INTERNATIONAL JOURNAL OF PAVEMENT ENGINEERING
2024, VOL. 25, NO. 1, 2337916
https://doi.org/10.1080/10298436.2024.2337916

Taylor & Francis
Taylor & Francis Group

Check for updates

Soft computing to predict the porosity and permeability of pervious concrete based on mix design and ultrasonic pulse velocity

Navaratnarajah Sathiparan [a], Sathushka Heshan Wijekoon[a], Pratheeba Jeyananthan [b] and Daniel Niruban Subramaniam [a]

[a]Department of Civil Engineering, Faculty of Engineering, University of Jaffna, Jaffna, Sri Lanka; [b]Department of Computer Engineering, Faculty of Engineering, University of Jaffna, Jaffna, Sri Lanka

ABSTRACT

The present study explores the potential of machine learning to predict the porosity and permeability of pervious concrete constructed on mix parameters (compaction energy, aggregate-to-cement ratio and aggregate size) and ultrasonic velocity. The prediction models use non-destructive measurements and mixed design variables, which can help the construction sector apply the models without any theoretical expertise. The study uses 225 data samples from an experimental study. This study used six machine learning algorithms, namely, linear regression, artificial neural networks, boosted decision tree regression, random forest regression, K-nearest neighbour and support vector regression, to determine the best predictive model. The results show that the ANN model is the best technique for predicting the porosity of pervious concrete ($R^2 = 0.9502$ for training datasets and $R^2 = 0.8958$ for testing datasets) and boosted decision tress for permeability of pervious concrete ($R^2 = 0.9323$ for training datasets and $R^2 = 0.7574$ for testing datasets). The sensitivity analysis of the random forest regression model reveals that ultrasonic pulse velocity is the most influential parameter for the prediction of porosity and permeability of pervious concrete. The proposed models provide a more accurate method for estimating the porosity and permeability of pervious concrete.

ARTICLE HISTORY
Received 24 October 2023
Accepted 26 March 2024

KEYWORDS
Pervious concrete; porosity; permeability; machine learning; SHAP analysis

Abbreviations used in this study: A/C: Aggregate to cement; ANN: Artificial neural network; BDT: boosted tree regression; CE: Compaction energy; LR: Linear regression; KNN: K nearest neighbours; MAE: mean absolute error; ML: Machine learning; NDT: Non-destructive testing; R^2 Coefficient of determination; RFR: random forest regression; RMSE: root mean squared error; SHAP: SHapley Additive exPlanations; SVR: support vector regression; UPV: Ultrasonic pulse velocity; W/C: Water to cement

1. Introduction

Pervious concrete is a construction material with many applications, such as parking lots, pedestrian walkways, residential streets, low-traffic areas and greenhouses. It has a porous structure because it uses little or no fine aggregate and very uniform (monosize) coarse aggregate in its mixture (Subramaniam and Sathiparan 2022). Three important properties of pervious concrete are porosity, permeability and compressive strength. Porosity is the ratio of the volume of voids (empty spaces between the cement-covered particles) to the total volume of the mixture. Pervious concrete usually has a porosity of 15 to 35% (Chandrappa and Biligiri 2016, Li *et al.* 2017, Xu *et al.* 2018), which makes it porous. Compressive strength is the ability to resist compression forces. The compressive strength of pervious concrete was usually between 3 to 30 MPa (Alemu *et al.* 2021). The properties of pervious concrete depend on the quality and quantity of the cement and aggregate used as raw materials, as well as the mix design factors such as the cement-to-aggregate ratio, aggregate size distribution, water-to-cement ratio, and compaction effort (Anburuvel and Niruban Subramaniam 2022, Anburuvel and Subramaniam 2022a). However, even with the same mix design, the properties of each component may vary significantly in an industrial application due to different compaction levels and the boundary effect (Ridengaoqier *et al.* 2021). Hence, it is essential to evaluate the properties of pervious concrete even after construction.

The practical significance of forecasting the strength and permeability of pervious concrete using non-destructive testing (NDT) techniques is its capacity to direct the design and enhancement of pervious concrete. For example, alterations in the pore structure during use, such as blockage, might decrease the essential pore size, diminishing water transportation. Additionally, the compressive strength of pervious concrete, influenced by its composition, is critical. A suitable prediction equation for the compressive strength of pervious concrete enables enhancing the proportioning process, considering both the desired porosity and compressive strength. Hence, non-destructive testing (NDT) techniques that can accurately forecast these characteristics hold significant practical significance since they offer vital insights into the functionality of pervious concrete without inflicting any harm to the concrete itself. This can result in enhanced resource utilisation,

CONTACT Navaratnarajah Sathiparan sakthi@eng.jfn.ac.lk

functionality, and prolonged durability of pervious concrete structures.

Non-destructive test (NDT) methods are suitable tools to estimate the properties of construction materials without taking in situ samples (Sathiparan *et al.* 2023a). Some NDT methods for concrete widely used worldwide are rebound hammer, electrical resistivity, and ultrasonic pulse velocity (UPV) (Singh and Singh 2018). These methods have been extensively studied for conventional concrete (Ahmmad *et al.* 2016, Rao *et al.* 2016, Shafigh *et al.* 2016, Farahani *et al.* 2017, Huda *et al.* 2018, Lu *et al.* 2021), self-compacting concrete (Calabrese *et al.* 2015, Wang *et al.* 2016), cement mortar (Wei *et al.* 2012), and masonry (Mandal *et al.* 2016, Martín-del-Rio *et al.* 2020, Sathiparan and Jeyananthan 2023). While several studies have focused on NDT measurement to predict the strength of pervious concrete (Delatte *et al.* 2009, Adato and Delatte 2013, Zaetang *et al.* 2015, Amini *et al.* 2018, Filho *et al.* 2020), studies on predicting the permeability of pervious concrete are limited.

Delatte *et al.* (2009) measured the UPV and the hydraulic conductivity of field-collected pervious concrete samples. They gathered samples from 16 locations and tested them for UPV and hydraulic conductivity (k). The hydraulic conductivity values ranged from 0 to 1400 in. per hour, while the UPV values ranged from 8,012 to 16,800 feet per second. Their experimental results and statistical analysis proposed Equation (1) for estimating the hydraulic conductivity (where UPV is in ft/s and k is in in./hr). The coefficient of determination (R^2) between the measured and predicted hydraulic conductivity was 0.63.

$$k = 10^8 e^{-0.0012UPV} \tag{1}$$

Amini *et al.* (2018) used ultrasonic pulse velocity (UPV) and void ratio (v) of pervious concrete to develop a statistical model for its hydraulic conductivity (k). They measured UPV, v, and k of 25 pervious concrete cylinder specimens. The data ranged from 19% to 35% for v, 900 to 6900 cm/hr for k, and 2700 to 3540 m/s for UPV. They found that a polynomial function of UPV alone could fit k reasonably well, with a coefficient of determination (R^2) of 0.73. However, the prediction accuracy improved when UPV and v were used as predictors to predict k. The average ratio of predicted to measured k was 1.00, indicating a respectable agreement between the model and the measured data. Based on their experimental data and statistical analysis, they proposed Equation (2) for k (where v is in%, UPV is in m/s, and k is in cm/hr).

$$k = -4.16UPV + 239.909v + 10299.36 \tag{2}$$

Filho *et al.* (2020) studied the UPV of pervious concrete. They prepared three pervious concrete mixes with a water-to-cement (w/c) ratio of 0.3 and a binder-to-aggregate (B/Ag) ratio of 0.45 to 0.65. They measured the void ratio, permeability, and UPV of the mixes. The data ranged from 18% to 30% for void ratio, 1.1 to 9.5 mm/s for permeability, and 3642 to 4262 m/s for UPV. Based on their data and statistical analysis, they developed Equation (3) for permeability (where UPV is in m/s and k is in mm/s). The equation had a coefficient of determination (R^2) of 0.87, which shows a good fit.

$$K = 65.842 + 10.089(B/Ag) - 0.01677UPV \tag{3}$$

The proposed formula in published works for the permeability of pervious concrete had reasonable statistical measures of fit, as shown by R^2 values between 0.63 and 0.87. However, they also had some average bias in the fit. Moreover, these studies used small data sets that did not cover the full range of aggregate size, A/C ratio, compaction energy, etc., which are critical in defining the properties of pervious concrete needed for various industrial applications. However, the proposed single equation is not applicable when a particular property depends on several parameters.

Machine learning algorithms are increasingly popular in engineering for predicting the properties of building materials (Feng *et al.* 2020, Marani and Nehdi 2020, Song *et al.* 2021, Ahmad *et al.* 2022, Zhang *et al.* 2022b). Since the permeability of pervious concrete is subject to mixture proportions and other factors, machine learning techniques are observed to be the best way to estimate these properties. More advanced methods could reduce the need for laboratory tests and offer engineers simple tools to predict experimental outcomes (Ahmed *et al.* 2022). Machine learning methods can also provide alternatives and answers for both linear and nonlinear situations where mathematical models cannot capture the relationship between the variables involved in a problem (Gao *et al.* 2019, Anburuvel and Subramaniam 2022b, Wijekoon *et al.* 2023). Table 1 briefly overviews the machine learning approach used to predict permeable concrete properties. The investigations have shown that machine-learning techniques are more accurate than regression-based models. To the authors' knowledge, the combination of mix design parameters, UPV measurement and machine learning has not yet been used to predict the porosity and permeability of permeable concrete. While some studies have focused on predicting permeability based on aggregate size and aggregate-to-cement ratio, none have considered compaction energy. Published literature indicates that compaction energy is a significant characteristic that affects porosity and permeability.

The present study intends to enhance the quality control and inspection of pervious concrete by using mix parameters, UPV and machine learning techniques to predict the porosity and permeability of the pervious concrete. To make the models accessible for the construction sector without prior theoretical knowledge, mix design and UPV are used to generate prediction models for the porosity and permeability of pervious concrete. Six different machine learning techniques: linear regression (LR), artificial neural networks (ANN), boosted decision tree regression (BDR), random forest regression (RFR), K-Nearest Neighbor (KNN) and support vector regression (SVR) models are employed as model development tools to the invention the utmost accurate model for estimating the permeability of pervious concrete. The proposed models can offer a more precise prediction for the permeability of pervious concrete.

Table 1. Summary of machine learning methods used to predict the characteristics of pervious concrete.

Ref.	Input parameters	Response parameters	Machine learning method used	No. of data used
Ahmad *et al.* (2023)	Cement content, Gravel content, Waste glass power content, W/C ratio, Curing period	Compressive strength	ANN	99
Huang *et al.* (2020)	A/C ratio, Aggregate grading	Permeability	RFR	69
Huang *et al.* (2022)	Aggregate size, Sample thickness, clogging sands	Permeability	SVR	84
Le *et al.* (2022)	Aggregate size, A/C ratio, W/C ratio, SF content, Sand content	Porosity and compressive strength	ANN, RFR, SVR, & XGB	164
Malami *et al.* (2022)	Curing days, RHA content, CCW content	Flexural strength	ANN, SVR, & ANFIS	42
Sathiparan *et al.* (2023b)	Aggregate size, A/C ratio, CE	Ultrasonic pulse velocity	LR, ANN, BDT, RFR, SVR & XGB	75
Sathiparan *et al.* (2023c)	A/C ratio, FA/C ratio, W/C ratio, Aggregate size, curing period	Compressive strength	LR, ANN, BDT, RFR, SVR & XGB	437
Sun *et al.* (2019)	Aggregate size, A/C ratio, W/C ratio	Permeability and compressive strength	SVR	90
Subramaniam *et al.* (2023)	Aggregate size, A/C ratio, CE	Electrical resistivity	LR, ANN, BDT, KNN, RFR & SVR	75
Sudhir Kumar *et al.* (2024)	Aggregate size, GBFS content, Curing days	Compressive strength, Flexural strength & Permeability	ANN	36

RHA: rice husk ash, CCW: Calcium carbide waste, GBFS: Ground Granulated Blast-furnace Slag, SF: Silica fume, FA/C: fly ash to cement, LR: Linear Regression, ANFIS: Adaptive Neuro-Fuzzy Inference System, ANN: Artificial Neural Network, BDT: Boosted Decision Tree regression, KNN: K-Nearest Neighbors' regression, RFR: Random Forest Regression, SVR: Support Vector Regression, and XGboost: EXtreme gradient boosting

2. Materials and methods

2.1. Material used

The primary binder used was ordinary Portland cement – Type 1 OPC, and it had a specific gravity of 3.15 and a bulk density of 1280 kg/m^3 (Sathiparan 2023). The aggregate consisted of natural granite-gneisses rock of igneous origin. Figure 1 illustrates the particle size distribution of the coarse aggregate. The aggregates were sieved and separated into three categories in the size range of 5–12, 12–18, and 18–25 mm. Aggregates used in contemporary literature typically range from 5 to 20 mm, affecting the performance of pervious concrete (Pradhan and Behera 2022). However, it is essential to note that the size distribution of aggregates can vary. Studies have shown that aggregate size and distribution significantly impact pervious concrete performance. Comparing the performance of pervious concrete between studies conducted in different environments, including particle size distribution, may not be suitable for attributing performance variations to aggregate size distributions. This study aims to assess the impact of aggregate size on the performance of pervious concrete in the same controlled environment. Additionally, more uniform-sized particles are envisaged for use in the casting of samples. The aggregates used in this study were divided into 5–12 mm, 12–18 mm, and 18–25 mm. Table 2 shows the results of the evaluation of the aggregate quality. To construct the pervious concrete specimen, the aggregates had to satisfy the requirements for coarse aggregate.

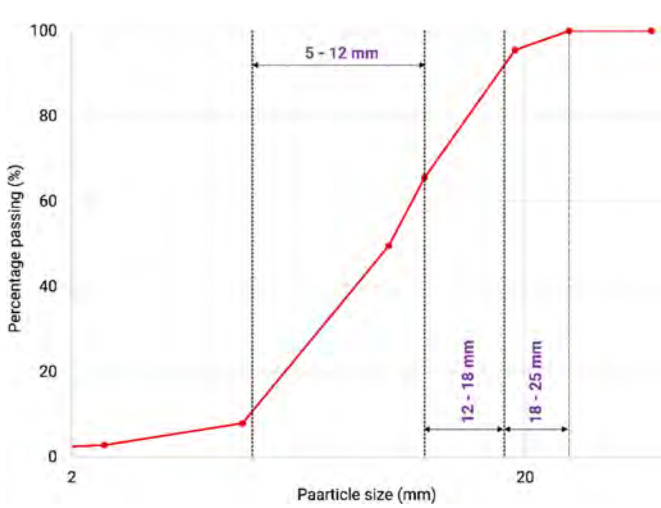

Figure 1. Particle size distribution of coarse aggregates.

2.2. Mix design

Based on published literature, the water-to-cement (W/C) ratio was fixed at 0.3 for all mixes to ensure zero slumps, while the aggregate-to-cement (A/C) ratio varied from 3.0 to 5.0 in steps of 0.5. The literature commonly uses an aggregate-to-cement ratio of approximately 3.0 to 8.0 for casting pervious concrete samples (Anburuvel and Subramaniam 2023). A study analysed the impact of aggregate-to-cement ratio on the performance of pervious concrete and optimum packing. The study found that the theoretical optimum aggregate-to-cement ratio was 2.65 for aggregates ranging from 12 mm to 25 mm in size. The study also observed an optimum aggregate-to-cement ratio of approximately 4.0 (Anburuvel and Niruban Subramaniam 2022). This study adopts a mix design that varies the aggregate-to-cement ratio from 3.0 to 5.0, with an increment of 0.5. Previous studies on pervious concrete have recommended zero slump and zero compaction to maintain open pores for fluid transmission. However, recent research has shown high uncertainty in pervious concrete's porosity and other performance factors. It has been found that an optimised degree of compaction can improve uniformity in the performance of pervious concrete. The optimum compaction degree was around 30 blows from a standard Proctor rammer (Subramaniam *et al.* 2022). This study varies the compaction from 0 to 60 blows in increments of 15 blows to analyse the impact of aggregate size on the optimum compaction in relation to other mix design parameters.

Pervious concrete cylindrical samples of a dimension of 150 mm cubes were cast with five different compaction efforts (0, 15, 30, 45 and 60 blows) and cylindrical core samples of 100

Table 2. Mechanical properties of aggregates.

Properties	Aggr. 5–12 mm	Aggr. 12–18 mm	Aggr. 18–25 mm	Test standard
Bulk density (kg/m^3)	1498	1491	1480	BS-EN-1097 (2020)
Specific gravity	2.73	2.72	2.72	BS-EN-1097 (2020)
Moisture content (%)	0.40	0.35	0.32	BS-EN-1097 (2020)
Absorption (%)	1.35	1.20	1.10	BS-EN-1097 (2020)
Flakiness Index	14.8	14.7	14.2	BS-EN-933 (2017)
Aggregate Impact value – AIV (%)	25.6	25.5	25.3	BS-EN-1097 (2020)
Aggregate Crushing Value – ACV (%)	30.5	30.5	30.4	BS-EN-1097 (2020)
Elongation Index (%)	23.95	24.02	24.80	BS-EN-933 (2017)

mm (diameter) x 150 mm (height) were obtained from that. The compaction process followed the BS 1377–4 (1990) standard, which uses a typical proctor compaction hammer. The Standard Proctor rammer with a face diameter of 5.08 cm, weight of 2.5 kg, and a plunge depth of 300 mm is employed to determine optimum sample moisture content according to BS 1377–4 (1990). The number of blows to the sample in the standard proctor mould with varying moisture content receives compaction energy in three layers (25 blows to each layer), so the sample assumes optimum packing within the mould. The standard proctor rammer gives a well-defined compaction energy to the sample in the mould. Although pervious concrete is cast with a zero-slump wet mixture and no compaction, some compaction is shown using a tamping rod when laboratory samples are fabricated for subsequent testing. Compaction given by tamping rods may not have definitive energy, as it may vary between tampings and laboratories, increasing uncertainty in the performance characteristics of hardened pervious concrete, such as porosity and compressive strength. Therefore, it is usual to employ standard or modified proctor rammer to provide well-defined compaction energy that may be uniformly maintained between rammings and laboratories. Pervious concrete having zero slumps, the binder-coated aggregates behave similarly to discrete soil particles. Table 3 shows the mix design for the pervious concrete mixture. The cement, water, and coarse aggregates were measured separately using an electronic balance with a precision of 0.1 g, based on the mix design in Table 2. A 120-litre electric concrete mixer was used to mix these ingredients uniformly. The mixing process took approximately 20 minutes to ensure a uniform mixture. A total of 225 cylindrical samples with three different aggregate sizes were used in this study. Figure 2 illustrates the previous concrete cube preparation procedure.

Table 3. Mix design for 1m^3 of pervious concrete mixture.

Mix ID	A/C ratio	Cement (kg)	Aggr. 5-12 mm (kg)	Aggr. 12-18 mm (kg)	Aggr. 18-25 mm (kg)	Water (L)
S-3.0	3.0	466.8	1400.3			140.0
S-3.5	3.5	413.3	1446.5			124.0
S-4.0	4.0	370.8	1483.2			111.2
S-4.5	4.5	336.2	1513.0			100.9
S-5.0	5.0	307.6	1537.8			92.3
M-3.0	3.0	466.8		1400.3		140.0
M-3.5	3.5	413.3		1446.5		124.0
M-4.0	4.0	370.8		1483.2		111.2
M-4.5	4.5	336.2		1513.0		100.9
M-5.0	5.0	307.6		1537.8		92.3
L-3.0	3.0	466.8			1400.3	140.0
L-3.5	3.5	413.3			1446.5	124.0
L-4.0	4.0	370.8			1483.2	111.2
L-4.5	4.5	336.2			1513.0	100.9
L-5.0	5.0	307.6			1537.8	92.3

2.3. Experimental programme

Standard methods were employed to regularise the experimental procedures to ensure reproducibility. Material properties were assessed using standard techniques consistent with the apparatus used. The aggregate size distribution was evaluated using BS standard sieve sizes, while the permeability apparatus was set according to the ACI standard method. UPV measurements were obtained using (an instrument model) that operates according to ASTM standards. However, these methods were widely used in literature and have consistency in assessments. This study assessed all parameters in the same batch and controlled environment, whereas good laboratory practices and controlled environments were followed throughout to ensure the quality of the observations.

2.3.1. UPV measurement

The portable ultrasonic NDT digital indication tester was utilised to measure the UPV of pervious concrete according to ASTM C597 (2010). Figure 3(a) shows the equipment setup. The instrument had a transit time range of 16 ms, a sampling rate of 2 MHz, a precision of 0.1 s, and a transmitter output of 1200 V. It detected the speed of ultrasonic pulses through the materials. The data analysis used the average value from three measurements in two horizontal and one vertical directional profile.

2.3.2. Porosity

The researchers measured the porosity of the pervious concrete specimen by the water displacement method, as shown in Figure 3(b) (Li *et al.* 2021). The samples were soaked in water for 24 hours and surface dried before permeability and porosity tests. Generally, there are four methods: the Corelok method (ASTM-D7063/D7063M 2018), the soak/tap method (Montes *et al.* 2005), the vacuum method and the image analysis method. The soak/tap method was explicitly developed for porous pavements and is based on the fundamental principle of porosity. In the current study, the method proposed by Montes *et al.* (2005) was used to modify volume instead of weight for the calculation. This technique is also relatively straightforward and can be carried out using standard laboratory equipment regularly used for many other test methods. The volume of water that filled the voids in the pervious concrete was calculated by the amount of water volume displaced by submerged oven-dried samples. The formulation for the

Figure 2. Pervious concrete cube preparation procedure.

porosity is given in Equation (4).

$$v = \left[1 - \frac{V_s}{V}\right] * 100 \quad (4)$$

where, v: porosity, V_s: volume of water moved upward due to submerged pervious concrete sample [m^3] (Volume = cross-section area of the container × increase in water height due to sample submerged), and V: volume of the sample [m^3]

2.3.3. Permeability

The constant head permeability test suggested by ACI 522R (2010) was used to measure the volume of water that passes through a sample in a given time. Figure 4 illustrates a schematic diagram of the test. The key components in a constant head permeability test apparatus for pervious concrete are as follows:

- Permeameter Assembly: The permeameter cell holds the pervious concrete specimen. It is a cylindrical shape with a known cross-sectional area. Sealed caps at both ends of the permeameter cell to prevent water from leaking. These end caps often have connections for water inflow and outflow.
- Constant Head cylinder: A container that holds a constant water head. It is connected to the permeameter cell to maintain a constant water level during the test.
- Water Inlet: Pipes that allow water to enter the constant head tank.

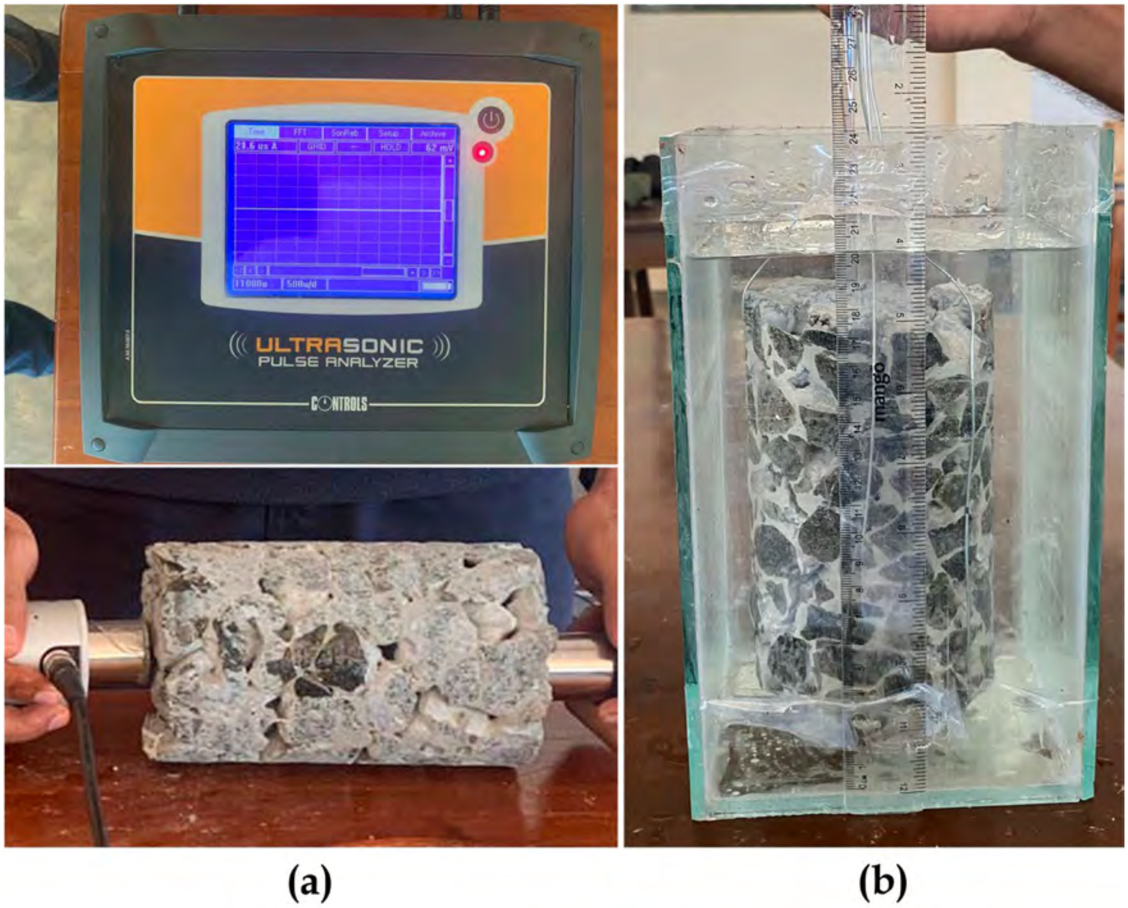

Figure 3. Test setup for (a) UPV measurement and (b) porosity measurement.

- Control Valves and Flow Regulators: It is the flow of water into and out of the permeameter cell, ensuring a constant head is maintained.
- Water Outlet: Pipes that allow water to exit the constant head tank.
- Graduated cylinder: It is used to collect the water from the outlet.

A cylindrical sample of pervious concrete with a height of 150 mm and a diameter of 100 mm was used for this test. The sample was wrapped laterally with PVC film and secured with regular scotch tape. It was also covered with silver tape (polyethylene film reinforced with cotton-laminated fabric) to ensure impermeability. The specimen was positioned between two PVC pipes with a diameter of 4 in. each. A hose was attached to the top PVC pipe to maintain the water level. The top pipe had an overflow drain to provide enough water for the test. We applied Equation (5) to calculate the permeability coefficient from the test results. The water was kept at a constant head of 300 mm for all experiments. The constant head permeability test used three samples for each type.

$$k = \frac{q \cdot L}{A \cdot h \cdot t} \tag{5}$$

Where k is the permeability coefficient (in cm/s), q is the water volume through the sample for the particular period (in cm^3), L is the length of the cylinder (in cm), A is the cross-sectional area of the cylinder (in cm^2), h is the water head height, and t is the time (in sec.) for water go through the sample. The flow rate (q) was measured by collecting the outflow over a period of time, measuring the total volume of fluid percolated through and then dividing it by the time lapsed. The measuring flask (bucket) had a perfect cylindrical shape with a gradation of a least count of 0.1 mm (vernier scale).

2.4. Machine learning modelling

Figure 5 shows the summary of the modelling flow chart diagram. Two hundred twenty-five data points were collected from the experimental programme to develop a reliable model for forecasting the porosity and permeability of pervious concrete. As the UPV value varied slightly for the subset of particular mixes, which consisted of 75 different mix types (comprising three types of aggregate size, five types of A/C ratio and five levels of compaction energy), the complete set of 255 samples was used to analyse the machine learning models. A stochastic allocation method was used to divide the data into two cohorts to pre-process the data for machine learning. The training dataset consisted of 150 datasets, representing two-thirds of the total datasets used to train the machine learning model. A test dataset of 75 datasets was allocated to evaluate the machine learning model's performance. This represents one-third of the total datasets. Six ML models

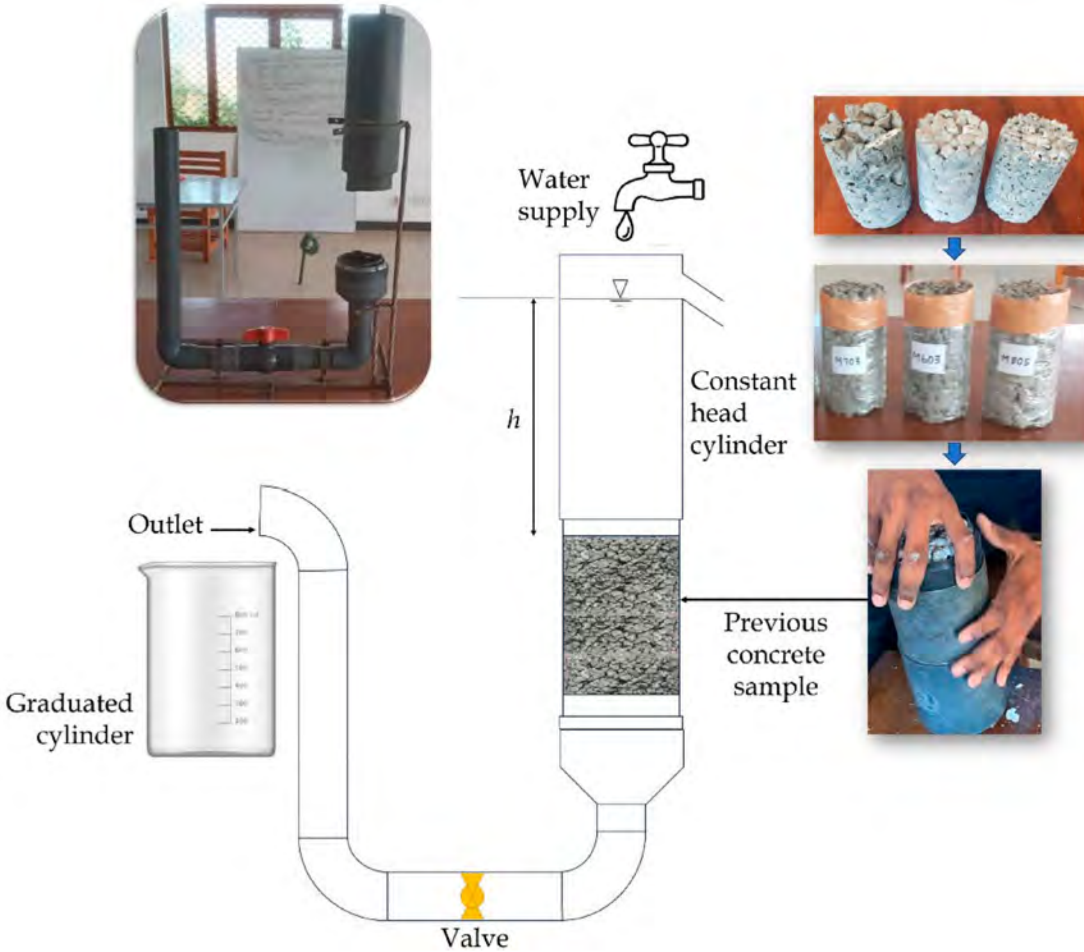

Figure 4. Schematic outline of apparatuses used to measure the permeability.

were used in this study to forecast porosity and permeability: linear regression (LR), artificial neural networks (ANN), boosted decision tree regression (BDT), random forest regression (RFR), K-Nearest Neighbor (KNN) and support vector regression (SVR). The selection of specific soft computing techniques for prediction often involves considering various factors: linearity and simplicity, non-linearity and complex patterns, ensemble learning and robustness, local patterns and instance-based learning, and flexibility and kernel methods. Linear regression was chosen for its simplicity and interpretability, which is suitable for capturing linear relationships in the data. ANN was selected for its ability to model complex, non-linear relationships and handle intricate patterns in data.

Boosted Decision Tree Regression and Random Forest Regression were selected for their effectiveness in handling non-linearity, capturing interactions, and providing robust predictions. The K-nearest neighbour was chosen for its suitability to capture local patterns and relationships in the data, especially when the underlying structure is not globally consistent. Support vector regression was chosen for its flexibility in capturing complex relationships through kernel functions and effectiveness in high-dimensional spaces. JMP Pro v.17 and Python version 3.11.5 were used for the machine learning analysis. JMP Pro was primarily used for hyperparameter tuning and model training. Python was used for model validation, SHAP analysis and dependence plots for critical features. All models use pre-built functions in Python libraries; none are implemented from scratch here.

2.4.1. Linear regression

Linear regression is a usually used supervised machine learning method that displays the linear correlation between one or more independent variables and a dependent variable. Certain conditions are necessary for this technique, including a linear relationship between the input and output values, minimal correlation among the features, and clean, noise-free data. Let's consider X as our feature matrix (or input) and y as the target vector measured on N samples, denoted by $\{(X^i, t^i)\}_{i=1}^N$. Equation stating the predicted value $\hat{y}$ of the above input data using a linear regression model can be written as Equation (6).

$$\hat{y} = \sum_{i=0}^{n} w_i x_i + b \tag{6}$$

Where w is the weight vector, and b is the bias term.

This can be equally written as Equation (7).

$$\hat{y} = w^T X + b \tag{7}$$

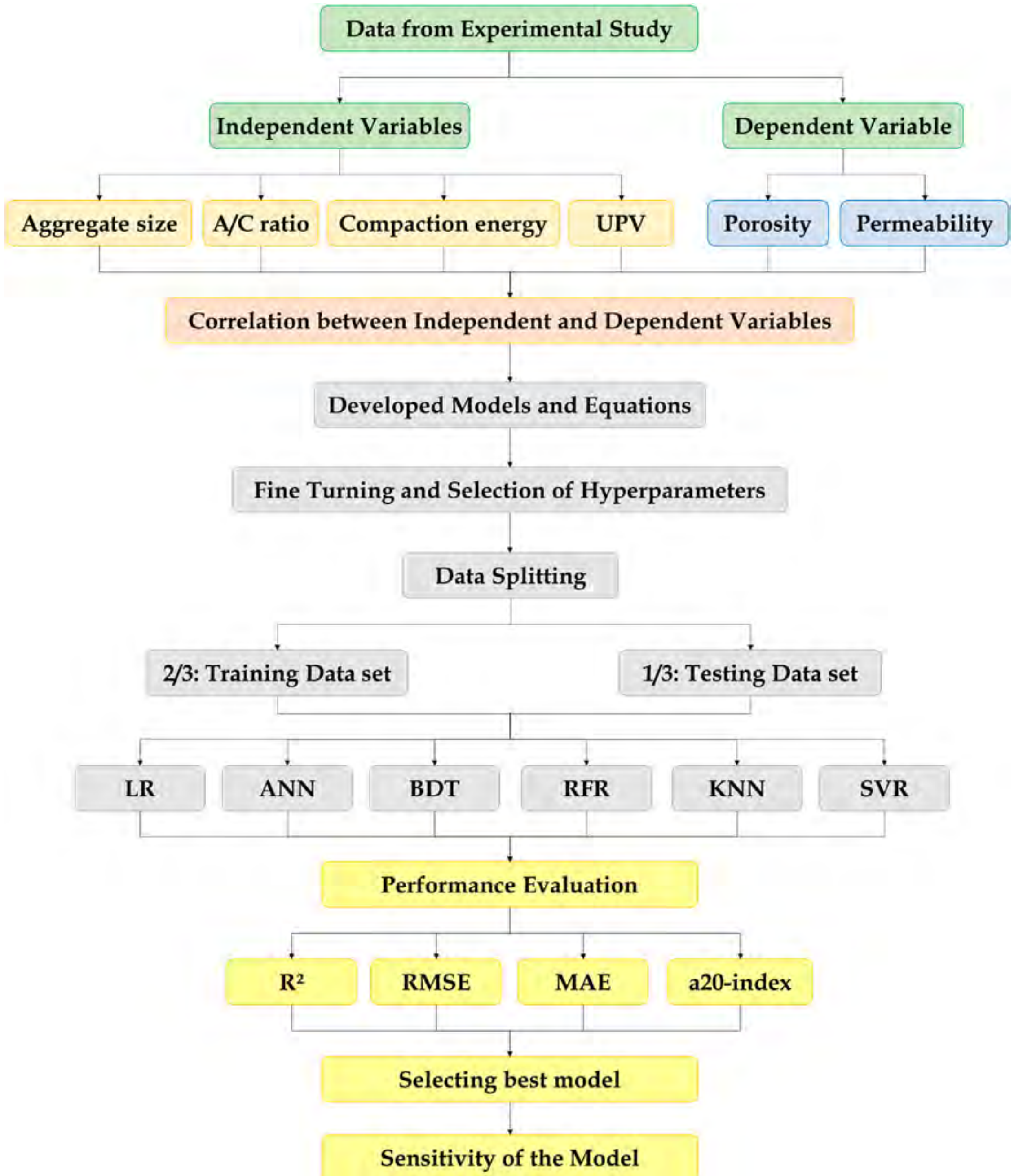

Figure 5. Flow chart for machine learning modelling procedure.

From the above equation, one can derive a simple equation for calculating the optimal weights, denoted by Equation (8).

$$w = (X^T X)^{-1} X^T t \tag{8}$$

2.4.2. *Artificial neural networks*

ANNs are computer systems developed based on the biological neural networks present in the animal brain. They are made up of interconnected nodes called artificial neurons, which process and transmit signals to one another. Each node has a weight and a threshold that determines its influence on the output of other nodes. ANNs can learn from data and perform classification, regression, clustering, and pattern recognition tasks. ANNs employ different learning algorithms, including supervised, unsupervised, and reinforcement learning, to adjust their weights and thresholds based on the training data and the desired output. Figure 6 illustrates the typical ANN structure. The output y of a single neuron is determined by applying an activation function f to the weighted sum of its input features plus a bias term, as shown in Equation (9).

$$y = f(\sum_{i=1}^{n} (w_i.x_i) + b) \tag{9}$$

Where, w_i represents the weights associated with each input feature x_i, and b is the bias.

ANNs also use various activation functions, such as linear and sigmoid, to introduce non-linearity and complexity into

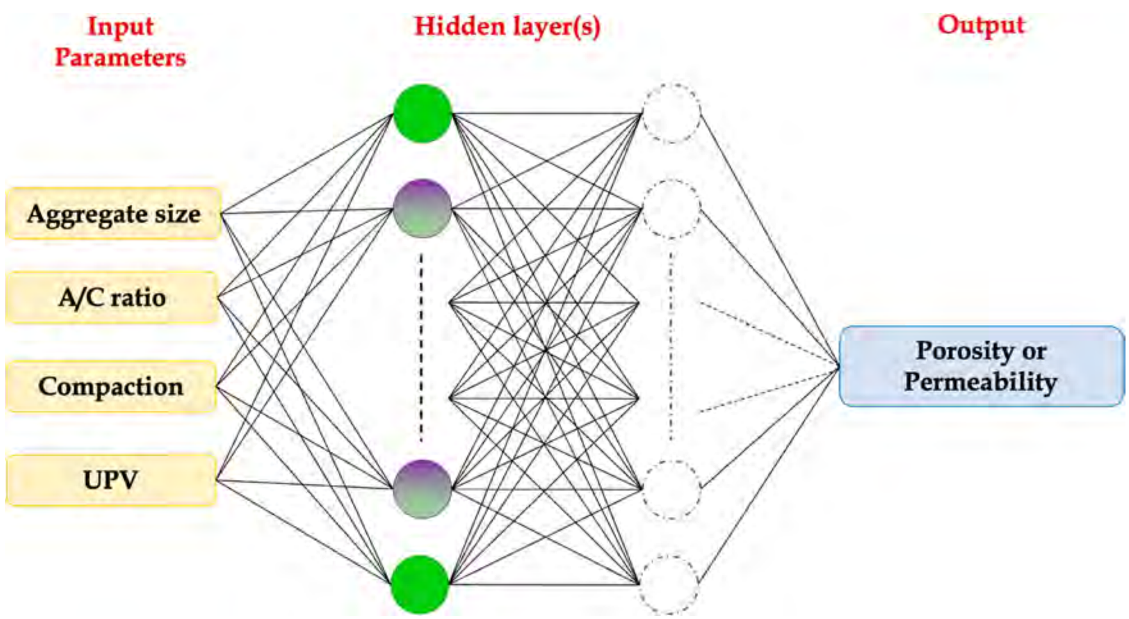

Figure 6. Typical ANN structure.

the network. The architecture of ANNs can vary, such as feed-forward, recurrent, or convolutional, depending on the structure and flow of information between the nodes and layers. The Root Mean Square Prop (RMSProp) optimiser is employed to accelerate the training process, as suggested in previous studies. This optimiser is particularly effective when dealing with limited datasets, which is often the case in civil engineering applications. To minimise the RMSE between the predicted and actual values, both forward and backward propagation are essential steps. Forward propagation calculates the final output by passing the input data through the network. Conversely, backward propagation adjusts the weights and biases within the network to minimise the RMSE. This is achieved by calculating the gradient of the loss function (RMSE in this case) with respect to each weight, utilising the chain rule. In the present study, the tanh (hyperbolic tangent) activation function was employed within the artificial neural network (ANN) architecture.

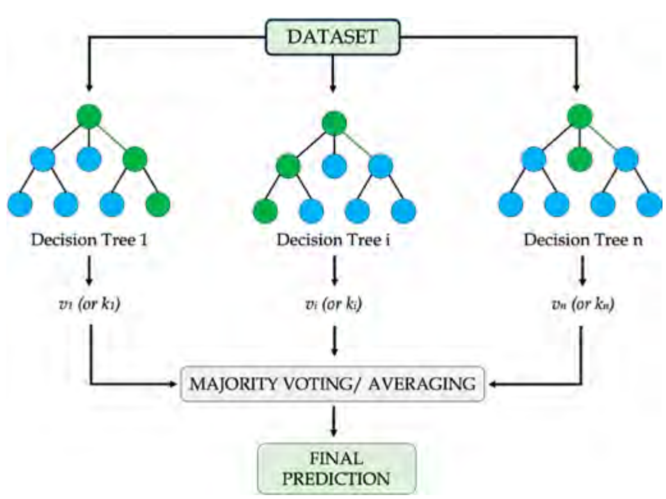

Figure 7. Sample of random forest regression.

2.4.3. Random forest regression

Random forest regression is a widely used ML technique that collaborates decision trees to forecast a continuous target variable. This technique is effective because each decision tree is trained on a different subset of the data and features, decreasing the correlation between the trees and boosting the ensemble's diversity. The random forest generates its final forecast by averaging the predictions of all the individual trees. The method can handle non-linear relationships, missing values, and feature interactions. Random forest regression offers feature importance metrics and error estimation, making it a valuable tool in many applications (Lin *et al.* 2022). The development of a random forest involves the creation of multiple decision trees by training on random subsets of the training data, known as bootstrap samples. At each split, a random subset of features is considered for each tree, encouraging diversity among the trees. The final prediction in a regression job often involves calculating the mean (or weighted mean) of the predictions made by each tree. The Random Forest regression algorithm uses the predictions of each tree to aggregate and predict the output for a new input. The primary goal of random forest regression is to minimise the mean squared error between the predicted and actual values in the training set. This objective is fine-tuned during training to identify the optimal parameters for each decision tree. The final prediction is obtained by combining the trees' forecasts, with each tree assigned a weight based on its performance, as shown in Figure 7.

2.4.4. Boosted tree regression

Boosted tree regression is a powerful ML technique that employs a collection of decision trees to forecast a continuous target variable. In contrast to random forest regression, which creates many uncorrelated trees, boosted tree regression generates trees in sequence, with each subsequent tree attempting to correct the errors of the preceding one. This is accomplished

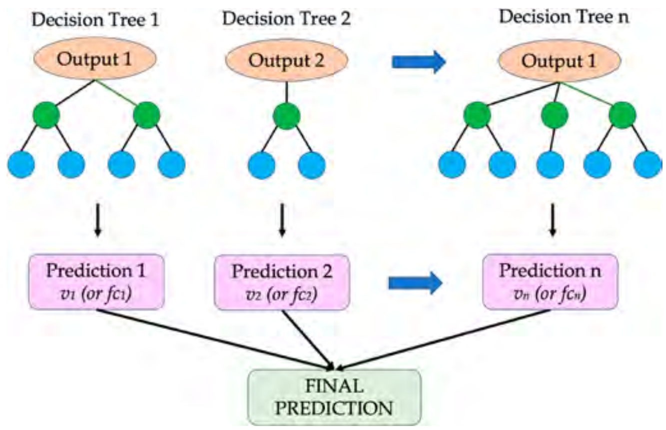

Figure 8. Typical boosted decision tree regression structure.

via a gradient descent algorithm that minimises the loss function (Guillen *et al.* 2023). Boosted tree regression can handle nonlinear relationships, feature interactions, and absent values in the data. Furthermore, it offers metrics of attribute significance and error estimation (Elith *et al.* 2008). Boosted decision tree regression employs decision trees as the fundamental learners, similar to random forest regression. The process starts with an initial forecast for each data point, typically using a basic model like the average value of the target variable. Boosting constructs trees sequentially, with each tree aiming to correct mistakes made by the previous ensemble. Each tree relies on the previous trees, as shown in Figure 8. In each iteration, a new decision tree is trained using the negative gradient of the loss function in relation to the current ensemble's predictions. The loss function is established at each stage to quantify the error, enabling the subsequent correction of the next tree until the optimal tree is achieved. In boosted decision tree regression, the prediction is a combination of multiple weaker prediction models that work together to create a more powerful prediction model, as demonstrated in the Equations (10) and (11):

$$\hat{y}(x) = \sum_{t} w_t h_t(x) \tag{10}$$

$$O(x) = \sum_{i} l(\hat{y}_i, y_i) + \sum_{t} \Omega(f_t) \tag{11}$$

In this context, $h(x)$ represents the output of the tree, w represents the weight, and $l(\hat{y}_i, y_i)$ represents the loss function, which measures the difference between the predicted value and the actual value of the ith sample. The function $\Omega(f_t)$ serves as a regularisation function.

Additionally, the prediction of each tree is multiplied by the learning rate before being incorporated into the collective ensemble. This helps to reduce overfitting and ensures a more gradual learning process. Boosting continues until a predetermined number of trees (iterations) are reached or a certain performance level is achieved.

2.4.5. K-nearest neighbour

KNN is a popular ML algorithm that can be utilised for both classification and regression tasks. It is based on the concept that similar data points tend to have similar labels or outputs. This algorithm does not require assumptions about the underlying data distribution and does not need any training phase. Instead, it stores all the data points available and their corresponding labels. KNN uses a distance metric, such as Euclidean distance, to classify a new query point and identify the most similar data points. In KNN regression, first, it needs to find the k nearest neighbours in the dataset. Euclidean distance is widely used in finding these k nearest neighbours. Euclidean distance between two points x^a *and* x^b can be calculated using Equation (12).

$$||x^a - x^b||_2 = \sqrt{\sum_{j=1}^{d} (x_j^a - x_j^b)^2} \tag{12}$$

Then, make a prediction based on the labels of those nearest neighbours. As the value is continuous in regression problems, a common approach for computing that continuous target is calculating the average value over all those k nearest neighbours, as shown in Equation (13).

$$h(X) = \frac{1}{k}\sum_{i=1}^{k} f(X^i) \tag{13}$$

The label or output of the query point is then assigned based on the majority vote (for classification) or the average (for regression) of the k nearest neighbours (Hu *et al.* 2016). Although KNN is easy to implement and interpret, it can be computationally expensive and susceptible to noise and irrelevant features (Saini *et al.* 2013).

2.4.6. Support vector regression

SVR is an ML technique that can be utilised for both linear and nonlinear regression problems. The method finds a function that fits data points with a small error margin while being as flat as possible (Xia 2020). SVR applies a kernel function to map data points into a higher dimensional space, making it possible to find a linear function. The linear function is defined by a subset of data points called support vectors, and their coefficients are called dual variables. The technique minimises a regularised objective function that balances the trade-off between function flatness and error margin, measured by an epsilon-insensitive loss function that considers errors within a specified tolerance epsilon. Optimal performance requires tuning several parameters, including regularisation parameter C, kernel function and its parameters, and the epsilon value (An and Liang 2013). The regression equation for SVR is given by Equation (14).

$$f(x) = w^T x_i + b \tag{14}$$

Where $f(x)$ is the predicted value, w is the weight vector, x_i is the feature vector on the transformed space, and b is the bias term.

In order to improve the robustness of the model, epsilon-intensive loss which is defined as Equation (15).

$$J_\epsilon(y, f(x)) = \max(0, |y - f(x)| - \epsilon) \tag{15}$$

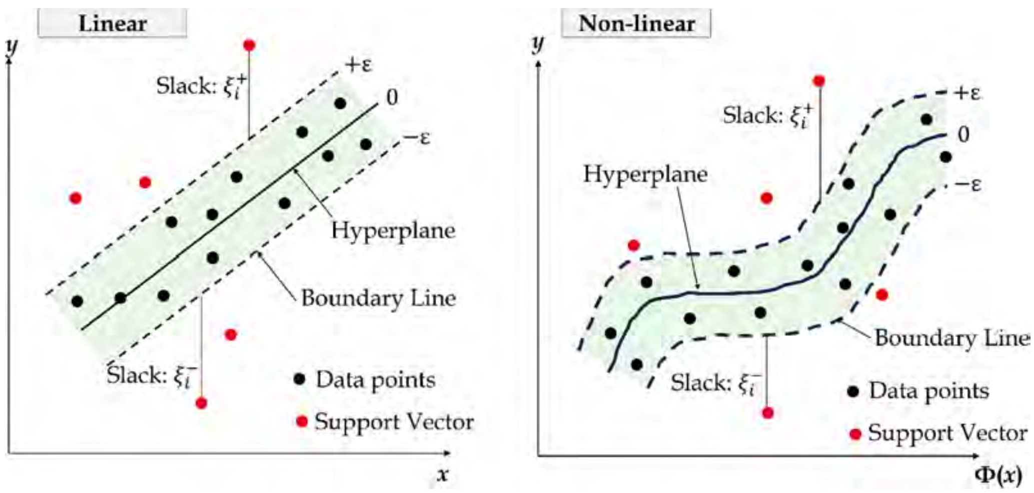

Figure 9. Linear and Non-linear SVR structure.

Where y is the actual value, ϵ is a user-defined parameter controlling the width of the insensitive zone.

SVR seeks to minimise the empirical risk, the sum of the training errors subject to a regularisation term. The objective function for SVR can be formulated as Equation (16).

$$min_{w,b,\xi_i^+,\xi_i^-} \frac{1}{2}w^2 + C\sum_i (\xi_i^+ + \xi_i^-) \tag{16}$$

SVR introduces constraints on the slack variables to ensure that errors are within a permissible range. The constraints are formulated as Equations (17)-(19).

$$y_i - f(x_i) \leq \varepsilon + \xi_i^+ \tag{17}$$

$$f(x_i) - y_i \leq \varepsilon + \xi_i^- \tag{18}$$

$$\xi_i^+, \quad \xi_i^- \quad \geq 0 \tag{19}$$

Where w is the weight vector, b is the bias term, ξ_i^+ and ξ_i^- are slack variable, and C is the regulation parameter

These constraints ensure that errors are within the epsilon-insensitive tube, and the regularisation term controls the trade-off between achieving a small error and keeping the margin wide, as shown in Figure 9. The choice of kernel function and tuning parameters such as C and ε play crucial roles in the performance of SVR.

2.5. K-fold validation and performance indicators

Validation methods in machine learning assess a model's performance on unseen data. Common approaches include a simple train-test split, where data is divided for training and testing. However, this method's reliance on a single split can introduce biases based on the randomness of that division. K-fold cross-validation addresses this limitation by dividing the data into K subsets and iteratively using each subset as a test set while the remaining data serves as the training set. This process is repeated K times, reducing the impact of a single split. By averaging performance over multiple folds, K-fold cross-validation provides a more robust and unbiased estimate of a model's generalisation ability. It ensures that the model is evaluated on various subsets of the data, smoothing out variability and preventing overfitting to specific splits. In summary, while simple train-test splits may lead to biased evaluations, K-fold cross-validation minimises such biases by considering multiple splits and averaging performance, offering a more reliable assessment of a model's performance. The stratified K-fold validation technique is commonly used because it is considered accurate and efficient. This technique divides the data into K groups; K is set to five in this study. It splits the data into five subsets and uses four to train the model. The remaining subset is used to validate the model. This process is repeated five times to obtain the average result.

Four different indicators are used to evaluate the performance of the proposed model: Coefficient of determination (R^2), Root Mean Squared Error ($RMSE$), Mean Absolute Error (MAE) and α20-index are represented by Eqs. (20), (21), (22) and (23), respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(P_i - E_i)^2}{\sum_{i=1}^{n}(E_i - \bar{E})^2} \tag{20}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(P_i - E_i)^2}{n}} \tag{21}$$

$$MAE = \frac{\sum_{i=1}^{n}|P_i - E_i|}{n} \tag{22}$$

where Ei is the measured value; Pi is the predicted value; $\bar{E}$ is the average of measured values; n is the number of the data used.

The α20-index is a new measure for assessing the reliability of the soft computing techniques developed, as defined in Equation (9) (Asteris *et al.* 2020, Armaghani and Asteris 2021).

$$\alpha 20 - \text{index} = n_{20}/N \tag{23}$$

N is the number of datasets and n_{20} is the number of data, where the predicted/measured strength ratio was between 0.80 to 1.20.

Table 4. Summary of UPV, porosity and permeability values for each mix.

Aggregate size (mm)	A/C ratio	Compaction (No. of blows)	UPV (km/s)	Porosity (%)	Permeability (cm/s)
5–12	3.0	0	3.31 ± 0.06	39.00 ± 0.78	2.35 ± 0.21
5–12	3.0	15	4.45 ± 0.02	20.14 ± 1.36	0.69 ± 0.07
5–12	3.0	30	4.69 ± 0.06	16.82 ± 1.36	0.37 ± 0.11
5–12	3.0	45	4.79 ± 0.01	14.32 ± 1.18	0.16 ± 0.04
5–12	3.0	60	4.96 ± 0.06	8.22 ± 1.04	0.09 ± 0.01
5–12	3.5	0	3.70 ± 0.06	35.12 ± 2.35	1.86 ± 0.60
5–12	3.5	15	4.50 ± 0.02	21.81 ± 1.36	1.12 ± 0.05
5–12	3.5	30	4.63 ± 0.06	20.14 ± 1.36	0.43 ± 0.08
5–12	3.5	45	4.65 ± 0.06	16.82 ± 1.36	0.34 ± 0.10
5–12	3.5	60	4.90 ± 0.01	13.21 ± 0.39	0.23 ± 0.04
5–12	4.0	0	3.86 ± 0.02	30.40 ± 1.57	1.70 ± 0.16
5–12	4.0	15	4.34 ± 0.14	20.42 ± 1.96	1.03 ± 0.25
5–12	4.0	30	4.59 ± 0.05	19.03 ± 3.42	0.62 ± 0.04
5–12	4.0	45	4.66 ± 0.03	16.54 ± 2.18	0.49 ± 0.15
5–12	4.0	60	4.76 ± 0.06	13.21 ± 1.04	0.36 ± 0.14
5–12	4.5	0	3.42 ± 0.1	36.78 ± 1.36	2.56 ± 0.17
5–12	4.5	15	4.09 ± 0.01	24.58 ± 0.78	1.25 ± 0.14
5–12	4.5	30	4.43 ± 0.10	17.65 ± 1.18	0.99 ± 0.17
5–12	4.5	45	4.42 ± 0.07	22.36 ± 0.78	1.12 ± 0.26
5–12	4.5	60	4.60 ± 0.06	19.18 ± 0.70	0.71 ± 0.02
5–12	5.0	0	3.63 ± 0.06	37.33 ± 1.57	2.47 ± 0.06
5–12	5.0	15	3.95 ± 0.07	27.35 ± 0.78	2.02 ± 0.15
5–12	5.0	30	4.14 ± 0.10	21.81 ± 1.36	1.14 ± 0.13
5–12	5.0	45	4.28 ± 0.09	19.03 ± 2.83	0.98 ± 0.22
5–12	5.0	60	4.39 ± 0.13	20.70 ± 2.07	1.11 ± 0.15
12–18	3.0	0	4.18 ± 0.06	23.47 ± 2.35	0.75 ± 0.30
12–18	3.0	15	4.74 ± 0.19	13.49 ± 2.35	0.33 ± 0.17
12–18	3.0	30	4.80 ± 0.02	12.93 ± 2.07	0.19 ± 0.08
12–18	3.0	45	5.02 ± 0.11	6.28 ± 2.07	0.27 ± 0.08
12–18	3.0	60	5.24 ± 0.17	10.16 ± 1.36	0.13 ± 0.08
12–18	3.5	0	3.88 ± 0.07	26.24 ± 3.92	2.68 ± 0.83
12–18	3.5	15	4.55 ± 0.02	17.37 ± 1.57	0.67 ± 0.19
12–18	3.5	30	4.83 ± 0.08	11.82 ± 0.10	0.31 ± 0.10
12–18	3.5	45	4.78 ± 0.21	9.05 ± 4.15	0.19 ± 0.04
12–18	3.5	60	5.02 ± 0.09	9.05 ± 1.57	0.20 ± 0.01
12–18	4.0	0	3.85 ± 0.04	36.22 ± 0.78	2.91 ± 0.14
12–18	4.0	15	4.48 ± 0.03	21.81 ± 0.10	1.74 ± 0.42
12–18	4.0	30	4.52 ± 0.04	20.70 ± 0.78	1.09 ± 0.07
12–18	4.0	45	4.73 ± 0.03	16.82 ± 0.68	0.81 ± 0.04
12–18	4.0	60	4.72 ± 0.03	14.60 ± 1.71	0.59 ± 0.11
12–18	4.5	0	3.69 ± 0.05	35.67 ± 0.78	3.27 ± 0.69
12–18	4.5	15	4.20 ± 0.09	16.82 ± 1.36	0.96 ± 0.16
12–18	4.5	30	4.36 ± 0.07	18.38 ± 0.14	0.95 ± 0.07
12–18	4.5	45	4.54 ± 0.04	19.59 ± 2.83	0.99 ± 0.41
12–18	4.5	60	4.61 ± 0.04	17.37 ± 0.78	0.66 ± 0.12
12–18	5.0	0	3.60 ± 0.08	36.22 ± 0.39	3.08 ± 0.21
12–18	5.0	15	3.85 ± 0.03	25.69 ± 0.78	2.06 ± 0.48
12–18	5.0	30	4.01 ± 0.07	21.25 ± 1.04	1.52 ± 0.08
12–18	5.0	45	4.19 ± 0.07	20.42 ± 1.04	1.23 ± 0.09
12–18	5.0	60	4.59 ± 0.09	19.03 ± 0.78	1.31 ± 0.21
18–25	3.0	0	3.78 ± 0.09	37.33 ± 3.42	2.57 ± 0.45
18–25	3.0	15	4.63 ± 0.11	26.24 ± 2.07	1.28 ± 0.22
18–25	3.0	30	4.88 ± 0.11	15.71 ± 0.78	1.20 ± 0.10
18–25	3.0	45	4.98 ± 0.14	14.60 ± 1.57	0.41 ± 0.22
18–25	3.0	60	5.37 ± 0.07	11.27 ± 2.07	0.24 ± 0.19
18–25	3.5	0	3.84 ± 0.04	36.22 ± 4.15	4.79 ± 2.32
18–25	3.5	15	4.70 ± 0.08	21.25 ± 1.57	2.70 ± 0.22
18–25	3.5	30	4.70 ± 0.04	19.59 ± 0.78	1.21 ± 0.34
18–25	3.5	45	4.86 ± 0.05	16.26 ± 0.78	1.15 ± 0.24
18–25	3.5	60	5.06 ± 0.04	16.26 ± 1.57	0.29 ± 0.10
18–25	4.0	0	3.61 ± 0.05	39.55 ± 4.37	4.59 ± 0.27
18–25	4.0	15	3.87 ± 0.13	25.69 ± 0.78	4.42 ± 0.69
18–25	4.0	30	4.08 ± 0.05	23.47 ± 1.00	2.50 ± 0.34
18–25	4.0	45	4.36 ± 0.08	20.70 ± 0.78	2.22 ± 0.66
18–25	4.0	60	4.6 ± 0.04	15.15 ± 1.36	1.27 ± 0.19
18–25	4.5	0	3.58 ± 0.01	45.10 ± 3.59	3.82 ± 0.71
18–25	4.5	15	3.60 ± 0.04	24.02 ± 1.57	2.79 ± 0.52
18–25	4.5	30	4.11 ± 0.05	22.36 ± 2.07	2.20 ± 0.79
18–25	4.5	45	4.23 ± 0.07	21.25 ± 0.78	1.86 ± 0.42
18–25	4.5	60	4.49 ± 0.03	19.03 ± 0.78	1.94 ± 0.20
18–25	5.0	0	3.39 ± 0.19	39.55 ± 4.37	5.58 ± 0.39
18–25	5.0	15	3.58 ± 0.06	29.02 ± 1.57	4.28 ± 0.74
18–25	5.0	30	3.72 ± 0.07	24.58 ± 0.78	3.76 ± 0.60
18–25	5.0	45	4.02 ± 0.07	22.92 ± 0.78	3.08 ± 0.45
18–25	5.0	60	4.23 ± 0.08	22.92 ± 1.57	2.48 ± 0.42

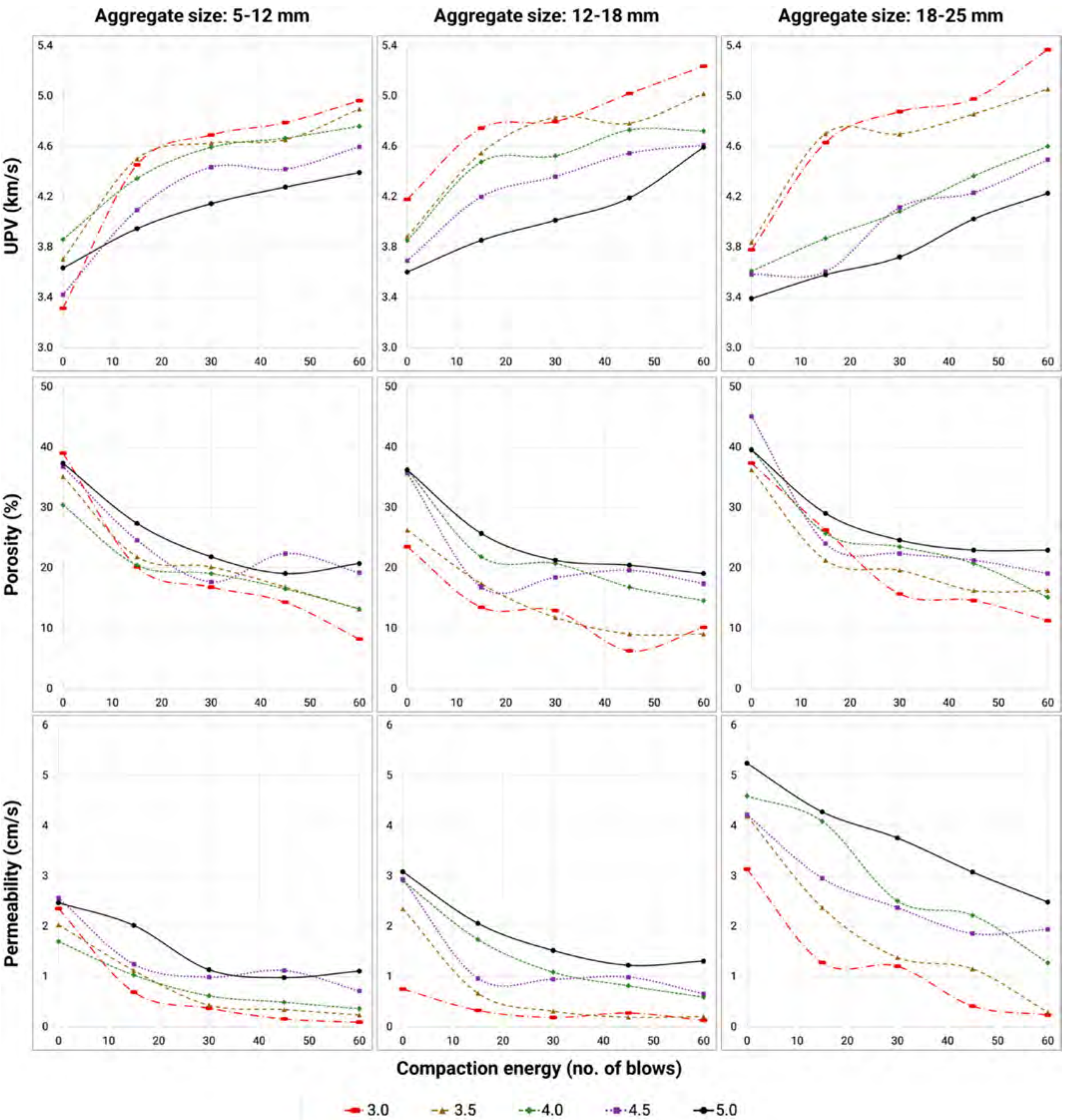

Figure 10. UPV, porosity and permeability variation with compaction energy, A/C ratio and aggregate size.

Furthermore, the Taylor diagram and error distribution were employed to assess the performance of the models thoroughly. The Taylor diagram is a more advanced approach for evaluating model performance than conventional measures. This diagram was overlooked in numerous machine learning research due to the complexity of graphing. The Taylor diagram facilitates the concurrent evaluation of numerous metrics, which is advantageous for comparing models that may exhibit disparate performance in many aspects. The Taylor diagram helps compare datasets with different means and variances. It utilises the standard deviation and correlation of observed data as reference points. Hence, the Taylor diagram is a tool for comparing datasets undergoing distinct normalisation or standardisation methods. The Taylor diagram employs orthogonal axes, circular lines, and a radial axis to depict the standard deviation, root mean square error, and correlation coefficient. The model nearest to the reference point on the Taylor diagram is regarded as the most precise model

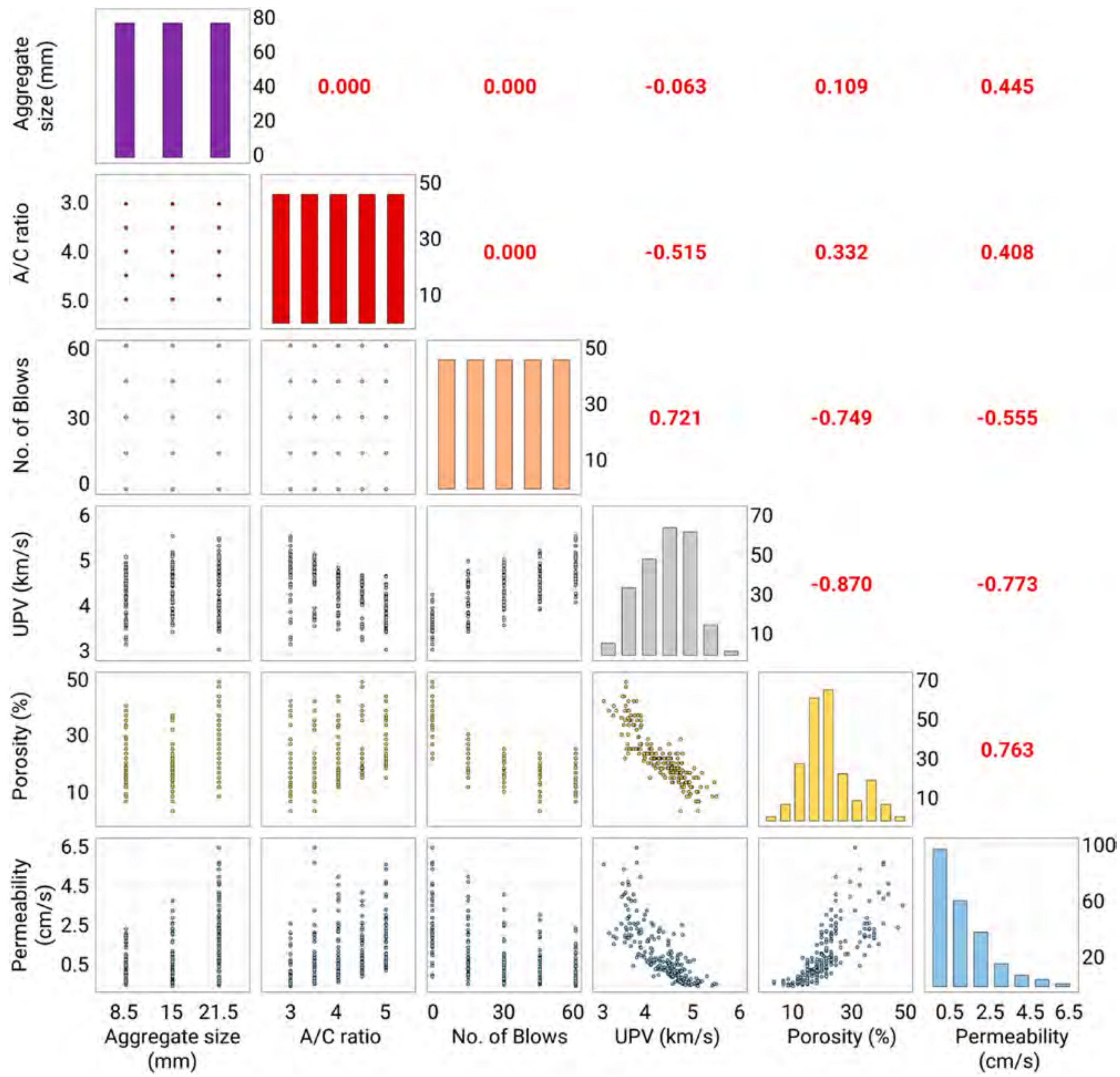

Figure 11. Data distribution of UPV, porosity and permeability of previous concrete.

in terms of its capacity to match the observed data (Ziamiavaghi and Toufigh 2023).

3. Experimental results

Table 4 summarises the mix design and corresponding UPV, porosity and permeability of the previous concrete. Figure 10 shows the effects of compaction effort, A/C ratio and aggregate size on UPV, porosity, and permeability. The UPV value was inversely proportional to the porosity of pervious concrete as it predominantly depended on the solid content of the mixture (Ndagi *et al.* 2019). Therefore, UPV increased with higher compaction energy and a lower A/C ratio. Low A/C ratios had enough cement paste to produce more hydration products, fill more pores, and increase the solid fraction of pervious concrete. This made the ultrasonic pulse travel faster through the pervious concrete. Higher compaction also reduced the pores in pervious concrete, and the UPV increased with compaction energy. Furthermore, the solid content of pervious concrete decreased as porosity increased with aggregate size. As a result, the UPV decreased as aggregate size increased.

Porosity values showed a tendency to decrease with higher compaction energy. For 30 blows, the porosity decreased very rapidly, but as the number of blows increased further, the rate of decrease slowed down. On the other hand, an increase in porosity was observed with an increase in the A/C ratio. This could also be due to the low A/C ratios filling the pores in pervious concrete and reducing the voids. However, insufficient cement paste left the pores in high A/C ratio specimens open. Low A/C ratios showed the most significant effect of compaction effort on porosity. As mentioned, more

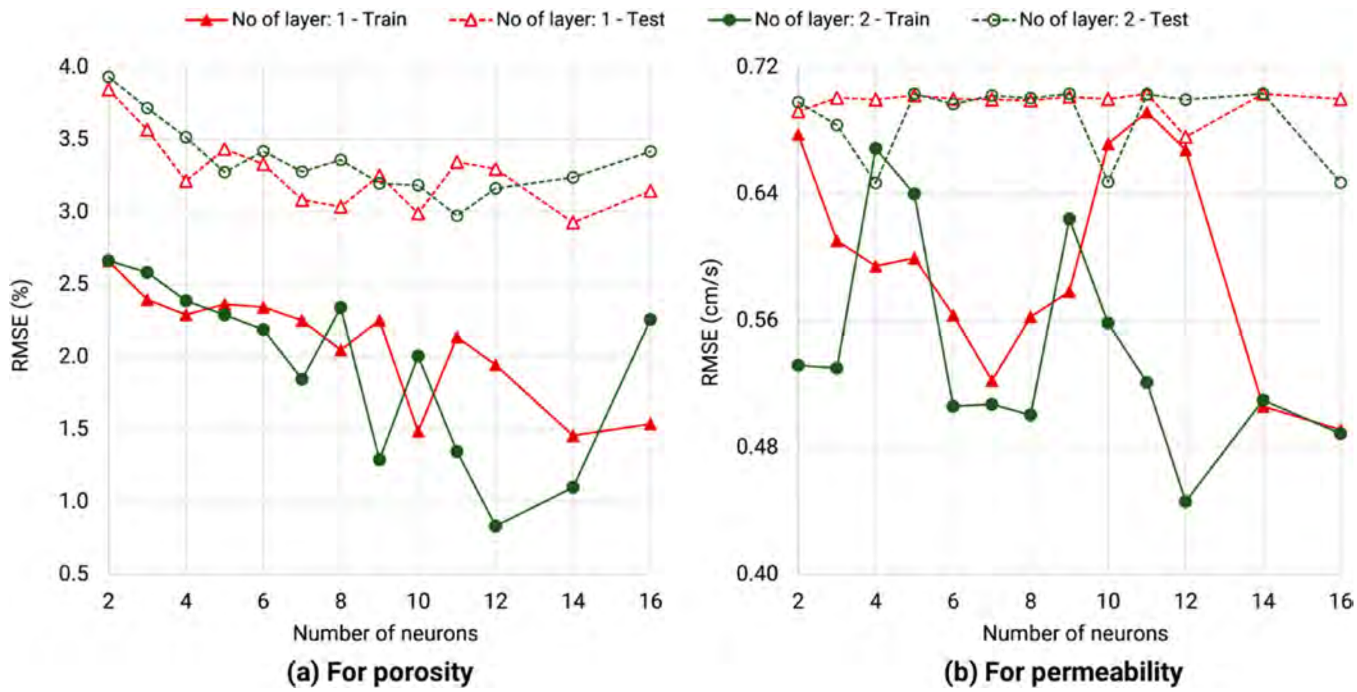

Figure 12. Effect of number of layers and neurons per layer on performance of models for prediction of (a) porosity and (b) permeability.

compaction helped pack the aggregates more densely to compensate for the lack of cement paste in low A/C ratios. The aggregates could be packed more tightly in high A/C ratios, but the lack of cement paste left the pores between the aggregates open.

On the other hand, porosity values varied randomly between aggregate sizes for samples with zero compaction effort, primarily owing to the random packing of binder-coated aggregates. When no compaction energy is supplied to the sample, and due to zero slump of concrete mix, the filling of binder-coated aggregates assumes a somewhat random packing compared to samples with compaction energy supplied. The observations of porosity values among the samples that had compaction energy delivered did not yield any significant impact of aggregate size on porosity, specifically on samples with 30 or higher blows. A change in aggregate size would result in a difference in the total surface area of aggregate particles. Hence, the binder thickness coating aggregate particles is expected to increase as the aggregate size increases.

The other observation on porosity is that it tends to asymptote with compaction effort for all mix designs and aggregate sizes used in the experiments. Owing to the dependency of UPV predominantly on porosity, it could be expected that the UPV values also tended to asymptote at approximately 30 blows and stayed constant after that for all mix designs. However, a continuously increasing trend was observed in UPV in contrast to the saturation observed in porosity. The UPV value depends not solely on the porosity fraction but also on the solid fraction's properties. An increase in compaction effort, although it does not change the porosity of the samples, may induce significant differences in the pore sizes, pore distribution and number of pores. This would significantly impact the path of UPV signals through the solid fraction, specifically in terms of interface resistances. This may have contributed to the continuously increasing trend of UPV with compaction effort.

Permeability showed a decreasing trend with higher compaction energy and a lower A/C ratio. The aggregate with the highest permeability (5.05 cm/s) had a size of 18–25 mm, an A/C ratio of 5.0, and a compaction effort of 0 blows. The aggregate with the lowest permeability (0.03 cm/s) had a size of 5 to 12 mm, an A/C ratio of 3, and a compaction effort of 60 blows. Cement paste would be reduced as the aggregate-to-cement ratio increased. As a result, the binder thickness would be thinner around the particles, subsequently increasing porosity. This would allow water to flow through easily. Therefore, a higher A/C ratio led to an increase in permeability. The permeability decreased as the number of blows increased as the voids became smaller in size and presence. When permeability values are compared across different aggregate sizes, samples with 5–12 and 12–18 aggregates showed a saturation trend in permeability with compaction, while samples with 18–25 mm aggregates showed a continuously decreasing trend with compaction. As discussed above, although porosity remained consistent with different aggregate sizes, the pore size, distribution and the number of pores would significantly change across aggregate sizes. Changes in pore characteristics (size, distribution and number) would significantly affect permeability as permeability is affected by the capillary effect.

From the above discussion, it could be observed that although dependent parameters, including porosity, UPV and permeability, show a high correlation among them, the degree of their association is affected by design parameters. Therefore, predicting the performance of pervious concrete (porosity and permeability) from mix-design parameters alone or from individual dependent parameters

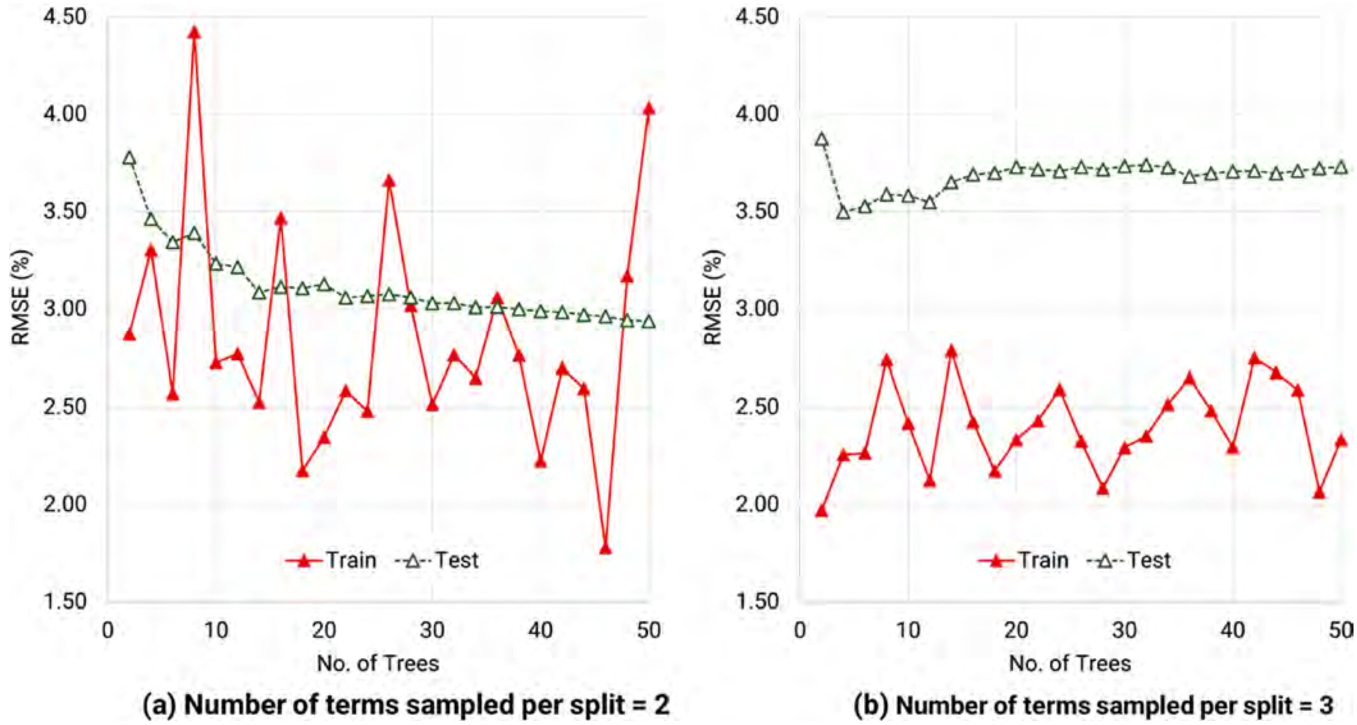

Figure 13. Effect of number of trees and number of terms sampled per split on performance of models for prediction of porosity.

(e.g. predicting permeability from porosity) may yield low accuracy. This will increase the uncertainty, reducing the effective prediction of models based on such approaches. It is, therefore, congruent to develop a multicriteria model to include the composite impact of parameters to predict the performance of pervious concrete.

4. Statistical analysis

In the present study, aggregate size (S), aggregate-to-cement ratio (A/C), number of compaction blows (B), and ultrasonic pulse velocity (UPV) were considered as input (independent) variables. In contrast, porosity (v) and permeability (*k*) were considered as desired output (dependent) variables. The

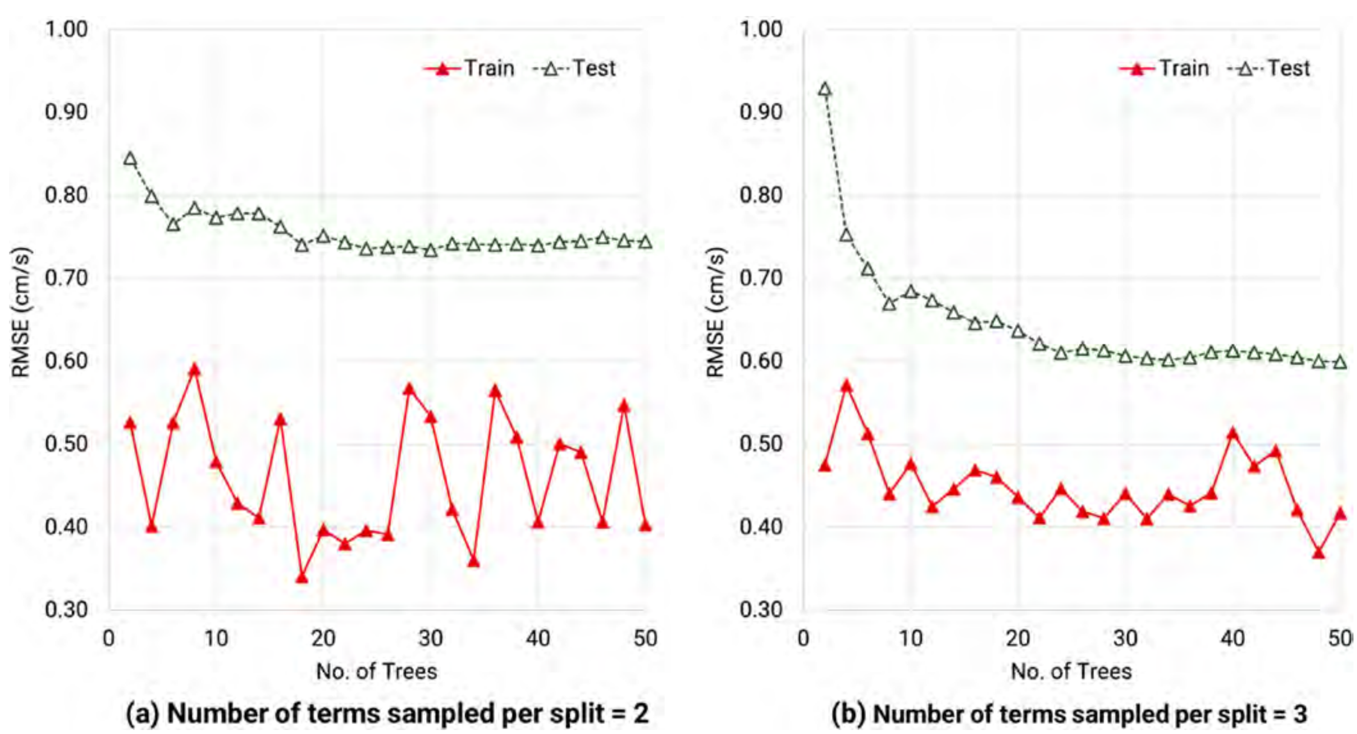

Figure 14. Effect of number of trees and number of terms sampled per split on performance of models for prediction of permeability.

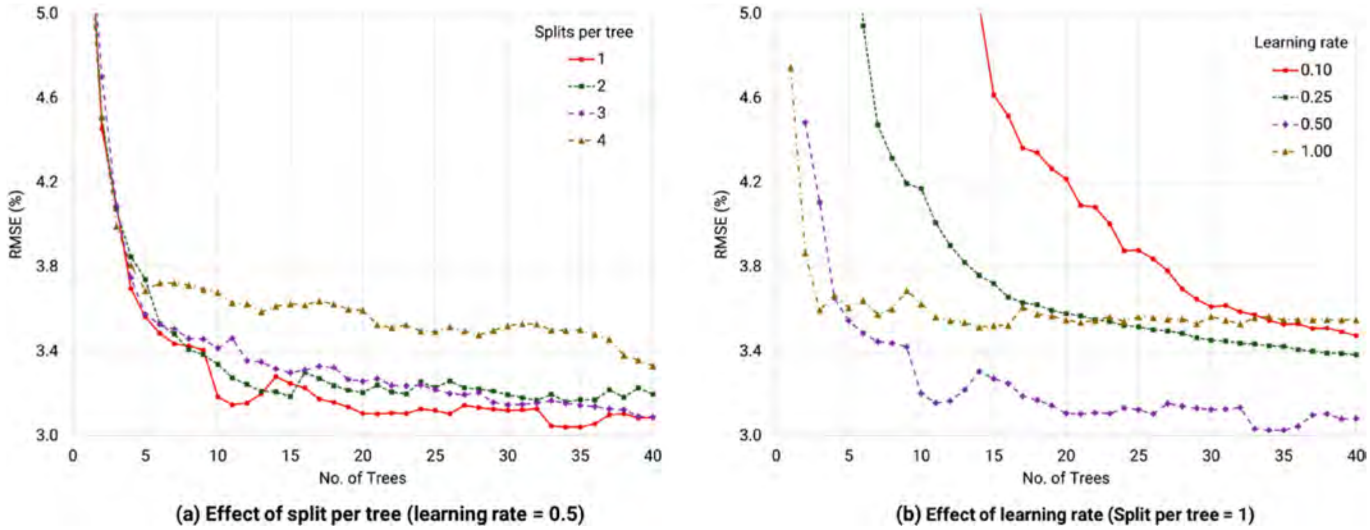

Figure 15. Effect of number of trees, split per tree and learning rate on performance of models for prediction of porosity.

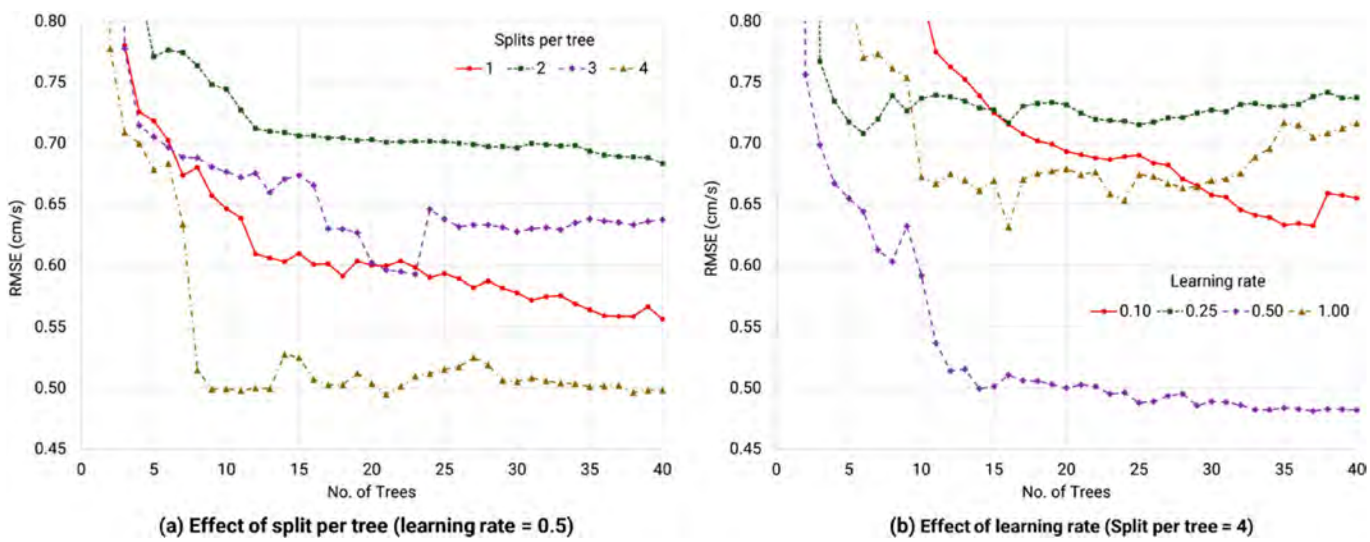

Figure 16. Effect of number of trees, split per tree and learning rate on performance of models for prediction of permeability.

descriptive statistics for the input and output data are illustrated in Figure 11. This Figure demonstrates that the parameter S ranged from 5–25 mm, the AC parameter ranged from 3.0 to 5.0, the B ranged from 0 to 60, and UPV ranged from 2.59 to 4.41 km/s. The output parameters: porosity ranged from 11 to 48%, and permeability ranged from 0.03 to 6.78 cm/s, respectively.

The correlation equation between the UPV and porosity of pervious concrete is given by Equation (24)

$$v = 464.51e^{-0.723UPV} \tag{24}$$

This equation is valid when the porosity varies from 3.5 to 48.4% and the UPV differs from 3.11 to 5.48 km/s. It reveals that the UPV and porosity of pervious concrete are well correlated with a negative exponential relationship with a correlation coefficient of 0.767.

The correlation equation between the UPV and porosity of pervious concrete is given by Equation (25).

$$k = 10.76 - 2.12UPV \tag{25}$$

This equation is valid when the permeability varies from 0.03 to 6.78 cm/s and the UPV varies from 3.11 to 5.08 km/s. It is revealed that there is a negative linear relationship between the permeability and UPV of pervious concrete. However, the correlation is moderate, with a correlation coefficient of 0.598.

The correlation between porosity and permeability, as shown in Equation (26) shows that porosity is directly proportional to permeability.

$$k = 0.12v{-}1.01 \tag{26}$$

This equation is valid when the permeability varies from 0.03 to 6.78 cm/s and the porosity varies from 8.4 to 48.4%. However, the correlation is moderate, with a correlation coefficient of 0.582. The increasing porosity results in increasing

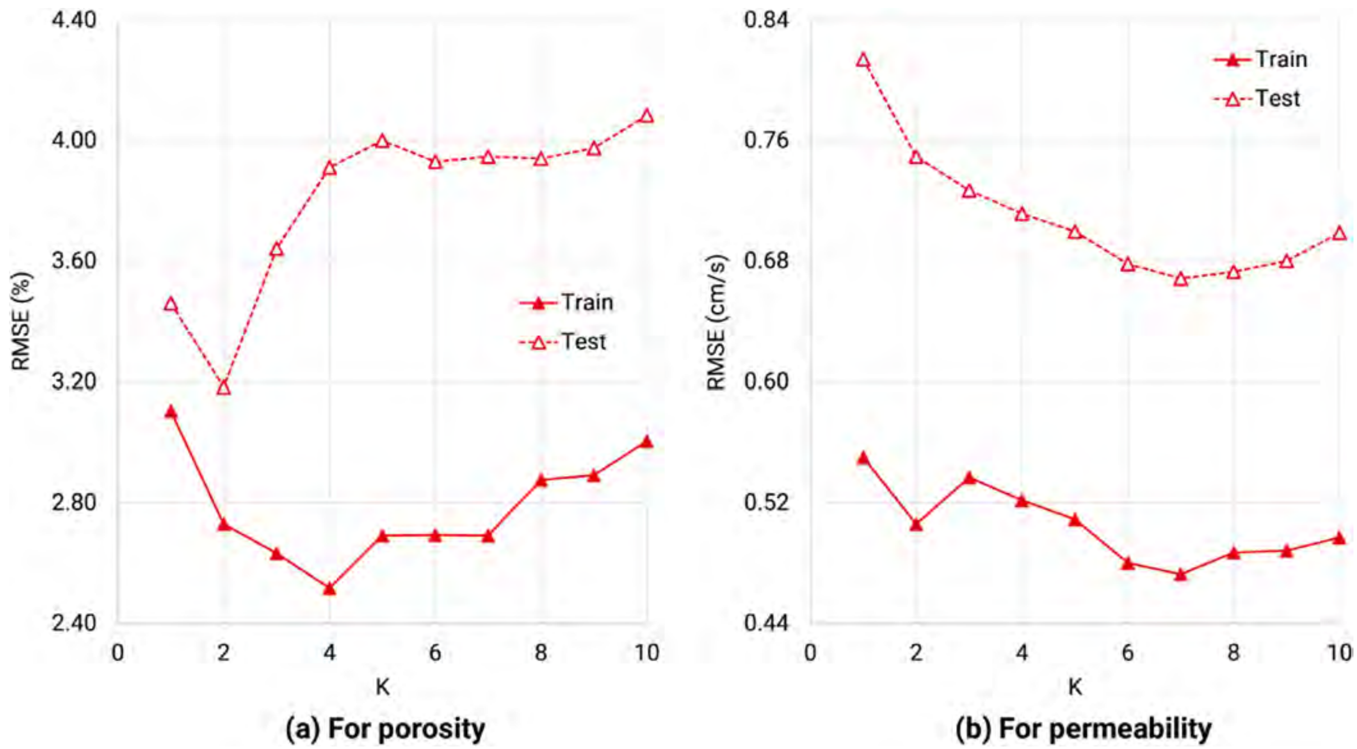

Figure 17. Effect of K on performance of models for prediction of (a) porosity and (b) permeability.

Table 5. Effect of kernel function, cost and gamma on performance of models for prediction of porosity.

Method	Kernel Function	Cost	Gamma	RMSE – Train (%)	RMSE – Test (%)	R^2 – Test	Remarks
1	Linear	0.04	.	3.88	4.38	0.7529	
2		0.29	.	3.79	4.37	0.7538	
3		0.55	.	3.79	4.38	0.7534	
4		0.81	.	3.79	4.38	0.7534	
5		1.12	.	3.79	4.38	0.7534	
6		1.43	.	3.79	4.38	0.7534	
7		1.71	.	3.79	4.38	0.7534	
8		2.01	.	3.79	4.38	0.7534	
9		2.25	.	3.79	4.38	0.7534	
10		2.46	.	3.79	4.38	0.7533	
11		2.68	.	3.79	4.38	0.7536	
12		2.90	.	3.79	4.38	0.7537	
13		3.15	.	3.79	4.38	0.7536	
14		3.38	.	3.79	4.38	0.7537	
15		3.64	.	3.79	4.38	0.7537	
16		3.89	.	3.79	4.38	0.7537	
17		4.13	.	3.79	4.37	0.7537	
18		4.36	.	3.79	4.38	0.7537	
19		4.56	.	3.79	4.37	0.7537	
20		4.96	.	3.79	4.38	0.7537	
21	Radial Basis Function	0.02	0.03	7.51	7.77	0.2237	
22		0.73	0.31	2.70	3.67	0.8270	
23		4.43	0.16	2.44	3.22	0.8662	Selected
24		2.86	0.07	2.90	3.45	0.8465	
25		3.33	0.48	1.89	3.55	0.8383	
26		4.77	0.32	2.03	3.51	0.8416	
27		0.34	0.20	3.19	3.85	0.8096	
28		3.06	0.28	2.20	3.36	0.8544	
29		4.95	0.46	1.84	3.58	0.8349	
30		2.50	0.44	2.00	3.49	0.8434	
31		3.52	0.13	2.56	3.24	0.8645	
32		2.39	0.02	3.41	3.95	0.7993	
33		3.86	0.39	1.97	3.54	0.8389	
34		1.33	0.12	2.88	3.49	0.8430	
35		4.65	0.09	2.67	3.28	0.8616	
36		1.79	0.40	2.18	3.58	0.8355	
37		4.11	0.23	2.26	3.26	0.8629	
38		2.01	0.25	2.40	3.34	0.8566	
39		1.11	0.49	2.35	3.80	0.8143	
40		0.16	0.36	3.74	4.46	0.7436	

Table 6. Effect of kernel function, cost and gamma on performance of models for prediction of permeability.

Method	Kernel Function	Cost	Gamma	RMSE – Train (cm/s)	RMSE – Test (cm/s)	R^2 – Test	Remarks
1	Linear	0.05	.	0.61	0.82	0.6103	
2		3.63	.	0.59	0.81	0.6156	
3		2.73	.	0.59	0.81	0.6155	
4		4.97	.	0.59	0.81	0.6156	
5		4.13	.	0.59	0.81	0.6156	
6		3.41	.	0.59	0.81	0.6156	
7		0.31	.	0.60	0.82	0.6132	
8		2.96	.	0.59	0.81	0.6156	
9		2.24	.	0.59	0.81	0.6156	
10		4.70	.	0.59	0.81	0.6156	
11		3.86	.	0.59	0.81	0.6156	
12		2.52	.	0.59	0.81	0.6156	
13		1.09	.	0.59	0.82	0.6152	
14		3.19	.	0.59	0.81	0.6156	
15		4.42	.	0.59	0.81	0.6156	
16		1.40	.	0.59	0.81	0.6156	
17		1.66	.	0.59	0.81	0.6156	
18		1.96	.	0.59	0.81	0.6155	
19		0.83	.	0.59	0.82	0.6144	
20		0.58	.	0.60	0.82	0.6136	
21	Radial Basis Function	4.22	0.49	0.30	0.71	0.7111	
22		2.74	0.11	0.45	0.71	0.7102	
23		0.40	0.10	0.54	0.76	0.6681	
24		0.11	0.28	0.65	0.81	0.6204	
25		4.36	0.25	0.36	0.70	0.7148	
26		4.50	0.02	0.51	0.75	0.6722	
27		4.93	0.13	0.42	0.70	0.7131	
28		2.66	0.46	0.33	0.70	0.7197	
29		2.37	0.22	0.40	0.70	0.7185	
30		0.61	0.36	0.46	0.71	0.7084	
31		3.94	0.17	0.40	0.70	0.7177	
32		4.67	0.35	0.33	0.71	0.7110	
33		3.00	0.31	0.36	0.70	0.7135	
34		0.27	0.50	0.54	0.75	0.6785	
35		3.73	0.41	0.32	0.70	0.7131	
36		1.53	0.07	0.49	0.72	0.6967	
37		3.57	0.04	0.48	0.73	0.6952	
38		0.92	0.19	0.46	0.71	0.7120	
39		2.00	0.39	0.37	0.70	0.7187	
40		1.63	0.45	0.37	0.69	0.7224	Selected

permeability. When the porosity is less than 8.4%, the permeability becomes zero.

Statistical analysis was performed to evaluate the degree of correlation among the parameters above after the descriptive analysis. Although this was previously reported, the current study had a wide range of experimental data. Pearson's correlation in Figure 4 shows less correlation among the mix parameters (S, AC, and B). On the other side, it was discovered that there was a far higher degree of association between UPV, porosity, and permeability. As expected, the independent parameters are not dependent on each other, while the dependent parameters had a dependency on different variables, including both dependent and independent variables.

5. Selection of hyperparameters

Fine-tuning refers to the meticulous process of making little modifications to attain the desired outcome or level of

Table 7. Hyperparameter optimisation for prediction models.

Model	Hyperparameters	Analyzed Values	Selected value for Porosity	Selected value for Permeability
Artificial Neural Network	Number of layers	[1, 2]	2	2
	Nodes per layer	[1–16]	12	12
Boosted Decision Tree Regression	Number of layers	[1–40]	34	21
	Split per tree	[1, 2, 3, 4]	1	4
	Learning rate	[0.1, 0.25, 0.5, 1]	0.5	0.5
KNN	K	[1–10]	2	7
Random Forest regression	Number of trees	[2–50]	46	48
	Number of terms sampled per split	[2, 3]	2	3
	Number of terms	5	5	5
	Minimum splits per tree	10	10	10
	Minimum size split	5	5	5
Support Vector Regression	Kernel	[linear, Radial basis function]	Radial basis function	Radial basis function
	Cost	[0.01–5]	4.43	1.63
	Gamma	[0.001–0.5]	0.16	0.45

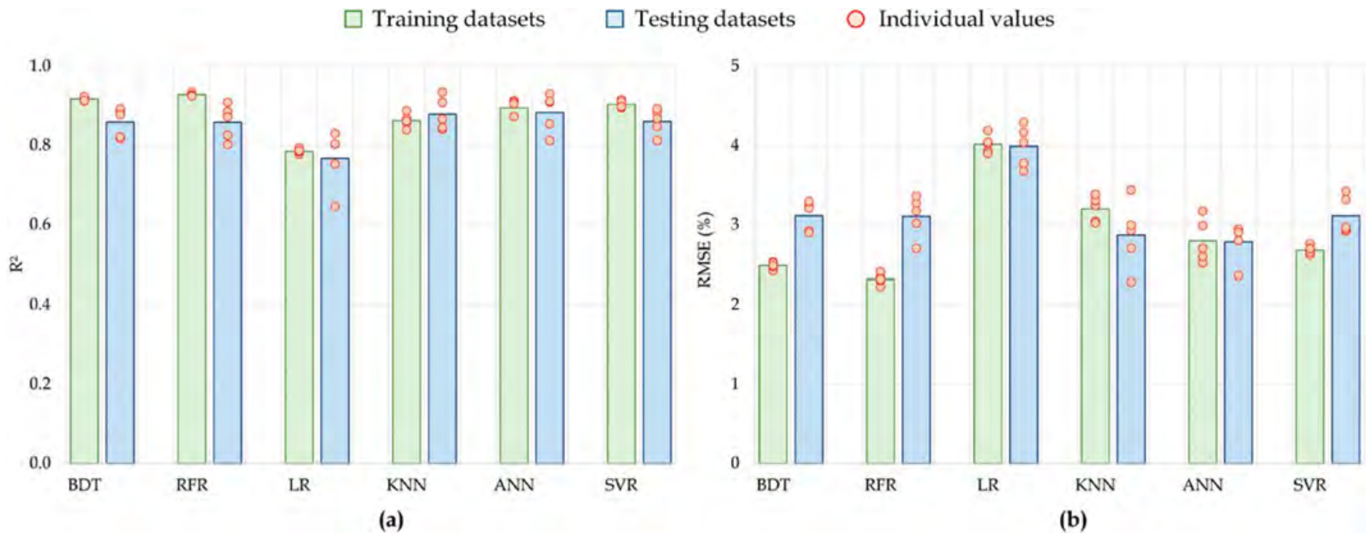

Figure 18. K-fold validation results for (a) R^2 and (b) RMSE variation across the folds.

performance. This procedure is commonly known as grid search or hyperparameter optimisation. To establish the most effective values for these parameters, it is recommended to experiment with various combinations of parameters across the entire range. This can be done by conducting cross-validation tests to identify the parameters that yield the highest performance (Elgeldawi *et al.* 2021). The ideal hyperparameters are subject to variation based on the particular dataset and task. Hence, engaging in experimentation and iteration on model is crucial to discovering the optimal settings.

5.1. Artificial neural networks

In the present study, fine-tuning an ANN involved experimenting with the number of layers and neurons in each layer. The number of layers considered was up to 2, and the number of neurons in each layer was up to 16. Figure 12 illustrates the variation of RMSE with different hyperparameters. For the porosity prediction also, the best performance was achieved with two layers and 12 neurons. The best performance for predicting permeability was with a two-layer neural network containing 12 neurons. Optimum hyperparameters were selected based on the performance of both the training and testing datasets.

5.2. Random forest regression

Fine-tuning a Random Forest regression model involves adjusting various hyperparameters such as the number of trees in the forest, number of terms sampling per split, bootstrap sample rate, minimum splits per tree, maximum splits per tree and minimum split. Published literature shows that the number of trees in the forest and the number of terms sampling per split highly influence the prediction's accuracy (Badarna and Shimshoni 2019, Makariou *et al.* 2021).

Figures 13 and 14 show the relationship between RMSE and several hyperparameters used to predict porosity and permeability, respectively. For porosity prediction, using a sampling rate of 3 terms per split gives superior performance for training datasets. However, the performance for test datasets is inferior compared to using a sampling rate of 2 terms per split. Performance is improved when the number of terms sampled per split is 2, and the number of trees in the forest is 46. Performance improves when using a split of 3 sampled words and a forest of 48 trees to predict permeability.

5.3. Boosted tree regression

Fine-tuning a boosted tree regression model involves adjusting various hyperparameters such as the number of tresses, splits per tree, learning rate and minimum size split. Figures 15 and 16 show the relationship between RMSE and several hyperparameters used to predict porosity and permeability, respectively. For the fine-tuning, the number of trees ranged from 1 to 40, split per tree ranged from 1 to 4, and a learning rate of 0.1, 0.25, 0.5 and 1 were considered. For porosity prediction, when the learning rate is fixed as 0.5, the best performance occurred for split per tree equal to one for most cases. When varying the learning rate, the learning rate of 0.5 provides the best performance. Based on these results, the number of trees of 34, split per tree of 1 and learning rate of 0.5 were used to predict porosity. Similarly, the number of trees of 21, split per tree of 4 and learning rate of 0.5 were used to predict permeability.

5.4. K-nearest neighbour

To fine-tune a K-nearest neighbours (KNN) model, different values of k (number of neighbours) were used to find the optimal balance between bias and variance. Cross-validation was used to assess the model's performance with different k values and select the one that gave the best results. Figure 17 shows the variation of RMSE with varying values of k. For porosity prediction, $k = 4$ performs best for the training data set. However, when both training and test data sets are considered, $k = 2$ serves best. On the other hand, for permeability prediction, $k = 7$ gives the best performance for both training and test data sets. Based on these results, $k = 2$ and $k = 7$ for porosity and

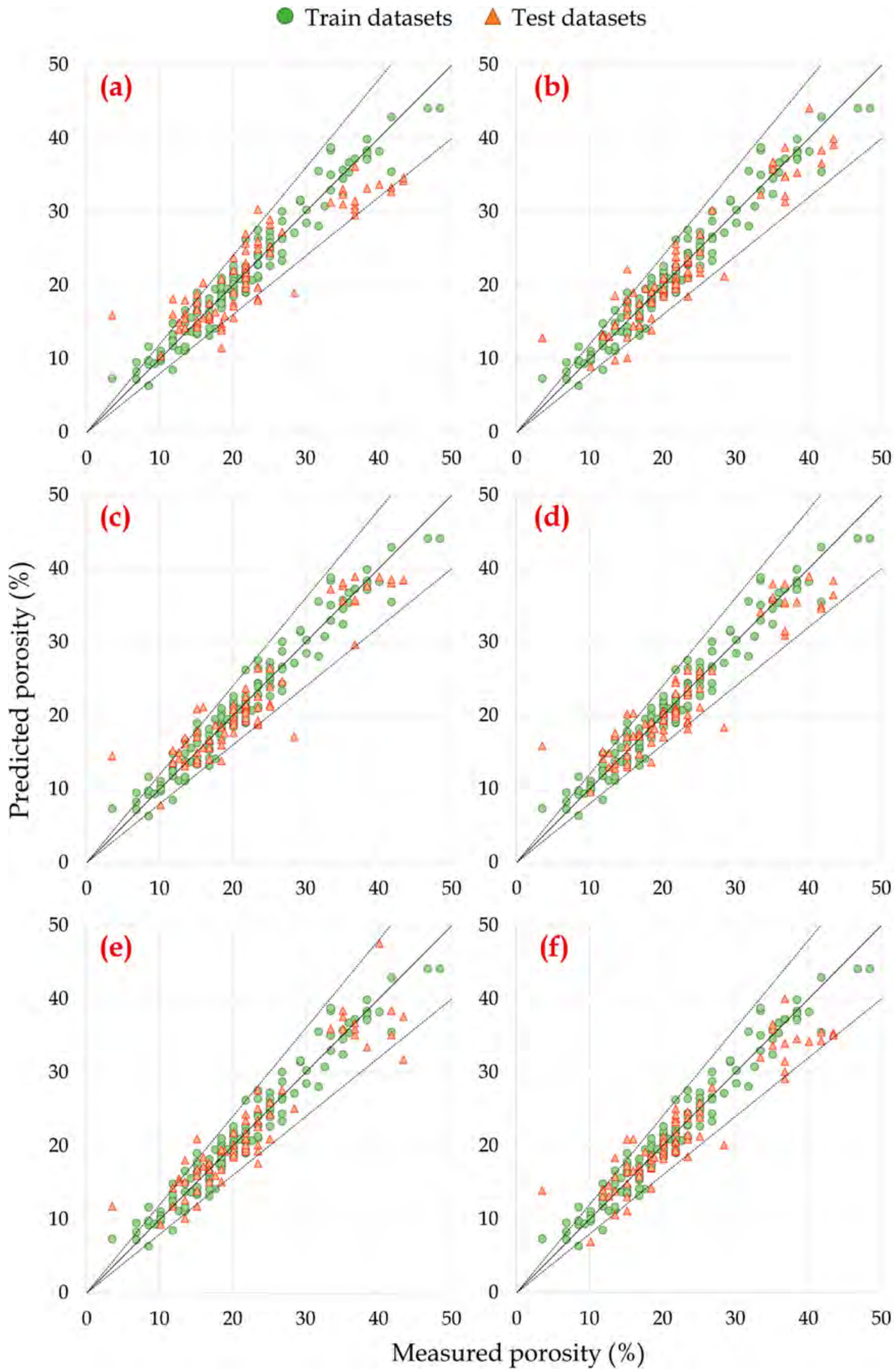

Figure 19. Predicted porosity vs. measured porosity for (a) LR, (b) ANN, (c) BDT, (d) RFR, (e) KNN and (f) SVR models (Note that middle continuous line indicates the predicted porosity = measured porosity, upper dotted line indicates predicted porosity = 1.2 measured porosity and lower dotted line indicated the predicted porosity = 0.8 measured porosity).

Table 8. Performance indicators for ML techniques used.

	Train				Test			
	R^2	RMSE (%)	MAE (%)	a20	R^2	RMSE (%)	MAE (%)	a20
ANN	0.9502	1.91	1.48	92.67	0.8958	2.87	2.23	89.33
KNN	0.9000	2.75	2.05	87.33	0.8805	3.06	2.20	88.00
SVR	0.9131	2.56	1.69	90.00	0.8685	3.47	2.57	85.33
RFR	0.9217	2.47	1.82	88.00	0.8721	3.29	2.33	88.00
BDT	0.9147	2.50	1.84	88.00	0.8693	3.26	2.40	88.00
LR	0.7992	3.83	2.88	72.00	0.7915	4.26	3.32	72.00

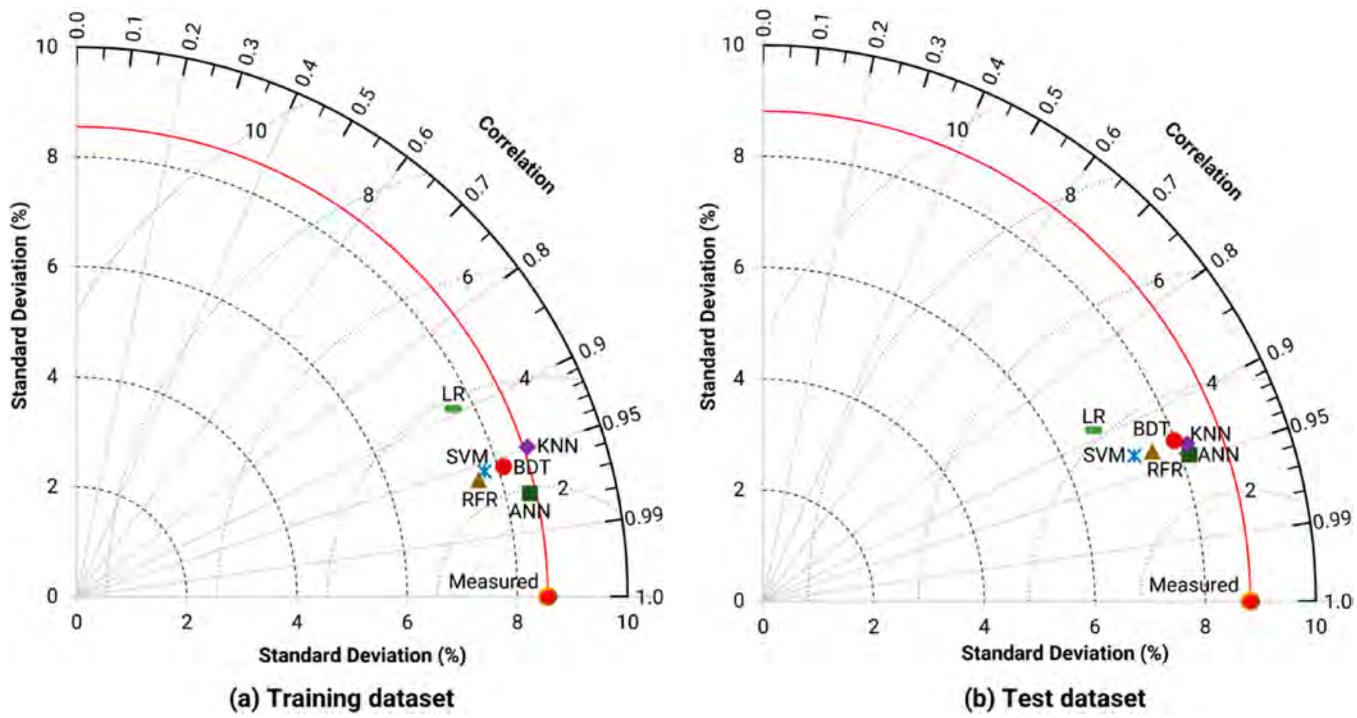

Figure 20. Taylor diagram showing the performance of the ML models. The red line represents the measured porosity.

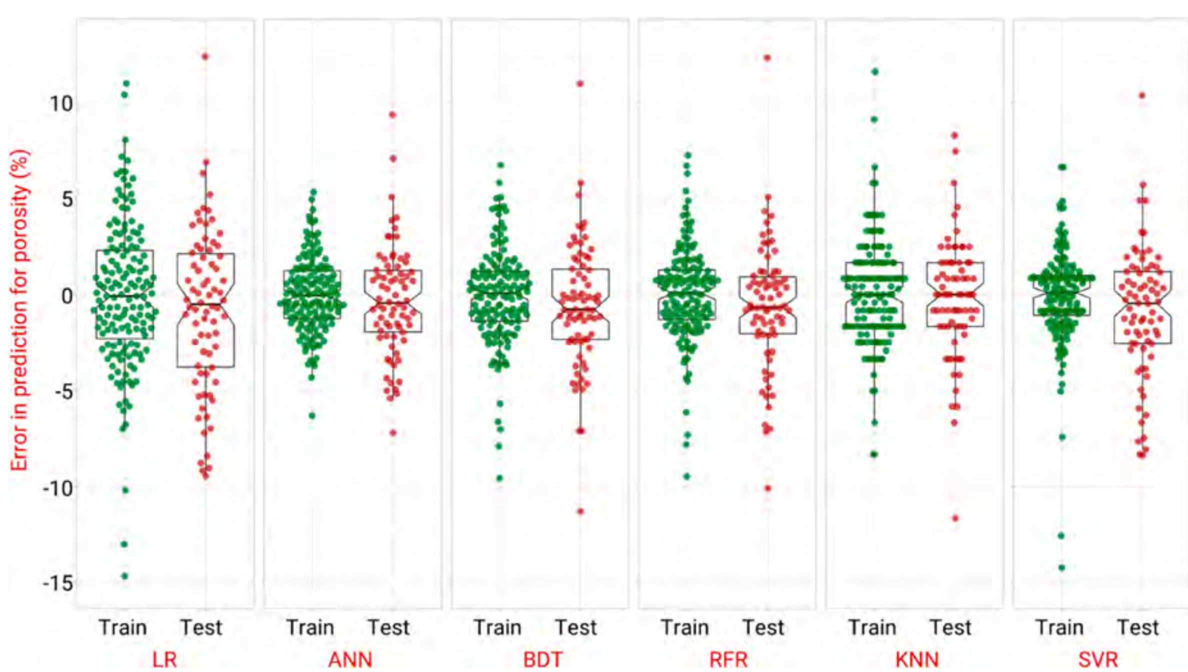

Figure 21. Error distribution between predicted vs. measured porosity for various ML models.

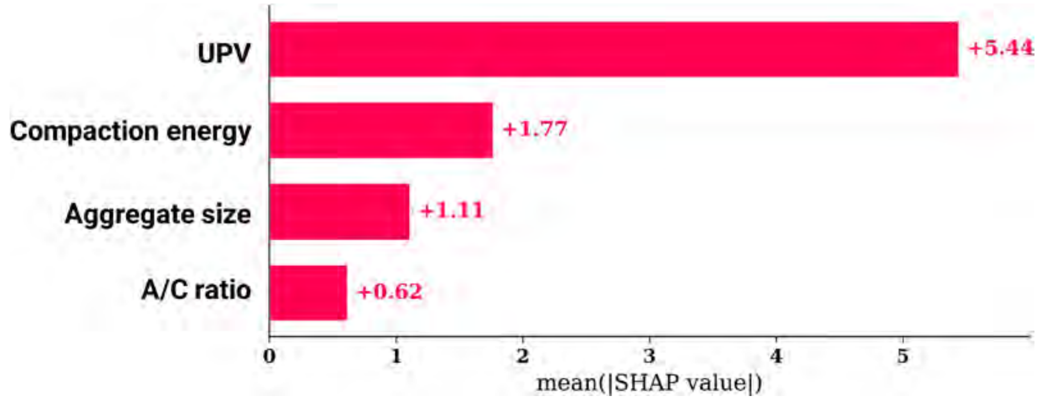

Figure 22. Mean SHAP values for the RFR model.

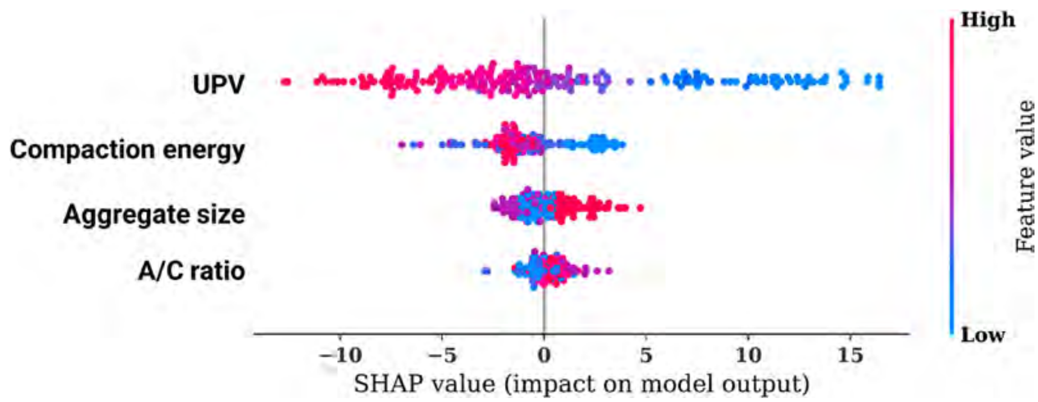

Figure 23. SHAP Summary Plot for compressive strength using RFR model.

permeability prediction, respectively, were used for further analysis.

5.5. Support vector regression

Its parameters must be optimised to improve the performance of a Support Vector Regression (SVR). SVR can use linear and non-linear kernels, such as the Radial Basis Function. A linear kernel is a simple dot product between two input vectors, while a Radial Basis Function is a more complex function that can capture more intricate patterns in the data (Amari and Wu 1999). SVR has several hyperparameters, namely cost and gamma, that can be adjusted to control the model's behaviour. Cost and Gamma are hyperparameters set before the training model to control error and give curvature weight to the decision boundary. Tables 5 and 6 summarise the variation of R^2 and RMSE with different hyperparameters for predicting porosity and permeability, respectively. The Radial Basis Function, with a cost of 4.43 and gamma of 0.16, performs better in predicting permeability. For permeability prediction, the radial basis function with a cost of 1.63 and gamma of 0.45 offers better performance.

5.6. Selected hyperparameters

Based on the optimal features derived from several feature selection models, six distinct machine learning algorithms have been employed to ascertain the algorithm that exhibits the highest level of accuracy in forecasting deck conditions. The analysis has focused on the most often observed ranges of hyperparameters, given the large number of potential combinations. Table 7 summarises the hyperparameter value range and optimal value employed in the current investigation for the machine learning method. The evaluation of each model's performance was conducted by assessing the R^2 values on both the training and test datasets to mitigate the risk of overfitting.

Standardising or normalising predictor variables in machine learning is often beneficial when they vary on significantly different scales. While standardising predictors is usually recommended for many machine learning algorithms, it may not be strictly necessary in some cases. Decision trees and random forest models are mainly not sensitive to the scale of predictor variables. These algorithms make decisions based on relative comparisons between variables at each split, so the scale of the variables does not affect their performance. However, Algorithms that rely on distances between data points, such as k-nearest neighbours, can be sensitive to the scale of the variables. However, in KNN, the distances may cancel out the scale differences if the features are on different scales but have similar variances. Some algorithms are inherently robust to differences in scale. For instance, support vector regression with a radial basis function kernel can often handle varying scales, and robust scaling techniques might be sufficient. While standardisation is a good default

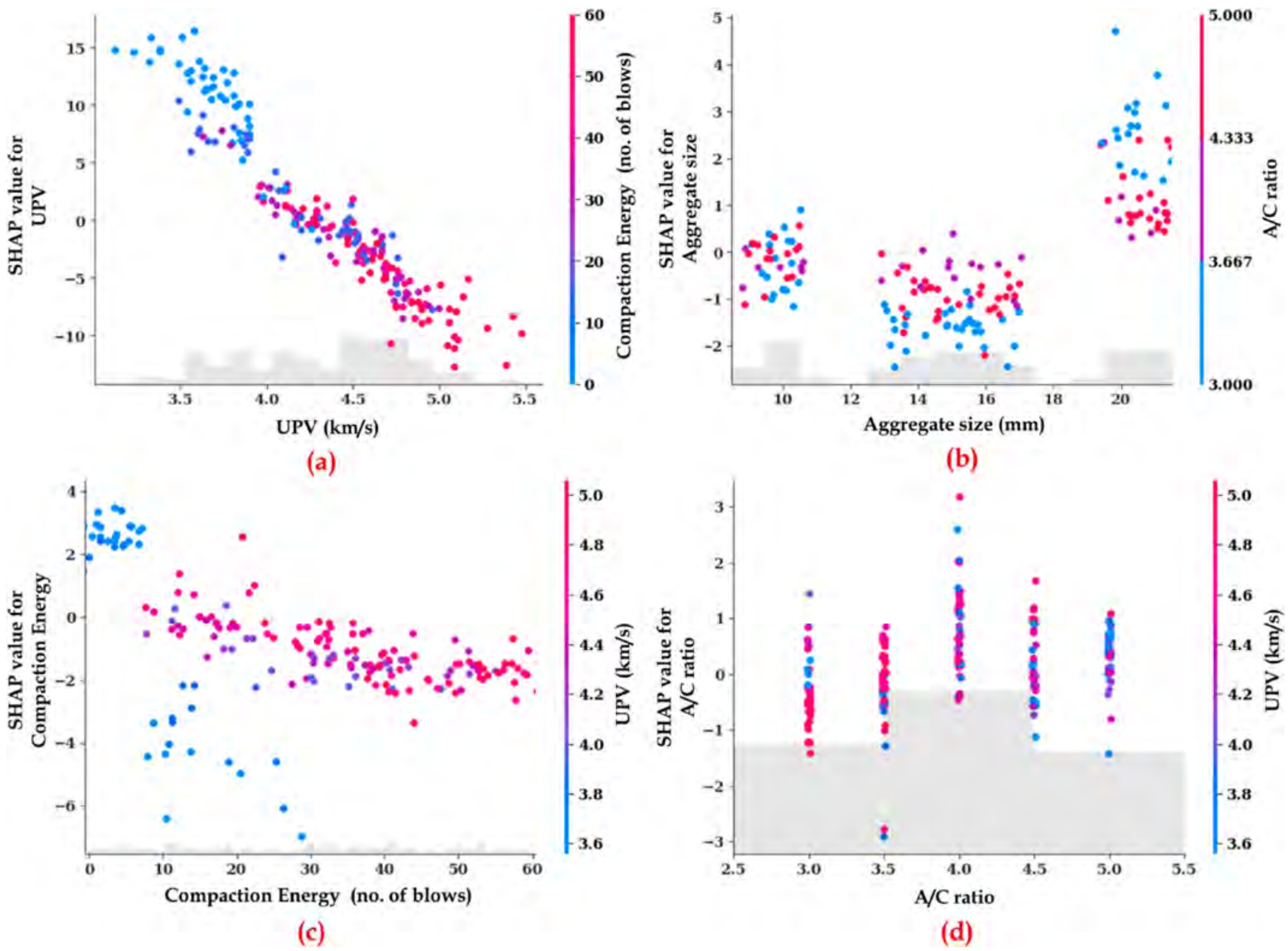

Figure 24. Feature interaction plot: (a) UPV, (b) aggregate size, (c) compaction energy, and (d) A/C ratio.

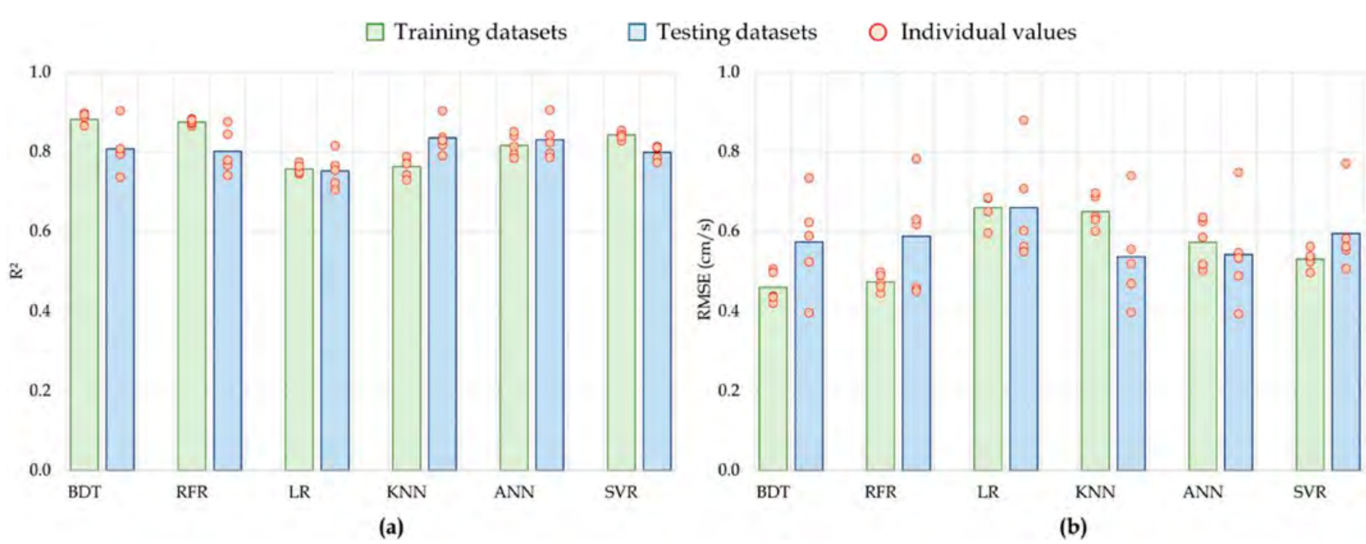

Figure 25. K-fold validation results for (a) R^2 and (b) RMSE variation across the folds.

Table 9. Performance indicators for each model.

	Train R^2	RMSE (cm/s)	MAE (cm/s)	a20	Test R^2	RMSE (cm/s)	MAE (cm/s)	a20
ANN	0.8893	0.43	0.28	58.00	0.7553	0.70	0.41	52.00
KNN	0.8533	0.51	0.33	49.33	0.7026	0.77	0.48	45.33
SVM	0.9145	0.38	0.23	62.00	0.7616	0.69	0.44	42.67
RFR	0.8677	0.48	0.32	49.33	0.7856	0.66	0.41	45.33
BDT	0.9323	0.34	0.23	63.33	0.7574	0.70	0.46	42.67
LR	0.7914	0.59	0.44	40.67	0.7048	0.78	0.52	41.33

pre-processing step, the necessity depends on the specific characteristics of both the algorithm and the data.

6. Prediction model for porosity

6.1. K-fold validation

The results of K-fold cross-validation are evaluated by an R^2 and RMSE, as illustrated in Figure 18. All the models except the LR model show a narrow range of R^2 and lower RMSE for both train and test datasets. The RFR model shows an R^2 between 0.92 and 0.94 for all the folds for training datasets, which is the best out of all the ML models. However, KNN offers better accuracy for test datasets than other ML models. Overall, the RFR model performs exceedingly well in training and testing datasets. These results demonstrate no overfitting issue for ML models except the LR model, and they can be considered more reliable machine learning models for prediction.

6.2. Performance indicators and correlations

Figure 19 shows how well the six ML models predicted the porosity values compared to the actual measurements. The RFR model had the most data points within the 20% error margin, which means it was the most accurate model. The ANN model had an α20-index of 92.67%, followed by the SVR model with 90% for train datasets. The ANN model has an α20-index of 89.33% for test datasets, followed by the KNN, RFR and BDT models, all shown with 88%. These results reveal the presence of outliers, suggesting some inaccuracy. This could be due to errors in the experimental measurement of porosity. Table 4 shows variations in porosity even for a specific mix, particularly at zero compaction level. This is affecting the accuracy of the prediction models. Using the average porosity value instead of individual values may be helpful to reduce error. However, this would also reduce the number of datasets available for analysis.

Table 8 presents the performance evaluation metrics for each of the six ML techniques. The ANN model had the highest R^2 values of 0.9502 and 0.8958 for train and test datasets, respectively. It indicates that the ANN model predicted porosity very well. The linear regression model had the lowest R^2 value and the worst prediction. The other models had moderate R^2 values between 0.90 and 0.92 for train datasets and 0.86 to 0.88 for test datasets. The KNN model also had the lowest RMSE and MAE values, with high accuracy and precision among all the tested models.

6.3. Performance of the ML models

This study assessed the model's predictability using a Taylor diagram. The model prediction and the actual porosity are more consistent when the pentagram is closer to the measured point, which represents similar statistical features as shown in Figure 20. Among all the ML models, the ANN model is the closest to the red line, while the linear regression model is the farthest. The ANN model also has a higher correlation value than other models, as indicated by its proximity to the horizontal black line. The relative performance of different machine learning models can depend on the specific characteristics of the dataset and the nature of the prediction task. ANNs are particularly well-suited for tasks where the relationships between input features and the target variable are complex and non-linear. If the porosity prediction problem involves intricate dependencies that a complicated non-linear model better captures, ANNs can excel. Also, ANNs can automatically learn hierarchical representations of features through their hidden layers. This feature learning ability allows ANNs to discover and capture relevant patterns and representations in the data, which may be challenging for ensemble methods like RFR and BDT. A comparison of the closeness revealed that the performance of the ML models was in the order of ANN > KNN > BDT > RFR > SVR > LR.

6.4. Error distribution

Figure 21 displays each ML method's error (predicted porosity – observed porosity) for training and testing datasets. The graph reveals that some of the observed data points are outliers, which means they have a significant error. This could be attributed to some error in the experimental measurement of porosity. The ANN model has more consistent prediction accuracy than the other models, which have fewer scattered errors. The ANN, KNN and RFR models have errors almost equally distributed on both sides of zero, meaning they sometimes overestimate and underestimate the porosity. However, the SVM model has an average error of – 0.36%, often underestimating the porosity. Therefore, using multiple statistical metrics and graphical representations can help evaluate the performance of predictive models more thoroughly.

6.5. Sensitivity analysis

Non-linear and complex models like ML sometimes act similar to 'black boxes' due to their complexity (Shah *et al.* 2022). SHAP (SHapley Additive exPlanations) is proven effective in analysing complex machine-learning models with multiple parameters (Quan Tran *et al.* 2022, Zhang *et al.* 2022a).

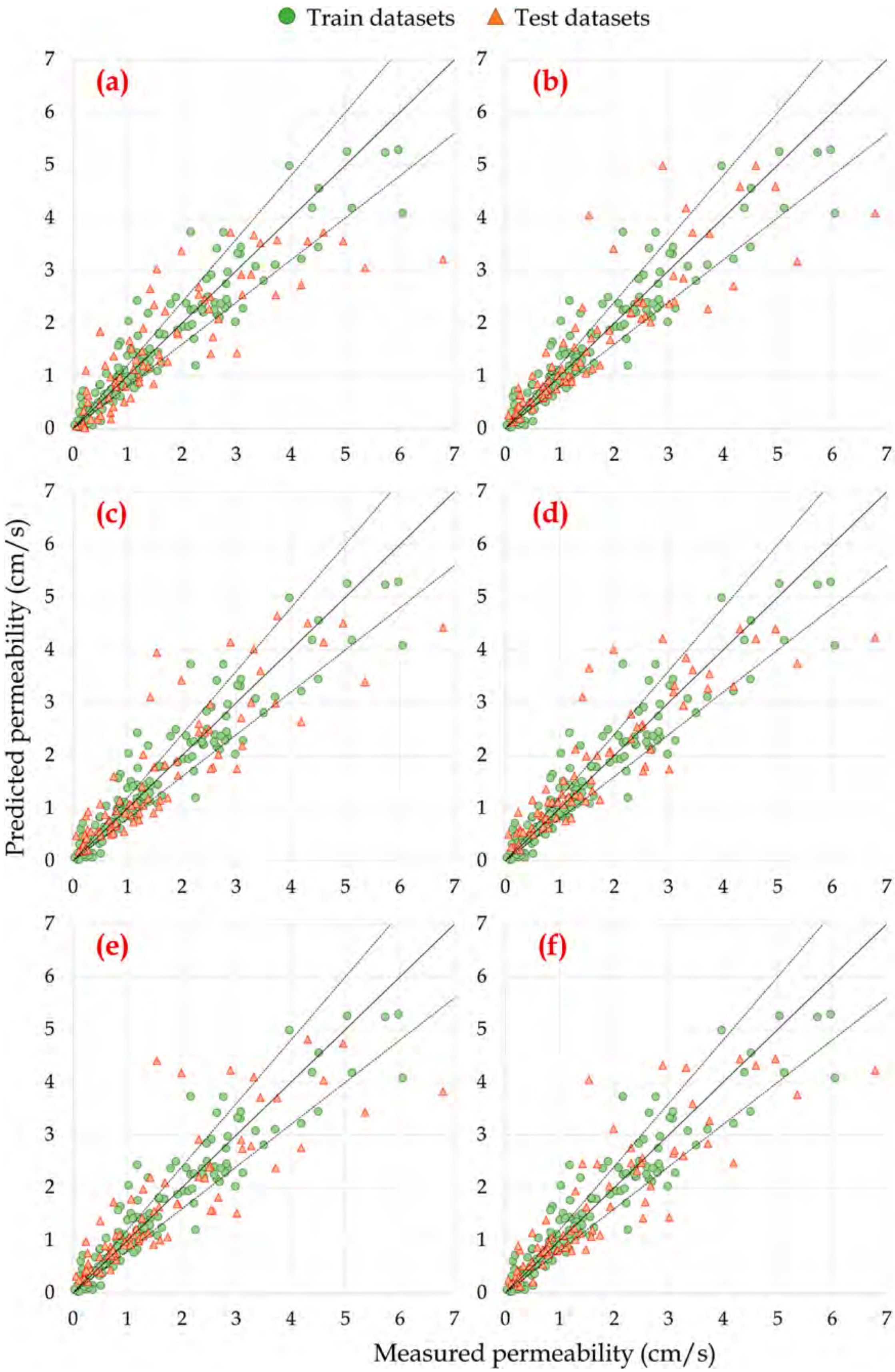

Figure 26. Predicted permeability vs. measured permeability for (a) LR, (b) ANN, (c) BDT, (d) RFR, (e) KNN and (f) SVR models (Note that middle continuous line indicates the predicted permeability = measured p permeability, upper dotted line indicates predicted permeability = 1.2 measured permeability and lower dotted line indicated the predicted permeability = 0.8 measured permeability).

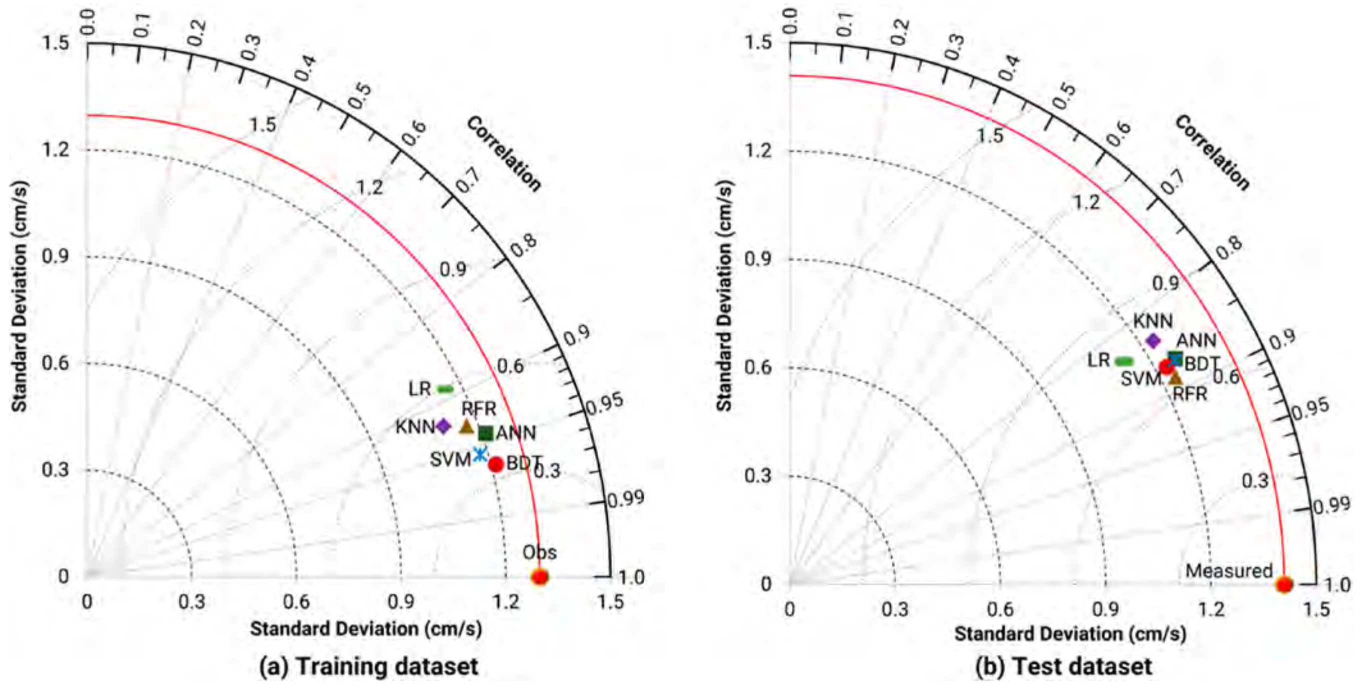

Figure 27. Taylor diagram showing the performance of the ML models for input parameters such as compaction energy, A/C ratio, aggregate size and UPV.

SHAP is a unified approach to explaining the output of any machine learning model. It is based on cooperative game theory and assigns a unique contribution value to each feature for a given prediction. The first step in SHAP analysis is to define input and output parameters. In the context of machine learning, the process involves predicting the output for a given instance based on the model's features. The second step is to generate a power set, which creates all possible combinations (subsets) of features. This represents all possible coalitions of features. The third step is to evaluate the contributions of each coalition of features, calculating the model's prediction with and without the features in that coalition. The difference in the predictions represents the marginal contribution of those features to the model's output. The fourth step is to calculate Shapley values using the Shapley value formula, which calculates the average marginal contribution of each feature

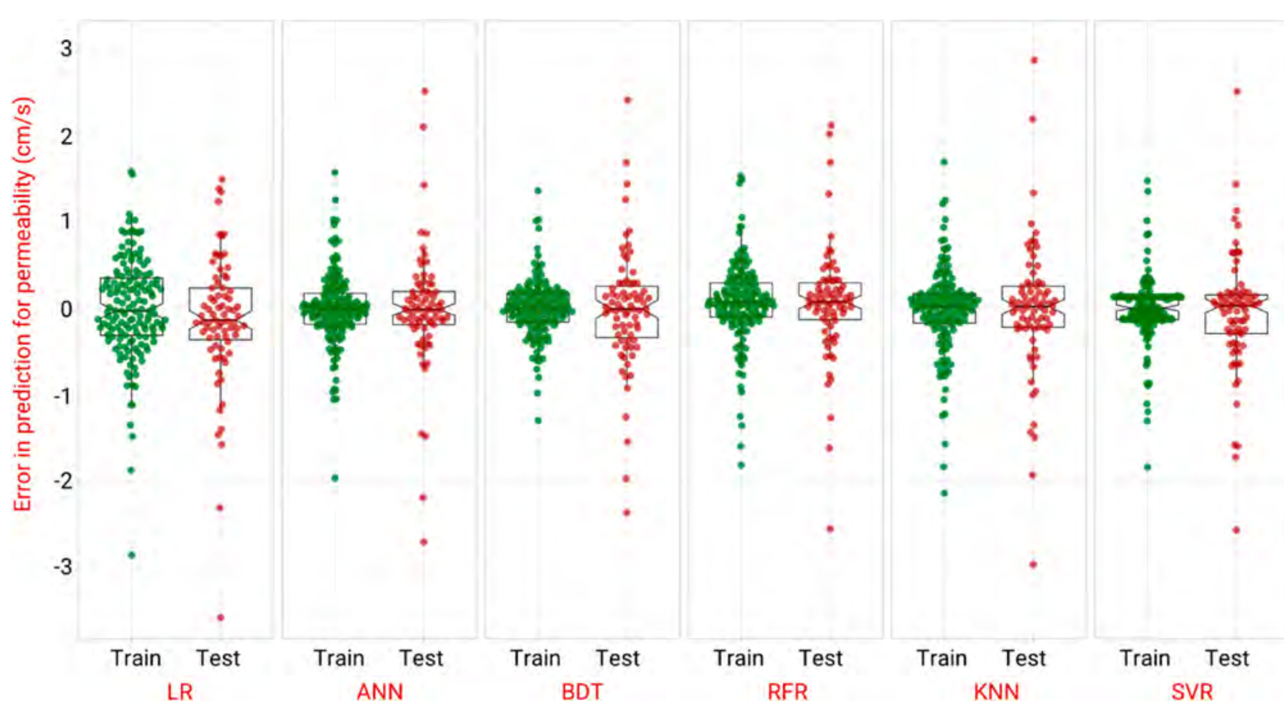

Figure 28. Error distribution of predicted value and measured permeability by ML models for considering input parameters as compaction energy, A/C ratio, aggregate size and UPV.

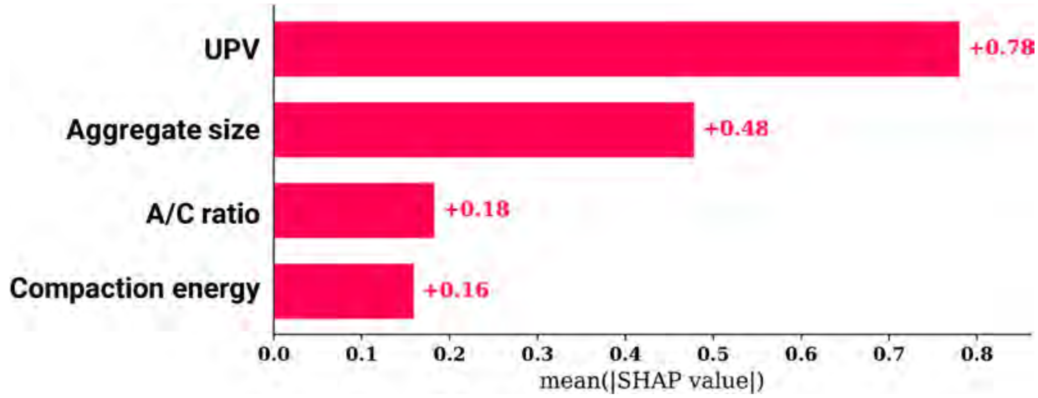

Figure 29. Mean SHAP values using RFR model by considering input parameters as compaction energy, A/C ratio, aggregate size and UPV.

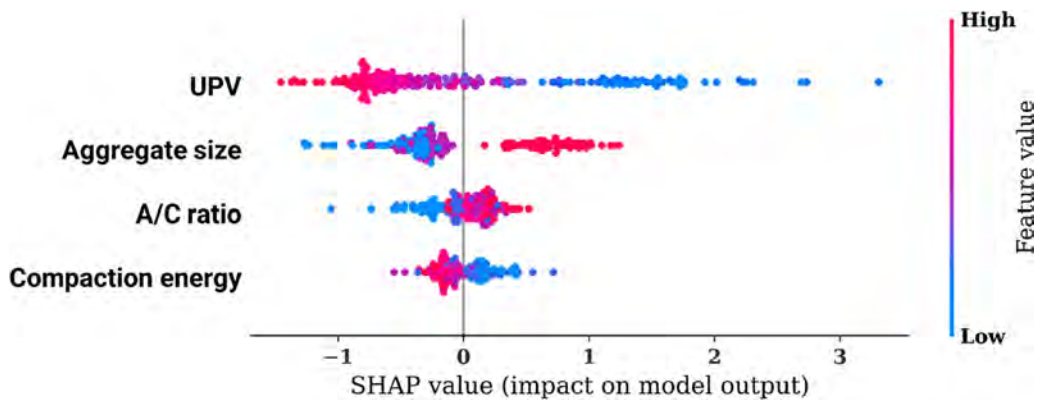

Figure 30. SHAP Summary Plot for compressive strength using RFR model by considering input parameters as compaction energy, A/C ratio, aggregate size and UPV.

across all possible coalitions. The Shapley value for feature ‘i’ is the average marginal contribution of feature ‘i’ across all possible coalitions and is calculated using Equation (27) (Lundberg and Lee 2017).

$$\emptyset_i(f) = \sum_{S \subseteq N\backslash\{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}[f(SU\{i\}) - f(S)] \quad (27)$$

Where

N is the set of all features.

S is a subset of features that does not include ‘i’.

$f(S)$ is the model prediction for the subset S.

$f(SU\{i\})$ is the model prediction for the subset S, including feature ‘i’.

Shapley values represent the fair contribution of each feature to the model’s prediction for a given instance. The final step is to summarise and visualise the results. The Shapley values can be used to understand the impact of each feature on the model’s output for a particular instance, and features with higher Shapley values have a more significant influence on the prediction. Visualise the Shapley values using different plots, such as summary plots, power plots or waterfall charts, to help interpret the results.

Since ANN had the best performance in predicting porosity, the SHAP results were obtained using the prediction model of ANN model. Figure 22 shows the average SHAP values for the porosity predictions based on the ANN model for different features (independent variables). The results indicate that UPV has the highest SHAP value (i.e. the most impact on prediction) for the porosity prediction. UPV is highly sensitive to changes in the structure and composition of materials. Variations in porosity often result in changes in the speed of sound waves through a material, making UPV an effective tool for detecting and quantifying porosity. The compaction energy also directly affects the voids within a material. High compaction generally leads to low porosity as they reduce the voids. Therefore, UPV and compaction energy were the most influential parameters in predicting porosity. While the aggregate size and aggregate-to-cement ratio influences are essential in determining pervious concrete’s overall composition and strength, they may not be as directly related to porosity or permeability as aggregate size or UPV. They are more closely related to compressive strength. Therefore, UPV shall be considered the critical factor in predicting porosity. Moreover, all other input parameters had the lowest SHAP values, suggesting they had little effect on porosity prediction compared to UPV.

Figure 23 presents the SHAP summary plot for the porosity predictions from the ANN model. The colour specifies the range of the feature values. The horizontal axis shows the SHAP value, representing the feature’s contribution to the porosity prediction. A red dot means the feature has a high and positive SHAP value. This implies that a higher aggregate

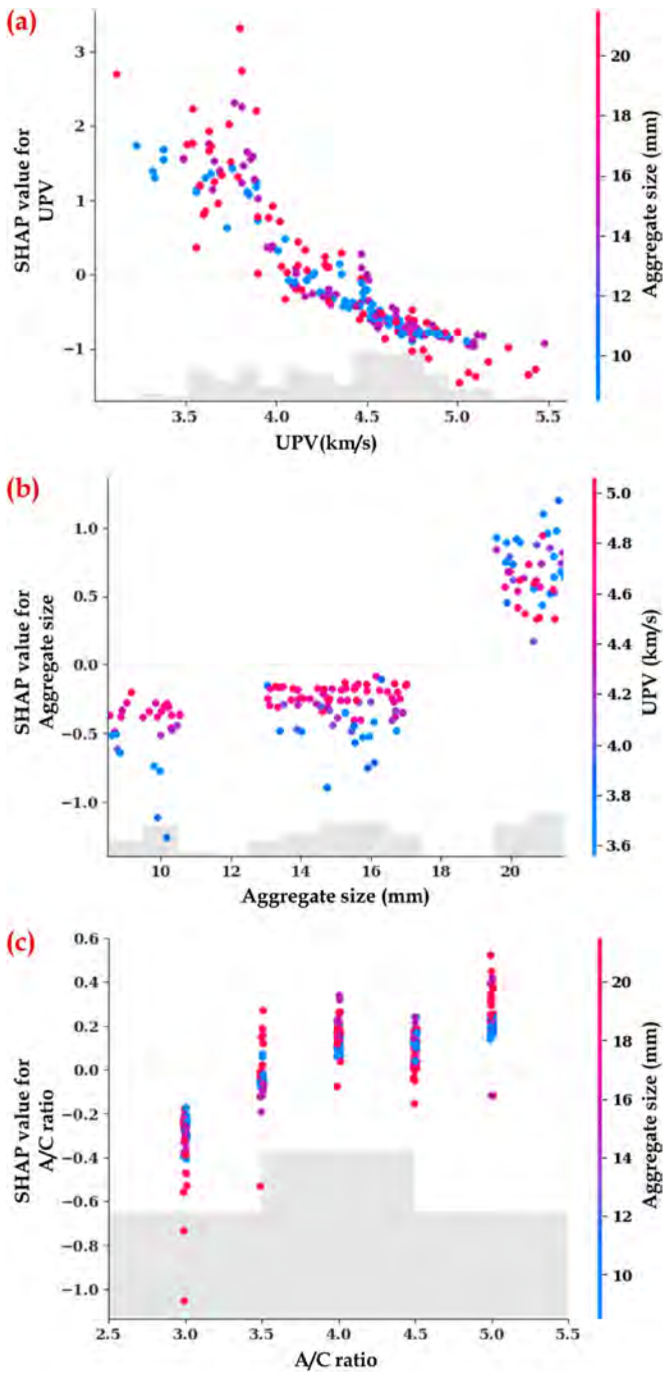

Figure 31. Feature interaction plot: (a) UPV, (b) aggregate size, and (c) A/C ratio.

size has a higher porosity, while higher UPV and compaction energy have a lesser porosity. The highest SHAP value, which is 16 on the right side (positive), suggests that compared to the average value, the UPV range used in this study can increase porosity by 16%. Similarly, a SHAP value of – 13 on the far left (negative) means that the porosity may decrease by 13% below the average value if the UPV is increased. These SHAP results show that the proposed hybrid ML models can be interpreted more clearly using the game theory method, and they also indicate that the model's accuracy is acceptable.

Figure 24 shows the correlation between characteristics and the porosity of pervious concrete. Figure 24(a) demonstrates the relationship between UPV feature interaction and its direct impact on porosity. The combination of low compaction and low UPV has a beneficial effect on porosity. Figure 24(b) illustrates the negative correlation between aggregate size and A/C ratio, which leads to reduced porosity. The aggregate size feature exhibited positive and negative consequences, contingent upon the A/C ratio. The impact of UPV on compaction energy and A/C ratio is demonstrated in Figure 24(c) and (d), respectively. Low UPV positively influenced porosity in cases with insufficient compaction, while high UPV had a detrimental effect on porosity as compaction energy increased. The influence of UPV on porosity exhibited a varied tendency in relation to the A/C ratio.

7. Prediction model for permeability

7.1. K-fold validation

The results of K-fold cross-validation are evaluated by an R^2 and RMSE, as illustrated in Figure 25. The RFR and BDT models show a narrow range of R^2 and lower RMSE for train datasets and KNN models show better accuracy for test datasets. The RFR model shows an R^2 between 0.86 and 0.89 for all the folds for training datasets, which is the best out of all the ML models. However, KNN offers better accuracy for test datasets than other ML models. Overall, the RFR model performs exceedingly well in training and testing datasets. These results demonstrate no overfitting issue for ML models except the LR model, and they can be considered more reliable machine learning models for prediction.

7.2. Performance indicators and correlations

The performance indicators of six ML models are presented in Table 9. The permeability of pervious concrete predictions made by all machine learning models except the linear regression model had a strong correlation, as indicated by the R^2 values ranging from 0.85 to 0.93 for train datasets and 0.70 to 0.78 for test datasets. Among the analysed models, the BDT model exhibits a reduced dispersion of data points, as evidenced by its R^2 value of 0.9323 for train datasets and 0.7574 for test datasets. Also, it had lower RMSE and MAE.

Figure 26 shows the predicted and measured permeability values for each of the six machine-learning models. The RFR displays more points inside the 20% error envelope. The BDT model had inliers for permeability of 63.33% for train datasets and 42.67% for test datasets. The LR model shows only 40.67% of data points inside the 20% error envelope for train datasets. Similar to porosity, this result reveals the presence of outliers, suggesting some inaccuracy. This could be due to errors in the experimental measurement of porosity. Table 4 shows variations in permeability even for a specific mix, particularly for a mix with a larger aggregate size. This is affecting the accuracy of the prediction models. Using the average permeability value instead of individual values may be helpful to reduce error.

7.3. Performance of the ML models

Figure 27 shows Taylor diagrams to evaluate the accuracy of the permeability prediction. The model prediction and the

Table 10. Summary of published literature on prediction models for porosity and permeability of pervious concrete.

Ref	Input parameters	No of datasets	ML algorithms used[b]	Response parameter	Best ML models	Accuracy
Sun *et al.* (2019)	Aggregate size, A/C ratio, W/C ratio	90	• LR • MLR • SVR	Permeability	SVR	$R^2 = 0.9635$, RMSE = 0.218 mm/s
Ibrahim *et al.* (2014)	Cement content, Water content, Aggregate content, Aggregate size	24	• MLR	Permeability	MLR	$R^2 = 0.75$
Adewumi *et al.* (2016)	Aggregate size, Cement content, W/C ratio, Aggregate content	24	• SVR	Porosity	SVR	$R^2 = 0.8697$, RMSE = 3.452 mm/s
Le *et al.* (2022)	A/C ratio, W/C ratio, Aggregate size, Sand content, Silica fume content	164	• ANN • RFR • SVM • XGB	Porosity	XGB	$R^2 = 0.88$, RMSE = 3.2%
Huang *et al.* (2020)	A/C ratio, Aggregate size, Aggregate gradation	69	• RFR	Permeability	RFR	$R^2 = 0.8478$
Almasaeid and Salman (2022)[a]	A/C ratio, W/C ratio, Density	369	• MLR • ANN	Porosity Permeability	ANN ANN	$R^2 = 0.886$ $R^2 = 0.934$
Shirgir *et al.* (2016)[a]	Fine aggregate content, W/C ratio, Aggregate size, Aggregate gradation, Porosity	20	• ANN	Permeability	ANN	$R^2 = 0.912$ RMSE = 0.798 mm/s
Present study	ggregate size, A/C ratio, Compaction energy, UPV	225	• ANN • KNN • SVR • RFR • BDT • LR	Porosity Permeability	ANN BDT	$R^2 = 0.8958$ $R^2 = 0.7574$

[a]Performance indicator for all datasets, for others it is for test datasets.
[b]ANN – Artificial Neural Network, BDT – Boosted Decision Tree Regression, GPR – Gaussian Process Regression, KNN – K-Nearest Neighbour, MLR – Multiple Linear Regression, NLR – Non-Linear Regression, LR – Linear regression, RFR – Random Forest Regression, SVR – Support Vector Regression, XGB – Extreme Gradient Boosting

observed permeability are more similar when the pentagram is closer to the observation point. The linear regression model is the most distant from the observation point among all the ML models, while the BDT model is the closest. Moreover, the BDT model has a higher correlation value than the others, as indicated by its proximity to the horizontal line. An assessment of the closeness demonstrated that the ranking of the ML model performance is described in the following order BDT > SVM > RFR > ANN > KNN > LR. Therefore, results indicate that the BDT model outperforms other ML models regarding prediction accuracy.

7.4. Error distribution

Figure 28 shows the error (difference between predicted and observed permeability) of each ML method for training and testing datasets. The graph indicates that some of the observed data points are outliers, especially for the LR model, which means they have a significant error. The BDT models has more consistent prediction accuracy than other models with more varied errors. The BDT models have the slightest error range for permeability, from – 1.30 to 1.36 cm/s for train datasets and – 2.36 to 2.41 cm/s for test datasets, which are the narrowest among all the ML models. Overall, BDT and RFR models have the slightest error range compared to others.

7.5. Sensitivity analysis

Since BDT had the best overall performance, we used SHAP analysis to explain the results of the BDT model. The main idea behind SHAP is to compute the Shapley values for each instance feature that needs to be explained. The Shapley values indicate the contribution of each feature to the prediction. Figure 29 shows the mean SHAP values for input parameters based on the BDT model for different input parameters. The UPV has the highest SHAP value for predicting permeability among the input parameters. This means that UPV is the most critical factor compared to other parameters for predicting permeability. The speed of sound waves propagating through a material, as measured by UPV, can be related to the material's permeability. Variations in UPV can occur due to changes in permeability, which is influenced by factors such as pore size and connectivity. Aggregate size also has a significant effect on the ability to predict the permeability of pervious concrete. Pervious concrete is characterised by substantial porosity and a network of interconnected voids. The choice of aggregate size directly affects the size and arrangement of voids within the concrete structure. Larger aggregate sizes result in larger voids, which in turn increase permeability. Increased particle size in permeable concrete results in larger voids between particles, allowing for increased water permeability. This increased permeability improves the overall permeability of the pervious concrete. Therefore, UPV and particle size significantly influence permeable concrete's permeability. The A/C ratio and compaction energy can affect the density and compaction of materials, but their direct correlation with permeability may not be as strong. However, even the lowest value observed for the compaction energy, which was 0.16, indicates that other input parameters also have significant importance in predicting permeability.

Figure 30 shows the SHAP summary for permeability prediction with different input values. The colour specifies the range of the feature value, which represents how much the feature affects the expected outcome. The red dots on the right indicate that increasing a particular input parameter increases permeability. The blue dots, on the other hand, suggest that increasing a specific parameter of input decreases permeability. For permeability prediction with all the input variables, UPV has the highest positive SHAP value of 3.5, which means that permeability could be 3.5 cm/s higher than the average value for the lower UPV values used in this study. On the other hand, UPV has a negative SHAP value of – 1.5, which means that permeability could be 1.5 cm/s lower than the average value for the higher UPV values used in this study. It is consistent with the SHAP mean values, which measure the individual contribution of each feature. The UPV strongly influences permeability deviation, as small changes in UPV can lead to significant differences in permeability. On the contrary, the A/C ratio weakly influences permeability deviation, as changes in the A/C ratio less affect the permeability prediction. The SHAP analysis demonstrates that using game theory to calculate SHAP values can help interpret the ML models and validate their prediction accuracy.

Figure 31 shows the correlation between characteristics and the porosity of pervious concrete. Figure 31(a) demonstrates the relationship between UPV and aggregate size feature interaction and its direct impact on permeability. There is no clear trend on the partial dependency between aggregate size and UPV with permeability. Lower UPV negatively impacts permeability more than higher UPV for small and medium aggregate sizes. However, for larger aggregate sizes, lower UPV positively impacts permeability more than higher UPV. When the A/C ratio increases, an increase in aggregate size positively influences the permeability, as shown in Figure 31(c).

8. Comparison with previous works

Table 10 summarises the published literature and presents a study on prediction models for the porosity and permeability of pervious concrete utilising various machine learning techniques. When multiple machine learning techniques were used for comparison, the research conducted by Almasaeid and Salman (2022) demonstrated that ANN models exhibit high accuracy for predicting the porosity and permeability of pervious concrete. Le *et al.* (2022) recommended XGBoost for porosity prediction, and Sun *et al.* (2019) recommended SVR for permeability prediction. Except for these studies, all other studies are conducted with limited datasets. Also, except Le *et al.* (2022), all other studies used limited machine-learning techniques. In contrast to prior research, the present study possesses many data sets and machine-learning methods. Furthermore, the ANN model, which was suggested, exhibited superior accuracy for porosity prediction with R^2 equal to 0.8958 for testing datasets. The BDT model performs better for permeability prediction with $R^2 = 0.7574$ for testing datasets.

9. Limitations of the models

The proposed analytical model equation consists of four input variables: compaction energy, A/C/ratio, aggregate size, and UPV. This study's ML models for permeability show a respectable agreement with the experimental values. However, it should be noted that the parameters considered are within a limited range. The aggregate size varied from 5 to 25 mm, the A/C ratio varied from 3 to 5, and the compaction varied from 0 to 60 blows. In future research, it is essential to consider a wide range of values for these parameters, as parameters outside this range may be used for onsite applications. Also, several other factors influence the permeability of pervious concrete, such as aggregate gradation, fine content, water-to-cement ratio, etc. The models derived in this study may reflect the effect of those missing variables that are implicitly included in the model. Moreover, adding more data could improve the machine learning model's performance. Therefore, it is required to maintain the extensive data set for mixed parameters, UPV, porosity and permeability of pervious concrete.

Conventional soft computing techniques such as ANN, random forest regression, boosted decision tree regression and support vector regression used in this study have their strengths and limitations. Although these techniques have shown excellent performance in predicting porosity, their performance in predicting permeability is not up to the mark. Due to their limited expressiveness and reliance on hand-crafted rules, these techniques can struggle to capture complex and non-linear relationships in the data. They also often require manual feature engineering, which can be time-consuming and may not capture all the relevant information in the data. Deep learning models, with their deep neural architectures, excel at learning complex hierarchical representations, automatically extracting features and capturing non-linear patterns in data, allowing for more flexible and powerful representations. Deep learning models can also automatically learn hierarchical features from raw data, eliminating the need for extensive manual feature engineering. This capability is particularly beneficial when dealing with high-dimensional or unstructured data.

10. Conclusion

This study proposes a framework for predicting the porosity and permeability of pervious concrete using machine learning techniques. Two hundred twenty-five experimental specimens (data points) were used to train and test the models. Six machine learning techniques were customised and trained as baseline predictors: linear regression, artificial neural networks, boosted tree regression, random forest regression, K-Nearest Neighbor and support vector regression. The main findings of the study are:

- ANN showed best-predicted porosity among the ML models evaluated in this study ($R^2 = 0.9502$, RMSE = 1.91% for training datasets and $R^2 = 0.8959$, RMSE = 2.87% for testing datasets). The scatter plots showing the relationship between the actual and predicted porosity values show that the models accurately fit both the training and test datasets, as indicated by the observed values. Nevertheless, the ANN model demonstrates superior performance with an accuracy of 92.6% on the training datasets and 89.3% on the test datasets, which fall below a 20%

margin of error. It, therefore, gives the most accurate predictions.

- The boosted decision tree regression model performed better than other models for permeability prediction. It has performance indicators of $R^2 = 0.9323$ and RMSE = 0.34 cm/s for training datasets and $R^2 = 0.7574$, RMSE = 0.70 cm/s for testing datasets, respectively. The scatter plots showing the relationship between the actual and predicted porosity values indicate that the models fit the training and test datasets less accurately. Nevertheless, the boosted decision tree regression model demonstrates better performance with an accuracy of 63.3% on the training datasets and 42.7% on the test datasets, both of which fall below a 20% margin of error.
- A sensitivity analysis aims to examine the impact of explanatory variables on the porosity and permeability. The findings indicated that the order of relative significance of explanatory variables employed in ANN for predicting porosity was as follows: UPV > compaction energy > aggregate size > A/C ratio. The order for predicting permeability using a boosted decision tree regression model is UPV > aggregate size > A/C ratio > compaction energy. According to the feature importance analysis using SHAP, the parameter that has the most significant impact on predicting the porosity of pervious concrete is UPV.

The present work systematically evaluates the permeability prediction of pervious concrete, which can contribute to the knowledge base and be applied in the real world. Material scientists and designers can use reliable prediction models to select the optimal input variables and the suitable mix parameters for producing pervious concrete with the desired properties.

Declarations

Disclosure statement

No potential conflict of interest was reported by the author(s).

Availability of data and material

Data can be made available on request by interested parties.

Authors' contributions

N. Sathiparan: Conceptualization, Data curation, Analysis, Writing – original draft, Writing – review & editing

S.H. Wijekoon: Conceptualization, Data curation, Analysis, Writing – original draft

P. Jeyananthan: Machine learning modelling, Analysis, Writing – original draft, Writing – review & editing

D.N. Subramaniam: Conceptualization, Data curation, Writing – review & editing

ORCID

Navaratnarajah Sathiparan http://orcid.org/0000-0001-8570-0580
Pratheeba Jeyananthan http://orcid.org/0000-0003-0476-3964
Daniel Niruban Subramaniam http://orcid.org/0000-0002-1023-1617

References

Aci-522r, 2010. *Report on pervious concrete*. Farmington Hills, MI: ACI Committee.

Adato, J. and Delatte, N.J., 2013. *Using ultrasonic pulse velocity (upv) to predict properties and performance of pervious concrete*. Ohio: Cleveland State University.

Adewumi, A.A., *et al.*, 2016. Estimation of physical, mechanical and hydrological properties of permeable concrete using computational intelligence approach. *Applied Soft Computing*, 42, 342–350. Available from: https://www.sciencedirect.com/science/article/pii/S1568494616300527.

Ahmad, A., *et al.*, 2022. Compressive strength prediction of fly ash-based geopolymer concrete via advanced machine learning techniques. *Case Studies in Construction Materials*, 16, e00840. Available from: https://www.sciencedirect.com/science/article/pii/S2214509521003557.

Ahmad, S.A., *et al.*, 2023. Mathematical modeling techniques to predict the compressive strength of pervious concrete modified with waste glass powders. *Asian Journal of Civil Engineering*, doi:10.1007/s42107-023-00811-1.

Ahmed, H.U., *et al.*, 2022. Mathematical modeling techniques to predict the compressive strength of high-strength concrete incorporated metakaolin with multiple mix proportions. *Cleaner Materials*, 5, 100132. Available from: https://www.sciencedirect.com/science/article/pii/S2772397622000922.

Ahmmad, R., *et al.*, 2016. Performance evaluation of palm oil clinker as coarse aggregate in high strength lightweight concrete. *Journal of Cleaner Production*, 112, 566–574. Available from: https://www.sciencedirect.com/science/article/pii/S0959652615011336.

Alemu, A.S., *et al.*, 2021. Practical considerations of porosity, strength, and acoustic absorption of structural pervious concrete. *Case Studies in Construction Materials*, 15, e00764. Available from: https://www.sciencedirect.com/science/article/pii/S2214509521002795.

Almasaeid, H.H. and Salman, D.G., 2022. Application of artificial neural network to predict the properties of permeable concrete. *Civil Engineering and Architecture*, 10 (6), 2305–2305.

Amari, S. and Wu, S., 1999. Improving support vector machine classifiers by modifying kernel functions. *Neural Networks*, 12 (6), 783–789. Available from: https://www.sciencedirect.com/science/article/pii/S0893608099000325.

Amini, K., Wang, X., and Delatte, N., 2018. Statistical modeling of hydraulic and mechanical properties of pervious concrete using nondestructive tests. *Journal of Materials in Civil Engineering*, 30 (6), 04018077. [Accessed 20 June 2023]. doi:10.1061/(ASCE)MT.1943-5533.0002268.

An, W. and Liang, M., 2013. Fuzzy support vector machine based on within-class scatter for classification problems with outliers or noises. *Neurocomputing*, 110, 101–110. Available from: https://www.sciencedirect.com/science/article/pii/S0925231213000106.

Anburuvel, A. and Niruban Subramaniam, D., 2022. Investigation of the effects of compaction on compressive strength and porosity characteristics of pervious concrete. *Transportation Research Record*, 2676 (9), 513–525. [Accessed 2 January 2023]. doi:10.1177/03611981221087236.

Anburuvel, A. and Subramaniam, D.N., 2022a. Influence of aggregate gradation and compaction on compressive strength and porosity characteristics of pervious concrete. *International Journal of Pavement Engineering*, 1–14. doi:10.1080/10298436.2022.2055022.

Anburuvel, A. and Subramaniam, D.N., 2022b. A novel multi-variable model for the estimation of compressive strength of pervious concrete. *International Journal of Pavement Research and Technology. Available from*, doi:10.1007/s42947-022-00266-8.

Anburuvel, A. and Subramaniam, D.N., 2023. Influence of aggregate gradation and compaction on compressive strength and porosity characteristics of pervious concrete. *International Journal of Pavement Engineering*, 24 (2), 2055022. doi:10.1080/10298436.2022.2055022.

Armaghani, D.J. and Asteris, P.G., 2021. A comparative study of ann and anfis models for the prediction of cement-based mortar materials compressive strength. *Neural Computing and Applications*, 33 (9), 4501–4532. doi:10.1007/s00521-020-05244-4.

Asteris, P.G., *et al.*, 2020. A novel heuristic algorithm for the modeling and risk assessment of the covid-19 pandemic phenomenon. *Computer Modeling in Engineering \& Sciences*, 125 (2), 815–828.

Astm-C597, 2010. *Standard test method for pulse velocity through concrete*. West Conshohocken, PA: ASTM International.

Astm-D7063/D7063m, 2018. *Standard test method for effective porosity and effective air voids of compacted asphalt mixture samples*. West Conshohocken, PA: ASTM International.

Badarna, M. and Shimshoni, I., 2019. Selective sampling for trees and forests. *Neurocomputing*, 358, 93–108. Available from: https://www.sciencedirect.com/science/article/pii/S0925231219306617.

Bs-1377, 1990. *Methods of test for soils for civil engineering purposes*. London: British Standards Institution (BSI).

Bs-En-1097, 2020. *Tests for mechanical and physical properties of aggregates*. London: British Standards Institution (BSI).

Bs-En-933, 2017. *Tests for geometrical properties of aggregates*. London: British Standards Institution (BSI).

Calabrese, L., *et al*., 2015. Identification of damage evolution during scc on 17-4 ph stainless steel by combining electrochemical noise and acoustic emission techniques. *Corrosion Science*, 98, 573–584. Available from: https://www.sciencedirect.com/science/article/pii/S0010938X15002644.

Chandrappa, A.K. and Biligiri, K.P., 2016. Pervious concrete as a sustainable pavement material – research findings and future prospects: a state-of-the-art review. *Construction and Building Materials*, 111, 262–274. Available from: https://www.sciencedirect.com/science/article/pii/S0950061816301131.

Delatte, N., Mrkajic, A., and Miller, D.I., 2009. Field and laboratory evaluation of pervious concrete pavements. *Transportation Research Record*, 2113 (1), 132–139. [Accessed 05 November 2023]. doi:10.3141/2113-16.

Elgeldawi, E., *et al*., 2021. Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis. *Informatics*, 8 (4), 79.

Elith, J., Leathwick, J.R., and Hastie, T., 2008. A working guide to boosted regression trees. *Journal of Animal Ecology*, 77 (4), 802–813.

Farahani, J.N., *et al*., 2017. Engineering properties of lightweight aggregate concrete containing binary and ternary blended cement. *Journal of Cleaner Production*, 149, 976–988. Available from: https://www.sciencedirect.com/science/article/pii/S0959652617302913.

Feng, D.-C., *et al*., 2020. Machine learning-based compressive strength prediction for concrete: an adaptive boosting approach. *Construction and Building Materials*, 230, 117000. Available from: https://www.sciencedirect.com/science/article/pii/S0950061819324420.

Filho, S.T.M., *et al*., 2020. Characterization of pervious concrete focusing on non-destructive testing. *Revista IBRACON de Estruturas e Materiais*, 13 (3), 483–500. [Accessed 04 February 2021].

Gao, W., *et al*., 2019. Development of a novel soft-computing framework for the simulation aims: a case study. *Engineering with Computers*, 35 (1), 315–322. doi:10.1007/s00366-018-0601-y.

Guillen, M.D., Aparicio, J., and Esteve, M., 2023. Gradient tree boosting and the estimation of production frontiers. *Expert Systems with Applications*, 214, 119134. Available from: https://www.sciencedirect.com/science/article/pii/S0957417422021522.

Hu, L., *et al*., 2016. Centrifuge-less dispersive liquid-liquid microextraction base on the solidification of switchable solvent for rapid on-site extraction of four pyrethroid insecticides in water samples. *Journal of Chromatography A*, 1472, 1–9. Available from: https://www.sciencedirect.com/science/article/pii/S0021967316313450.

Huang, J., *et al*., 2020. Predicting the permeability of pervious concrete based on the beetle antennae search algorithm and random forest model. *Advances in Civil Engineering*, 2020, 8863181. doi:10.1155/2020/8863181.

Huang, J., Zhang, J., and Gao, Y., 2022. Evaluating the clogging behavior of pervious concrete (pc) using the machine learning techniques. *Computer Modeling in Engineering \& Sciences*, 130 (2), 805–821.

Huda, M.N., *et al*., 2018. Performance of high strength lightweight concrete using palm wastes. *IIUM Engineering Journal*, 19 (2), 30–42. Available from: https://journals.iium.edu.my/ejournal/index.php/iiumej/article/view/919 [Accessed 23 February 2022].

Ibrahim, A., *et al*., 2014. Experimental study on portland cement pervious concrete mechanical and hydrological properties. *Construction and Building Materials*, 50, 524–529. Available from: https://www.sciencedirect.com/science/article/pii/S0950061813008623.

Le, B.-A., *et al*., 2022. Predicting the compressive strength and the effective porosity of pervious concrete using machine learning methods. *KSCE Journal of Civil Engineering*, 26 (11), 4664–4679. doi:10.1007/s12205-022-1918-z.

Le, B.-A., Vu, V.-H., Seo, S.-Y., Tran, B.-V., Nguyen-Sy, T., Le, M.-C., and Vu, T.-S., 2022. Predicting the compressive strength and the effective porosity of pervious concrete using machine learning methods. *KSCE Journal of Civil Engineering*, 26 (11), 4664–4679. doi:10.1007/s12205-022-1918-z.

Li, J., *et al*., 2017. Preparation and performance evaluation of an innovative pervious concrete pavement. *Construction and Building Materials*, 138, 479–485. Available from: https://www.sciencedirect.com/science/article/pii/S0950061817301666.

Li, L.G., *et al*., 2021. Roles of mortar volume in porosity, permeability and strength of pervious concrete. *Journal of Infrastructure Preservation and Resilience*, 2 (1), 19. doi:10.1186/s43065-021-00033-2.

Lin, P., *et al*., 2022. Machine learning-enabled estimation of crosswind load effect on tall buildings. *Journal of Wind Engineering and Industrial Aerodynamics*, 220, 104860. Available from: https://www.sciencedirect.com/science/article/pii/S0167610521003342.

Lu, X., *et al*., 2021. Equivalent method for obtaining concrete age on the basis of electrical resistivity. *Scientific Reports*, 11 (1), 21720. doi:10.1038/s41598-021-00447-8.

Lundberg, S.M. and Lee, S.-I., 2017. A unified approach to interpreting model predictionsed. *Advances in Neural Information Processing Systems*. doi:10.48550/arXiv.1705.07874.

Makariou, D., Barrieu, P., and Chen, Y., 2021. A random forest based approach for predicting spreads in the primary catastrophe bond market. *Insurance: Mathematics and Economics*, 101, 140–162. Available from: https://www.sciencedirect.com/science/article/pii/S0167668721001153.

Malami, S.I., *et al*., 2022. Implementation of soft-computing models for prediction of flexural strength of pervious concrete hybridized with rice husk ash and calcium carbide waste. *Modeling Earth Systems and Environment*, 8 (2), 1933–1947. doi:10.1007/s40808-021-01195-4.

Mandal, T., Tinjum, J.M., and Edil, T.B., 2016. Non-destructive testing of cementitiously stabilized materials using ultrasonic pulse velocity test. *Transportation Geotechnics*, 6, 97–107. Available from: https://www.sciencedirect.com/science/article/pii/S2214391215000240.

Marani, A. and Nehdi, M.L., 2020. Machine learning prediction of compressive strength for phase change materials integrated cementitious composites. *Construction and Building Materials*, 265, 120286. Available from: https://www.sciencedirect.com/science/article/pii/S0950061820322911.

Martín-Del-Rio, J.J., Canivell, J., and Falcón, R.M., 2020. The use of non-destructive testing to evaluate the compressive strength of a lime-stabilised rammed-earth wall: Rebound index and ultrasonic pulse velocity. *Construction and Building Materials*, 242, 118060. Available from: https://www.sciencedirect.com/science/article/pii/S0950061820300659.

Montes, F., Valavala, S. and Haselbach, L.M., 2005. A new test method for porosity measurements of portland cement pervious concrete. *Journal of ASTM International*, 2 (1), JAI12931. Available from: http://www.astm.org/cgi-bin/doiLink.cgi?JAI12931.

Ndagi, A., *et al*., 2019. Non-destructive assessment of concrete deterioration by ultrasonic pulse velocity: a review. *IOP Conference Series: Earth and Environmental Science*, 357 (1), 012015. doi:10.1088/1755-1315/357/1/012015.

Pradhan, S.K. and Behera, N., 2022. Performance assessment of pervious concrete road on strength and permeability by using silica fume. *Materials Today: Proceedings*, 60, 559–568. Available from: https://www.sciencedirect.com/science/article/pii/S2214785322005831.

Quan Tran, V., Quoc Dang, V., and Si Ho, L., 2022. Evaluating compressive strength of concrete made with recycled concrete aggregates using machine learning approach. *Construction and Building Materials*, 323, 126578. Available from: https://www.sciencedirect.com/science/article/pii/S0950061822002707.

Rao, S.K., Sravana, P., and Rao, T.C., 2016. Experimental studies in ultrasonic pulse velocity of roller compacted concrete pavement containing fly ash and m-sand. *International Journal of Pavement Research and Technology*, 9 (4), 289–301. Available from: https://www.sciencedirect.com/science/article/pii/S1996681416300049.

Ridengaoqier, E., *et al.*, 2021. Experimental study on the porosity evaluation of pervious concrete by using ultrasonic wave testing on surfaces. *Construction and Building Materials*, 300, 123959. Available from: https://www.sciencedirect.com/science/article/pii/S0950061821017190.

Saini, I., Singh, D., and Khosla, A., 2013. Qrs detection using k-nearest neighbor algorithm (knn) and evaluation on standard ecg databases. *Journal of Advanced Research*, 4 (4), 331–344. Available from: https://www.sciencedirect.com/science/article/pii/S209012321200046X.

Sathiparan, N., *et al.*, 2023a. Prediction of characteristics of cement stabilized earth blocks using non-destructive testing: Ultrasonic pulse velocity and electrical resistivity. *Materialia*, 29, 101794. Available from: https://www.sciencedirect.com/science/article/pii/S2589152923001217.

Sathiparan, N., 2023. Performance of sustainable cement mortar containing different types of masonry construction and demolition wastes. *Clean Technologies and Environmental Policy*, doi:10.1007/s10098-023-02681-2.

Sathiparan, N. and Jeyananthan, P., 2023. Prediction of masonry prism strength using machine learning technique: Effect of dimension and strength parameters. *Materials Today Communications*, 35, 106282. Available from: https://www.sciencedirect.com/science/article/pii/S235249282300973X.

Sathiparan, N., Jeyananthan, P., and Subramaniam, D.N., 2023b. Effect of aggregate size, aggregate to cement ratio and compaction energy on ultrasonic pulse velocity of pervious concrete: Prediction by an analytical model and machine learning techniques. *Asian Journal of Civil Engineering*, doi:10.1007/s42107-023-00790-3.

Sathiparan, N., Jeyananthan, P., and Subramaniam, D.N., 2023c. Prediction of compressive strength of fly ash blended pervious concrete: a machine learning approach. *International Journal of Pavement Engineering*, 24 (2), 2287146. doi:10.1080/10298436.2023.2287146.

Shafigh, P., *et al.*, 2016. Engineering properties of lightweight aggregate concrete containing limestone powder and high volume fly ash. *Journal of Cleaner Production*, 135, 148–157. Available from: https://www.sciencedirect.com/science/article/pii/S0959652616307594.

Shah, S.F.A., *et al.*, 2022. Compressive strength prediction of one-part alkali activated material enabled by interpretable machine learning. *Construction and Building Materials*, 360, 129534. Available from: https://www.sciencedirect.com/science/article/pii/S0950061822031907.

Shirgir, B., Mamdoohi, A.R., and Hassani, A., 2016. Prediction of pervious concrete permeability and compressive strength using artificial neural networks. *International Journal of Transportation Engineering*, 2 (4), 307–316. Available from: https://www.icevirtuallibrary.com/doi/abs/10.1680macr.15.00076.

Singh, N. and Singh, S.P., 2018. Evaluating the performance of self compacting concretes made with recycled coarse and fine aggregates using non destructive testing techniques. *Construction and Building Materials*, 181, 73–84. Available from: https://www.sciencedirect.com/science/article/pii/S0950061818314296.

Song, H., *et al.*, 2021. Predicting the compressive strength of concrete with fly ash admixture using machine learning algorithms. *Construction and Building Materials*, 308, 125021. Available from: https://www.sciencedirect.com/science/article/pii/S0950061821027677.

Subramaniam, D.N., Hareindirasarma, S., and Janarthanan, B., 2022. An alternative approach to optimize aggregate-to-cement ratio and compaction in pervious concrete. *Arabian Journal for Science and Engineering*, 47 (10), 13063–13071. doi:10.1007/s13369-022-06737-1.

Subramaniam, D.N., Jeyananthan, P., and Sathiparan, N., 2023. Soft computing techniques to predict the electrical resistivity of pervious concrete. *Asian Journal of Civil Engineering. Available from*, doi:10.1007/s42107-023-00806-y.

Subramaniam, D.N. and Sathiparan, N., 2022. Comparative study of fly ash and rice husk ash as cement replacement in pervious concrete: Mechanical characteristics and sustainability analysis. *International Journal of Pavement Engineering*, 1–18. doi:10.1080/10298436.2022.2075867.

Sudhir Kumar, B., Srikanth, K., and Eeshwar, T., 2024. Implementation of soft computing techniques in forecasting compressive strength and permeability of pervious concrete blended with ground granulated blast-furnace slag. *Journal of Soft Computing in Civil Engineering*, 8 (2), 19–45. Available from: https://www.jsoftcivil.com/article_172918.html.

Sun, J., *et al.*, 2019. Prediction of permeability and unconfined compressive strength of pervious concrete using evolved support vector regression. *Construction and Building Materials*, 207, 440–449. Available from: https://www.sciencedirect.com/science/article/pii/S095006181930412X.

Wang, X., *et al.*, 2016. Monitoring of setting time of self-consolidating concrete using ultrasonic wave propagation method and other tools. *Magazine of Concrete Research*, 68 (3), 151–162. Available from: https://www.icevirtuallibrary.com/doi/abs/10.1680macr.15.00076.

Wei, X., Xiao, L., and Li, Z., 2012. Prediction of standard compressive strength of cement by the electrical resistivity measurement. *Construction and Building Materials*, 31, 341–346. Available from: https://www.sciencedirect.com/science/article/pii/S0950061812000074.

Wijekoon, S.H., *et al.*, 2023. A mathematical model to predict the porosity and compressive strength of pervious concrete based on the aggregate size, aggregate-to-cement ratio and compaction effort. *Asian Journal of Civil Engineering*, doi:10.1007/s42107-023-00757-4.

Xia, Y., 2020. Chapter eleven - correlation and association analyses in microbiome study integrating multiomics in health and disease. *In*: J. Sun, ed. *Progress in molecular biology and translational science*. Massachusetts: Academic Press, 309–491.

Xu, G., *et al.*, 2018. Investigation on the properties of porous concrete as road base material. *Construction and Building Materials*, 158, 141–148. Available from: https://www.sciencedirect.com/science/article/pii/S095006181731975X.

Zaetang, Y., *et al.*, 2015. Use of coal ash as geopolymer binder and coarse aggregate in pervious concrete. *Construction and Building Materials*, 96, 289–295. Available from: https://www.sciencedirect.com/science/article/pii/S0950061815302701.

Zhang, J., *et al.*, 2022a. Machine learning prediction models for compressive strength of calcined sludge-cement composites. *Construction and Building Materials*, 346, 128442. Available from: https://www.sciencedirect.com/science/article/pii/S095006182202102X.

Zhang, L.V., Marani, A., and Nehdi, M.L., 2022b. Chemistry-informed machine learning prediction of compressive strength for alkali-activated materials. *Construction and Building Materials*, 316, 126103. Available from: https://www.sciencedirect.com/science/article/pii/S0950061821038356.

Ziamiavaghi, B. and Toufigh, V., 2023. Fracture toughness evaluation of ground granulated blast furnace slag concrete using experimental study and machine learning techniques. *Engineering Fracture Mechanics*, 291, 109577. Available from: https://www.sciencedirect.com/science/article/pii/S0013794423005350.