Estimation of Permeability of Soil-Fly Ash Mix using Machine Learning Algorithms

Galupino, Joenel
Department of Civil Engineering, De La Salle University : Academic Service Faculty

Dungca, Jonathan
Department of Civil Engineering, De La Salle University : Full Professor

https://doi.org/10.5109/7157938

出版情報：Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 9, pp.28-33, 2023-10-19. 九州大学大学院総合理工学府
バージョン：

Estimation of Permeability of Soil-Fly Ash Mix using Machine Learning Algorithms

Joenel Galupino[1*], Jonathan Dungca[2]

[1]Academic Service Faculty, Department of Civil Engineering, De La Salle University, Manila, Philippines
[2] Full Professor, Department of Civil Engineering, De La Salle University, Manila, Philippines [3]Third affiliation
*Corresponding author email: joenel.galupino@dlsu.edu.ph

Abstract: *This study demonstrates the potential of machine learning to predict the permeability of soil-fly ash mixtures, thereby promoting fly ash as a sustainable building material. Due to its environmental benefits and enhanced engineering properties when added to mixtures, fly ash, a byproduct of coal combustion, is gaining popularity. Several machine learning algorithms were evaluated, with the linear regression model proving to be the most precise and straightforward. It captured the linear relationship between percentage of fly ash and permeability (RMSE of 6.42 x 10^{-06}cm/s and R^2 of 0.811). Training and testing of models utilized a comprehensive database of soil-fly ash mixtures. The implications of the study's findings for engineering and environmental applications are substantial. The model's accuracy in estimating soil-fly ash mixture permeability is validated by the excellent correlation between predicted and actual permeability values.*

Keywords: Permeability; Index Properties; Fly Ash; Silty Sand; Machine Learning;

1. INTRODUCTION

In recent years, the construction industry has undergone a paradigm shift toward environmentally friendly and sustainable practices [1] [2] [3]. As the world faces the challenges of climate change and declining natural resources, there is a growing need to investigate alternative materials that can reduce the ecological footprint of construction projects [4]. The unique properties and environmental benefits of fly ash, a byproduct of coal combustion in power plants, make it a promising construction material [5]. By incorporating machine learning techniques, it is possible to estimate the permeability characteristics of soil-fly ash mixtures.

Before applying machine learning to estimate permeability characteristics, it is essential to comprehend the environmental effects of fly ash and its viability as a building material. Fly ash is a powdery substance produced in coal-fired power plants during the combustion of pulverized coal [5]. Historically, fly ash was viewed as a waste product requiring extensive disposal measures. Despite this, the construction industry has recognized fly ash as a valuable resource due to a growing awareness of its potential environmental impacts [6].

Fly ash has several characteristics that make it an environmentally viable alternative. In the first place, its use reduces the demand for conventional cement, a major contributor to carbon dioxide emissions. By substituting a portion of cement with fly ash, the production of concrete can significantly reduce its carbon footprint. In addition, fly ash improves construction materials' workability, durability, and chemical resistance [7]. Its application can reduce water permeability, reduce alkali-silica reaction, and increase the strength of concrete structures.

In numerous engineering applications, including embankments, road pavements, and landfill liners, the permeability characteristics of soil-fly ash mixtures are crucial [8]. Accurately estimating permeability is crucial for ensuring the stability and long-term performance of these structures. Traditional laboratory-based methods for determining permeability are time-consuming, costly, and frequently constrained by sample availability [9].

Machine learning techniques offer a promising alternative for overcoming such challenges. It is possible to develop robust and effective algorithms for estimating the permeability characteristics of soil-fly ash mixtures by taking advantage of the power of artificial intelligence and data-driven models. Large datasets containing geotechnical properties, particle size distribution, and fly ash content, among other variables, can be analyzed by machine learning algorithms to establish patterns and relationships [10].

Using machine learning to estimate permeability characteristics has several advantages over traditional methods. First, machine learning algorithms can process vast quantities of data quickly and efficiently, enabling the analysis of intricate relationships and patterns. This capability enables accurate permeability predictions in situations where conventional models fail due to inherent assumptions and limitations.

Second, machine learning models can adapt and learn from new data, enhancing their accuracy and performance continuously over time [11]. As more data from field measurements and laboratory tests becomes available, the models can be updated and refined to improve their predictive capabilities. This iterative procedure ensures that estimates of permeability remain relevant and accurate, even in environments of dynamic construction.

The incorporation of machine learning techniques into the estimation of the permeability characteristics of soil-fly ash mixtures is a significant step toward sustainable construction practices and environmental preservation.

The use of fly ash in construction results in enhanced performance characteristics. The addition of fly ash to soil mixtures improves their stability, decreases their permeability, and increases their chemical resistance. These characteristics make fly ash an ideal material for applications such as embankments, road pavements, and landfill liners, where the permeability characteristics of the soil-fly ash mixtures are crucial to the long-term performance.

Learning machine techniques provide an effective method for accurately estimating permeability characteristics. By analyzing large datasets containing geotechnical properties, particle size distribution, and fly ash content, machine learning algorithms can identify patterns and relationships that are challenging to determine using traditional methods. These algorithms can then generate models with high accuracy for predicting permeability.

The benefits of machine learning extend beyond the accurate estimation of permeability. The adaptability and self-learning capabilities of these algorithms allow them to improve their performance continuously over time. As new data from field measurements and laboratory tests become available, the models can be updated to improve their predictive capabilities and ensure that the permeability estimates remain accurate and relevant.

Thus, the construction industry has been adopting sustainable practices in response to climate change and depletion of natural resources. Fly ash, a coal combustion byproduct, has emerged as a promising environmentally friendly building material. It reduces the demand for carbon-intensive cement and improves concrete's workability, durability, and chemical resistance. Permeability characteristics of soil-fly ash mixtures are essential for structural integrity and performance in a variety of engineering applications. Traditional methods for estimating permeability are time-consuming and expensive, but machine learning offers a promising alternative. By analyzing massive datasets, machine learning algorithms can accurately predict permeability and continuously enhance their performance with new information. This incorporation of machine learning into construction practices is an important step toward sustainable construction and environmental preservation, as it promotes the use of fly ash for improved construction performance.

2. METHODOLOGY

2.1 Soil Mixtures for Permeability Test

In the laboratory testing protocol, various soil and fly ash mixtures were formulated to determine the effect of fly ash on permeability. By adjusting the proportions of the constituents, these mixtures were created. The objective was to determine how the presence of fly ash affected the soil's permeability.

Table 1. Number of Specimens for Permeability Tests

Soil Mixture	No. of Specimens
100FA	24
75FA25S	24
50FA50S	24
25FA50S	24
100S	24

To precisely evaluate the impact of fly ash on the soil, various amounts of fly ash were incorporated into the mixtures. The following sums were applied: 0% (100S), 25% (25FA75S), 50% (50FA50S), 75% (75FA25S), and 100% (100FA). These ratios indicate the proportion of fly ash present in each mixture, with the remainder being soil (denoted "S"). Using these varying quantities of fly ash, the researchers sought to determine the effect of fly ash content on the permeability characteristics of soil.

2.2 Sample Preparation

To calculate the required mass for each dry soil mixture, the relative density requirement of 90% was utilized. The first step is to unscrew the permeameter's cap and upper chamber in order to remove them. Next, calculate the inside diameter of the permeameter by measuring the inside diameter of the chambers. Then, position porous stones inside the chamber's base and ensure that they are properly positioned. Place filter paper on top of the porous stones to create a barrier. Using a scoop and funnel, carefully pour the prepared soil in a circular motion into the lower chamber until it reaches a depth of 1.5 cm. Now, position the rubber gasket between the chambers before placing the upper chamber on the lower chamber. Utilize a tamping device and perform approximately 30 to 40 tamps per layer to compact the soil layer. Repeat this compacting procedure until all the soil meets the required 10 cm height. It is imperative to practice this step for precision. After the soil has been compacted, place another layer of filter paper followed by the upper porous stones on top of the soil. Place the compression spring on top of the upper porous stone, followed by the chamber cap and its sealing gasket. To firmly secure the cap, tighten the nuts by screwing them together. The final set-up is shown in Fig. 1.

Fig. 1. Permeability Sample

2.3 Permeability Test

The soil permeability test is a laboratory test that measures the rate at which water can flow through a soil sample under a hydraulic gradient. It provides valuable information regarding the soil's ability to allow water to pass through it, which is a crucial parameter for a variety of geotechnical and engineering applications [12]. In this study, falling head was employed. This test is similar to the constant head permeability test; however, instead of maintaining a constant hydraulic head, the head is allowed to fall at a known rate. In a permeameter, the time required for the water level to fall between two specific points is measured. Using the permeameter's dimensions and the time recorded, the permeability of the soil can be determined.

2.4 Machine Learning Modelling

Machine learning is a subfield of artificial intelligence (AI) concerned with the development of algorithms and models that enable computers to learn and make predictions or decisions without being explicitly programmed [13]. Utilizing statistical techniques to enable computer systems to automatically learn and improve based on experience or data. In this study, the following are used for modelling: decision tree, linear regression, quadratic regression, neural network, and ensemble methods.

A decision tree is a hierarchical model that uses a tree-like structure to make decisions or predictions by segmenting the data based on feature values and creating a series of if-then conditions that lead to a final outcome [14]. The objective of linear regression is to establish a linear relationship between a dependent variable and one or more independent variables using a line that best fits the data to make predictions or estimate the value of the dependent variable [15]. Quadratic regression is a type of regression analysis that uses a quadratic equation to model the relationship between a dependent variable and independent variables, allowing for a curved relationship between the variables [16]. A neural network is a computational model that emulates the structure and operation of the human brain. It is composed of interconnected nodes or artificial neurons organized in layers, which learn and process information in order to make predictions or solve complex problems by adjusting the weights between nodes [17]. Ensemble methods combine multiple individual models to enhance overall performance in terms of prediction. These methods, such as Random Forest and Gradient Boosting, aggregate predictions from multiple models, capitalizing on the diversity and collective intelligence of the individual models to make more accurate and robust predictions [18].

In this study, the dependent variable s permeability in cm/s, while the independent variable is the percentage of fly ash. The data split ratio is 70% training, 15% validation, and 15% testing.

2.5 Validation

Two evaluation metrics will be used to compare the models: coefficient of determination (R^2) and root mean square error (RMSE). R^2 measures the proportion of the variance in the dependent variable that is explained by the model, whereas RMSE quantifies the average deviation between the predicted and actual values [19]. Additionally, an equality line representing the ideal relationship between predicted and actual values will be generated for each model. This will provide a visual reference for evaluating the accuracy and performance of each model in relation to the equality line [20].

3. RESULTS AND DISCUSSIONS

3.1 Results of Permeability

Typically, the soil component comprises the majority of the soil mixture. If the soil is contaminated or does not meet the required specifications, it can be excavated and treated. Typically, the permeability of pure soil falls between 1.47×10^{-05} cm/s and 2.70×10^{-05} cm/s. In the case of the '100S' microfabric, which consists of extremely elongated grains, large angular grains, and an abundance of silt grains with a rough surface, these characteristics contribute to efficient drainage.

Fly ash is a recommended soil additive because it permits the recycling of waste materials. However, the addition of fly ash modifies the inter-particle void ratio, which is a crucial factor in the microscopic test for characterizing the '100F' mixture. This particular mixture consists of silt grains of varying sizes, which form the microfabric. The presence of silt particles of nearly identical sizes increases permeability by creating larger inter-particle voids. Typically, the permeability range for pure fly ash is between 1.93×10^{-05} cm/s and 7.29×10^{-05} cm/s.

The 75FA25S, 50FA50S, and 25FA75S mixtures consist of a combination of fly ash and soil. The microstructure of these mixtures is comprised of a combination of extremely elongated grains, large angular grains, and a significant presence of both larger and smaller silt grains with rough surfaces. As the percentage of fly ash in these mixtures increases, the drainage capacity also increases. While complete replacement of bentonite with a mixture of fly ash may not be feasible, it is worthwhile to consider substituting a certain amount of fly ash for bentonite in the cut-off mixture, as detailed in the following sections of the study. The addition of fly ash to soil mixtures increases the inter-particle voids, resulting in a permeability range of 1.93×10^{-05} to 5.02×10^{-05} cm/s, shown in Fig. 2.

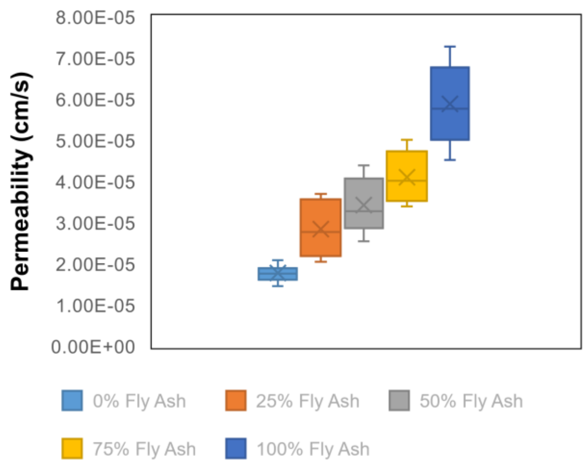

Fig. 2. Range of Permeability of Soil-Fly Ash Mixture

3.2 Machine Learning Models for Permeability

In this study, the following are used for modelling: decision tree, linear regression, quadratic regression, neural network, and ensemble methods.

The Linear Regression model emerges as the victor due to its superior performance in terms of both RMSE and R^2. It achieves the lowest RMSE value of 6.42×10^{-06}, which indicates minimal deviation between the predicted and actual values. This illustrates the model's precision in making accurate predictions.

Moreover, the Linear Regression model has the highest R^2 value, 0.811. This indicates that the model can explain approximately 81.1% of the variance in the target variable. The high R^2 value highlights the model's capacity to capture and account for the underlying relationships between input features and the target variable.

Comparatively, the Tree, Quadratic Regression, Ensemble, and Neural Network models trail the Linear Regression model slightly in terms of RMSE and R^2 values. Although these models exhibit adequate performance, they do not surpass the precision of the Linear Regression model. The responses in shown in Fig. 3.

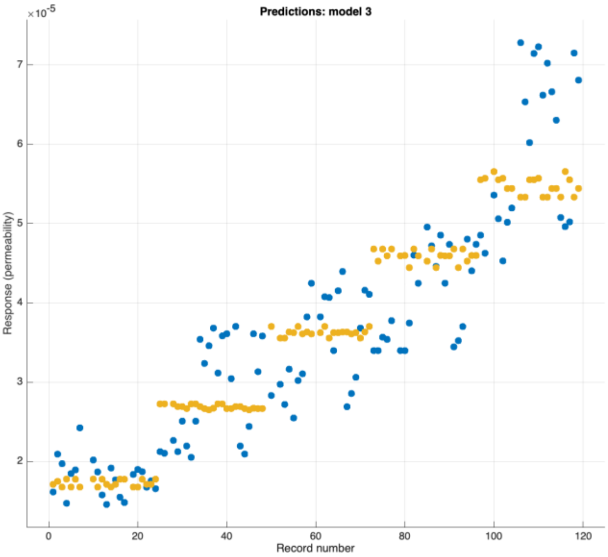

Fig. 3. Responses of ML Linear Model

On a given dataset, we compare the performance of multiple regression models. A Tree model, Linear Regression, Quadratic Regression, Neural Network, and Ensemble model are under consideration. The assessment is based on two key metrics: root mean squared error (RMSE) and the coefficient of determination (R^2). Table 2 contains the RMSE and R^2 values obtained for each model.

The RMSE of the tree model is 6.85×10^{-06} cm/s, and its R^2 is 0.796. It captures intricate data interactions and patterns with a high degree of predictive accuracy. In terms of RMSE and R^2, it lags slightly behind the Linear Regression model, despite its excellent performance.

The Linear Regression model has an impressive RMSE of 6.42×10^{-06} cm/s and R^2 of 0.811. These results indicate that the model is a good fit for the data, explaining a substantial portion of the target variable's variance. Its performance is superior to that of the Tree model, making it a formidable rival.

The RMSE of the Quadratic Regression model is 7.22×10^{-06} cm/s and the R^2 is 0.774. Incorporating quadratic terms, this model represents nonlinear relationships. However, it performs marginally worse than both the Tree and Linear Regression models.

The Neural Network model has an RMSE of 0.000780034 and an R^2 of -2639.049, which is an extremely low negative value. These results indicate that the model does not adequately represent the data and may be affected by overfitting or other issues. It significantly underperforms the other models evaluated.

The Ensemble model yields an RMSE of 7.04×10^{-06} cm/s and an R^2 of 0.785 respectively. It combines multiple predictive models to improve performance. In terms of R-squared, it performs adequately but is slightly inferior to the Linear Regression model.

Based on the evaluation of RMSE and R^2 values, the best-performing model is the Linear Regression model. It has the lowest RMSE, which indicates minimal deviation from the actual values, and the highest R^2, which indicates a strong ability to explain the variance in the target variable. The Linear Regression model performs better than the other models, including the Tree model, which is renowned for its ability to capture complex interactions. When selecting the most appropriate model for a particular use case, it is necessary to also consider other factors, such as interpretability and computational complexity.

Table 2. Summary of Model Performances

Model	RMSE	R^2
Tree	6.85E-06	0.79634207
Linear Regression	6.42E-06	0.81095208
Quadratic Regression	7.22E-06	0.7739044
Neural Network	7.80E-04	0.0037892
Ensemble	7.04E-06	0.78502235

In the given data, we compared the stability of the performance of various models (Tree, Linear Model, Quadratic Model, Ensemble, and Neural Network). Stability is the consistency of the model's predictions when presented with identical inputs or data.

Observing the provided data, we can see that the linear model consistently predicts values within a small range for each and every data point. Fly Ash percentage predictions from the linear model range between 1.78×10^{-05} cm/s and 5.89×10^{-05} cm/s, indicating a stable behavior.

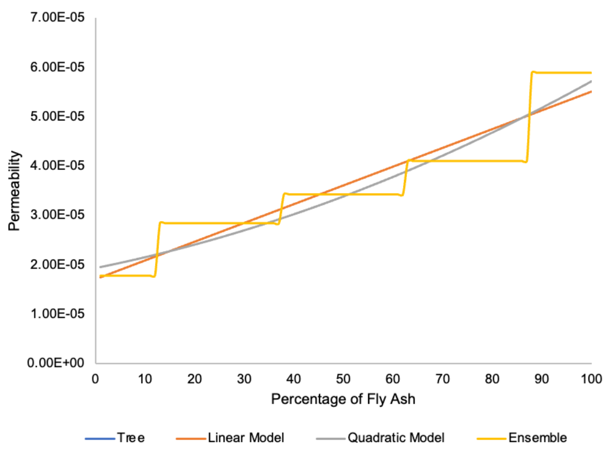

Fig. 4. Parametric Analysis of each model

In contrast, the predictions of the other models (Quadratic Model, Ensemble, and Neural Network) are more variable. For instance, the predictions of the quadratic model range from 1.75×10^{-05} cm/s to 5.05×10^{-05} cm/s, those of the ensemble model range from 1.95×10^{-05} cm/s to 5.08×10^{-05} cm/s, and those of the neural network range from -0.000775689 to 0.004707021. These larger ranges imply that these models are more sensitive to the input data and may not generate consistent predictions for similar inputs.

The linear model's stability is attributable to its simplicity and linearity. The linear relationship between the input variables and the target variable is assumed by linear models. This simplicity makes the linear model less susceptible to overfitting and more tolerant of input data variations.

In practical applications, stability is an essential model characteristic. When presented with new data that is similar to the training data, a stable model is less likely to produce erroneous or unexpected outcomes. This can be especially important in domains where consistency and reliability are crucial, such as financial forecasts, medical diagnoses, and safety-critical systems.

Based on the provided data, we can conclude that the linear model is more stable than the other evaluated models.

The linear model successfully derived an equation, denoted by Eq. 1, which accurately predicts the permeability of the soil mixture based on its fly ash content. This equation demonstrates the capability of the model to establish a consistent relationship between these two variables, shown in Eq. 1.

$$k = 3.789x10^{-07}FA + 1.71410^{-05} \quad (1)$$

Where,
k = Permeability in cm/s;
FA = Percentage of Fly Ash in %.

3.3 Validation

Examining the equality line on a scatter plot where the x-axis represents the observed permeability values and the y-axis represents the predicted permeability values can assess the linear model's validity. In this instance, the equality line is a 45-degree line where the values of x and y are equal, shown in Fig. 5.

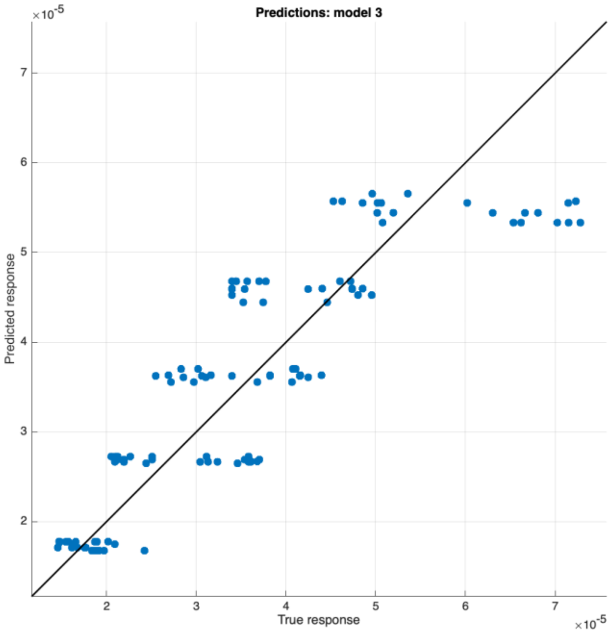

Fig. 5. Equality Line of the linear model

On the scatter plot of observed permeability values versus predicted permeability values, the data points corresponding to the experimental percentages of 0%, 25%, 50%, 75%, and 100% are still prevalent. On the scatter plot, these specific percentages form distinct clusters or points.

4. CONCLUSIONS

This study concludes by highlighting the potential of machine learning techniques for predicting the permeability of soil-fly ash mixtures and the significance of using fly ash as a sustainable building material. By incorporating fly ash into soil mixtures, not only can waste from power plant fly ash be managed effectively, but the engineering properties of the mixtures can also be enhanced.

To estimate the permeability of soil-fly ash mixtures, various machine learning algorithms, including decision tree, linear regression, quadratic regression, neural network, and ensemble methods, were investigated. Compared to other models, the linear regression model demonstrated superior predictive accuracy and simplicity. With an RMSE value of 6.42 x10^{-06} cm/s and an R2 value of 0.811, it accurately captured the linear relationship between the percentage of fly ash and permeability.

The equality line analysis further validated the linear model's performance by comparing the predicted and actual permeability values. The scatter plot revealed that the data points were closely aligned with the equality line, indicating that the linear model accurately estimated the permeability of soil-fly ash mixtures.

This study highlights the potential of machine learning in predicting the permeability of soil mixtures and emphasizes the significance of incorporating sustainable materials such as fly ash into construction practices. The results contribute to a greater understanding of soil-fly ash mixtures and provide important insights for engineering and environmental applications.

5. REFERENCES

[1] C. Maraveas, "Production of sustainable construction materials using agro-wastes," *Materials*, 2020.

[2] K. Sindol, A. Gadat and J. Sanchez, "Development and Characterization of Briquettes made from Unsalable Banana Peel Wastes: A Preliminary Evaluation," in *IEICES Proceedings*, Kyushu University, 2022.

[3] S. Zuraida, B. Dewancker, R. Bramantyo Margono and M. Irfan, "Study on prefabrication method for housing in Indonesia by utilization undegradable waste for building material," in *IEICES Proceedings*, Kyushu University, 2022.

[4] J. Correa, J. Montalvo-Navarrete and M. Hidalgo-Salazar, "Carbon footprint considerations for biocomposite materials for sustainable products: A review," *Journal of Cleaner Production*, 2019.

[5] G. Xu and X. Shi, "Characteristics and applications of fly ash as a sustainable construction material: A state-of-the-art review," *Resources, Conservation and Recycling*, 2018.

[6] A. Bieliatynskyi, S. Yang and V. Pershakov, "Study of crushed stone-mastic asphalt concrete

using fiber from fly ash of thermal power plants," *Case Studies in Construction Materials,* p. 16, 2022.

[7] Y. Cho, S. Jung and Y. Choi, "Effects of chemical composition of fly ash on compressive strength of fly ash cement mortar," *Construction and Building Materials,* vol. 204, p. 255–264, 2019.

[8] J. Galupino and J. Dungca, "Permeability characteristics of soil-fly ash mix," *ARPN Journal of Engineering and Applied Sciences,* vol. 10, no. 15, p. 6440–6447, 2015.

[9] J. Dungca and J. Galupino, "Modelling of permeability characteristics of soil-fly ash-bentonite cut-off wall using response surface method," *International Journal of GEOMATE,* vol. 10, no. 4, p. 2018–2024, 2016.

[10] J. Galupino and J. Dungca, "Machine learning models to generate a subsurface soil profile: a case of Makati City, Philippines," *International Journal of GEOMATE,* vol. 23, no. 95, p. 57–64, 2022.

[11] C. Pylianidis, V. Snow and H. Overweg, "Simulation-assisted machine learning for operational digital twins," *Environmental Modelling and Software,* p. 148, 2022.

[12] G. Wrzesiński, "Permeability coefficient tests in non-cohesive soils," *Scientific Review Engineering and Environmental Sciences,* vol. 29, no. 1, p. 72–80, 2020.

[13] C. Janiesch, P. Zschech and K. Heinrich, "Machine learning and deep learning," *Electronic Markets,* vol. 31, no. 3, p. 685–695, 2021.

[14] C. Lee, P. Cheang and M. Moslehpour, "Predictive Analytics in Business Analytics: Decision Tree," *Advances in Decision Sciences,* vol. 26, no. 1, p. 1–29, 2022.

[15] S. Kim, S. Bae and M. Jang, "Linear Regression Machine Learning Algorithms for Estimating Reference Evapotranspiration Using Limited Climate Data," *Sustainability,* vol. 14, no. 18, 2022.

[16] J. Kiani, C. Camp and S. Pezeshk, "On the application of machine learning techniques to derive seismic fragility curves," *Computers and Structures,* vol. 218, p. 108–122, 2019.

[17] S. Lyu and J. Liu, "Convolutional recurrent neural networks for text classification," *Journal of Database Management,* vol. 32, no. 4, p. 65–82, 2021.

[18] J. Paireau, A. Andronico and N. Hozé, "An ensemble model based on early predictors to forecast COVID-19 health care demand in France," in *Proceedings of the National Academy of Sciences of the United States of America*, 2022.

[19] S. Tajik, S. Ayoubi and M. Zeraatpisheh, "Digital mapping of soil organic carbon using ensemble learning model in Mollisols of Hyrcanian forests, northern Iran," *Geoderma Regional*, vol. 20, 2022.

[20] S. Deori, R. Choudhary and D. Tiwari, "HDM-4 deterioration modelling: Validation and adoption for flexible pavements with modified bituminous road surfacing," *Baltic Journal of Road and Bridge Engineering,* vol. 14, no. 2, p. 208–226, 2019.