence scores, since they reveal the most critical drivers and vulnerabilities of regional development.

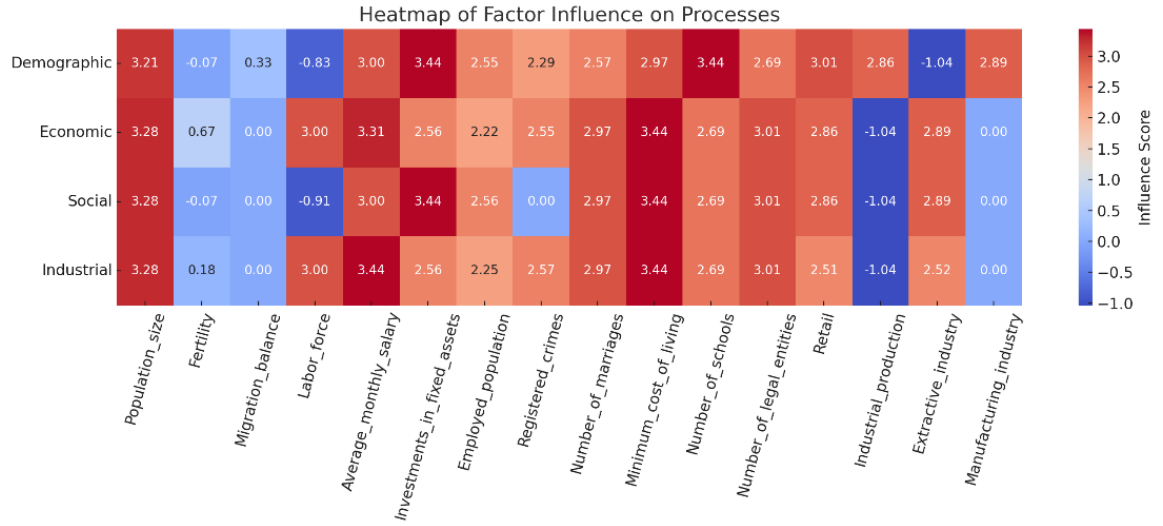

Figure 3. Heatmap of factor influence on processes

As the heatmap demonstrates, the strongest positive influence is observed for factors such as investments in fixed assets (3.44), number of schools (3.44), and minimum cost of living (3.44), which remain consistently high across multiple processes. Similarly, average monthly salary (3.31) and retail trade (3.01) also show stable significance, confirming their role as systemic drivers of socio-economic stability. At the same time, some factors reveal negative or unstable effects. For example, migration balance (-0.07 to +0.67) and manufacturing industry (-1.04) display weak or even inverse influence, especially within the social and industrial processes. This suggests that while core socio-economic indicators drive development uniformly, sectoral indicators can highlight structural risks and imbalances.

Overall, three important conclusions can be drawn:

- High-impact factors (investments, education, and salaries) demonstrate strong cross-process influence, with scores consistently above 3.0.
- Moderate factors (retail, employment, and population size) fluctuate around 2.5–3.0, indicating stable but not dominant roles.
- Low or negative factors (migration balance and manufacturing) remain below 1.0, showing weak or destabilizing effects that require deeper examination.

Numerical analysis reveals variations in the influence indices of the socioeconomic factors under consideration. Population size, average monthly wage, fixed capital investment, and educational infrastructure consistently demonstrated higher positive influence indices across demographic, economic, social, and business processes. Birth rate, labor force size, and mining output demonstrated negative or relatively low influence indices across several process categories. The distribution of influence indices reflects differences in the intensity of factor contributions within the analyzed regional system. The hybrid analytical framework combines correlation coefficients, principal component loadings, and fuzzy inference results to generate normalized influence indices for each factor. During model implementation, preprocessing parameters,

including standardization coefficients and principal component loadings, were maintained to ensure consistent application of the procedure to new input data. The computational pipeline allows for repeated model runs without recalculating transformation parameters, as previously estimated scaling and dimensionality reduction components are reused. This ensures methodological consistency when analyzing additional datasets in the same configuration.

4. DISCUSSION

The results of this study confirm the effectiveness of a hybrid analytical framework that integrates correlation analysis, PCA, and fuzzy inference in evaluating the influence of socio-economic factors on regional development. Compared to traditional single-method approaches, the hybrid model provides a more comprehensive view of complex, nonlinear, and uncertain relationships between indicators.

First, the correlation analysis stage revealed strong linear dependencies between key socio-economic indicators and target processes. For example, investments in fixed assets (r=0.8963 and q=0.000010), average monthly salary (r=0.8907 and q=0.000010), and retail trade (r=0.8885 and q=0.000010) were found to be the most influential factors. These findings are consistent with earlier studies that highlighted the role of investment and income in ensuring sustainable socio-economic growth [7], [8]. At the same time, factors such as migration balance and manufacturing industry demonstrated weak or even negative correlations, underscoring the existence of structural imbalances that require further investigation.

Second, the application of PCA allowed for dimensionality reduction and elimination of multicollinearity among variables. This step aggregated correlated indicators into principal components, enabling the identification of the most significant contributors without losing essential variance in the dataset. Similar applications of PCA in socio-economic modeling have demonstrated its robustness in handling high-dimensional data and improving prediction accuracy [12]-[15].

Third, the fuzzy inference system provided a flexible mechanism for capturing the uncertainty and vagueness inherent in socio-economic datasets. Unlike crisp models that classify factor influence as strictly "high" or "low," fuzzy logic accommodated intermediate states and enabled a nuanced interpretation of factor impacts. This approach resonates with other works applying fuzzy methods in economics and decision-making, where uncertainty and partial knowledge dominate [14], [19], [24], [25].

The visualizations presented in Figures 2 and 3 reinforced these conclusions. The line chart demonstrated that factors such as population size, average monthly salary, and Investments in fixed assets maintain consistently high influence across demographic, economic, social, and industrial processes, positioning them as universal drivers of systemic development. In contrast, the heatmap highlighted process-specific sensitivities, revealing that while education-related indicators (e.g., number of schools) have cross-cutting effects, industrial outputs are prone to negative fluctuations, signaling sectoral vulnerabilities. Comparatively, the hybrid model aligns with recent research where integrated approaches outperformed classical regression or machine learning models in socio-economic forecasting [20]-[23]. Studies in energy planning, waste management, and CO_2 emissions prediction similarly emphasized that hybrid methods enhance accuracy and interpretability, bridging the gap between statistical rigor and real-world applicability.

Overall, this discussion underscores three main insights:

a. Universal growth drivers: investments, education, and wage levels consistently demonstrate strong positive effects across domains.
b. Process-specific variation: demographic and economic processes exhibit stable trends, whereas social and industrial processes show higher variability.
c. Structural imbalances: weak or negative influence of factors such as migration and industrial outputs indicates areas requiring targeted policy interventions.

These findings validate the potential of the hybrid fuzzy-PCA approach as a decision-support tool for regional planners and policymakers, offering both methodological robustness and practical applicability.

4.1. Limitations and future work

While the proposed hybrid fuzzy-PCA model has shown effectiveness in analyzing socio-economic factors that affect regional development, it has some limitations that need to be recognized. First, the quality and scope of the input data limit the model's accuracy and how well it can be applied generally. The analysis used aggregated annual data for the city of Taraz, which might overlook yearly trends and differences across various districts or population groups. More detailed data, like monthly statistics or broken-down indicators by age, sector, or location, could improve the model's sensitivity and allow for detailed diagnostics. Second, the model assumes that relationships between factors and outcomes are linear. This might not fully reflect the complex, nonlinear interactions at play. While using fuzzy logic provides some flexibility and clarity, it could be further improved by adding techniques like nonlinear dimensionality reduction, such as t-SNE or UMAP,

or by using hybrid neuro-fuzzy systems. Third, the current study is limited to a single region, Taraz. Although the method can be adjusted, its effectiveness in different regions, socio-economic situations, or countries hasn't been tested. Applying the model in various geographic areas and administrative units would help verify its reliability and make it more useful for national planning and policy development. Additionally, the model does not currently account for changes over time or forecasting abilities, which are important for planning and evaluating long-term policy effects. Future research should look into combining time series models or recurrent neural networks to enable trend analysis and predict regional outcomes under different policy scenarios. Lastly, the model now relies on fuzzy rules and membership functions set by experts. While these are easy to understand, they can bring in subjective bias. Future versions could improve by using data-driven methods to create fuzzy rules or by incorporating adaptive learning to optimize inference parameters based on past outcomes. In future research, attention should focus on:

- Expanding the dataset with better spatial and temporal detail.
- Comparing the hybrid model's performance with other methods, like AHP, neural networks, and regression trees.
- Improving the automation of fuzzy rule tuning through machine learning.
- Applying the model in other regions of Kazakhstan and globally.
- Developing a web-based or GIS-integrated decision support system based on the model framework.

Addressing these areas will help create a more reliable, dynamic, and broadly useful decision-making tool for regional planning and socio-economic policy evaluation.

5. CONCLUSION

This study proposed and applied a hybrid analytical framework integrating correlation analysis, PCA, and fuzzy inference to evaluate the influence of socio-economic factors on regional development processes. The model enables a multidimensional assessment by combining statistical relationships with fuzzy reasoning, thereby producing structured influence scores for each factor. The results confirm the applicability of the proposed hybrid approach for analyzing complex socio-economic systems characterized by multivariate dependencies and uncertainty.

The developed methodology demonstrates reproducibility and adaptability, allowing consistent application to regional datasets under unified preprocessing and modeling procedures. In this sense, the framework contributes to methodological advancement in regional socio-economic analysis by offering an integrated and systematic evaluation tool suitable for data-driven assessment and comparative regional studies.

At the same time, the present study has several limitations, including the use of aggregated annual data, the focus on a single region, and the dependence on expert-defined fuzzy rules. Therefore, future research should focus on extending the dataset with greater spatial and temporal detail, validating the framework across different regions, and improving the automation of fuzzy rule tuning through data-driven methods. In addition, the integration of forecasting components and GIS- or web-based decision-support tools may further increase the practical value of the proposed framework for regional planning and socio-economic policy evaluation.

FUNDING INFORMATION

This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP19677451).

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Ayman Akynbekova	✓				✓		✓			✓	✓		✓	✓
Raikhan Muratkhan		✓		✓	✓				✓	✓	✓			
Zhanar Lamasheva		✓		✓		✓		✓		✓				
Ayagoz Mukhanova	✓		✓		✓			✓	✓		✓			
Gulbakhar Yussupova		✓		✓	✓				✓	✓	✓			
Serik Eslyamov		✓		✓	✓				✓		✓			
Saya Santeyeva	✓		✓	✓		✓		✓		✓	✓		✓	✓
Alfiya Abdrakhmanova		✓		✓	✓				✓		✓			

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, Ayagoz Mukhanova, upon reasonable request. Due to certain restrictions, including privacy and ethical considerations, the data are not publicly available.

REFERENCES

[1] A. Menezes *et al.*, "Socio-Economic Impact of the Brumadinho Landslide: A Hybrid MCDM-ML Approach," *Sustainability*, vol. 16, no. 18, pp. 1–32, Sep. 2024, doi: 10.3390/su16188187.

[2] Z. Li, G. Wang, D. Lin, and A. Mashhadi, "Hybrid approach for accurate water demand prediction using socio-economic and climatic factors with ELM optimization," *Heliyon*, vol. 10, no. 3, pp. 1–14, Feb. 2024, doi: 10.1016/j.heliyon.2024.e25028.

[3] W. F. Mbasso *et al.*, "Hybrid modeling approach for precise estimation of energy production and consumption based on temperature variations," *Scientific Reports*, vol. 14, no. 1, pp. 1–22, Oct. 2024, doi: 10.1038/s41598-024-75244-0.

[4] M. Abou-Zeid and M. Ben-Akiva, "Hybrid choice models," in *Handbook of Choice Modelling, Second Edition*, Edward Elgar Publishing, 2024, pp. 489–521, doi: 10.4337/9781800375635.00026.

[5] A. Zanellati, D. Di Mitri, M. Gabbrielli, and O. Levrini, "Hybrid Models for Knowledge Tracing: A Systematic Literature Review," *IEEE Transactions on Learning Technologies*, vol. 17, pp. 1021–1036, 2024, doi: 10.1109/TLT.2023.3348690.

[6] S. Suravi, "Training and development in the hybrid workplace," *Learning Organization*, vol. 31, no. 1, pp. 48–67, Feb. 2024, doi: 10.1108/TLO-10-2022-0119.

[7] M. Z. Salim, Y. Qiang, B. Dixon, and J. Collins, "A Disparate Disaster: Spatial Patterns of Building Damage Caused by Hurricane Ian and Associated Socio-Economic Factors," *Remote Sensing*, vol. 16, no. 20, pp. 1–18, Oct. 2024, doi: 10.3390/rs16203792.

[8] Z. Karakayaci, "Study of Socio-Economic Development Effects on Agricultural Lands' Value," *Pakistan Journal of Agricultural Sciences*, vol. 61, no. 3, pp. 743–755, 2024, doi: 10.21162/PAKJAS/24.148.

[9] K. Kraus, N. Kraus, and O. Marchenko, "Forecasting the Innovative and Digital Strength of Ukraine'S Economy on the Basis of Correlation-Regress Analysis," *Baltic Journal of Economic Studies*, vol. 10, no. 3, pp. 180–192, Sep. 2024, doi: 10.30525/2256-0742/2024-10-3-180-192.

[10] T. W. Nurdiani, Nofirman, M. R. Aulia, G. W. Putra, and I. H. Kusnadi, "Analysis of Digital Literacy Sources to Identify The Relationship Between Population Income, Socio-Economic and Subjective Well-Being," *Jurnal Sistim Informasi dan Teknologi*, pp. 42–47, May 2024, doi: 10.60083/jsisfotek.v6i2.350.

[11] J. Lin, K. Wei, and Z. Guan, "Exploring the connection between morphological characteristic of built-up areas and surface heat islands based on MSPA," *Urban Climate*, vol. 53, p. 101764, Jan. 2024, doi: 10.1016/j.uclim.2023.101764.

[12] X. He, "Principal Component Analysis (PCA)," in *Geographic Data Analysis Using R*, Singapore: Springer Nature Singapore, 2024, pp. 155–165, doi: 10.1007/978-981-97-4022-2_8.

[13] W. K. Härdle, L. Simar, and M. R. Fengler, "Principal component analysis," in *Applied multivariate statistical analysis, Cham: Springer International Publishing*, 2024, pp. 309–345, doi: 10.1017/9781009218276.009.

[14] C. Challoumis, "Fuzzy Logic Concepts and the Q.E. (Quantification of Everything) Method in Economics," *Web of Scholars: Multidimensional Research Journal*, vol. 3, no. 4, pp. 1–25, 2024.

[15] A. H. Alamoodi *et al.*, "A Novel Evaluation Framework for Medical LLMs: Combining Fuzzy Logic and MCDM for Medical Relation and Clinical Concept Extraction," *Journal of Medical Systems*, vol. 48, no. 1, p. 81, Aug. 2024, doi: 10.1007/s10916-024-02090-y.

[16] K. Alimhan, N. Otsuka, M. N. Kalimoldayev, and N. Tasbolat, "Practical output tracking for a class of uncertain nonlinear time-delay systems via state feedback," in *MATEC Web of Conferences*, vol. 189, pp. 1–8, Aug. 2018, doi: 10.1051/matecconf/201818910027.

[17] G. Bakhadirova, N. Tasbolatuly, A. Tanirbergenova, A. Dautova, A. Akanova, and Y. Ulikhina, "Computer Simulation Control of High-Order Nonlinear Systems using Feedback," *Journal of Applied Data Sciences*, vol. 5, no. 3, pp. 1096–1109, Sep. 2024, doi: 10.47738/jads.v5i3.275.

[18] N. Tasbolatuly, K. Alimhan, A. Yerdenova, G. Bakhadirova, A. Nazyrova, and M. Kaldarova, "Using Computer Modeling for Tracking high-order Nonlinear Systems with Time-Delay," in *2024 IEEE 4th International Conference on Smart Information Systems and Technologies (SIST)*, Astana, Kazakhstan: IEEE, May 2024, pp. 154–158, doi: 10.1109/SIST61555.2024.10629397.

[19] D. Tešić, D. Božanić, and M. Khalilzadeh, "Enhancing Multi-Criteria Decision-Making with Fuzzy Logic: An Advanced Defining Interrelationships Between Ranked II Method Incorporating Triangular Fuzzy Numbers," *Journal of Intelligent Management Decision*, vol. 3, no. 1, pp. 56–67, Mar. 2024, doi: 10.56578/jimd030105.

[20] R. Tian *et al.*, "Socio-economic correlation analysis and hybrid artificial neural network model development for provincial waste electrical and electronic equipment generation forecasting in China," *Journal of Cleaner Production*, vol. 418, p. 138076, Sep. 2023, doi: 10.1016/j.jclepro.2023.138076.

[21] S. Kumar, "A novel hybrid machine learning model for prediction of CO2 using socio-economic and energy attributes for climate change monitoring and mitigation policies," *Ecological Informatics*, vol. 77, p. 102253, Nov. 2023, doi: 10.1016/j.ecoinf.2023.102253.

[22] J. Qian, S. Zeritis, M. Larsen, and M. Torok, "The application of spatial analysis to understanding the association between area-level socio-economic factors and suicide: a systematic review," *Social Psychiatry and Psychiatric Epidemiology*, vol. 58, no. 6, pp. 843–859, Jun. 2023, doi: 10.1007/s00127-023-02441-z.
[23] N. Kumar, K. Namrata, and A. Samadhiya, "Techno socio-economic analysis and stratified assessment of hybrid renewable energy systems for electrification of rural community," *Sustainable Energy Technologies and Assessments*, vol. 55, p. 102950, Feb. 2023, doi: 10.1016/j.seta.2022.102950.
[24] A. Y. Gunal and R. Mehdi, "Application of a new fuzzy logic model known as 'SMRGT' for estimating flow coefficient rate," *Turkish Journal of Engineering*, vol. 8, no. 1, pp. 46–55, Jan. 2024, doi: 10.31127/tuje.1225795.
[25] T. Meng *et al.*, "New Framework for Fuzzy Logic Reasoning: A Robust Control Theoretic Approach," *International Journal of Fuzzy Systems*, vol. 26, no. 2, pp. 463–481, Mar. 2024, doi: 10.1007/s40815-023-01606-x.
[26] P. Dhivya, A. Karthikeyan, S. Pradeep, and H. Umamaheswari, "Natural Language Processing in Generative Adversarial Network," *Generative Artificial Intelligence: Concepts and Applications*, pp. 53–79, 2025, doi: 10.1002/9781394209835.ch4.
[27] U. Aitimova *et al.*, "Data generation using generative adversarial networks to increase data volume," *International Journal of Electrical and Computer Engineering*, vol. 14, no. 2, pp. 2369–2376, Apr. 2024, doi: 10.11591/ijece.v14i2.pp2369-2376.
[28] V. Vlasenko *et al.* "Devising a fast median filtering procedure for aligning the noise background of a digital frame," *Eastern-European Journal of Enterprise Technologies*, vol. 2, 2025, doi: 10.15587/1729-4061.2025.324680.

BIOGRAPHIES OF AUTHORS

Ayman Akynbekova in 1998, she graduated from the Kazakh State Academy of Management, majoring in "Information Systems in the Economy". In 2006, she received a master's degree in information systems. In 2022, she entered doctoral studies at the L.N. Gumilyov Eurasian National University, majoring in "Information Systems". Since 2022, she is currently a doctoral student at this university. The topic of the dissertation work is "Implementation of fuzzy models of decision-making in social processes". She can be contacted at email: ataiman77@mail.ru.

Raikhan Muratkhan in 2002 he graduated from the Karaganda State University named after E.A. Buketov with a degree in Applied Mathematics. In 2019 he defended her dissertation in the specialty "6D060200–computer sciences" and received a Ph.D. Currently, he is Associate Professor at the Department of Applied mathematics and Informatica of Karaganda Buketov University. He is the author of more than 70 scientific papers, including 1 monograph, 8 articles in the Scopus database. Scientific interests-image processing, pattern recognition theory, data mining, and information security. He can be contacted at email: raykhan.muratkhan@mail.ru.

Zhanar Lamasheva in 2006, she graduated from Kazakh National Technical University named after K.I. Satbayev, Almaty, Kazakhstan. In 2004, she received a Master's degree in Information systems. In 2015, she graduated from the doctoral program "Kazakh National Technical University named after K.I. Satbayev", specialty 6D070300 - " Information systems ". From 2017 to the present, she has been a Ph.D. of the L.N. Gumilyov Eurasian National University, Astana, Kazakhstan. She is a co-author of 6 publications. Her research interests include knowledge bases, big data, artificial intelligence, and machine learning. She can be contacted at email: lamasheva_zhb@enu.kz.

Ayagoz Mukhanova received her Ph.D. in 2015 in Information Systems from L.N. Gumilyov Eurasian National University, Kazakhstan. Currently, she is an associate professor of the Department of Information Systems at the same university. Her research interests include artificial intelligence and decision making. She can be contacted at email: ayagoz198302@mail.ru.

Gulbakhar Yussupova graduated from the Almaty Institute of Energy and Communications in 2005 with a degree in Multichannel Telecommunication Systems. In 2017 she defended her doctoral dissertation “6D071900 – Radio engineering, electronics and telecommunications” and received a Ph.D. degree. Currently she is an associate professor at the Department of Electronics, Telecommunications and Space Technologies at the Kazakh National Research University named after K. Satpayev". She is the author of more than 80 scientific works, including 2 monographs, 9 articles in the Scopus database. Scientific interests – development of fiber-optic bragg meshes for use in telecommunication systems. She can be contacted at email: gulbaharusupova6@gmail.com.

Serik Eslyamov graduated from Karaganda State University named after E.Buketov in 1987 with a degree in Physics. In 1993, he defended his Ph.D. thesis in the specialty "05.25.05 - Information Systems and Processes" at the V. Glushkov Institute of Cybernetics (Kiev) and received the degree of Candidate of Technical Sciences. He began his career in 1987 as a lecturer at the Department of Computer Science and Computer Systems of Kostanay State University. Currently, he is an acting professor at the Department of Radio Engineering, Electronics and Telecommunications of the L.N. Gumilyov Eurasian National University, a non-profit Joint-Stock Company. He is the author of more than 120 scientific papers, including 3 monographs, 2 articles in the Scopus database. Research interests – information systems, artificial intelligence, robotics, neural networks, and machine learning. He can be contacted at email: eslyamov@gmail.com.

Saya Santeyeva in 2015, she graduated from the L.N. Gumilyov Eurasian National University with a degree in Information Systems. In 2008, he received a Master's degree in Computer Science. In 2021, she graduated from the doctoral program "L.N. Gumilyov Eurasian National University", specialty "6D070200 – Automation and control". From 2022 to the present, he has been a Doctor of Philosophy Ph.D. in the specialty "6D070200 – Automation and control" of the L.N. Gumilyov Eurasian National University. She is the author of more than 40 works. Her research interests include engineering in telecommunications, computer networks, information security, and data transmission security over networks. She can be contacted at email: saya_santeeva@mail.ru.

Alfiya Abdrakhmanova received a bachelor's degree in technical sciences in specialty 6M070300 - "Information systems" at the Eurasian National University (ENU) named after L.N. Gumilyov, Astana, Kazakhstan, 2022. She is a co-author of 3 publications. Her research interests include knowledge bases, big data, artificial intelligence and machine learning. She received a Master's degree in Technical Sciences in the field of information systems from the Caspian State University of Technology and Engineering named after Sh. Yesenov in 2009. She can be contacted at email: alfiya_zagievna@mail.ru.