Developing a Robust Modeling Approach for Pavement Performance Prediction and Optimization

Parnian Ghasemi[a], Mohamad Aslani[b], Derrick K. Rollins[c], R. Chris Williams[a]

[a]*Civil, Construction, and Environmental Engineering Department, Iowa State University, Ames, Iowa*

[b] *Department of Mathematics, Florida State University, Tallahassee, Florida*

[c]*Chemical and Biological Engineering Department, Iowa State University, Ames, Iowa*

ABSTRACT: *In pavement technology, performance models are mathematical expressions that relate pavement condition, surface distresses and structural properties as response variables to a set of predictors including material properties, traffic loading, environmental factors, etc. In the existence of numerous important predictors and their interrelationships, developing a predictive model for pavement performance is not a trivial task. In this study, a machine learning-based framework is developed for predicting pavement performance. The framework starts with a preparation step of data pre-processing and data wrangling. After removing outliers, the framework will conduct principal component analysis (PCA) to firstly reduce the dimensionality of the problem and secondly eliminate pairwise correlation between the inputs by producing orthogonal pseudo-inputs. These pseudo-inputs are used to develop two predictive models using multivariate regression analysis and artificial neural networks (ANN). In empirical predictive models, mapping input space to response space can be threatened by extrapolation. However, it is often disregarded by design engineers. In this study to confront extrapolation, a method is implemented to determine a hyperspace based on the inputs. The hyperspace determines where the predictive model is valid up to given thresholds and is then added as a constraint to the modeling problem. Two of the performance-related characteristics of asphalt mixtures, including rut resistance and dynamic modulus, are considered to examine the robustness of the proposed approach. The developed predictive models are then compared to conventional models for each case and indicated superior performance (r_{fit} of 0.97 and 0.99 for rutting and dynamic modulus, respectively). A global variable importance analysis is also conducted to obtain the most effective variable in each case. Percent air voids and binder shear properties appeared to be the most effective variables in predicting rutting and dynamic modulus, respectively. To indicate an application of the developed framework in asphalt pavement design, for each of these two cases a design-related optimization problem is defined and solved using a mean-variance mapping optimization (MVMO) algorithm. The obtained optimal design parameters are within the acceptable range of current asphalt pavement design specifications and thus can be used as an appropriate starting point in a design procedure.*

KEYWORDS*: pavement performance modeling, principal component analysis, artificial neural networks, multivariate regression analysis, optimization, variable importance analysis.*

This paper was presented by title only.

1.0 Introduction

Maintenance, rehabilitation, and reconstruction of the highway system are the major expenses in a state general expenditure. Therefore, seeking to develop an accurate and efficient performance model to predict the remaining service life of a pavement and to provide its rehabilitation or reconstruction requirements punctually is beneficial. Relating pavement condition, surface distresses, and structural properties to a set of predictors including material properties, traffic loading, environmental factors, etc. via mathematical expressions is called performance modeling (Morovatdar et al., 2019; Jalali et al., 2019; Hosseini et al., 2020). According to the American Association of State Highway and Transportation Officials (AASHTO), pavement performance is the pavement ability to serve traffic over time sufficiently. To measure and predict pavement performance, a reproducible, authoritative, and field calibrated condition evaluating system is required. Several researchers have attempted to develop pavement performance predictive models but almost all of the performance models are site specific and also restricted to the materials used in the AASHO road test.

One of the performance-related properties of asphalt pavement is its resistance to rutting. Rutting or permanent deformation often happens under the wheel path and appears as a depression worn into a pavement with uplift occurring along the sides (Ghasemi et al., 2018; Notani et al., 2019). To analyze asphalt mixture rut susceptibility, performance testing along with mechanistic-empirical regression-based modeling appears to be a common approach (Bashin et al., 2012). To simulate rutting in the laboratory a rut resistance index called flow number (FN) is defined. In a repeated loading and unloading test FN is the point at which the strain rate starts to increase with loading. According to AASHTO TP 79-13, "Standard Method of Test for Determining the Dynamic Modulus and Flow Number for Asphalt Mixtures Using the Asphalt Mixture Performance Tester (AMPT)," this parameter has demonstrated a strong correlation with rutting that happens in asphalt pavement due to traffic. In asphalt pavement design procedures, the amount of rutting should generally be limited to 0.4 in. (10.16 mm) regarding the total deformation of a pavement structure.

It has been demonstrated that the amount of rutting is a function of binder viscosity, volumetric properties of the asphalt mixture, and testing temperature (Kaloush et al., 2003; Witczak et al., 2002). Kvasnak et al. (2007) proposed a list of the efficacious factors in rut susceptibility of asphalt mixture. The list includes nominal maximum aggregate size (NMAS); voids in mineral aggregate (VMA); percentage aggregate passing through sieve sizes No. 4, No. 16, No. 200; binder grade; binder viscosity; asphalt content; testing temperature; and the number of gyrations. Rodezno et al. (2010) selected 12 parameters (i.e., testing temperature; maximum shear stress; normal stress; binder viscosity; percentage aggregate passing through sieve sizes 3/4-inch, 3/8-inch, and No. 4; percentage air voids; effective binder content; binder content; VMA; and voids filled with asphalt (VFA)) to be important in estimating asphalt pavement rutting behavior. It was illustrated by Apeagyei (2011) that dynamic modulus test results at specific temperatures and loading

frequencies along with aggregate gradation appears to have strong correlation with FN test results. However, there are some discrepancies on the existence of correlation between rut susceptibility of asphalt mixture and its dynamic modulus value (Birgisson et al., 2004; Pellinen and Witczak, 2002; Timm et al., 2006).

Another widely used pavement performance characteristic is dynamic modulus, $|E^*|$, which defines the stress-strain relationship of asphalt mixtures under sinusoidal loading. Dynamic modulus represents the stiffness characteristic of an asphalt mixture, and it has a significant role in pavement design. Therefore, several researchers have attempted to predict asphalt mixture dynamic modulus as a function of material's components properties, loading rate, and temperature (Nobakht and Sakhaeifar, 2018; Peng et al., 2019; Shu and Huang, 2008).

There are several well-known predictive models for dynamic modulus amongst which some use regression analysis while newer ones use other techniques including genetic programming and artificial neural networks (Ziari et al., 2018). Witczak developed a predictive model using material component properties including binder viscosity. Andrei et al. (1999) modified the original Witczak model. The developed model has then been modified to use binder shear modulus instead of binder viscosity (Bari and Witczak, 2007). Christensen et al. (2003) created a predictive model based on the law of mixtures. Al-Khateeb et al. (2006) created a model from the law of mixtures to be used over a wide range of temperatures and loading frequencies. Sakhaeifar et al. (2017) created separate temperature-based models that can predict dynamic modulus over a wide range of temperature. The predictor variables of the aforementioned models are selected from the following list: cumulative percentage aggregate retained on sieve sizes 3/4-inch, 3/8-inch, No. 4, and percent aggregate passing the No. 200; VMA; VFA; percentage air voids; effective binder content; binder shear modulus ($|G^*|$); and binder phase angle (δ).

The predictor variables used in the conventional performance predictive models, e.g., rutting and dynamic modulus, are not admitted being independent sets of variables and therefore may not be suitable to be used in modeling. Pairwise correlated predictors in the data set can decrease the estimation accuracy of their effects on the response variable. Therefore, a data preparation step is useful to assure that the input variables are qualified to be used in model development (Rollins, 2015). The process of transforming raw data into another format with the intention of making it more appropriate and valuable for the main analysis is called data wrangling. One of the data wrangling techniques is reducing the dimension of the data especially when predictor variables are highly correlated. There are several dimensionality reduction techniques including exploratory factor analysis (EFA), variable clustering and principal component analysis (PCA). To reduce the dimension of the data, variable clustering and EFA tend to eliminate some of the predictors (Thompson, 2004) while PCA transforms the data into a new coordinate system and preserves more information of observed variables with less tendency for information loss. Therefore, PCA is preferred over other techniques for the particular application presented in this study.

PCA is a multivariate statistical procedure that reduces the size of a data set by transforming a large set of variables into a smaller set of orthogonal (i.e. zero correlation) pseudo-variables called principal components (PCs). The produced pseudo-inputs can be used as input variables in developing predictive models. They not only make the prediction analysis easier but also contain most of the information of the large set (Ghasemi et al., 2018).

Another issue within performance models is that since they are developed based on empirical data, they can be prone to extrapolation which is defined as the process of estimating beyond the original observation range. In case of extrapolation, predictive models are subject to major uncertainty and high risk of producing meaningless results. To prevent extrapolation from happening, a hyper-space containing all the data points can be found and added as a constraint to the desired modeling problem.

The focus of this study is on developing a machine learning-based framework to predict pavement performance using orthogonal pseudo-inputs obtained from principal component analysis. Unlike most of the conventional performance models, the proposed framework utilizes different data for model training and performance testing. An n-dimensional hyperspace is determined and added as a constraint to the modeling problem to guard against extrapolation. The authority of the proposed framework is illustrated by solving two separate problems of predicting rutting behavior and predicting dynamic modulus of asphalt mixtures. In order to find the most effective variable, global variable importance analysis is performed on the developed models. This research also claims to determine the optimal design and a successful approach to perform inverse design of asphalt mixture as some of the applications of the framework using a state-of-the-art evolutionary optimization algorithm.

2.0 Methodology

2.1. Data Preprocessing

Data preprocessing is a crucial task in every machine learning and data mining project. Irrelevant and redundant information, as well as correlated and unreliable predictor variables, can produce misleading results. Therefore, the representation and quality of data should be verified prior to running the main analysis.

A parsimonious model is a model that accomplishes a great explanatory predictive power with as few predictor variables as possible. The obstacles in creating such model can be the availability of numerous, highly correlated, and weakly related or unrelated predictor variables (Fodor, 2002). In a general model, the expectation function is given by:

$$\eta_i = f(x_i, \theta), \qquad i = 1,2,\dots,n \qquad [1]$$

Where vector xi is input values at the i^{th} sampling time, ηi is the expected value of the response, and $\theta = [\theta 1 \dots \theta q]^T$ is a vector containing unknown parameters of the model.

The corresponding element of the i^{th} row and j^{th} column in the Jacobian matrix, J, is defined by $\partial\eta_i/\partial\theta_j$. In this case if two columns (e.g., m and n) are orthogonal, their correlation coefficient should be zero. In other words, if the two columns are independent, the information used to estimate θ_m is separate from the information used to estimate θ_n. Such an approach will solidify the inputs-output relationship which leads to enhancing the accuracy of analysis. Correlated columns in the Jacobian Matrix are a consequence of pairwise correlated inputs. Therefore, to decrease standard parameter errors and increase accuracy of input-output mapping, the pairwise correlation of inputs could be eliminated by implementing PCA (Fodor, 2002).

2.2. Data Wrangling: Dimensionality Reduction using Principal Component Analysis (PCA)

Data wrangling is a main step in creating a machine learning model. During this step the data will be converted into a suitable format which can be used by any machine learning algorithm. During data wrangling step PCA can be implemented to remove correlated predictors, improve algorithm performance, and reduce overfitting.

PCA is a statistical technique often used to reduce the dimensionality of the data by selecting the most important features that capture maximum information about the data set. In other words, PCA is a technique of extracting important features (in the form of PCs) from a large set of available features in a data set. The features are selected based on variance that they cause in the response. During this orthogonal transformation, original inputs of the data set are converted to the principal components (PCs) which are linear combinations of the original inputs (Rollins et al., 2006). PCA works in a way that the variable that causes highest variance is the first PC, the variable responsible for the second highest variance is considered the second PC, and so on.

PCA can be implemented by either eigenvalue decomposition of a data covariance (or correlation) matrix or singular value decomposition (SVD), often after normalizing the data matrix (mean centering) for each procedure (Jolliffe, 2002). In the present study, the eigenvalue decomposition of the data correlation matrix is used. PCA reduces the dimension of a given data set, X, by representing the l original variables $(x1, \ldots, xl)$ as p new pseudo-variables (PCs), where $p < l$. For a given data set, X, this analysis is summarized in the following steps.

Standardize original data, X, by transforming it to Z using the following equations:

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x_{11}} & \boldsymbol{x_{12}} & \cdots & \boldsymbol{x_{1l}} \\ \boldsymbol{x_{21}} & \boldsymbol{x_{22}} & \cdots & \boldsymbol{x_{2l}} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x_{n1}} & \boldsymbol{x_{n2}} & \cdots & \boldsymbol{x_{nl}} \end{bmatrix}, \quad \boldsymbol{Z} = [\boldsymbol{z_1} \quad \boldsymbol{z_2} \quad \cdots \quad \boldsymbol{z_l}] = \begin{bmatrix} \frac{x_{11}-\overline{x_1}}{s_1} & \frac{x_{12}-\overline{x_2}}{s_2} & \cdots & \frac{x_{1l}-\overline{x_l}}{s_l} \\ \frac{x_{21}-\overline{x_1}}{s_1} & \frac{x_{22}-\overline{x_2}}{s_2} & \cdots & \frac{x_{2l}-\overline{x_l}}{s_l} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_{n1}-\overline{x_1}}{s_1} & \frac{x_{n2}-\overline{x_2}}{s_2} & \cdots & \frac{x_{nl}-\overline{x_l}}{s_l} \end{bmatrix} \quad [2]$$

Where, for $k = 1$ to n and $j = 1$ to l, x_{kj} is the k^{th} measurement for the j^{th} variable, x_k is sample mean for the k^{th} variable, and s_k is sample standard deviation for the k^{th} variable.

Determine the unit eigenvectors $e_1, \ldots, e_l$ of Z.

Determine the corresponding eigenvalues $\lambda_1, \ldots, \lambda_l$.

Rank the eigenvectors (in descending order) according to their eigenvalues.

Select the p PCs according to their eigenvalues.

Considering the number of eigenvalues (or components) versus their fraction of total represented variance is used to select the appropriate number of PCs. One might stop adding more PCs when little variance is gained by retaining additional eigenvalues. The selected PCs will be used in the main analysis.

2.3. Cross Validation

To assure that the results of a statistical analysis can generalize to an independent data set a model validation technique called cross validation should be implemented. This technique is mainly used to prevent overfitting in prediction problems, where a model is usually trained with a data set called training data and is tested against a first-seen data set called testing data (Refaeilzadeh et al., 2009). According to the size of an available database and the desired computational time, the most suitable cross validation technique should be selected to be used in the modeling.

2.4 Principal Component Regression (PCR)

Recalling from section 2.2, the selected PCs are used as new predictors in the modeling procedure. All possible regression structures should be considered and examined for mapping the predictors (pc's) to the response variable (pavement performance). To estimate the values of unknown coefficients of the model, the least squares criterion of minimizing the sum of squared residuals (SSE) is implemented. Finally, after eliminating the redundant terms the reduced model is developed and selected as the best fitted model. The developed model is called "Principal Component Regression (PCR)".

2.5 Principal Component Neural Network (PCNN)

A predictive model called "Principal Component Neural Network (PCNN)" is developed using artificial neural networks (ANN). ANNs consist of a collection of connected units or nodes called artificial neurons which learn to perform tasks by considering examples (Cheng and Titterington, 1994; Fathi et al., 2019; Majidifard et al., 2019). A three-layer feed-forward neural network, consisting of an input layer of n neurons, with n being the

number of principal components (pc's), a hidden layer of ten neurons, and an output layer of one neuron, which is the pavement performance, is developed using the MATLAB software. The number of hidden neurons is selected to balance between the cost function and computational time using a trial and error procedure.

To initiate the training process the input of each processing neuron, pc_i, is multiplied by a randomly assigned (and adaptable) connection weight w_{ij} and the weighted inputs are summed and added to a threshold value, b_0. The result crossed through a nonlinear transfer function (sigmoid in this study) resulted in the output of the first layer, v_i, which establishes the input for the next layer. During each iteration in the training process, the network adjusts its weights and biases to minimize the loss function which is the difference between the predicted and observed values of response variable. The iterative procedure continues until the convergence criterion is satisfied. The performance of the trained network is then validated against an unseen set of data (test data set). For network training efficiency the Bayesian Regularization algorithm is implemented.

The output vj from the j^{th} hidden node is given by:

$$v_j = f_1(pc_i, w_{ij}), \quad i = 1, \dots, m \quad and \quad j = 1, \dots, 10 \qquad [3]$$

and the single output $\hat{y}$ is:

$$\hat{y} = f_2\left(f_1(pc_i, w_{ij})\right) \qquad [4]$$

Then the expression of $\hat{y}$ as a function of PC becomes a complicated nonlinear regression function with the *j* sets of weights, as parameters. It is assumed that:

$$f_2(v_j, w_{Hj}) = b_0 + \sum_j v_j \, w_{Hj} \qquad [5]$$

and for each j,

$$f_1(pc_i, w_{ij}) = b_{Hj} + \sum_j pc_i \, w_{ij} \qquad [6]$$

so a general form of the feed forward neural network is described in Equation 7.

$$\hat{y} = f_2\{b_0 + \sum_{j=1}^{n}[w_{Hj}. f_1(b_{Hj} + \sum_{i=1}^{m} pc_i \, w_{ij})]\} \qquad [7]$$

where b_0 is bias at output layer; w_{Hj} is weight of connection between neuron j of the hidden layer and output layer neuron; b_{Hj} is bias at neuron *j* of the hidden layer (for $j = 1$ to 10); wij is weight of connection between input variable *i* (for $i = 1 \; to \; m$) and neuron *j* of the hidden layer; pci is pseudo input parameter *i*; $f_1(t)$ is transfer function of the hidden layer, and $f_2(t)$ is transfer function of the output layer.

For an arbitrary variable t, the transfer functions used in the network, $f_1(t)$ and $f_2(t)$, are defined in Equation 8.

$$f_k(t) = \frac{1}{1+e^{-t}} \quad \text{for} \quad k = 1,2 \qquad [8]$$

2.6 Effective Variable Space

A major shortcoming in most research articles where a predictive model is developed is the space (range of input variables) where the developed empirical model is valid. The answer to this question is important because the behavior of such an empirical model is arbitrary when the inputs are not inside the n-dimensional hyper ellipsoid covering the original data (in this case the training data set). A number of valid assumptions make the above conclusion possible. One of these assumptions is based on the normal distribution of the input variables and their joint distribution which is bi-variate normal (Devore, 2012). In this way, any slice of such a domain (at a constant density function) will result in a n-dimensional hyper ellipsoid. Thus, one should perform a test to determine if the desired data is inside this appropriate space. This concept is better visualized in Figure 1 (Neter et al., 1989).

The n-dimensional space can be located using a number of approaches (Todd and Yildirim, 2007). The problem of finding the n-dimensional ellipsoid that contains m-dimensional data has been a subject to a thorough classical computational complexity analysis (Sun and Freund, 2002) and is not the purpose of this article.

Ghasemi et al. (2019) showed that an iterative scheme where a modified dual optimization problem is solved numerically to find the details of the desired n-dimensional hyperspace (in this case an n-dimensional ellipsoid) is an effective solution to this problem. This approach is only dependent on an inverse matrix step which might become numerically cumbersome for large data sets.

Another approach to this problem is possible through the usage of interior-point algorithms as described in [Sun and Freund, (2002)]. To summarize the approach used in this paper, we start by defining the following optimization (minimization) problem:

$$\begin{aligned} &\text{minimize} && Volume\ (Ellip.) \\ &\text{with respect to} && v \in \mathbb{R}^n, A \in \mathbb{R}^n \\ &\text{subject to} && (M_i - v)^T A\ (M_i - v) \leq 1\ for\ i = 1, 2, \dots, p \end{aligned} \quad [9]$$

where p is the number of points given in n dimensions and M_i is the ith point. The unknowns are vector v and the matrix A. The above problem is reformulated in order to fit an interior point type algorithm as follows:

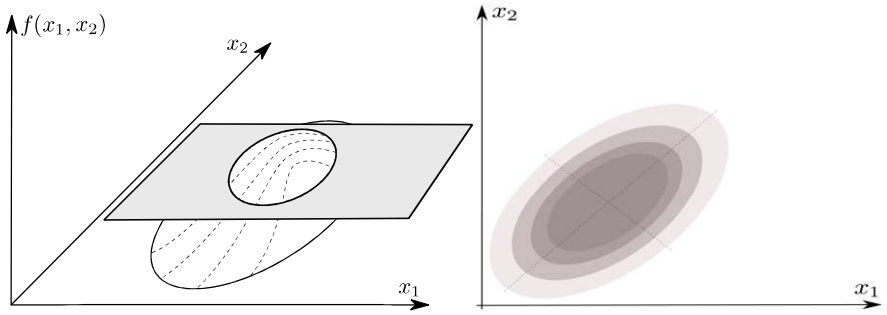

Figure 1. *Schematics of a bi-variate (normal) distribution (left) and their cross-sections (right) (Neter et al., 1989).*

$$\begin{aligned} &\text{minimize} && -\ln det \mathrm{M} \\ &\text{with respect to} && \omega_i, y_i \in \mathbb{R}^p \\ &\text{subject to} && \mathrm{Ma_i} - z - y_i = 0,\ \mathrm{i} = 1, \dots, \mathrm{p} \\ & && \omega_\mathrm{i} = 1, i = 1, \dots, p \end{aligned} \qquad [10]$$

To use an available interior-point method, one should reformulate the above problem to a barrier function type:

$$B = -ln\, det\, M - \theta \sum_i^m \mathrm{n}(\omega_i^2 - y_i^T y_i) \qquad [11]$$

An interior point method is used where at each step Schur-complement matrix of size $p^2(n+1)^2$ is factorized.

The above algorithm is used to determine the p-dimensional and d-dimensional hyper ellipsoid in the original and pseudo space of the data set, respectively.

A summary of the proposed methodology is presented in Figure 2. For every performance prediction problem, the framework starts with creating an empirical data base using field or laboratory-produced data. The framework continues with a data preprocessing and data wrangling step to make the data more appropriate and valuable for the main analysis which is developing the predictive models (PCR and PCNN). The framework confronts extrapolation with using the developed predictive models over their allowed variable space. The developed models can be used in further performance prediction, design and optimization problems, etc.

The rest of the paper focuses on using the developed framework to define and solve two separate performance related problems: (1) predicting rutting behavior and (2) predicting dynamic modulus of asphalt mixture. These two examples are presented, solved, and discussed in section 3 of this paper.

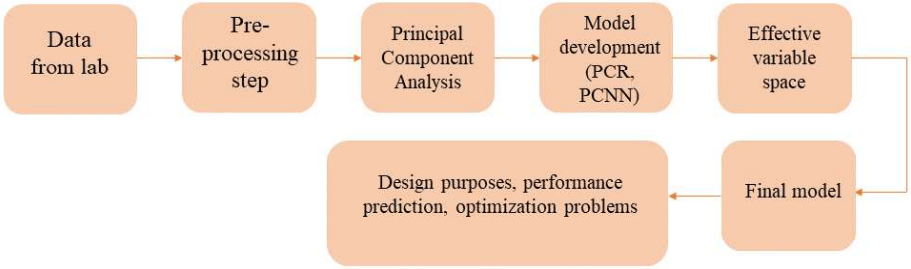

Figure 2. *A summary of the sequential tasks implemented to produce the machine learning based framework for predicting pavement performance.*

3.0 Results and Discussion

The problem of predicting rutting behavior, as one of the most important performance related properties of asphalt mixture, is defined and solved in this section.

3.1 Problem (1): Predicting Rutting Behavior

3.1.1 Material and Laboratory Testing

To create a data set for predicting rutting behavior of asphalt mixture, specimens were collected from different locations in the state of Wisconsin. Eighty-three specimens from 21 different mixtures were used to perform laboratory testing.

Asphalt mixtures were collected directly from the back of the delivery trucks at the plant site, and for each asphalt mixture the corresponding asphalt binder was sampled during mix plant production. Maximum theoretical specific gravity (G_{mm}) was measured in accordance with AASHTO T 209/ASTM D2041. The measured G_{mm} was used to obtain other volumetric properties of the asphalt mixtures. Specimens are compacted using a Superpave gyratory compactor to the following dimensions, 150 mm in diameter by 170 mm in height. Specimens were compacted to three different air voids including 4.0%, 7.0%, and 10.0%. Following AASHTO T 166/ASTM D2726, the bulk specific gravity values of specimens were determined.

To conduct the dynamic modulus test, a 100-mm diameter by 150-mm height cylindrical specimen was cored out of the laboratory compacted specimens. The specimens were trimmed and prepared for the dynamic modulus test. The specimens were tested at an effective test temperature of 36.6°C under repeated sinusoidal load with 25, 10, 1, and 0.1 Hz loading. The same specimens were used to perform flow number tests under repeated haversine load with the load being applied for a duration of 0.1 second and a dwell period of 0.9 second. The same effective temperature of 36.6°C was selected for the flow number test. No confining pressure was applied, and the axial stress was similar to the deviator stress (600 kPa). The accumulated strain at the FN was considered to be the response variable during the modeling procedure.

To study binder shear properties, the complex shear modulus test was performed using a dynamic shear rheometer (DSR) based on ASTM D7552-09 at 36.6°C and similar loading frequencies used in the dynamic modulus tests (25, 10, 1, and 0.1 Hz).

It is worth pointing out that 36.6°C is selected as the temperature that all the laboratory tests were performed at based on climate conditions in the Midwestern area of the United States. In other words, this temperature is considered as a reasonable temperature at which permanent deformation happens in this area, equivalent to a seasonal correction throughout the year. However, what makes the machine learning based models so special is that the model can be retrained and modified based on any test temperatures suiting any climate conditions. Once the data is fed into the framework it will learn the data pattern and the network will modify its weights and biases to fit the new data.

3.1.2 Step 1: Data Preprocessing

According to the literature (Kaloush et al., 2003; Kvasnak et al., 2007; Rodezno et al., 2010; Witczak et al., 2002), the rutting behavior of an asphalt mixture can be accurately estimated as a function of its component properties. The input variables are selected amongst those properties that have already proven to be important in predicting rutting behavior. However, their importance was also re-examined before selection by performing a multi-factor analysis of variance (ANOVA). The selected material component properties and their ranges that are measured and used in this section are presented in Table 1. These properties are selected based on the existing literature and used as the original input variables to predict accumulated strain value at the FN.

To study the quality of the input variables and their interrelationships, the correlation analysis is performed, and the cross-correlation matrix of the input variables is obtained and presented by Table 2. Within the matrix, there are 273 elements with absolute values greater than 0.1. This means that the corresponding variables are not independent. Besides, there are 41 elements with absolute values are greater than 0.5 (elements in bold text). This means that several of the input variables appear to have strong correlation. Therefore, to eliminate the existing pairwise correlation, PCA should be implemented.

Table 1. *Original input variables of problem (1): predicting rutting behavior.*

Variable	Identity	Values in the database			
		Min.	**Max.**	**Ave.**	**Std. Dev.**
x_1	**Binder %**	3.40	6.60	5.09	0.77
x_2	**G***	210800.54	1163559.91	612179.57	265903.65
x_3	**NMAS**	12.50	25.00	15.92	3.76
x_4	**Passing 3/4"**	81.30	100.00	98.55	4.09
x_5	**Passing 1/2"**	38.30	98.80	87.13	15.20
x_6	**Passing 3/8"**	34.10	89.90	76.34	15.10
x_7	**Passing #4**	26.20	72.50	56.25	13.74
x_8	**Passing #8**	17.50	54.00	42.25	10.51
x_9	**Passing #16**	14.20	47.40	32.18	8.74
x_{10}	**Passing #30**	9.60	39.10	23.20	7.00
x_{11}	**Passing #50**	5.70	18.60	12.02	3.18
x_{12}	**Passing #100**	3.70	9.80	6.19	1.42
x_{13}	**Passing #200**	2.80	8.50	4.32	1.12
x_{14}	**VMA**	10.32	21.00	16.45	2.50
x_{15}	**VFA**	46.45	91.72	65.19	9.06
x_{16}	**Va%**	1.02	9.83	5.87	2.09
x_{17}	**E***	395.70	2299.40	869.41	411.52

Table 2. *Pairwise correlation matrix for the input variables of problem (1): predicting rutting behavior.*

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}
x_1	1	0.33	**-0.73**	**0.60**	**0.71**	**0.63**	0.47	0.45	0.46	0.45	0.41	0.43	0.42	**0.61**	-0.18	0.35	-0.43
x_2	0.33	1	-0.11	0.08	-0.03	-0.23	-0.38	-0.32	-0.18	-0.06	0.01	0.23	0.35	0.02	-0.02	0.02	0.24
x_3	**-0.73**	-0.11	1	**-0.64**	**-0.76**	**-0.69**	**-0.51**	-0.41	-0.39	-0.32	-0.25	-0.30	-0.39	**-0.53**	0.13	-0.30	0.37
x_4	**0.60**	0.08	**-0.64**	1	**0.88**	**0.78**	**0.58**	**0.51**	0.42	0.32	0.25	0.23	0.24	0.46	-0.21	0.32	-0.45
x_5	**0.71**	-0.03	**-0.76**	**0.88**	1	**0.95**	**0.77**	**0.71**	**0.61**	0.49	0.43	0.39	0.36	**0.56**	-0.33	0.44	**-0.51**
x_6	**0.63**	-0.23	**-0.69**	**0.78**	**0.95**	1	**0.92**	**0.86**	**0.75**	**0.61**	0.49	0.30	0.16	**0.53**	-0.27	0.38	-0.49
x_7	0.47	-0.38	**-0.51**	**0.58**	**0.77**	**0.92**	1	**0.95**	**0.84**	**0.69**	0.49	0.11	-0.14	0.43	-0.16	0.26	-0.37
x_8	0.45	-0.32	-0.41	**0.51**	**0.71**	**0.86**	**0.95**	1	**0.95**	**0.83**	**0.59**	0.11	-0.15	0.42	-0.16	0.26	-0.41
x_9	0.46	-0.18	-0.39	0.42	**0.61**	**0.75**	**0.84**	**0.95**	1	**0.96**	**0.74**	0.18	-0.11	0.40	-0.16	0.25	-0.41
x_{10}	0.45	-0.06	-0.32	0.32	0.49	**0.61**	**0.69**	**0.83**	**0.96**	1	**0.84**	0.28	-0.06	0.35	-0.15	0.22	-0.37
x_{11}	0.41	0.01	-0.25	0.25	0.43	0.49	0.49	**0.59**	**0.74**	**0.84**	1	**0.59**	0.17	0.21	-0.20	0.20	-0.25
x_{12}	0.43	0.23	-0.30	0.23	0.39	0.30	0.11	0.11	0.18	0.28	**0.59**	1	**0.82**	0.24	-0.28	0.28	-0.15
x_{13}	0.42	0.35	-0.39	0.24	0.36	0.16	-0.14	-0.15	-0.11	-0.06	0.17	**0.82**	1	0.29	-0.30	0.32	-0.22
x_{14}	**0.61**	0.02	**-0.53**	0.46	**0.56**	**0.53**	0.43	0.42	0.40	0.35	0.21	0.24	0.29	1	**-0.62**	**0.83**	**-0.71**
x_{15}	-0.18	-0.02	0.13	-0.21	-0.33	-0.27	-0.16	-0.16	-0.16	-0.15	-0.20	-0.28	-0.30	**-0.62**	1	**-0.94**	0.52
x_{16}	0.35	0.02	-0.30	0.32	0.44	0.38	0.26	0.26	0.25	0.22	0.20	0.28	0.32	**0.83**	**-0.94**	1	**-0.63**
x_{17}	-0.43	0.24	0.37	-0.45	**-0.51**	-0.49	-0.37	-0.41	-0.41	-0.37	-0.25	-0.15	-0.22	**-0.71**	**0.52**	**-0.63**	1

3.1.2 Step 2: Data wrangling: Dimensionality reduction using PCA

For the pairwise correlation matrix, the eigenvalues and their corresponding percent variance were calculated and presented in Tables 3 which indicates the fraction of total variation in the data expressed by each eigenvalue. In Table 3 the eigenvalues were sorted in descending order meaning that the first eigenvalue represents the highest portion of the total variation, the second one has the second highest portion and so on. By selecting the first five PCs, 89.72% of the variation in the original data will be represented. Adding the sixth PC has an insignificant impact on the overall represented variation. Thus, the first five PCs are selected to be used as the pseudo-input variables.

Table 3. *Eigenvalues of the normalized matrix and the corresponding percent variance for problem (1): predicting rutting behavior.*

Component Number	Eigenvalue	Percent Variance	Cumulative Percent
1	7.98	46.94	46.94
2	2.88	16.92	63.86
3	1.95	11.46	75.31
4	1.54	9.05	84.36
5	0.91	5.36	89.73
6	0.58	3.42	93.15
7	0.42	2.48	95.62
8	0.26	1.53	97.15
9	0.20	1.17	98.31
10	0.15	0.89	99.21
11	0.06	0.37	99.58
12	0.04	0.22	99.80
13	0.02	0.10	99.90
14	0.01	0.05	99.95
15	0.01	0.03	99.97
16	0.00	0.02	99.99
17	0.00	0.01	100.00

PCs can be obtained using Equation 12:

$$pc_i = \sum_{j=1}^{17} \alpha_{ij} \mathrm{x_j} + \beta_i \quad [12]$$

Where $i = 1, \dots, 5$, the αij is the corresponding coefficients, the βi are constants, and the xj's are the original input variables. Equation 12 can be presented in matrix notation as in Equation 13.

$$\mathrm{p} = \mathrm{Mz} + \mathrm{n} \quad [13]$$

Where $\mathbf{p} = \begin{bmatrix} pc_1 \\ pc_2 \\ pc_3 \\ pc_4 \\ pc_5 \end{bmatrix}$, $\mathbf{n} = \begin{bmatrix} -1.72E+01 \\ 5.87E-01 \\ -8.21E+00 \\ 7.36E+00 \\ -4.85E-01 \end{bmatrix}$

$$M^T = \begin{bmatrix} 3.41E-01 & -2.04E-01 & 2.89E-01 & -1.94E-01 & 3.71E-01 \\ -7.60E-08 & -1.17E-06 & 1.31E-06 & 1.05E-07 & 2.47E-06 \\ -6.59E-02 & 3.22E-02 & -4.53E-02 & 9.10E-02 & -7.97E-03 \\ 6.34E-02 & -1.31E-02 & 1.86E-02 & -9.38E-02 & -7.13E-03 \\ 2.11E-02 & -2.33E-03 & 5.26E-03 & -1.63E-02 & -9.03E-03 \\ 2.17E-02 & 7.23E-03 & 1.55E-03 & -1.23E-02 & -1.13E-02 \\ 2.13E-02 & 2.01E-02 & -2.82E-03 & -7.27E-03 & -7.11E-03 \\ 2.79E-02 & 2.89E-02 & -2.98E-03 & 3.79E-03 & 1.41E-03 \\ 3.28E-02 & 3.22E-02 & 4.43E-03 & 2.35E-02 & 1.71E-02 \\ 3.73E-02 & 3.39E-02 & 1.53E-02 & 4.98E-02 & 3.14E-02 \\ 7.08E-02 & 3.30E-02 & 7.38E-02 & 1.54E-01 & -9.05E-03 \\ 1.08E-01 & -2.03E-01 & 2.33E-01 & 2.30E-01 & -2.72E-01 \\ 9.18E-02 & -4.06E-01 & 2.29E-01 & 6.69E-02 & -3.18E-01 \\ 9.98E-02 & -8.77E-02 & -1.12E-01 & -1.14E-02 & 8.85E-02 \\ -1.73E-02 & 3.21E-02 & 4.57E-02 & -2.69E-02 & -9.43E-04 \\ 9.75E-02 & -1.42E-01 & -1.95E-01 & 7.31E-02 & 4.70E-02 \\ -5.50E-04 & 2.73E-04 & 8.20E-04 & -8.92E-06 & 1.36E-04 \end{bmatrix}$$

The PCs obtained in this way will be used in the main analysis.

3.1.3. Step 3: Model Development

To validate the stability of the machine learning model and examine how well it would generalize to new data, a cross validation technique is implemented. In this case of having limited amount of data (sample size of 83), a k-fold cross validation technique leads to a less biased model compared to other methods, because it ensures that every observation from the original data set has the chance of appearing in training and test sets. K-fold cross validation technique works in a way that the given data set is randomly partitioned into k subsets. K-1 of these subsets are used to train the model and the remaining subset is used to test the model. The process is repeated k times and each subset can be used as test data exactly once.

Based on the size of rutting data set (83 data points), the data set is randomly partitioned into three folds. Results of the developed models are presented in this section, and their performance in predicting the response variable over the defined effective variable space is examined and discussed.

After examining all the possible regression structures, the second-order quadratic linear regression model fitting the measured response the best, is presented in Equation 14:

$$\hat{y} = c_0 + c_1 * pc_1 + c_2 * pc_2 + c_3 * pc_3 + c_4 * pc_4 + c_5 * pc_5 + c_6 * pc_1 * pc_2 + c_7 * pc_2 * pc_4 + c_8 * pc_1 * pc_3 + c_9 * pc_2 * pc_3 + c_{10} * pc_3 * pc_5 \quad [14]$$

Where $c_0 = 1.64 \times 10^4, c_1 = -8.89 \times 10^2, c_2 = -1.24 \times 10^3, c_3 = -1.24 \times 10^3, c_4 = -92.41, c_5 = 6.55 \times 10^2, c_6 = 1.58 \times 10^2, c_7 = -4.35 \times 10^2, c_8 = 2.35 \times 10^2, c_9 = 3.74 \times 10^2,$ and $c_{10} = -8.28 \times 10^2$.

Using Equation 7 the developed network's connection weights and biases are presented by the following matrices.

$$W^T = \begin{bmatrix} -0.45 & 1.70 & -0.81 & -0.85 & -1.30 \\ 0.01 & -1.46 & 0.28 & 0.11 & -1.93 \\ 1.32 & -1.39 & -0.03 & -0.22 & 0.90 \\ -0.09 & 0.18 & -0.14 & -0.32 & -0.15 \\ 0.09 & -0.18 & 0.14 & 0.32 & 0.15 \\ -0.61 & 0.26 & -1.19 & -0.81 & 0.12 \\ 0.09 & -0.18 & 0.14 & 0.32 & 0.15 \\ 0.44 & -1.50 & -0.41 & 1.05 & -0.64 \\ 0.28 & -0.63 & 0.83 & 0.74 & 1.16 \\ -0.00 & 0.17 & 0.00 & -1.38 & -0.29 \end{bmatrix}$$

$$W_H = \begin{bmatrix} 0.67 \\ 0.35 \\ -1.25 \\ -0.29 \\ 0.29 \\ -1.71 \\ 0.29 \\ -0.26 \\ 0.97 \\ 0.99 \end{bmatrix} \quad B_H = \begin{bmatrix} -0.08 \\ -1.21 \\ -1.74 \\ -1.53 \\ 0.57 \\ -0.57 \\ -1.32 \\ -0.57 \\ 1.46 \\ -1.75 \end{bmatrix} \quad , B_0 = [1.33]$$

Using several statistics, the performance results of the PCR and PCNN models are discussed and presented in Table 4. The first statistic is the "average difference (AD)" defined as:

$$\text{AD} = \frac{1}{\text{n}\sum_{\text{i}=1}^{\text{n}}(\text{y}_\text{i} - \hat{y}_\text{i})} \qquad [15]$$

AD is an estimate of systematic model bias, n is the number of input vectors, y_i is the i^{th} measured response value, and $\hat{y}$ is the i^{th} fitted response value. The second statistical component is the "average absolute difference (AAD)" which shows the average closeness of the fitted value to the measured response value. AAD is defined as:

$$AAD = \frac{1}{n\sum_{i=1}^{n}|y_i - \hat{y}_i|} \qquad [16]$$

The third statistical component, $rfit$, is the correlation of y and $\hat{y}$ defined as:

$$r_{fit} = \frac{n\sum_{i=1}^{n} y_i\hat{y}_i - (\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} \hat{y}_i)}{\sqrt{n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}\sqrt{n\sum_{i=1}^{n} \hat{y}_i^{\,2} - (\sum_{i=1}^{n} \hat{y}_i)^2}} \qquad [17]$$

Higher r_{fit} (with maximum value of 1) indicates better fit. The last statistical component is R-squared (R^2), also known as the coefficient of determination. R^2 is the portion of the variance in the dependent variable that is predictable from the independent variables and can be explicated by the fitted model. It is applicable to the PCR (training set) since it is linear in its parameters and not to PCNN due to the non-linear nature of its parameters.

Table 4. *Statistical analysis of PCR and PCNN models for problem (1): predicting rutting behavior.*

Subset	Statistics	PCR			PCNN		
		Fold 1	Fold 2	Fold 3	Fold 1	Fold 2	Fold 3
Training	AD	0	0	0	34.99	-242.99	46.19
	AAD	1497.02	1705.55	1514.59	729.41	1350.87	944.83
	r_{fit}	0.83	0.82	0.85	0.96	0.87	0.94
	R^2	0.69	0.68	0.72	na*	na*	na*
Testing	AD	626.73	-129.91	-226.1	-98.24	149.2	-169.6
	AAD	2007.47	1515.74	2110.64	694.53	719.9	1037.28
	r_{fit}	0.79	**0.80**	0.73	**0.97**	0.95	0.92
	R^2	na*	na*	na*	na*	na*	na*

**not applicable*

Researchers have developed several predictive models for rutting in asphalt mixture (Kaloush et al., 2003; Leahy, 1989; Ayres and Witczak, 1998; Kaloush and Witczak, 1991; Rodezno et al., 2010; Witczak and El-Basyouny, 2004; Kvasnak et al., 2007; Andrei et al., 1999; ARA, 2004). A summary of the most well-known conventional models, their parameters and prediction accuracy expressed in R^2 is presented in Table 5.

Although the reported R^2 values for the Leahy, Ayres, and Kaloush models seem reasonable, they did not use separate data sets for training and testing possibly resulting in biased and over fitted models. Comparing the results obtained from the PCR and PCNN models with the previous prediction models used in the AASHTO design procedure, the developed models are performing significantly better with $r_{fit} = 0.8$ for PCR model and $r_{fit} = 0.97$ for PCNN model. In comparison between PCR and PCNN, one can see that although the PCR works well in predicting the response variable, PCNN provides the best fit for both training and test sets. Measured values of accumulated strain and the fitted values are presented in Figure 3. The measured and fitted values are close to the line of equality meaning that the fitted values by PCR and PCNN have a strong correlation with the measured one.

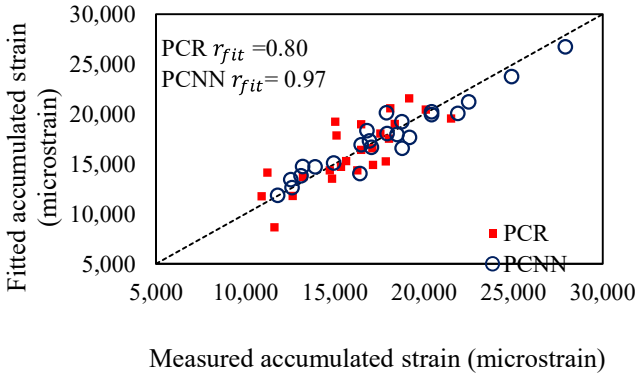

Figure 3. *Measured values of accumulated strain versus fitted values by PCR and PCNN.*

Table 5. *Summary of the well-known existing rutting prediction models.*

Kaloush et al. (2003)	$\log\frac{\epsilon_p}{\epsilon_r} = -3.15552 + 1.734\log T + 0.39937\log N$ $R^{2^*} = 0.664$ $\log\frac{\epsilon_p}{\epsilon_r} = 0.3082 + 0.3534\log N$ $R^{2^*} = 0.55$	ϵ_p = accumulated permanent strain ϵ_r = resilient strain N = number of load repetitions T = mix temperature (°F)
Leahy (1989)	$\log\frac{\epsilon_p}{\epsilon_r} = -6.631 + 0.435\log N + 2.767\log T$ $+ 0.11\log S$ $+ 0.118\log\eta$ $+ 0.93\log V_{beff} + 0.501\log V_a$ $R^{2^*} = 0.76$	ϵ_p = accumulated permanent strain ϵ_r = resilient strain N = number of load repetitions T = mix temperature (°F) S = deviatoric stress (psi_ η = viscosity at70 °F ($10^6\ poise$) V_{beff} = effective binder content (%volume) V_a = percent air voids
Ayres and Witczak (1998)	$\log\frac{\epsilon_p}{\epsilon_r} = -4.80661 + 2.58155\log T + 0.42956\log N$ $R^{2^*} = 0.725$	ϵ_p = accumulated permanent strain ϵ_r = resilient strain N = number of load repetitions T = mix temperature (°F)
Kaloush and Witczak (1991)	$FN = (432367000)T^{-2.215}Visc^{0.312}V_{beff}^{-2.6604}V_a^{-0.1525}$ $\frac{S_e}{S_y} = 0.534,\ R^{2^*} = 0.72$	T = test temperature (^oF) $Visc$ = binder viscosity at 70°F $(10^6\ poise)$ V_{beff} = effective asphalt content (%volume) V_a = air voids (%)
Kvasnak et al. (2007)	$logFN = 2.866 + 0.00613\ Gyr + 3.86\ Visc -$ $0.072VMA + 0.282\ P_4 - 0.051\ P_{16} + 0.075\ P_{200}$ $R^{2^*} = 0.91$	Gyr = number of gyrations $Visc$ = binder viscosity at test temperature $(0.02\ to\ 0.286 * 10^6\ poise)$ VMA = voids in mineral aggregates P_4, P_{16}, P_{200} = percent aggregate passing from sieve sizes #4, #16, and #200
2002 design guide (ARA, 2004)	$\frac{\epsilon_p}{\epsilon_r} = a\ T^b N^c$	ϵ_p = accumulated plastic strain at N repetition of load ϵ_r = resilient strain of the asphalt material as a function of mix properties, temperature, and time rate of loading N = number of load repetitions a, b, c = non-linear regression coefficients

** It should be mentioned that the ability of the developed model in fitting to the empirical data should be expressed in terms of the r_{fit} value and not the R^2 value.*

3.2. Problem (2): Predicting Dynamic Modulus

The problem of predicting dynamic modulus, as one of the most important performance characteristics of asphalt mixture, is solved and discussed in this section.

3.2.1 Material and Laboratory Testing

To create a data set for dynamic modulus prediction, 27 specimens from nine different asphalt mixtures are selected and used to perform laboratory testing. AASHTO T 209/ASTM D2041 was used to measure maximum theoretical values of specific gravity (G_{mm}). The measured G_{mm} values were used to obtain other volumetric properties of the asphalt mixtures.

The samples were 150 mm in diameter and 38 mm in thickness cut from Superpave gyratory compacted specimens. The dynamic modulus test in indirect tension mode is performed at three temperatures (0.4, 17.1, and 33.8 °C) and nine loading frequencies (25, 20, 10, 5, 2, 1, 0.5, 0.2, 0.1 Hz) in accordance with AASHTO TP 62-07, "Standard Method of Test for Determining Dynamic Modulus of Hot-Mix Asphalt Concrete Mixtures." Based on ASTM D7552-09, the dynamic shear rheometer (DSR) test was conducted to measure complex shear modulus of asphalt binder. The test was performed at a wide variety of temperatures (-10 to 54 °C) and frequencies (0.1 Hz to 25 Hz), including the exact test temperatures and loading frequencies which were used in the mixture dynamic modulus test. It is important to mention that the present research uses a consistent definition of frequency, meaning that to predict the dynamic modulus value of an asphalt mixture for example at 4 °C and 25 Hz, one should input in the model the complex shear modulus of asphalt binder, $|G^*|$, at 4 °C and 25 Hz. A summary of the nine different mixtures' properties is presented in Table 6. Using the laboratory test results on 27 specimens, a database of 243 data points was created to be used in further modeling.

3.2.2. Step 1: Data Preprocessing

According to the literature, dynamic modulus of asphalt mixture can be estimated by the properties of the material components (Andrei et al., 1999; Bari and Witczak, 2008; Christensen et al., 2003; Al-Khateeb et al., 2006; Sakhaeifar et al., 2015). The input variables are selected amongst those properties that have already proven to be important in predicting dynamic modulus. However, their importance was also re-examined before selection by performing a multi-factor analysis of variance (ANOVA). The selected component properties and their ranges that are obtained from laboratory testing and used in the present study are summarized in Table 7.

To evaluate the quality of the predictors, correlation analysis was performed, and the results are presented in Table 8. Fifty elements of the pairwise correlation matrix with absolute values of greater than 0.5 are shown in bold text indicating that there is strong correlation between the predictors. Therefore, to eliminate the existing correlation, PCA was implemented.

Table 6. *Properties of nine asphalt mixtures used in problem (2): predicting dynamic modulus.*

Properties	Asphalt mixtures								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Binder PG	58-28	58-28	58-28	58-34	58-34	58-34	64-28	64-34	64-28
% V_{beff}	4.20	4.10	4.10	3.90	3.50	4.30	4.20	4.00	4.60
%VMA	13.50	13.50	13.60	13.10	12.50	13.90	13.70	13.40	14.40
% VFA	70.30	70.40	70.60	69.60	68.10	71.20	70.80	70.20	72.30
G_{mb}	2.30	2.30	2.30	2.30	2.30	2.30	2.30	2.30	2.30
G_{mm}	2.40	2.50	2.50	2.50	2.60	2.50	2.50	2.50	2.40
% V_A	4.01	4.00	4.00	3.98	3.99	4.00	4.00	3.99	3.99
% Passing 3/4”	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
% Passing 1/2”	93.90	96.40	87.20	93.50	95.10	96.40	94.10	94.40	94.20
% Passing 3/8”	77.50	84.60	73.70	76.40	83.10	87.30	83.40	82.00	80.90
% Passing #4	49.80	53.10	48.40	52.20	52.20	60.90	63.80	48.20	58.60
% Passing #8	34.40	38.40	35.10	43.60	38.80	46.90	47.10	34.90	46.00
% Passing #30	16.70	18.70	17.90	20.90	18.80	23.40	21.70	19.20	25.90
% Passing #50	10.30	10.80	10.90	11.40	9.90	12.40	11.90	11.80	13.80
% Passing #100	6.10	5.90	6.40	5.80	5.40	6.10	6.60	6.10	7.20
% Passing #200	3.60	3.30	6.20	3.30	3.50	3.40	4.00	3.10	4.00

Table 7. *Original input variables of problem (2): predicting dynamic modulus.*

Variable	Identity	Min.	Max.	Ave.	Std. Dev.
x_1	**Cum. % retained on 3/4"**	3.60	13.00	6.11	2.63
x_2	**Cum. % retained on 3/8"**	12.68	26.29	19.01	4.11
x_3	**Cum. % retained on #4**	36.20	51.76	45.86	5.32
x_4	**Cum. % retained on #8**	52.87	65.70	59.42	5.06
x_5	**Cum. % retained on #30**	74.06	83.30	79.63	2.76
x_6	**Cum. % retained on #50**	86.22	90.12	88.57	1.15
x_7	**Cum. % retained on #100**	92.81	94.59	93.83	0.48
x_8	**% Passing from #200**	3.07	6.18	3.81	0.89
x_9	**Log |G*|**	-2.29	3.03	0.50	1.26
x_{10}	**Phase angle (degree)**	28.15	79.17	52.86	11.54
x_{11}	**V_{beff}%**	3.50	4.60	4.10	0.29
x_{12}	**VMA**	12.50	14.40	13.51	0.49
x_{13}	**VFA**	68.10	72.30	70.40	1.08
x_{14}	**Va%**	3.98	4.01	3.99	0.01

Table 8. *Correlation matrix for the input variables for problem (2): predicting dynamic modulus.*

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
x_1	1	**0.832**	0.412	0.366	0.294	0.119	-0.269	**0.905**	-0.044	-0.058	0.003	0.04	0.049	0.013
x_2	**0.832**	1	**0.597**	0.458	0.391	0.246	-0.109	**0.583**	-0.035	0.106	-0.061	-0.099	-0.089	-0.115
x_3	0.412	**0.597**	1	**0.918**	**0.756**	**0.596**	0.425	0.133	-0.019	0.154	-0.465	-0.485	-0.49	-0.111
x_4	0.366	0.458	**0.918**	1	**0.87**	**0.687**	0.375	0.169	-0.028	0.237	-0.388	-0.412	-0.424	0.212
x_5	0.294	0.391	**0.756**	**0.87**	1	**0.919**	**0.618**	0.112	-0.021	0.235	**-0.585**	**-0.631**	**-0.633**	0.3
x_6	0.119	0.246	**0.596**	**0.687**	**0.919**	1	**0.794**	-0.009	0.003	0.203	**-0.741**	**-0.796**	**-0.806**	0.209
x_7	-0.269	-0.109	0.425	0.375	**0.618**	**0.794**	1	-0.414	0.036	0.047	**-0.854**	**-0.886**	**-0.892**	-0.087
x_8	**0.905**	**0.583**	0.133	0.169	0.112	-0.009	-0.414	1	-0.032	-0.102	0.179	0.238	0.238	0.142
x_9	-0.044	-0.035	-0.019	-0.028	-0.021	-0.003	0.036	-0.032	1	**-0.808**	0.021	0.016	0.013	0.034
x_{10}	-0.058	0.106	0.154	0.237	0.235	0.203	0.047	-0.102	**-0.808**	1	0.09	0.024	0.014	0.3
x_{11}	0.003	-0.061	-0.465	-0.388	**-0.585**	**-0.741**	**-0.854**	0.179	0.021	0.09	1	**0.988**	**0.985**	0.372
x_{12}	0.04	-0.099	-0.485	-0.412	**-0.631**	**-0.796**	**-0.886**	0.238	0.016	0.024	**0.988**	1	**0.998**	0.321
x_{13}	0.049	-0.089	-0.49	-0.424	**-0.633**	**-0.806**	**-0.892**	0.238	0.013	0.014	**0.985**	**0.998**	1	0.301
x_{14}	0.013	-0.115	-0.111	0.212	0.3	0.209	-0.087	0.142	0.034	0.3	0.372	0.321	0.301	1

3.2.3. Step 2: Data Wrangling: Dimensionality Reduction Using PCA

The eigenvalues of the correlation matrix are calculated and presented in Table 9. The eigenvalues and their corresponding contribution to the total variance are sorted in descending order. According to the table, 95.8% of the variation in the original data is expressed by the first five PCs. Adding the sixth PC has an insignificant impact on the overall represented variation. Therefore, the first five PCs are selected to be used as input variables in the modeling problem.

Table 9. *Eigenvalues of the normalized matrix and corresponding percent variance for problem (2): predicting dynamic modulus.*

Component Number	Eigenvalue	Percent Variance	Cumulative Percent
1	6.02	43.02	43.02
2	3.23	23.00	66.02
3	1.97	14.10	80.12
4	1.42	10.12	90.24
5	0.79	5.61	95.85
6	0.32	2.27	98.12
7	0.11	0.78	98.90
8	0.08	0.56	99.46
9	0.06	0.39	99.85
10	0.02	0.15	100.00

The PCs are obtained using Equation12 and 13 where matrices M and n are as follows

$$M^T = \begin{bmatrix} 0.03 & 0.19 & -0.08 & -0.06 & -0.09 \\ 0.03 & 0.11 & -0.04 & -0.05 & 0.04 \\ 0.06 & 0.04 & 0.00 & 0.00 & 0.01 \\ 0.06 & 0.05 & 0.02 & 0.04 & 0.07 \\ 0.13 & 0.06 & 0.05 & 0.09 & -0.01 \\ 0.33 & 0.01 & 0.09 & 0.16 & -0.20 \\ 0.71 & -0.54 & 0.05 & 0.13 & -0.20 \\ -0.03 & 0.52 & -0.20 & -0.70 & -0.53 \\ -0.01 & -0.05 & -0.40 & 0.43 & 0.12 \\ 0.00 & 0.01 & 0.06 & -0.02 & 0.00 \\ -1.26 & 0.64 & 0.52 & 0.61 & 0.67 \\ -0.75 & 0.37 & 0.21 & 0.28 & 0.31 \\ -0.34 & 0.17 & 0.09 & 0.12 & 0.14 \\ -4.41 & 18.38 & 47.08 & 76.24 & -40.48 \end{bmatrix}, \quad n = \begin{bmatrix} -55.95 \\ -58.54 \\ -218.20 \\ -352.79 \\ 174.78 \end{bmatrix}$$

3.2.4. Step 3: Model Development

To examine the stability of the developed model against an unseen data set and based on the size of data set, a holdout cross validation technique was used. During this procedure the given data set was randomly assigned to two subsets, d0 and d1, called the training set

and the test set, respectively. The training set contains 80% of the data points and the test set contains 20% of the data points which means 80% of the data points are used to train the model and the rest is used to evaluate the performance of the trained model.

After examining all the possible regression structures, the best reduced third-order cubic linear regression model fitting the measured response, y, is presented in Equation 18:

$$\hat{y} = c_0 + c_1\,pc_1 + c_2\,pc_2 + c_3\,pc_3 + c_4\,pc_4 + c_5\,pc_5 + c_6\,pc_1\,pc_2 + c_7\,pc_1\,pc_3 + c_8\,pc_1\,pc_4 + c_9\,pc_1\,pc_5 + c_{10}\,pc_2\,pc_3 + c_{11}\,pc_2\,pc_4 + c_{12}\,pc_2\,pc_5 + c_{13}\,pc_3\,pc_4 + c_{14}\,pc_3\,pc_5 + c_{15}\,pc_4\,pc_5 + c_{16}\,pc_1\,pc_2\,pc_3 + c_{17}\,pc_1\,pc_2\,pc_4 + c_{18}\,pc_1\,pc_2\,pc_5 + c_{19}\,pc_1\,pc_3\,pc_4 + c_{20}\,pc_1\,pc_3\,pc_5 + c_{21}pc_2\,pc_3\,pc_4 + c_{22}\,pc_1\,pc_4\,pc_5 + c_{23}\,pc_2\,pc_4\,pc_5 + c_{24}\,pc_3\,pc_4\,pc_5 \quad [18]$$

Where $c_0 = 6.59;\ c_1 = 2.58;\ c_2 = 4.4;\ c_3 = -0.36;\ c_4 = 0.49;\ c_5 = 1.93;\ c_6 = -0.33;\ c_7 = -0.77;\ c_8 = -1.69;\ c_9 = 0.15;\ c_{10} = -1.65;\ c_{11} = -4.68;\ c_{12} = 4.81;\ c_{13} = 0.7;\ c_{14} = -0.85;\ c_{15} = -1.58;\ c_{16} = -0.17;\ c_{17} = -0.79;\ c_{18} = 1.83;\ c_{19} = 0.04;\ c_{20} = 0.18;\ c_{21} = 0.42;\ c_{22} = 0.05;\ c_{23} = 0.32;\ c_{24} = 0.06.$

The networks weights and biases are presented in the following matrices.

$$W^T = \begin{bmatrix} -0.511 & 0.134 & 0.654 & -1.064 & -0.267 \\ -0.315 & -0.147 & -0.267 & 0.177 & -1.047 \\ -0.060 & -1.266 & 0.759 & -1.248 & -0.331 \\ -0.075 & 0.022 & 0.208 & 0.015 & 0.167 \\ -0.074 & 0.022 & 0.206 & 0.015 & 0.165 \\ 0.103 & -0.177 & 1.253 & -1.045 & 0.535 \\ 0.078 & -0.020 & -0.231 & -0.014 & -0.172 \\ 0.238 & 0.070 & -0.885 & 0.848 & 0.943 \\ 0.123 & 0.456 & -0.387 & 1.547 & -0.017 \\ -0.079 & 0.020 & 0.213 & 0.014 & 0.173 \end{bmatrix}$$

$$W_H = \begin{bmatrix} 0.869 \\ -0.886 \\ 0.632 \\ -0.291 \\ -0.288 \\ -0.859 \\ 0.299 \\ 0.556 \\ 0.971 \\ -0.299 \end{bmatrix} \qquad B_H = \begin{bmatrix} 0.162 \\ 0.710 \\ 0.319 \\ -0.008 \\ -0.009 \\ 0.570 \\ 0.007 \\ 0.290 \\ -0.373 \\ -0.007 \end{bmatrix} \qquad B_0 = [0.148]$$

Performance results of the PCR and PCNN models were compared with three of the well-known predictive models for dynamic modulus based on the statistics introduced in section 3.1.3. A summary of these models' equations and definitions of parameters is presented in Table 10 (Bari and Witczak, 2008; Christensen et al., 2003; Al-Khateeb et al., 2006).

According to the obtained r_{fit} values, the predicted values of dynamic modulus by PCR and PCNN models have a strong correlation (0.99) with the measured ones thus, both PCR and PCNN are capable of predicting the response. The corresponding values of r_{fit} fit for the modified Witczak, Hirsch, and Alkhateeb models are 0.93, 0.95, and 0.95 respectively, which seem reasonable. However, r_{fit} can be biased since it shows if the measured response increases then the predicted response will increase and vice versa. Considering

other statistics, AD and AAD, shows that these values are significantly higher for the modified Witczak, Hirsch, and Alkhateeb models, meaning that the fitted values by these models are not as close as the ones fitted by PCR and PCNN to the response value. In other words, AD and AAD indicate that the Hirsch and Alkhateeb models are overpredicting and the modified Witczak model is underpredicting the response variables. Measured values of dynamic modulus and the fitted values are presented in Figure 4. The measured and fitted values are close to the line of equality meaning that the fitted values by PCR and PCNN have a strong correlation with the measured one.

Table 10. *Summary of the well-known conventional dynamic modulus prediction model*

Modified Witczak	$\log\|E^*\| = -0.349 + 0.754(\|G_b^*\|^{-0.0052})(6.65 - 0.032\,\rho_{200} - 0.0027\,(\rho_{200})^2 + 0.011\rho_4 - 0.0001\,(\rho_4)^2 + 0.006\,\rho_{\frac{3}{8}} - 0.00014\left(\rho_{\frac{3}{8}}\right)^2 - 0.08\,V_a - 1.06\,(\frac{V_{beff}}{V_{beff}+V_a})) + \frac{2.558+0.032\,V_a+0.713\left(\frac{V_{beff}}{V_{beff}+V_a}\right)+0.0124\,\rho_{\frac{3}{8}}-0.0001\left(\rho_{\frac{3}{8}}\right)^2-0.0098\,\rho_{\frac{3}{4}}}{1+\exp(-0.7814-0.5785\log\|G_b^*\|+0.8835\log\delta_b)}$	$\|E^*\|$ = dynamic modulus (psi) $\|G^*\|$ = binder shear modulus (psi) δ_b = binder phase angle (degree) $\rho_{\frac{3}{4}}$, $\rho_{\frac{3}{8}}$, ρ_4 = cumulative percent aggregate retained on sieve seizes ¾'', 3/8" and no.4 ρ_{200} = percent aggregate passing from sieve no.200 V_a = percent air voids V_{beff} = effective binder content
Hirsch	$\|E^*\| = P_C\left[4200000\left(1-\frac{VMA}{100}\right)+3\|G_b^*\|\left(\frac{VFA*VMA}{1000}\right)\right] + \frac{1-P_c}{\frac{1-VMA}{100}+\frac{VMA}{3\|G_b^*\|(VFA)}}$ $P_c = \frac{\left(20+\frac{3\|G_b^*\|(VFA)}{VMA}\right)^{0.58}}{650+\left(\frac{3\|G_b^*\|(VFA)}{VMA}\right)^{0.58}}$	$\|E^*\|$ = dynamic modulus (psi) $\|G^*\|$ = binder shear modulus (psi) VMA = voids in mineral aggregate VFA = voids filled with asphalt Pc = aggregate contact volume
Al-Khateeb	$\|E^*\| = 3\left(\frac{100-VMA}{100}\right)\left(\frac{\left(90+1.45\frac{\|G^*\|}{VMA}\right)^{0.66}}{1100+\left(0.13\frac{\|G^*\|}{VMA}\right)^{0.66}}\right)\|G_g^*\|$	$\|E^*\|$ = dynamic modulus (psi) $\|G^*\|$ = binder shear modulus (psi) VMA = voids in mineral aggregate $\|G_g^*\|$ = complex shear modulus of binder in glassy state

Table 11. *Statistical comparison of PCR and PCNN and the existing predictive models for problem (2): predicting dynamic modulus.*

Model		Average difference (MPa)	Average abs difference (MPa)	rf_{it}	R^2
PCR	**Training**	3.90	575.30	0.99	0.99
	Testing	-162.30	718.90	0.99	na
PCNN	**Training**	13.20	380.70	0.99	na
	Testing	9.70	337.50	0.99	na
Modified Witczak		-2460	3152.10	0.93	0.88
Hirsch		1241.60	1785.70	0.95	0.91
Alkhateeb		2844.50	2984.50	0.95	0.9

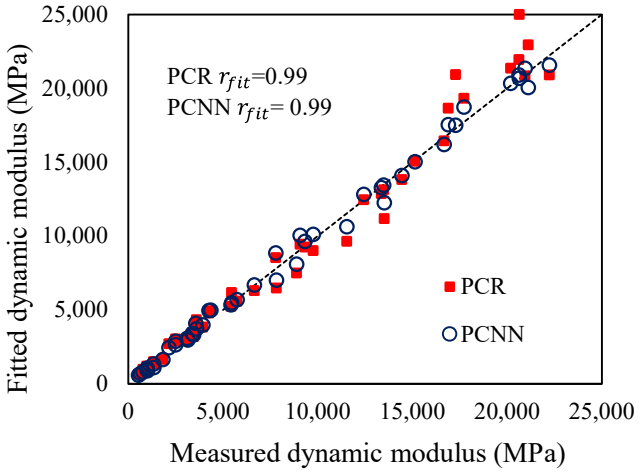

Figure 4. *Measured values of dynamic modulus versus fitted values by PCR and PCNN.*

4.0 Model Validation

The task of confirming that the outputs of a statistical model have enough fidelity to the output of data generating process is called model validation. The difference between estimated and measured values of the response (residual) is assumed to be a random error which is normally distributed with a mean of zero and unknown variance (Devore, 2012). To examine the adequacy of these assumptions, two sets of residual diagnostic analyses are implemented and presented in Figure 5. If the residuals are random error terms, the residual plot should contain no obvious pattern. According to the residual plot this assumption is satisfied for both PCR and PCNN in rutting as well as dynamic modulus prediction models. The assumption of normality can be checked by a normal probability plot in a way that if the residual distribution is normal, their plot will resemble a straight line. According to Figure 6 the data points are located around a straight line. Therefore, the normality assumption does not appear to be violated.

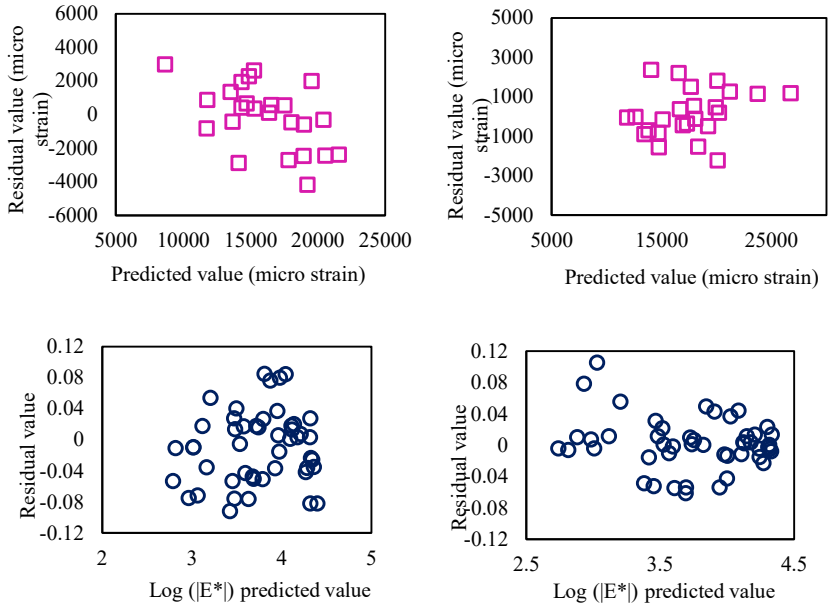

Figure 5. *Plot of the residuals for (a) PCR rutting predictive model, (b) PCNN rutting predictive model, (c) PCR dynamic modulus predictive model, and (d) PCNN dynamic modulus predictive model.*

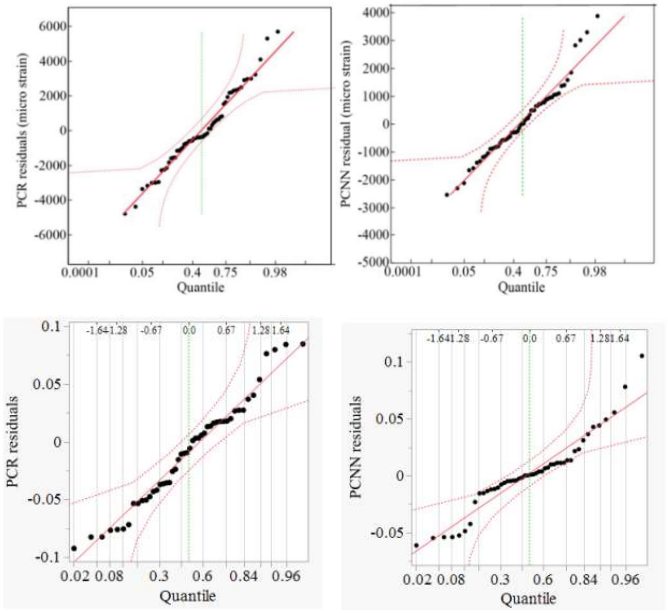

Figure 6. *Normal probability plot of the residuals for (a) PCR rutting predictive model, (b) PCNN rutting predictive model, (c) PCR dynamic modulus predictive model, and (d) PCNN dynamic modulus predictive model.*

5.0 Variable Importance Analysis (VIA)

Once the performance model has been identified, design engineers may desire to know which factor (predictor variable) included in the model has the strongest influence on the

response variable. If variation of a specific factor causes high variability in the response, that effect is important relative to the model. Variable importance analysis is one of the global sensitivity analysis methods and is based on the functional decomposition idea by Sobol (1990). He proved that one can decompose a function $y = f(x1, \dots, xp) = f0 + f1(x1) + \dots + fp(xp) + f12(x1, x2) + \dots$ into the sum of lower dimensional functions. The variability of these lower dimensional functions assesses the importance of input variables in terms of effect indexes which indicate relative importance of the variables and are presented as the main effect,

$$S_i = \frac{V_i}{V(y)} = \frac{V(E(y|x_i))}{V(y)} \quad [19]$$

and the total effect

$$S_{Ti} = \frac{V_{Ti}}{V(y)} = \frac{V(y) - V(E(y|x_{-i}))}{V(y)} \quad [20]$$

Where xi represents a random input (vector) variable and x_{-i} indicates all input variables except xi. The unique contribution of the input variable xi to the total variation of the response, y, is represented by main effect, while the total effect represents the overall contribution of xi on y which includes all interaction terms (Wei et al., 2015).

In the case of correlated factors (inputs), the contribution of an individual input variable to the variation of the response should be divided into two parts of uncorrelated and correlated contributions. In order to account for correlation, in the present study, factor values were constructed from observed combinations using a k-nearest neighbors (KNN) approach. Observed variance and co-variance were treated as representative of the co-variance structure of the factors. For rutting and dynamic modulus prediction models, VIA was conducted, and the results are presented in this section. Table 12 and Figure 7 indicate the main and total effects for the rutting prediction model. For the developed rutting prediction model, percent air voids, percent passing the #200 sieve, VFA and VMA are in turn the most effective variables, while for dynamic modulus prediction model the most effective variables are in turn complex shear modulus and phase angle as presented in Table 13 and Figure 8.

Table 12. *Variable importance analysis results for rutting prediction model.*

Variable	Main Effect	Total Effect
Va%	0.155	0.195
Passing #200	0.024	0.181
VFA	0.155	0.155
VMA	0.128	0.134
Passing #16	0.042	0.119
Passing #100	0.033	0.112
E*	0.091	0.091
NMAS	0.068	0.068
Passing #50	0.03	0.067
Passin 3/4''	0.052	0.052
Passing #8	0.043	0.043
G*	0.03	0.043
Passing 1/2''	0.041	0.041
Passing 3/8''	0.035	0.035
Passing #30	0.03	0.03
Passing #4	0.02	0.025
Binder%	0.021	0.021

Table 13. *Variable importance analysis results for dynamic modulus prediction model.*

Variable	Main Effect	Total Effect
Log |G*|	0.221	0.685
Phase angle	0.156	0.322
VMA	0.057	0.057
Cumulative % Retained on 1/2''	0.052	0.052
Cumulative % Retained on 3/8''	0.052	0.052
Cumulative % Retained on #4	0.052	0.052
Cumulative % Retained on #8	0.052	0.052
Cumulative % Retained on #30	0.052	0.052
Cumulative % Retained on #50	0.052	0.052
Cumulative % Retained on #100	0.052	0.052
Passing #200	0.052	0.052
VFA	0.052	0.052
%Va	0.052	0.052
%Vbeff	0.042	0.042

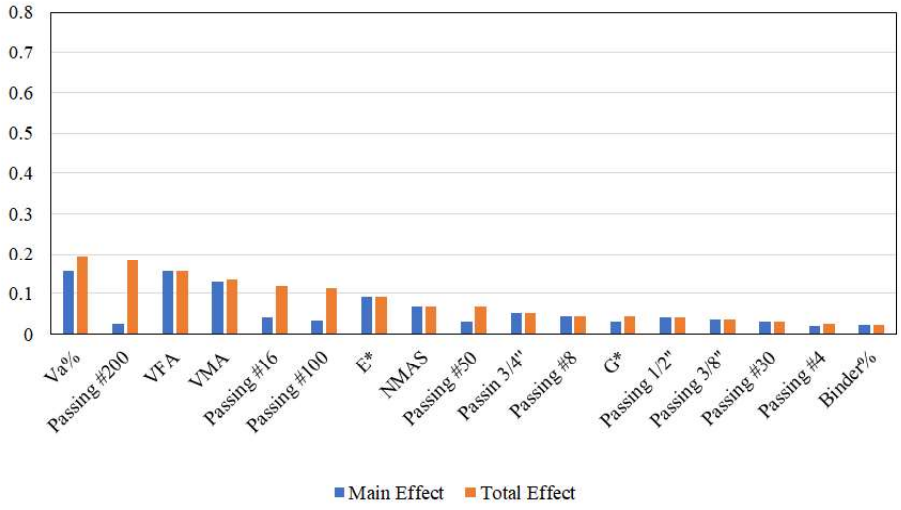

Figure 7. *Variable importance analysis results for rutting prediction model.*

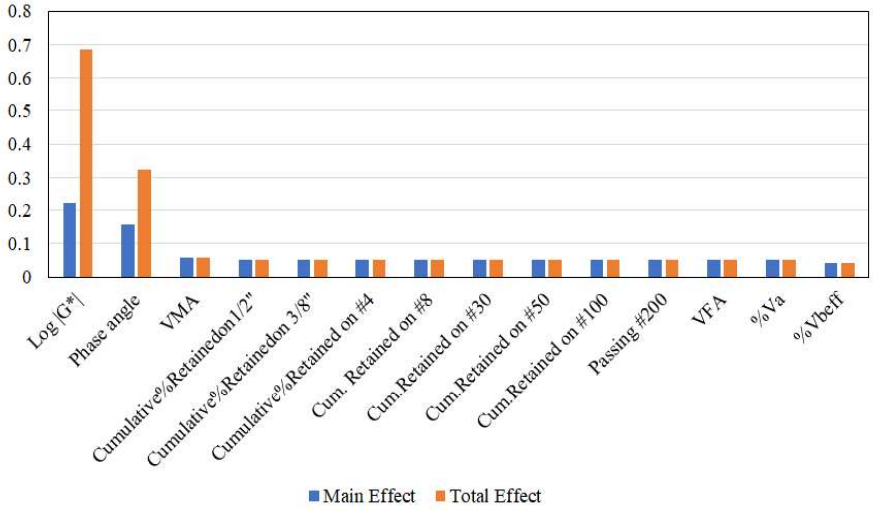

Figure 8. *Variable importance analysis results for dynamic modulus prediction model.*

6.0 Application of the Framework

In this section we present two applications of the developed framework in the design and optimization of flexible pavement.

6.1 Application 1: Minimizing Accumulated Strain

A proper formulation of minimizing accumulated strain in the flexible pavement design can be formulated based on the above framework. The objective function is the output of the trained network based on the laboratory data as explained in the previous section. The next step is adding the appropriate constraints to this problem. These constraints ensure

that the desired point remains in the space where the trained data set was originally located and that the model remains reliable in this domain. It should be noted that the ANN network makes a nonlinear objective function and unlike convex function, their behavior tends to promote multiple local minimizing points. The problem becomes more complicated by the introduction of constraints.

Since the overall optimization problem is not convex, gradient-based algorithms are prone to being trapped in a local optimal point. Many researchers have shown that a proper application of evolutionary-based algorithms for engineering-based problems can result in desirable solutions in a finite amount of computational time (Cai and Wang, 2006; He and Yao, 2002; Thiele and Zitzler, 1999). In this study, we used a variant of an evolutionary-based search algorithm called mean variance mapping optimization (MVMO). MVMO (Erlich et al., 2010) employs a number of operators to locate the global optimum of a given function. More specifically, MVMO preserves an archive of points during its evolution and extracts information to proceed forward from this archive. MVMO also behaves adaptively so that it can explore the domain initially and exploit the specific domain of interest to better locate the global optimum.

MVMO was originally designed to solve unconstrained problems. In an attempt to optimize truss structures with multiple constraints, Aslani et al. (2017) provided an adaptive penalty function which can transfer a given constrained problem into an unconstrained one. This strategy has been shown to be robust, especially in the context of the problems with many constraints involved. In the problem of minimizing accumulated strain, the constraints limiting the space to an enclosing ellipsoid enter the problem through constraints and then are added to the objective function through the penalty approach that was described above. Thus, the problem is formulated as following:

$$\begin{aligned} &\text{minimize} && \epsilon = \epsilon_{ANN}(\mathrm{x}) \\ &\text{with respect to} && x = (x_1, \dots, x_{17}) \\ &\text{subject to} && (x - v)^T A\,(x - v) \le 1 \\ & && \left(x_{pca} - v'\right)^T A' \left(x_{pca} - v'\right) \le 1 \end{aligned} \qquad [21]$$

Where ε is the objective function (in this case, it is the accumulated strain) which is a function of material property indicated by x. Finally, the set of constraints are represented by $(x - v)^T\, A(x - \, v) \le 1$. The uncertainty in the process of finding the enclosing can be found as a function of principal semi-axes of the ellipsoid (s_i):

$$(\mathrm{x} - v)^T A\,(\mathrm{x} - v) \le 1 - |\Delta| \max(s_i) \qquad [22]$$

Where Δ is pre-defined threshold in the effective variable space section. As indicated before, the constrained problem is changed to an unconstrained one through application of a penalty function.

The MVMO convergence plot is show in Figure 9. The initial point is random, and the algorithm is evolved through the application multiple routines that model exploration initially and later on exploitation to narrow down the domain of interest.

PCNN predictive model used in the objective function along with the constraint results in a minimum of 1772 micro strain. Table 14 summarizes the material properties associated with this solution. The obtained aggregate gradation graph in presented in Figure 10.

Table 14. *Design parameters associated with minimum accumulated strain for problem (1): predicting rutting behavior.*

Variable	Identity	Value Obtained from PCNN	Superpave Design Specification			
			Control Points		Restricted Zone	
			Lower	Upper	Lower	Upper
x_1	**Binder %**	4	-	-	-	-
x_2	**G* (Pa)**	270,190	-	-	-	-
x_3	**NMAS**	19	-	-	-	-
x_4	**Passing 3/4"**	92	90	100	-	-
x_5	**Passing 1/2"**	66	-	90	-	-
x_6	**Passing 3/8"**	65	-	-	-	-
x_7	**Passing #4**	58	-	-	-	-
x_8	**Passing #8**	50	23	49	34.6	34.6
x_9	**Passing #16**	39	-	-	22.3	28.3
x_{10}	**Passing #30**	27	-	-	16.7	20.7
x_{11}	**Passing #50**	9	-	-	13.7	13.7
x_{12}	**Passing #100**	4	-	-	-	-
x_{13}	**Passing #200**	3	2	8	-	-
x_{14}	**VMA**	16	13	-	-	-
x_{15}	**VFA**	76	65	80	-	-
x_{16}	**Va%**	4	4		-	-
x_{17}	**E*(Mpa)**	713	-	-	-	-

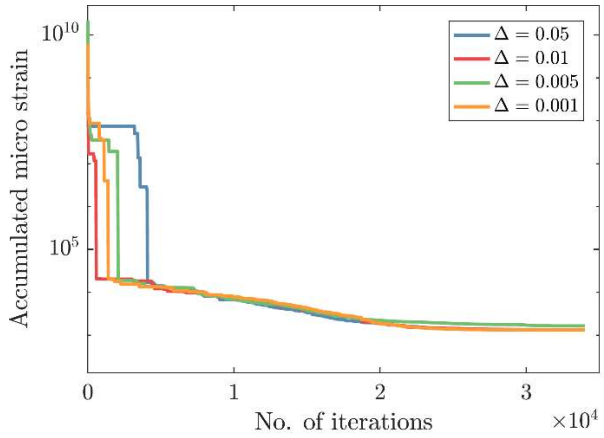

Figure 9. *Convergence plot for minimization problem of accumulated strain. For different values of the threshold parameter the algorithm converges to the same output.*

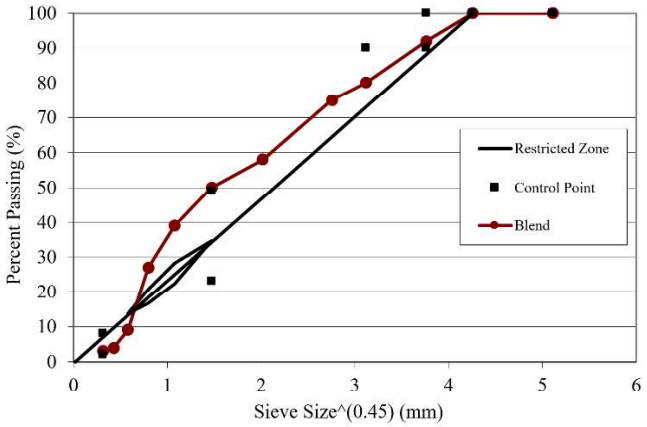

Figure 10. *Optimal aggregate gradation graph with 19.0mm NMAS particle size distribution associated with minimum accumulated strain.*

6.2. Problem 2

The developed predictive model for dynamic modulus is used along with an optimization algorithm to answer the following two central questions.

- What design parameters result in the maximum $|E^*|$?
- What design parameters result in a pre-specified $|E_0^*|$?

One can see that the first item corresponds to the optimal design problem whereas the second one corresponds to the so-called inverse design. Recalling the PCNN's higher prediction capability, solving optimization problems based on PCNN would be more reliable. The optimal design problem is formulated as follows:

$$\begin{aligned} &\text{maximize} && |E^*| = F_{ANN}(\mathrm{x}) \\ &\text{with respect to} && \mathrm{x} = (x_1, \dots, x_{14}) \\ &\text{subject to} && (\mathrm{x} - v)^T A\,(\mathrm{x} - v) \le 1 \end{aligned} \qquad [23]$$

$$\left(\mathrm{x}_{\mathrm{pca}} - v'\right)^T A' \left(\mathrm{x}_{\mathrm{pca}} - v'\right) \leq 1$$

$$(\mathrm{X} - v)^T A\ (\mathrm{x} - v) \leq 1 - |\Delta| \max(s_i)$$

Similar to the rutting predicting problem, the vector x represents the predictor variables, and $(x - v)^T A(x - v) \leq 1$ is the constrain equation. This optimization problem will be solved by implementing a penalty function that penalizes (decreases, in the case of the maximization problem) the objective value for each constraint regarding its degree of closeness/violation of the corresponding constraint.

Compared to optimal design problem, the inverse design problem is a minimization problem defined as follows:

minimize	$error = \left\|\|E^*\| - \|E_0^*\|\right\|$	
with respect to	$\mathrm{x} = (x_1, \ldots, x_{14})$	
subject to	$(\mathrm{x} - v)^T A\ (\mathrm{x} - v) \leq 1$	[24]
	$\left(\mathrm{x}_{\mathrm{pca}} - v'\right)^T A' \left(\mathrm{x}_{\mathrm{pca}} - v'\right) \leq 1$	

Where $|\mathrm{E}_0{}^*|$ is the desired dynamic modulus. A similar penalization method is used to address the constraint in this case as well. As mentioned in previous section, constrained MVMO algorithm is implemented to solve the problem.

Figure 11 (left) shows the convergence plot for the optimal design problem solved using the constrained MVMO algorithm. The algorithm initiates with a random initial point (heavily penalized as can be seen from the graph) and the objective function increases with every iteration. It should be noted that Δ = 0.05 is used as the threshold in Figure 9. Solving the maximization problem resulted in $|E^*max| = 53{,}703\ MPa$.The obtained optimal design parameters are presented in the first column of Table 15. The maximization problem is solved one more time with an additional constraint of $G^* \sin(\delta) \leq 5000$ kPa to find the maximum dynamic modulus one could design for without low temperature failure in the asphalt binder. Solving this problem resulted in $|E_{max}| = 36{,}307\ MPa$. Corresponding design parameters are presented in the second column of Table 15 as the optimal design 2.

Figure 11 (right) shows the convergence of the algorithm for the inverse design problem which was started randomly from three different initial points. The algorithm is terminated when the error reaches around 10^{-9}. A pre-specified $|\mathrm{E}_0{}^*|$ of 20,417 MPa. is considered and the inverse problem of finding the corresponding design parameters is solved. Due to the non-linearity of the function, the problem does not have a unique solution. Three of the possible solutions are presented as designs 1 to 3 in Table 15. The five sets of design parameters are compared with current design specification in Table 15. The percentage of aggregate passing by sieve size located in the allowable range of the gradation specification. Gradation charts are presented in Figure 12. The obtained percentage for air void is 4% which is the target value of the design specification. The obtained values for

VMA are slightly less than 14% for nominal maximum aggregate size (NMAS) of 12.5 mm. The reason is that VMA values of the nine mixtures used to train the PCNN are slightly less than 14% (see Table 6). As discussed in previous section, the acceptable range for VFA depends on the amount of traffic. The obtained VFAs for all of the five sets of design are satisfied for all of the traffic categories.

Table 15. *Design parameters associated with maximum and predefined values of dynamic modulus for problem (2): predicting dynamic modulus.*

Identity	Optimal Design	Optimal Design restricted by low temperature performance criteria	Design 1	Design 2	Design 3	Design Specification			
						Control Points		Restricted Zone	
						Lower	Upper	Lower	Upper
%Passing from 3/4"	100.00	100.00	100.00	100.00	100.00	-	100	-	-
%Passing from 1/2"	93.38	94.03	92.25	91.88	91.80	90	100	-	-
%Passing from 3/8"	81.74	81.72	79.57	79.92	80.70	-	90	-	-
%Passing from #4	53.00	53.90	55.36	55.23	54.39	-	-	-	-
%Passing from #8	39.56	40.51	41.37	41.08	40.92	28	58	39.1	39.1
%Passing from #30	20.75	20.68	21.02	20.87	20.83	-	-	19.1	23.1
%Passing from #50	11.66	11.60	12.08	11.81	12.02	-	-	15.5	15.5
%Passing from #100	6.22	6.21	6.52	6.38	6.40	-	-	-	-
%Passing from #200	4.10	3.85	4.38	4.58	4.56	2	10	-	-
G* (Mpa)	103.13	7.81	133.51	30.20	11.82	-	-	-	-
Phase angle (degree)	35.71	39.60	47.69	47.27	44.77	2	8	-	-
Vbeff%	4.11	4.18	4.02	4.06	4.05	-	-	-	-
VMA	13.47	13.56	13.41	13.45	13.44	-	-	-	-
VFA	70.29	70.50	70.11	70.24	70.24	-	-	-	-
Va%	4.00	4.00	3.99	4.00	4.01	4		-	-

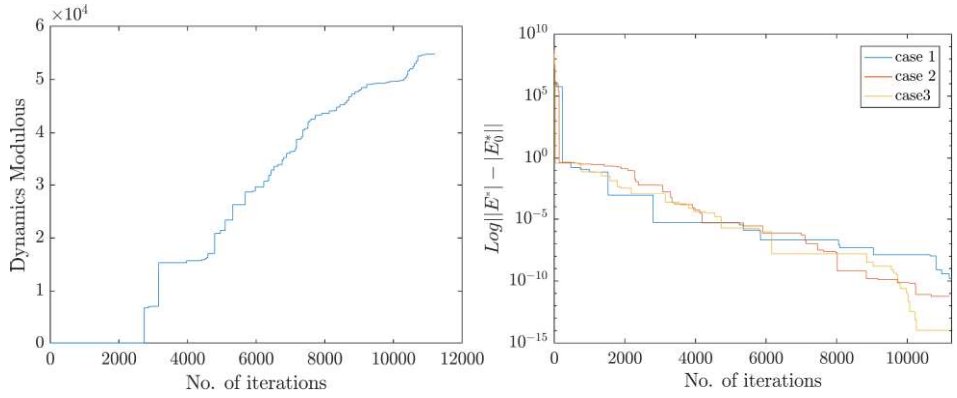

Figure 11. *Convergence plot for maximization (left) and inverse design (right) problems of dynamic modulus. For different values of the threshold parameter the algorithm converges to the same output.*

It should be mentioned that these are examples of the framework's application in pavement design indicating how to apply appropriate constraints to the modeling problem. Other examples can be finding optimal design parameters when a specific source of aggregate (or specific aggregate size) is missing. Multi-objective design problems (minimizing rut depth while maximizing pavement fatigue life) can also be solved with this framework.

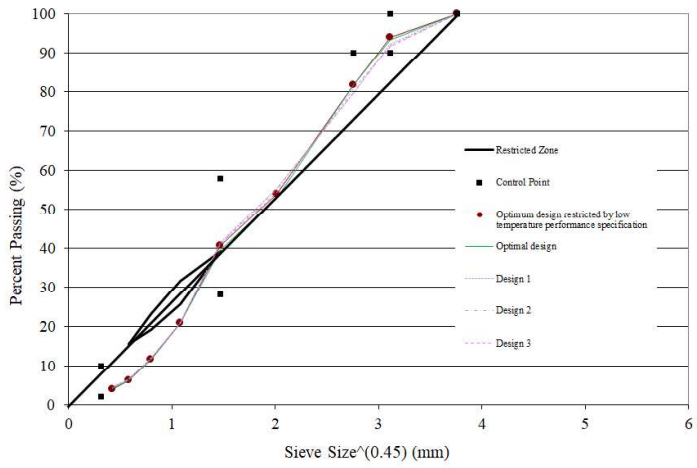

Figure 12. *Aggregate gradation graphs with 12.5-mm NMAS particle size distribution obtained from PCNN.*

7.0 Conclusions and Recommendations

The purpose of this study was to develop a predictive framework for pavement performance that could be reproducible and easy to use for every data base. The proposed framework in a data preprocessing step, evaluates and qualifies the input variables to use them in further analysis. It identifies cross correlated input variables using correlation

analysis and substitutes them by orthogonal pseudo-variables (PCs) using a multivariate technique called PCA. This transformation not only eliminates the existing pairwise correlation between the original inputs but also it reduces the dimensionality of the data set and increases the accuracy of the prediction. The framework uses pseudo-inputs (PCs) and develops two predictive models using multivariate regression and ANN (models are called PCR and PCNN respectively.)

According to the size of the available data set, the framework implements a cross-validation technique to prevent developing biased or overfitted models. In case of limited amount of data points (e.g. rutting data set) the network performs k-fold cross validation technique to assure that the developed model is stable against an unseen data set. Empirical predictive models can lead to inaccurate results under extrapolation. A simple method is implemented to define the effective variable space in which both predictive models can be used. The defined hyperspace is added as a constraint to the modeling problem.

To illustrate the authority of the proposed framework, two separate performance prediction problems are defined and solved. In the first problem, the rutting behavior of asphalt mixture is predicted as a function of asphalt binder, aggregate, and mixture properties using experimental data from the flow number test. In the second problem, the dynamic modulus of asphalt mixture is predicted using asphalt binder, aggregate, and mixture volumetric properties using experimental data of the dynamic modulus test. In both of the problems, the developed models, PCR and PCNN, indicated satisfactory performance in terms of modeling the amount of permanent deformation and dynamic modulus value, with PCNN being significantly better in fitting the test data than the conventional performance predictive models.

To indicate one application of the proposed framework in pavement performance prediction, the problem of finding the optimal design parameters is solved using mean-variance mapping optimization algorithm for both of the performance prediction problems over their effective variable spaces. The value of 1772 micro-strain is obtained as the minimum accumulated strain. For the problem of predicting dynamic modulus, the inverse design problem of finding the design parameters corresponding to any pre-specified value of dynamic modulus is also solved over its effective variable space. In all the optimization problems the design parameters corresponding to optimum or inverse designs are obtained for PCNN. The obtained optimal design parameters satisfy the current asphalt pavement design specifications and could be used as an appropriate starting point in the design procedure.

It is also worth pointing out that selection of study materials was based on the availability of asphalt mixtures for laboratory testing and like every empirical model the obtained results are based on the available empirical database. Eventually, for creating a more reliable predictive model, a larger database is required. However, what makes machine learning-based models special is that the model will be retrained and modified when a new data set is fed into the framework. Thus, unlike other empirical predictive

models there is no need to calibrate the model for each specific location based on its climate condition and available materials. Moreover, the proposed framework can be used to predict any performance-related characteristics of asphalt mixture including rutting, fatigue, low temperature behavior, etc. In other words, having several performance predictive models for asphalt mixture is no longer necessary. Besides, due to the high accuracy of the developed models in predicting pavement performance characteristics, the framework can be implemented to improve level 2 and level 3 inputs in the MEPDG design procedure. The proposed framework can make a stand-alone software for predicting pavement performance which is highly beneficial for asphalt agencies when a large amount of performance data is available.

Future research should address the capability of the model in handling larger data sets. Considering the high accuracy of the developed predictive models, these PCA-based approaches are strongly recommended as appropriate modeling approaches in this application. Moreover, these methodologies appear to be capable of modeling asphalt binder chemical properties as well as finding performance volumetric relationships and such investigations are recommended as future studies.

8.0 Authors Contribution

The authors confirm contribution to the paper as follows: study conception and design:

P. Ghasemi, D.K. Rollins; data collection: P. Ghasemi; analysis and interpretation of results: P. Ghasemi, M. Aslani, D.K. Rollins, R.C. Williams; draft manuscript preparation: P. Ghasemi; M. Aslani; D.K. Rollins; R.C. Williams. All authors reviewed the results and approved the final version of the manuscript. The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

9.0 References

Al-Khateeb, G., A. Shenoy, N. Gibson, and T. Harman, "A New Simplistic Model for Dynamic Modulus Predictions of Asphalt Paving Mixtures," *Journal of the Association of Asphalt Paving Technologists* Vol. 75, 2006, pp.1–40.

Andrei, D., M. W. Witczak, and M. W. Mirza, "Development of a Revised Predictive Model for the Dynamic (Complex) Modulus of Asphalt Mixtures," *Development of the 2002 Guide for the Design of New and Rehabilitated Pavement Structures,* NCHRP, 1999.

Apeagyei, A. K., "Rutting as a Function of Dynamic Modulus and Gradation," *American Society of Civil Engineers,* Vol 23, 2011, no. 9, pp. 1302–10.

ARA, "Calibration of Permanent Deformations Models for Flexible Pavements," *NCHRP 1-37A: Development of the 2002 Guide for the Design of New and Rehabilitated*

Pavement Structures, Transportation Research Board of the National Academies, Washington, D.C., 2004.

Aslani, M., P., Ghasemi, and A.H., Gandomi, "Constrained Mean-Variance Mapping Optimization for Truss Optimization Problems," *Structural Design of Tall and Special Buildings*, Vol 27, 2018, no. 6, e1449.

Ayres, M. Jr., and M. W. Witczak, "A Mechanistic Probabilistic System to Evaluate Flexible Pavement Performance," *Transportation Research Board*, Washington, D.C., 1998.

Bari, J., and M. W. Witczak, "New Predictive Models for Viscosity and Complex Shear Modulus of Asphalt Binders," *Transportation Research Record: Journal of the Transportation Research Board,* 2001, no. 1, 2007, pp. 9–19.

Bashin, A., E. Masad, M. E. Kutay, W. Buttlar, Y. Kim, M. Marasteanu, Y. R. Kim, C. W. Schwartz, and R. L. Carvalho, *Applications of Advanced Models to Understand Behavior and Performance of Asphalt Mixtures,* Transportation Research Board, Washington, DC., 2012.

Birgisson, B., R. Roque, J. Kim, and L. V. Pham, *The Use of Complex Modulus To Characterize Performance of Asphalt Mixtures and Pavements in Florida,* Florida Department of Transportation, 2004.

Cai, Z., and Y. Wang, "A Multiobjective Optimization-Based Evolutionary Algorithm for Constrained Optimization," *IEEE Transactions on Evolutionary Computation* Vol. 10, 2006, no. 6, pp. 658–675.

Cheng, B, and D.M. Titterington, "Neural Networks: A Review from a Statistical Perspective," *Statistical Science*, 1994, pp. 2–30.

Christensen Jr, D. W, T. Pellinen, and R. F. Bonaquist, "Hirsch Model for Estimating the Modulus of Asphalt Concrete," *Journal of the Association of Asphalt Paving Technologists* Vol. 72, 2003, pp. 97–121.

Devore, J., *Probability and Statistics for Engineering and the Sciences*, 2012.

Erlich, I., G. K. Venayagamoorthy, and N. Worawat, "A Mean-Variance Optimization Algorithm," *2010 IEEE World Congress on Computational Intelligence, WCCI,* 2010, pp. 1–6.

Fathi, A., M. Mazari, M. Saghafi, A. Hosseini, and S. Kumar, "Parametric Study of Pavement Deterioration Using Machine Learning Algorithms," *Innovation and Sustainability in Highway and Airfield Pavement Technology*, American Society of Civil Engineers, 2019, pp. 31–41.

Fodor, I. K., *A Survey of Dimension Reduction Techniques,* U.S. Department of Energy, 2002.

Ghasemi, P., M. Aslani, D. K. Rollins, R. C. Williams, and V.R. Schaefer, "Modeling Rutting Susceptibility Of Asphalt Pavement Using Principal Component Pseudo Inputs

In Regression And Neural Networks," *International Journal of Pavement Research and Technology*, Vol. 11, 2018, pp. 679–690.

Ghasemi, P., M. Aslani, D. K. Rollins, and R. C. Williams, "Principal Component Analysis-Based Predictive Modeling and Optimization of Permanent Deformation in Asphalt Pavement: Elimination of Correlated Inputs and Extrapolation in Modeling," *Structural and Multidisciplinary Optimization,* Vol. 59, 2019, no. 4, pp. 1335–1353.

He, J., and X. Yao, "From an Individual to a Population: An Analysis of the First Hitting Time of Population-Based Evolutionary Algorithms," *IEEE Transactions on Evolutionary Computation,* Vol. 6, 2002, no. 5, pp. 495–511.

Hosseini, A., A. Faheem, H. Titi, and S. Schwandt. "Evaluation of the long-term performance of flexible pavements with respect to production and construction quality control indicators," *Construction and Building Materials,* Vol. 230, 2020.

Jalali, F., A. Vargas-Nordcbeck, and M. Nakhaei, "Role of Preventive Treatments in Low-Volume Road Maintenance Program: Full-Scale Case Study," *Transportation Research Record: Journal of the Transportation Research Boar*d, 2019.

Jolliffe, I. T., *Principal Component Analysis, Second Edition,* Springer Series in Statistics, 2002.

Kaloush, K. E., M.W. Witczak, and B. W. Sullivan,"Simple Performance Test for Permanent Deformation Evaluation of Asphalt Mixtures," *Sixth International RILEM Symposium on Performance Testing and Evaluation of Bituminous Materials*, 2003, pp. 498–505.

Kaloush, K., and, M.W. Witczak, *Development of Permanent to Elastic Strain Ratio Model for Asphalt Mixtures. Development of the 2002 Guide for the Design of New and Rehabilitated Pavement Structure*, University of Maryland, College Park, MD, 1999.

Kvasnak, A., C., Robinette, and R.C., Williams, "A Statistical Development of a Flow Number Predictive Equation for the Mechanistic-Empirical Pavement Design Guide." *Transportation Research Board*, Washington, DC., 2007.

Leahy, R.B., *Permanent Deformation Characteristics of Asphalt Concrete*, Ph.D. Dissertation, University of Maryland, College Park, MD., 1989.

Neter, J., W. Wasserman, and M. H. Kutner, *Applied Linear Regression Models,* Irwin Homewood, IL, 1989.

Nobakht, M., and M. S. Sakhaeifar, "Dynamic Modulus and Phase Angle Prediction of Laboratory Aged Asphalt Mixtures," *Construction and Building Materials* Vol. 190, 2018, pp. 740–751.

Notani, M. A., P. Hajikarimi, F. Moghadas Nejad, and A. Khodaii, "Rutting resistance of toner-modified asphalt binder and mixture," *International Journal of Pavement Research and Technology*, 2019, pp. 1–9.

Majidifard, H., B. Jahangiri, W. G. Buttlar, and A. H. Alavi. "New machine learning-based prediction models for fracture energy of asphalt mixtures," *Measurement,* Vol.135, 2019, pp. 438–451.

Morovatdar, A., S.R. Ashtiani, C. Licon, and C. Tirado,"Development of a mechanistic approach to quantify pavement damage using axle load spectra from south Texas overload corridors," G*eo-Structural Aspects of Pavements, Railways, and Airfields Conference*, 2019, (In Press).

Pellinen, T., and M.W. Witczak,"Use of Stiffness of Hot-Mix Asphalt as a Simple Performance Test," *Transportation Research Record: Journal of the Transportation Research Board,* No. 1789, 2002, pp. 80–90.

Peng, C., J. Feng, S. Feiting, Z. Changjun, and F. Decheng, "Modified Two-Phase Micromechanical Model and Generalized Self-Consistent Model for Predicting Dynamic Modulus of Asphalt Concrete," *Construction and Building Materials* Vol. 201, 2019, pp. 33–41.

Refaeilzadeh, P., L. Tang, and H. Liu, "Cross-Validation," *Encyclopedia of Database Systems*, 2009, pp. 532–538.

Rodezno, M., K. E. Kaloush, and M. Corrigan, "Development of a Flow Number Predictive Model," *Transportation Research Record: Journal of the Transportation Research Board* 2181, 2010, pp. 79–87.

Rollins, D. K., D. Zhai, A. L. Joe, J. W Guidarelli, A. Murarka, and R. Gonzalez, "A Novel Data Mining Method to Identify Assay-Specific Signatures in Functional Genomic Studies," *BMC Bioinformatics,* Vol. 7, 2006, no.1, pp. 377–397.

Rollins, D. K.,"A One-Dimensional PCA Approach for Classifying Imbalanced Data," *Journal of Computer Science & Systems Biology*, Vol. 8, 2015, no. 1, pp. 5–11.

Sakhaeifar, M. S., Y. R. Kim, and B. E. Garcia Montano, "Individual Temperature Based Models for Nondestructive Evaluation of Complex Moduli in Asphalt Concrete," *Construction and Building Materials*, Vol. 137, 2017, pp. 117–27.

Sakhaeifar, M. S., Y. R. Kim, and P. Kabir, "New Predictive Models for the Dynamic Modulus of Hot Mix Asphalt," *Construction and Building Materials,* Vol. 76, 2015, pp. 221–231.

Shu, X., and B. Huang, "Dynamic Modulus Prediction of HMA Mixtures Based on the Viscoelastic Micromechanical Model," *Journal of Materials in Civil Engineering* , Vol. 20, 2008, no. 8, pp. 530–38.

Sobol, I. Y.M, "On Sensitivity Estimation for Nonlinear Mathematical Models," *Matematicheskoe Modelirovanie*, Vol. 2, 1990, no.1, pp. 112–18.

Sun, P., and R. M. Freund, "Computation of Minimum Volume Covering Ellipsoids," *Operational Research*, Vol. 52, 2004, no. 5, pp. 690–706.

Thiele, L., and E. Zitzler, "Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach," *IEEE Transactions on Evolutionary Computation*, Vol. 3, 1999, no. 4, pp. 257–271.

Thompson, B., *Exploratory and Confirmatory Factor Analysis: Understanding Concepts and Applications,* American Psychological Association, 2004.

Timm, D., I. Selvaraj, R. Brown, R. C. West, and A. Priest, NCAT Report 06-05*: Phase II NCAT Test Track Results,* 2006.

Todd, M. J., and E. A. Yildirim, "On Khachiyan's Algorithm for the Computation of Minimum-Volume Enclosing Ellipsoids," *Discrete Applied Mathematics,* Vol. 155, 2007, no. 13, pp. 1731-1744.

Wei, P., Z. Lu, and J. Song, "Variable Importance Analysis: A Comprehensive Review," *Reliability Engineering and System Safety*, Vol. 142, 2015, pp. 399–432.

Witczak, M. W., K. E. Kaloush, T. Pellinen, M. El-Basyouny, and H. V. Quintus, *Simple Performance Test for Superpave Mix Design,* Transportation Research Board, Washington, D.C., 2002.

Witczak, M. W., and M. El-Basyouny, NCHRP Report 1-37A- *Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures. Appendix A: Calibration of Permamnent Deformation Models For Flexible Pavements*, Transportation Research Board, Washington, DC, 2004.

Ziari, H., A. Amini, A. Goli, and D. Mirzaiyan. "Predicting rutting performance of carbon nano tube (CNT) asphalt binders using regression models and neural networks," *Construction and Building Materials,* Vol.160, 2018, pp. 415–426.