Hybrid Model of LightGBM Regression and Grid Search Optimization for the Estimation of Permanent Deformation of Asphalt Mixtures Pavements

Hoang Ha
University of Transport and Communications, Ha Noi, Vietnam
hoangha.utc@gmail.com

Tran Thi Thu Trang
Le Quy Don University, Ha Noi, Vietnam
trangtran1979@lqdtu.edu.vn

Dam Duc Nguyen
University of Transport Technology, Ha Noi, Vietnam
damnd@utt.edu.vn

Amjad Islam
Department of Civil Engineering, Iqra National University, Peshawar, Pakistan
amjadtaraki@gmail.com

Hieu Trung Tran
University of Transport Technology, Ha Noi, Vietnam
hieutrantrung@utt.edu.vn

Indra Prakash
DDG (R) Geological Survey of India, Gandhinagar, India
indra52prakash@gmail.com

Binh Thai Pham
University of Transport Technology, Ha Noi, Vietnam
binhpt@utt.edu.vn (corresponding author)

Received: 1 February 2025 | Revised: 5 March 2025 | Accepted: 9 March 2025

ABSTRACT

This study aimed to estimate the permanent deformation of Asphalt Mixtures (AMs) of pavements (Fn) utilizing a hybrid LGBM-GSO machine learning model, which combines Light Gradient-Boost Machine (LGBM) regression and Grid Search Optimization (GSO). In this study, input physical parameters, namely Filler (FP), fine aggregate (S), coarse aggregate (C), bitumen percent (BP), Marshall stability (M), Voids in Mineral Aggregate (VMA), air voids (Va), and Marshall flow (F) were used to predict Fn. Laboratory data from 118 AMs were analyzed. Model validation was carried out using various standard evaluation indicators, namely Mean Absolute Error (MAE), Root Mean Square Error (RMSE), R^2 (determination coefficient), and learning curve. The proposed LGBM-GSO model ($R^2 = 0.943$) performed well for the estimation of the Fn compared with the base model LGBM ($R^2 = 0.909$), showing that LGBM-GSO is an effective tool for accurate Fn estimation, and GSO is an effective optimization technique for LGBM to improve prediction performance.

Keywords-permanent deformation; asphalt mixtures; machine learning; grid search optimization; light gradient-boost machine

I. INTRODUCTION

The performance of asphalt concrete used in road and bridge pavements is largely influenced by the permanent deformation of the Asphalt Mixture (AM) layers [1]. Under repeated traffic loads, stress accumulates in these layers at the contact area between vehicle tires and the pavement surface, leading to progressive deformation. This phenomenon, commonly known as asphalt rutting, is particularly severe in high-temperature regions. Rutting significantly affects pavement durability, reducing the functional lifespan of roads, potentially compromising vehicle stability, and posing safety risks to highway users [2].

Since asphalt is a composite material, aggregates, which constitute over 90% of its weight and 80% of its volume, play a critical role in determining the extent of permanent deformation [3]. Among various aggregate properties, coarse aggregate geometry has a particularly strong influence on AM's deformation behavior [4]. The morphological characteristics of aggregates, such as angularity, are directly linked to the mechanical strength of AMs, affecting skeleton stability, interlocking force, shear resistance, and overall deformation resistance [3]. In addition to aggregate properties, permanent deformation in AMs is influenced by factors such as air voids, binder content, material selection, mix design, pavement structural design, environmental conditions (e.g., extreme temperatures), and overloading [5-7].

Various laboratory tests and analytical methods have been proposed to evaluate the permanent deformation behavior of AMs [8-10]. These include the softening point (ring and ball test) for hardening properties, penetration, resilience, and apparent viscosity tests for traditional asphalt 50/70 [11], triaxial repeated, wheel tracking, and recently developed partial triaxial tests and single penetration repeated shear test [12]. Various tests have been accelerated, including loading, static and dynamic creep, dynamic modulus, uniaxial and triaxial loading, and wheel tracking [2, 13, 14]. However, since the permanent deformation of AMs is affected by various complex and highly nonlinear elements, the development of a correlation model to effectively explain their performance has proved challenging. Early models based on mechanistic-empirical rutting studied the strains in the subgrade [15], involving some mechanistic approaches such as deflection level or computed strain, and regression models on permanent strain and strain ratio [16, 17]. HDM-III rutting performance models are considered the most well-known among other regression approaches, and several other terms could also be considered for environmental variables, mixture characteristics, and other factors [11]. In [18], a multiple linear regression model was used as a function of the mix volumetric, stiffness in terms of resilient modulus, and gradation parameters for predicting PD behavior. In [19], the rutting resistance of AM was examined as a function of gradation and dynamic modulus. Furthermore, statistical study of the connection between elastic and plastic compressive stresses determined by repeated-load uniaxial/triaxial testing often yields the general form of hot-mix asphalt rutting models [20].

Nowadays, machine learning (ML) methods are well-known as more advanced than conservative methods in prediction problems, due to their outstanding learning features in diverse research fields [21]. Predictive modeling can estimate the results of a given system using probability theory approaches and data mining tools. Each model is built utilizing various predictors that could have an impact on future events. In [22], machine learning-based models were used to predict AM rutting behavior. To prevent extrapolation, these models tried to specify the *n*-dimensional input space, taking into account the issues that were mostly ignored in previous pavement predictive models. Such frameworks can utilize optimization techniques to discover the optimal design. In [23], Multi-Objective Particle Swarm Optimization (MOPSO) was proposed for multivariate, multilevel optimization problems with multiple constraints. In [24], an automatic tool was developed to optimize AMs using the local experience of road agencies, an Artificial Neural Network (ANN), mix properties from construction projects, and an optimization strategy using a Genetic Algorithm (GA) to automate the mix design process.

Among ML models, Light Gradient-Boost Machine (LGBM) regression is an effective distributed gradient-boosting system. This method has been successfully utilized in a variety of research in various scientific fields and in the prediction of different modes of deformation and failure in materials and systems. As the accuracy of the LGBM model relies on its parameters, an optimization technique should be used to optimize and improve it. The main objective of this study is to evaluate the performance of a hybrid model, namely LGBM-GSO, a hybridization of LGBM and Grid Search Optimization (GSO), to estimate the permanent deformation of AM. A dataset of 118 asphalt mixture samples, including eight key input parameters, namely Filler (FP), fine aggregate (S), coarse aggregate (C), bitumen percent (BP), Marshall stability (M), voids in mineral aggregate (VMA), air voids (Va), and Marshall flow (F), was used to train and validate the model. Model performance was assessed using statistical indicators, namely Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the coefficient of determination (R^2). The Python programming language was used to implement and evaluate the model.

II. MATERIALS AND METHOD

A. Data Used

This study used data from [11], analyzing a total of 118 samples to model the permanent deformation of AM. These data considered one output (Fn), a dependent variable, and a set of inputs, including FP, S, C, BP, M, VMA, Va, and F, as independent variables. The reason for selecting these input parameters was presented specifically in [11]. To train and validate the models, the data were divided into two parts, one (70%) used for training and constructing them and the remaining (30%) used to validate their predictive ability. Figure 1 shows the histogram of the parameters used, and Figure 2 indicates the correlation analysis of the parameters.

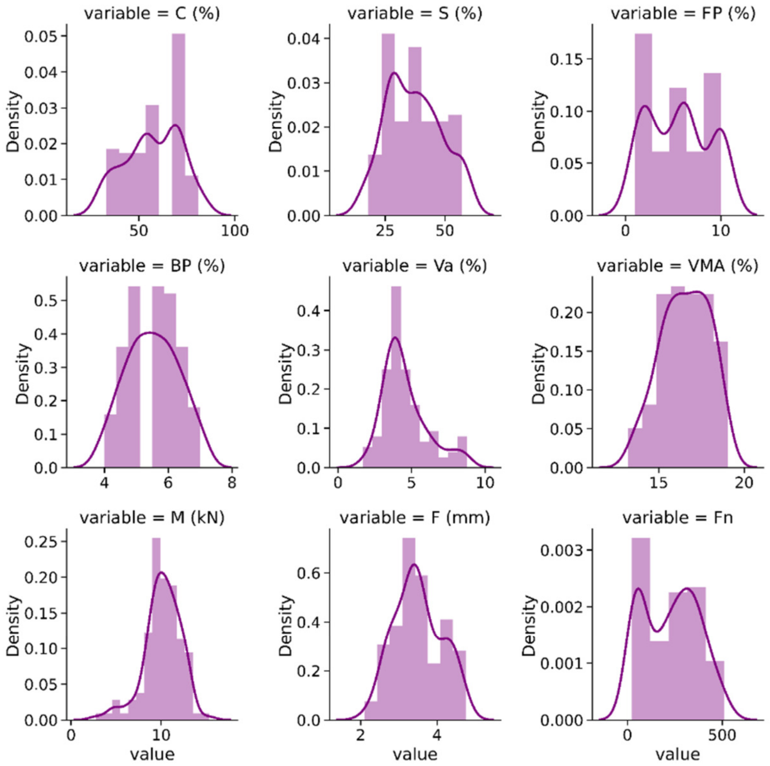
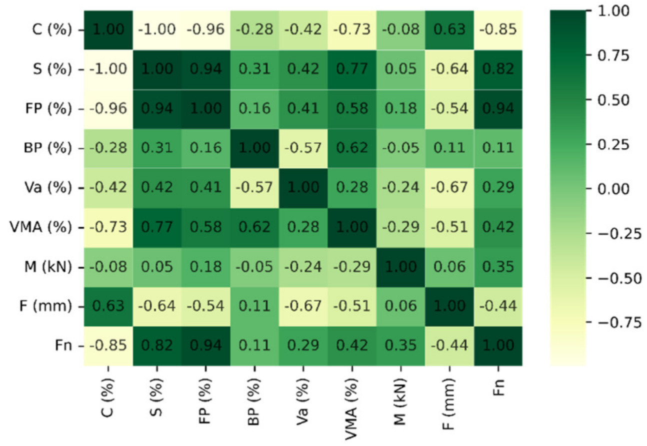


Fig. 1. Histogram of the variables used for modeling the permanent deformation of AM**.**



	C (%)	S (%)	FP (%)	BP (%)	Va (%)	VMA (%)	M (kN)	F (mm)	Fn
C (%)	1.00	-1.00	-0.96	-0.28	-0.42	-0.73	-0.08	0.63	-0.85
S (%)	-1.00	1.00	0.94	0.31	0.42	0.77	0.05	-0.64	0.82
FP (%)	-0.96	0.94	1.00	0.16	0.41	0.58	0.18	-0.54	0.94
BP (%)	-0.28	0.31	0.16	1.00	-0.57	0.62	-0.05	0.11	0.11
Va (%)	-0.42	0.42	0.41	-0.57	1.00	0.28	-0.24	-0.67	0.29
VMA (%)	-0.73	0.77	0.58	0.62	0.28	1.00	-0.29	-0.51	0.42
M (kN)	-0.08	0.05	0.18	-0.05	-0.24	-0.29	1.00	0.06	0.35
F (mm)	0.63	-0.64	-0.54	0.11	-0.67	-0.51	0.06	1.00	-0.44
Fn	-0.85	0.82	0.94	0.11	0.29	0.42	0.35	-0.44	1.00

Fig. 2. Correlation matrix analysis input variables in this study.

B. Methods Used

1) Light Gradient-Boost Machine (LGBM)

ML and artificial intelligence are the fastest-spreading fields in the modern world. LGBM is one of the most successful ML algorithms. As the accuracy of LGBM regression models is much better than that of others due to their extremely effective process and perfectly accurate performance, it has been widely employed in data modeling [25]. LGBM works on the principle of building a strong learner with the training data, including a large number of decision tree learners. Every new regression tree is generated by using the residual error of preceding weak learners. The final predictive output is then derived from the summation of projected value from several decision trees [25]. The advantages of LGBM include reduced memory utilization, robust training, speed, greater effectiveness, support of equivalent and GPU learning, ease of handling large-scale data, and enhanced precision [26]. In [27], these advantages were described as fast training speed by turning large feature values into discrete bins, which accelerates the training process. Replacing these discrete bins with continuous feature values requires low memory usage using a leaf-wise split approach, producing a more complex tree that achieves high accuracy. The training process is even faster because of its capacity to support features and data parallelism using GPUs. Although LGBM is one of the most efficient procedures and performs well on small datasets, it is sensitive to overfitting, especially in small datasets (with rows less than 10,000), as it splits leaf-wise and produces much more complex trees.

2) Cross Validation-based Grid Search Optimization (GSO)

GSO can help classifiers find the ideal model parameters as an alternative to successfully predict unlabeled testing data. This method can be categorized as exhaustive, as the procedure uses trials and iterations to explore the best parameter values. First, the prediction values are sorted. The method then shows the score for each parameter value to select for modeling [1]. The GSO is valid in the case the required maximum value is known to be within the upper and lower limits of each of the independent variables. To obtain the best values for model training, GSO can be applied along with cross-validation, which is a technique to resample existing data to assess machine learning models. Various pairs are examined and the one with the best cross-validation accuracy is selected. The advantages of the GSO approach are: (a) avoiding exhaustive search due to approximation heuristics, (ii) having only two parameters, and (iii) less computation time to find decent parameters by grid search [28]. In this study, GSO was used to optimize the hyperparameters used in LGBM to improve the model for the prediction of Fn.

C. Validation Indicators

1) Popular Statistical Validation Indicators

This study used popular indicators for the validation of regression analysis and comparison of models developed, including MAE, RMSE, and R^2. R^2 is defined as the proportion of the variation in the output which is predictable from the inputs, with values ranging from 0 to 1. A closer R^2 value to 1 shows better model performance. RMSE shows the differences between predicted and actual output values, while MAE measures the errors between predicted and actual output values. Lower values of MAE and RMSE show better accuracy of the model. The equations to calculate these indicators are [29-31]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_{i-}\ \check{y}_i\right| \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \check{y}_i)^2}{n}} \quad (2)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \check{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (3)$$

where, y_i denotes the actual values, $\hat{y}_i$ denotes the ML-predicted values, $\bar{y}$ is the mean of all actual values, and n is the sample size.

2) Learning Curve Analysis

Generally, a learning curve is a direct proportion on a graph that shows the relationship between a learner's performance on a task and the number of tries or times needed to finish it [32]. This work used learning curve analysis to evaluate the performance of the model with different training/testing ratios.

III. RESULTS AND DISCUSSION

A. Evaluation Comparison of the Models

LGBM-GSO was trained and validated using training and testing datasets. In the training process, the hyperparameters in LGBM, including max_depth, and learning_rate, were optimized using GSO, and their optimal values are presented in Table I. In addition, a 10-fold cross-validation was also used for training the LGBM-GSO to get the best fit of the model. Figures 3, 4, and 5 along with Table II show the validation results. Figure 3 shows that the performance of the models varied with the change in training data size, and it was the best with 70% training data and 30% test data. Therefore, it can be concluded that the best-split rate to divide the training and testing datasets is 70/30, which is also in line with other studies [33, 34].

TABLE I. BEST VALUES OF LGBM HYPERPARAMETERS (USING GSO)

No	Hyperparameters	LGBM-GSO
1	Maximum of depth	4
2	Learning rate	0.1
3	Number of cross-validations	10

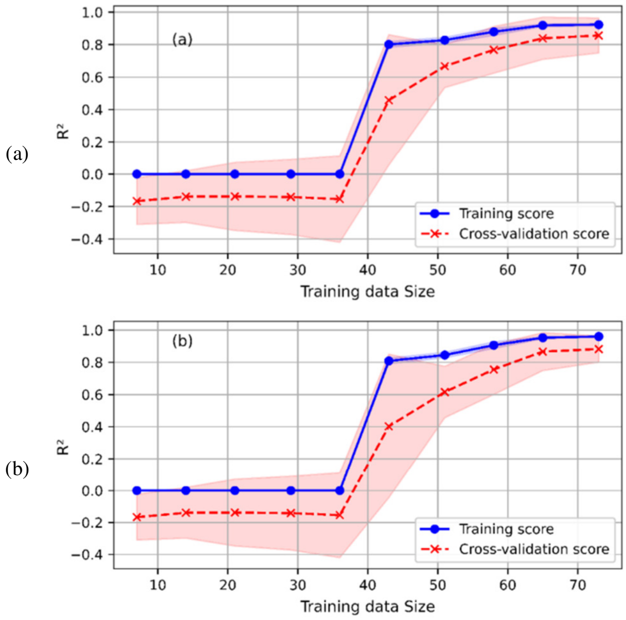


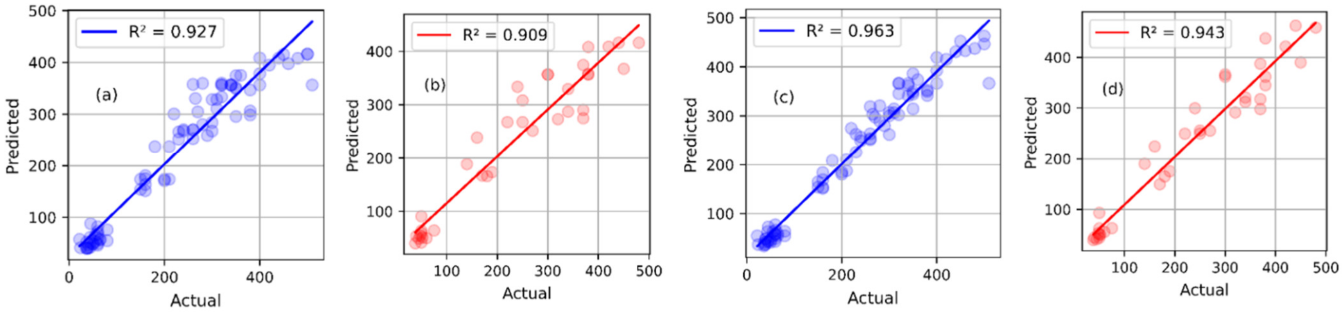
Fig. 3. Learning curve analysis of (a) LGBM and (b) LGBM-GSO.

Figure 4 shows the performance of the models using R^2, indicating that both models (LGBM and LGBM-BSO) have high R^2 values. More specifically, the LGBM R^2 values are 0.927 for training and 0.909 for testing, while the LGBM-GSO R^2 values are 0.963 for training and 0.943 for testing. When comparing the R^2 of these two models, those of LGBM-GSO are higher than those of the single LGBM model.

Figure 5 and Table II show the performance of the models using RMSE and MAE, indicating low error values for both of them. More specifically, the RMSE values of LGBM were 38.761 for training and 43.143 for testing while the RMSE values of LGBM-GSO were 27.634 for training and 43.143 for testing. The MAE values for the LGBM model were 27.299 and 32.624 for training and testing, respectively, whereas for the LGBM-GSO model, the MAE was 18.754 for training and 25.843 for testing. When comparing the RMSE and MAE values of these two models, the RMSE and MAE values of LGBM-GSO were lower than those of the single LGBM model.

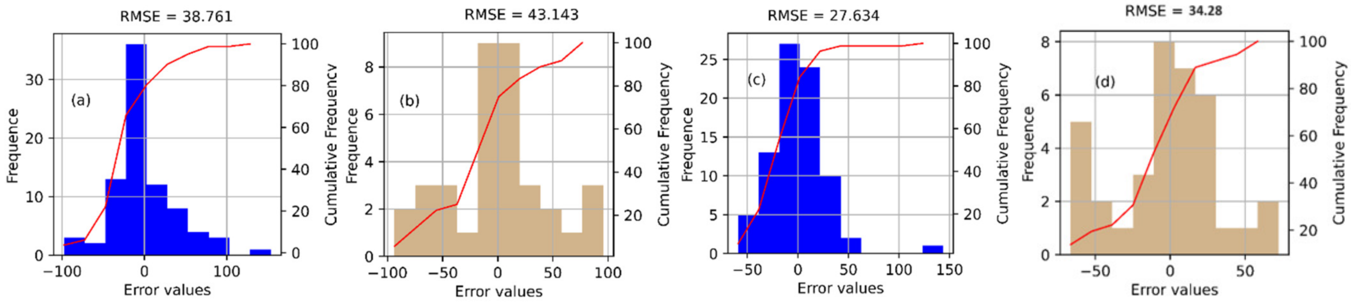
TABLE II. SUMMARY OF THE PERFORMANCE OF THE LGBM-GSO AND LGBM MODELS

Parameter	Train			Test		
	R^2	MAE	RMSE	R^2	MAE	RMSE
LGBM	0.927	27.299	38.761	0.909	32.624	43.143
LGBM-GSO	0.963	18.754	27.634	0.943	25.843	34.280



(a) (b) (c) (d)

Fig. 4. R^2 of the models: (a) training LGBM, (b) testing LGBM, (c) training LGBM-GSO, and (d) testing LGBM-GSO.



(a) (b) (c) (d)

Fig. 5. RMSE values of: (a) training LGBM, (b) testing LGBM, (c) training LGBM-GSO, and (d) testing LGBM-GSO.

The LGBM-GSO model demonstrated high accuracy in predicting the permanent deformation of AMs. The key reason for its superior performance lies in the combination of the strengths of LGBM and GSO. LGBM is widely recognized as one of the most effective ML techniques for predictive tasks due to its high parallelism efficiency, scalability with large datasets, and fast learning speed [25]. On the other hand, GSO is a powerful optimization technique that helps identify the optimal hyperparameters for machine learning models. It enhances model accuracy by allowing parallel processing during model training, improving the predictive capability of models [28]. This hybrid approach optimizes both the prediction algorithm and the hyperparameter tuning process, leading to a more robust and accurate model for predicting permanent deformation in asphalt mixtures.

When comparing with the performance of other models, namely Multi-Layer Perceptron (MLP) and Multi-Expression Programming (MEP) in [11] using the same data, the performance of LGBM-GSO ($R^2 = 0.943$) is equivalent to MEP and MLP.

B. Importance of Input Variables Used

To evaluate the importance of input variables in the estimation of Fn, one of the most popular feature selection methods, namely the Pairwise Correlation-based method (PC), was selected. The main principle of this method is that good feature datasets include variables that are highly correlated [35]. Figure 5 shows the results of feature selection, indicating that of the eight input variables, FP (AM = 6.214), C (AM = 6.036) and S (AM = 5.844) are the most important factors for the prediction of the permanent deformation of AM, followed by M (AM = 4.531), VMA (AM = 4.437), F (AM = 4.36), Va (AM = 4.16), and BP (AM = 3.957), respectively. These feature selection results show that out of eight input variables, the six input variables C, M, S, VMA, Va, F, F, and BP have contributed to the predictive capacity of the model for correct prediction of permanent deformation of AM.

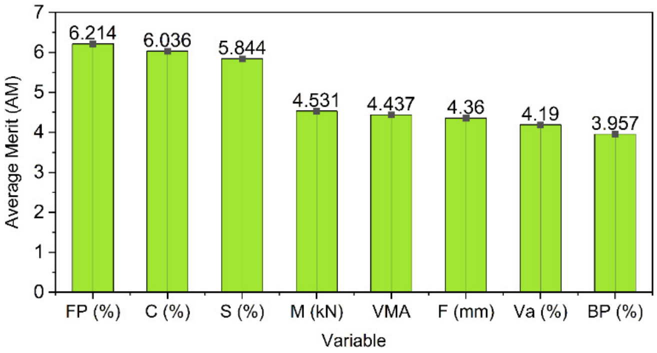


Fig. 6. Feature importance using LGBM-GSO.

IV. CONCLUSION

This study used an LGBM-GSO model to accurately predict the permanent deformation (Fn) of AMs, a crucial parameter in pavement design to mitigate rutting in flexible road and bridge pavements. The model was trained using laboratory data from 118 asphalt mixtures, incorporating eight key input parameters: C, M, VMA, Va, F, FP, BP, and S [11]. Validation using statistical metrics such as R^2, RMSE, MAE, and learning curve analysis confirmed that the LGBM-GSO model ($R^2 = 0.943$) outperformed the base LGBM ($R^2 = 0.909$), demonstrating the effectiveness of GSO in improving model performance. Looking ahead, there is potential for other ML models, such as Support Vector Machines, Bayesian Ridge Regression, and Random Forest, for further comparative studies in AM prediction. Additionally, feature selection techniques such as ANOVA or Extra Tree Regressor (ETR) can be used to identify the most informative subset of input features, addressing multicollinearity issues and enhancing model interpretability. In general, this study underscores the importance of hybrid ML techniques in pavement engineering and highlights the effectiveness of GSO in optimizing predictive models for AM properties.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are available in [11].

REFERENCES

[1] T. L. Nguyen, V. P. Le, Q. P. Nguyen, G. B. Dang, and M. H. N. Nguyen, "Performance Evaluation of Asphalt Mixtures with Rediset LQ-1200 Additive," *Engineering, Technology & Applied Science Research*, vol. 14, no. 4, pp. 15724–15728, Aug. 2024, https://doi.org/10.48084/etasr.7848.

[2] L. P. T. L. Fontes, G. Trichês, J. C. Pais, and P. A. A. Pereira, "Evaluating permanent deformation in asphalt rubber mixtures," *Construction and Building Materials*, vol. 24, no. 7, pp. 1193–1200, Jul. 2010, https://doi.org/10.1016/j.conbuildmat.2009.12.021.

[3] D. Zhang, L. Gu, and J. Zhu, "Effects of Aggregate Mesostructure on Permanent Deformation of Asphalt Mixture Using Three-Dimensional Discrete Element Modeling," *Materials*, vol. 12, no. 21, Jan. 2019, Art. no. 3601, https://doi.org/10.3390/ma12213601.

[4] J. Ren, Y. Xu, J. Huang, Y. Wang, and Z. Jia, "Gradation optimization and strength mechanism of aggregate structure considering macroscopic and mesoscopic aggregate mechanical behaviour in porous asphalt mixture," *Construction and Building Materials*, vol. 300, Sep. 2021, Art. no. 124262, https://doi.org/10.1016/j.conbuildmat.2021.124262.

[5] M. F. Oliveira, Bessa I. S., Vasconcelos R. R., Vasconcelos K. L., and L. L. B. Bernucci, "A new approach to laboratory roller compaction method and its influence on surface texture and permanent deformation of asphalt mixtures," *International Journal of Pavement Engineering*, vol. 23, no. 11, pp. 3867–3878, Sep. 2022, https://doi.org/10.1080/10298436.2021.1924377.

[6] M. Junaid, M. Irfan, S. Ahmed, and Y. Ali, "Effect of binder grade on performance parameters of asphaltic concrete paving mixtures," *International Journal of Pavement Research and Technology*, vol. 11, no. 5, pp. 435–444, Sep. 2018, https://doi.org/10.1016/j.ijprt.2017.11.006.

[7] Y. Ali, M. Irfan, S. Ahmed, and S. Ahmed, "Empirical Correlation of Permanent Deformation Tests for Evaluating the Rutting Response of Conventional Asphaltic Concrete Mixtures," *Journal of Materials in Civil Engineering*, vol. 29, no. 8, Aug. 2017, Art. no. 04017059, https://doi.org/10.1061/(ASCE)MT.1943-5533.0001888.

[8] D. M. Pires, S. L. Schuster, L. P. Specht, D. da Silva Pereira, and S. A. Biancardo, "Study of the permanent deformation of asphalt mixtures in the field: A multiscale approach," *Construction and Building Materials*, vol. 325, Mar. 2022, Art. no. 126763, https://doi.org/10.1016/j.conbuildmat.2022.126763.

[9] L. P. Leon and D. Gay, "Gene expression programming for evaluation of aggregate angularity effects on permanent deformation of asphalt mixtures," *Construction and Building Materials*, vol. 211, pp. 470–478, Jun. 2019, https://doi.org/10.1016/j.conbuildmat.2019.03.225.

[10] A. Ghanbari, B. S. Underwood, and Y. R. Kim, "Development of a rutting index parameter based on the stress sweep rutting test and permanent deformation shift model," *International Journal of Pavement Engineering*, vol. 23, no. 2, pp. 387–399, Jan. 2022, https://doi.org/10.1080/10298436.2020.1748190.

[11] M. R. Mirzahosseini, A. Aghaeifar, A. H. Alavi, A. H. Gandomi, and R. Seyednour, "Permanent deformation analysis of asphalt mixtures using soft computing techniques," *Expert Systems with Applications*, vol. 38, no. 5, pp. 6081–6100, May 2011, https://doi.org/10.1016/j.eswa.2010.11.002.

[12] Y. Xu and L. Sun, "Study on Permanent Deformation of Asphalt Mixtures by Single Penetration Repeated Shear Test," *Procedia - Social and Behavioral Sciences*, vol. 96, pp. 886–893, Nov. 2013, https://doi.org/10.1016/j.sbspro.2013.08.101.

[13] A. Khodaii and A. Mehrara, "Evaluation of permanent deformation of unmodified and SBS modified asphalt mixtures using dynamic creep test," *Construction and Building Materials*, vol. 23, no. 7, pp. 2586–2592, Jul. 2009, https://doi.org/10.1016/j.conbuildmat.2009.02.015.

[14] J. Jiang, F. Ni, L. Gao, and S. Lou, "Developing an optional multiple repeated load test to evaluate permanent deformation of asphalt mixtures based on axle load spectrum," *Construction and Building Materials*, vol. 122, pp. 254–263, Sep. 2016, https://doi.org/10.1016/j.conbuildmat.2016.05.006.

[15] J. F. Shook, F. N. Finn, M. W. Witczak, and C. L. Monismith, "Thickness Design of Asphalt Pavements – The Asphalt Institute Methods," presented at the 5th International Conference on the Design of Asphalt Pavements, 1982.

[16] Y. Ali, M. Irfan, S. Ahmed, and S. Ahmed, "Permanent deformation prediction of asphalt concrete mixtures – A synthesis to explore a rational approach," *Construction and Building Materials*, vol. 153, pp. 588–597, Oct. 2017, https://doi.org/10.1016/j.conbuildmat.2017.07.105.

[17] M. Irfan, Y. Ali, S. Iqbal, S. Ahmed, and I. Hafeez, "Rutting Evaluation of Asphalt Mixtures Using Static, Dynamic, and Repeated Creep Load Tests," *Arabian Journal for Science and Engineering*, vol. 43, no. 10, pp. 5143–5155, Oct. 2018, https://doi.org/10.1007/s13369-017-2982-4.

[18] H. Bin Tahir, M. Irfan, A. Hussain, Y. Ali, and E. Hussain, "Predicting the permanent deformation behaviour of the plant produced asphalt concrete mixtures: A first order regression approach," *Construction and Building Materials*, vol. 189, pp. 629–639, Nov. 2018, https://doi.org/10.1016/j.conbuildmat.2018.08.164.

[19] A. K. Apeagyei, "Rutting as a Function of Dynamic Modulus and Gradation," *Journal of Materials in Civil Engineering*, vol. 23, no. 9, pp. 1302–1310, Sep. 2011, https://doi.org/10.1061/(ASCE)MT.1943-5533.0000309.

[20] H. Wang and I. L. Al-Qadi, "Impact Quantification of Wide-Base Tire Loading on Secondary Road Flexible Pavements," *Journal of Transportation Engineering*, vol. 137, no. 9, pp. 630–639, Sep. 2011, https://doi.org/10.1061/(ASCE)TE.1943-5436.0000245.

[21] M. Y. Cheng, D. Prayogo, and Y. W. Wu, "A self-tuning least squares support vector machine for estimating the pavement rutting behavior of asphalt mixtures," *Soft Computing*, vol. 23, no. 17, pp. 7755–7768, Sep. 2019, https://doi.org/10.1007/s00500-018-3400-x.

[22] P. Ghasemi, M. Aslani, D. K. Rollins, and R. C. Williams, "Principal Component Neural Networks for Modeling, Prediction, and Optimization of Hot Mix Asphalt Dynamics Modulus," *Infrastructures*, vol. 4, no. 3, Sep. 2019, Art. no. 53, https://doi.org/10.3390/infrastructures4030053.

[23] C. Liang *et al.*, "Machine Learning Approach to Develop a Novel Multi-Objective Optimization Method for Pavement Material Proportion," *Applied Sciences*, vol. 11, no. 2, Jan. 2021, Art. no. 835, https://doi.org/10.3390/app11020835.

[24] H. Sebaaly, S. Varma, and J. W. Maina, "Optimizing asphalt mix design process using artificial neural network and genetic algorithm," *Construction and Building Materials*, vol. 168, pp. 660–670, Apr. 2018, https://doi.org/10.1016/j.conbuildmat.2018.02.118.

[25] Y. Hou, Z. Zhang, P. Liu, C. Song, and Z. Wang, "Research on a novel data-driven aging estimation method for battery systems in real-world electric vehicles," *Advances in Mechanical Engineering*, vol. 13, no. 7, Jul. 2021, Art. no. 16878140211027735, https://doi.org/10.1177/16878140211027735.

[26] Z. Zhang, C. Wang, Y. Gao, J. Chen, and Y. Zhang, "Short-Term Passenger Flow Forecast of Rail Transit Station Based on MIC Feature Selection and ST-LightGBM considering Transfer Passenger Flow," *Scientific Programming*, vol. 2020, no. 1, 2020, Art. no. 3180628, https://doi.org/10.1155/2020/3180628.

[27] E. A. Minastireanu and G. Mesnita, "Light GBM Machine Learning Algorithm to Online Click Fraud Detection," *Journal of Information Assurance & Cybersecurity*, pp. 1–12, Apr. 2019, https://doi.org/10.5171/2019.263928.

[28] Y. Sun, S. Ding, Z. Zhang, and W. Jia, "An improved grid search algorithm to optimize SVR for prediction," *Soft Computing*, vol. 25, no. 7, pp. 5633–5644, Apr. 2021, https://doi.org/10.1007/s00500-020-05560-w.

[29] S. Pal, L. H. Trang, V. T. Hieu, D. D. Nguyen, D. Q. Vu, and I. Prakash, "Investigation of Support Vector Machines with Different Kernel Functions for Prediction of Compressive Strength of Concrete," *Journal of Science and Transport Technology*, pp. 55–68, Jun. 2024, https://doi.org/10.58845/jstt.utt.2024.en.4.2.55-68.

[30] M. V. Le, I. Prakash, and D. D. Nguyen, "Predicting Load-Deflection of Composite Concrete Bridges Using Machine Learning Models," *Journal of Science and Transport Technology*, pp. 43–51, Dec. 2023, https://doi.org/10.58845/jstt.utt.2023.en.3.4.43-51.

[31] I. Prakash, D. D. Nguyen, N. T. Tuan, T. V. Phong, and L. V. Hiep, "Landslide Susceptibility Zoning: Integrating Multiple Intelligent Models with SHAP Analysis," *Journal of Science and Transport Technology*, pp. 23–41, Mar. 2024, https://doi.org/10.58845/jstt.utt.2024.en.4.1.23-41.

[32] T. Viering and M. Loog, "The Shape of Learning Curves: A Review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 6, pp. 7799–7819, Jun. 2023, https://doi.org/10.1109/TPAMI.2022.3220744.

[33] S. Yang and H. Zhang, "Comparison of Several Data Mining Methods in Credit Card Default Prediction," *Intelligent Information Management*, vol. 10, no. 05, Sep. 2018, Art. no. 115, https://doi.org/10.4236/iim.2018.105010.

[34] B. T. Pham and I. Prakash, "A novel hybrid model of Bagging-based Naïve Bayes Trees for landslide susceptibility assessment," *Bulletin of Engineering Geology and the Environment*, vol. 78, no. 3, pp. 1911–1925, Apr. 2019, https://doi.org/10.1007/s10064-017-1202-5.

[35] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, University of Waikato, Hamilton, New Zealand, 1999.