An intelligent model for predicting the behavior of soil conditions depending on external weather conditions

Oleslav Antamoshkin[1,2]*, *Anton* Mikhalev[2], *Andrey* Menshenin[1], and *Alexander* Lukishin[1]

[1]Siberian Federal University, 79 Svobodny pr., Krasnoyarsk, 660041, Russian Federation
[2]Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Ave, Krasnoyarsk, 660037, Russia

Abstract. The study focuses on the development of an intelligent system for monitoring and forecasting the condition of road surfaces in Eastern Siberia, addressing challenges posed by extreme climate fluctuations. Seasonal variations in temperature and soil moisture critically impact the load-bearing capacity of road structures, leading to accelerated wear, deformations, and safety risks. This research integrates advanced machine learning models, including LSTM, Transformer, TCN, and XGBoost, to predict changes in road conditions based on meteorological and soil data. Field measurements of soil elasticity modules were analyzed to assess seasonal impacts, with LSTM demonstrating the highest accuracy (MSE: 0.025, MAE: 0.0045). The findings confirm that freezing increases soil stability during winter, while spring thawing causes significant weakening due to over-saturation. Strengthening road bases with 30% sludge improved their durability and resilience under heavy loads. The proposed system combines real-time monitoring with predictive analytics, offering a practical tool for infrastructure management in extreme climates. Key outcomes include optimized maintenance schedules, recommendations for spring traffic restrictions, and strategies to mitigate road degradation. This work highlights the potential of machine learning in enhancing the efficiency and safety of road infrastructure, contributing to sustainable transportation in cold regions.

1 Introduction

Road infrastructure plays a critical role in enabling transportation, economic activities, and social interaction, particularly in vast regions like Eastern Siberia. However, the extreme climate conditions in this region, characterized by sharp seasonal variations in temperature and soil moisture, pose significant challenges to maintaining road durability and safety. These fluctuations adversely affect the load-bearing capacity of road surfaces, leading to accelerated wear, deformations, and frequent repairs. The spring thaw, in particular, causes over-saturation of soil, drastically reducing its strength and increasing maintenance costs.

* Corresponding author: oleslav@mail.ru

This study aims to develop an intelligent system for monitoring and forecasting the condition of road surfaces in Eastern Siberia. By leveraging machine learning algorithms, the system will predict seasonal variations in road conditions using meteorological and soil elasticity data. The ultimate goal is to provide actionable insights for road maintenance planning and traffic safety measures, especially during critical periods such as spring thaw.

Integrating advanced technologies, such as machine learning and real-time data from sensors, offers a promising solution to the region's road infrastructure challenges. This research not only contributes to the scientific understanding of road behavior under extreme climates but also provides practical tools for optimizing maintenance efforts, reducing costs, and improving safety for heavy transport operations in cold regions.

2 Related works

Modern studies in road condition monitoring and forecasting focus on the integration of advanced technologies such as machine learning, big data, and IoT systems. These technologies enable more precise assessments and effective management of road infrastructure, especially in regions with challenging climatic conditions.

A significant portion of the research emphasizes the role of smart sensors and artificial intelligence in road condition monitoring. Studies [1, 2, 6] highlight the potential of computer vision, smartphone-based data collection, and various sensor technologies for detecting road surface defects. These methods not only enhance the accuracy of evaluations but also reduce maintenance costs by allowing targeted interventions.

Predictive models, especially those based on machine learning, have shown promising results for forecasting temporal changes in road conditions. Articles [3, 7, 9] demonstrate the superiority of deep learning models such as LSTM and Transformer over traditional approaches due to their ability to capture nonlinear relationships and temporal dependencies in the data. For instance, [3] discusses the use of LSTM and Transformer models to evaluate real-time crash risks by analyzing traffic flow and driver behavior.

Climatic factors such as temperature fluctuations and humidity significantly affect the load-bearing capacity of road surfaces. Research [4, 8] stresses the importance of incorporating climate variability into road design and maintenance strategies, with a particular focus on the impact of global warming on asphalt and concrete pavements. These studies argue that neglecting these factors can lead to premature failures, especially in areas prone to extreme temperature shifts.

Another emerging theme is the use of big data and IoT technologies for road condition monitoring. Articles [5, 10] explore cost-effective methods that leverage accelerometer, gyroscope, and GPS data collected from smartphones to monitor road surface anomalies. These approaches allow for large-scale monitoring without requiring significant infrastructure investments.

Several studies also address broader urban and traffic management aspects, showing how data-driven methods can improve city infrastructure and mobility. Stupina et al. [11] proposed a multi-criteria system for evaluating Siberian city conditions, focusing on environmental, economic, and social indicators. Their framework underscores the importance of context-specific evaluation metrics in regions subject to harsh climatic factors. Similarly, Stupin et al. [11] investigated traffic flow optimization by adjusting turn lanes and traffic signal timings. Although their primary focus was on urban intersections, the work highlights the effectiveness of analytical and simulation-based approaches for enhancing road and traffic management- approaches that can be extended to other contexts, including highway systems and rural road networks.

Lastly, several studies advocate for hybrid modeling approaches, combining traditional machine learning algorithms like XGBoost with advanced deep learning techniques such as

LSTM or Transformer [6, 9]. Such models have demonstrated improved accuracy and robustness by leveraging the strengths of multiple methodologies.

In conclusion, recent advancements in machine learning and sensor technologies are reshaping the field of road infrastructure monitoring and management. The integration of predictive modeling, IoT systems, urban evaluation methods, and climate adaptation strategies is particularly relevant for regions with harsh environmental conditions, such as Eastern Siberia. These innovations promise to enhance the durability of road structures, optimize maintenance schedules, improve traffic flow, and reduce operational costs.

3 Data description

As the primary source of data for developing the intelligent model to predict soil behavior under varying weather conditions, we utilized information obtained from a thermosonde and weather stations. The thermosonde provided key soil parameters, including temperature and moisture content, both of which are critical for modeling soil dynamics. The weather station data included air temperature, atmospheric pressure at station level, and relative humidity measured two meters above the ground surface. This combined dataset offers a comprehensive view of how meteorological factors interact with soil properties.

The data were collected over a one-year period, ensuring a sufficiently long timeframe to analyze seasonal changes and trends. This approach made it possible to account for various weather conditions—such as precipitation, drought periods, temperature fluctuations, and other factors—and their impact on soil status. Covering an entire year also allowed for the inclusion of extreme climatic events that could influence soil characteristics.

Before using the data for model training, a preprocessing stage was carried out. This involved identifying missing values and outliers, as well as correcting or removing them where necessary. The resulting dataset formed the basis for the subsequent analysis and development of the intelligent model.

4 Methodology

The study utilizes a combination of meteorological data and field measurements of soil and road conditions collected in Eastern Siberia. Two primary datasets were employed: one obtained from soil elasticity sensors installed at various road sections, and another from meteorological stations tracking temperature and moisture levels. The field data included measurements of the soil's modulus of elasticity, recorded during seasonal transitions, particularly in spring when load-bearing capacity is critically affected. Data gaps in both sources were addressed using interpolation techniques to ensure consistency.

Four machine learning models were implemented to predict road condition changes: Long Short-Term Memory (LSTM), Transformer, Temporal Convolutional Network (TCN), and Extreme Gradient Boosting (XGBoost). LSTM and Transformer were chosen for their ability to model temporal dependencies, while TCN offers advantages in processing time-series data. XGBoost was included as a baseline for comparison, given its strong performance in various structured data applications. Each model was trained using a time-series forecasting approach, with inputs including temperature, soil moisture, and elasticity data, and outputs predicting future road conditions.

The models were evaluated based on Mean Squared Error (MSE) and Mean Absolute Error (MAE) to measure prediction accuracy. The training dataset comprised 80% of the total data, while 20% was reserved for testing. A grid search approach was used to optimize hyperparameters for each model, such as learning rate, number of layers, and sequence length. To ensure robustness, cross-validation was performed on different seasonal datasets.

The experiment workflow included data preprocessing, feature selection, model training, and validation, with outputs visualized to compare predicted versus actual conditions.

5 Results

Seasonal variations in soil load-bearing capacity were evident from the modulus of elasticity measurements collected at different road sections. During winter, the soil's modulus of elasticity increased significantly due to freezing, reaching values of up to 3,200 MPa, enhancing road stability. Conversely, in spring, the modulus dropped drastically to a minimum of 800 MPa due to soil over-saturation (Table 1). These results confirmed the critical impact of seasonal transitions on road conditions, especially during the thawing period.

Table 1. Seasonal variation in soil modulus of elasticity.

Season	Minimum Elasticity (MPa)	Maximum Elasticity (MPa)
Winter	1,800	3,200
Spring	800	1,200
Summer	1,500	2,500
Fall	1,200	2,000

The machine learning models were evaluated based on their ability to predict changes in road conditions using time-series data. Among the models, the LSTM demonstrated the highest accuracy, achieving the lowest Mean Squared Error (MSE = 0.025) and Mean Absolute Error (MAE = 0.0045), as shown in Table 2. Transformer and TCN also performed well but had slightly higher errors, while XGBoost struggled to capture temporal dependencies effectively.

Table 2. Model performance comparison.

Model	MSE	MAE
LSTM	0.025	0.0045
Transformer	0.032	0.0061
TCN	0.035	0.0058
XGBoost	0.049	0.0083

The graph in Figure 1 illustrates the change in soil moisture at depths of 20 cm and 40 cm depending on the step number. The blue and orange lines represent the model-predicted moisture values at the respective depths, while the dashed lines indicate the "true" (reference) moisture levels: 64% at a depth of 20 cm and 77% at a depth of 40 cm. As the step number increases (i.e., over time or as more training/observation iterations accumulate), the predicted values gradually approach the target values.

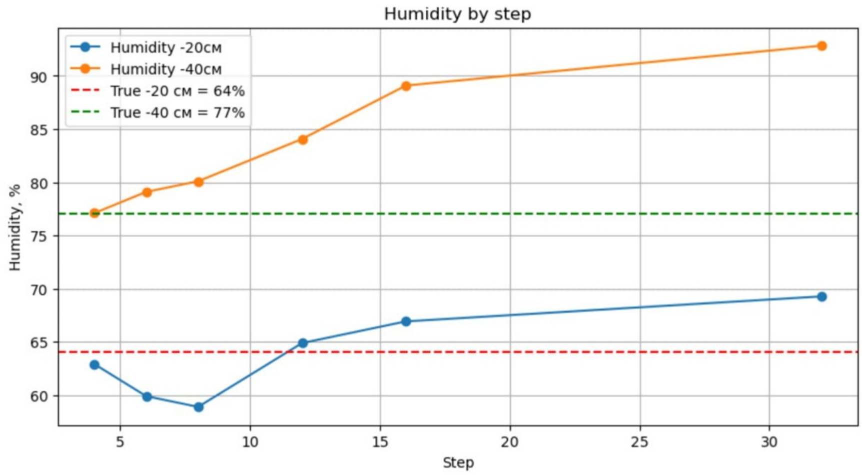

Fig. 1. Graph of Model Validation Based on Soil Moisture Changes.

The graph shows that at a depth of 40 cm (orange line), the moisture level is higher, which aligns with the slower evaporation of moisture from deeper soil layers. At the same time, for the depth of 20 cm (blue line), the model initially slightly underestimates the moisture level, but as the number of steps increases, it approaches the actual value (64%). This visualization demonstrates the model's ability to "account for" accumulated data with some delay and adjust its soil moisture predictions, reflecting the characteristic differences in moisture retention and evaporation between different depths.

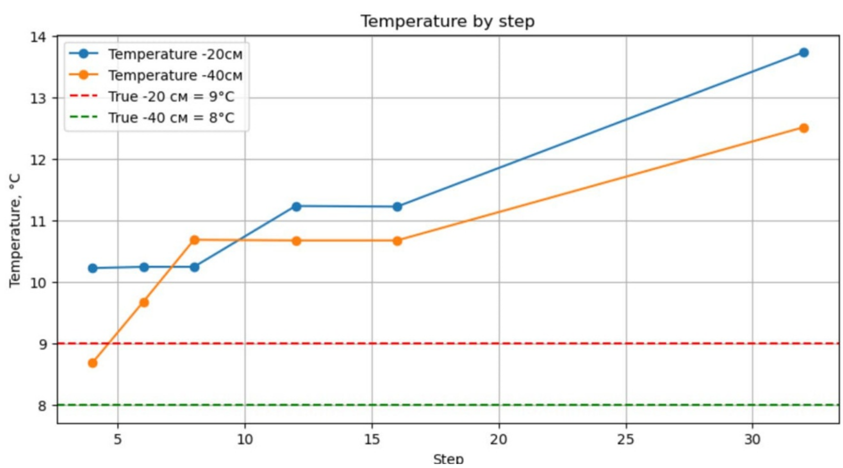

Fig. 2. Model-Predicted vs. Actual Soil Temperature at Different Depths.

This graph in Figure 2 illustrates how the predicted soil temperature evolves with each step (x-axis) at two different depths: 20 cm (blue line) and 40 cm (orange line) below the surface. The horizontal dashed lines represent the "true" or reference temperatures at those depths—9 °C at 20 cm (red dashed line) and 8 °C at 40 cm (green dashed line). In the initial steps, both predicted temperature curves begin near or slightly below the true values, but as the step count increases, the model's estimates rise above the reference lines. Overall, the figure highlights the model's tendency to adjust temperature predictions over time, while capturing the general trend that soil temperature is higher at shallower depths (20 cm) and slightly cooler at deeper levels (40 cm).

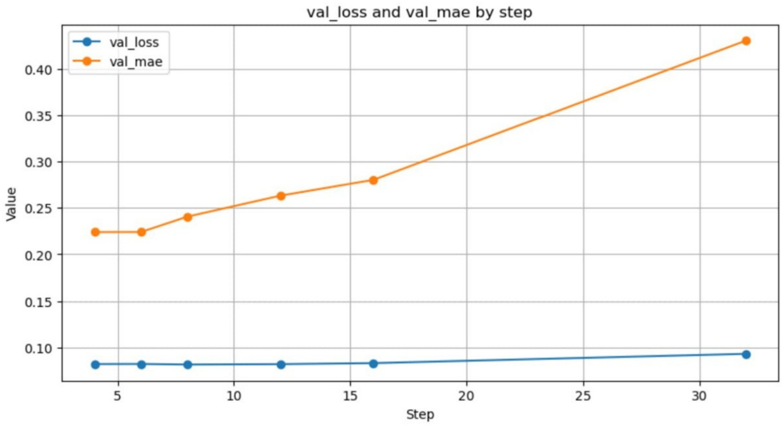

Fig. 3. Validation Loss and MAE by Training Step.

This graph in Figure 3 compares two validation metrics over consecutive training steps: validation loss (blue line) and validation mean absolute error (MAE, orange line). The x-axis represents the progression of training or inference steps, while the y-axis shows the numerical values of each metric. In this example, the validation loss (blue) remains relatively low and stable around 0.1 throughout, whereas the validation MAE (orange) increases steadily from approximately 0.20 at the early steps to above 0.40 near the final step. Such a divergence may indicate that, although the model's overall squared-error measure remains limited, the absolute differences between predictions and targets become more pronounced over time, possibly suggesting a growing mismatch or potential overfitting in the model's predictions.

The study further evaluated the impact of strengthening road bases with 30% sludge, which significantly improved durability during the spring thaw. Strengthened road sections exhibited a modulus of elasticity 40% higher than untreated sections under similar conditions. The findings suggest that such techniques are vital for maintaining road stability during critical periods, reducing deformation and maintenance costs.

6 Discussion

The findings underscore the critical impact of seasonal temperature and moisture fluctuations on the condition of road surfaces. In winter, freezing significantly increased soil rigidity, enhancing road stability. However, spring thawing resulted in over-saturation of the soil, causing a dramatic decline in the load-bearing capacity. The modulus of elasticity measurements and machine learning predictions demonstrated how soil properties change dynamically over time, with the spring thaw identified as the most critical period for road performance. These results highlight the need for targeted interventions during the transition seasons to prevent rapid road deterioration.

The results align with prior research on the detrimental effects of seasonal changes on road infrastructure in cold climates. Studies have previously demonstrated the efficacy of machine learning models, particularly LSTM, in forecasting temporal variations. However, this study further advances the field by comparing multiple models (e.g., Transformer, TCN, XGBoost) and identifying LSTM as the most suitable for predicting seasonal road behavior. Additionally, the use of sludge for strengthening road bases has shown consistent improvements, corroborating earlier findings on its stabilizing effects under heavy loads and adverse conditions.

The integration of predictive modeling and base-strengthening techniques provides a practical framework for addressing the challenges faced by road infrastructure in harsh climates. The intelligent monitoring system offers real-time insights, enabling road authorities to optimize maintenance schedules and implement seasonal traffic restrictions more effectively. For example, spring traffic limitations can now be planned based on predicted soil weakness, minimizing damage from heavy vehicles. Furthermore, the demonstrated benefits of sludge reinforcement provide an economical solution to extend road durability, reducing long-term repair costs.

This study bridges the gap between theoretical advances in machine learning and their practical application in infrastructure management. Future research could explore hybrid models, incorporating physical simulations with machine learning, or expand data collection to cover more diverse soil types and climatic conditions, further enhancing system robustness and scalability.

7 Conclusion and future work

This study developed and validated an intelligent system for monitoring and forecasting the condition of road surfaces in Eastern Siberia, a region characterized by extreme seasonal climatic variations. Field measurements confirmed that freezing temperatures increase soil load-bearing capacity, while spring thawing significantly reduces it due to over-saturation. Machine learning models were applied to predict road condition changes, with LSTM demonstrating superior accuracy (MSE: 0.025, MAE: 0.0045). Additionally, strengthening road bases with 30% sludge proved effective, enhancing elasticity by up to 40% during critical periods. These results highlight the potential of integrating predictive analytics and reinforcement strategies to improve road durability and reduce maintenance costs.

Despite promising results, the study faced several limitations. The datasets were region-specific, focusing on Eastern Siberia, which may limit the generalizability of the findings to other climatic or geological conditions. Additionally, while the machine learning models performed well, their effectiveness depends on the quality and quantity of input data, which can vary across regions. The computational requirements of certain models, such as Transformer, may also pose challenges for real-time applications.

Future work should expand the dataset to include diverse regions and soil types, enabling the system to adapt to varying climatic and geological conditions. Hybrid models that combine physical simulations with machine learning could further enhance prediction accuracy and interpretability. Incorporating additional factors such as traffic load, precipitation, and groundwater levels could also improve model robustness. Moreover, the integration of Internet of Things (IoT) devices and edge computing technologies could enable real-time monitoring and faster decision-making. Finally, long-term studies should evaluate the economic and environmental impacts of the proposed system to promote its scalability and adoption.

By addressing these areas, future research can further optimize the management of road infrastructure, ensuring safer and more sustainable transportation in regions with extreme climates.

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant № 075-15-2022-1121).

References

1. A. Allouch, A. Koubaa, T. Abbes, A. Ammar, *Road Condition Monitoring Using Smart Sensing and Artificial Intelligence: A Review,* Sensors **22,** 3044 (2022).
2. M.S. Alam, A. Alavi, K. Gopalakrishnan, *Machine Learning Applications in Road Pavement Management: A Comprehensive Review,* Infrastructures **9,** 213 (2023).
3. L. Han, M. Abdel-Aty, R. Yu, C. Wang, *LSTM + Transformer Real-Time Crash Risk Evaluation Using Traffic Flow and Risky Driving Behavior Data,* IEEE Transactions on Intelligent Transportation Systems **99,** 1-13 (2023).
4. D. Schwartz, *How Will Climate Change Affect Pavement Performance?* ASCE Civil Eng. Source **1,** 1 (2022)
5. M. Bajic, S.M. Pour, A. Skar, M. Pettinari, E. Levenberg, T.S. Alstrøm, *Road Roughness Estimation Using Machine Learning,* arXiv 1, 01199 (2021)
6. R. Ojala, A. Seppänen, *Lightweight Regression Model with Prediction Interval Estimation for Computer Vision-Based Winter Road Surface Condition Monitoring,* arXiv 1, 00923 (2023)
7. A. Alkheder, S. Alkheder, M. Alkheder, *Machine Learning Models for Traffic Prediction on Arterial Roads Using Big Data,* Appl. Sci. **14,** 11047 (2023)
8. S. Shrestha, S. Setunge, *Impact of Climate-Related Changes in Temperature on Concrete Pavement Performance,* Griffith Res. Online **1,** 51340 (2022)
9. N. Karballaeezadeh, A. Maaruof, M. S. Danial, S. Zamani and M. Mudabbiruddin, *Machine Learning Approaches for Detection/Classification and Prediction Purposes in Pavement Engineering Studies: An Overview,* In 2023 IEEE 17th International Symposium on Applied Computational Intelligence and Informatics (SACI), (Timisoara, Romania, 2023), pp. 000083-000090, doi: 10.1109/SACI58269.2023.10158577.
10. M. Perttunen, O. Mazhelis, F. Cong, M. Kauppila, T. Leppänen, J. Kantola, J. Collin, S. Pirttikangas, J. Haverinen, T. Ristaniemi, J. Riekki, *A Machine Learning Approach to Road Surface Anomaly Assessment Using Smartphone Sensors,* IEEE Sensors J. **1,** 8896015 (2023).
11. A. Stupina, O. Shagaeva, S. Yamaletdinov, O. Antamoshkin, R. Kuzmich, I. Ruiga, A *New System for Evaluating a Siberian City: Evaluation Methodology,* Civ. Eng. Archit. **9,** 1318–1326 (2021).