

Interpretable Empirical Modelling of Subsurface Pavement Temperature for an Instrumented Test Section in Edmonton, Alberta Using Gene Expression and Multiple Expression Programming

Malik Noor Ul Amin Awan

PhD Student, Department of Civil and Environmental Engineering,
University of Alberta, Edmonton, Alberta, Canada

mnawan@ualberta.ca

Shristi Adhakari

MSc Student, Department of Civil and Environmental Engineering,
University of Alberta, Edmonton, Alberta, Canada

shristi3@ualberta.ca

Leila Hashemian, PhD, PEng (corresponding author)

Professor, Department of Civil and Environmental Engineering,
University of Alberta, Edmonton, Alberta, Canada.

hashemia@ualberta.ca

Paper prepared for presentation at the PV - A Changing Climate: Pavement Resilience in the Era of a Changing Global Landscape session of the 2026 Transportation Association of Canada (TAC) Conference & Exhibition
Winnipeg, Manitoba

Acknowledgements

The authors would like to thank the Artificial Intelligence for Logistics (AI4L) program of the National Research Council (NRC) Canada for providing funding for this research.

Abstract

Accurate prediction of subsurface pavement temperature is essential for understanding thermal behaviour and improving pavement performance, particularly in cold regions. This study develops interpretable empirical models to predict pavement temperature at multiple depths using Gene Expression Programming (GEP) and Multiple Expression Programming (MEP). Air temperature, incoming solar radiation, and sine and cosine transformations of the day of year were evaluated as candidate predictors. Solar radiation was not retained in any of the final equations; consequently, the reported formulations depend on air temperature and seasonal harmonic terms.

Model performance was evaluated at depths of 0.25 m, 0.70 m, and 2.25 m within the pavement structure. GEP achieved test R^2 values of approximately 0.83, 0.84, and 0.92 at the three depths, with corresponding RMSE values of 5.21°C, 5.02°C, and 1.89°C. MEP produced lower test R^2 values of approximately 0.77, 0.76, and 0.79. GEP performance improved with depth, whereas MEP performance was not monotonic between 0.25 m and 0.70 m. The lower RMSE at 2.25 m coincided with a smoother, lower-variability temperature signal.

Overall, GEP outperformed MEP for this one-year, single-site dataset and produced explicit, depth-specific equations. The results represent a site-specific proof of concept. Application to pavement design, seasonal load restrictions, or winter weight programs would require independent multi-year and multi-site validation, including evaluation of prediction errors near the freezing threshold.

Keywords: Pavement temperature; Gene Expression Programming; Multiple Expression Programming; Machine Learning; Granular base course; Subgrade.

1. Introduction

Temperature is a primary factor governing the response and performance of flexible pavement systems, as it directly influences the mechanical properties of hot mix asphalt (HMA), as well as the supporting base and subgrade layers¹. Due to its viscoelastic nature, HMA exhibits temperature-dependent stiffness and load-bearing capacity, resulting in varying structural behaviour under different climatic conditions. These temperature variations are responsible for distinct distress mechanisms: elevated temperatures during summer reduce stiffness and increase susceptibility to rutting, whereas low temperatures in winter induce brittleness and lead to thermal cracking¹. In addition to the asphalt layer, the thermal condition of the base and subgrade plays a crucial role in overall pavement performance. In cold regions such as Alberta, seasonal freeze–thaw cycles, particularly during spring thaw, significantly weaken these layers, reducing structural capacity and accelerating pavement deterioration¹.

Roadway pavements and their surrounding environment form a coupled microclimate system, in which pavement surface and subsurface temperatures are governed by atmospheric conditions and the thermal properties of pavement materials, while also influencing near-surface air temperature². Accurately characterizing this interaction is essential for understanding pavement behaviour and for developing reliable predictive models. Consequently, extensive research has been conducted across various climatic regions to develop pavement temperature prediction models, which are widely used in asphalt binder selection, pavement design, and performance evaluation.

Traditional temperature prediction approaches, including empirical correlations and mechanistic heat transfer models, often require extensive input parameters and calibration efforts, and may not adequately capture site-specific variability. In recent years, data-driven techniques, particularly machine learning models, have been increasingly employed to improve prediction accuracy. However, many of these approaches operate as “black-box” models, limiting their interpretability and practical application in engineering analysis and decision-making^{3,4}.

Mechanistic models, such as Pavement ME and the Enhanced Integrated Climatic Model (EICM), are physically based approaches that support depth-wise temperature prediction; however, they are often complex and require extensive environmental and material input parameters⁵. Empirical models are simple and practical for routine applications, but they are often site-specific and exhibit limited accuracy, particularly for deeper pavement layers. The most widely used of these is the BELLS model, named after its developers Baltzer, Ertman-Larsen, Lukanen and Stubstad, which was developed from Long-Term Pavement Performance (LTPP) Seasonal Monitoring Program data and estimates the temperature of an asphalt layer at a given depth from the infrared surface temperature measured at the time of testing, the previous day’s mean air temperature and the time of day. Because BELLS requires a measured surface temperature and was calibrated only for bound asphalt layers, it cannot be applied to the granular base and subgrade depths considered in this study⁶. Machine learning models, such as Random Forest, XGBoost, and CatBoost, offer high predictive accuracy and are capable of capturing nonlinear relationships, although they lack physical interpretability⁷. Deep learning models, including Artificial Neural Network, Long Short-Term Memory, and deep neural networks, demonstrate strong performance in modelling complex temporal patterns, but they are data-intensive and require fine tuning⁸. More recent studies on the same instrumented section and in comparable cold-region settings have applied machine-learning methods to predict temperature and moisture in pavement base and subgrade layers and to determine seasonal load-restriction and winter-weight-premium periods, reporting high accuracy across depths^{20,21,22}.

Recent work has extended this trend in three directions relevant to the present study. First, ensemble and deep-learning models have continued to improve raw accuracy: Kebede et al.²³ reported that a gradient-

boosted ensemble predicted pavement temperature at multiple depths with an R^2 of 0.98 and reduced error by more than 35% relative to conventional learners, with SHAP analysis attributing over 40% of the model response to air temperature, while Bai et al.²⁴ raised winter road-surface temperature accuracy by roughly 30% by pre-training a long short-term memory network on a long-record station and fine-tuning it at a data-scarce site. Ma et al.²⁵ showed that the same family of models can predict freezing depth to within 0.05–0.20 m and thaw date to within one to three days up to two weeks ahead using air temperature alone, which is directly relevant to load-restriction timing.

Second, and most closely related to this work, evolutionary symbolic regression has begun to be benchmarked against these black-box learners on cold-region instrumented roads: Oyeyi et al.²⁶ and Huyan et al.²⁷ compared genetic and gene expression programming with ensemble methods for layer temperature in Canadian test sections and found that the evolved closed-form expressions were only marginally less accurate than XGBoost ($R^2 > 0.97$ against $R^2 > 0.98$) while remaining fully transparent. Padala et al.²⁸ applied gene expression programming to pavement air-temperature prediction for binder grade selection, using an optimum tree-depth criterion to keep the resulting equations simple enough for hand calculation, and multi expression programming has produced explicit design equations of comparable accuracy in related civil-engineering applications²⁹.

Third, work on cold-region climate adaptation³⁰ has shown that freeze–thaw-linked performance models reproduce seasonal behaviour that current mechanistic-empirical procedures do not, reinforcing the need for temperature models that are both accurate and inspectable as freeze–thaw regimes shift.

To address these limitations, interpretable modelling techniques such as Gene Expression Programming (GEP) and Multiple Expression Programming (MEP) have emerged as effective alternatives, as they can generate explicit mathematical relationships while maintaining high predictive performance. These approaches provide a transparent framework for understanding the influence of key environmental variables on subsurface temperature behaviour⁹.

This study develops an interpretable empirical modelling framework for predicting subsurface pavement temperature using GEP and MEP. Model performance is evaluated using statistical indicators, and the final equations are examined to determine which candidate variables were retained. In this study, incoming solar radiation was available to the evolutionary search but was not retained in the selected equations.

The contribution of this work is the development and comparison of explicit, depth-specific GEP and MEP equations for the monitored Edmonton section. The equations provide a transparent alternative to black-box models, but their present evidence base is limited to one site and one annual cycle. Their possible use in seasonal load restriction, winter weight program planning, pavement design, or infrastructure management should therefore be treated as a future application requiring threshold-specific and external validation.

2. Edmonton Smart Road Test Section

The smart road test section, located between Aurum Road and the Edmonton Waste Management Centre in Edmonton, Alberta, was constructed to evaluate pavement performance under heavy traffic conditions. The pavement structure consists of a 90 mm surface course, a 160 mm binder course, and a 450 mm granular base course supported by the natural subgrade.

Construction began with saw cutting the existing asphalt surface, followed by pulverization using a concrete breaker and removal of materials with a bucket attachment. Throughout the process, safety measures such as traffic cones and warning signage were implemented to minimize disruption and ensure safe operations.

Temperature sensors were installed at nominal depths of 0.25 m, 0.70 m, and 2.25 m to monitor temperature variations within the pavement system. Based on the stated layer thicknesses, the 0.25 m and 0.70 m locations correspond nominally to the asphalt–granular-base and granular-base–subgrade interfaces, respectively, while the 2.25 m location is within the subgrade. At each depth, two sensors were installed in the inner and outer wheel paths to capture spatial variation across the lane. Four boreholes were also drilled using a screw auger to collect samples from the granular base and subgrade. The test-section cross-section and a temperature sensor are shown in Figure 1.

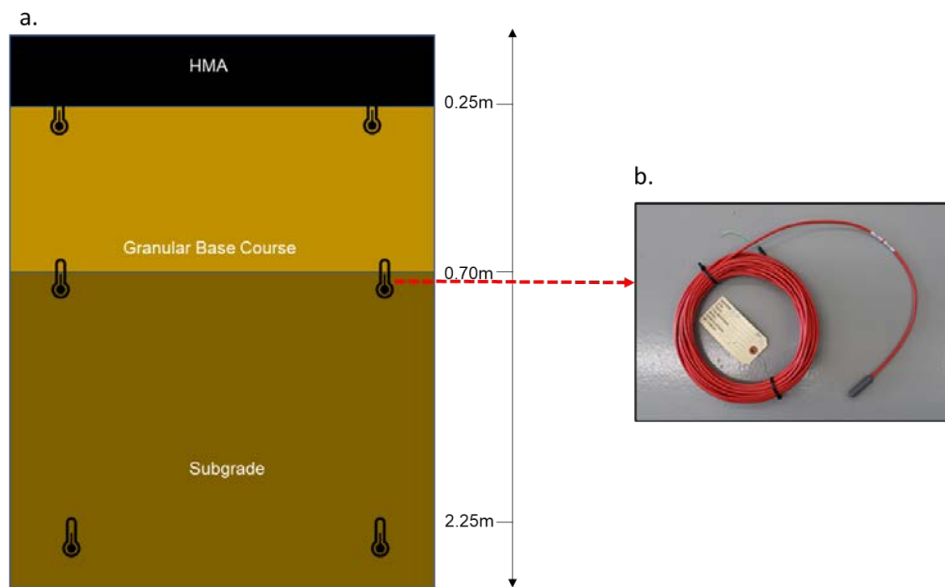


Figure 1. (a) Test-section cross-section; (b) temperature sensor

In the final stage, a data acquisition system was installed on a reinforced concrete platform measuring 150 cm by 150 cm by 75 cm. Sensor cables were routed through conduits, and after the concrete had cured, the data logger cabinet was securely mounted. Ethernet and power connections were established to enable continuous data collection. The system was configured to record temperature readings from all sensors at approximately 15-minute intervals, with the data stored locally and accessible remotely for analysis.

2.2 Material Composition and Properties

The granular base course, composed of crushed aggregate with a maximum particle size of 19 mm, is classified as well-graded gravel under the Unified Soil Classification System. This material was selected

for its high load-bearing capacity and effective drainage properties, which are essential for maintaining pavement stability. Its well-graded structure reduces voids and enhances strength under traffic loading¹⁰.

The subgrade soil is classified as clayey sand, consisting of sand particles mixed with a significant clay fraction. This composition influences its compaction behaviour, permeability, and response to moisture changes. The subgrade is moderately frost-susceptible, with approximately 27% and 21% of particles passing the 0.075 mm and 0.02 mm sieves, respectively. The liquid limit is 25 percent, and the plasticity index is 9 percent, indicating moderate plasticity and sensitivity to moisture variations. These properties affect the subgrade's overall mechanical performance under environmental and loading conditions¹⁰.

The grain-size distributions of both the granular base and subgrade layers are presented in Table 1, highlighting the role of finer particles in the subgrade and the stabilizing contribution of the granular base within the pavement structure.

Table 1 Grain-Size Distribution of Subgrade Soil and Granular Base Course (GBC)¹⁰

Subgrade Soil		GBC	
Grain Size, D (mm)	Percent Finer (%)	Grain Size, D (mm)	Percent Finer (%)
0.250	98.0	19	99.15
0.150	83.9	12.5	75.17
0.075	46.9	9.5	61.23
0.044	37.1	4.75	42.73
0.032	32.5	2.38	32.08
0.023	27.3	1.19	25.61
0.012	24.1	0.595	18.93
0.009	22.4	0.297	7.09
0.004	19.9	0.149	1.46
0.001	15.0	0.075	0.51

3. Objectives

The primary objective of this study is to develop robust and interpretable empirical models that predict subsurface temperature in the granular base course (GBC) and subgrade layers using two evolutionary symbolic-regression techniques: GEP and MEP. Specifically, the study aims to (i) develop temperature-prediction models at three depths (0.25 m, 0.70 m, and 2.25 m); (ii) compare the accuracy of the GEP and MEP formulations using the coefficient of determination (R^2), RMSE, and MAE; and (iii) derive explicit, physically interpretable equations suitable for practical pavement-engineering applications, thereby improving predictive accuracy while reducing model complexity.

4. Machine Learning Models

4.1. Gene Expression Programming (GEP)

GEP is an evolutionary algorithm derived from genetic programming, designed to develop mathematical models through symbolic regression¹¹. The fundamental components of GEP include a function set, terminal set, fitness function, control parameters, and termination criteria. The function and terminal sets define the search space of the model, while the fitness function evaluates solution quality. Control parameters and stopping criteria govern the efficiency and convergence of the search process¹¹.

GEP operates through a population of chromosomes, each representing a potential solution encoded in linear strings known as Karva notation (K-expressions). These chromosomes are subsequently translated into expression trees (ETs), which provide a visual and mathematical representation of the solution. The translation process begins from the first element of the K-expression, forming the root of the expression tree and expanding hierarchically into branches. Conversely, expression trees can be converted back into linear strings by traversing nodes level by level¹².

The algorithm initiates by generating an initial population of candidate solutions and evaluating their fitness. The most suitable chromosomes are then selected and modified using genetic operators such as mutation and recombination to produce new generations. This evolutionary process continues until a predefined termination condition is satisfied. The final solution is represented as a mathematical expression derived from the expression tree, which can be interpreted by traversing the tree from left to right and top to bottom¹⁴. Figure 2 shows the working mechanism of the Gene Expression Programming.

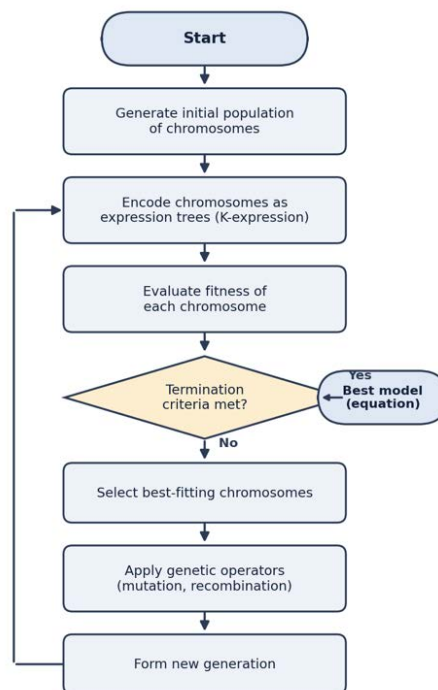


Figure 2. Gene Expression Programming workflow

4.2. Multiple Expression Programming (MEP)

MEP is an advanced evolutionary algorithm that builds upon the concept of GEP. Unlike traditional approaches, MEP uses linear chromosomes that can encode multiple mathematical expressions within a

single structure. This allows each chromosome to represent several potential solutions at once, from which the most suitable one is selected based on its performance¹⁵.

The process begins with the generation of a random population of chromosomes, each representing a possible solution to the problem. From this population, parent chromosomes are selected using a competitive selection method. These parents then undergo genetic operations such as crossover and mutation to produce new offspring¹⁶.

Each newly generated solution is evaluated using a fitness function, which measures how well it predicts the target variable. The best-performing solutions are retained, while weaker ones are gradually replaced through successive iterations. This evolutionary process continues until a stopping criterion is reached. The final outcome is the most optimal mathematical expression that accurately represents the relationship between the input variables and the target response¹⁶. Figure 3 shows the working mechanism of the Multiple expression programming (MEP).

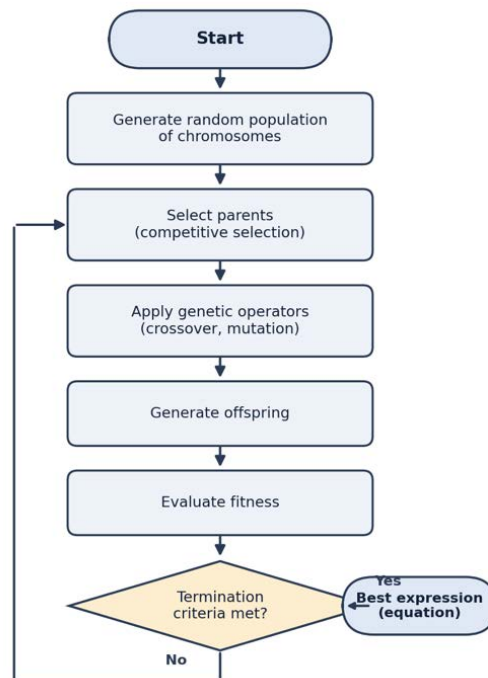


Figure 3. Multiple Expression Programming workflow

The overall research methodology consists of a systematic framework for developing interpretable empirical models to predict subsurface pavement temperature. The process includes data collection, preprocessing, feature analysis, model development, equation formulation, and model validation.

5.1 Data Collection

The study is based on data collected from an instrumented pavement system in Edmonton, Alberta. One year of continuous data, spanning June 2023 to May 2024, was used to represent one annual cycle. Candidate inputs supplied to the modelling process were air temperature, incoming solar radiation, and seasonal indicators represented by sine and cosine transformations of the day of year (DOY). The final equations retained air temperature and the DOY terms but did not retain incoming solar radiation.

5.2 Data Preprocessing

Initial data preprocessing was performed to ensure the dataset's quality and consistency. This step involved data cleaning to address missing values and outliers, followed by normalization of the temperature data. Measurements from both inner and outer wheel paths were incorporated to represent realistic field conditions. The final dataset was structured for subsequent modelling and analysis.

5.3 Identification of Influential Factors

To identify the key variables influencing subsurface temperature, exploration and statistical analyses were conducted. The relationships between input variables and target temperatures were examined through correlation analysis and preliminary functional evaluations. Arithmetic operations were utilized to assess interactions among variables and to guide the selection of significant predictors for model development.

5.4 Model Development

Two interpretable machine learning techniques, namely Gene Expression Programming (GEP) and Multiple Expression Programming (MEP), were employed to develop predictive models. These methods were selected due to their ability to generate explicit mathematical relationships between input variables and target outputs.

The models were trained using the prepared dataset, and hyperparameter tuning was performed to optimize performance. The GEP approach utilizes expression trees to evolve mathematical equations, while MEP provides compact linear representations of predictive relationships.

Each depth was modelled independently. The candidate predictor set comprised air temperature, incoming solar radiation, and the sine and cosine day-of-year terms. None of the models received measured or predicted temperature from another depth. Because solar radiation does not appear in Equations 1–6, the selected formulations ultimately depend only on air temperature and day of year. The intended use is site-specific estimation from available meteorological data; the present analysis does not establish transferability to other sites or years.

5.5 Equation Formulation

The trained models were used to derive empirical equations representing subsurface temperature at different depths. The GEP-based expression trees were converted into explicit mathematical formulations, while MEP provided simplified linear and nonlinear relationships. These equations enable direct application in engineering practice and improve model interpretability.

5.6 Model Evaluation and Validation

A comprehensive evaluation of model performance was conducted using statistical indicators, including the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). Model validation was performed using unseen test data to assess generalization capability. A comparative analysis of GEP and MEP models was also conducted to identify the most accurate and reliable approach. The one-year record (June 2023–May 2024) was partitioned chronologically rather than randomly: the first eight months, from June 2023 to the end of January 2024, were used for model training, and the remaining four months, from February to May 2024, were withheld entirely and used as the unseen testing and validation set. This corresponds to an approximate 70/30 split, and the split point is marked by the vertical dash-dot line in Figures 5 to 7. Validation was therefore performed on a held-out portion of the same one-year record and not on a separate year of data. A chronological split was preferred to a random one because randomly shuffling a 15-minute time series places highly autocorrelated records

from the same day in both subsets.

Figure 4 summarizes the methodology.

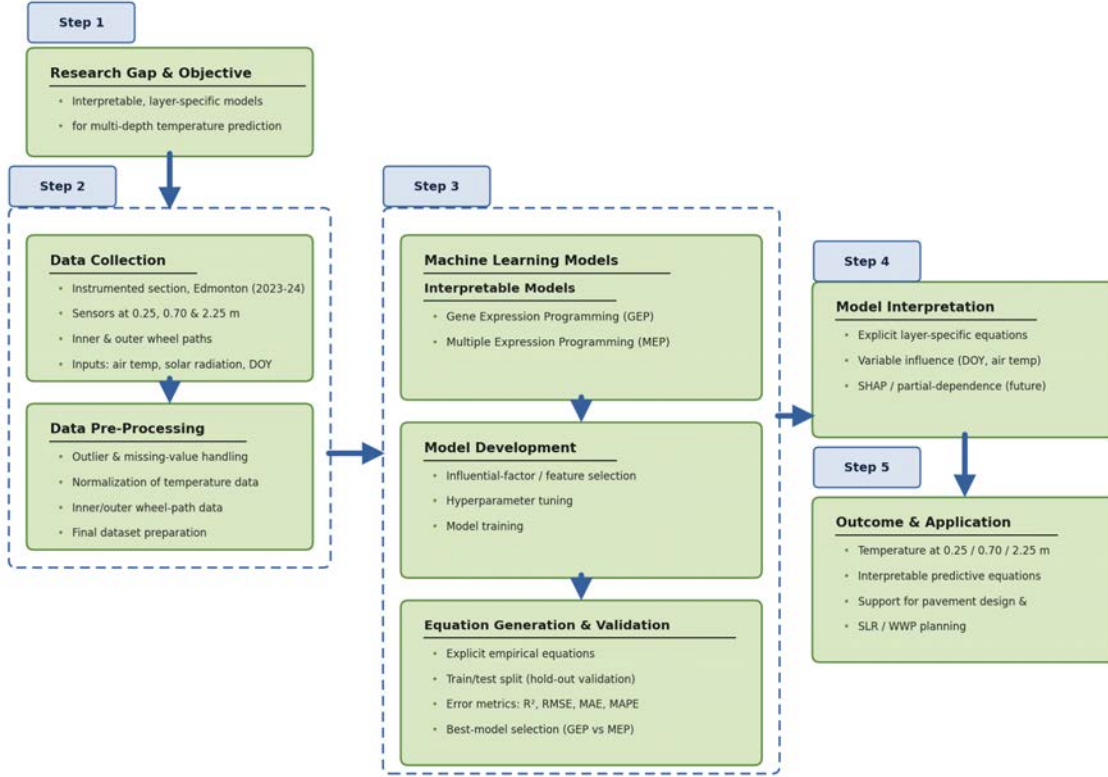


Figure 4. Research methodology.

6.0. Results and Discussion

6.1. Model at Depth 0.25 m:

At 0.25 m, the two models produced similar aggregate statistics in the training and testing periods. Model selection was based on higher R² and lower RMSE and MAE; practical adequacy near the freezing threshold was not evaluated.

For the GEP model at 0.25 m, the training R² was 0.825, with RMSE and MAE values of 5.24°C and 4.14°C, respectively. The corresponding test R² was 0.826, with RMSE of 5.21°C and MAE of 4.12°C. The similarity of the aggregate training and test statistics does not constitute external validation. The GEP equation is shown below.

$$T_{0.25} = \frac{0.7472 \cdot T_{\text{air}} - 0.8234 \cdot \cos(\text{DOY}) - 3.9074 \cdot \sin(\text{DOY}) + 7.5535}{\cos(\text{DOY}) + 1.744} \quad (1)$$

where $T_{0.25}$, $T_{0.70}$ and $T_{2.25}$ are the predicted pavement temperatures (°C) at depths of 0.25 m, 0.70 m and 2.25 m below the surface; T_{air} is the measured air temperature (°C); and $\sin(\text{DOY})$ and $\cos(\text{DOY})$ are the seasonal harmonic terms $\sin(2\pi d/365)$ and $\cos(2\pi d/365)$, in which d is the day of the year ($d = 1$ on 1 January). Equations 1, 3 and 5 take a rational form, but their denominators remain bounded away from zero over the full annual cycle $0.744 \leq \cos(\text{DOY}) + 1.744 \leq 2.744$, and $4.74 \leq (\cos(\text{DOY}) + 3.177)^2 \leq 17.45$ so the equations are numerically stable for every day of the year.

For the MEP model at 0.25 m, the training R^2 was 0.767, with RMSE and MAE values of 6.05°C and 4.65°C, respectively. The corresponding test R^2 was 0.767, with RMSE of 6.01°C and MAE of 4.61°C. MEP produced larger errors than GEP at this depth. The MEP equation is shown below.

$$T_{0.25} = T_{\text{air}} - 2 \cdot \cos(\text{DOY}) + 5.6723 \quad (2)$$

Overall, GEP produced lower aggregate errors than MEP at 0.25 m; however, the GEP test RMSE of 5.21°C should not be described as low without an application-specific acceptance criterion.

Figure 5 shows the measured and GEP-predicted temperature profile at 0.25 m.

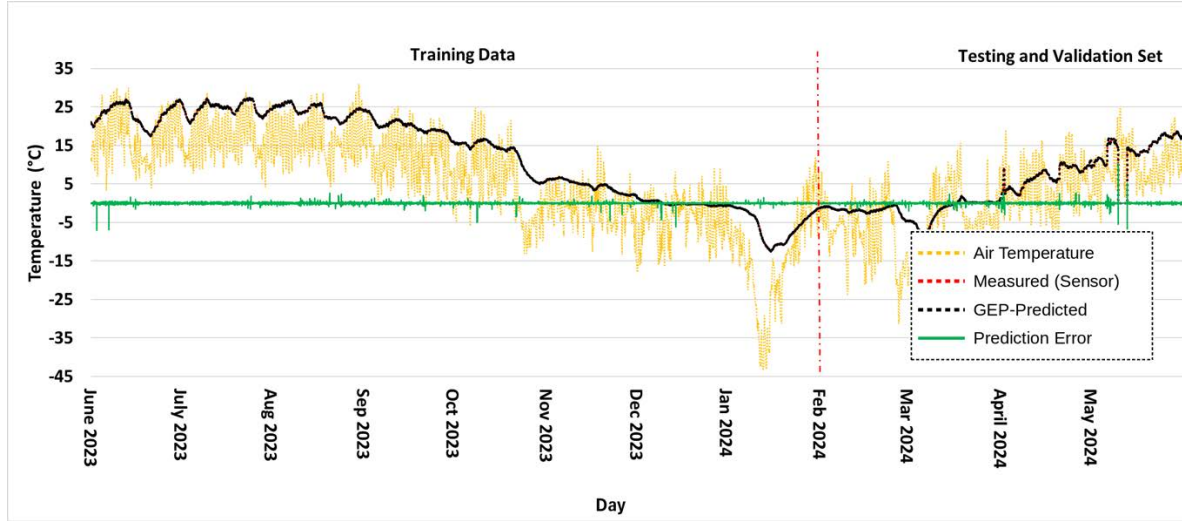


Figure 5. Measured and GEP-predicted temperature profiles at 0.25 m

6.2. Model at Depth 0.70 m:

At the nominal 0.70 m granular-base–subgrade interface, GEP and MEP were evaluated using the same statistical indicators. Practical adequacy near freezing was not assessed.

For the GEP model at 0.70 m, the training R^2 was 0.837, with RMSE and MAE values of 5.07°C and 3.99°C, respectively. The corresponding test R^2 was 0.838, with RMSE of 5.02°C and MAE of 3.97°C. The similarity of the aggregate training and test statistics does not establish performance outside the monitored annual cycle. The GEP equation is shown below.

$$T_{0.70} = \frac{0.7472 \cdot T_{\text{air}} - 2.7472 \cdot \sin(\text{DOY}) - (\cos(\text{DOY}) + 1.106)(\cos(\text{DOY}) + \sin(\text{DOY})) + 9.6164}{\cos(\text{DOY}) + 1.744} \quad (3)$$

For the MEP model at 0.70 m, the training R^2 was 0.754, with RMSE and MAE values of 6.22°C and 4.86°C, respectively. The corresponding test R^2 was 0.755, with RMSE of 6.17°C and MAE of 4.79°C. MEP produced larger errors than GEP at this depth. The MEP equation is shown below.

$$T_{0.70} = T_{\text{air}} - \sin(\text{DOY}) + 5.9634 \quad (4)$$

GEP produced lower aggregate errors than MEP at 0.70 m. The GEP test RMSE remained 5.02°C, so the result should be interpreted in relation to the intended decision threshold rather than characterized as uniformly accurate. Figure 6 shows the corresponding temperature profile.

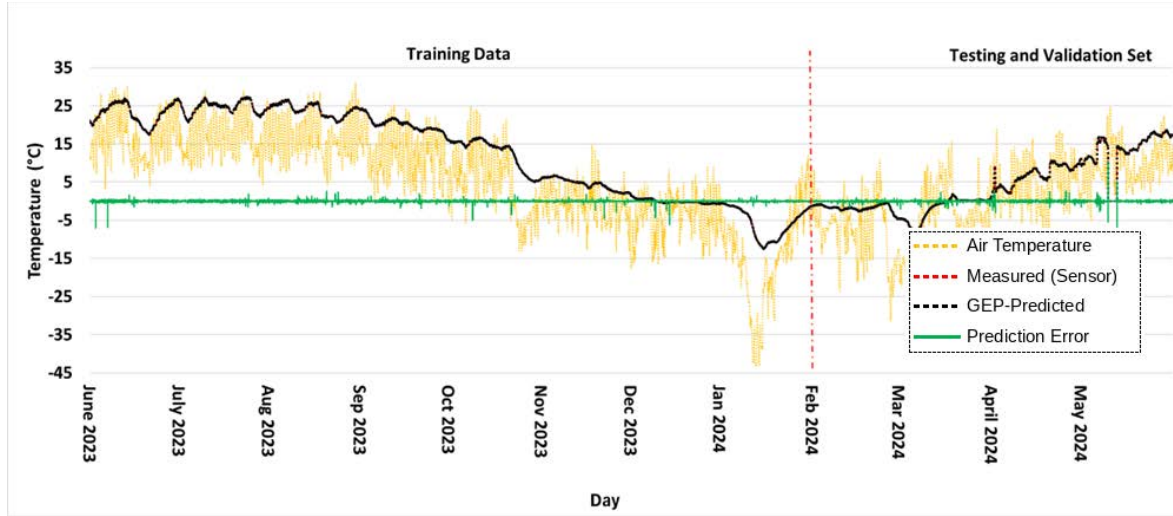


Figure 6. Measured and GEP-predicted temperature profiles at 0.70 m

6.3. Model at Depth 2.25 m:

At 2.25 m in the subgrade, the models produced lower absolute errors than at the shallower locations, coinciding with a smoother and less variable annual temperature signal.

For the GEP model at 2.25 m, the training R^2 was 0.923, with RMSE and MAE values of 1.90°C and 1.48°C, respectively. The corresponding test R^2 was 0.924, with RMSE of 1.89°C and MAE of 1.48°C. These are the strongest aggregate statistics among the evaluated depths, although they do not constitute external validation. For compactness, let $s = \sin(\text{DOY})$, $c = \cos(\text{DOY})$, and $C = c + 3.177$. The algebraically equivalent GEP equation is shown below.

$$T_{2.25} = \frac{T_{\text{air}}}{c^2} + \frac{7.922 - 4.695s}{c} + 7.922 - 2c - 4.695s \quad (5)$$

For the MEP model at 2.25 m, the training R^2 was 0.797, with RMSE and MAE values of 3.08°C and 2.48°C, respectively. The corresponding test R^2 was 0.795, with RMSE of 3.09°C and MAE of 2.49°C. MEP produced larger errors than GEP at this depth. Using the same s and c notation, the MEP equation is shown below.

$$T_{2.25} = -0.0766T_{\text{air}}s + 0.3127T_{\text{air}} + 0.5150cs - 4.1641s + 9.2475 \quad (6)$$

Temperature variation at 2.25 m was more gradual than at the shallower locations. GEP produced lower RMSE and MAE than MEP at this depth. The higher R^2 and lower RMSE do not by themselves establish greater transferability, because the deep temperature series is strongly seasonal and was evaluated using data from the same site and annual cycle. Figure 7 shows the GEP-based temperature profile at 2.25 m.

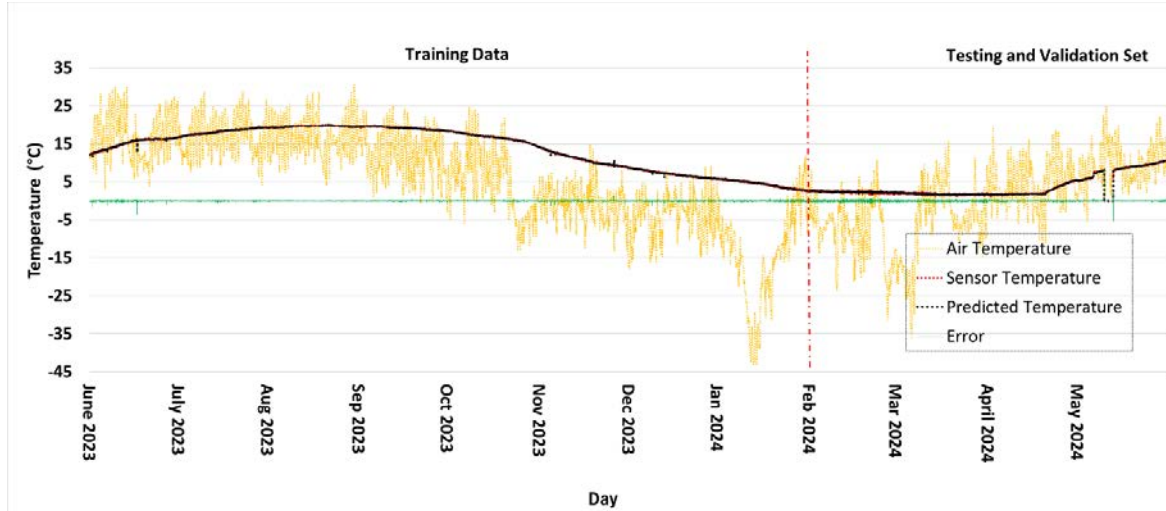


Figure 7. Measured and GEP-predicted temperature profiles at 2.25 m

The statistical analysis of the model at each depth is shown in Table 2.

Table 2. Model performance statistics

Depth and model	Train R^2	Train RMSE (°C)	Train MAE (°C)	Test R^2	Test RMSE (°C)	Test MAE (°C)
$T_{0.25}$ – GEP	0.825	5.244	4.144	0.826	5.207	4.125
$T_{0.25}$ – MEP	0.767	6.054	4.652	0.767	6.014	4.605
$T_{0.70}$ – GEP	0.837	5.071	3.994	0.838	5.022	3.966
$T_{0.70}$ – MEP	0.754	6.224	4.857	0.755	6.17	4.79
$T_{2.25}$ – GEP	0.923	1.897	1.483	0.924	1.887	1.477
$T_{2.25}$ – MEP	0.797	3.082	2.477	0.795	3.091	2.491

6.4. Comparison of the Models with the Existing Models:

Existing approaches for predicting subsurface pavement temperature can be broadly classified into mechanistic, empirical, and machine learning-based models. Mechanistic models, such as those implemented in Pavement ME Design and Enhanced Integrated Climatic Model (EICM), rely on heat transfer equations and require detailed material thermal properties, making them computationally intensive and data-demanding. Empirical models, including regression-based and BELLS equation approaches, provide simplified relationships but are often limited in accuracy and generalizability, particularly for deeper layers such as the granular base course (GBC) and subgrade. Recent machine learning models have improved prediction accuracy; however, their black-box nature limits interpretability and practical use in engineering applications.

The models developed in this study generate explicit mathematical expressions, but their performance

cannot be ranked directly against published models trained at different sites, depths, temporal resolutions, and validation periods. In the selected equations, air temperature and the day-of-year terms dominate because incoming solar radiation was not retained. This outcome should be confirmed through an ablation comparison and independent validation. Related machine-learning work on the same instrumented section provides context but not a controlled benchmark²⁰.

Table 3 therefore summarizes model scope and the performance reported in this study without treating accuracy values from different datasets as directly comparable.

Table 3. Scope of existing approaches and the proposed models

Model type	Examples	Typical inputs	Depth scope	Performance statement	Ref.
Mechanistic	Pavement ME; EICM	Weather and material thermal properties	Multiple depths	Requires model inputs and local calibration	17
Empirical	BELLS	Surface temperature, air temperature and time	Bound asphalt layers at depth	Study- and calibration-specific	6
Empirical	Regression	Selected weather variables	Usually depth-specific	Not directly comparable across datasets	18
Machine learning	RF and related learners	Multiple environmental variables	Single or multiple depths	Not directly comparable across datasets	19
Machine learning	Deep learning	Time-series and environmental variables	Single or multiple depths	Not directly comparable across datasets	8
Proposed	GEP and MEP	Candidate: air temperature, solar radiation and DOY; retained: air temperature and DOY	Depth-specific	Test R^2 : GEP 0.826–0.924; MEP 0.755–0.795	This study

6.5. Effect of Depth on Reported Performance Metrics

Depth affected GEP and MEP differently. For GEP, test R^2 increased from 0.826 at 0.25 m to 0.838 at 0.70 m and 0.924 at 2.25 m, while RMSE decreased from 5.21°C to 5.02°C and 1.89°C. MEP did not improve monotonically: test R^2 decreased from 0.767 at 0.25 m to 0.755 at 0.70 m before increasing to 0.795 at 2.25 m, and RMSE increased slightly from 6.01°C to 6.17°C before falling to 3.09°C. The general statement that model accuracy increased continuously with depth is therefore not supported for both methods.

R^2 and RMSE describe explained variance and absolute error; neither directly measures the physical coupling between surface weather and subsurface temperature. Thermal damping reduces the amplitude and short-term variability of temperature with depth, which can reduce RMSE. However, reduced amplitude alone does not necessarily increase R^2 because R^2 is normalized by the variance of the measured response.

The 2.25 m record appears smoother and more strongly seasonal than the shallower records, allowing the sine and cosine DOY terms to reproduce a larger portion of its variation. This is a plausible explanation for the improved GEP statistics, but it was not tested through an ablation experiment. A DOY-only baseline and a model excluding air temperature are needed to quantify how much skill comes from the seasonal

terms.

Accordingly, the higher GEP R^2 at 2.25 m should not be interpreted as evidence that the deepest model is more physically complete or more transferable. External validation is required for sites with different soil thermal properties, moisture, groundwater, and snow-cover conditions. Performance should also be reported by season and near the 0°C threshold before the equations are considered for operational use.

6.6. Model Limitations and Practical Considerations

The models presented here are based on preliminary findings from a single instrumented section and one year of record, and several limitations should be stated explicitly so that the equations are not applied beyond the conditions for which they were derived.

GEP and MEP use population-based searches over candidate symbolic expressions, so their computational demand depends on the population size, chromosome structure, number of generations, function set, stopping criterion, and number of independent runs. This study did not report runtime or benchmark computational cost against ANN, random-forest, or gradient-boosting models; therefore, no comparative efficiency conclusion can be drawn. The principal practical benefit evaluated here is the availability of an explicit equation rather than demonstrated computational efficiency.

In terms of scope, the models were calibrated on one instrumented section in Edmonton with a specific pavement structure, a well-graded granular base and a moderately frost-susceptible clayey-sand subgrade, and on a single annual cycle. They do not include moisture content, traffic loading, snow cover or material thermal properties, all of which influence subsurface temperature and any of which may differ at another site. The reported statistics should accordingly be regarded as a preliminary proof of concept and as site-specific relationships suitable for the Edmonton test section and comparable cold-region structures, rather than as a generalized predictive tool. Independent validation on additional sections and additional years is required before the equations are used in design or in seasonal load-restriction decisions.

7. Conclusions

The following conclusions can be drawn from the present study:

- GEP and MEP produced explicit, site-specific equations for temperatures at three monitored depths using air temperature and seasonal day-of-year terms; incoming solar radiation was evaluated but not retained in the final equations.
- GEP consistently outperformed MEP, producing higher R^2 values and lower RMSE and MAE across all depths.
- GEP performance improved with depth, while MEP performance was not monotonic between 0.25 m and 0.70 m. Lower absolute errors at 2.25 m coincided with a smoother, lower-variability temperature signal.
- Independent depth-specific modelling avoided the direct use of temperatures from other depths, but it does not by itself establish external generalizability or eliminate all forms of temporal dependence.
- Future research should incorporate moisture, traffic, material properties, multi-year data, and additional test sites to improve model robustness and generalizability.
- Cascaded, hybrid, and metaheuristic-optimized models could further improve predictive performance.
- Overall, the findings demonstrate the potential of evolutionary modelling for pavement temperature prediction, while representing a preliminary proof of concept requiring broader validation.

References

1. Huang, Y., Molavi Nojumi, M., Ansari, S., and Hashemian, L. "Multi-Depth Temperature Prediction Using Machine Learning for Pavement Sections." *Journal of Applied Remote Sensing*. 18 (1): 1-26. (2024) <https://doi.org/10.1117/1.JRS.18.014517>
2. Gui, J.G., Phelan, P.E., Kaloush, K.E., and Golden, J.S. "Impact of Pavement Thermophysical Properties on Surface Temperatures." *Journal of Materials in Civil Engineering*. 19 (8): 683-690. (2007) [https://doi.org/10.1061/\(ASCE\)0899-1561\(2007\)19](https://doi.org/10.1061/(ASCE)0899-1561(2007)19)
3. Asefzadeh, A., Hashemian, L., and Bayat, A. "Development of Statistical Temperature Prediction Models for a Test Road in Edmonton, Alberta, Canada." *International Journal of Pavement Research and Technology*. 10 (5): 369-382. (2017) <https://doi.org/10.1016/j.ijprt.2017.05.004>
4. Molavi Nojumi, M., Huang, Y., Hashemian, L., and Bayat, A. "Application of Machine Learning for Temperature Prediction in a Test Road in Alberta." *International Journal of Pavement Research and Technology*. 15 (2): 303-319. (2022) <https://doi.org/10.1007/s42947-021-00023-3>
5. Witczak, M., Andrei, D., and Houston, W. "Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures, Appendix DD-1: Resilient Modulus as Function of Soil Moisture - Summary of Predictive Models." Final Report, NCHRP Project 1-37A. p. 424. (2000)
6. Li, Y., Liu, L., and Sun, L. "Temperature Predictions for Asphalt Pavement with Thick Asphalt Layer." *Construction and Building Materials*. 160: 802-809. (2018) <https://doi.org/10.1016/j.conbuildmat.2017.12.145>
7. Saad, M., Farghaly, H., Yang, S.X., and Gharabaghi, B. "Hourly Road Pavement Surface Temperature Forecasting Using Deep Learning Models." *Journal of Hydrology*. 603. (2021) <https://doi.org/10.1016/j.jhydrol.2021.126877>
8. Fahad, M. "Data-Driven Pavement Performance: Machine Learning-Based Predictive Models." (2025)
9. Shah, H.A. "Prediction of Compressive and Splitting Tensile Strength of Concrete with Fly Ash by Using Gene Expression Programming." *Structural Concrete*. pp. 2435-2449. (2021) <https://doi.org/10.1002/suco.202100213>
10. Awan, M.N.U.A., Waim, A., Hashemian, L., Shafiee, M., and Bayat, A. "Comparative Analysis of Machine Learning Models for Multi-Depth Pavement Temperature Prediction." *Canadian Journal of Civil Engineering*. (2026) <https://doi.org/10.1139/cjce-2025-0158>
11. Farjad, M., Liu, Q.-F., Azim, I., Zhu, X., and Yang, J. "Prediction of Mechanical Properties of Green Concrete Incorporating Waste Foundry Sand Based on Gene Expression Programming." *Journal of Hazardous Materials*. 384: 121322. (2020) <https://doi.org/10.1016/j.jhazmat.2019.121322>
12. Khan, M.A., Zafar, A., Akbar, A., Javed, M.F., and Mosavi, A. "Application of Gene Expression Programming (GEP) for the Prediction of Compressive Strength of Geopolymer Concrete." *Materials*. 14 (5): 1-23. (2021) <https://doi.org/10.3390/ma14051106>
13. Ferreira, C. "Genetic Algorithms, Genetic Programming and Decision Functions." In *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*. Berlin: Springer, pp. 21-56. (2006)
14. Shah, M.I., Javed, M.F., and Abunama, T. "Proposed Formulation of Surface Water Quality and Modelling Using Gene Expression, Machine Learning, and Regression Techniques." *Environmental Science and Pollution Research*. 28 (11): 13202-13220. (2021) <https://doi.org/10.1007/s11356->

020-11490-9

15. Farjad, M., Faisal, M., Rauf, M., Azim, I., Ashraf, M., Yang, J., and Liu, Q.-F. "Sustainable Utilization of Foundry Waste: Forecasting Mechanical Properties of Foundry Sand Based Concrete Using Multi-Expression Programming." *Science of the Total Environment*. 780: 146524. (2021) <https://doi.org/10.1016/j.scitotenv.2021.146524>
16. Ilyas, I., Zafar, A., Javed, M.F., Farooq, F., Aslam, F., Musarat, M.A., and Vatin, N.I. "Forecasting Strength of CFRP Confined Concrete Using Multi Expression Programming." (2021)
17. National Cooperative Highway Research Program. "Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures under Traffic Loadings, Appendix AA." NCHRP Project 1-37A. (1999)
18. Swarna, S.T., and Hossain, K. "Climate Change Impact and Adaptation for Highway Asphalt Pavements: A Literature Review." *Canadian Journal of Civil Engineering*. 49 (7): 1109-1120. (2022) <https://doi.org/10.1139/cjce-2021-0209>
19. Tabrizi, S.E., Xiao, K., Van Griensven The, J., Saad, M., Farghaly, H., Yang, S.X., and Gharabaghi, B. "Hourly Road Pavement Surface Temperature Forecasting Using Deep Learning Models." *Journal of Hydrology*. 603. (2021) <https://doi.org/10.1016/j.jhydrol.2021.126877>
20. Huang, Y., Molavi Nojumi, M., Hashemian, L., and Bayat, A. "Evaluation of a Machine Learning Approach for Temperature Prediction in Pavement Base and Subgrade Layers in Alberta, Canada." *Journal of Transportation Engineering, Part B: Pavements*. 149 (1): 1-12. (2023) <https://doi.org/10.1061/JPEODX.PVENG-1010>
21. Huang, Y., Molavi Nojumi, M., Ansari, S., Hashemian, L., and Bayat, A. "Evaluating the Use of Machine Learning for Moisture Content Prediction in Base and Subgrade Layers." *Road Materials and Pavement Design*. 24 (12): 2910-2928. (2023) <https://doi.org/10.1080/14680629.2023.2182135>
22. Huang, Y., Baghaee Moghaddam, T., Hashemian, L., and Bayat, A. "Application of the Machine Learning Method to Determine Spring Load Limits and Winter Weight Premium." *Transportation Research Record*. 2678 (11): 1935-1948. (2024) <https://doi.org/10.1177/03611981241246780>
23. Kebede, Y.B., Yang, M.-D., and Huang, C.-W. "Real-Time Pavement Temperature Prediction Through Ensemble Machine Learning." *Engineering Applications of Artificial Intelligence*. 135: 108870. (2024) <https://doi.org/10.1016/j.engappai.2024.108870>
24. Bai, S., Dai, B., Yang, Z., Zhu, F., Yang, W., and Li, Y. "Enhancing Road Surface Temperature Prediction: A Novel Approach Integrating Transfer Learning with Long Short-Term Memory Neural Networks." *Journal of Transportation Engineering, Part B: Pavements*. 151 (1): 04024063. (2025) <https://doi.org/10.1061/JPEODX.PVENG-1616>
25. Ma, Y., Park, S., Bae, A., Kwon, K., and Choi, H. "Long Short-Term Memory-Based Real-Time Prediction Models for Freezing Depth and Thawing Time in Unbound Pavement Layers." *Computer-Aided Civil and Infrastructure Engineering*. 40 (20): 3182-3196. (2025) <https://doi.org/10.1111/mice.13516>
26. Oyeyi, A.G., Khan, A., Huyan, J., Zhang, W., Ni, F.M.-W., and Tighe, S.L. "Ensemble and Evolutionary Prediction of Layers Temperature in Conventional and Lightweight Cellular Concrete Subbase Pavements." *International Journal of Pavement Engineering*. 25 (1): 2322525. (2024) <https://doi.org/10.1080/10298436.2024.2322525>

27. Huyan, J., Oyeyi, A.G., Khan, A., Zhang, W., and Tighe, S.L. "Predicting Layer Temperatures in Flexible Pavement with Lightweight Cellular Concrete Subbase Using Explainable Machine Learning." *Road Materials and Pavement Design*. 27 (6): 1849-1878. (2026) <https://doi.org/10.1080/14680629.2025.2531221>
28. Padala, S.K., Swamy, A.K., and Bhattacharjee, B. "Air Temperature Prediction Models for Pavements Based on the Gene Expression Programming Approach." *Journal of Transportation Engineering, Part B: Pavements*. 151 (1): 04024059. (2025) <https://doi.org/10.1061/JPEODX.PVENG-1496>
29. Khan, M., Ali, M., Najeh, T., and Gamil, Y. "Computational Prediction of Workability and Mechanical Properties of Bentonite Plastic Concrete Using Multi-Expression Programming." *Scientific Reports*. 14: 6105. (2024) <https://doi.org/10.1038/s41598-024-56088-0>
30. Jiang, Y., Ullah, S., Zapata, C., and Yu, X. "Assessment of the Impacts of Climate Change on Cold Region Pavement Design and Operation." *Journal of Infrastructure Preservation and Resilience*. 6 (1): 20. (2025) <https://doi.org/10.1186/s43065-025-00126-2>