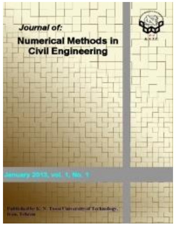
Numerical Methods in Civil Engineering, 10-1 (2025) 01-12



Numerical Methods in Civil Engineering

Journal Homepage: https://nmce.kntu.ac.ir/

A machine learning algorithm for the prediction of the viscoelastic properties of asphalt mixtures

Fernando Martinez[*], **Silvia Angelone**[**], **Marina Cauhapé Casaux**[***], **Luis Zorzutti**[****]

ARTICLE INFO

RESEARCH PAPER

Article history:
Received:
April 2025
Revised:
June 2025
Accepted:
July 2025

Keywords:
Asphalt Mixtures,
Machine Learning,
Viscoelastic Properties,
Dynamic Modulus,
Phase Angle

Abstract:

The dynamic modulus |E*| and phase angle ϕ are the most important properties for the viscoelastic characterization of asphalt mixtures. Their experimental determination requires time-consuming procedures and expensive laboratory equipment. Hence, different prediction procedures have been developed for the estimation of these rheological properties, with Witczak and Hirsch models being the most widely accepted. Nowadays machine learning (ML) techniques are applied to various engineering problems because of their abilities in data processing, optimization and estimation. This paper proposes the K-Nearest Neighbors algorithm as an ML method for the prediction of the viscoelastic properties of asphalt mixtures. The training and validation of the algorithm was based on a database containing the bitumen characteristics, volumetric properties and dynamic modulus and phase angle values at different frequencies and temperatures with more than 5500 data points. The obtained results indicate that the ML algorithm developed in this study is accurate and it could be an effective approach to predict the viscoelastic properties of asphalt mixtures.

1. Introduction

The dynamic modulus |E*| and the phase angle ϕ are the most important viscoelastic properties of asphalt mixtures characterizing their rheological behavior. The dynamic modulus measures the stiffness of a mixture, defined by the test temperature and frequency of loading conditions.

It is the primary material property for modern pavement design procedures based on mechanistic principles, as it determines the distribution of stresses and strains within the pavement structure and can be correlated with the rutting and fatigue cracking of bituminous layers.

Additionally, the phase angle characterizes the phase lag between applied stress and strain response in viscoelastic materials, serving as an indicator of the viscous and elastic components under specified test conditions.

Typically, these dynamic viscoelastic properties are determined in the laboratory using sophisticated equipment and skilled technicians. When such equipment is unavailable, the dynamic modulus can be estimated using predictive models based on the mixture's volumetric properties, gradation, and binder characteristics.

Various |E*| predictive models have been developed to estimate the dynamic modulus as an alternative to laboratory testing. Among the most widely used models are the Witczak predictive model [1, 2], the Hirsch model [3, 4] and the Al-Khateeb model [5].

Various authors have evaluated these predictive models and generally concluded that reliable first-order dynamic modulus estimates can be obtained using these procedures.

Less attention has been given to predicting the phase angle, and only a few research studies have established predictive models for this viscoelastic property in asphalt materials [6, 7, 8, 9, 10].

**Corresponding author: Professor, Institute of Applied Mechanics and Structures, Faculty of Exact Sciences, Engineering and Surveying, National University of Rosario, Rosario, Argentina. Email:* fermar@fceia.unr.edu.ar

*** Professor, Institute of Applied Mechanics and Structures, Faculty of Exact Sciences, Engineering and Surveying, National University of Rosario, Rosario, Argentina. Email:* sangelon@fceia.unr.edu.ar

**** Assistant Professor, Institute of Applied Mechanics and Structures, Faculty of Exact Sciences, Engineering and Surveying, National University of Rosario, Rosario, Argentina. Email:* mccasaux@fceia.unr.edu.ar

***** Lecturer, Institute of Applied Mechanics and Structures, Faculty of Exact Sciences, Engineering and Surveying, National University of Rosario, Rosario, Argentina. Email:* zorzutti@fceia.unr.edu.ar

https://doi.org/10.61882/NMCE.2504.1085

More recently, machine learning (ML) techniques have been applied to various engineering problems due to their capabilities in data processing, optimization, and estimation. ML algorithms use computational methods to "learn" information directly from historical data or experience. They have become increasingly popular because they can process large data sets, identify natural patterns, and make informed decisions based on improved predictions [11].

Chaabene et al. [12] reviewed and compared different ML techniques used for forecasting the mechanical properties of concrete materials and structures. The advantages and drawbacks of those techniques were critically discussed and compared. The authors concluded that the performance of the models is influenced by various factors, such as the nature of the relationship between inputs and outputs and the size of the training data set.

The study of Khambra and Shukla [13] presents the development and application of machine learning techniques on fly ash-based concrete using different models, algorithms, and approaches for predicting engineering properties. This paper states that machine learning is a useful and powerful technique for predicting material engineering properties and represents a scientific challenge in the construction and infrastructure sectors.

Different ML algorithms and techniques were applied in engineering problems like gradient boosting, k-nearest neighbors and support vector regression to predict the compressive strength of concrete [14]; gradient boosted regression tree (GBRT) algorithms were used to predict the compression modulus of soils in geotechnical foundations [15].

Neural network, stochastic gradient descent, support vector machine and decision tree ML algorithms were used to predict the frost depth in asphalt pavements in South Korea [16]. Bajic et al. [17] utilized different types of Neural Networks to estimate the road roughness of asphalt pavements. The paper demonstrates that machine learning methods can accurately predict road roughness, using the recordings of the cost approachable in-vehicle sensors installed in conventional passenger cars and the technology is well suited to meet future pavement condition monitoring, by enabling continuous monitoring of a wide road network.

Suleymanov et al. [18], Huang et al. [19], and Ghorbani et al. [20] employed various machine learning algorithms to estimate the resilient modulus of fine-grained soils and the moisture content in base and subgrade layers. The predictive performance of these algorithms was evaluated using statistical indices, and the results demonstrated the robustness of machine learning methods in predicting these soil properties.

Related to asphalt mixtures and binders, Atakan and Yıldız [21] predicted Marshall design parameters of asphalt mixtures using machine learning algorithms based on literature data. Tangga et al. [22] have used an integrated machine learning methodology to improve the prediction of Marshall characteristics of asphalt pavement layers. Upadhya et al. [23] investigated the potential of soft computing-based models, such as Artificial neural networks, Support vector machines, Gaussian process, M5P tree, Random forest, and Random tree-based models, for the prediction of Marshall Stability of carbon–fiber asphalt mixtures.

These three studies demonstrate that the proposed models effectively capture the influence of various features on the target parameters and predict their actual values with high accuracy.

Also Alnaqbi et al. [24] used a range of machine learning models for the IRI prediction in flexible pavements. Highly accurate IRI prediction model that surpasses traditional empirical approaches could be obtained with these selected algorithms.

Gul et al. [25] developed prediction models for the Marshall Stability and Marshall Flow of asphalt pavements using three different soft computing techniques. Uwanuakwa et al. [26] have compared the performance of five different machine learning models with the empirical prediction model developed by Witczak and collaborators and they concluded that the proposed ML models have higher predictive accuracy compared to the Witczak regression model.

Martinez and Angelone [27] used Artificial Neural Network (ANN) techniques in order to develop a robust prediction model of the dynamic modulus of asphalt mixtures. A biogeography-based optimization (BBO) algorithm of ML was used to develop a predictive |E*| procedure with improved accuracy compared to previously empirically developed models [28].

Leiva-Villacorta and Vargas-Nordcbeck [29] developed an Artificial Neural Network (ANN) model to estimate |E*| for asphalt mixtures in Costa Rica.

Useche-Castelblanco et al. [30] applied three different ML techniques to predict the rheological properties of wax-modified asphalt binders. Rahman et al. [31] used support vector regression analysis and decision tree-based ensemble ML methods to predict metrics related to asphalt mixture performance.

Artificial Neural Network, Adaptive Neuro-Fuzzy Inference System, and Multi Expression Programming as Supervised ML algorithms were used for the prediction of Marshall Stability and Marshall Flow of asphalt mixtures [25]. Botella et al. [32] applied three different ML techniques to estimate the degree of binder activity of reclaimed asphalt pavement. Majidifard et al, [33] utilized innovative machine learning methods known as gene expression programming and hybrid artificial neural network/simulated annealing to predict the fracture energy of asphalt mixture specimens.

Shu and Huang [34] employed a differential method to predict the dynamic modulus and phase angle of asphalt mixtures. The major characteristics of asphalt mixtures (viscoelastic effect, aggregate gradation, and air voids) were taken into account in the predicting procedures.

Rondinella et al. [35] proposed an innovative ML approach for the simultaneous prediction of the dynamic modulus and the phase angle using data from nine different asphalt mixtures tested by the 4-point bending test. These authors concluded that ML-based methods outperformed the implemented empirical equations, showing much better performance.

Additionally, Rondinella et al. [36] used an artificial neural network (ANN) methodology for the simultaneous prediction of the dynamic modulus and phase angle of asphalt concrete mixtures. In this study, the empirical Witczak 1-37A equation, a well-established regression model, was used as a reference to compare the performance obtained by the model in terms of dynamic modulus. Machine learning predictions demonstrated remarkable accuracy leading to outperforming regression-based ones with respect to all the evaluation metrics used. Both in terms of dynamic modulus and phase angle, Pearson correlation coefficients and coefficients of determination achieved by the model were higher than 0.98, resulting in a powerful and reliable predictive tool.

This paper proposes the K-Nearest Neighbors (KNN) algorithm as an ML method for the simultaneous prediction of the dynamic modulus $|E^*|$ and the phase angle ϕ of asphalt mixtures.

Compared to other machine learning applications (e.g., artificial neural networks (ANN), support vector machines (SVM), and gradient-boosted regression trees (GBRT)), the K-nearest neighbors (KNN) model was selected due to its conceptual simplicity, ease of implementation, and low computational cost. This study aims to develop a dual-property prediction model for the main rheological characteristics of asphalt mixtures by integrating a large database encompassing a wide range of binder properties, volumetric characteristics, and testing conditions

The KNN method was successfully used for the characterization of cracking in pavement distress [37], predicting phases in high-entropy alloys [38], modeling Marshall Stability test data [39], predicting fuel consumption of agricultural machinery [40] and for text categorization [41]. Ahmed et al. [42] have proposed a spatial k-nearest neighbour method for nonparametric prediction of real-valued spatial data and supervised classification for categorical spatial data.

Ikeagwuani et al. [43] have used the KNN algorithm among other algorithms for the prediction of resilient modulus of fine-grained soil

It is considered that the KNN algorithm is simple to develop and intuitive to understand, sufficiently robust for prediction purposes, and could be implemented in a spreadsheet.

The development of the KNN model for estimating the viscoelastic properties of asphalt mixtures, the obtained results, and the evaluation of its predictive performance is presented and discussed in the following sections.

2. The K-Nearest Neighbours (KNN) algorithm

The K-Nearest Neighbours (KNN) algorithm is a nonparametric, supervised machine learning method used for regression and classification problems using 'feature similarity' to predict the value of any new data points. It is considered a lazy learning algorithm, with a low computational cost and straightforward implementation. The algorithm is based on the intuitive assumption that objects close in distance are potentially similar.

The K-Nearest Neighbour method has the advantage that it is completely nonparametric, meaning that no assumptions are made with regards to the statistical distribution of the data [44]. KNN is also well suited to handling discontinuities in the relationships between variables [45]. An important advantage of KNN is the independence of any data distribution since it uses local neighbourhoods. Finally, the K Nearest Neighbour method is computationally inexpensive for the implementation, since this merely involves the assembly of a database.

However, the disadvantage in computational time comes when predictions are made, since similarity distances must be computed for every example in the database each time a prediction is made. In addition, since the entire database needs to be available to use the method, it is not particularly compact or portable when incorporating into larger applications.

The KNN algorithm requires a training dataset containing a given number of objects with inputs and outputs. The training objects are vectors in multidimensional feature space with n inputs variables (predictors) and m outputs such as $x_i = \{x_{i1}, x_{i2} \ldots . x_{in}, y_{i1}, y_{i2}, y_{im}\}$.

Figure 1 shows an example with only two predictors named x_1 and x_2. Each data point has an output y. The point 1 has inputs x_{11} and x_{21} and an output y_1. The point 2 has inputs x_{12} and x_{22} and an output y_2. The estimated output value y_U for a new data point U with inputs x_{U1} and x_{U2} is calculated as a weighted average of the distances d_i and the outputs y_i of the four nearest neighbours if K is 4.

The value of K is a user-defined parameter that determines the number of nearest neighbours used to make the prediction based on the distance between the new data point and its neighbours.

A commonly used distance metric d_i is the Euclidean distance as the length of a line segment between two points that can be calculated from the Cartesian coordinates of the points as:.

$$\mathrm{d}_i = \sqrt{(x_{U1} - x_{i1})^2 + (x_{U2} - x_{i2})^2} \tag{1}$$

with i = 1, 2, 3 and 4.

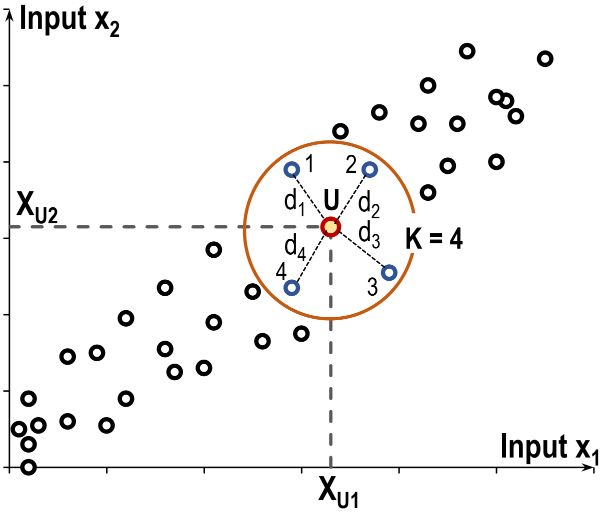


Fig. 1: Example of the KNN algorithm

The algorithm uses a weighted average of the K nearest neighbours, weighted by the inverse of their distance in a manner that the nearer neighbours contribute more to the average than the more distant ones. For K = 4, the estimation results:

$$\mathrm{d}_m = \frac{d_1 + d_2 + d_3 + d_4}{4} \tag{2}$$

$$y_U = \frac{y_1 \cdot {}^{d_1}/_{d_m} + y_2 \cdot {}^{d_2}/_{d_m} + y_3 \cdot {}^{d_3}/_{d_m} + y_4 \cdot {}^{d_4}/_{d_m}}{\left(\frac{d_1}{d_m} + \frac{d_2}{d_m} + \frac{d_3}{d_m} + \frac{d_4}{d_m}\right)} \tag{3}$$

The performance of the algorithm strongly depends on the parameter K. An optimal value for K is typically determined empirically, by cross-validation with different K values and determining which provides the lowest error.

3. Materials and procedures

3.1 Compiled data

In order to support the development of the ML algorithm, a unique database was compiled from four different datasets reported in the literature.

The first dataset (called the Arizona dataset) with 2183 data points was developed at the Arizona State University by Dr. Witczak and collaborators as a part of Task C of the NCHRP 9-19 Project, Superpave Support and Performance Models Management [46]. It contains data from 57 asphalt mixtures with conventional, polymer modified and asphalt-rubber binders.

The second dataset (called the Maryland dataset) with 2132 data points was developed at the University of Maryland and used for the calibration of the predictive model for the Dynamic Modulus of asphalt mixtures [47]. It contains results of 41 different dense asphalt mixtures made with conventional asphalt binders.

The third dataset (called the Virginia dataset) with 985 data points was developed by Dr. Flintsch and collaborators at the Virginia Tech Transportation Institute [48]. This dataset contains data of 31 asphalt concretes of base, intermediate and surface layers with conventional asphalt binders.

Finally, the fourth dataset (called the Florida dataset) with 265 data points was developed by Dr. Birgisson and collaborators at the University of Florida as a part of a project to develop complex modulus capabilities for the state of Florida [49]. This database contains information of 27 different asphalt mixtures of varying gradations and aggregate types.

The data contained in the different sources was compiled and the main volumetric characteristics of the mixtures, binder properties and, testing conditions were retained for the analysis presented in this paper. No homogenization of the different datasets was necessary, as all the data in the database were obtained using well-established experimental standard procedures.

Specifically, the percentage by volume of effective binder content (Vb), the percentage by volume of air voids (Va), the Voids in the Mineral Aggregates (VMA), the Voids Filled with Asphalt (VFA), the viscosity of the binder valued by their A and VTS parameters, the testing temperature (T) and the frequency (f). Descriptive statistics for the retained data in these four databases are listed in Table 1.

3.2 Selected model inputs

The selection of the model inputs for the KNN algorithm was basically intuitive, based on a survey of the literature. Based on a general overview of the sources, the most commonly cited factors critical to the viscoelastic properties of the asphalt mixtures are the asphalt binder consistency, the volumetric properties, and the testing conditions (temperature and frequency).

From the A and VTS parameters, the logarithmic value of the binder viscosity Visc (cP) at the testing temperature T was calculated as:

$$\log(Visc) = 10^{[A+VTS \cdot \log(T_R)]} \tag{4}$$

$$T_R = \frac{9}{5} \cdot T + 491.67 \tag{5}$$

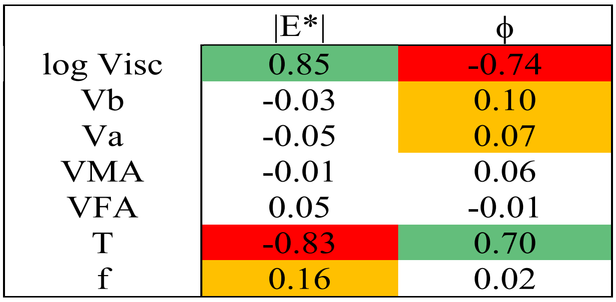
with T_R, temperature in °Rankine.

Table 1: Descriptive statistics for the reported data

		Asphalt binder		**Volumetric properties**				**Test conditions**		**Test results**	
Dataset		**A**	**VTS**	**Vb**	**Va**	**VMA**	**VFA**	**T**	**f**	**|E*|**	**ϕ**
Arizona	Max.	11.3750	-2.7384	22.2	18.5	40.7	79.0	54.4	25.0	64463	48.4
	Min.	8.3904	-3.8220	10.2	3.8	15.5	54.6	-10.0	0.1	118	2.8
	Ave.	10.4026	-3.4702	14.2	7.4	21.0	66.5	21.8	7.1	11035	23.0
	Std. Dev.	0.9060	0.3309	3.3	2.9	5.9	6.8	23.0	8.8	11256	11.4
Maryland	Max.	11.1165	-2.8989	16.3	11.3	22.9	85.5	54.4	25.0	42816	48.6
	Min.	8.8325	-3.7237	8.5	2.3	10.7	46.5	-17.8	0.1	86	4.0
	Ave.	10.3345	-3.4395	12.0	6.3	15.9	60.3	19.8	7.6	8339	24.6
	Std. Dev.	0.5723	0.2109	1.9	1.5	2.0	7.3	24.8	9.2	8416	10.3
Virginia	Max.	10.9800	-3.6800	15.2	7.6	22.8	69.6	54.4	25.0	40311	38.1
	Min.	10.9800	-3.6800	9.6	5.1	15.9	58.2	-12.2	0.1	125	0.7
	Ave.	10.9800	-3.6800	12.0	6.7	18.7	64.1	21.0	7.0	9022	19.8
	Std. Dev.	0.0000	0.0000	1.6	0.5	1.9	2.8	23.5	8.8	9355	10.6
Florida	Max.	9.8466	-3.2640	14.8	4.5	18.5	78.5	40.0	16.0	14398	48.9
	Min.	9.8466	-3.2640	7.3	3.7	10.5	63.8	10.0	1.0	317	14.1
	Ave.	9.8466	-3.2640	11.1	4.0	14.7	72.4	24.9	7.6	3845	29.7
	Std. Dev.	0.0000	0.0000	1.7	0.2	1.7	3.3	13.3	5.7	2937	7.1

A, VTS : Asphalt binder viscosity-temperature parameters [50]; Vb: Binder volume (%); Va: Air voids (%); VMA: Voids in the Mineral Aggregate (%); VFA: Voids Filled with Asphalt (%); T: Test temperature (°C); f: Test frequency (Hz); |E*|: Dynamic modulus (MPa); ϕ: Phase angle (°)

For the selection of the model inputs, the Pearson correlation coefficient has been used for a preliminarily examination of the strength of the relationships between the available inputs and the dynamic modulus and phase angle. Pearson correlation coefficient between two variables always ranges between +1 and −1. An absolute value around 1 represents a perfect correlation, whereas an absolute value around 0 represents no correlation. Plus and minus signs stand for direct and inverse proportionality, respectively. The Pearson coefficients between the available inputs, |E*| and ϕ are presented in Figure 2.



	|E*|	ϕ
log Visc	0.85	-0.74
Vb	-0.03	0.10
Va	-0.05	0.07
VMA	-0.01	0.06
VFA	0.05	-0.01
T	-0.83	0.70
f	0.16	0.02

Fig. 2: Pearson correlation coefficients

Figures 3(a) and 3(b) show the variation of |E*| with the logarithmic values of binder viscosity and temperature, respectively, as these represent the most significant correlations identified through Pearson analysis. A strong positive correlation exists between binder viscosity and the dynamic modulus (as binder viscosity increases, |E*| also increases), while a strong negative correlation is observed between temperature and the dynamic modulus (as testing temperature increases, |E*| decreases).

Figures 4(a) and 4(b) show the variation of ϕ with the logarithmic values of binder viscosity and temperature, respectively. A strong positive correlation exists between the testing temperature and the phase angle (if the temperature increases, ϕ increases and the binder becomes less elastic). On the other hand, a negative correlation occurs between the binder viscosity and the phase angle. A moderate correlation exists between the binder volume Vb and the air void content Va with the phase angle and between the testing frequency f and the dynamic modulus. Based on these observations, log Visc, Vb, Va, T and f were selected as inputs for the development of the first KNN model named as Reduced Model.

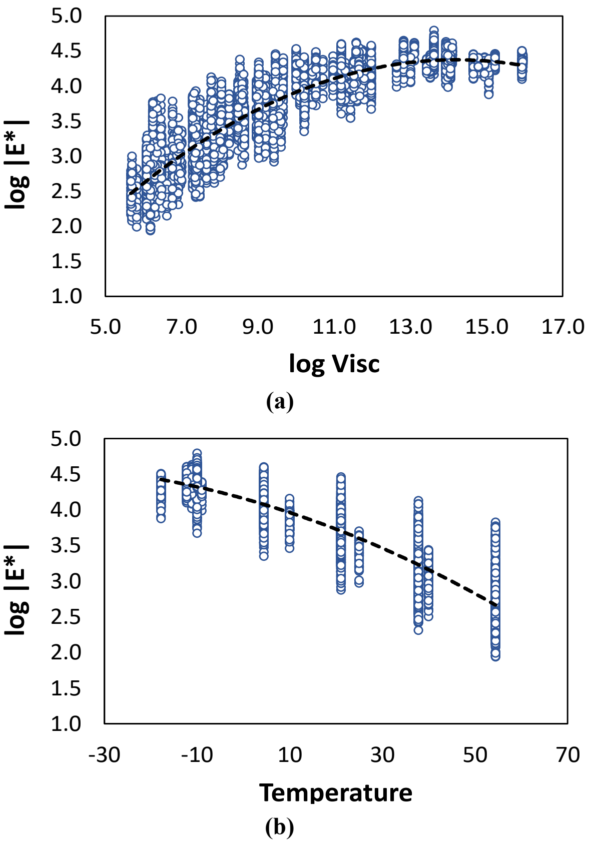


Fig. 3: Variation of |E*| with binder viscosity (a) and temperature (b)

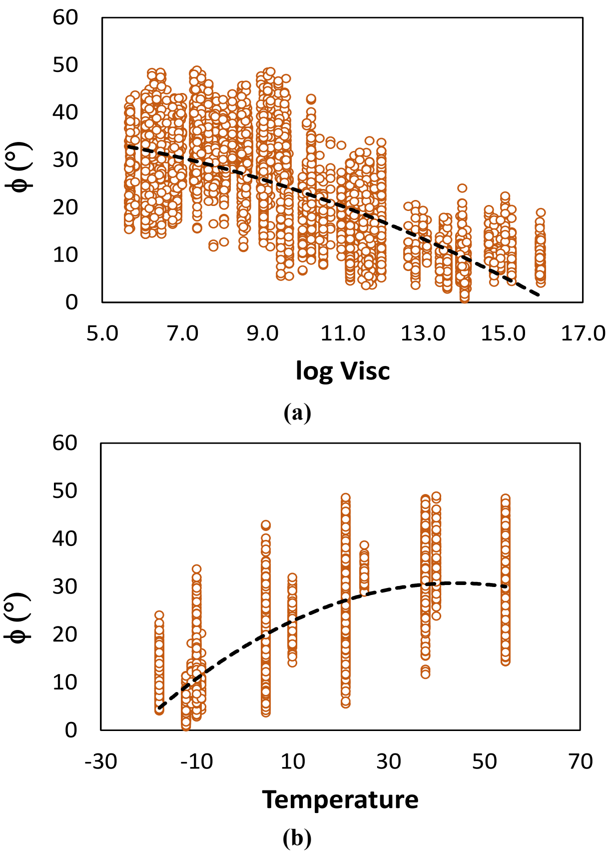


Fig. 4: Variation of ϕ with binder viscosity (a) and temperature (b)

The pavement community commonly adopts the Hirsch and Al-Khateeb models, which use VMA and VFA as key properties influencing the viscoelastic characteristics of asphalt mixtures. Therefore, a second KNN model, named Complete Model, was developed with VMA and VFA as additional inputs.

For both models, the outputs were |E*| and ϕ. Thus, the choice of the retained inputs was based solely on the Pearson analysis and empirical evidence of the influence of these volumetric parameters on the mechanical properties of asphalt mixtures.

3.3 Training and testing datasets

When using the KNN algorithm, the dataset should be divided into two subsets: the training dataset, on which the algorithm bases its predictions; and the validation dataset, which is used to test the algorithm's performance on previously unseen data [51].

Thus, the data was randomized and divided into two datasets. The first dataset, consisting of 75% of the data points (4174 inputs and outputs), was used as a training dataset, and the remaining 25% of the data points (1391 inputs and outputs) were utilized as a validation dataset. Thus, the results were based on a single split of the available data. The algorithm was implemented in an Excel spreadsheet where, for a given set of inputs in the validation dataset, the predicted values of |E*| and ϕ are automatically calculated.

3.4 Optimal K value

The choice of the K parameter is important because it influences the predicted performance of the model. If K is small then the variability of the predictions could be high, but when K is large then the predictions could be based on points that are relatively far away from our point of interest [52]. An optimal K value is usually determined empirically through a cross-validation procedure with different K values and determining which provides the lowest error. The normalized root mean square error (NRMSE), shown in Equation 6, was chosen as the metric for measuring the relative performance of the algorithm with different K values.

$$NRMSE = NRMSE_{|E*|} + NRMSE_{\emptyset} \tag{6}$$

$$NRMSE_{|E*|} = \left[\frac{\sqrt{\frac{\sum_{1}^{N}\left(|E^{*}|_{i} - |E^{*}|_{pred}\right)^{2}}{N}}}{|E^{*}|_{mean}} \right] \tag{7}$$

$$NRMSE_{\emptyset} = \left[\frac{\sqrt{\frac{\sum_{1}^{N}(\emptyset_i - \emptyset_{pred})^2}{N}}}{\emptyset_{mean}} \right] \qquad (8)$$

where $NRMSE_{|E^*|}$ = normalized root mean square error for the |E*| values; $NRMSE_{\phi}$ = normalized root mean square error for the ϕ values; N = number of cases in the dataset; $|E^*|_i$ and ϕ_i = measured |E*| and ϕ values for the ith case in the dataset; $|E^*|_{pred}$ and ϕ_{pred} = predicted |E*| and ϕ values for the ith case in the dataset; $|E^*|_{mean}$ and ϕ_{mean} = mean of the measured |E*| and ϕ values in the dataset.

Table 2 shows the variations of $NRMSE_{|E^*|}$, $NRMSE_{\phi}$ and NRMSE for different K values for the Reduced Model while Table 3 shows the corresponding variations for the Complete Model. Variation of NRMSE for both models is presented in Figure 5.

Table 2: Variation of NRMSE for the Reduced Model

K	$NRMSE_{\|E^*\|}$	$NRMSE_{\phi}$	NRMSE
3	0.392	0.223	0.615
5	0.372	0.217	0.589
6	0.378	0.217	0.595
7	0.385	0.218	0.603
9	0.382	0.222	0.604
11	0.382	0.223	0.605

Table 3: Variation of NRMSE for the Complete Model

K	$NRMSE_{\|E^*\|}$	$NRMSE_{\phi}$	NRMSE
3	0.377	0.196	0.573
5	0.361	0.197	0.558
6	0.363	0.197	0.560
8	0.368	0.200	0.568
10	0.374	0.207	0.581

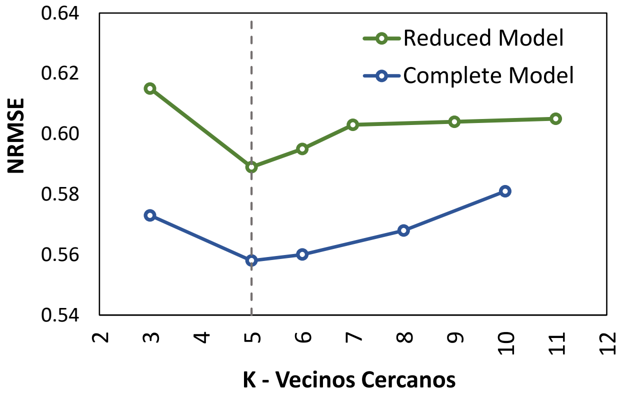


Fig. 5: NRMSE for both models

An optimal K value equal to 5 was selected for both models providing the minimum NRMSE value.

3. 5 Model evaluation

Six different goodness-of-fit metrics were selected and implemented in order to evaluate the performance of the developed models. Specifically, they included: mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), root mean squared error (RMSE), correlation coefficient (R^2) and the relationship between the standard error of predicted values and the standard deviation of measured values (Se/Sy). Singh et al. [53] have proposed a subjective criterion shown in Table 4 as a classification measure of the overall performance of the models.

Table 4: Criteria for goodness-of-fit statistical parameters

Criteria	R^2	Se/Sy
Excellent	≥ 0.90	≤ 0.35
Good	0.70 – 0.89	0.36 - 0.55
Fair	0.40 – 0.69	0.56 – 0.75
Poor	0.20 – 0.39	0.76 – 0.89
Very Poor	≤ 0.19	≥ 0.90

4. Results and discussion

Measured and predicted values were compared in order to examine the scattering of the data along the line of equality (LOE). Figure 6 shows the comparison of measured and predicted |E*| values while Figure 7 shows the comparison of measured and predicted ϕ values for the Reduced Model. Figures 8 and 9 show the same comparison of measured and predicted values for the Complete Model.

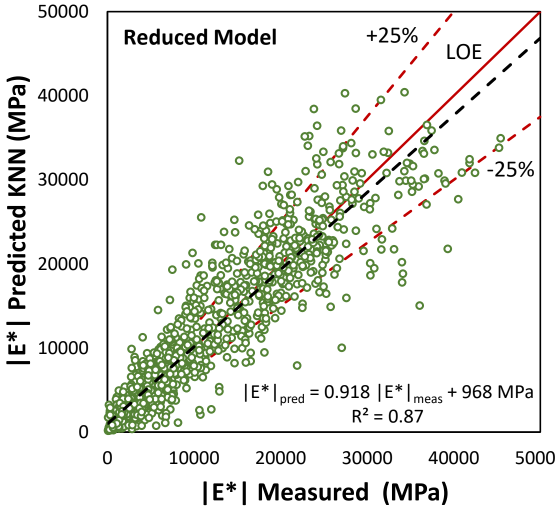


Fig. 6: Comparison of predicted and measured |E*| values for the Reduced Model

For both models, |E*| data points are located along the line of equality in a relatively narrow band and distributed on both sides of the LOE without a remarkable bias. Phase angle data points are also located along the LOE and tightly clustered around it.

Figure 10 shows the Normalized Residuals for the |E*| values, while Figure 11 presents the Normalized Residuals for the ϕ values obtained with the Reduced Model. Figures 12 and 13 show the corresponding results for the Complete Model.

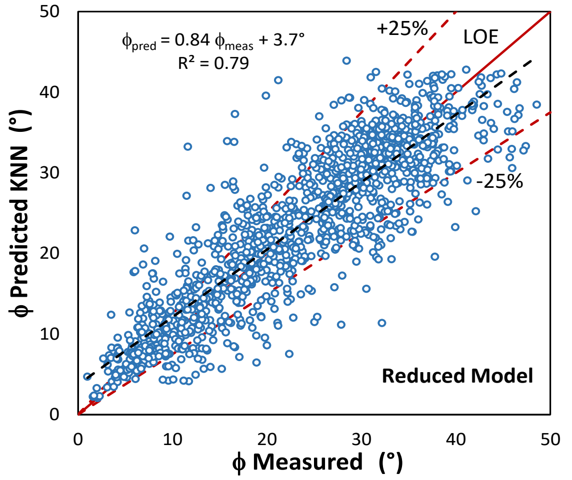


Fig. 7: Comparison of predicted and measured ϕ values for the Reduced Model

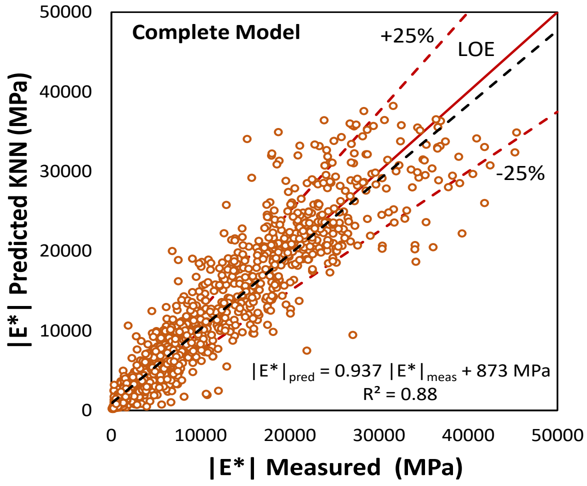


Fig. 8: Comparison of predicted and measured |E*| values for the Complete Model

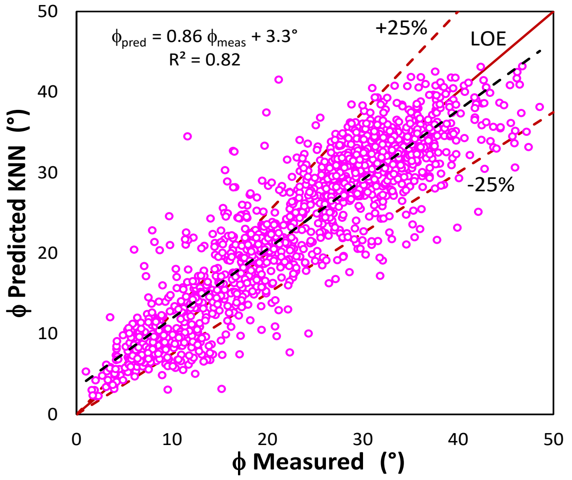


Fig. 9: Comparison of predicted and measured ϕ values for the Complete Model

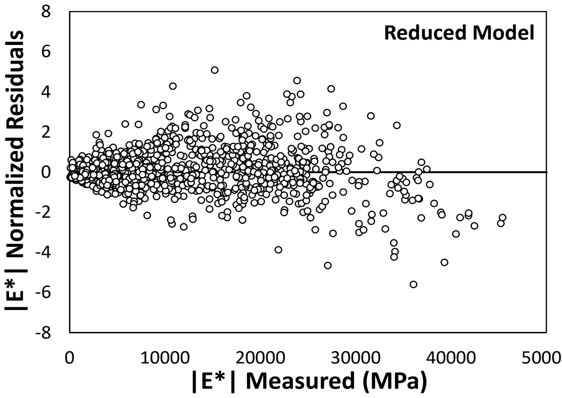


Fig. 10: Normalized Residual for the |E*| values and the Reduced Model

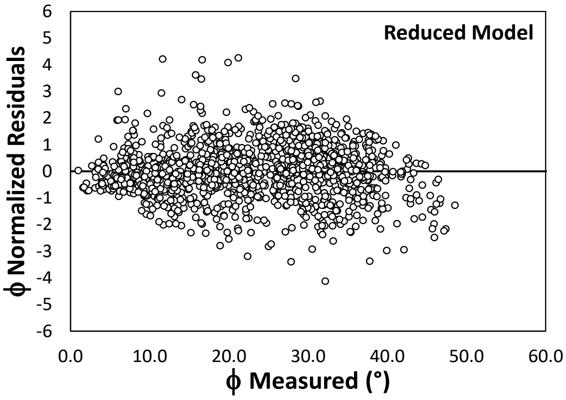


Fig. 11: Normalized Residual for the ϕ values and the Reduced Model

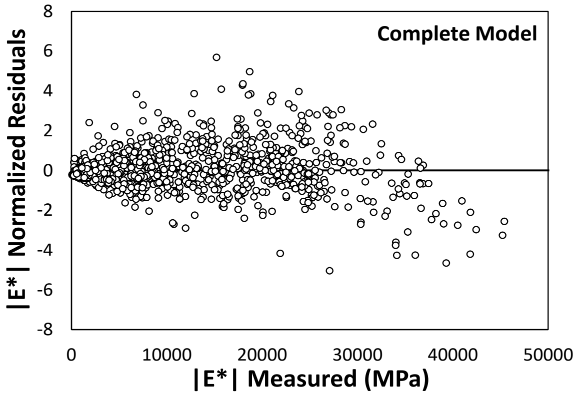


Fig. 12: Normalized Residual for the |E*| values and the Complete Model

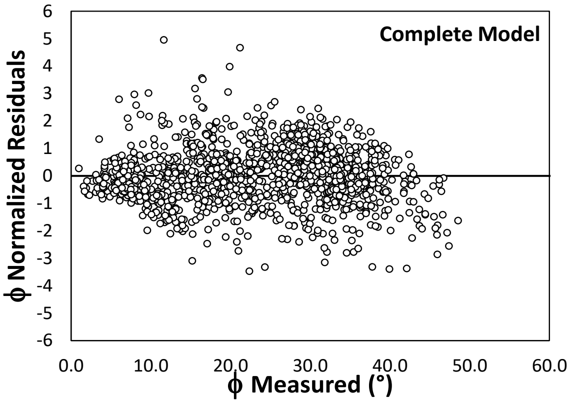


Fig. 13: Normalized Residual for the ϕ values and the Complete Model

In the four cases, no significant bias was detected. These observations indicate that the predicted values are good matches with the measured values. The relationship between predicted and measured values was investigated through the slope and intercept of the regression line.

For the Reduced Model and |E*| values, Figure 6 shows a low intercept (968 MPa) and a slope close to 1 (0.918); for the ϕ values, Figure 7 shows a low intercept (3.7º) and a slope also close to 1 (0.84).

For the Complete Model and |E*| values, Figure 8 shows also a low intercept (873MPa) and a slope close to 1 (0.937); for the ϕ values, Figure 9 shows a low intercept (3.3º) and a slope equal to 0.86. Both models could be rated as very good if they are compared with similar observations reported by other researchers [54, 55].

Although VMA and VFA showed very poor correlation with |E*| and ϕ in the Pearson analysis, including these volumetric properties improved the predictions, resulting in slopes closer to one and smaller intercept values.

In Figures 6 to 9, two additional lines were added, showing a scatter range of ±25% around the LOE. For both models, more than 86% of the predicted |E*| values fall within this range, while over 70% of the phase angle values differ by less than ±25%.

Table 5 shows a comparison between the predictive performance achieved by both KNN models, expressed in terms of error metrics (MAE, MAPE, MSE, RMSE), correlation coefficient (R^2), the relationship between the standard error of predicted values and the standard deviation of measured values (Se/Sy) and the overall performance evaluation according to the model evaluation criteria showed in Table 4.

The comparisons of the error metrics for the |E*| and ϕ predictions show a slight improvement for the Complete Model compared to the Reduced Model. The same observation applies to the correlation coefficient R^2 and the relationship Se/Sy. The resulting overall evaluation varies between Good and Excellent for the prediction of the |E*| and both models with R^2 values greater than 0.87 and Se/Sy values smaller than 0.35. For the phase angle ϕ, the evaluation results Good with R^2 values greater than 0.79 and Se/Sy values smaller than 0.44 for both models.

Comparing the two KNN models developed in this paper, the Complete Model with VMA and VFA as complementary inputs performs slightly better than the Reduced Model with respect of all six-performance metrics.

Table 5: Performance evaluation of the KNN algorithm

	Dynamic Modulus |E*|		**Phase Angle ϕ**	
	Reduced Model	**Complete Model**	**Reduced Model**	**Complete Model**
MAE	2138	2066	3.7	3.3
MAPE	0.37	0.35	0.20	0.19
MSE	12241410	11547556	25	21
RMSE	3499	3398	5.0	4.5
R^2	0.87	0.88	0.79	0.82
Se	3404	3333	4.7	4.3
Sy	9722	9722	10.8	10.8
Se/Sy	0.35	0.34	0.44	0.40
Evaluation	Good Excellent	Good Excellent	Good Good	Good Good

In addition, it should be emphasized that both ML models have been trained to simultaneously predict the Dynamic Modulus |E*| and also the Phase Angle ϕ. Thereby, they are able to provide a more detailed insight about the fundamental viscoelastic properties of the asphalt mixtures.

5. Conclusions

The present study outlines a detailed methodology in order to develop an innovative ML algorithm, namely K Nearest Neighbors (KNN) for the simultaneous prediction of both the Dynamic Modulus |E*| and the Phase Angle ϕ of asphalt mixtures. The major advantage of the KNN algorithm over other conventional predicting procedures is its ability to formulate the predictions with no presumptions, simplifications o empirical considerations about the relationship among their different influencing factors.

A unique database was compiled from four different datasets reported in the literature combining the main volumetric characteristics of the mixtures, binder properties and testing conditions. The variables required as input by the predictive models can be easily determined during a preliminary mix design procedure or derived from it.

The first KNN model, named the Reduced Model, uses log Visc, Vb, Va, T and f as inputs while the second model, called as Complete Model, included the same plus VMA and the VFA as complementary inputs. Both models were implemented in an Excel spreadsheet and the full database was divided into a training dataset with 75% of the data points and a training dataset with the remaining 25% of the data points. A K value equal to 5 was adopted producing the minimum NRMSE.

The performance achieved by both models ranges between Good and Excellent with the Complete Model performing slightly better according to error metrics and the goodness-of-fit statistical parameters.

With regard to |E*|, R^2 and Se/Sy resulted equal to 0.88 and 0.34 for the Complete Model. Similarly, with regard to ϕ, the same performance metrics resulted equal to 0.82 and 0.40, respectively.

Although Pearson's analysis shows a low correlation between the volumetric parameters VMA and VFA and the viscoelastic properties of asphalt mixtures, their inclusion has improved the accuracy of the |E*| and ϕ predictions. These volumetric parameters are routinely calculated during the formulation of asphalt mixtures, and their determination does not require additional effort. This supports the empirical evidence of the significant importance of these parameters on the mechanical behavior of asphalt mixtures.

Finally, both developed KNN models are capable of simultaneously providing effective estimates of |E*| and ϕ, avoiding the need for additional expensive and time-consuming laboratory tests. However, the need to generalize the database and the relative computational cost of the implemented models must be acknowledged.

The outlined methodology could be adopted to analyze larger and more varied datasets in order to obtain an increasingly powerful and high-performance tool capable of making accurate predictions to be applied in mechanistic-empirical pavement design procedures.

Acknowledgements

This research received funding from the National University of Rosario as a qualified Research Project.

Declaration of Competing Interest

The authors declare that they have no known competing interests or personal relationships that could have influenced the work reported in this paper.

Data availability

The database and the spreadsheet used in this study are available upon request from the corresponding author.

References

[1] Andrei, D., Mirza, W. and Witczak, M.W. (1999). Development of a Revised Predictive Model for the Dynamic (Complex) Modulus of Asphalt Mixtures. NCHRP Report 1-37A.

[2] Bari, J. and Witczak, M.W. (2006). Development of a New Revised Version of the Witczak E* Predictive Model for Hot Mix Asphalt Mixtures. Journal Association of Asphalt Paving Technologists: 75, 381-424.

[3] Christensen, D. W, Pellinen, T., Bonaquist, R.F. (2003). Hirsch Model for Estimating the Modulus of Asphalt Concrete. Journal Association of Asphalt Paving Technologists: 72, 97-121.

[4] Zhang, C. and Shen, S. (2017). Modification of the Hirsch dynamic modulus prediction model for asphalt mixtures. Journal of Materials in Civil Engineering, 29(12): 04017241. https://doi.org/10.1061/(ASCE)MT.1943-5533.0002099

[5] Al-Khateeb G., Shenoy A., Gibson N., and Harman T. (2006). A New Simplistic Model for Dynamic Modulus Predictions of Asphalt Paving Mixtures. Journal of the Association of Asphalt Paving Technologists: 75, 1254–1293.

[6] Picado-Santos, L., Capitao, S. D. and Pais, J. C. (2003). Stiffness modulus and phase angle prediction models for high modulus bituminous mixtures. International Journal of Pavements: Volume 2, issue 3.

[7] Biligiri, K. P., Kaloush, K., and Uzan, J. (2010). Evaluation of asphalt mixtures' viscoelastic properties using phase angle relationships. International Journal of Pavement Engineering, 11(2), 143-152. https://doi.org/10.1080/10298430903033354

[8] Naik, A. K. and Biligiri, K. P. (2014). Predictive Models to Estimate Phase Angle of Asphalt Mixtures. Journal of Materials in Civil Engineering: Volume 27, Issue 8. https://doi.org/10.1061/(ASCE)MT.1943-5533.0001197

[9] Yang, X., and You, Z. (2015). New Predictive Equations for Dynamic Modulus and Phase Angle Using a Nonlinear Least-Squares Regression Model. Journal of Materials in Civil Engineering: 27(3). https://doi.org/10.1061/(ASCE)MT.1943-5533.0001070

[10] Venudharan, V. and Biligiri, K. P. (2015). Estimation of phase angles of asphalt mixtures using resilient modulus test. Construction and Building Materials: Volume 82, 274-286. https://doi.org/10.1016/j.conbuildmat.2015.02.061

[11] Justo-Silva, R., Ferreira, A. and Flintsch, G. (2021). Review on Machine Learning Techniques for Developing Pavement Performance Prediction Models. Sustainability: 13, 5248. https://doi.org/10.3390/su13095248

[12] Chaabene, W. B., Flah, M. and Nehdi, M. L. (2020). Machine learning prediction of mechanical properties of concrete: Critical review. Construction and Building Materials: 260(7), 1-18. https://doi.org/10.1016/j.conbuildmat.2020.119889

[13] Khambra, G. and Shukla, P. (2021). Novel machine learning applications on fly ash based concrete: An overview. Materials Today: Proceedings: 80(6245). https://doi.org/10.1016/j.matpr.2021.07.262

[14] Beskopylny, A.N., Stelmakh, S.A., Shcherban, E.M., Mailyan, L.R., Meskhi, B., Razveeva, I., Chernilnik, A. and Beskopylny, N. (2022). Concrete Strength Prediction Using Machine Learning Methods: CatBoost, k-Nearest Neighbors, Support Vector Regression. Applied Sciences. 12, 10864. https://doi.org/10.3390/app122110864

[15] Zhang, D., Hang, J., Zhang, H, Huang, C., Qi, C. and Chang, E. (2020). Machine learning-based prediction of soil compression modulus with application of 1D settlement. Journal of Zhejiang University-Science A (Applied Physics &

Engineering): 21(6):430-444. https://doi.org/10.1631/jzus.A1900515

[16] Choi, H.J., Kim, S., Kim, Y. and Won, J. (2022). Predicting Frost Depth of Soils in South Korea Using Machine Learning Techniques. Sustainability: 14, 9767. https://doi.org/10.3390/su14159767

[17] Bajic, M., Pour, S. M., Skar, A., Pettinari, M., Levenberg, E. and Alstrøm, T. S. (2021). Road Roughness Estimation Using Machine Learning. arXiv preprint arXiv:2107.01199.. https://doi.org/10.48550/arXiv.2107.01199

[18] Suleymanov, A., Tuktarova, I., Belan, L. et al. (2023). Spatial prediction of soil properties using random forest, k-nearest neighbors and cubist approaches in the foothills of the Ural Mountains, Russia. Model. Earth Syst. Environ. 9, 3461–3471. https://doi.org/10.1007/s40808-023-01723-4

[19] Huang, Y., Molavi Nojumi, M., Ansari, S., Hashemian, L., & Bayat, A. (2023). Evaluating the use of machine learning for moisture content prediction in base and subgrade layers. Road Materials and Pavement Design, 24(12), 2910–2928. https://doi.org/10.1080/14680629.2023.2182135

[20] Ghorbani, B., Yaghoubi, E., Wasantha, P. L. P., van Staden, R., Guerrieri, M., & Fragomeni, S. (2023). Machine learning-based prediction of resilient modulus for blends of tire-derived aggregates and demolition wastes. Road Materials and Pavement Design, 25(4), 694–715. https://doi.org/10.1080/14680629.2023.2222176

[21] Atakan, M., & Yıldız, K. (2023). Prediction of Marshall design parameters of asphalt mixtures via machine learning algorithms based on literature data. Road Materials and Pavement Design, 25(3), 454–473. https://doi.org/10.1080/14680629.2023.2213774

[22] Tangga, A.A., Mufargi, H.A.L., Milad, A. et al. (2024). Utilising machine learning algorithms to predict the Marshall characteristics of asphalt pavement layers. Innov. Infrastruct. Solut. 9, 381. https://doi.org/10.1007/s41062-024-01698-w

[23] Upadhya, A., Thakur, M.S. & Sihag, P. (2024). Predicting Marshall Stability of Carbon Fiber-Reinforced Asphalt Concrete Using Machine Learning Techniques. Int. J. Pavement Res. Technol. 17, 102–122. https://doi.org/10.1007/s42947-022-00223-5

[24] Alnaqbi, A., Zeiada, W. & Al-Khateeb, G.G. (2024). Machine learning modeling of pavement performance and IRI prediction in flexible pavement. Innov. Infrastruct. Solut. 9, 385. https://doi.org/10.1007/s41062-024-01688-y

[25] Gul, M. A., Islam, M. K., Awan, H. H., Sohail, M., Al Fuhaid, A. F., Arifuzzaman, M. and Qureshi, H. J. (2022). Prediction of Marshall Stability and Marshall Flow of Asphalt Pavements Using Supervised Machine Learning Algorithms. Symmetry: 14, 2324. https://doi.org/10.3390/sym14112324

[26] Uwanuakwa, I.D., Busari, A., Ali, S.I.A. et al. (2022). Comparing Machine Learning Models with Witczak NCHRP 1-40D Model for Hot-Mix Asphalt Dynamic Modulus Prediction. Arab J Sci Eng 47, 13579–13591. https://doi.org/10.1007/s13369-022-06935-x

[27] Martínez, F. and Angelone, S. (2010). The Estimation of the Dynamic Modulus of Asphalt Mixtures Using Artificial Neural Networks. 11th International Conference on Asphalt Pavements: Volume 1, 354-363.

[28] Behnood, A. and Golafshan, E. M. (2021). Predicting the dynamic modulus of asphalt mixture using machine learning techniques: An application of multi biogeography-based programming. Construction and Building Materials, Volume 266, Part A. https://doi.org/10.1016/j.conbuildmat.2020.120983

[29] Leiva-Villacorta, F. and Vargas-Nordcbeck, A. 2019. Neural network based model to estimate dynamic modulus E* for mixtures in Costa Rica. J. Soft Comput. Civ. Eng.: 3(2):01–15. https://doi.org/10.22115/scce.2019.188006.1110

[30] Useche-Castelblanco, J.S., Reyes-Ortiz, O. J. and Alvarez, A. E. (2023). Application of machine learning models for prediction of rheological properties of wax-modified asphalt binders. Construction and Building Materials: Volume 395, 132352. https://doi.org/10.1016/j.conbuildmat.2023.132352

[31] Rahman, S., Bhasin A. and Smit, A. (2021). Exploring the use of machine learning to predict metrics related to asphalt mixture performance. Construction and Building Materials: Volume 295, 123585. https://doi.org/10.1016/j.conbuildmat.2021.123585

[32] Botella, R., Lo Presti, D., Vasconcelos, K. et al. (2022). Machine learning techniques to estimate the degree of binder activity of reclaimed asphalt pavement. Materials and Structures: 55, 112. https://doi.org/10.1617/s11527-022-01933-9

[33] Majidifard, H., Jahangiri, B., Buttlar, W. G. and Alavi, A. H. (2019). New Machine Learning-based Prediction Models for Fracture Energy of Asphalt Mixtures. Measurement: 135, 438–451. https://doi.org/10.1016/j.measurement.2018.11.081

[34] Shu, X., and Huang, B. (2009). Predicting Dynamic Modulus of Asphalt Mixtures with Differential Method. Road Materials and Pavement Design, 10(2), 337–359. https://doi.org/10.1080/14680629.2009.9690198

[35] Rondinella, F., Daneluz, F., Baldo, N. and Hofko, B. (2023). Improved predictions of asphalt concretes' dynamic modulus and phase angle using decision-tree based categorical boosting model. Construction and Building Materials: 400 (2023) 132709. https://doi.org/10.1016/j.conbuildmat.2023.132709

[36] Rondinella, F., Daneluz, F., Hofko, B., Baldo, N. (2024). A Machine Learning Approach for the Simultaneous Prediction of Dynamic Modulus and Phase Angle of Asphalt Concrete Mixtures. Advanced Research in Technologies, Information, Innovation and Sustainability. ARTIIS 2023. Communications in Computer and Information Science, vol 1935. Springer, Cham. . https://doi.org/10.1007/978-3-031-48858-0_40

[37] Ibrahim, A., Osman, M. K. Yusof, N.A., Ahmad, K., Harun, N. H. and Raof. R.A. (2019). Characterization of cracking in pavement distress using image processing techniques and k-Nearest neighbour. Indonesian Journal of Electrical Engineering and Computer Science. Vol. 14, No. 2, pp. 810-818. http://doi.org/10.11591/ijeecs.v14.i2.pp810-818

[38] Ghouchan Nezhad Noor Nia, R.; Jalali, M.; Houshmand, M. A. (2022). Graph-Based k-Nearest Neighbor (KNN) Approach for Predicting Phases in High-Entropy Alloys. Applied Sciences, 12, 8021. https://doi.org/10.3390/app12168021

[39] Aksoy, A., Iskender, E. and Kahraman, H. T. (2012). Application of the intuitive KNN Estimator for prediction of the Marshall Test (ASTM D1559) results for asphalt mixtures. Construction and Building Materials, 34, pp. 561–569. https://doi.org/10.1016/j.conbuildmat.2012.02.091

[40] Al-Dosary, N. M. N, Al-Hamed, S. A. and Aboukarima, A.M. (2019). K-Nearest Neighbors method for prediction of fuel consumption in tractor-chisel plow systems. Eng. Agric. (Online): 39 (6), 729–736. https://doi.org/10.1590/1809-4430-Eng.Agric.v39n6p729-736/2019

[41] Wang, Y. and Wang, Z. (2005). Text categorization rule extraction based on fuzzy decision tree. Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Vol 4, pp. 2122-2127, China. https://doi.org/10.1109/ICMLC.2005.1527296

[42] Ahmed, MS., N'diaye, M., Attouch, M.K. et al. (2023). K-nearest neighbors prediction and classification for spatial data. J. Spat. Econometrics. 4, 12. https://doi.org/10.1007/s43071-023-00041-2

[43] Ikeagwuani, C.C., Nweke, C.C. & Onah, H.N. (2023) Prediction of resilient modulus of fine-grained soil for pavement design using KNN, MARS, and random forest techniques. Arabian Journal of Geosciences. 16, 388. https://doi.org/10.1007/s12517-023-11469-z

[44] Weiss, S.M., and Kulikowski, C.A. (1991). Computer systems that learn: classification and prediction methods from statistics, neural nets, machine learning, and expert systems. Morgan Kaufmann Publishing, San Mateo, San Mateo, CA. United States. ISBN: 978-1-55860-065-2

[45] Dhar, V., and Stein, R. (1997). Intelligent Decision Support Methods: The Science of Knowledge Work. Prentice-Hall, Upper Saddle River, NJ. ISBN-13978-0135199350

[46] Witczak, M. (2005). NCHRP Project No. 9-19, Superpave Support and Performance Models Management, Database Dynamic Modulus (E*) Test and Master Curves (AC Mixture Simple Performance Test). Final Report.

[47] Witczak, M. W., and Fonseca, O. A. (1996). Revised Predictive Model for Dynamic (Complex) Modulus of Asphalt Mixtures. Transportation Research Record: 1540 (1), 15–23. https://doi.org/10.1177/0361198196154000010

[48] Fintsch, G., Loulizi, A., Diefenderfer, S. D., Galal, K.A. and Diefenderfer, B. K. (2007). Asphalt Materials Characterization in Support of Implementation of the Proposed Mechanistic-Empirical Pavement Design Guide. Report No.: VTRC 07-CR10, Virginia Tech Transportation Institute.

[49] Birgisson, B., Roque, R., Kim, J. and Pham, L. J. (2004). The use of complex modulus to characterize the performance of asphalt mixtures and pavements in Florida. UF Project No.: 4910-4504-784-12, Department of Civil and Coastal Engineering, College of Engineering, University of Florida.

[50] ASTM International. (2009). ASTM D2493. Standard Viscosity–Temperature Chart for Asphalts. American Society for Testing and Materials, Pennsylvania, USA.

[51] Imandoust, S. B. and Bolandraftar, M. (2013). Application of knearest neighbor (KNN) approach for predicting economic events: theoretical background. International Journal of Engineering Research and Applications: 3(5).605-610.

[52] Seman, P., Shoop, S., McGrath and Rollings, R. (2006). Soil strength prediction with K-nearest neighbour method. Proceedings of the 59th Canadian Geotechnical Conference. Canadian Geotechnical Society.

[53] Singh, D., Zaman, M. and Commuri, S. 2011. Evaluation of predictive models for estimating dynamic modulus of Hot-Mix Asphalt in Oklahoma. Transportation Research Record: 2210, 57-72. https://doi.org/10.3141/2210-07

[54] Tran, N. and Hall, K. (2005). Evaluating the Predictive Equation in Determining Dynamic Moduli of Typical Asphalt Mixtures Used in Arkansas. Journal of the Association of Asphalt Paving Technologists, 74, 1–17.

[55] Abdo, A., Bayomy, F., Nielsen, R., Weaver, T. and Jung, S. (2009). Prediction of the Dynamic Modulus of Superpave Mixes. Proc., 8th International Conference on the Bearing Capacity of Roads, Railways and Airfields. 305–314.