



**SYSTEM FOR PREDICTING TRAIN MOVEMENTS AND  
REDUCING DELAYS BASED ON ARTIFICIAL INTELLIGENCE IN  
RAILWAY TRANSPORT LOGISTICS.**

*Islomov Diyorbek Zafar o'g'li, teacher, Orcid ID: 0009-0009-2982-8630,  
[diyorbekshaxrisabz@gmail.com](mailto:diyorbekshaxrisabz@gmail.com),*

*Abdugalilov Farrux, Karimov Hotam, Geldiyorova Malika,  
Jurayev Abdumalik, students,  
University of Economics and Pedagogy, Karshi, Uzbekistan*

**Annotation.** *This article investigates the application of artificial intelligence and machine learning technologies for predicting train delays in the Uzbekistan railway transport system. The study comparatively analyzes Random Forest, Gradient Boosting, and LSTM neural network models. The experimental data includes train movement statistics for the Tashkent—Samarkand—Bukhara route during 2022—2024, weather data, and infrastructure condition indicators. Results demonstrate that the Gradient Boosting model can predict delays with 87.3% accuracy, showing 23% higher efficiency compared to traditional statistical methods. The research findings can be applied within the framework of the “Digital Uzbekistan – 2030” strategy for railway system digitalization.*

**Keywords:** *artificial intelligence, machine learning, railway logistics, delay prediction, Random Forest, Gradient Boosting, LSTM, Uzbekistan railways.*

**Аннотация.** *В данной статье исследуется применение технологий искусственного интеллекта и машинного обучения для прогнозирования задержек поездов в железнодорожной транспортной системе Узбекистана. В исследовании проводится сравнительный анализ моделей нейронных сетей Random Forest, Gradient Boosting и LSTM. Экспериментальные данные включают статистику движения поездов по маршруту Ташкент—Самарканд—Бухара за 2022–2024 годы, данные о погоде и показатели*



состояния инфраструктуры. Результаты показывают, что модель Gradient Boosting может прогнозировать задержки с точностью 87,3%, демонстрируя на 23% большую эффективность по сравнению с традиционными статистическими методами. Результаты исследования могут быть применены в рамках стратегии «Цифровой Узбекистан – 2030» по цифровизации железнодорожной системы.

**Ключевые слова:** искусственный интеллект, машинное обучение, железнодорожная логистика, прогнозирование задержек, Random Forest, Gradient Boosting, LSTM, железные дороги Узбекистана.

**Annotatsiya.** Ushbu maqolada O'zbekiston temiryo'l transport tizimida poyezdlar kechikishlarini bashorat qilish uchun sun'iy intellekt va mashinaviy o'rganish texnologiyalarini qo'llash imkoniyatlari tadqiq etilgan. Tadqiqotda Random Forest, Gradient Boosting va LSTM neyron tarmoq modellari solishtirma tahlil qilingan. Tajriba ma'lumotlari sifatida Toshkent—Samarqand—Buxoro yo'nalishi bo'yicha 2022—2024 yillar orasidagi poyezd harakati statistikasi, ob-havo ma'lumotlari va infratuzilma holati ko'rsatkichlari ishlatilgan. Natijalar shuni ko'rsatdiki, Gradient Boosting modeli 87,3% aniqlik bilan kechikishlarni bashorat qilishga qodir bo'lib, an'anaviy statistik usullarga nisbatan 23% yuqori samaradorlik ko'rsatdi. Tadqiqot natijalari O'zbekiston temiryo'l tizimini raqamlashtirish va "Raqamli O'zbekiston – 2030" strategiyasi doirasida amaliy qo'llanilishi mumkin.

**Kalit so'zlar:** sun'iy intellekt, mashinaviy o'rganish, temiryo'l logistika, kechikishlarni bashorat qilish, Random Forest, Gradient Boosting, LSTM, O'zbekiston temir yo'llari.

## INTRODUCTION

Railway transport is one of the most important sectors of the economy of Uzbekistan. Joint-Stock Company "Uzbekistan Railways" manages the country's railway network, which is more than 2740 km long and transports millions of passengers and millions of tons of cargo annually. However, train delays remain a serious problem for railway systems around the world. Delays negatively affect



passenger comfort, the efficiency of the logistics chain, and the economic performance of the organization [1, 2].

International experience shows that machine learning algorithms are an effective tool for predicting railway delays. In particular, the Random Forest Classifier model has shown high results in detecting delays in the Indian railway system [3]. In European railway networks, the distribution of delays has been modeled using Spatio-Temporal Graph Convolutional Networks [4, 5].

Digitalization processes in the railway system of Uzbekistan are actively being carried out within the framework of the “Digital Uzbekistan – 2030” strategy. The “Single Window” national information system has been introduced, freight documents have been digitized, and an electronic ticket system has been launched. However, an intelligent system designed to predict train delays has not yet been implemented [6].

Research objective: To develop a prediction system based on artificial intelligence on the example of the railway routes of Uzbekistan (Tashkent—Samarkand—Bukhara) and to conduct a comparative analysis of the effectiveness of various machine learning algorithms.

Research objectives: (1) to identify factors affecting railway delays; (2) to conduct a comparative analysis of various machine learning models; (3) to propose a prediction system model adapted to the conditions of Uzbekistan.

## METHODS

**Data set.** The study used train traffic data on the Tashkent-Samarkand-Bukhara route from January 2022 to December 2024. The dataset consists of the following components:

*Table 1. Dataset composition*

T/r	Data type	Number of variables	Source
1	Train traffic statistics	8	Uzbekistan TY JSC
2	Weather data	6	Uzgidromet



3	Infrastructure status	5	Technical reports
4	Route characteristics	4	GIS data
	Total number of records: 18,450	<b>23</b>	

Train movement variables: train number, type (passenger/freight), departure time (scheduled and actual), arrival time, distance between stations, number of wagons and load capacity. Weather variables: temperature, precipitation, wind speed, visibility distance, snow cover and humidity level. Infrastructure variables: rail condition index, date of last repair, type of signaling system, electrification level and station capacity.

Data processing. The following operations were performed during the data preparation stage: missing values were filled with the median value (in 5.2% of records); categorical variables were converted to numeric form using the One-Hot Encoding method; numeric variables were normalized to the range [0, 1] using the Min-Max normalization method; data were randomly divided into 80% training and 20% test sets.

Machine learning models. Three machine learning models were tested in the study:

Random Forest (RF) — an ensemble learning method that combines multiple decision trees. Parameters: `n_estimators=200`, `max_depth=15`, `min_samples_split=5`. Random Forest reduces the problem of overfitting by learning based on randomly selected features in each tree [7].

Gradient Boosting (GB) — a method in which each subsequent model is aimed at correcting the errors of the previous model by building successive decision trees. Parameters: `n_estimators=300`, `learning_rate=0.05`, `max_depth=8`. Implemented using the XGBoost library [8].

LSTM (Long Short-Term Memory) — a recurrent neural network architecture specialized in processing time series data. Structure: 2 LSTM layers (64



and 32 neurons), Dropout (0.2), Dense output layer. Implemented in the TensorFlow/Keras library [9].

Evaluation criteria. The following criteria were used to evaluate the performance of the models: Accuracy — the proportion of correct predictions; Precision — the proportion of true positives among positive predictions; Recall — the proportion of correctly identified among all true positives; F1-Score — the harmonic mean of Precision and Recall; MAE (Mean Absolute Error) — the error in predicting the delay time in minutes.

## RESULTS

Analysis of delay factors. Data analysis showed that during the studied period, the average delay of trains on the Tashkent-Samarkand-Bukhara route was 12.7 minutes. 68.4% of delays were less than 15 minutes, 24.1% were between 15 and 45 minutes, and 7.5% were more than 45 minutes. The main reasons for delays were: technical malfunctions (31.2%), weather conditions (22.7%), increased passenger flow (19.4%), infrastructure repair work (15.8%) and other factors (10.9%).

Comparison of model efficiency.

**Table 2. Comparative results of machine learning models**

Model	Accuracy	Precision	Recall	F1-Score	MAE (daq)
Random Forest	84,1%	82,5%	81,3%	81,9%	8,4
<b>Gradient Boosting</b>	<b>87,3%</b>	<b>86,1%</b>	<b>85,7%</b>	<b>85,9%</b>	<b>6,2</b>
LSTM	85,8%	84,2%	83,9%	84,0%	7,1
Traditional statistics	64,2%	61,8%	59,5%	60,6%	14,7

As can be seen from Table 2, the Gradient Boosting model showed the highest results across all evaluation criteria. Its accuracy (87.3%) was 3.2% higher



than Random Forest, 1.5% higher than LSTM, and 23.1% higher than traditional statistical methods. Gradient Boosting also has the lowest error rate in terms of MAE — 6.2 minutes, meaning the model predicts the delay time with an average error of 6.2 minutes. Feature Importance Analysis. The results of the feature importance analysis for the Gradient Boosting model revealed the following most influential factors:

*Table 3. The most important factors affecting delay.*

T/r	Factor Name	Importance	Category
1	Delay from previous station	0,247	Traffic
2	Rail condition index	0,189	Infrastructure
3	Weather conditions (precipitation)	0,156	Weather
4	Time of day (hours)	0,132	Time
5	Passenger flow	0,118	Demand

The results of Table 3 show that the most important factor in predicting delays is the delay from the previous station (0.247). This result is also consistent with the results of international studies - the cascade effect of delays (delay propagation) is one of the main problems of railway systems [4, 5].

## DISCUSSION

The results of the study confirm the high efficiency of artificial intelligence technologies in predicting delays in the Uzbek railway system. The Gradient Boosting model showed the best result with an accuracy of 87.3%. This result is comparable to the results obtained in the Indian railway system — the Random Forest Classifier showed an accuracy of 82-85% [3]. The difference is that in our study, the inclusion of infrastructure condition and weather data significantly increased the accuracy of the model.

The LSTM model showed a strong result in time series analysis (85.8%), but slightly lower than Gradient Boosting. This is due to the relatively small data size



(18,450 records), and deep learning models usually work better on larger datasets. It is expected that the results of the LSTM model will improve further as the data size increases.

The feature importance analysis showed that the cascade effect of delays (delay propagation) is the most important factor. This result is consistent with the results of international studies, in particular studies conducted on European railway networks [4, 5]. It was also found that track condition and weather conditions are particularly important factors for Uzbekistan - this is explained by the climatic conditions of Central Asia (high temperatures in summer, cold in winter).

Practical significance of the study. The proposed forecasting system can be integrated into the “Situation Center” of the Unified Dispatch Center of “Uzbekistan Railways” JSC. This will allow dispatchers to predict train delays, dynamically adjust schedules and inform passengers in a timely manner.

Limitations of the study. The study is based on data from only one route (Tashkent-Samarkand-Bukhara). In the future, it is advisable to conduct a larger study covering all major routes. Also, additional technical infrastructure will be required to implement a real-time forecasting system.

## CONCLUSION

Tadqiqot natijalariga asoslanib quyidagi xulosalar chiqarildi:

1. Artificial intelligence and machine learning technologies allow for effective prediction of train delays in the Uzbek railway system. Three models were tested, and all of them showed superior results compared to traditional statistical methods.
2. The Gradient Boosting (XGBoost) model showed the highest efficiency with an accuracy of 87.3% and an average error of 6.2 minutes. This model is recommended for implementation in the Uzbek railway system.
3. The most important factors in predicting delays are: delay from the previous station (cascade effect), track condition and weather conditions. Monitoring these factors can increase dispatching efficiency.



4. The proposed system can be integrated into the digitalization projects of “Uzbekistan Railways” JSC within the framework of the “Digital Uzbekistan – 2030” strategy and serve to increase the efficiency of transport logistics.

## REFERENCES

1. Chopra S., Meindl P. Supply Chain Management: Strategy, Planning, and Operation. – Pearson Education, 2021. – 528 p.
2. Bowersox D.J., Closs D.J., Cooper M.B. Supply Chain Logistics Management. – McGraw-Hill, 2020. – 496 p.
3. Kumar V., Sharma S.K. Improving On-Time Performance: Predicting Train Delays with Machine Learning Techniques // Springer, 2024. – P. 112–125.
4. Nguyen V.D.A., Li Z. Deep Learning to Identify the Spatio-Temporal Cascading Effects of Train Delays in a High-Density Network // GeoAI '25, ACM, 2025.
5. Marković N. et al. Railway Network Delay Evolution: A Heterogeneous Graph Neural Network Approach // arXiv:2303.15489, 2023.
6. O'zbekiston Respublikasi Prezidentining 2020 yil 5 oktyabrdagi PF-6079-son Farmoni. “Raqamli O'zbekiston – 2030” strategiyasi. – lex.uz.
7. Breiman L. Random Forests // Machine Learning. – 2001. – Vol. 45. – P. 5–32.
8. Chen T., Guestrin C. XGBoost: A Scalable Tree Boosting System // Proceedings of KDD, ACM, 2016. – P. 785–794.
9. Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. – 1997. – Vol. 9(8). – P. 1735–1780.
10. Minbashi N. et al. Enhancing Freight Train Delay Prediction with Simulation-Assisted Machine Learning // IET Intelligent Transport Systems. – 2024.
11. “O'zbekiston temir yo'llari” AJ rasmiy sayti. Raqamlashtirish va zamonaviy axborot texnologiyalarini joriy etish. – railway.uz, 2024.
12. Бобомуродов Б. С., Исломов Д. З. ИНФОРМАЦИОННАЯ СИСТЕМА СБОРА И ВИЗУАЛИЗАЦИИ ДАННЫХ С ГЕОРАСПРЕДЕЛЁННЫХ ДАТЧИКОВ //Modern education and development. – 2026. – Т. 45. – №. 2. – С. 263-273.



13. Бобомуродов Б. С., Исломов Д. З. СИСТЕМА МОНИТОРИНГА КАЧЕСТВА ВОЗДУХА НА ОСНОВЕ РАСПРЕДЕЛЁННОЙ СЕТИ IoT-ДАТЧИКОВ AIR QUALITY MONITORING SYSTEM BASED ON A DISTRIBUTED IoT SENSOR NETWORK //Modern education and development. – 2026. – Т. 45. – №. 2. – С. 251-262.
14. Islomov D. Z. ARGO UML DASTURI YORDAMIDA AXBOROT TIZIMINI LOYIHALASHTIRISH USULLARI //Экономика и социум. – 2024. – №. 12-2 (127). – С. 367-372.
15. Zafar o'g'li I. D. et al. WEBASSEMBLY (WASM): BRAUZERDA YUQORI UNUMDORLIKKA ERISHISH VA NATIVE TEZLIKDA ISHLAYOTGAN VEB-ILOVALARNI QURISH IMKONIYATLARI //Modern education and development. – 2026. – Т. 47. – №. 4. – С. 11-18.
16. Zafar o'g'li I. D. et al. DIFFERENSIAL TENGLAMALARNI SONLI YECHISH: RUNGE-KUTTA METODLARINING TAQQOSLAMA TAHLILI VA AMALIY QO'LLANILISHI //Modern education and development. – 2026. – Т. 47. – №. 4. – С. 3-10.
17. Zafar o'g'li I. D. et al. KVANT HISOBLASH VA KRIPTOGRAFIYAGA TA'SIRI: POST-KVANT KRIPTOGRAFIYASI VA MAVJUD ALGORITMLAR XAVFSIZLIGI //Ta'lim innovatsiyasi va integratsiyasi. – 2026. – Т. 66. – №. 3. – С. 331-334.
18. Zafar o'g'li I. D. et al. EDGE COMPUTING VA IOT INTEGRATSIYASI: REAL-TIME MA'LUMOTLAR QAYTA ISHLASH //Ta'lim innovatsiyasi va integratsiyasi. – 2026. – Т. 66. – №. 3. – С. 326-330.