https://www.sworldjournal.com/index.php/swj/article/view/swj33-02-138
DOI: 10.30888/2663-5712.2025-33-02-138

UDC 625.7/.8: 004.8

PREDICTING FROST HEAVE IN ROAD EMBANKMENTS: A MACHINE LEARNING APPROACH INTEGRATING SOIL COMPOSITION AND WEATHER

Balashova Yu.B.
c.t.s., as.prof.
ORCID: 0000-0002-2286-9263
Balashov A.O.
student
ORCID ID: 0009-0007-5833-0888
Ukrainian State University of Science and Technologies ESI «Prydniprovska State Academy of Civil Engineering and Architecture», 24-a, Architect Oleh Petrov St., Dnipro, 49005

Abstract. *Frost heave in road embankments is a critical issue in cold-region geotechnical engineering, leading to surface distresses, differential settlements, and maintenance challenges. As climate patterns evolve, accurate frost heave prediction becomes increasingly essential for designing resilient pavements. This study presents an advanced Support Vector Machine (SVM) framework to predict frost heave based on comprehensive soil parameters (grain size distribution, Atterberg limits, moisture content, organic matter) and meteorological variables (air temperature, precipitation, freezing index). A multi-year dataset was compiled from laboratory tests and field instrumentation at selected cold-region sites, capturing variations in soil types and climatic conditions. The SVM model, utilizing a Radial Basis Function kernel, was optimized through grid search and ten-fold cross-validation to balance complexity and generalization. Performance metrics, including the coefficient of determination (R^2), mean absolute error (MAE), and mean squared error (MSE), demonstrated a significant improvement over conventional multiple linear regression (MLR) models. The findings underscore the advantages of leveraging machine learning to capture the non-linear interplay of soil and environmental factors, ultimately guiding the design of more durable road embankments in regions prone to frost heave.*

Key words: *frost heave, Support Vector Machines (SVM), cold regions, soil composition, weather data, road embankments, Machine Learning (ML)*

Introduction.

Frost heave is a geotechnical process where soils expand upon freezing due to the formation of ice lenses. It is particularly problematic for pavements in cold-climate regions, as it can induce longitudinal and transverse cracks, faulting at joints, and significant roughness on road pavements [1]. The critical factors controlling frost heave include soil grain size distribution, mineralogy, water availability, and thermal gradients [2].

With shifting climate patterns, freeze-thaw cycles in many regions are becoming more frequent or irregular, posing further challenges to pavement engineers [3]. Traditional empirical or semi-empirical methods, while offering initial estimates, often lack the flexibility to adapt to site-specific complexities or variable climatic conditions [4-6]. Moreover, designing embankments resistant to frost heave often demands a holistic integration of soil properties, in-situ measurements, and weather predictions.

Recent advancements in computational modeling have paved the way for machine learning approaches to tackle geotechnical problems [7]. Among these, Support Vector Machines (SVMs) have gained attention for their ability to handle non-linear relationships and avoid overfitting through robust regularization [8]. The objective of this research is to develop a framework for using SVM-based modeling to predict frost heave in road embankments. Specifically, we explore how advanced SVM configurations can integrate detailed soil characterization and meteorological parameters to yield accurate, site-specific frost heave predictions.

The remainder of this paper is organized as follows. Section 1 provides an expanded overview of the fundamental mechanisms of frost heave and relevant predictive methods. Section 2 details data acquisition, preprocessing steps, and the theoretical foundations of SVM regression. Section 3 presents and discusses the results of hyperparameter tuning, model performance, and sensitivity analyses. Section 4 offers a deeper interpretation of the findings and contextualizes them within existing literature. Finally, Section 5 concludes the study, offering implications for future research and practical engineering applications.

Main text.

1.Background on Frost Heave and Predictive Approaches

Frost Heave Mechanisms. Frost heave occurs when subfreezing temperatures induce ice crystal growth within the pore structure of soils. Water is drawn from unfrozen zones through capillary and cryogenic suction forces, leading to the progressive development of ice lenses. Fine-grained soils, especially silts with high capillarity, are particularly susceptible [1-6]. The amount of heave depends on: (1) soil texture and structure, (2) the rate of freezing, (3) thermal gradient depth, and (4) soil-water interaction, including water table level [2].

Traditional and Emerging Predictive Methods:

1) Empirical and Semi-Empirical Methods. Early methods relied on index

parameters such as frost-susceptibility classifications based on grain size distribution and plasticity indices. Empirical correlations often link measured frost heave to a freezing index or seasonal freezing depth. While straightforward, these methods may be unreliable when site conditions or climate variables deviate from the original calibration datasets [9, 10].

2) Mechanistic Modeling. More rigorous approaches incorporate heat and mass transfer models to simulate ice lens formation [2]. These models capture thermal gradients and water migration but require extensive soil-specific parameters (e.g., thermal conductivity, hydraulic conductivity, unfrozen water content) and detailed climatic data. Calibration can be complex, and errors in parameter estimation propagate through the model.

3) Machine Learning. Machine learning methods, including Artificial Neural Networks, Random Forests, and SVMs, bypass many limitations of purely physics-based or empirical models by learning patterns directly from data. SVMs, in particular, excel in handling non-linear relationships through kernel functions and have demonstrated robustness in various geotechnical applications such as slope stability and soil classification [7]. However, successful implementation requires careful data preprocessing, parameter tuning, and validation to ensure that the model generalizes well beyond the training dataset.

2.Methodology

Data Collection and Preprocessing. A multi-year, multi-site dataset was collated from ongoing geotechnical monitoring projects in cold regions. Comprehensive laboratory and field investigations were performed on representative soil samples used in road embankment construction. Table 1 summarizes the geotechnical parameters for each soil sample.

Meteorological data were gathered from local weather stations, recording average monthly temperature, total precipitation, and a calculated freezing index (accumulated degree-days below 0◦C).

Frost heave displacements were continuously recorded via extensometers and settlement plates installed at varying depths in the embankment sections. The measured heave data were averaged to single values (in mm) per winter season for each tested location. Where sensor readings appeared inconsistent (e.g., calibration errors), they were filtered through cross-checks with nearby gauges or omitted if irreparable.

Table 1 - Representative Soil Composition Data

Sample ID	Grain Size Distribution (%)			OM Content (%)
	Sand	Silt	Clay	
S1	34.2	54.8	11.0	2.58
S2	41.6	41.1	17.3	3.12
S3	22.7	58.3	19.0	3.94
S4	48.5	29.4	22.1	2.09
S5	26.3	59.8	13.9	4.87
S6	14.8	66.7	18.5	5.96
S7	38.9	44.1	17.0	3.41
S8	53.2	31.6	15.2	2.16
S9	19.4	64.3	16.3	3.84
S10	31.5	57.2	11.3	3.62

A total of 10 soil types × 5 winter seasons × 4 replicate measurements were compiled. Each entry combined soil parameters, meteorological factors, and the measured frost heave. Numerical variables were standardized to zero mean and unit variance to facilitate model training [7]. The dataset was then randomly split into training (80%) and testing (20%) subsets, ensuring each soil category and winter season was represented proportionally.

Support Vector Machine (SVM) Formulation. Support Vector Regression (SVR) with the Radial Basis Function (RBF) kernel was used to model frost heave *(y)* as a function of input features $x \in R^d$ (soil and meteorological parameters). In the primal form, the objective is to minimize:

$$\frac{min}{w,b,\xi_i,\xi_i^*} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{1}$$

$$subject\ to \quad \begin{cases} y_i - w \cdot \phi(x_i) - b \le \varepsilon + \xi_i \\ w \cdot \phi(x_i) + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{2}$$

where C is the penalty parameter, ε is the epsilon-tube defining the margin of tolerance, and ξ_i, ξ_i^* represent positive and negative errors, respectively. The non-linear mapping $\phi(x)$ is implicitly handled through the RBF kernel:

$$K(x, x') = \exp(-\gamma\|x - x'\|^2), \tag{3}$$

where γ is the kernel parameter controlling the width of the Gaussian function.

Optimal hyperparameters (C, γ, ε) were selected via a grid search with ten-fold cross-validation, maximizing the R^2 on the validation folds [11]. This process systematically varied each parameter over a predefined range to identify the best combination.

Model Evaluation Metrics. To robustly assess predictive accuracy, three standard metrics were employed:

Coefficient of Determination (R^2):

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{4}$$

where y_i is the observed frost heave, $\hat{y}_i$ is the predicted frost heave, and $\bar{y}$ is the mean of observed frost heave.

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

providing an intuitive measure of average absolute deviation from observed values.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}|(y_i - \hat{y}_i)^2| \tag{6}$$

emphasizing larger errors due to the square term.

These metrics collectively capture both the variance explained by the model and the magnitude of predictive error [7].

3. Results

Hyperparameter Tuning. The grid search explored $C \in \{1, 10, 100, 1000\}$, $\gamma \in \{0.001, 0.01, 0.1, 1\}$, and $\varepsilon \in \{0.001, 0.01, 0.1\}$. The final choice was $C = 100$, $\gamma = 0.01$, and $\varepsilon = 0.01$ based on a balance of high R^2 and low MAE across cross-validation folds. Table 2 highlights a portion of the hyperparameter search results.

Although $C = 1000$ offered a marginally higher R^2, it showed signs of overfitting (larger variance in fold-to-fold error). Thus, $C = 100$ was chosen to maintain generalizability.

Table 2 - Selected Hyperparameter Tuning Results (Cross-Validation)

Combination	C	γ	ε	CV R^2	CV MAE (mm)
1	10	0.1	0.01	0.85	1.96
2	100	0.01	0.01	0.90	1.57
3	100	0.01	0.1	0.88	1.72
4	1000	0.01	0.01	0.91	1.51

Model Performance on Test Set. Table 3 contrasts the final SVM model with a benchmark multiple linear regression (MLR) approach, tested on the held-out 20% data.

Table 3 - Test Set Performance Comparison

Model	R^2	MAE (mm)	MSE (mm^2)
SVM (RBF)	0.905	1.43	3.19
Multiple Linear Regression	0.761	2.17	6.08

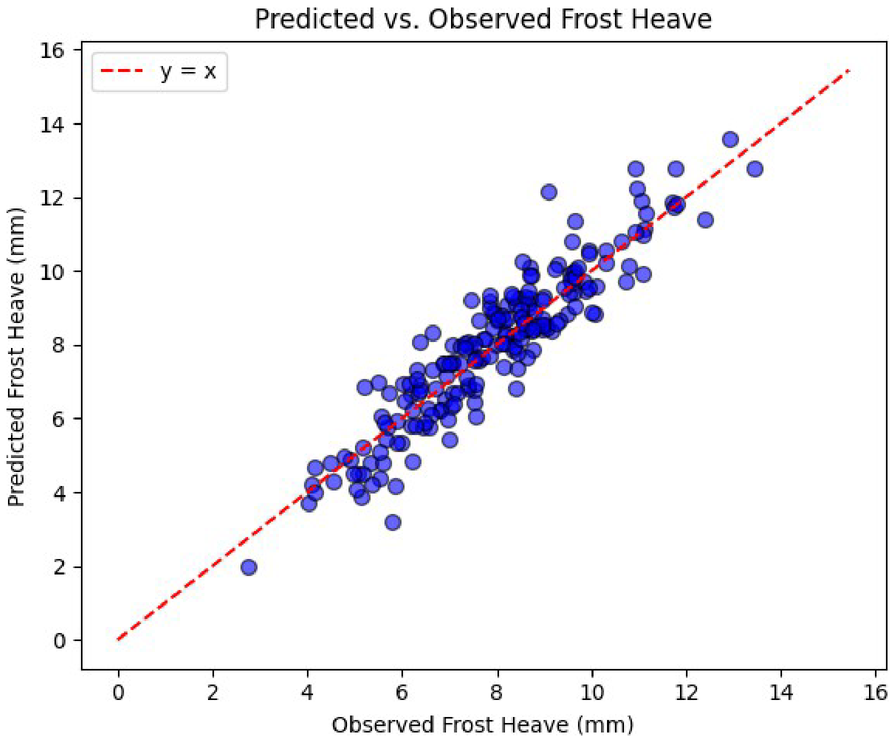

Figure 1 - Predicted vs. Observed Frost Heave for the Test Set (SVM)

The SVM showed a substantial improvement over MLR in terms of variance explained (R^2 increased by ≈ 0.14) and error metrics (MAE and MSE reduced by notable margins). Figure 1 depicts a scatter plot of observed vs. predicted frost heave for the SVM model, illustrating a close alignment with the $y = x$ reference line.

Sensitivity Analysis. A sensitivity analysis was performed to quantify the influence of each predictor on frost heave. Each input was systematically varied across its observed range while holding all other inputs at their mean values. Figure 2 shows the relative importance scores, reflecting how changes in a single variable affect the model's output.

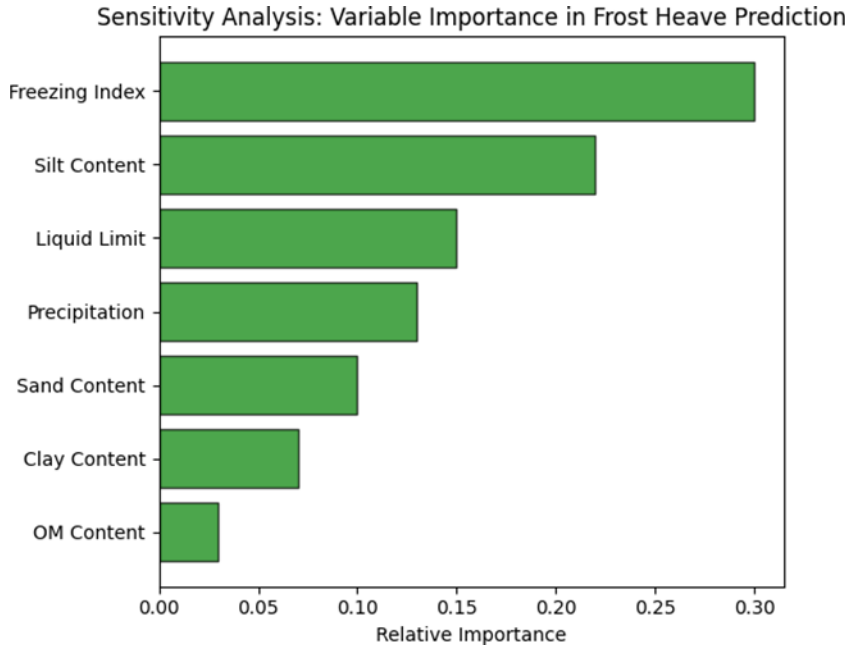

Figure 2 - Sensitivity Analysis: Relative Importance of Input Variables in Frost Heave Prediction

The freezing index exhibited the highest impact, highlighting the critical role of prolonged subfreezing conditions in driving frost heave. Silt content, liquid limit, and precipitation also showed notable effects, consistent with the well-known influence of soil capillarity and moisture on ice lens formation [2].

4.Discussion

Interpretation of SVM Performance. The SVM's superior performance over MLR underscores the non-linear interplay among soil properties (e.g., silt content, plasticity) and climatic factors (e.g., cumulative degree-days of freezing). Linear approaches struggle to capture this complexity, especially where threshold or multiplicative effects emerge (e.g., once soils exceed certain moisture or plasticity limits, frost heave may escalate disproportionately) [8].

Furthermore, the robust regularization inherent in SVM helps prevent overfitting, leading to stable generalization across varied soil-climate combinations. The grid search and cross-validation procedure were essential in identifying an optimal parameter set that balanced accuracy with model complexity.

Practical Implications for Road Embankment Design. Material Selection and Layer Configuration. By linking geotechnical and climatic factors to expected heave magnitudes, practitioners can screen prospective borrow materials more effectively. For instance, soils with lower silt fractions or stabilized soils with reduced plasticity indices can be strategically used in upper layers.

Climatic Adaptation. As freezing index and precipitation patterns shift, the predictive model allows engineers to anticipate potential increases in heave severity. This proactive approach supports the incorporation of additional protective measures such as insulation layers, drainage improvements, or specialized additives in susceptible areas.

Life-Cycle Cost Analysis. Reliable frost heave forecasts feed into life-cycle cost models, allowing agencies to estimate the economic benefits of investing in robust designs or timely maintenance. By quantifying future risks, the decision-making process is more data-driven, ultimately reducing the total cost of ownership for roadway assets.

Limitations and Future Directions. Despite promising outcomes, several limitations must be acknowledged. First, the accuracy of frost heave measurements depends on the precision and calibration of in-situ sensors. Second, while the freezing index is a valuable proxy, it may not fully capture transient freeze-thaw cycles or soil moisture changes in real time. Incorporating advanced thermal-hydraulic modeling or higher-resolution climate data could further refine predictions.

Future research could explore:

• Ensemble Learning: Methods like Gradient Boosted Trees or Random Forests could be compared against SVM to investigate gains in predictive performance.

• Physics-Informed Machine Learning: Hybrid approaches that embed physical constraints within the model architecture, ensuring consistency with established frost heave theories.

• Geospatial Generalization: Large-scale studies spanning multiple geographic regions would validate model scalability, capturing broader variations in soil mineralogy, drainage conditions, and weather extremes.

5.Conclusion and Future Work

This study demonstrates a robust machine learning methodology for predicting frost heave in road embankments by integrating detailed soil characterization and meteorological variables. Key findings include:

• Enhanced Accuracy: The best SVM model achieved an R^2 of 0.905 on the test set, notably outperforming multiple linear regression.

• Dominant Influence of Freezing Index: Sensitivity analysis identified cumulative cold exposure as a primary driver of frost heave, with silt content and soil plasticity also playing significant roles.

• Design and Maintenance Applications: The model's predictive capability aids in material selection, layer configuration, and proactive maintenance strategies, potentially lowering life-cycle costs.

Future work should incorporate ensemble machine learning methods, higher-resolution climate inputs, and expanded datasets from diverse geoclimatic regions. Integrating mechanistic insights with data-driven techniques holds promise for refining frost heave models, ultimately guiding more resilient infrastructure in cold environments.

References:

1.Yuliia Balashova, Viktor Demianenko, Petro Sankov, Vladislav Lukianenko and Khadija Youb / New construction solutions and materials for panels of road pavements / Innovative Technologies in Construction, Civil Engineering and Architecture. AIP Conference Proceedings 2678, 020001 (2023); 15 February, 2023. – p. 1-7. URL: https://doi.org/10.1063/5.0118620 / https://aip.scitation.org/doi/pdf/10.1063/5.0118620

2. Yajun Wu, Encheng Zhai, Xudong Zhang, Gang Wang, Yitian Lu / A study on

frost heave and thaw settlement of soil subjected to cyclic freeze-thaw conditions based on hydro-thermal-mechanical coupling analysis / Cold Regions Science and Technology / Volume 188, August 2021, 103296
URL: https://doi.org/10.1016/j.coldregions.2021.103296

3. Yuliia Balashova, Viktor Demianenko, Nataliia Tkach, and Hennadii Karasev / Ensuring the sustainability of the roadbeds in the zones of the underground mine works / Scopus, ISSN: 25550403, DOI: 10.1051/e3sconf/201912301041. Volume 123. – EDP Sciences, 2019. - p. 01041. - 13c. URL: https://www.scopus.com/authid/detail.uri?authorId=57211522441&eid=2-s2.0-85074287353.

4. Balashova Yu.B. Enhancing the bearing capacity of soft clay subgrades of roads by using lime and xanthan gum / Monographic series «European Science» Book INTELLECTUAL AND TECHNOLOGICAL POTENTIAL OF THE XXI CENTURY: Karlsruhe, Germany. Book 33. Part 1. CHAPTER 6, 2024, P. 96-108.
URL: https://desymp.promonograph.org/index.php/sge/issue/view/sge33-01/sge33-01
https://doi.org/10.30890/2709-2313.2024-33-00-007

5. Ю.Б. Балашова, В.В. Дем'яненко, О.Ю. Усиченко, П.А. Тиквенко, А.О. Балашов / Підвищення несучої здатності слабких водонасичених глинистих основ земляного полотна / Науково-практичний журнал «Український журнал будівництва та архітектури», № 4 (016), 2023, Дніпро: ПДАБА, С. 7-14. URL: http://uajcea.pgasa.dp.ua/issue/view/17138;
DOI:10.30838/J.BPSACEA.2312.290823.7.965
URL: http://uajcea.pgasa.dp.ua/article/view/288626/282288'

6. Yu.B. Balashova, V.V. Demianenko, A.O. Balashov / Enhancing the stability of soft soil foundations in transportation infrastructure through the application of lime and xanthan gum / SWorld-Ger Conference Proceedings, 1/ No. gec34-00 (2024): Future in the results of modern scientific research '2024 / Karlsruhe, Germany: p. 37–40. ISBN 978-3-98924-057-5. DOI: https://doi.org/10.30890/2709-1783.2024-34-00
URL: https://www.proconference.org/index.php/gec/issue/view/gec34-00
URL: https://doi.org/10.30890/2709-1783.2024-34-00-010 (gec34-01, p. 37)

7. Nitish Puri, Harsh Deep Prasad, Ashwani Jain / Prediction of Geotechnical Parameters Using Machine Learning Techniques / Procedia Computer Science: Volume 125, 2018, Pages 509-517 /URL: https://doi.org/10.1016/j.procs.2017.12.066

8. Chih-Chung Chang, Chih-Jen Lin / LIBSVM: A library for support vector machines / ACM Transactions on Intelligent Systems and Technology (TIST), Volume 2, Issue 3 Article No.: 27, Pages 1 – 27 /
URL: https://doi.org/10.1145/1961189.1961199

9. Andrii Balashov / Attention-integrated convolutional neural networks for enhanced image classification: a comprehensive theoretical and empirical analysis / International periodic scientific journal "Modern engineering and innovative technologies" ISSN 2567-5273, Issue №35, Part 2, October 2024, Karlsruhe, Germany, p. 18-27. / URL: https://doi.org/10.30890/2567-5273.2024-35-00-030
http:// www.moderntechno.de/index.php/meit/article/view/meit35-00-030
https://www.moderntechno.de/index.php/meit/issue/view/meit35-02/meit35-02

10. Andrii Balashov, Olena Ponomarova, Xiaohua Zhai / Multi-Stage Prompt Inference Attacks on Enterprise LLM Systems / https://arxiv.org/pdf/2507.15613
URL: https://doi.org/10.48550/arXiv.2507.15613

11.Wengang Zhang, Hongrui Li, Liang Han, Longlong Chen, Lin Wang / Slope stability prediction using ensemble learning techniques: A case study in Yunyang County, Chongqing, China / Journal of Rock Mechanics and Geotechnical Engineering / Volume 14, Issue 4, August 2022, Pages 1089-1099
URL: https://doi.org/10.1016/j.jrmge.2021.12.011

sent: 24.09.2025