Earth Science Informatics (2024) 17:3287–3332
https://doi.org/10.1007/s12145-024-01336-0

RESEARCH

Assessment of hydraulic conductivity of compacted clayey soil using artificial neural network: an investigation on structural and database multicollinearity

Jitendra Khatti[1] · **Kamaldeep Singh Grover**[1]

Received: 8 December 2023 / Accepted: 15 May 2024 / Published online: 5 June 2024

Abstract
This work reveals the effect of hidden layers (HL) and neurons (N) on the performance of artificial neural network (ANN) models in predicting clayey soil's hydraulic conductivity (K). To achieve this, three databases, i.e., training, validation, and testing, have been created from a published database, i.e., 104. Each ANN model has been configured with one to five hidden layers interconnected with each 5, 10, and 15 neurons. Interestingly, this research also presents a comparative study among six backpropagation algorithms. Thus, ninety ANN models (fifteen for each backpropagation algorithm) have been developed, learned, and analyzed using seventeen performance metrics. The performance comparison illustrates that the Levenberg–Marquardt algorithm-based ANN model LM_K15 (configured with five HL interconnected with 15) has attained a correlation coefficient of 0.9959, root mean square error of 0.0487, the variance accounted for of 99.16, a20-index of 100, and performance index of 1.9348 in the testing phase, close to ideal values, and is introduced as an optimal performance model. This work concludes that the performance of the ANN model increases with hidden layers. Still, the effect of multicollinearity can't be neglected. The database used in this work has moderate to problematic multicollinearity, and the impact of such multicollinearity has been analyzed for gradient descent backpropagation algorithm-based ANN models. On the other side, ANOVA, Z, and Anderson–Darling tests reject the null hypothesis of normality. Model LM_K15 has gained the highest score in rank (= 291) and uncertainty (= 1) analyses. Wilcoxon test also presents the prediction capabilities of model LM_K15. The sensitivity analysis reveals that the maximum dry density influences the prediction of K, followed by sand and specific gravity. On the other side, it was observed that specific gravity has no direct relationship with K, obtained in terms of a correlation coefficient of 0.05468, comparatively lower than other input variables.

Keywords Hydraulic conductivity · Artificial neural network · Structural and Database multicollinearity · Performance and hidden layers relationship

Introduction

An essential characteristic of soil that describes its capacity to carry water through it is known as hydraulic conductivity. In hydrogeology, soil science, and civil engineering, hydraulic conductivity is a crucial quantity because it aids in understanding and predicting groundwater flow, drainage, and soil–water interactions. The type of soil, porosity, particle size distribution, compaction, and other factors all affect hydraulic conductivity. Sands and gravels, which contain larger particles, often have higher hydraulic conductivity, but clays, which contain smaller particles, have lower hydraulic conductivity and obstruct water flow. Hydraulic conductivity can be assessed in the lab or in the field, and it is frequently used in groundwater modeling, investigations of soil permeability, the design of drainage systems, and groundwater remediation strategies. The value of saturated hydraulic conductivity (K) for a specific soil is critical in assessing its suitability for various engineering and environmental applications. The experimental procedures for determining the hydraulic conductivity of fine-grained soil

Communicated by: Hassan Babaie

✉ Jitendra Khatti
jitendrakhatti197@gmail.com

Kamaldeep Singh Grover
ksgrover@rtu.ac.in

[1] Department of Civil Engineering, Rajasthan Technical University, Kota, Rajasthan, India

are arduous. Therefore, several investigators and scientists evolved and employed different empirical and advanced computational approaches.research on the assessment of hydraulic conductivity

Teng et al. (2023) implemented the Kozeny-Carman (KCn) equation to assess the K of coarse-grained soil during frozen conditions. Tan et al. (2023) established an artificial neural network (ANN), gradient boosting decision tree (GB_DT), multiple linear regression (MLR), and random forest (RF) utilizing 329 soil sample results to predict K. The researchers predicted K with 93% accuracy using plasticity index (PI), degree of saturation (D_S), specific gravity (SG), fine contents (FG), and void ratio (e) as input variables. Zeitfogel et al. (2023) used a Feedforward neural network (FNN) and extreme gradient boosting (XGBoost) using organic matter content, clay, silt, and sand content. The authors noted that XGBoost predicts K more efficiently than the FNN model with a test RMSE of 12.0%. Azarhoosh and Koohmishi (2023) employed RF, ANN, and adaptive neuro-fuzzy inference system (ANFIS) computational models to assess K of coarse-grained soil. The investigators reported that air void content is the most influential variable in assessing K. Peters et al. (2023a) assessed K_{UST} using a water retention curve. The researchers proposed a function for unsaturated K (K_{UST}) by replacing the soil-specific saturated K. The mean error between real and predicted K_{UST} was less than half an order of magnitude. Mufti and Das (2023) implemented a hybrid approach using pore-network and discrete element methods to estimate the K_{UST} of granular soils. Wang et al. (2023) proposed a model capable of assessing K with RMSE of 0.78 cm/day. Zhang and Wang (2023) assessed saturated hydraulic conductivity (K) soil (having bulk densities of 1.25, 1.3, 1.35, 1.4, 1.45, 1.5, 1.55, 1.6, 1.65, 1.7, 1.75, and 1.8 g/cm^3) using CT scanning technology for mining areas. The investigators found that the number of macro-porosity and macro-pores decreases due to increased bulk density. Using CT scan technology for soil samples prepared at different bulk densities, the model predicted K with a determination coefficient (R^2) of 0.84. Li et al. (2023) predicted K of clay and sand using a modified KCn equation based on Poiseuille's law. Peters et al. (2023b) employed Mualem (Mual), Chlds and Collis-George (CGG), Alexander and Skaggs (AS), and Burdine (Br) models to measure the K of soil. It was noted that the Mualen model measured K better than other models. Zhang et al. (2023) used the K results of 329 soil samples to employ the random forest model. The proposed model predicted K with high precision (i.e., 92%) using compaction parameters, hydraulic characteristics, and soil physical properties. Kim et al. (2023) used a regional database of 68 soil samples with geotechnical, geological, and hydrological parameters to assess K. The authors noted that the K is directly affected by the infiltration process of rainfalls into the soil. Also, the authors have mapped the best correlation between groundwater level and the moving precipitation average. Chandel et al. (2023) constructed a feedforward neural network (FFNN) to assess the K of porous media. The researchers predicted K with root mean square error (RMSE) of 0.016, mean bias error (MBE) of 0.006, and determination coefficient (R^2) of 0.94 using the FFNN model. The performance of the FFNN model was compared with MLR and KSOM models, and it found that the FFNN model is robust in predicting K. Khaja et al. (2023) conducted research to identify the relationship among K, field dry density, porosity, and gradational parameters of sandy soil. For this aim, the authors used the results of 60 soil samples. It was concluded that (i) particle size at 50% fine (D50) has a strong relationship with K, (ii) coefficient of uniformity (C_U) and curvature (C_C) poorly correlate with K. In addition, a significant relationship has been mapped between K, field dry density, and porosity. The regression model attained an RMSE of 0.67 and a mean square error (MSE) of 0.45. Piri et al. (2023) compared RF, Chi-Squared Automatic Interaction Detection (CHAID), and Geo-statistics models to assess the K using 130 soil sample results. It was noted that the RF model outperformed the CHAID and other models with the least residuals of 0.0019. Singh et al. (2023) employed support vector machine (SVM), RF, GPR, gene expression programming (GEP), and multivariate adaptive regression splines (MARS) to assess the K_{UST}. The investigators used the results of 240 soil samples to complete the published work. The investigators concluded that Pearson VII universal kernel (PUK) based SVM models are highly capable of assessing the K_{UST} of soil. The published research used moisture content, bulk density, silt, clay, and sand as input variables. Tseng et al. (2023) constructed GPR and Bayesian models to compute the K in a watershed. The authors found that the model's accuracy depends on the high- and low-fidelity data and location distribution. Báťková et al. (2023) predicted K for agricultural soil using the pedo-transfer function. The researchers developed ten models using 56 data points, including results of organic matter/ organic carbon content, dry bulk density, clay, silt, and sand particles. Singh and Sharma (2023) implemented Zamarin, NAVFACDM7, Sauerbrei, Kruger, Slitcher, Hazen, Terzaghi, and KCn equations for assessing K of soil using surface NMR porosity and particle size distribution. The researchers noted that the modified KCn equation assessed K with R^2 of 0.904 and RMSE of 6.36. Emberga et al. (2023) predicted K for aquifer based on the grain-size database using the MLR technique. The authors reported that Slitcher, ANN, and MLR models estimated K with RMSE of 5.14, 2.57, and 1.00, respectively. Veloso et al. (2022) estimated K using MARS, RF, SVR, and k-nearest neighbors (kNN). The researchers prepared different combinations of input variables, i.e., sand, silt, clay, bulk density, particle density, total porosity, microporosity,

microporosity, soil moisture at the permanent wilting point, and soil moisture at field capacity. The investigators noted that the RF and SVR models predicted K with higher R^2 and the least residuals using all input variables. Chandel et al. (2022) derived seven empirical equations using grain size parameters for predicting K. Using the Hazen equation, the authors noted a good agreement between real and predicted K. Khalili-Maleki et al. (2022) used Hybrid Wavelet-ANN (WANN), Least square support vector machine (LSSVM), and Larsen Fuzzy Logic (LFL) models to predict the K using grain size database. Model WANN was identified as a more accurate model than LSSVM and LFL models in predicting K. Albalasmeh et al. (2022) employed an optimized ANN model to compute the K for arid and semi-arid regions. The investigators implemented a generalized regression neural network (GRNN) model using depth, texture, organic matter, pH, bulk density, and electric conductivity as input variables of 165 soil samples. The investigators concluded that the GRNN model gives a reliable prediction of K with a limited database. Ruan and Fu (2022) assessed the K of compacted bentonite in confined conditions using a modified KCn equation. Hedayati-Azar and Sadeghi (2022) developed a semi-empirical model to assess the K of clayey soil. The authors reported that the estimation of K becomes erroneous if solute concentrations of permeating fluid are ignored. Shan et al. (2022) employed Weibull distribution models to estimate the relative K. Faloye et al. (2022) constructed MLR, ANN, and ANFIS models using biochar levels and soil moisture content. The models ANFIS, ANN, and MLR attained R^2 of 0.95, 0.98, and 0.92 in the validation phase. Therefore, the authors concluded ANN models are the most potent tool for predicting K_{UST} of biochar-amended soil. Singh et al. (2022) developed a genetic algorithm-optimized ANN and SVM models to estimate soil K. Furthermore, the pedo-transfer function (PTF) was implemented with developed models. The performance comparison demonstrated that the SVM_GA PTF model is more capable of predicting K than the ANN models. Using empirical relationships, Chandel and Shankar (2022) predicted K for borehole soil samples. The authors found that the KCn equation has better agreements between predicted and real K values than other equations, i.e., Alyamani & Sen, Hazen, and Beyer. ur Rehman et al. (2022) compared the multi-expression programming, GEP, and ANN in assessing K using a large database. The researchers concluded that GEP predicted K with high accuracy. In addition, it was noted that particle size at 10% finer (D_{10}) is the most influencing input variable in assessing K. Granata et al. (2022) used ANN, RF, and SVM approaches to compute the K of soil. Based on the performance comparison, the authors found that SVM and RF models are more accurate than ANN. Pham and Won (2022) reported that the extreme gradient boosting (XGB) approach based on PTF is highly capable of predicting soil K. It was also noted that clay content is the most significant variable in assessing the K. Hosseini et al. (2022) used soil texture to predict the K of soil by applying genetic and neural network approaches. The authors recorded a residual of 1.22 and regression coefficient of 0.997 for the neural network model in predicting K of soil, comparatively better than geo-statistics and genetic models. Thakur et al. (2022) assessed the hydraulic conductivity of porous media using ANFIS, triangular, GPR, and SVM models. In this published work, model ANFIS outperformed the triangular, GPR, and SVM models with an RMSE of 0.0010. Tan et al. (2022) predicted K of geosynthetic clay liners with a validation performance of 85%. More et al. (2022) applied extreme learning machine (ELM), SVM, and ANFIS approach to estimate saturated hydraulic conductivity for tropical semi-arid zones. The researchers reported that model ELM achieved Nash–Sutcliffe efficiency (NSE) of 0.90, better than the other two approaches. Tao et al. (2022) mapped the relationship between particle size and the K_{UST} of soil. Gupta et al. (2021) used RF to assess soil K. The RF model predicted the K of soil with an accuracy of 79% and RMSE of 0.72. Williams and Ojuri (2021) compared ANN and MLR models in predicting the hydraulic conductivity of soil. The authors reported that model ANN has gained an accuracy of 95.5%, higher than the MLR model. Mujtaba et al. (2021) mapped a relationship between hydraulic conductivity and gradational parameters of sandy soil. The researchers noted that D_{10} has a healthy relationship with K. Peters et al. (2021) estimated the hydraulic conductivity of medium to dry soil using a water retention curve. Yan et al. (2021) predicted the effect of biochar on the saturated K of natural and artificial media. The researchers noted that the hydraulic conductivity decreases because of an increase in inter-porosity due to bio-char and a decrease in mean pore radii. Rout and Singh (2021) introduced empirical models using hydraulic conductivity and basic soil properties. The proposed empirical model predicted K with $\pm$ 20% intervals.

Chen and Zhang (2020) estimated the K of frozen soil. A discontinuous noncircular capillary bundle model was introduced for this aim using modified Hagen-Poiseuille, Kelvin, and Campbell equations. Kashani et al. (2020) implemented MARS, M5 tree, SVM, ELM, and ANN approaches to assess the hydraulic conductivity of soil using electrical conductivity, pH, bulk density, organic matter, clay, and silt parameters as input variables. Based on the performance metrics, model ANN achieved the highest performance compared to other models, i.e., NSE = 0.939 (in training) and = 0.917 (in the testing). Arshad et al. (2020) derived empirical models and reported that void ratio and grain size characteristics are significant parameters in predicting the hydraulic conductivity of sandy soils. Trejo-Alonso et al. (2020) introduced a pedo-transfer function using 900 data points to assess the K of the soil. The proposed models assessed K with over 99%

accuracy. Babaoglu and Simms (2020) improved K estimation for soft clayey soil. The authors reported that (i) the K – a high-void ratio can improve the void ratio, and (ii) the compressibility curve can be a predictor. Sihag et al. (2020) employed ANN, GPR, GEP, and GRNN approaches to predict the infiltration process using 155 data points. Ming et al. (2020) assessed the K of frozen soil from the soil freezing characteristics curve. Sihag et al. (2019a) employed ANFIS, firefly (FFA), and particle swarm (PSO) algorithm-optimized ANFIS models to assess the hydraulic conductivity. These models were trained and tested by 170 and 70 data points. The ANFIS-PSO model outperformed the ANFIS-FFA and traditional ANFIS models with a correlation coefficient of 0.9816 in the testing phase. In addition, Sihag et al. (2019b) compared RF, M5P, and regression models in estimating the K_{UST} field. The RF model attained the highest performance, i.e., 0.819 in the testing phase, then other models. Sihag et al. (2019c) mapped a comparison between regression analysis, ANN, and ANFIS and found that the regression model MLR (RMSE = 4.5578) is better than other models. Naganna and Deka (2019) compared SVM, ANN, and ANFIS models to introduce the best prediction approach. The comparison of performance metrics shows that the SVM model is the best approach for predicting streambed hydraulic conductivity. Al-Dosary et al., (2019) implemented GPR, linear regression (LR), and multilayer perceptron (MLP) approaches to assess the K_{UST} of sandy loam soil. In the published work, the GPR model outperformed the LR and MLP models. Sihag (2018) estimated the K_{UST} of soil by implementing fuzzy logic-FL (based on triangular and Gaussian) and ANN models. The researcher concluded that the fuzzy logic model based on Gaussian attained R of 0.9270 and RMSE of 7.4393, better than ANN and fuzzy logic (based on triangular) models. More and Deka (2018) employed fuzzy neural networks (FNN), ANN, FL, and MLR using 175 data points to measure the K for murum soils. The authors concluded that the FNN model attained an accuracy of over 85%, higher than the accuracy of ANN, FL, and MLR models. Nematolahi et al. (2018) employed GA and PSO-optimized fuzzy inference system (FIS) models to assess the K. The PSO-optimized FIS model attained an accuracy of over 70%, higher than conventional FIS and GA-optimized FIS models.

Also, Mady and Shein (2018), Qaderi et al. (2018), Fatoba et al. (2018), and Shi and Yin (2018) reported that the SVM, nonlinear regression, GMDH, harmony search-optimized GMDH, and ANN models can predict the K of soil. Table 1 summarizes the published research on the assessment of hydraulic conductivity of soil.

The published research reveals that most researchers employed MLR, GPR, GEP, MEP, SVM, DT, MARS, ANFIS, LSSVM, GMDH, ANN, and hybrid (WANN, SVM_GA, ANFIS_FFA, ANFIS_PSO, and HS_GMDH) approaches to predict the K of soil. These researchers also concluded that the ANN approach gives the most promising results of soil hydraulic conductivity. Still, the effect of structural multicollinearity on the performance of ANN models in predicting hydraulic conductivity has not been studied and analyzed. In addition, the backpropagation algorithms of neural networks have not been compared for designing the optimal performance ANN model. Also, the effect of multicollinearity levels on the ANN model has not been studied and analyzed. Based on the gap identified in the published work, the present research has the following novelty:

- This research illustrates the effect of structural multicollinearity, considering the one to five hidden layers interconnected with each 5, 10, and 15 neurons, on artificial neural network models in predicting the hydraulic conductivity of clayey soil.
- This research compares Gradient Descent with Adaptive Learning (GDA), Gradient Descent (GD), Gradient Descent with Momentum (GDM), Scaled Conjugate Gradient (SCG), Broyden, Fletcher, Goldfarb, and Shanno (BFGs), and Levenberg–Marquardt (LM) backpropagation algorithms to design an optimal performance ANN model.
- The effect of multicollinearity levels is studied and analyzed for each artificial neural network in predicting the hydraulic conductivity of clayey soil.
- This research introduces an optimal performance ANN model with the best hyperparameters for predicting the hydraulic conductivity of clayey soil.

The hydraulic conductivity of clayey soil is determined by performing the falling head test. The falling head hydraulic conductivity test is time-consuming. Therefore, several investigators applied traditional and advanced methods to assess the hydraulic conductivity of soil. These advanced methods are based on machine learning. However, an artificial neural network is an ML technique that can predict accurately. Still, selecting the number of hidden layers and neurons is much more important to achieve certain accuracy. The present research helps engineers choose the number of hidden layers and neurons for artificial neural networks to assess the hydraulic conductivity of compacted clayey soil. This research will also reduce the laboratory efforts of the geotechnical engineers in assessing hydraulic conductivity. This research also introduces the best backpropagation algorithm for developing neural network models.

Research methodology

This research introduces an optimal-performance artificial neural network model for predicting the hydraulic conductivity of clayey soil. In addition, this research compares the

Table 1 Summary of ML models used in the literature

S.No	References	Saturated	Database	Input Variables	Approach	R Test
1	Zeitfogel et al. (2023)	Yes	15,641	S, M, C, OC	XGBoost	0.80
2	Zeitfogel et al. (2023)	Yes	14,674	S, M, C, OC	XGBoost	0.82
3	Zeitfogel et al. (2023)	Yes	10,493	S, M, C, OC	XGBoost	0.74
4	Zeitfogel et al. (2023)	Yes	15,641	S, M, C, OC	XGBoost	0.65
5	Zeitfogel et al. (2023)	Yes	14,674	S, M, C, OC	XGBoost	0.77
6	Zeitfogel et al. (2023)	Yes	10,493	S, M, C, OC	XGBoost	0.54
7	Chandel et al. (2023)	No	165	D_{10}, D_{50}, P, C_U	FFNN	0.9711
8	Thakur et al. (2022)	-	56	D_{10}, D_{50}, P, C_U	GA_ANFIS	0.9661
9	Hoseini et al. (2022)	Yes	134	S, M, C	ANN	0.9988
10	Granata et al. (2022)	Yes	5023	S, M, C, OCR, dg, Sg, Db, WCs	MLP_RF_SVR	0.9105
11	Granata et al. (2022)	Yes		OCR, Sg, dg, Db, WCs	RF_SVR	0.8955
12	Granata et al. (2022)	Yes		WCs, Db, dg, Sg	RF_SVR	0.8712
13	Granata et al. (2022)	Yes		OCR, Db, dg, Sg	RF_SVR	0.7987
14	Granata et al. (2022)	Yes		Dg, Sg, Db	RF_SVR	0.7714
15	Williams & Ojuri (2021)	-	144	PI, S, FC, C, MDD, OMC	ANN	0.9770
16	Gupta et al. (2021)	Yes	6814	Db, S, M, C	RF	0.7900
17	Sihag et al. (2020)	-	155	S, C, M, Db, MC	ANN	0.9133
18	Sihag et al. (2019a)	No	240	S, C, M	ANFIS	0.9633
19	Sihag et al. (2019a)	No	240	S, C, M	ANFIS_FFA	0.9794
20	Sihag et al. (2019a)	No	240	S, C, M	ANFIS_PSO	0.9816
21	Sihag et al. (2019b)	No	240	S, C, M, Db, MC	RF	0.9050
22	Sihag et al. (2019c)	No	46	S/FA, S/RHA, SH, Db	MLR	0.8450
23	Naganna & Deka (2019)	Yes	-	L, D, SP (data 2016)	SVM	0.9716
24	Naganna & Deka (2019)	Yes	-	L, D, SP (data 2017)	ANFIS	0.9742
25	Al-Dosary et al. (2019)	No	48	EC, SAR, MC, Db, SR	GPR_PUK	0.9646
26	Sihag (2018)	No	46	S, RHA, FA, SH, Db, MC	G_FL	0.9270
27	Qaderi et al. (2018)	Yes	151	S, M, C, OCR, CCC, pH, EC, DS, P, PD, Db	HS_GMDH	0.9710
28	Nematolahi et al. (2018)	Yes	113	S, M, C, OCR, CCC, EC, Db	PSO_FIS	0.8485
29	More & Deka (2018)	Yes	175	Db, P, SG, S, C, M, OC	FNN	0.9234
30	Mady & Shein (2018)	Yes	33	S, M, C, Db, OC	SVM	0.9644

Where: S is sand, M is silt, C is clay, OC is organic matter content, XGBoost is extreme gradient boosting, FFNN is feedforward neural network, D_{10} is particle size 10% finer, D_{50} is particle size 50% fine, P is porosity, C_U is coefficient of uniformity, GA_ANFIS is Gaussian kernel based adaptive neuro-fuzzy inference system, ANN is artificial neural network, MLP_RF_SVR is hybrid multilayer perceptron random forest support vector regressor model, dg is geometric mean diameter, Sg is standard deviation of soil particle diameter, OCR is soil organic carbon content, Db is soil bulk density, WCs is saturated soil water content, RF_SVR is hybrid random forest support vector regressor model, PI is plasticity index, FC is fine, MDD is maximum dry density, OMC is optimum moisture content, MC is moisture content, FA is fly ash content, RHA is rice husk ash content, SH is suction head, S/FA is sand to fly ash ratio, S/RHA is sand to rice husk ash ratio, MLR is multiple linear regression, L is latitude, D is departure, SP is sample locations, EC is electric conductivity, SAR is sodium absorption ratio, SR is suction rate, GPR_PUK is Pearson VII function-based universal kernel based gaussian process regression model, G_FL is gaussian kernel based fuzzy logic, HS_GMDH is harmony search optimized group method of data handling model, CCC is calcium carbonate content, DS is degree of saturation, PD is particle density, PSO_FIS is particle swarm optimized fuzzy inference system, SG is specific gravity, FNN is fuzzy neural network model.

predictive capabilities of Gradient Descent with Adaptive Learning (GDA), Gradient Descent (GD), Gradient Descent with Momentum (GDM), Scaled Conjugate Gradient (SCG), Broyden, Fletcher, Goldfarb, and Shanno (BFGs), and Levenberg–Marquardt (LM) backpropagation algorithms to find the best backpropagation algorithm. For this aim, a database with results of soil texture, consistency limits, compaction parameters, and hydraulic conductivity of 104 soil specimens has been compiled from the published articles by Benson et al. (1994) and Benson and Trast (1995). The multicollinearity analysis has been performed to determine the collinearity levels for input variables. In addition, ANOVA and Z tests have been performed to determine the hypothesis for the present research. A cosine amplitude sensitivity analysis has been performed to determine the significant input variables in predicting the hydraulic conductivity of

soil. The training, validation, and testing databases have been created by arbitrarily selecting 80, 12, and 12 data points (soil samples). One to five hidden layers interconnected with 5, 10, and 15 neurons have been selected for developing ANN models. Thus, fifteen ANN models have been developed for each backpropagation algorithm. Fourteen performance metrics, RSR, LMI, MBE, WI, NMBE, BF, PI, NS, WMAPE, VAF, MAPE, R, MAE, and RMSE, have measured the performance and accuracy of learned ANN models. In addition, three novel performance metrics, a20-index, index of scatter, and index of agreement, have been implemented for measuring performance. One best architectural model is identified from each backpropagation algorithm by comparing the performance metrics. Thus, the six best architectural ANN models have been obtained and further analyzed by REC curve, rank, uncertainty, and Wilcoxon analysis. The research hypothesis (H_R) for the normality of predicted hydraulic conductivity of clayey soil has been checked by performing the Anderson–Darling (AD) test. Finally, one optimal performance ANN model has been identified for predicting the hydraulic conductivity of clayey soil. The accuracy of the optimal performance ANN model has been validated by published models. The robustness of the optimal performance ANN model has been determined by cross-validation (cost computation) and external validation (generalizability). The logic behind this methodology is to select the hyperparameters to design the optimal performance ANN model to predict the K of soil without the hit and trial method. Figure 1 depicts the flow chart for the execution of the work.

Data collection and analysis

A raw database from the published research by Benson et al. (1994), and Benson and Trast (1995) has been compiled to execute this research (refer Appendix, Table D). The database consists of soil texture (S, M, C), consistency limits (LL, PI), compaction parameters (OMC, MDD), and hydraulic conductivity (K) results of clayey soil. Most researchers implemented ML models using plastic limits in their published work to assess the geotechnical properties of fine-grained soil. Still, the plasticity index has not used to predict the hydraulic conductivity of soil. It is known that high PI shows low hydraulic conductivity of soil. Therefore, this research uses the plasticity index (PI) as an input variable for ML models. From removing outliers and missing data points, one hundred and four data points have been collected and used in this research. Three databases, training, validation, and testing, have been constructed by arbitrarily selecting 80, 12, and 12 data points, respectively. The descriptive statistics of the 104, 80, 12, and 12 databases are summarized in Table 2, along with the frequency plot of data points, as shown in Fig. 2.

Figure 2 depicts the frequency distribution of variables using the Lorentz curve. This curve helps to understand the distribution of the data variables. Gamma represents the anticipated change in Delta, with a maximum value of 1. The Gini coefficient (γ) varies from 0 to 1, presenting no inequality to complete inequality. Moreover, the complete database, i.e., 104, has been classified as per IS 1498: 1970, as illustrated in Fig. 3.

Figure 3 shows that the database consists of results of inorganic silts with none to low plasticity (ML), inorganic clays of low plasticity (CL), organic silts of low plasticity (OL), inorganic silts of medium plasticity (MI), inorganic clays of medium plasticity (CI), organic silts of medium plasticity (OI), inorganic silts of high compressibility (MH), inorganic clays of high plasticity (CH), and organic clays of medium to high plasticity (OH). Because of the number of different soils available in the database, Pearson's product-moment correlation coefficient has been calculated for each variable and presented in Fig. 4.

Figure 4 illustrates the relationship between the variables in terms of correlation coefficient. A correlation of ± 1.0 to ± 0.81, ± 0.80 to ± 0.61, ± 0.60 to ± 0.41, ± 0.40 to ± 0.21, and ± 0.20 to ± 0.00 represents the very strong, strong, moderate, weak, and no relationship between the variables (Hair et al. 2017). Figure 4 shows that (a) S content very strongly (=-0.9323) correlates with F content, (b) LL (=0.7537), PI (=0.7089), OMC (=0.6612), and MDD (=-0.6663) strongly correlates with F content, (c) S content also strongly correlates with LL (=-0.7668), PI (=-0.7207), OMC (=-0.7315), and MDD (=0.7167), (d) LL (=-0.2851) and PI (=-0.2516) weakly correlates with specific gravity, (e) OMC (=-0.1474) and MDD (=0.1675) have no relationship with SG, (f) LL (=0.9190) very strongly correlates with PI, (g) OMC (=0.7532) and MDD (=-0.6962) strongly correlates with PI, (h) OMC, and MDD very strongly (=-0.8703) correlates with each other, (i) F content (=-0.4122), LL (=-0.4972), PI (=-0.5620), OMC (=-0.4354), and MDD (=0.5295) moderately correlate with hydraulic conductivity of clayey soil. The pairwise scatterplot, correlation coefficient matrix, variance inflation factor (VIF), and eigenvalue methods are used to determine the multicollinearity levels of the database (Shrestha 2020). The correlation coefficient values for independent and dependent variables show multicollinearity. Therefore, another method, variance inflation factor (VIF), has been used to determine the multicollinearity levels for independent variables.

Multicollinearity analysis

Multicollinearity or collinearity occurs between the variables during regression analysis. However, an extensive

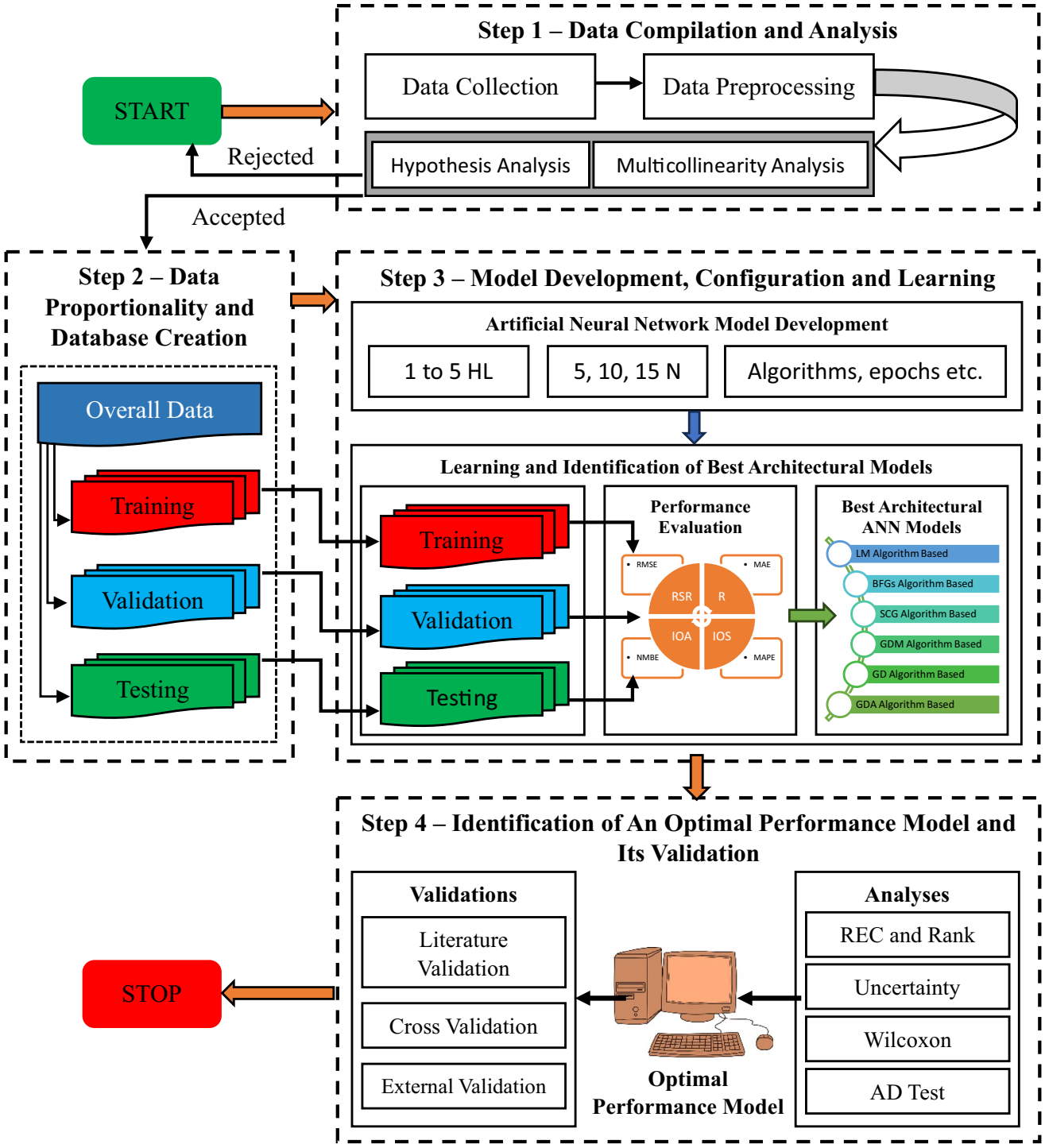

Fig. 1 Flow chart of the work

database is used for artificial intelligence techniques, increasing the chances of multicollinearity (Chan et al. 2022). The reasons for occurring multicollinearity are as follows: (a) variables are not significantly correlated, (b) multiple regression analysis is performed, and (c) variables are highly correlated. For determining the multicollinearity levels, the variance inflation factor ($VIF = 1/(1 - R^2)$) method has been used. Gareth et al. (2013) and Vittinghoff et al. (2006) introduced problematic multicollinearity levels if a VIF value is more than 10. Menard (2002) suggested a considerable multicollinearity level based on VIF value. Khatti and Grover

Table 2 Descriptive statistics of databases

Parameters	F (%)	S (%)	SG	LL (%)	PI (%)	OMC (%)	MDD (g/cc)	K (10^{-6} cm/s)
Complete Database								
Mean	77.37	20.95	2.68	38.73	16.44	17.61	1.69	1.40
Median	79.79	17.47	2.69	37.62	16.60	17.37	1.67	1.02
Standard Deviation	13.41	12.74	0.06	7.76	4.10	3.47	0.10	1.38
Sample Variance	179.84	162.31	0.00	60.29	16.77	12.04	0.01	1.91
Kurtosis	-1.09	-0.90	0.18	-0.84	-0.66	-0.65	-0.02	9.38
Skewness	-0.27	0.42	-0.44	0.20	-0.04	0.09	0.55	2.64
Minimum	51.00	0.00	2.53	23.85	7.91	10.60	1.50	0.05
Maximum	100.00	49.00	2.80	54.30	24.70	25.00	1.93	8.04
Confidence Level (95.0%)	2.61	2.48	0.01	1.51	0.80	0.67	0.02	0.27
Training Database								
Mean	77.13	21.57	2.68	37.98	16.31	17.13	1.71	1.35
Median	79.42	17.92	2.69	36.70	16.46	16.55	1.69	0.94
Standard Deviation	13.36	13.22	0.06	7.84	4.28	3.43	0.10	1.34
Sample Variance	178.59	174.78	0.00	61.52	18.28	11.76	0.01	1.81
Kurtosis	-1.17	-1.04	0.29	-0.89	-0.70	-0.74	-0.10	7.60
Skewness	-0.29	0.45	-0.74	0.23	-0.03	0.12	0.54	2.33
Minimum	51.00	1.28	2.53	23.85	7.91	10.60	1.50	0.05
Maximum	98.72	49.00	2.80	54.30	24.70	25.00	1.93	8.04
Confidence Level (95.0%)	2.97	2.94	0.01	1.75	0.95	0.76	0.02	0.30
Validation Database								
Mean	78.37	18.14	2.70	39.44	16.67	18.49	1.67	1.57
Median	80.87	16.20	2.71	38.10	16.70	19.05	1.66	0.98
Standard Deviation	11.29	8.20	0.06	6.19	3.68	2.32	0.08	2.15
Sample Variance	127.54	67.18	0.00	38.33	13.56	5.38	0.01	4.60
Kurtosis	0.35	-1.21	0.09	-0.54	-0.78	1.23	2.07	9.06
Skewness	-0.73	0.40	0.09	0.16	0.02	-1.10	1.17	2.89
Minimum	54.19	7.50	2.60	28.85	11.16	13.10	1.57	0.21
Maximum	92.50	30.94	2.80	50.29	22.57	21.00	1.86	8.04
Confidence Level (95.0%)	7.18	5.21	0.04	3.93	2.34	1.47	0.05	1.36
Testing Database								
Mean	78.00	19.58	2.67	42.99	17.06	19.94	1.65	1.59
Median	79.50	20.50	2.65	42.87	16.89	19.54	1.65	1.41
Standard Deviation	16.49	13.56	0.06	7.74	3.40	3.79	0.09	0.47
Sample Variance	271.82	183.90	0.00	59.94	11.55	14.40	0.01	0.22
Kurtosis	-1.33	-1.28	0.59	-1.51	-1.32	-1.92	-1.91	4.46
Skewness	-0.09	-0.19	1.32	0.25	0.32	0.13	-0.05	2.01
Minimum	52.00	0.00	2.62	32.98	12.46	15.21	1.53	1.19
Maximum	100.00	40.00	2.80	54.18	22.08	24.72	1.76	2.85
Confidence Level (95.0%)	10.48	8.62	0.04	4.92	2.16	2.41	0.06	0.30

(2023a) introduced five multicollinearity levels, i.e., problematic multicollinearity ($10 < \text{VIF}$), moderate multicollinearity ($5 < \text{VIF} \leq 10$), considerable multicollinearity ($2.5 < \text{VIF} \leq 5$), weak multicollinearity ($0 < \text{VIF} \leq 2.5$), and no multicollinearity ($0 = \text{VIF}$) based on VIF values using the published statement. Table 3 presents the multicollinearity levels for F, S, SG, LL, PI, OMC, and MDD variables in predicting the hydraulic conductivity of clayey soil.

Table 3 reveals that F (%), S (%), PI (%), and OMC (%) variables have moderate multicollinearity. Conversely, specific gravity and MDD (g/cc) have weak and considerable multicollinearity levels, respectively. The liquid limit of clayey soil shows the problematic

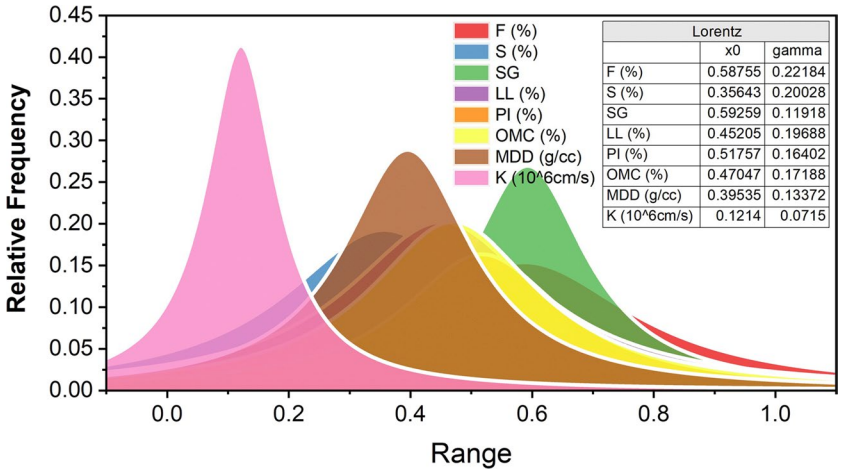

Fig. 2 Distribution of variables

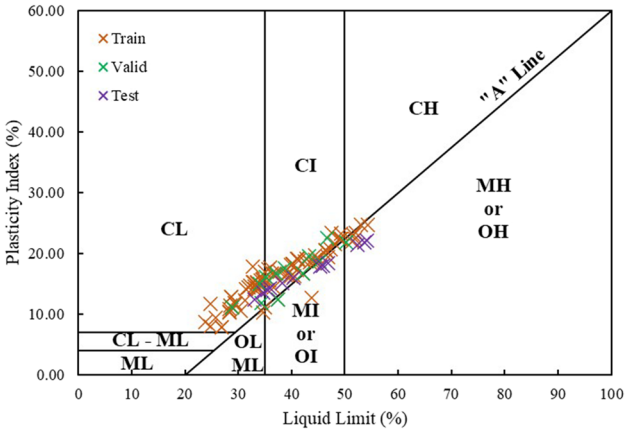

Fig. 3 Classification of database

multicollinearity in predicting the hydraulic conductivity of the soil.

Hypothesis analysis

The hypothesis analysis is performed for decision-making, inference, quality control, decision evaluation, risk assessment, and statistical inference. So, hypothesis testing is necessary to make informed judgments, draw inferences from data, and ensure that outcomes do not result from chance (Khatti and Grover 2023d, 2023e, 2023f). It does all of these things systematically and thoroughly. The following statements have been mapped for the present research for selecting the research hypothesis:

- The soil textures, i.e., F and S contents, are the significant variables in assessing the hydraulic conductivity of clayey soil.
- The liquid limit of soil increases due to an increase in fine content and a decrease in sand content.

For this purpose, ANOVA and Z tests have been performed in this research. The statistical test, Analysis of Variance (ANOVA), examines the variations in group means in a sample (Khatti et al. 2023). When comparing more than two groups, it is beneficial. ANOVA evaluates whether there is a statistically significant difference between these groups' means. The results of the ANOVA test are summarized in Table 4.

Table 4 demonstrates that each input variable, i.e., F ($3302.25 > 3.89$), S ($241.84 > 3.89$), SG ($88.20 > 3.89$), LL ($2329.37 > 3.89$), PI ($1258.38 > 3.89$), OMC ($1957.13 > 3.89$), and MDD ($4.55 > 3.89$), follows the research hypothesis (H_R) clause (F > F crit). Hence, the ANOVA test **ACCEPTS** the H_R for the present work. Moreover, another statistical hypothesis test, the Z test, has been performed to determine whether the sample mean is significantly different from a known population mean when the population standard deviation is known (Hosseini et al. 2023). The Z test results are summarized in Table 5.

Table 5 presents that each input variable follows the research hypothesis clause, i.e., Z > Z critical two-tail > Z critical one-tail and P one-tail < 0.05 > P two-tail. Hence, the Z test confirms the **REJECTION** of the null hypothesis for the present research.

Cosine amplitude sensitivity analysis

This analysis reveals the most significant input variables in predicting the hydraulic conductivity of clayey soil. The nonlinear cosine amplitude method (CAM) has been used for this aim. The sections mentioned earlier show that this work uses F, S, LL, PI, OMC, and MDD as input variables to assess the hydraulic conductivity of clayey soil. The sensitivity of input variables is determined by applying the following equation (Hasanzadehshooiili et al. 2012):

$$x = \{x_1, x_2, x_3 \ldots x_m, \ldots, x_n\} \quad (1)$$

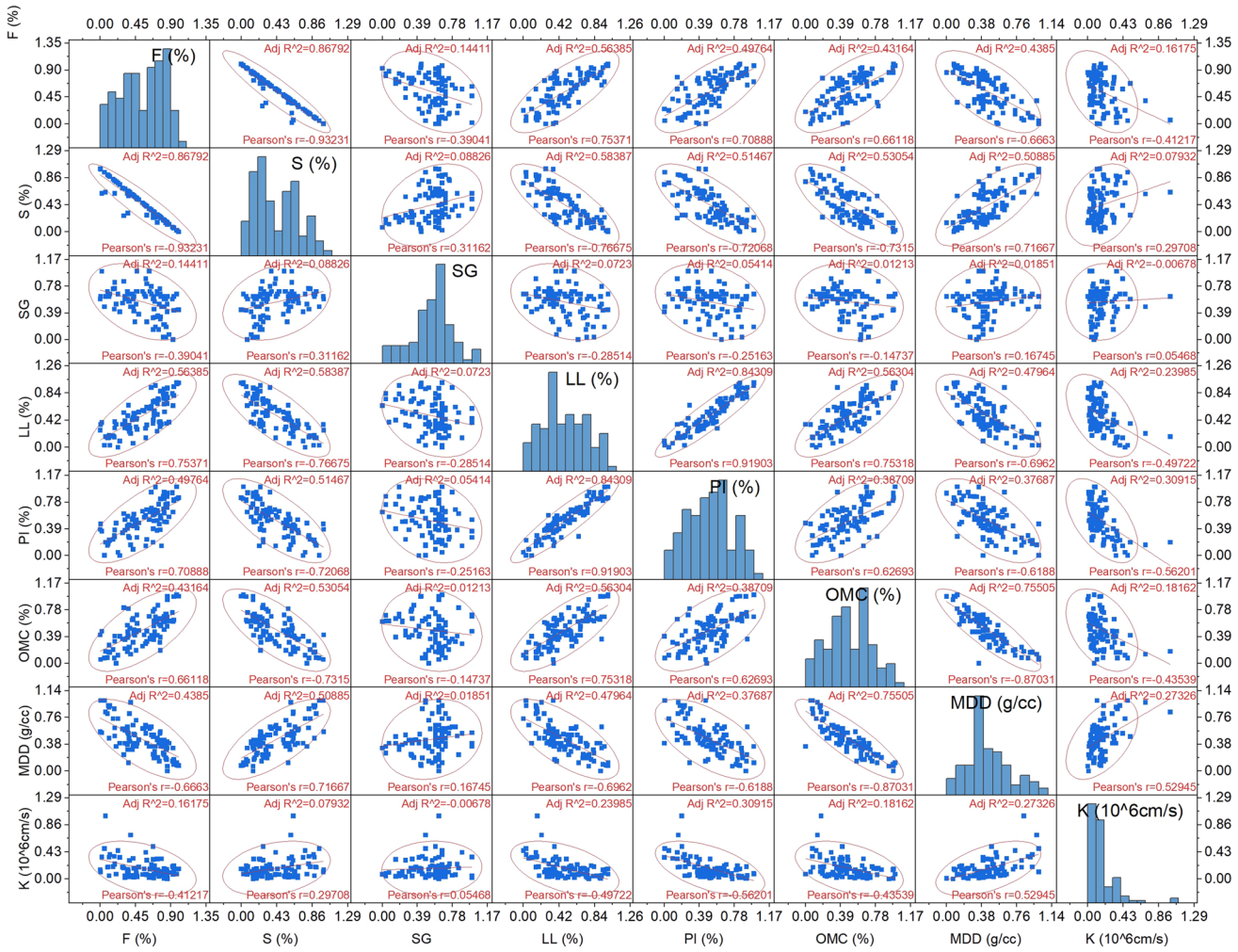

Fig. 4 Correlation coefficient for variables available in the complete database

The variable x_i in array, X is a length of vector m as:

$$x_i = \{x_{i1}, x_{i2}, x_{i3} \ldots x_{im}\} \quad (2)$$

The relationship between CAM (strength of the relation) and database of (x_i) and (x_j) is presented by the following equation (Ghorbani et al. 2020):

Table 3 Multicollinearity levels for variables

Parameters	Coeff	Std Error	t Stat	P-value	Lower 95%	Upper 95%	R^2	VIF	Levels
Intercept	10.1404	6.6938	1.5149	0.1331	-3.1468	23.4275	-	-	-
F (%)	-0.1062	0.0198	-5.3654	0.0000	-0.1455	-0.0669	0.8860	8.77	Moderate
S (%)	-0.1474	0.0218	-6.7604	0.0000	-0.1907	-0.1041	0.8959	9.61	Moderate
SG	-2.7710	1.6568	-1.6725	0.0977	-6.0597	0.5177	0.1951	1.24	Weak
LL (%)	0.0667	0.0371	1.7985	0.0752	-0.0069	0.1404	0.9032	10.33	Problematic
PI (%)	-0.2728	0.0591	-4.6155	0.0000	-0.3901	-0.1555	0.8628	7.29	Moderate
OMC (%)	-0.0349	0.0625	-0.5587	0.5777	-0.1591	0.0892	0.8294	5.86	Moderate
MDD (g/cc)	7.3835	1.9442	3.7977	0.0003	3.5243	11.2427	0.7771	4.49	Considerable

Table 4 ANOVA test results

Source of Variation	SS	df	MS	F	P-value	F crit
ANOVA for F (%)						
Between Groups	300,090.07	1.00	300,090.07	3302.25	0.00	3.89
Within Groups	18,720.13	206.00	90.87			
ANOVA for S (%)						
Between Groups	19,857.58	1.00	19,857.58	241.84	0.00	3.89
Within Groups	16,914.69	206.00	82.11			
ANOVA for SG						
Between Groups	84.46	1.00	84.46	88.20	0.00	3.89
Within Groups	197.28	206.00	0.96			
ANOVA for LL (%)						
Between Groups	72,441.08	1.00	72,441.08	2329.37	0.00	3.89
Within Groups	6406.40	206.00	31.10			
ANOVA for PI (%)						
Between Groups	11,754.36	1.00	11,754.36	1258.38	0.00	3.89
Within Groups	1924.22	206.00	9.34			
ANOVA for OMC (%)						
Between Groups	13,652.35	1.00	13,652.35	1957.13	0.00	3.89
Within Groups	1436.99	206.00	6.98			
ANOVA for MDD (g/cc)						
Between Groups	4.37	1.00	4.37	4.55	0.03	3.89
Within Groups	197.89	206.00	0.96			

Table 5 Z test results

Parameters	F (%)	S (%)	SG	LL (%)	PI (%)	OMC (%)	MDD (g/cc)	K (10^{-6} cm/s)
Mean	77.37	20.95	2.68	38.73	16.44	17.61	1.69	1.40
Known Variance	179.84	162.31	0.00	60.29	16.77	12.04	0.01	1.91
Observations	104.00	104.00	104.00	104.00	104.00	104.00	104.00	104.00
Mean Difference	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
z	57.47	15.55	9.40	48.26	35.48	44.24	2.13	
P(Z<=z) one-tail	0.00	0.00	0.00	0.00	0.00	0.00	0.02	
z Critical one-tail	1.64	1.64	1.64	1.64	1.64	1.64	1.64	
P(Z<=z) two-tail	0.00	0.00	0.00	0.00	0.00	0.00	0.03	
z Critical two-tail	1.96	1.96	1.96	1.96	1.96	1.96	1.96	

$$CAM = \frac{\sum_{k=1}^{m} x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{ik}^2 \sum_{k=1}^{m} x_{jk}^2}} \quad (3)$$

The CAM value close to one presents that the specific input variable is highly significant in the prediction. The CAM value close to zero shows the least significance. Figure 5 depicts the sensitivity of input variables F, S, LL, PI, OMC, and MDD in predicting the hydraulic conductivity of clayey soil. Figure 5 shows that sand (=0.7187), specific gravity (=0.7150), and MDD (=0.7343) are significant input variables in predicting the hydraulic conductivity of clayey soil. It can be seen that the fine content (=0.6548) also influences the hydraulic conductivity of clayey soil, followed by OMC (=0.6423), LL (=0.6324), and PI (=0.5986).

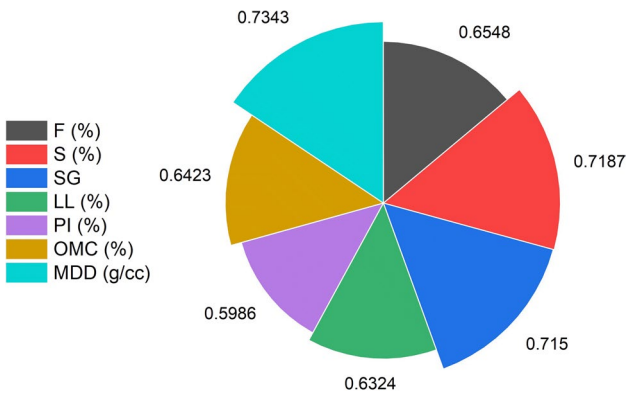

Fig. 5 Depiction of sensitivity analysis

Performance metrics

Performance metrics, which are diverse statistical factors, are used to assess the efficiency of soft computing. Both linear and nonlinear indicators of performance are used. Sixteen performance metrics have been used in this study to evaluate the performance of machine learning models and check the reliability of the best architectural model. The following is how the performance determined expressed mathematically (Kumar and Samui 2020; Asteris et al. 2021a, 2021b; Khatti and Grover 2021, 2023b, 2023c):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\alpha-\beta)^2} \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(\omega-\alpha)| \tag{5}$$

$$R^2 = \frac{\sum_{i=1}^{r}(\alpha-\beta)^2 - \sum_{i=1}^{r}(\alpha-\omega)^2}{\sum_{i=1}^{r}(\alpha-\beta)^2} \tag{6}$$

$$R = \frac{\sum\left(\alpha_i-\overline{\beta}\right)\left(\omega_i-\overline{\omega}\right)}{\sqrt{\sum\left(\alpha_i-\overline{\beta}\right)^2\sum\left(\omega_i-\overline{\omega}\right)^2}} \tag{7}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\alpha-\omega}{\alpha}\right| * 100 \tag{8}$$

$$WMAPE = \frac{\sum_{i=1}^{n}\left|\frac{\alpha-\omega}{\alpha}\right| * \alpha}{\sum_{i=1}^{n}\alpha} \tag{9}$$

$$VAF = \left(1-\frac{var(\alpha-\omega)}{var(\alpha)}\right) * 100 \tag{10}$$

$$PI = R^2 + \left({}^{VAF}/_{100}\right) - RMSE \tag{11}$$

$$RSR = \frac{RMSE}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\alpha-\beta)^2}} \tag{12}$$

$$NS = 1\frac{\sum_{i=1}^{n}(\alpha-\omega)^2}{\sum_{i=1}^{n}(\alpha-\beta)^2} \tag{13}$$

$$BF = \frac{1}{n}\sum_{i=1}^{n}\frac{\omega}{\alpha} \tag{14}$$

$$a20index = \frac{m20}{H} \tag{15}$$

$$IOA = 1-\frac{\sum_{i=1}^{n}(\omega-\alpha)}{2\sum_{i=1}^{n}(\alpha-\beta)} \tag{16}$$

$$IOS = \frac{RMSE}{Avg.\ of\ Actual\ Values} \tag{17}$$

$$LMI = 1-\left[\frac{\sum_{i=1}^{n}|\alpha-\omega|}{\sum_{i=1}^{n}|\alpha-\beta|}\right] \tag{18}$$

$$NMBE = \frac{\frac{1}{N}\sum_{i=1}^{n}(\omega-\alpha)^2}{\frac{1}{N}\sum_{i=1}^{n}\alpha} \tag{19}$$

$$MBE = \frac{1}{N}\sum_{i=1}^{n}(\omega-\alpha) \tag{20}$$

where α and ω are the real and assessed i^{th} value, n is the total number of data, β presents the mean of the real values, $\overline{\omega}$ presents the mean of the assessed value, k presents the number of independent variables, m20 is the ratio of real to the assessed value, varies from 0.8 to 1.2, and H presents the total data samples. On the other hand, the index of agreement (IOA) is bounded by -1.0 and 1.0 (Willmott et al. 2012). Moreover, the scatter index (IOS) value close to zero presents an excellent prediction and accuracy (Mentaschi et al. 2013). A computational model is reliable and accurate if it achieves R^2 over 0.95. Also, a weak, good, and strong relationship between actual and computed data is presented if a pair of data has R less than 0.2, between 0.2 to 0.8, and more than 0.8 (Smith 1986). A perfect predictive model always has performance indicators' values equal to the ideal value, as in Table 6 (Bahmed et al. 2024; Daniel et al. 2024).

Table 6 Ideal value of the different performance indicators

Indicators	Value	Indicators	Value
RMSE	0	NMBE	0
MAE	0	NS	1
R2	1	LMI	0
R	1	RSR	0
MAPE	0–100	a20-index	100
WMAPE	0	IOA	1
VAF	100	IOS	0
PI	2	BF	0

Computational approach

A computational model called an Artificial Neural Network (ANN), also known simply as a neural network, is modeled after the structure and operation of biological neural networks, such as the human brain. Artificial intelligence (AI) and machine learning techniques, including neural networks, are utilized for various tasks, such as classification, regression, pattern recognition, and more. The fundamental components of an artificial neural network are the neurons, layers, weights/ Biases, activation functions, feedforward process, backpropagation algorithm, loss functions, and cost functions. A backpropagation algorithm (BA) must be carefully selected because it distributes the prediction error by updating the neuron's weight. The GDM, GD, GDA, SCG, BFG, and LM algorithms have been compared to find the best BA. Ninety artificial neural network models (fifteen for each BA) have been developed, learned, and analyzed. Each ANN model has been configured with a min-gradient of 10e-7, max fail of 6, momentum of 0.001, multilayer perceptron class, feedforward backpropagation, 1000 epochs, train: valid: test of 76%: 12%: 12%, sigmoid activation function at hidden layers, linear activation function at output layer, log function for output normalization, min–max function for input normalization, one to five hidden layers (HL) interconnected with each 5, 10, and 15 neurons (N). The reasons for selecting the sigmoid function in this research are (a) most of the published work used the sigmoid function, (b) the smoothness of the sigmoid allows for stable and continuous updates to model parameters, and (c) it is easy to set a threshold (e.g., 0.5) for decision-making. Table 7 summarizes the designations of the ANN model for LM, BFG, SCG, GDM, GD, and GDA algorithms.

Results and discussion

Simulation of results

The training (TRG), validation (VDN), and testing (TSG) performance results for developed, learned, and analyzed ANN models are summarized and presented in Appendix – I (Tables A, B, and C). Table A presents that model LM_K15 has an excellent TRG performance, i.e., R = 0.9924, NS = 0.9846, PI = 1.8038, BF = 0.9923, WI = 0.9758, MBE = 0.0175, LMI = 0.1390. Model LM_K15 assesses the hydraulic conductivity with a minor prediction error, i.e., NMBE = 0.0204, WMAPE = 0.0978, MAPE = 0.1984, MAE = 0.1321, and RMSE = 0.1658. The TRG performance comparison reveals that model LM_K15 achieves VAF of 98.48, RSR of 0.1241, IOA of 0.9305, IOS of 0.1227, and a20 of 68.75. The hydraulic conductivity of twelve soil samples is computed to validate model LM_K15. The VDN performance comparison reveals that model LM_K15 gains R of 0.9719, IOA of 0.8834, IOS of 0.3258, a20 of 72.00, RSR of 0.2601, LMI of 0.2332, MBE of 0.0227, WI of 0.9791, BF of 1.1231, PI of 1.3859, NS of 0.9668, and VAF of 96.89, better than other LM_ANN models and close to ideal values. Model LM_K15 predicts the hydraulic conductivity with the least validation residuals, i.e., NMBE = 0.1202, RMSE = 0.4928, WMAPE = 0.2168, MAE = 0.3212, and MAPE = 0.4536. The TSG phase reveals

Table 7 Designation of ANN models

Number of HL	Number of Neurons	Backpropagation Algorithms					
		LM	BFG	SCG	GDM	GD	GDA
1	5	LM_K1	BFG_K1	SCG_K1	GDM_K1	GD_K1	GDA_K1
1	10	LM_K2	BFG_K2	SCG_K2	GDM_K2	GD_K2	GDA_K2
1	15	LM_K3	BFG_K3	SCG_K3	GDM_K3	GD_K3	GDA_K3
2	5	LM_K4	BFG_K4	SCG_K4	GDM_K4	GD_K4	GDA_K4
2	10	LM_K5	BFG_K5	SCG_K5	GDM_K5	GD_K5	GDA_K5
2	15	LM_K6	BFG_K6	SCG_K6	GDM_K6	GD_K6	GDA_K6
3	5	LM_K7	BFG_K7	SCG_K7	GDM_K7	GD_K7	GDA_K7
3	10	LM_K8	BFG_K8	SCG_K8	GDM_K8	GD_K8	GDA_K8
3	15	LM_K9	BFG_K9	SCG_K9	GDM_K9	GD_K9	GDA_K9
4	5	LM_K10	BFG_K10	SCG_K10	GDM_K10	GD_K10	GDA_K10
4	10	LM_K11	BFG_K11	SCG_K11	GDM_K11	GD_K11	GDA_K11
4	15	LM_K12	BFG_K12	SCG_K12	GDM_K12	GD_K12	GDA_K12
5	5	LM_K13	BFG_K13	SCG_K13	GDM_K13	GD_K13	GDA_K13
5	10	LM_K14	BFG_K14	SCG_K14	GDM_K14	GD_K14	GDA_K14
5	15	LM_K15	BFG_K15	SCG_K15	GDM_K15	GD_K15	GDA_K15

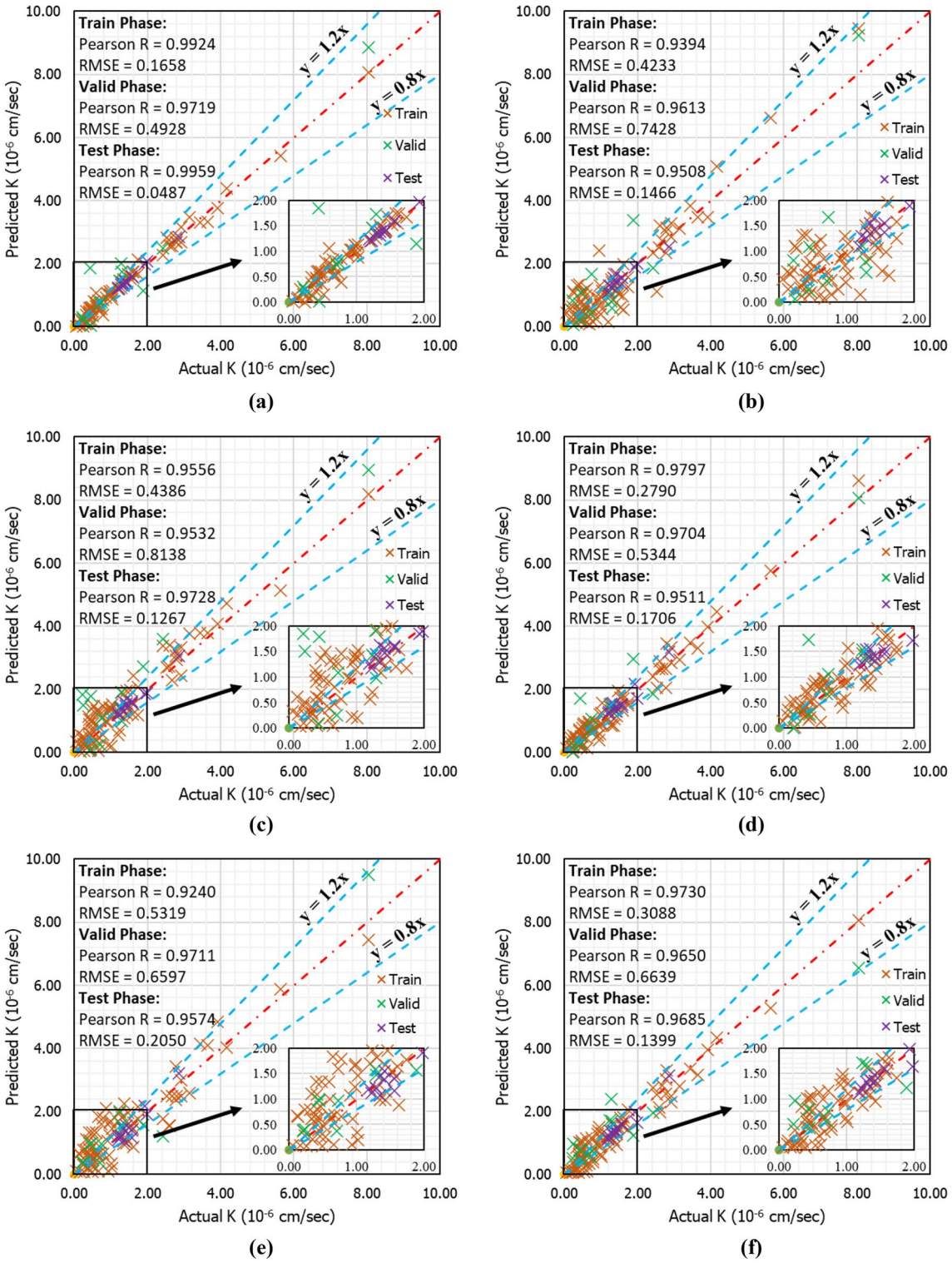

Fig. 6 Presentation of statistical relationship between experimental and predicted hydraulic conductivity of clayey soil using model (**a**) LM_K15, (**b**) BFG_K9, (**c**) SCG_K10, (**d**) GDM_K12, (**e**) GD_K7, and (**f**) GDA_K12

that model LM_K15 estimates the hydraulic conductivity with R of 0.9959, a20 of 100, IOA of 0.9364, IOS of 0.0306, VAF of 99.16, NS of 0.9882, PI of 1.9348, RSR of 0.1084, LMI of 0.1272, MBE of 0.0260, WI of 0.9339, BF of 1.0163. The remaining TSG performance metrics confirm the prediction capabilities of model LM_K15, i.e., WMAPE = 0.0264, MAPE = 0.0279, NMBE = 0.0015, RMSE = 0.0487, and MAE = 0.0419. A thorough analysis demonstrates that model LM_K15, configured with five hidden layers and 15 neurons, attains higher TRG, VDN, and TSG performance. The overall performance analysis for LM_ANN models reveals that the predictive capabilities increase with an increasing number of neurons (i.e., 15) and hidden layers (i.e., 5). Figure 6 (a) presents a statistical relationship between experimental and predicted hydraulic conductivity using the LM_K15 model in the TRG, VDN, and TSG phases.

The performance comparison for the BFG_ANN models reveals that model BFG_K9, configured with three hidden layers interconnected by 15 neurons, attains higher performance (PI = 1.2661, NS = 0.8670, BF = 1.3353, WI = 0.9390, a20 = 43.75, R = 0.9394, RMSE = 0.4233, NMBE = 0.1203, MAPE = 0.6616, MAE = 0.3947, and WMAPE = 0.2921) in the TRN phase. Similarly, model BFG_K9 has VDN (BF = 1.3080, NMBE = 0.3505, WI = 0.9642, MBE = 0.1959, LMI = 0.4289, RSR = 0.3617, a20 = 56.67, IOA = 0.7955, IOS = 0.4719, PI = 1.0595, NS = 0.8692, WMAPE = 0.3295, VAF = 89.83, MAPE = 0.7589, R = 0.9613, MAE = 0.5213, RMSE = 0.7428) and TSG (R = 0.9508, PI = 1.6514, a20 = 100, RMSE = 0.1466, VAF = 89.39, NMBE = 0.0135, MAE = 0.1057) performances higher than other BFG_ANN models and close to ideal values. Figure 6 (b) depicts the relationship between actual and predicted hydraulic conductivity using model BFG_K9. The TRG, VDN, and TSG comparisons show that the performance of BFG_ANN models increases with an increasing number of hidden layers and neurons, up to 3 hidden layers interconnected with 15 neurons. The performance of BFG_ANN models decreases with increasing the hidden layers and neurons.

The performance comparison for SCG_ANN models shows that model SCG_K10 achieves a good predictive performance in the TRG (R = 0.9556, VAF = 90.52, NS = 0.8923, PI = 1.3797, BF = 1.1227, a20 = 46.25, WI = 0.9386, RMSE = 0.4386, IOA = 0.8079, WMAPE = 0.5843, MAE = 0.3651, IOS = 0.3246,

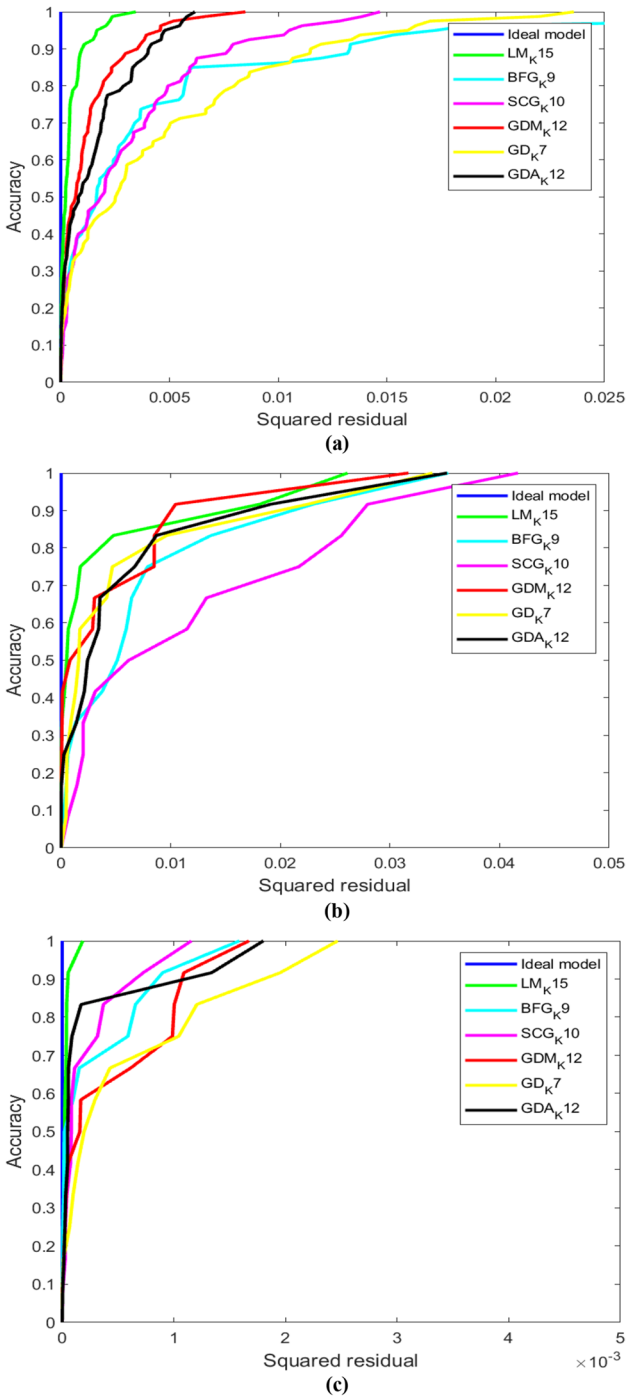

Fig. 7 Representation of REC curve for the best architectural models in (**a**) training, (**b**) validation, and (**c**) testing phase

Table 8 AOC details for the best architectural models

Phase	Actual	LM_K15	BFG_K9	SCG_K10	GDM_K12	GD_K7	GDA_K12
Train	0.00E + 00	**4.09E-04**	4.08E-03	2.92E-03	1.17E-03	4.28E-03	1.45E-03
Valid	0.00E + 00	**3.38E-03**	7.17E-03	1.13E-02	4.19E-03	5.41E-03	5.44E-03
Test	0.00E + 00	**2.94E-05**	2.70E-04	2.03E-04	4.15E-04	5.56E-04	2.31E-04

*Bold value represents the optimum performance model

MAPE = 0.5843), VDN (RMSE = 0.8138, VAF = 88.91, PI = 0.8838, a20 = 36.67, MAE = 0.6906, IOS = 0.5181, NS = 0.8320), and TSG (a20 = 100, IOS = 0.0797, IOA = 0.8444, R = 0.9728, MAPE = 0.0663, MAE = 0.1026, WMAPE = 0.0646, NS = 0.9203, PI = 1.7593, RMSE = 0.1267, RSR = 0.2822) phases. Figure 6 (c) illustrates the regression relationship between actual and assessed hydraulic conductivity of clayey soil using model SCG_K10, configured with four hidden layers interconnected with five neurons.

In the case of the GDM_ANN models, the TRG performance comparison presents that model GDM_K12 assesses the hydraulic conductivity with a minimum residual (NMBE = 0.0576, WMAPE = 0.1683, MAE = 0.2274, RMSE = 0.2790, MAPE = 0.3413) and high performance (RSR = 0.2088, MBE = 0.0145, WI = 0.9605, NS = 0.9564, PI = 1.6372, R = 0.9797). Model GDM attains RMSE of 0.5344, R of 0.9704, VAF of 94.08, PI of 1.3482, RSR of 0.2602, IOA of 0.8627 and MAE of 0.3487, comparatively better than other GDM_ANN models, in the VDN phase. Model GDM_K12 also attains outstanding performance (RMSE = 0.1706, MAE = 0.1039, MAPE = 0.0683, WMAPE = 0.0688, NMBE = 0.0179, R = 0.9511, VAF = 86.22, PI = 1.5908, RSR = 0.3192) compared to other GDM_ANN models in the TSG phase. Model GDM_K12 is configured with three hidden layers interconnected with fifteen neurons. A statistical relationship between experimental and predicted hydraulic conductivity is shown in Fig. 6 (d).

Furthermore, model GD_K7 outperforms the other GD_ANN models with an acceptable TRG (IOS = 0.3937, RSR = 0.3980, BF = 1.3788, VAF = 85.02, R = 0.9240, NMBE = 0.2094, LMI = 0.4525, IOS = 0.7737, RMSE = 0.5319, PI = 1.1720, WMAPE = 0.3183), VDN (NMBE = 0.2765, WI = 0.9703, LMI = 0.4085, RSR = 0.3212, VAF = 90.73, R = 0.9711, MAE = 0.5186, RMSE = 0.6597, a20 = 35.00 and IOS = 0.4191), and TSG (NS = 0.7913, PI = 1.5055, BF = 0.9718, NMBE = 0.0264, WI = 0.8006, LMI = 0.4991, RSR = 0.4568, a20 = 100, MAE = 0.1645, RMSE = 0.2050) performance, close to the ideal values. Figure 6 (e) depicts the relationship between actual and predicted hydraulic conductivity using the GD_K7 model.

Moreover, the performance comparison for model GDA_ANN reveals that model GDA_K12 predicts hydraulic conductivity with a good TRN (RMSE = 0.3088, MAE = 0.2515, R = 0.9730, MAPE = 0.3451, VAF = 94.66, WMAPE = 0.1862, NS = 0.9466, PI = 1.5846, BF = 1.0411, NMBE = 0.0706, WI = 0.9530, MBE = -0.0035, LMI = 0.2647, RSR = 0.2310, a20 = 48.75, IOS = 0.2285 and IOA = 0.8677), VDN (R = 0.9650, VAF = 89.62, a20 = 66.67, RMSE = 0.6639, PI = 1.1634, IOS = 0.4218, MAE = 0.5150, IOA = 0.7972, NS = 0.8955, WMAPE = 0.3271), and TSG (R = 0.9685, WAF = 90.29, NS = 0.9029, PI = 1.7009, a20 = 91.67, RMSE = 0.1399, MAPE = 0.0551, RSR = 0.3117) performance. The overall performance analysis for GDA_ANN models reveals that the predictive capabilities increase with an increasing number of neurons (i.e., 15) and hidden layers (i.e., 4). Figure 6 (f) presents a statistical relationship between experimental and predicted hydraulic conductivity using the GDA_K12 model in the TRG, VDN, and TSG phases. Finally, the six best architectural ANN models (one from each backpropagation algorithm) have been identified to predict the hydraulic conductivity of clayey soil.

In the continuity of results simulation, a visual interpretation of results has been graphically presented, analyzed, and discussed. The regression error characteristics (REC) curve has been discussed and analyzed, followed by rank, uncertainty, and Wilcoxon analysis. The Anderson–darling test has been performed to determine the normality of predicted hydraulic conductivity concerning the actual hydraulic conductivity of clayey soil.

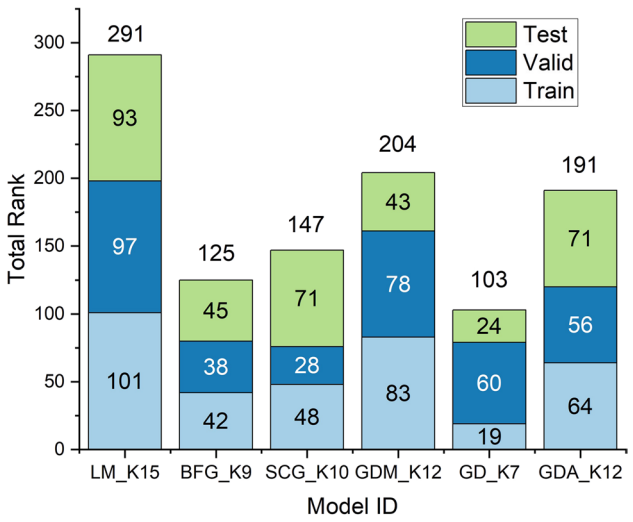

Fig. 8 Rank analysis for the best architectural models

REC curve

Regression Error Characteristic (REC) curves show how well the regressor model functions. The REC graph compares the percentage of exemplars correctly predicted within the tolerance interval against the absolute deviation tolerance. As a result, a curve is produced that calculates the error's cumulative distribution function. A biased estimate of the predicted error is provided by the area over the REC curve (AOC), created by subtracting the area under the REC curve from one. The coefficient of determination (R^2) concerning the AOC can also be computed (Tahmassebi et al. 2018; Bi and Bennett 2003). The MATLAB R2020a framework has been used to plot the REC curve for the best

Table 9 Score of the best architecture model for rank analysis

Algorithm	Model ID	Phase	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS	Total
LM	**LM_K15**	**Train**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**5**	**6**	**6**	**6**	**6**	**6**	**101**
		Valid	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**1**	**6**	**6**	**6**	**6**	**6**	**97**
		Test	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**6**	**3**	**6**	**6**	**4**	**6**	**6**	**2**	**6**	**6**	**93**
BFG	BFG_K9	Train	3	2	2	1	4	2	2	2	2	3	3	6	2	2	2	2	2	42
		Valid	2	2	2	1	3	2	2	2	2	2	2	4	2	2	4	2	2	38
		Test	3	2	1	4	3	3	3	3	4	3	2	2	3	2	2	2	3	45
SCG	SCG_K10	Train	2	3	3	3	2	3	3	3	3	2	2	4	3	3	3	3	3	48
		Valid	1	1	1	2	1	1	1	1	5	1	1	6	1	1	2	1	1	28
		Test	5	4	5	3	5	4	5	5	1	5	4	6	4	5	2	3	5	71
GDM	GDM_K12	Train	5	5	5	5	5	5	5	5	5	5	5	3	5	5	5	5	5	83
		Valid	5	5	4	5	5	5	5	5	3	5	5	3	5	5	3	5	5	78
		Test	2	3	2	2	2	2	2	2	2	2	3	5	2	3	2	5	2	43
GD	GD_K7	Train	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	1	1	19
		Valid	4	3	5	4	4	3	4	4	1	4	4	5	3	4	1	3	4	60
		Test	1	1	3	1	1	1	1	1	5	1	1	1	1	1	2	1	1	24
GDA	GDA_K12	Train	4	4	4	4	3	4	4	4	4	4	4	1	4	4	4	4	4	64
		Valid	3	4	3	3	2	4	3	3	4	3	3	2	4	3	5	4	3	56
		Test	4	5	4	5	4	5	4	4	6	4	5	3	5	4	1	4	4	71

*Bold values represent the optimal performance model

Table 10 Details of parameters of UA

Model ID	MOE	SD	SE	ME	LB	UB	WCB	Rank
Training Phase								
LM_K15	**0.0164**	**0.0126**	**0.0014**	**0.0028**	**0.0125**	**0.0203**	**0.0078**	**1**
BFG_K9	0.0494	0.0433	0.0048	0.0095	0.0189	0.0799	0.0611	5
SCG_K10	0.0456	0.0307	0.0034	0.0067	0.0023	0.0935	0.0913	4
GDM_K12	0.0284	0.0204	0.0023	0.0045	0.0440	0.1007	0.0567	2
GD_K7	0.0538	0.0394	0.0044	0.0086	0.0161	0.0916	0.0755	6
GDA_K12	0.0315	0.0225	0.0025	0.0049	0.0289	0.0340	0.0050	3
Validation Phase								
LM_K15	**0.0427**	**0.0534**	**0.0154**	**0.0302**	**0.0278**	**0.0577**	**0.0299**	**1**
BFG_K9	0.0778	0.0532	0.0154	0.0301	0.0422	0.1977	0.1555	5
SCG_K10	0.0649	0.0533	0.0154	0.0302	0.0063	0.1361	0.1298	4
GDM_K12	0.0521	0.0546	0.0158	0.0309	0.1283	0.2326	0.1042	2
GD_K7	0.0989	0.0599	0.0173	0.0339	0.0549	0.2527	0.1979	6
GDA_K12	0.0645	0.0548	0.0158	0.0310	0.2065	0.3354	0.1289	3
Testing Phase								
LM_K15	**0.0052**	**0.0032**	**0.0009**	**0.0018**	**0.0061**	**0.0166**	**0.0105**	**1**
BFG_K9	0.0132	0.0133	0.0038	0.0075	0.0061	0.0204	0.0143	4
SCG_K10	0.0128	0.0097	0.0028	0.0055	0.0049	0.0306	0.0257	3
GDM_K12	0.0174	0.0141	0.0041	0.0080	0.0442	0.0790	0.0348	5
GD_K7	0.0206	0.0160	0.0046	0.0091	0.0427	0.0839	0.0412	6
GDA_K12	0.0122	0.0131	0.0038	0.0074	0.0024	0.0268	0.0244	2

*Bold values represent the optimal performance model

architectural models, i.e., LM_K15, BFG_K9, SCG_K10, GDM_K12, GD_K7, and GDA_K12. Figure 7 (a-c) illustrates the REC plot of the best architectural models in the TRG, VDN, and TSG phases and AOC values for each best architectural model in Table 8. The AOC values given in Table 8 demonstrate that model LM_K15 has predicted the hydraulic conductivity of clayey soil with the least AOC (TRG = 4.09E-04, VDN = 3.38E-03, and TSG = 2.94E-05) and recognized as an optimal performance neural network model.

Rank analysis

Another easy method for contrasting model performance is "Rank Analysis." (Khatti et al. 2024) In this technique, the model with the best value for each performance parameter is given a score of "n" (in this study, n = 6; this refers to the number of computational models that are taken into account in the analysis), and the model with the worst value for the same performance parameter is given a score of 1 (one), separately for training and testing results. The next step is to add up each model's scores to determine the final score of the models. The model's final score is calculated using the combined scores from the training and testing phases (Asteris et al. 2021c). Table 9 presents the details of rank analysis for the best architectural model for the TRG, VDN, and TSG phases. Table 9 demonstrates that model LM_K15 has obtained 101, 97, and 93 ranks in the training, validation, and testing phases, comparatively higher than other best architectural models. Model GDM_K12 has secured second rank with 83, 78, and 43 scores in the TRG, VDN, and TSG phases, followed by models GDA_K12, SCG_K10, BFG_K9, and GD_K7. Figure 8 illustrates the overall rank of the best architectural models, and it is noted that model LM_K15 has the highest rank in predicting the hydraulic conductivity of clayey soil, i.e., 291. Model GD_K7 has the lowest rank, i.e., 103.

Uncertainty analysis

Any predictive model's credibility must be evaluated to estimate predictive outputs with accuracy and reliability. Uncertainty analysis (UA) has been used in the current work to quantify the error of the top architectural models in forecasting the hydraulic conductivity of soils. The training, validation, and testing datasets containing 80, 12, and 12 experimental data points related to clayey soils have been subjected to UA. Therefore, comparing prediction results with these experimental datasets is important in determining how reliable the constructed models are, and UA is perfectly suited for this task (Bardhan et al. 2021). The results of the UA analysis are summarized in Table 10.

Table 10 presents the UA results of the six best architectural models, i.e., LM_K15, BFG_K9, SCG_K10,

Table 11 Results of the Wilcoxon analysis

Model ID	Num	Median	Confidence Levels		Achieved Confidence
			LCL	UCL	
Training Phase					
Actual	80	1.0650	0.8650	1.4600	94.99%
LM_K15	**80**	**1.0758**	**0.8628**	**1.4632**	**94.99%**
BFG_K9	80	1.0399	0.8294	1.4073	94.99%
SCG_K10	80	1.2672	0.9981	1.6178	94.99%
GDM_K12	80	1.0801	0.8603	1.4129	94.99%
GD_K7	80	1.2728	1.0568	11.5222	94.99%
GDA_K12	80	1.1247	0.9014	1.4160	94.99%
Validation Phase					
Actual	12	1.0275	0.5650	1.8800	94.54%
LM_K15	**12**	**1.0566**	**0.5983**	**1.9714**	**94.54%**
BFG_K9	12	1.0943	0.8506	2.5230	94.54%
SCG_K10	12	1.6709	0.9893	2.7519	94.54%
GDM_K12	12	1.3631	0.7180	2.3381	94.54%
GD_K7	12	1.2136	0.6878	1.7873	94.54%
GDA_K12	12	1.2456	0.7640	2.0450	94.54%
Testing Phase					
Actual	12	1.4841	1.3155	1.9244	94.54%
LM_K15	**12**	**1.4914**	**1.3298**	**1.9306**	**94.54%**
BFG_K9	12	1.5260	1.3330	1.8315	94.54%
SCG_K10	12	1.5636	1.3952	1.8520	94.54%
GDM_K12	12	1.4978	1.3797	1.8597	94.54%
GD_K7	12	1.4147	1.2062	1.9098	94.54%
GDA_K12	12	1.4724	1.2982	1.7918	94.54%

*Bold values correspond to the optimal performance model

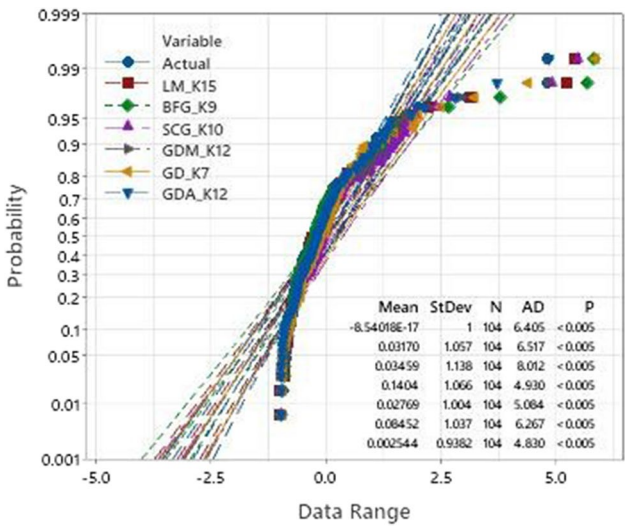

Fig. 9 Comparison of AD test results

GDM_K12, GD_K7, and GDA_K12. It is noted that model LM_K15 has the lowest value of the width of confidence bend (WCB) in the TRG, VDN, and TSG phases. The other parameters, margin error (ME), standard deviation (SD), mean of error (MOE), square error (SE), upper bound (UB), and lower bound (LB), have also obtained less than BFG_K9, SCG_K10, GDM_K12, GD_K7, and GDA_K12 models. Based on these parameters, model LM_K15 has gained first rank and shows superiority in predicting the hydraulic conductivity of clayey soil. Hence, model LM_K15 is the best architectural model recognized in this work.

Wilcoxon analysis

A non-parametric statistical test called the Wilcoxon test, commonly called the signed-rank test, compares the means of two related or paired groups. It is frequently applied when paired data or when the data deviates from the assumption of normalcy. In this research, the Wilcoxon analysis has been performed for models LM_K15, BFG_K9, SCG_K10, GDM_K12, GD_K7, and GDA_K12 to find the optimal performance model in predicting the hydraulic conductivity of clayey soil. Table 11 summarizes the results of the Wilcoxon analysis. The comparison of results reveals that model LM_K15 has an excellent confidence level, is close to actual values, and presents superiority in predicting the hydraulic conductivity of clayey soil.

Anderson darling test

In addition, the "Anderson–Darling" test (AD) has been run as a non-parametric statistical test to give a more in-depth understanding of the divergence of the results. The A-D test is a statistical procedure used to determine if a sample of data originated from a population with a particular distribution. In order to determine if the actual and anticipated values in the current study fit a normal distribution, the A-D test is used to assess the data. The Minitab Statistical Tool has been used to perform the AD test for the best architectural models. Figure 9 depicts the AD test results for models LM_K15, BFG_K9, SCG_K10, GDM_K12, GD_K7, and GDA_K12.

Figure 9 demonstrates that the best architectural model assessed the hydraulic conductivity of clayey soil with a p-value of 0.005, which is less than the significance value ($p = 0.05$). Still, model LM_K15 has an AD value of 6.517, close to the AD value of actual data, i.e., 6.405. Hence, model LM_K15 rejects the null hypothesis of normality and presents superiority over the other ANN models.

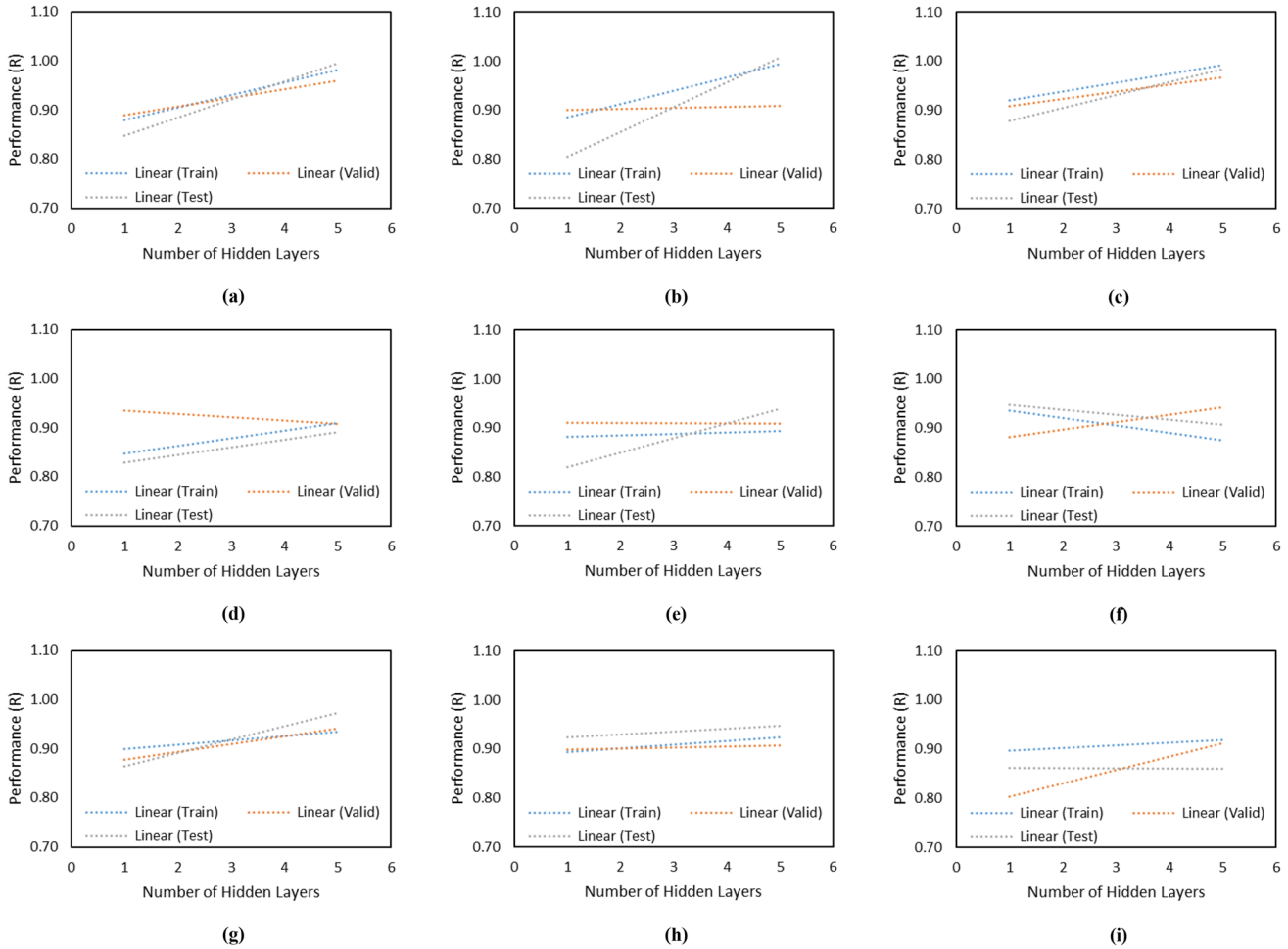

Fig. 10 Illustration of the relationship between hidden layers and performance for models (**a**-**c**) LM_ANN, (**d**-**f**) BFG_ANN, (**g**-**i**) SCG_ANN, (**j**-**l**) GDM_ANN, (**m**–**o**) GD_ANN, and (**p**-**r**) GDA_ANN. Note: Fig. 10 (a, d, g, j, m, p), (b, e, h, k, n, q), and (c, f, i, l, o, r) show the performance of ANN models configured with 5, 10, and 15 neurons

Analysis of results

This research uses artificial neural network models to predict the hydraulic conductivity of clayey soil. Ninety ANN models have been developed using one to five hidden layers interconnected with each 5, 10, and 15 neurons. The six backpropagation algorithms have been implemented and compared to find the best algorithm. This section analyses the performance of ANN models concerning the number of hidden layers and neurons. For that purpose, a statistical linear relationship has been drawn for varying numbers of hidden layers and constant neurons. Figure 10 (a-q) shows the relationship for LM_ANN, BFG_ANN, SCG_ANN, GDM_ANN, GD_ANN, and GDA_ANN models in the TRG, VDN, and TSG phases. Figure 10 (a-c) shows that the performance of LM_ANN models increases with hidden layers. It has also been observed that five and fifteen-neuron-based LM_ANN models give the most promising prediction. Figure 10 (d) reveals that the performance of BFG_ANN models increases with hidden layers interconnected with five neurons. Figure 10 (g-i) illustrates that the performance of SCG_ANN models increases with hidden layers and neurons. Figure 10 (j-l) demonstrates that GDM_ANN models give the most promising results with five and fifteen neurons. However, the performance of GDM_ANN models increases with hidden layers. Still, the GDM_ANN models based on 15 neurons achieve higher performance than the GDM_ANN models based on five neurons. Figure 10 (m–o) does not show a significant performance concerning hidden layers and neurons. Therefore, increasing the epochs to analyze the effect of hidden layers and neurons may be suggested. Figure 10 (p-r) shows insignificant results for GDA_ANN models. Figure 10 (r) depicts that the GDA_ANN model requires many neurons to attain high performance in predicting the hydraulic conductivity of clayey soil. The epochs may also be increased for GDA_ANN models to analyze the impact of hidden layers and neurons. The overall analysis presents that the

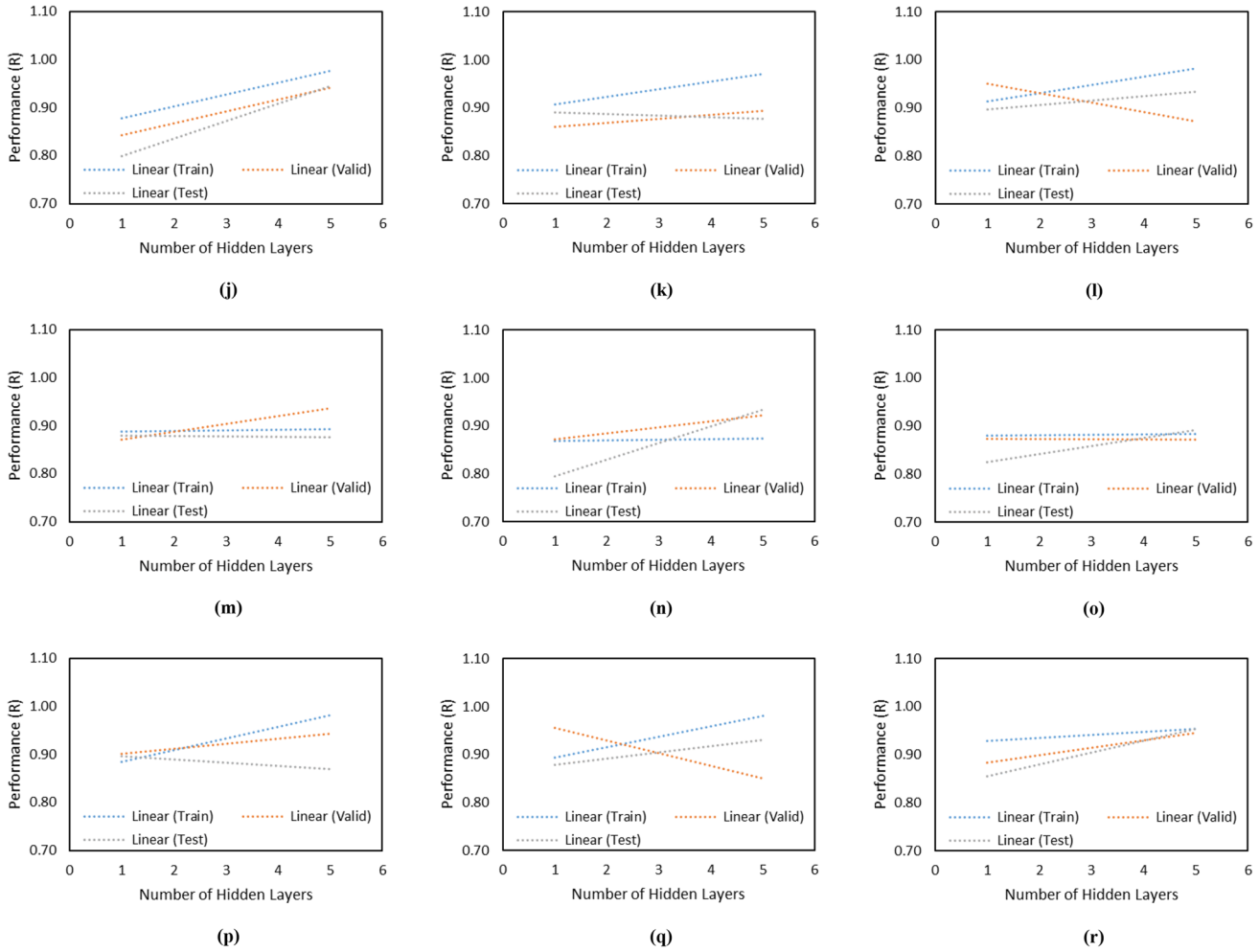

Fig. 10 (continued)

LM_ANN models achieve excellent performance with a significant number of hidden layers and neurons.

Furthermore, the database used in this research has been simulated to validate the model's capabilities. For this aim, one input variable varies to create the simulated database, while the other variables remain constant. For example, to ensure the trend of fine content and hydraulic conductivity, only the value of fine content linearly varies, and the other input variables are constant. However, a soil sample consists of 100% soil particles. These particles are gravel, sand, silt, and clay, and their summation can't exceed 100%. This research uses fine content terms, a summation of silt and clay content. In this study, the simulated hydraulic conductivity of clayey soil has been created using F (%), S (%), LL (%), PI (%), OMC (%), and MDD (g/cc). Table 12 presents the details of the simulated database.

Figure 11 presents the relationship between the simulated database of F, S, SG, LL, PI, OMC, and MDD with the hydraulic conductivity of clayey soil. Figure 11 (a1-a4) shows that the hydraulic conductivity of clayey soil is inversely proportional to fine content. Figure 11 (b1-b4) illustrates that the hydraulic conductivity decreases with increased sand content. Figure 11 (c1-c4) shows that specific gravity increases by decreasing the hydraulic conductivity. It can be seen that the specific gravity starts increasing with a decrease in fine content and an increase in sand content. Figure 11 (d1-d4) demonstrates that the liquid limit of clayey soil significantly affects the hydraulic conductivity of clayey soils. Figure 11 (e1, e2, e4) illustrates that hydraulic conductivity increases with the plasticity of clayey soil. Figure 11 (f1-f4) reveals that the hydraulic conductivity of clayey soil decreases due to an increase in optimum moisture content. Figure 11 (g1-g4) presents a significant impact of maximum dry density on the hydraulic conductivity of clayey soil. The sensitivity analysis has also reported that MDD of clayey soil influences the prediction of hydraulic conductivity of soil.

Table 12 Details of simulated database

Variable Inputs		Constant Input Variables	Data	References
Para	**Range**			
F	a1: 50–60	a1: SG = 2.71, LL = 29.96, PI = 12.15, OMC = 13.54, MDD = 1.83	16×5	Figure 11 (a1-a4)
	a2: 60–70	a2: SG = 2.70, LL = 34.24, PI = 14.35, OMC = 16.67, MDD = 1.69		
	a3: 70–80	a3: SG = 2.70, LL = 36.25, PI = 15.25, OMC = 17.16, MDD = 1.71		
	a4: 80–90	a4: SG = 2.67, LL = 41.20, PI = 17.43, OMC = 18.13, MDD = 1.67		
S	b1: 0–10	b1: SG = 2.64, LL = 47.16, PI = 20.79, OMC = 20.90, MDD = 1.61	11×5	Figure 11 (b1-b4)
	b2: 10–20	b2: SG = 2.68, LL = 41.09, PI = 14.54, OMC = 18.54, MDD = 1.67		
	b3: 30–40	b3: SG = 2.69, LL = 31.52, PI = 12.98, OMC = 14.58, MDD = 1.74		
	b4: 40–50	b4: SG = 2.69, LL = 28.84, PI = 11.96, OMC = 12.87, MDD = 1.86		
SG	c1: 2.60–2.65	c1: F = 82, S = 18, LL = 41.22, PI = 16.65, OMC = 18.89, MDD = 1.66	11×6	Figure 11 (c1-c4)
	c2: 2.65–2.70	c2: F = 76, S = 24, LL = 37.42, PI = 16.02, OMC = 16.88, MDD = 1.70		
	c3: 2.70–2.75	c3: F = 74, S = 26, LL = 36.71, PI = 16.22, OMC = 16.48, MDD = 1.72		
	c4: 2.75–2.80	c4: F = 70, S = 30, LL = 37.94, PI = 15.11, OMC = 19.79, MDD = 1.66		
LL	d1: 20–27	d1: F = 68, S = 32, SG = 2.69, PI = 13.26, OMC = 14.75, MDD = 1.77	13×6	Figure 11 (d1-d4)
	d2: 27–35	d2: F = 77, S = 23, SG = 2.69, PI = 16.53, OMC = 18.09, MDD = 1.67		
	d3: 35–42	d3: F = 87, S = 13, SG = 2.66, PI = 19.28, OMC = 19.78, MDD = 1.64		
	d4: 42–49	d4: F = 93, S = 07, SG = 2.65, PI = 22.74, OMC = 21.57, MDD = 1.60		
PI	e1: 7–12	e1: F = 65, S = 35, SG = 2.69, LL = 28.73, OMC = 14.29, MDD = 1.80	12×6	Figure 11 (e1-e4)
	e2: 16–19	e2: F = 80, S = 20, SG = 2.68, LL = 40.15, OMC = 18.61, MDD = 1.66		
	e3: 19–22	e3: F = 90, S = 10, SG = 2.65, LL = 46.78, OMC = 19.73, MDD = 1.65		
	e4: 22–25	e4: F = 92, S = 08, SG = 2.65, LL = 50.65, OMC = 21.14, MDD = 1.60		
OMC	f1: 13–16	f1: F = 72, S = 28, SG = 2.71, LL = 34.63, PI = 14.98, MDD = 1.76	12×6	Figure 11 (f1-f4)
	f2: 16–19	f2: F = 76, S = 24, SG = 2.68, LL = 36.90, PI = 15.46, MDD = 1.69		
	f3: 19–21	f3: F = 87, S = 13, SG = 2.65, LL = 43.79, PI = 18.88, MDD = 1.64		
	f4: 21–25	f4: F = 86, S = 14, SG = 2.68, LL = 47.48, PI = 19.91, MDD = 1.56		
MDD	g1: 1.5–1.58	g1: F = 92, S = 08, SG = 2.68, LL = 47.44, PI = 20.20, OMC = 22.36	10×6	Figure 11 (g1-g4)
	g2: 1.67–1.76	g2: F = 76, S = 24, SG = 2.69, LL = 36.67, PI = 15.36, OMC = 16.32		
	g3: 1.76–1.85	g3: F = 70, S = 30, SG = 2.71, LL = 33.26, PI = 14.30, OMC = 13.86		
	g4: 1.85–1.94	g4: F = 57, S = 43, SG = 2.68, LL = 28.64, PI = 11.60, OMC = 12.40		

Validation of optimal performance model

Literature validation

The literature study presents many computational models used for the hydraulic conductivity of soil. These published studies illustrate that researchers used soil textures, combined soil texture with PI, and combined soil texture with dg, Sg, Db, and WCs. For the first time, soil texture, LL, PI, and compaction parameters have been used to predict clayey soil's hydraulic conductivity and model's performance compared with published models, as shown in Table 13.

Cross validation This research introduces the six best architectural ANN models, one from each backpropagation algorithm, in predicting the hydraulic conductivity of clayey soil. These ANN models, i.e., LM_K15, BFG_K9, SCG_K10, GDM_K12, GD_K7, and GDA_K12 have been configured with 5 k-fold. The same models have been configured with ten k-fold for the cross-validation, and the computational costs have been computed and compared with five k-fold-based ANN models. The comparison of computational cost is presented in Table 14.

Table 14 shows that each ten k-fold-based model attains a higher computational cost than five k-fold-based models. The comparison of computational cost (program runs in MATLAB R2020a version with i3-2350 M @ 2.3 GHz, 4 GB RAM) reveals that model LM_K15 achieves the desired prediction at significantly less cost. Hence, model LM_K15 is identified as an optimal performance model in predicting the hydraulic conductivity of clayey soil.

External validation A model's generalizability is evaluated, and external validation is carried out to make sure the model isn't just overfitting the training set. Finding the most accurate model for predicting ground vibrations is made easier by the findings of external validation. Accuracy is the capacity of the model to correctly identify patients as having or not having the desired outcome. External validation checks for overfitting and guarantees that models are reliable. When a model is too tightly suited to the training data and does not generalize effectively to new data, it is said to overfit. By contrasting the model's performance on the training data with the test data, external validation can help to spot overfitting. The Golbraikh and Tropsha (2002) theory, which

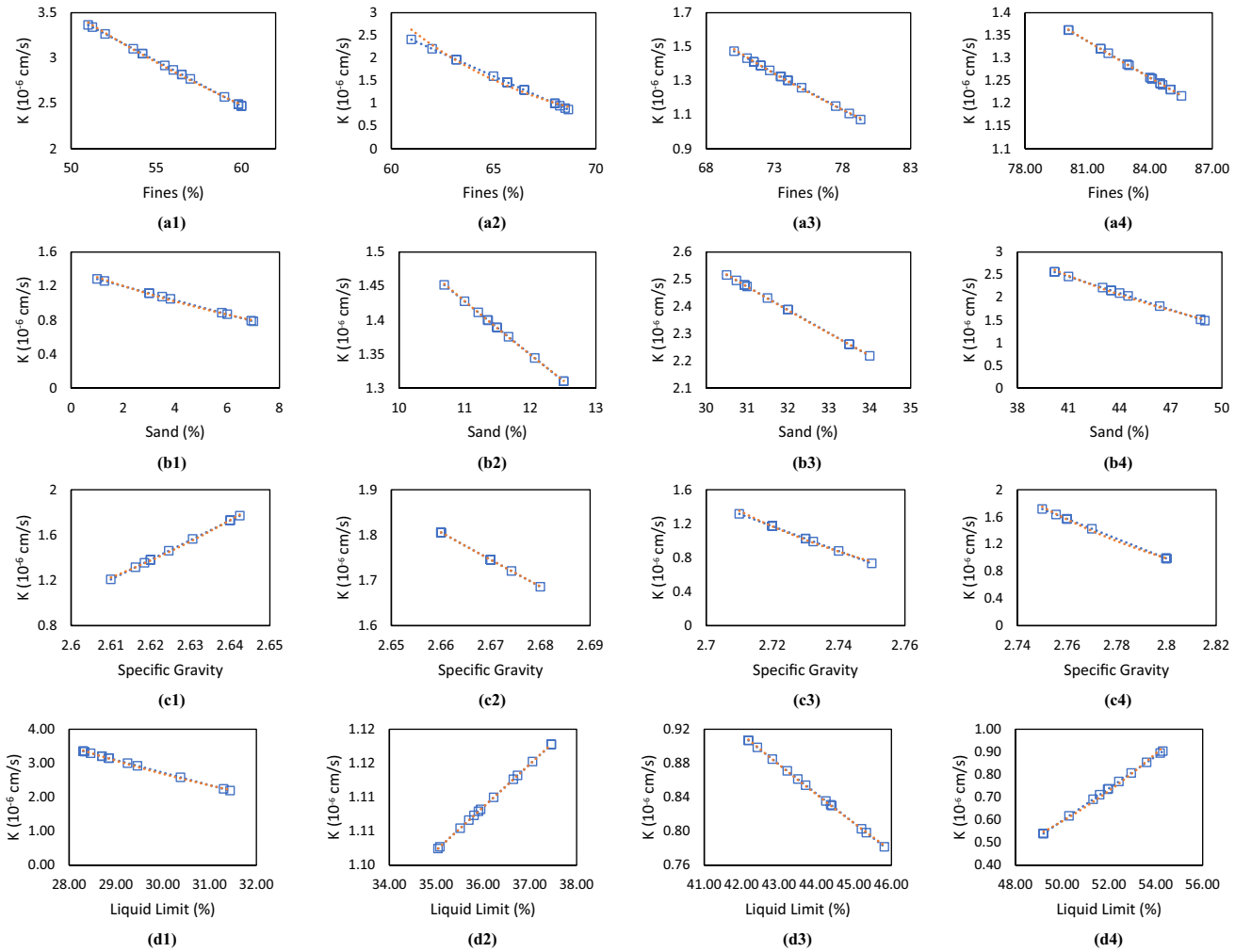

Fig. 11 Illustration of simulated hydraulic conductivity of clayey soil by varying (**a**) fine content, (**b**) sand content, (**c**) specific gravity, (**d**) liquid limit, (**e**) plasticity index, (**f**) optimum moisture content, and (**g**) maximum dry density

was proposed, is an accurate model in this investigation. Table 15 provides a summary of the theory's various mathematical expression-related aspects.

Where d_i denotes the experimental hydraulic conductivity and y_i denotes the predicted hydraulic conductivity, k and $k\prime$ represent the slopes of the predicted versus actual hydraulic conductivity and actual versus predicted hydraulic conductivity with respect to the origin. R_o^2 and $R\prime_o^2$ denotes the coefficients of determination of the predicted versus actual hydraulic conductivity and actual versus predicted hydraulic conductivity. m and n represent the factors for estimating the predictive power of the proposed models. The external validation results are presented in Table 16 for all proposed models in the training, validation, and testing phase. Table 16 demonstrates that model LM_K15 has attained excellent generalizability, showing superiority over all ANN models employed in this work.

Conclusions and summary

The hydraulic conductivity of clayey soil, an essential parameter for any Civil Engineering project, must be determined experimentally. The experimental procedure for determining the hydraulic conductivity of clayey soil is arduous and time-consuming. Therefore, the present work is motivated to replace the tedious laboratory procedures with computational models for predicting the hydraulic conductivity of soil. It is important to note that estimating hydraulic conductivity accurately and reliably can avoid the need for costly and time-consuming laboratory testing. For this aim, the hydraulic conductivity database of clayey soil has been compiled from published research. The database consists of the hydraulic conductivity of CH, CI, CL, OH, OL, OI, MH, OH, MI, and ML soils and is utilized to develop artificial neural network models. The following conclusions are mapped based on the novelty statements.

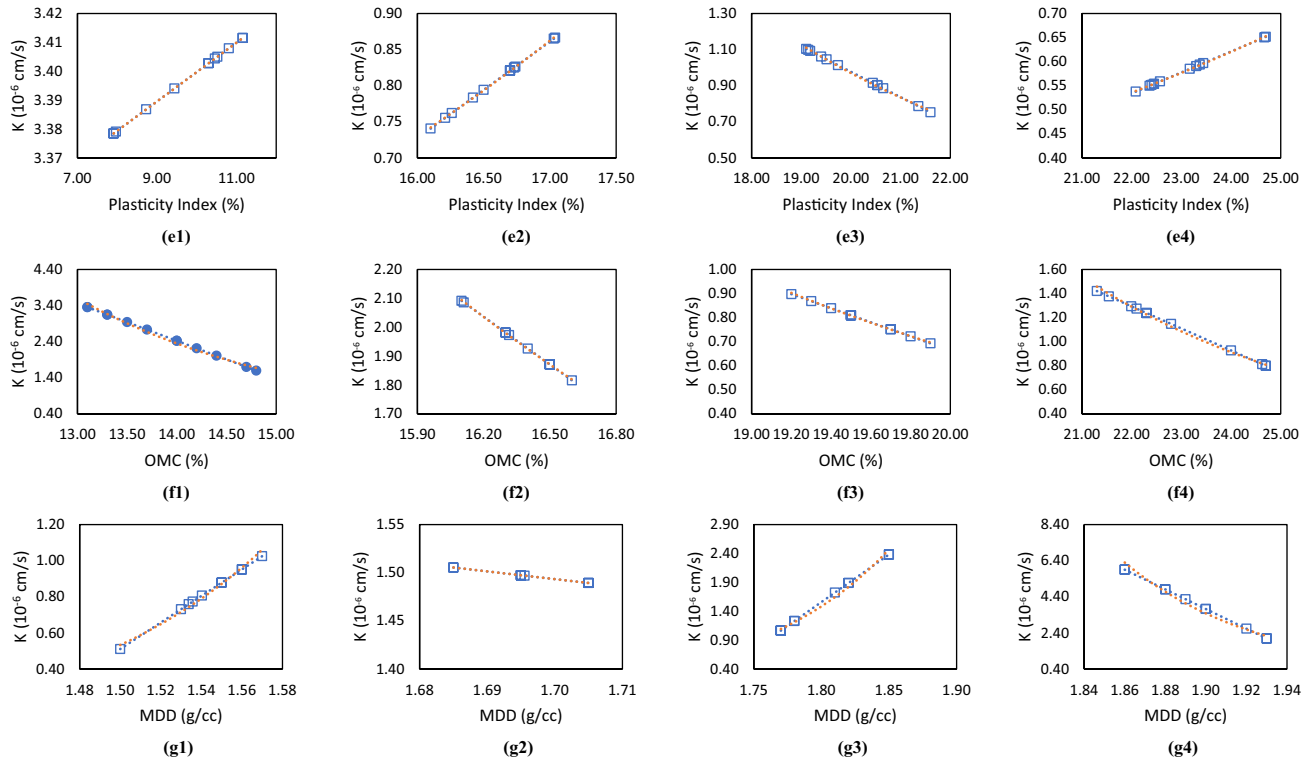

Fig. 11 (continued)

Table 13 Comparison of model LM_K15 with published models

S.No	References	Database	Input Variables	Approach	R Test
1	Zeitfogel et al. (2023)	15,641	S, M, C, OC	XGBoost	0.8000
2	Chandel et al. (2023)	165	D_{10}, D_{50}, P, C_U	FFNN	0.9711
3	Thakur et al. (2022)	56	D_{10}, D_{50}, P, C_U	GA_ANFIS	0.9661
4	Sihag et al. (2019a)	240	S, C, M	ANFIS_PSO	0.9816
5	**Present Case**	**104**	**S, C, M, LL, PI, OMC, MDD**	**LM_K15**	**0.9959**

*Bold values represent the optimum performance model

Table 14 Comparison of computational cost for cross-validation

Phase	Train		Valid		Test	
k-fold	**5**	**10**	**5**	**10**	**5**	**10**
LM_K15	**0.0254**	**0.0499**	**0.0219**	**0.0235**	**0.0212**	**0.0228**
BFG_K9	0.0101	0.0143	0.0103	0.0149	0.0135	0.0167
SCG_K10	0.0162	0.0199	0.0082	0.0115	0.0171	0.0193
GDM_K12	0.0237	0.0379	0.0189	0.0213	0.0103	0.0149
GD_K7	0.0061	0.0107	0.015	0.0187	0.0072	0.0111
GDA_K12	0.0182	0.0234	0.0139	0.0172	0.0179	0.0201

*Bold values present the optimal performance model

- The performance analysis reveals that the prediction performance and accuracy increase with neurons and hidden layers. It is also noted that the number of hidden layers increases the prediction accuracy compared to neurons. Still, the ANN model achieves a performance of over 85% by increasing hidden layers in moderate to problematic multicollinearity.

Table 15 Mathematical expression and condition of validation factors

Validation Parameters	Condition	Mathematical Expression
k	$0.85 < k < 1.15$	$k = \frac{\sum_{i=0}^{n}(d_i \times y_i)}{\sum_{i=0}^{n} y_i^2}$
k'	$0.85 < k\prime < 1.15$	$k\prime = \frac{\sum_{i=0}^{n}(d_i \times y_i)}{\sum_{i=0}^{n} d_i^2}$
R_o^2	Close to 1	$R_o^2 = 1 - \frac{\sum_{i=1}^{n} y_i^2 (1-k)^2}{\sum_{i=1}^{n}(y_i - \bar{y})}$
$R\prime_o^2$	Close to 1	$R\prime_o^2 = 1 - \frac{\sum_{i=1}^{n} d_i^2 (1-k\prime)^2}{\sum_{i=1}^{n}(d_i - \bar{d})}$
R_m	$R_m > 0.5$	$R_m = R^2 \times \left(1 - \sqrt{\lvert R^2 - R_o^2 \rvert}\right)$
$\lvert m \rvert$	$\lvert m \rvert < 0.1$	$m = \frac{R^2 - R_o^2}{R^2}$
$\lvert n \rvert$	$\lvert n \rvert < 0.1$	$n = \frac{R^2 - R\prime_o^2}{R^2}$

- It is noted that the Levenberg–Marquardt (LM) algorithm-based ANN model has achieved the highest performance in predicting the hydraulic conductivity of clayey soil. LM employs the notion of the neural neighborhood to enhance the behaviour of both memory and time limitations. Hence, this research suggests configuring the ANN models with the LM algorithm to solve geotechnical issues.
- Based on the VIF values, it is noted that the input variables, excluding specific gravity, have moderate to problematic multicollinearity. The impact of such multicollinearity has been observed on the accuracy of SCG, GDM, and GD backpropagation algorithm-based ANN models.
- The overall analysis of the ninety ANN models reveals that model LM_K15 is an optimal performance model in predicting the hydraulic conductivity of clayey soil. Model LM_K15 has achieved the highest testing performances, i.e., RMSE = 0.0487, a20 = 100, VAF = 99.16, R = 0.9959, IOA = 0.9364, and PI = 1.9348.

To conclude, this research introduces a highly capable of predicting the hydraulic conductivity of clayey soil neural network model. The concept developed in the present study may be implemented to assess the hydraulic conductivity of unsaturated soil. Also, the same configured neural network model may be used to solve the other geotechnical issues. This research is limited by determining the effect of hidden layers and neurons. This research may extend by drawing a relationship between epochs and the number of hidden layers. Also, the impact of the activation function may be analyzed in predicting the hydraulic conductivity of clayey soil. A comparative study for the configuration of non-optimized and optimized ANN models may be carried out in predicting the hydraulic conductivity of fine-grained soil. The present study may also be extended by implementing tanh, ReLU, leaky ReLU, PreLU, and ELU activation functions and comparing the results. For the first time, the effect of hidden layers and neurons has been studied on ANN models in predicting the hydraulic conductivity of soil.

Table 16 Result obtained from external validation

Model ID	Phase	k	k'	R_o^2	$R\prime_o^2$	R_m	$\lvert m \rvert$	$\lvert n \rvert$
LM_K15	**TRG**	**0.99**	**1.00**	**1.00**	**1.00**	**0.86**	**-0.02**	**-0.02**
	VDN	**0.94**	**1.03**	**0.99**	**1.00**	**0.73**	**-0.05**	**-0.06**
	TSG	**0.98**	**1.02**	**1.00**	**1.00**	**0.92**	**0.00**	**0.00**
BFG_K9	TRG	0.90	1.04	0.98	1.00	0.61	-0.11	-0.13
	VDN	0.84	1.13	0.96	0.97	0.74	-0.04	-0.05
	TSG	1.01	0.98	1.00	0.99	0.63	-0.10	-0.10
SCG_K10	TRG	0.90	1.06	0.98	0.99	0.67	-0.07	-0.09
	VDN	0.80	1.17	0.92	0.96	0.80	-0.02	-0.05
	TSG	0.96	1.04	0.98	0.98	0.78	-0.03	-0.04
GDM_K12	TRG	0.97	1.01	1.00	1.00	0.77	-0.04	-0.04
	VDN	0.88	1.10	0.98	0.99	0.77	-0.03	-0.04
	TSG	0.95	1.04	0.97	0.98	0.66	-0.08	-0.08
GD_K7	TRG	0.92	1.00	0.99	1.00	0.54	-0.16	-0.17
	VDN	0.85	1.13	0.97	0.97	0.80	-0.02	-0.03
	TSG	0.98	1.01	1.00	1.00	0.66	-0.09	-0.09
GDA_K12	TRG	1.00	0.98	1.00	1.00	0.73	-0.06	-0.06
	VDN	1.10	0.85	0.98	0.97	0.73	-0.05	-0.04
	TSG	0.98	1.01	1.00	1.00	0.71	-0.06	-0.07

*Bold values present the optimum performance model

Appendix I

Table 17 Performance metrics details for ANN models in training

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
1HL & 5N	LM_K1	0.7446	0.6117	0.8841	0.8944	72.12	0.4527	0.6896	0.7582	1.5063	0.4104	0.9054	0.2377	0.6436	0.5571	20.00	0.6782	0.5511
1HL & 10N	LM_K2	0.7020	0.5670	0.8661	1.0980	74.06	0.4197	0.7241	0.7888	1.7787	0.3647	0.8947	0.1716	0.5967	0.5253	27.50	0.7016	0.5196
1HL & 15N	LM_K3	0.5430	0.4376	0.9235	0.6340	83.77	0.3239	0.8349	1.1476	1.1652	0.2182	0.9230	-0.0704	0.4605	0.4063	28.75	0.7697	0.4019
2HL & 5N	LM_K4	0.6421	0.5277	0.8971	1.0795	78.43	0.3906	0.7691	0.9471	1.7354	0.3052	0.9111	0.1647	0.5553	0.4805	25.00	0.7224	0.4753
2HL & 10N	LM_K5	0.5363	0.4137	0.9272	0.6558	83.97	0.3062	0.8390	1.1632	1.2708	0.2129	0.9288	0.0364	0.4353	0.4013	41.25	0.7824	0.3969
2HL & 15N	LM_K6	0.5063	0.4124	0.9345	0.6527	85.81	0.3052	0.8565	1.2250	1.2144	0.1897	0.9291	-0.0528	0.4339	0.3788	28.75	0.7830	0.3747
3HL & 5N	LM_K7	0.5061	0.4055	0.9329	0.6785	85.73	0.3001	0.8566	1.2216	1.2822	0.1896	0.9290	0.0366	0.4267	0.3787	32.50	0.7867	0.3746
3HL & 10N	LM_K8	0.3952	0.3158	0.9582	0.4889	91.40	0.2337	0.9125	1.4369	1.1756	0.1156	0.9442	0.0516	0.3323	0.2957	45.00	0.8339	0.2926
3HL & 15N	LM_K9	0.3941	0.3130	0.9578	0.6384	91.45	0.2317	0.9130	1.4377	1.2079	0.1150	0.9429	-0.0506	0.3293	0.2949	38.75	0.8353	0.2917
4HL & 5N	LM_K10	0.4017	0.3253	0.9635	0.4417	91.56	0.2408	0.9097	1.4422	1.2036	0.1194	0.9470	0.1029	0.3423	0.3006	40.00	0.8289	0.2973
4HL & 10N	LM_K11	0.3496	0.2955	0.9652	0.6145	93.16	0.2187	0.9316	1.5137	1.3051	0.0904	0.9451	0.0050	0.3110	0.2616	43.75	0.8445	0.2587
4HL & 15N	LM_K12	0.3223	0.2709	0.9744	0.4640	94.23	0.2005	0.9418	1.5695	1.0526	0.0769	0.9530	-0.0283	0.2850	0.2412	47.50	0.8575	0.2386
5HL & 5N	LM_K13	0.2867	0.2369	0.9785	0.3457	95.40	0.1754	0.9540	1.6246	1.0046	0.0609	0.9581	0.0033	0.2493	0.2145	43.75	0.8754	0.2122
5HL & 10N	LM_K14	0.2470	0.2070	0.9832	0.3412	96.60	0.1532	0.9658	1.6857	1.0714	0.0452	0.9626	0.0195	0.2178	0.1848	56.25	0.8911	0.1828
5HL & 15N	**LM_K15**	**0.1658**	**0.1321**	**0.9924**	**0.1984**	**98.48**	**0.0978**	**0.9846**	**1.8038**	**0.9923**	**0.0204**	**0.9758**	**0.0175**	**0.1390**	**0.1241**	**68.75**	**0.9305**	**0.1227**
1HL & 5N	BFG_K1	0.8815	0.6832	0.7988	1.4047	59.24	0.5057	0.5650	0.3490	2.0360	0.5751	0.8761	0.2213	0.7190	0.6596	26.25	0.6405	0.6524
1HL & 10N	BFG_K2	0.7413	0.5840	0.8544	1.0943	70.54	0.4323	0.6923	0.6942	1.7301	0.4067	0.8941	0.1528	0.6146	0.5547	22.50	0.6927	0.5487
1HL & 15N	BFG_K3	0.6704	0.5252	0.9132	0.8392	77.38	0.3887	0.7484	0.9374	1.5318	0.3326	0.9231	0.2132	0.5526	0.5016	26.25	0.7237	0.4962
2HL & 5N	BFG_K4	0.6784	0.5574	0.8814	0.8688	74.41	0.4126	0.7424	0.8427	1.4472	0.3406	0.9058	0.0551	0.5865	0.5076	25.00	0.7067	0.5021
2HL & 10N	BFG_K5	0.6999	0.5613	0.8935	1.0404	74.74	0.4155	0.7257	0.8457	1.7128	0.3626	0.9135	0.1966	0.5907	0.5237	27.50	0.7047	0.5181
2HL & 15N	BFG_K6	0.5821	0.4587	0.9269	0.7564	83.08	0.3395	0.8103	1.1079	1.4904	0.2508	0.9269	0.1914	0.4826	0.4356	27.50	0.7587	0.4309

Table 17 (continued)

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
3HL & 5N	BFG_K7	0.4893	0.3981	0.9374	0.6675	87.02	0.2947	0.8659	1.2597	1.3662	0.1772	0.9305	0.0871	0.4189	0.3661	30.00	0.7905	0.3622
3HL & 10N	BFG_K8	0.5397	0.4434	0.9277	0.8624	84.45	0.3282	0.8369	1.1655	1.4812	0.2156	0.9242	0.1162	0.4666	0.4038	27.50	0.7667	0.3995
3HL & 15N	**BFG_K9**	**0.4233**	**0.3947**	**0.9394**	**0.6616**	**94.74**	**0.2921**	**0.8670**	**1.2661**	**1.3353**	**0.1203**	**0.9390**	**0.0343**	**0.4153**	**0.3392**	**43.75**	**0.7923**	**0.3287**
4HL & 5N	BFG_K10	0.5798	0.4267	0.9147	0.9051	82.59	0.3158	0.8118	1.0828	1.6726	0.2488	0.9268	0.1589	0.4490	0.4338	36.25	0.7755	0.4292
4HL & 10N	BFG_K11	0.6446	0.5005	0.8970	0.6772	78.13	0.3705	0.7674	0.9413	1.3624	0.3075	0.9149	0.1576	0.5267	0.4823	30.00	0.7367	0.4771
4HL & 15N	BFG_K12	0.6299	0.5051	0.8909	0.8150	79.11	0.3739	0.7778	0.9550	1.5458	0.2937	0.9043	0.1538	0.5315	0.4713	37.50	0.7342	0.4662
5HL & 5N	BFG_K13	0.7425	0.5564	0.8611	1.0732	71.57	0.4118	0.6913	0.7146	1.7684	0.4080	0.9036	0.2085	0.5855	0.5556	30.00	0.7073	0.5496
5HL & 10N	BFG_K14	0.7577	0.5930	0.8665	0.9150	71.12	0.4389	0.6786	0.7044	1.6122	0.4249	0.9017	0.2412	0.6240	0.5669	25.00	0.6880	0.5608
5HL & 15N	BFG_K15	0.7490	0.6362	0.8555	1.1789	71.16	0.4709	0.6859	0.6946	1.8275	0.4152	0.8883	0.2142	0.6695	0.5604	21.25	0.6653	0.5544
HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
1HL & 5N	SCG_K1	0.6668	0.5285	0.8819	0.9100	75.74	0.3912	0.7510	0.8684	1.4906	0.3291	0.9034	0.1070	0.5561	0.4990	30.00	0.7219	0.4936
1HL & 10N	SCG_K2	0.7179	0.5763	0.8630	0.9723	72.13	0.4266	0.7115	0.7481	1.6092	0.3814	0.8962	0.1325	0.6065	0.5372	26.25	0.6968	0.5314
1HL & 15N	SCG_K3	0.6794	0.5508	0.8851	1.0644	76.97	0.4077	0.7415	0.8738	1.7627	0.3417	0.9029	0.2245	0.5796	0.5084	28.75	0.7102	0.5029
2HL & 5N	SCG_K4	0.5354	0.4207	0.9251	0.6694	84.02	0.3114	0.8395	1.1606	1.2880	0.2122	0.9266	0.0345	0.4427	0.4006	36.25	0.7786	0.3963
2HL & 10N	SCG_K5	0.5052	0.4166	0.9306	0.6604	86.21	0.3084	0.8571	1.2228	1.3582	0.1889	0.9234	0.0940	0.4384	0.3780	31.25	0.7808	0.3739
2HL & 15N	SCG_K6	0.6222	0.5029	0.8891	0.9292	78.35	0.3722	0.7832	0.9517	1.5245	0.2866	0.9039	0.0200	0.5292	0.4656	27.50	0.7354	0.4606
3HL & 5N	SCG_K7	0.5594	0.4664	0.9151	0.9351	82.58	0.3452	0.8248	1.1038	1.5373	0.2316	0.9172	0.0431	0.4908	0.4186	28.75	0.7546	0.4141
3HL & 10N	SCG_K8	0.5322	0.4441	0.9303	0.7615	84.70	0.3287	0.8414	1.1802	1.4184	0.2097	0.9265	0.0998	0.4674	0.3983	26.25	0.7663	0.3940
3HL & 15N	SCG_K9	0.4839	0.3845	0.9443	0.7023	88.31	0.2846	0.8689	1.2909	1.4128	0.1733	0.9344	0.1593	0.4046	0.3621	37.50	0.7977	0.3582
4HL & 5N	**SCG_K10**	**0.4386**	**0.3651**	**0.9556**	**0.5843**	**90.52**	**0.2703**	**0.8923**	**1.3797**	**1.1227**	**0.1424**	**0.9386**	**0.0152**	**0.3842**	**0.3282**	**46.25**	**0.8079**	**0.3246**
4HL & 10N	SCG_K11	0.5935	0.4851	0.9098	0.8737	80.42	0.3590	0.8028	1.0383	1.4964	0.2607	0.9178	0.0506	0.5104	0.4441	30.00	0.7448	0.4393

Table 17 (continued)

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
4HL & 15N	SCG_K12	0.5251	0.4206	0.9276	0.7095	84.72	0.3114	0.8456	1.1825	1.3195	0.2041	0.9267	0.0531	0.4426	0.3929	33.75	0.7787	0.3887
5HL & 5N	SCG_K13	0.5575	0.4317	0.9109	0.7062	82.69	0.3195	0.8260	1.0990	1.3776	0.2301	0.9169	0.0407	0.4543	0.4172	28.75	0.7729	0.4127
5HL & 10N	SCG_K14	0.5984	0.4755	0.9105	0.8733	80.66	0.3519	0.7995	1.0372	1.4874	0.2651	0.9199	0.1122	0.5003	0.4478	36.25	0.7498	0.4429
5HL & 15N	SCG_K15	0.6374	0.4972	0.8937	0.7980	77.90	0.3680	0.7725	0.9403	1.4652	0.3007	0.9113	0.1078	0.5232	0.4769	30.00	0.7384	0.4718
1HL & 5N	GDM_K1	0.7100	0.5896	0.8773	1.0102	74.67	0.4364	0.7177	0.8063	1.6831	0.3731	0.8996	0.2273	0.6204	0.5313	27.50	0.6898	0.5255
1HL & 10N	GDM_K2	0.6161	0.4809	0.9092	0.7567	79.52	0.3560	0.7874	1.0057	1.3739	0.2810	0.9201	0.1175	0.5061	0.4610	23.75	0.7470	0.4561
1HL & 15N	GDM_K3	0.6085	0.4718	0.9126	0.7675	80.49	0.3492	0.7927	1.0293	1.4347	0.2740	0.9220	0.1477	0.4964	0.4553	32.50	0.7518	0.4504
2HL & 5N	GDM_K4	0.6332	0.5059	0.9005	0.7650	78.46	0.3745	0.7755	0.9624	1.3978	0.2967	0.9130	0.1274	0.5323	0.4738	26.25	0.7338	0.4687
2HL & 10N	GDM_K5	0.5784	0.4566	0.9262	0.6505	81.61	0.3380	0.8127	1.0955	1.2257	0.2476	0.9276	0.0776	0.4805	0.4328	30.00	0.7597	0.4281
2HL & 15N	GDM_K6	0.5501	0.4385	0.9183	0.7067	83.11	0.3246	0.8306	1.1242	1.3055	0.2240	0.9216	0.0308	0.4614	0.4116	27.50	0.7693	0.4072
3HL & 5N	GDM_K7	0.5047	0.4025	0.9308	0.7326	85.83	0.2979	0.8574	1.2199	1.3611	0.1885	0.9281	0.0394	0.4235	0.3776	30.00	0.7882	0.3736
3HL & 10N	GDM_K8	0.5250	0.4352	0.9219	0.9171	84.75	0.3221	0.8457	1.1724	1.5518	0.2040	0.9181	0.0565	0.4580	0.3929	27.50	0.7710	0.3886
3HL & 15N	GDM_K9	0.4145	0.3180	0.9616	0.5017	90.73	0.2354	0.9038	1.4174	1.2463	0.1272	0.9488	0.0788	0.3346	0.3101	42.50	0.8327	0.3068
4HL & 5N	GDM_K10	0.4335	0.3572	0.9496	0.5828	89.76	0.2644	0.8948	1.3658	1.2931	0.1391	0.9365	0.0704	0.3759	0.3244	35.00	0.8121	0.3209
4HL & 10N	GDM_K11	0.3720	0.3010	0.9659	0.4117	92.25	0.2228	0.9225	1.4834	1.0272	0.1025	0.9485	-0.0043	0.3167	0.2784	42.50	0.8417	0.2754
4HL & 15N	**GDM_K12**	**0.2790**	**0.2274**	**0.9797**	**0.3413**	**95.65**	**0.1683**	**0.9564**	**1.6372**	**1.0108**	**0.0576**	**0.9605**	**0.0145**	**0.2393**	**0.2088**	**57.50**	**0.8804**	**0.2065**
5HL & 5N	GDM_K13	0.2908	0.2384	0.9768	0.4317	95.31	0.1765	0.9526	1.6163	1.1817	0.0626	0.9574	0.0293	0.2509	0.2176	51.25	0.8745	0.2153
5HL & 10N	GDM_K14	0.3334	0.2689	0.9699	0.4714	93.80	0.1990	0.9378	1.5453	1.1318	0.0823	0.9519	-0.0206	0.2830	0.2494	45.00	0.8585	0.2468
5HL & 15N	GDM_K15	0.3494	0.2848	0.9680	0.4156	93.16	0.2108	0.9316	1.5194	1.1009	0.0904	0.9502	0.0015	0.2997	0.2614	47.50	0.8501	0.2586

Table 17 (continued)

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
1HL & 5N	GD_K1	0.7776	0.6304	0.8582	1.3000	71.18	0.4666	0.6615	0.6708	1.9976	0.4475	0.8946	0.2996	0.6634	0.5818	27.50	0.6683	0.5755
1HL & 10N	GD_K2	0.8337	0.6437	0.8307	1.0051	61.60	0.4765	0.6109	0.4724	1.5671	0.5144	0.8940	0.0959	0.6774	0.6238	20.00	0.6613	0.6171
1HL & 15N	GD_K3	0.7560	0.6086	0.8700	0.9770	69.28	0.4505	0.6800	0.6937	1.5596	0.4231	0.9045	0.1514	0.6404	0.5657	23.75	0.6798	0.5596
2HL & 5N	GD_K4	0.6287	0.5140	0.9018	0.8881	79.42	0.3804	0.7787	0.9788	1.6019	0.2926	0.9124	0.1667	0.5408	0.4705	32.50	0.7296	0.4654
2HL & 10N	GD_K5	0.6199	0.5013	0.8932	0.9358	78.90	0.3711	0.7848	0.9670	1.6020	0.2845	0.9095	0.0866	0.5276	0.4639	30.00	0.7362	0.4589
2HL & 15N	GD_K6	0.6175	0.5050	0.8925	0.8736	78.69	0.3738	0.7865	0.9661	1.4620	0.2822	0.9062	0.0280	0.5314	0.4621	25.00	0.7343	0.4571
3HL & 5N	**GD_K7**	**0.5319**	**0.4301**	**0.9240**	**0.6466**	**85.02**	**0.3183**	**0.8416**	**1.1720**	**1.3788**	**0.2094**	**0.9237**	**0.0124**	**0.4525**	**0.3980**	**35.00**	**0.7737**	**0.3937**
3HL & 10N	GD_K8	0.6272	0.4912	0.9012	0.7815	79.64	0.3636	0.7798	0.9814	1.5246	0.2912	0.9152	0.1724	0.5169	0.4693	28.75	0.7416	0.4642
3HL & 15N	GD_K9	0.6235	0.5029	0.8963	1.0366	78.41	0.3723	0.7824	0.9640	1.6693	0.2877	0.9126	0.0556	0.5292	0.4665	26.25	0.7354	0.4615
4HL & 5N	GD_ K10	0.6263	0.4757	0.9076	0.7673	78.88	0.3521	0.7804	0.9863	1.4208	0.2903	0.9228	0.1224	0.5006	0.4686	31.25	0.7497	0.4636
4HL & 10N	GD_ K11	0.7771	0.6156	0.8870	1.1775	75.53	0.4556	0.6619	0.7650	1.9862	0.4469	0.9045	0.4084	0.6478	0.5815	28.75	0.6761	0.5752
4HL & 15N	GD_ K12	0.7891	0.6439	0.8593	0.9559	67.00	0.4766	0.6514	0.6194	1.5603	0.4608	0.8978	0.1821	0.6775	0.5904	18.75	0.6612	0.5841
5HL & 5N	GD_ K13	0.7275	0.6002	0.8596	0.9846	71.81	0.4442	0.7037	0.7296	1.6006	0.3918	0.8916	0.1607	0.6316	0.5444	21.25	0.6842	0.5385
5HL & 10N	GD_ K14	0.8325	0.6713	0.8409	1.0809	62.22	0.4969	0.6120	0.4968	1.6342	0.5130	0.8951	0.1351	0.7064	0.6229	17.50	0.6468	0.6162
5HL & 15N	GD_ K15	0.7122	0.5724	0.8900	0.9119	73.90	0.4237	0.7160	0.8189	1.5814	0.3754	0.9115	0.2026	0.6023	0.5329	25.00	0.6988	0.5271
1HL & 5N	GDA_ K1	0.6795	0.5403	0.8715	1.0438	74.79	0.3999	0.7415	0.8278	1.6597	0.3418	0.9007	0.1070	0.5685	0.5085	30.00	0.7157	0.5030
1HL & 10N	GDA_ K2	0.6774	0.5541	0.8757	0.8082	75.15	0.4102	0.7431	0.8409	1.4440	0.3397	0.9003	0.1226	0.5831	0.5069	30.00	0.7084	0.5014
1HL & 15N	GDA_ K3	0.5876	0.4776	0.9100	0.9936	81.24	0.3535	0.8067	1.0530	1.6114	0.2556	0.9162	0.1013	0.5026	0.4397	33.75	0.7487	0.4349
2HL & 5N	GDA_ K4	0.5823	0.4614	0.9145	1.0188	81.29	0.3416	0.8101	1.0669	1.6128	0.2510	0.9234	0.0700	0.4856	0.4357	30.00	0.7572	0.4310
2HL & 10N	GDA_ K5	0.5473	0.4262	0.9200	0.7454	83.73	0.3155	0.8323	1.1365	1.3937	0.2217	0.9241	0.0942	0.4485	0.4095	35.00	0.7758	0.4051

Table 17 (continued)

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
2HL & 15N	GDA_K6	0.4516	0.3558	0.9448	0.5982	88.58	0.2634	0.8858	1.3268	1.2190	0.1510	0.9370	-0.0080	0.3744	0.3379	35.00	0.8128	0.3343
3HL & 5N	GDA_K7	0.4364	0.3579	0.9458	0.6043	89.34	0.2649	0.8934	1.3515	1.2537	0.1409	0.9329	-0.0097	0.3766	0.3265	40.00	0.8117	0.3230
3HL & 10N	GDA_K8	0.4133	0.3320	0.9574	0.5049	90.46	0.2457	0.9044	1.4079	1.1746	0.1264	0.9435	0.0191	0.3493	0.3093	40.00	0.8253	0.3059
3HL & 15N	GDA_K9	0.4215	0.3325	0.9491	0.6278	90.08	0.2461	0.9005	1.3802	1.2953	0.1315	0.9362	-0.0243	0.3499	0.3154	37.50	0.8251	0.3120
4HL & 5N	GDA_K10	0.3842	0.3112	0.9635	0.5302	92.71	0.2303	0.9174	1.4712	1.2763	0.1092	0.9443	0.1317	0.3275	0.2875	40.00	0.8363	0.2844
4HL & 10N	GDA_K11	0.3334	0.2738	0.9706	0.4527	93.78	0.2026	0.9378	1.5464	1.1231	0.0823	0.9514	0.0088	0.2881	0.2495	45.00	0.8560	0.2468
4HL & 15N	**GDA_K12**	**0.3088**	**0.2515**	**0.9730**	**0.3451**	**94.66**	**0.1862**	**0.9466**	**1.5846**	**1.0411**	**0.0706**	**0.9530**	**-0.0035**	**0.2647**	**0.2310**	**48.75**	**0.8677**	**0.2285**
5HL & 5N	GDA_K13	0.3505	0.2873	0.9699	0.5118	93.24	0.2127	0.9312	1.5227	1.1265	0.0909	0.9506	0.0471	0.3024	0.2623	46.25	0.8488	0.2595
5HL & 10N	GDA_K14	0.3910	0.3095	0.9591	0.5453	91.74	0.2291	0.9144	1.4462	1.2793	0.1131	0.9447	0.0725	0.3257	0.2925	45.00	0.8372	0.2894
5HL & 15N	GDA_K15	0.5087	0.4268	0.9280	0.9037	85.73	0.3159	0.8551	1.2098	1.5365	0.1916	0.9235	0.0624	0.4492	0.3807	28.75	0.7754	0.3765

*Bold values correspond to the best architectural model

Table 18 Performance metrics details for ANN models in validation

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
1HL & 5N	LM_K1	1.0046	0.7354	0.8991	1.2923	76.47	0.4672	0.7607	0.5684	1.8878	0.6411	0.9255	-0.1287	0.5792	0.4891	33.33	0.7104	0.6382
1HL & 10N	LM_K2	1.1622	0.9217	0.8768	1.5390	69.44	0.5855	0.6798	0.3010	2.1638	0.8580	0.9469	0.2484	0.7260	0.5659	33.33	0.6370	0.7383
1HL & 15N	LM_K3	1.0302	0.7711	0.8778	0.8232	74.85	0.4898	0.7484	0.4889	1.2887	0.6742	0.9454	0.0185	0.6074	0.5016	25.00	0.6963	0.6545
2HL & 5N	LM_K4	0.9339	0.7043	0.8996	1.4539	80.91	0.4474	0.7932	0.6845	2.2203	0.5540	0.9486	0.2589	0.5547	0.4547	33.33	0.7226	0.5933
2HL & 10N	LM_K5	1.1258	1.0122	0.9080	1.5715	76.50	0.6430	0.6995	0.4637	2.3081	0.8051	0.9447	0.5254	0.7973	0.5481	16.67	0.6014	0.7152
2HL & 15N	LM_K6	1.2578	1.0606	0.9602	1.8459	89.16	0.6737	0.6250	0.5558	2.8459	1.0050	0.9448	1.0606	0.8354	0.6124	25.00	0.5823	0.7990
3HL & 5N	LM_K7	0.8612	0.6968	0.9114	0.9581	82.50	0.4427	0.8242	0.7944	1.3212	0.4712	0.9499	-0.0583	0.5489	0.4193	25.00	0.7256	0.5471
3HL & 10N	LM_K8	0.8371	0.6614	0.9485	1.0137	86.85	0.4202	0.8339	0.9309	1.6330	0.4452	0.9606	0.3820	0.5210	0.4076	25.00	0.7395	0.5318
3HL & 15N	LM_K9	0.5725	0.4411	0.9637	0.7060	92.86	0.2802	0.9223	1.2848	1.3301	0.2082	0.9683	0.1631	0.3474	0.2788	41.67	0.8263	0.3637
4HL & 5N	LM_K10	0.7925	0.5715	0.9470	0.7688	87.07	0.3630	0.8511	0.9750	1.5285	0.3989	0.9649	0.2877	0.4501	0.3858	50.00	0.7749	0.5034
4HL & 10N	LM_K11	0.9894	0.7640	0.8931	1.1684	79.00	0.4853	0.7680	0.5982	1.7338	0.6218	0.9452	0.3049	0.6018	0.4817	33.33	0.6991	0.6285
4HL & 15N	LM_K12	1.0239	0.8987	0.9163	1.2619	83.78	0.5709	0.7515	0.6535	1.9631	0.6660	0.9368	0.6034	0.7079	0.4985	8.33	0.6460	0.6505
5HL & 5N	LM_K13	0.8714	0.6793	0.9630	1.6258	92.65	0.4315	0.8200	0.9824	2.6057	0.4824	0.9573	0.6702	0.5350	0.4243	41.67	0.7325	0.5536
5HL & 10N	LM_K14	0.9690	0.7840	0.8953	1.2861	79.73	0.4981	0.7774	0.6298	2.0428	0.5965	0.9339	0.2898	0.6176	0.4718	16.67	0.6912	0.6156
5HL & 15N	**LM_K15**	**0.4928**	**0.3212**	**0.9719**	**0.4536**	**96.89**	**0.2168**	**0.9668**	**1.3859**	**1.1231**	**0.1202**	**0.9791**	**0.0227**	**0.2332**	**0.2601**	**72.00**	**0.8834**	**0.3258**
1HL & 5N	BFG_K1	0.9731	0.6705	0.9235	1.6664	83.83	0.4259	0.7755	0.7181	2.4531	0.6016	0.9570	0.5149	0.5281	0.4738	25.00	0.7360	0.6182
1HL & 10N	BFG_K2	1.3531	1.1910	0.9014	2.0951	81.23	0.7566	0.5660	0.2717	2.9134	1.1630	0.9210	1.0193	0.9381	0.6588	16.67	0.5310	0.8596
1HL & 15N	BFG_K3	1.0598	0.8567	0.8567	0.9349	73.38	0.5442	0.7338	0.4079	1.4508	0.7134	0.9266	0.0155	0.6748	0.5160	0.00	0.6626	0.6732
2HL & 5N	BFG_K4	0.7847	0.6188	0.9281	0.7837	86.09	0.3931	0.8540	0.9375	1.3345	0.3912	0.9528	-0.1706	0.4874	0.3821	25.00	0.7563	0.4985

Table 18 (continued)

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
2HL & 10N	BFG_K5	0.9867	0.7913	0.9090	1.3627	81.63	0.5027	0.7692	0.6559	2.1910	0.6185	0.9489	0.4456	0.6233	0.4804	25.00	0.6884	0.6268
2HL & 15N	BFG_K6	1.1072	0.9138	0.8940	1.0800	72.64	0.5805	0.7094	0.4184	1.6844	0.7787	0.9472	0.2675	0.7197	0.5391	8.33	0.6401	0.7033
3HL & 5N	BFG_K7	0.7800	0.6476	0.9344	1.0030	86.09	0.4114	0.8558	0.9540	1.6176	0.3865	0.9582	0.1476	0.5101	0.3798	25.00	0.7450	0.4955
3HL & 10N	BFG_K8	0.7517	0.6550	0.9408	1.1225	88.51	0.4161	0.8660	1.0186	1.8188	0.3590	0.9535	0.2837	0.5159	0.3660	16.67	0.7420	0.4775
3HL & 15N	**BFG_K9**	**0.7428**	**0.5213**	**0.9613**	**0.7589**	**89.83**	**0.3295**	**0.8692**	**1.0595**	**1.3080**	**0.3505**	**0.9642**	**0.1959**	**0.4289**	**0.3617**	**56.67**	**0.7955**	**0.4719**
4HL & 5N	BFG_K10	0.8204	0.5796	0.9357	1.0962	86.42	0.3682	0.8404	0.9194	1.8250	0.4276	0.9617	0.3169	0.4565	0.3995	50.00	0.7717	0.5212
4HL & 10N	BFG_K11	1.3219	1.0213	0.8901	1.6288	76.47	0.6488	0.5857	0.2351	2.4057	1.1101	0.9362	0.8689	0.8044	0.6436	16.67	0.5978	0.8398
4HL & 15N	BFG_K12	0.8253	0.5568	0.9333	0.7673	86.42	0.3537	0.8385	0.9100	1.5612	0.4327	0.9625	0.3291	0.4385	0.4018	33.33	0.7807	0.5243
5HL & 5N	BFG_K13	1.1461	0.8992	0.8870	1.2760	74.38	0.5712	0.6886	0.3846	1.7133	0.8344	0.9415	0.4826	0.7083	0.5580	25.00	0.6459	0.7280
5HL & 10N	BFG_K14	1.0724	0.8394	0.9089	1.6998	81.64	0.5332	0.7274	0.5700	2.3979	0.7306	0.9451	0.6127	0.6612	0.5222	16.67	0.6694	0.6813
5HL & 15N	BFG_K15	1.2902	1.0503	0.9140	1.5908	80.12	0.6672	0.6054	0.3464	2.4390	1.0574	0.9394	0.9089	0.8273	0.6282	16.67	0.5863	0.8196
HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
1HL & 5N	SCG_K1	1.1564	0.9109	0.8611	1.8482	72.30	0.5787	0.6830	0.3082	2.4821	0.8495	0.9395	0.4110	0.7175	0.5630	25.00	0.6412	0.7346
1HL & 10N	SCG_K2	1.2448	1.0794	0.8892	1.6106	69.83	0.6857	0.6327	0.2441	2.3392	0.9843	0.9433	0.5259	0.8502	0.6061	16.67	0.5749	0.7907
1HL & 15N	SCG_K3	1.4939	1.1981	0.7343	2.9969	49.21	0.7611	0.4709	0.4625	3.4054	1.4177	0.9097	0.2990	0.9437	0.7274	16.67	0.5282	0.9490
2HL & 5N	SCG_K4	1.1839	0.8877	0.9326	1.3610	78.85	0.5639	0.6677	0.4745	2.1382	0.8903	0.9527	0.7137	0.6992	0.5764	33.33	0.6504	0.7521
2HL & 10N	SCG_K5	1.2990	1.0290	0.9025	1.2306	66.68	0.6537	0.6000	0.1823	1.9113	1.0719	0.9477	0.5307	0.8105	0.6325	25.00	0.5947	0.8252
2HL & 15N	SCG_K6	0.9953	0.7901	0.8765	1.2248	76.55	0.5019	0.7652	0.5384	1.8648	0.6293	0.9353	-0.0395	0.6224	0.4846	16.67	0.6888	0.6323
3HL & 5N	SCG_K7	1.2555	1.0960	0.8697	1.2641	62.78	0.6962	0.6263	0.1287	1.3802	1.0013	0.9387	0.0798	0.8633	0.6113	8.33	0.5684	0.7975
3HL & 10N	SCG_K8	1.3456	1.1125	0.9129	1.7510	67.76	0.7067	0.5708	0.1655	2.2917	1.1502	0.9457	0.6715	0.8763	0.6552	0.00	0.5619	0.8548

Table 18 (continued)

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
3HL & 15N	SCG_K9	1.1769	1.0185	0.9120	1.6483	83.13	0.6470	0.6716	0.4861	2.5125	0.8800	0.9317	0.8207	0.8023	0.5730	16.67	0.5989	0.7477
4HL & 5N	**SCG_K10**	**0.8138**	**0.6906**	**0.9532**	**0.6828**	**88.91**	**0.4502**	**0.8320**	**0.8838**	**1.1399**	**0.4305**	**0.9541**	**0.6062**	**0.5227**	**0.4045**	**36.67**	**0.7689**	**0.5181**
4HL & 10N	SCG_K11	0.8499	0.7193	0.9242	0.8751	85.41	0.4569	0.8288	0.8583	1.1723	0.4589	0.9441	-0.3269	0.5666	0.4138	16.67	0.7167	0.5399
4HL & 15N	SCG_K12	1.3235	1.1157	0.9175	1.9968	83.67	0.7088	0.5848	0.3550	2.9510	1.1127	0.9320	1.0309	0.8788	0.6444	25.00	0.5606	0.8407
5HL & 5N	SCG_K13	0.8680	0.7738	0.9299	1.3026	82.97	0.4916	0.8214	0.8265	2.0076	0.4786	0.9306	0.1875	0.6095	0.4226	8.33	0.6952	0.5514
5HL & 10N	SCG_K14	1.1451	0.9172	0.8888	1.3506	74.67	0.5826	0.6892	0.3916	2.0703	0.8329	0.9442	0.4925	0.7224	0.5575	25.00	0.6388	0.7274
5HL & 15N	SCG_K15	1.2338	0.9603	0.8500	2.0491	69.59	0.6100	0.6391	0.1845	2.8935	0.9671	0.9099	0.4894	0.7564	0.6007	25.00	0.6218	0.7838
1HL & 5N	GDM_K1	1.1211	0.8958	0.8936	1.3358	79.45	0.5691	0.7021	0.4720	2.1848	0.7984	0.9381	0.6246	0.7056	0.5458	25.00	0.6472	0.7122
1HL & 10N	GDM_K2	1.2738	1.1871	0.8703	1.6795	75.74	0.7541	0.6153	0.2410	2.4895	1.0308	0.9106	0.7740	0.9350	0.6202	8.33	0.5325	0.8092
1HL & 15N	GDM_K3	0.7564	0.6205	0.9374	0.9468	87.76	0.3941	0.8644	1.0000	1.7285	0.3635	0.9571	0.2364	0.4887	0.3683	16.67	0.7556	0.4805
2HL & 5N	GDM_K4	1.3219	1.0218	0.7821	2.0231	60.99	0.6491	0.5858	-0.1003	2.5344	1.1100	0.9075	0.3189	0.8048	0.6436	16.67	0.5976	0.8397
2HL & 10N	GDM_K5	1.2387	0.9624	0.8104	1.2631	63.64	0.6114	0.6362	0.0544	1.8561	0.9748	0.8983	-0.0295	0.7580	0.6031	16.67	0.6210	0.7869
2HL & 15N	GDM_K6	0.9376	0.7754	0.9273	1.6842	80.78	0.4925	0.7916	0.7300	2.2294	0.5585	0.9560	0.2616	0.6107	0.4565	33.33	0.6946	0.5956
3HL & 5N	GDM_K7	0.9017	0.7337	0.9086	1.4141	81.06	0.4661	0.8073	0.7345	1.8549	0.5165	0.9505	0.1177	0.5779	0.4390	16.67	0.7110	0.5728
3HL & 10N	GDM_K8	1.0007	0.7963	0.9223	1.7839	84.28	0.5059	0.7626	0.6929	2.6378	0.6361	0.9504	0.5818	0.6272	0.4872	16.67	0.6864	0.6357
3HL & 15N	GDM_K9	0.9565	0.8284	0.9011	1.8068	80.35	0.5262	0.7831	0.6590	2.2231	0.5812	0.9427	0.2933	0.6525	0.4657	16.67	0.6738	0.6076
4HL & 5N	GDM_K10	0.9963	0.7786	0.9416	1.3312	88.64	0.4946	0.7647	0.7767	2.2912	0.6305	0.9467	0.7166	0.6133	0.4851	8.33	0.6933	0.6329
4HL & 10N	GDM_K11	0.8122	0.7053	0.9339	1.1013	85.26	0.4480	0.8436	0.9126	1.7258	0.4191	0.9562	0.1948	0.5555	0.3955	33.33	0.7222	0.5160
4HL & 15N	**GDM_K12**	**0.5344**	**0.3487**	**0.9704**	**0.5063**	**94.08**	**0.2215**	**0.9323**	**1.3482**	**1.2347**	**0.1814**	**0.9759**	**0.1891**	**0.2747**	**0.2602**	**50.00**	**0.8627**	**0.3395**
5HL & 5N	GDM_K13	0.7722	0.5862	0.9370	1.4204	87.74	0.3724	0.8586	0.9832	2.1712	0.3788	0.9580	0.2815	0.4617	0.3760	33.33	0.7691	0.4905

Table 18 (continued)

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
5HL & 10N	GDM_K14	1.1156	0.8855	0.8509	1.4560	72.25	0.5625	0.7050	0.3310	2.1074	0.7906	0.9230	0.2717	0.6975	0.5432	25.00	0.6513	0.7087
5HL & 15N	GDM_K15	1.2931	1.1943	0.8178	1.5615	66.54	0.7587	0.6036	0.0412	2.2910	1.0622	0.8939	0.5107	0.9407	0.6296	0.00	0.5296	0.8215
HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
1HL & 5N	GD_K1	1.2321	1.0795	0.8726	2.2060	67.95	0.6858	0.6401	0.2087	2.5865	0.9644	0.9378	0.4075	0.8503	0.5999	16.67	0.5749	0.7827
1HL & 10N	GD_K2	1.2365	1.1026	0.8079	2.0379	65.15	0.7004	0.6375	0.0676	2.5767	0.9713	0.9104	0.2422	0.8685	0.6020	8.33	0.5658	0.7855
1HL & 15N	GD_K3	1.3693	1.0796	0.8170	1.2178	63.01	0.6858	0.5555	-0.0717	1.7243	1.1910	0.9224	0.5609	0.8503	0.6667	16.67	0.5748	0.8698
2HL & 5N	GD_K4	1.4495	1.2233	0.8348	1.9382	69.60	0.7771	0.5019	-0.0565	2.7975	1.3347	0.9037	0.9049	0.9635	0.7057	0.00	0.5182	0.9208
2HL & 10N	GD_K5	1.0967	0.9466	0.9291	1.3705	73.67	0.6013	0.7149	0.5032	1.7873	0.7640	0.9517	0.3032	0.7456	0.5340	0.00	0.6272	0.6967
2HL & 15N	GD_K6	1.0235	0.8616	0.8942	1.5830	79.24	0.5474	0.7516	0.5685	2.2469	0.6655	0.9417	0.4145	0.6787	0.4984	16.67	0.6607	0.6502
3HL & 5N	**GD_K7**	**0.6597**	**0.5186**	**0.9711**	**0.5539**	**90.73**	**0.3295**	**0.8968**	**1.1906**	**1.3644**	**0.2765**	**0.9703**	**0.2106**	**0.4085**	**0.3212**	**35.00**	**0.7958**	**0.4191**
3HL & 10N	GD_K8	0.8562	0.6993	0.9440	1.1404	85.88	0.4442	0.8262	0.8938	1.7978	0.4657	0.9584	0.3707	0.5508	0.4169	16.67	0.7246	0.5439
3HL & 15N	GD_K9	1.2139	1.0057	0.9113	1.9147	82.92	0.6389	0.6507	0.4459	2.7819	0.9361	0.9335	0.8678	0.7922	0.5910	16.67	0.6039	0.7711
4HL & 5N	GD_K10	0.8199	0.6661	0.9378	0.7856	84.46	0.4232	0.8406	0.9041	1.2334	0.4270	0.9587	0.1291	0.5247	0.3992	16.67	0.7377	0.5208
4HL & 10N	GD_K11	0.8178	0.6615	0.9356	0.6853	84.17	0.4202	0.8415	0.8991	1.2565	0.4249	0.9595	0.0305	0.5210	0.3982	16.67	0.7395	0.5195
4HL & 15N	GD_K12	0.6684	0.5540	0.9526	0.7159	89.77	0.3519	0.8941	1.1367	1.4420	0.2838	0.9645	0.1233	0.4364	0.3254	16.67	0.7818	0.4246
5HL & 5N	GD_K13	0.9124	0.7764	0.9025	0.9185	80.42	0.4932	0.8026	0.7063	1.4378	0.5289	0.9295	-0.0823	0.6115	0.4443	8.33	0.6942	0.5796
5HL & 10N	GD_K14	1.0930	0.7924	0.8683	0.8994	74.97	0.5034	0.7168	0.4106	1.7031	0.7589	0.9392	0.3724	0.6242	0.5322	25.00	0.6879	0.6943
5HL & 15N	GD_K15	1.3695	1.1509	0.7869	2.2277	61.90	0.7311	0.5554	-0.1312	2.9465	1.1915	0.8957	0.5182	0.9065	0.6668	8.33	0.5467	0.8700
1HL & 5N	GDA_K1	1.2848	0.8662	0.9050	0.8301	69.99	0.5503	0.6087	0.2342	1.5997	1.0485	0.9536	0.6202	0.6823	0.6255	16.67	0.6589	0.8161
1HL & 10N	GDA_K2	0.8433	0.6683	0.9516	1.0382	83.59	0.4245	0.8314	0.8981	1.7835	0.4517	0.9355	0.1378	0.5264	0.4106	25.00	0.7368	0.5357

Table 18 (continued)

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
1HL & 15N	GDA_K3	1.5941	1.3585	0.8747	2.5699	72.86	0.8630	0.3976	-0.1003	3.4849	1.6142	0.9240	1.1817	1.0700	0.7761	8.33	0.4650	1.0126
2HL & 5N	GDA_K4	1.1923	0.8656	0.9145	1.8224	70.03	0.5499	0.6630	0.3444	2.3092	0.9031	0.9551	0.3970	0.6818	0.5805	8.33	0.6591	0.7574
2HL & 10N	GDA_K5	0.9593	0.7990	0.9273	1.2834	78.71	0.5076	0.7818	0.6876	1.7278	0.5846	0.9557	0.1483	0.6294	0.4671	8.33	0.6853	0.6094
2HL & 15N	GDA_K6	0.8644	0.6583	0.9098	0.6788	82.73	0.4182	0.8229	0.7907	1.4073	0.4747	0.9477	0.1374	0.5185	0.4209	16.67	0.7407	0.5491
3HL & 5N	GDA_K7	1.2297	1.0915	0.9114	1.8973	82.12	0.6934	0.6415	0.4221	2.8224	0.9606	0.9191	0.8707	0.8597	0.5987	25.00	0.5701	0.7812
3HL & 10N	GDA_K8	1.0342	0.7181	0.9128	1.0354	77.92	0.4562	0.7464	0.5781	1.8708	0.6795	0.9587	0.3717	0.5656	0.5035	58.33	0.7172	0.6570
3HL & 15N	GDA_K9	1.0806	0.8784	0.8986	1.4063	76.68	0.5580	0.7232	0.4937	2.0375	0.7418	0.9461	0.4291	0.6919	0.5261	8.33	0.6541	0.6865
4HL & 5N	GDA_K10	0.8170	0.6382	0.9301	0.7354	84.96	0.4054	0.8418	0.8976	1.1823	0.4240	0.9424	-0.1820	0.5027	0.3978	16.67	0.7487	0.5190
4HL & 10N	GDA_K11	1.2234	0.9776	0.8792	1.6766	70.39	0.6210	0.6452	0.2534	2.4283	0.9508	0.9454	0.4976	0.7700	0.5957	33.33	0.6150	0.7772
4HL & 15N	**GDA_K12**	**0.6639**	**0.5150**	**0.9650**	**0.6711**	**89.62**	**0.3271**	**0.8955**	**1.1634**	**1.1513**	**0.2800**	**0.9654**	**0.0542**	**0.4056**	**0.3233**	**66.67**	**0.7972**	**0.4218**
5HL & 5N	GDA_K13	0.6955	0.5940	0.9483	1.0867	88.65	0.3773	0.8853	1.0904	1.5126	0.3073	0.9610	0.0717	0.4678	0.3386	16.67	0.7661	0.4418
5HL & 10N	GDA_K14	1.3662	1.1819	0.8450	2.1121	71.10	0.7508	0.5575	0.0588	2.9369	1.1858	0.9179	0.8047	0.9310	0.6652	16.67	0.5345	0.8679
5HL & 15N	GDA_K15	0.8391	0.6166	0.9254	0.6777	83.43	0.3917	0.8331	0.8517	1.3709	0.4472	0.9412	0.0725	0.4857	0.4085	33.33	0.7572	0.5330

*Bold values correspond to the best architectural model

Table 19 Performance metrics details for ANN models in testing

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
1HL & 5N	LM_K1	0.3432	0.3090	0.8142	0.2050	45.17	0.1944	0.4156	0.7715	0.9395	0.0741	0.5479	-0.0854	0.9374	0.7645	50.00	0.5313	0.2159
1HL & 10N	LM_K2	0.3817	0.3148	0.7869	0.2227	31.16	0.1980	0.2771	0.5492	1.0495	0.0916	0.5839	0.0834	0.9549	0.8502	41.67	0.5226	0.2401
1HL & 15N	LM_K3	0.2464	0.1779	0.9005	0.1190	74.97	0.1119	0.6987	1.3141	1.0679	0.0382	0.7600	0.1013	0.5398	0.5489	83.33	0.7301	0.1550
2HL & 5N	LM_K4	0.2715	0.2042	0.9348	0.1342	63.46	0.1284	0.6342	1.2371	0.9884	0.0464	0.7763	0.0094	0.6193	0.6048	83.33	0.6903	0.1708
2HL & 10N	LM_K5	0.2418	0.1768	0.8609	0.1305	72.25	0.1112	0.7099	1.2218	1.0473	0.0368	0.7243	0.0504	0.5364	0.5386	75.00	0.7318	0.1521
2HL & 15N	LM_K6	0.2111	0.1793	0.8847	0.1187	78.12	0.1128	0.7788	1.3528	0.9969	0.0280	0.6700	-0.0223	0.5440	0.4703	91.67	0.7280	0.1328
3HL & 5N	LM_K7	0.2190	0.1664	0.9161	0.1207	83.17	0.1047	0.7620	1.4519	1.0844	0.0302	0.7558	0.1186	0.5047	0.4879	83.33	0.7476	0.1378
3HL & 10N	LM_K8	0.1890	0.1499	0.9224	0.1050	82.32	0.0943	0.8227	1.4851	0.9919	0.0225	0.7575	-0.0098	0.4547	0.4210	83.33	0.7726	0.1189
3HL & 15N	LM_K9	0.1862	0.1474	0.9198	0.1049	82.79	0.0927	0.8279	1.4877	1.0017	0.0218	0.7687	-0.0014	0.4472	0.4148	75.00	0.7764	0.1171
4HL & 5N	LM_K10	0.1405	0.1261	0.9522	0.0872	90.59	0.0793	0.9021	1.6721	1.0228	0.0124	0.7958	0.0279	0.3826	0.3129	100.00	0.8087	0.0884
4HL & 10N	LM_K11	0.0998	0.0798	0.9771	0.0566	95.15	0.0502	0.9506	1.8064	1.0076	0.0063	0.8729	0.0132	0.2421	0.2223	100.00	0.8790	0.0628
4HL & 15N	LM_K12	0.1626	0.1409	0.9574	0.0923	86.93	0.0886	0.8688	1.6233	1.0009	0.0166	0.8103	0.0106	0.4275	0.3623	100.00	0.7862	0.1023
5HL & 5N	LM_K13	0.1048	0.0733	0.9882	0.0525	96.97	0.0461	0.9455	1.8416	0.9496	0.0069	0.8796	-0.0699	0.2223	0.2334	91.67	0.8888	0.0659
5HL & 10N	LM_K14	0.0834	0.0688	0.9834	0.0430	96.64	0.0433	0.9655	1.8500	1.0088	0.0044	0.8894	0.0134	0.2088	0.1857	100.00	0.8956	0.0524
5HL & 15N	**LM_K15**	**0.0487**	**0.0419**	**0.9959**	**0.0279**	**99.16**	**0.0264**	**0.9882**	**1.9348**	**1.0163**	**0.0015**	**0.9339**	**0.0260**	**0.1272**	**0.1084**	**100.00**	**0.9364**	**0.0306**
1HL & 5N	BFG_K1	0.4290	0.3502	0.7855	0.2346	47.79	0.2203	0.0865	0.6660	0.8211	0.1158	0.4555	-0.2808	1.0623	0.9558	41.67	0.4688	0.2699
1HL & 10N	BFG_K2	0.2934	0.2271	0.7886	0.1532	58.01	0.1429	0.5729	0.9086	1.0384	0.0541	0.5795	0.0382	0.6890	0.6535	66.67	0.6555	0.1845
1HL & 15N	BFG_K3	0.3114	0.2534	0.9394	0.1734	55.54	0.1594	0.5187	1.1264	0.9180	0.0610	0.7187	-0.0859	0.7688	0.6937	58.33	0.6156	0.1959
2HL & 5N	BFG_K4	0.2280	0.1635	0.9254	0.1064	74.21	0.1029	0.7421	1.3705	0.9899	0.0327	0.7902	0.0006	0.4960	0.5078	83.33	0.7520	0.1434
2HL & 10N	BFG_K5	0.2297	0.2005	0.8883	0.1335	74.73	0.1261	0.7382	1.3068	1.0300	0.0332	0.7074	0.0428	0.6081	0.5116	83.33	0.6960	0.1445

Table 19 (continued)

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
2HL & 15N	BFG_K6	0.2132	0.1797	0.9195	0.1148	77.67	0.1130	0.7745	1.4090	0.9821	0.0286	0.7486	-0.0211	0.5450	0.4749	91.67	0.7275	0.1341
3HL & 5N	BFG_K7	0.2766	0.2344	0.8078	0.1656	62.04	0.1475	0.6204	0.9964	1.0178	0.0481	0.6233	-0.0016	0.7111	0.6161	66.67	0.6444	0.1740
3HL & 10N	BFG_K8	0.2274	0.1746	0.8702	0.1068	75.23	0.1098	0.7433	1.2822	1.0432	0.0325	0.6256	0.0425	0.5295	0.5066	83.33	0.7352	0.1430
3HL & 15N	**BFG_K9**	**0.1466**	**0.1057**	**0.9508**	**0.0658**	**89.39**	**0.0665**	**0.8933**	**1.6514**	**1.0033**	**0.0135**	**0.8184**	**-0.0102**	**0.3206**	**0.3266**	**100.00**	**0.8397**	**0.0922**
4HL & 5N	BFG_K10	0.1943	0.1684	0.9085	0.1090	81.52	0.1059	0.8126	1.4462	0.9895	0.0237	0.7145	-0.0229	0.5109	0.4328	83.33	0.7446	0.1222
4HL & 10N	BFG_K11	0.1785	0.1441	0.9435	0.1026	84.22	0.0907	0.8419	1.5539	0.9968	0.0200	0.7872	0.0075	0.4372	0.3976	83.33	0.7814	0.1123
4HL & 15N	BFG_K12	0.1733	0.1435	0.9470	0.0880	85.33	0.0903	0.8510	1.5769	1.0109	0.0189	0.8057	0.0215	0.4353	0.3861	91.67	0.7824	0.1090
5HL & 5N	BFG_K13	0.2399	0.2022	0.8729	0.1436	71.71	0.1272	0.7143	1.2391	0.9921	0.0362	0.6948	-0.0237	0.6132	0.5345	58.33	0.6934	0.1509
5HL & 10N	BFG_K14	0.2752	0.2226	0.9089	0.1471	62.91	0.1400	0.6243	1.1801	0.9620	0.0476	0.7089	-0.0314	0.6753	0.6130	66.67	0.6624	0.1731
5HL & 15N	BFG_K15	0.3523	0.3210	0.8769	0.2116	39.04	0.2019	0.3839	0.8070	1.0032	0.0781	0.6455	0.0364	0.9738	0.7849	50.00	0.5131	0.2216

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
1HL & 5N	SCG_K1	0.2775	0.2133	0.8595	0.1532	61.80	0.1341	0.6179	1.0793	1.0020	0.0484	0.7018	0.0033	0.6469	0.6181	66.67	0.6765	0.1745
1HL & 10N	SCG_K2	0.2933	0.2384	0.9159	0.1558	58.42	0.1500	0.5730	1.1298	1.0147	0.0541	0.7323	0.0475	0.7233	0.6535	66.67	0.6383	0.1845
1HL & 15N	SCG_K3	0.3152	0.2720	0.7871	0.1899	50.91	0.1711	0.5071	0.8135	0.9952	0.0625	0.5484	-0.0203	0.8251	0.7021	41.67	0.5875	0.1982
2HL & 5N	SCG_K4	0.2329	0.1931	0.8928	0.1301	76.91	0.1215	0.7307	1.3332	1.0591	0.0341	0.6986	0.0879	0.5857	0.5189	75.00	0.7071	0.1465
2HL & 10N	SCG_K5	0.2451	0.2034	0.9292	0.1373	70.24	0.1279	0.7020	1.3207	0.9788	0.0378	0.7429	-0.0089	0.6169	0.5459	75.00	0.6916	0.1541
2HL & 15N	SCG_K6	0.2238	0.1777	0.9073	0.1253	76.86	0.1118	0.7515	1.3680	0.9574	0.0315	0.7142	-0.0587	0.5391	0.4985	75.00	0.7304	0.1408
3HL & 5N	SCG_K7	0.2848	0.2349	0.9116	0.1557	62.54	0.1478	0.5975	1.1715	1.0310	0.0510	0.7229	0.0749	0.7126	0.6344	75.00	0.6437	0.1791
3HL & 10N	SCG_K8	0.1587	0.1093	0.9456	0.0734	87.51	0.0687	0.8750	1.6107	0.9961	0.0158	0.8329	-0.0047	0.3314	0.3535	91.67	0.8343	0.0998

Table 19 (continued)

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
3HL & 15N	SCG_K9	0.2540	0.2148	0.9342	0.1394	72.13	0.1351	0.6798	1.3399	1.0439	0.0406	0.7405	0.0914	0.6515	0.5659	75.00	0.6742	0.1598
4HL & 5N	**SCG_K10**	**0.1267**	**0.1026**	**0.9728**	**0.0663**	**93.98**	**0.0646**	**0.9203**	**1.7593**	**1.0420**	**0.0101**	**0.8533**	**0.0626**	**0.3113**	**0.2822**	**100.00**	**0.8444**	**0.0797**
4HL & 10N	SCG_K11	0.2108	0.1929	0.9544	0.1226	79.91	0.1213	0.7795	1.4993	1.0288	0.0279	0.7725	0.0629	0.5851	0.4696	91.67	0.7074	0.1326
4HL & 15N	SCG_K12	0.2555	0.2232	0.8773	0.1480	68.58	0.1404	0.6759	1.1998	0.9686	0.0411	0.6464	-0.0446	0.6771	0.5693	58.33	0.6614	0.1607
5HL & 5N	SCG_K13	0.1998	0.1491	0.9568	0.0927	80.18	0.0938	0.8018	1.5175	0.9881	0.0251	0.8220	-0.0023	0.4523	0.4452	91.67	0.7739	0.1257
5HL & 10N	SCG_K14	0.3056	0.2472	0.9318	0.1539	54.26	0.1555	0.5366	1.1053	0.9976	0.0587	0.7415	0.0348	0.7499	0.6808	58.33	0.6251	0.1922
5HL & 15N	SCG_K15	0.2820	0.2238	0.7985	0.1479	61.86	0.1408	0.6054	0.9743	0.9815	0.0500	0.5437	-0.0516	0.6788	0.6282	75.00	0.6606	0.1774
1HL & 5N	GDM_K1	0.3695	0.2846	0.7509	0.1989	32.33	0.1790	0.3224	0.5176	0.9949	0.0859	0.5631	-0.0135	0.8634	0.8232	58.33	0.5683	0.2324
1HL & 10N	GDM_K2	0.3867	0.3148	0.8277	0.2031	28.43	0.1980	0.2577	0.5827	1.0425	0.0941	0.6735	0.0732	0.9550	0.8616	66.67	0.5225	0.2433
1HL & 15N	GDM_K3	0.1921	0.1344	0.9234	0.0932	82.81	0.0845	0.8169	1.4886	0.9678	0.0232	0.7747	-0.0473	0.4077	0.4279	75.00	0.7961	0.1208
2HL & 5N	GDM_K4	0.2186	0.1742	0.8860	0.1210	77.80	0.1096	0.7629	1.3443	1.0473	0.0301	0.7136	0.0552	0.5284	0.4870	83.33	0.7358	0.1375
2HL & 10N	GDM_K5	0.2305	0.1974	0.9228	0.1331	77.81	0.1242	0.7362	1.3992	0.9306	0.0334	0.6969	-0.0919	0.5988	0.5136	75.00	0.7006	0.1450
2HL & 15N	GDM_K6	0.2600	0.2234	0.8559	0.1552	67.27	0.1405	0.6645	1.1452	1.0297	0.0425	0.6584	0.0407	0.6777	0.5792	66.67	0.6611	0.1636
3HL & 5N	GDM_K7	0.2695	0.2373	0.8902	0.1659	64.15	0.1492	0.6396	1.1645	0.9791	0.0457	0.6838	-0.0195	0.7197	0.6003	66.67	0.6401	0.1695
3HL & 10N	GDM_K8	0.1718	0.1409	0.9361	0.0940	86.22	0.0886	0.8535	1.5667	0.9720	0.0186	0.7612	-0.0419	0.4274	0.3828	91.67	0.7863	0.1081
3HL & 15N	GDM_K9	0.1815	0.1447	0.9245	0.1023	84.60	0.0910	0.8365	1.5192	1.0349	0.0207	0.7812	0.0438	0.4389	0.4044	83.33	0.7806	0.1142
4HL & 5N	GDM_K10	0.1867	0.1391	0.9125	0.0981	82.87	0.0875	0.8270	1.4748	1.0202	0.0219	0.7694	0.0186	0.4219	0.4159	83.33	0.7890	0.1174
4HL & 10N	GDM_K11	0.1875	0.1184	0.9223	0.0767	84.25	0.0745	0.8254	1.5056	0.9666	0.0221	0.7968	-0.0586	0.3593	0.4178	83.33	0.8203	0.1180
4HL & 15N	**GDM_K12**	**0.1706**	**0.1039**	**0.9511**	**0.0683**	**86.22**	**0.0688**	**0.8546**	**1.5908**	**1.0345**	**0.0179**	**0.8209**	**0.0563**	**0.3422**	**0.3192**	**100.00**	**0.8897**	**0.1011**

Table 19 (continued)

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
5HL & 5N	GDM_K13	0.1905	0.1561	0.9185	0.1043	82.27	0.0982	0.8199	1.4759	0.9871	0.0228	0.7576	-0.0237	0.4736	0.4243	91.67	0.7632	0.1198
5HL & 10N	GDM_K14	0.3147	0.2715	0.8115	0.1819	52.23	0.1708	0.5086	0.8661	0.9633	0.0623	0.5375	-0.0526	0.8235	0.7010	58.33	0.5882	0.1979
5HL & 15N	GDM_K15	0.2487	0.2210	0.9225	0.1427	69.29	0.1390	0.6929	1.2952	0.9858	0.0389	0.7172	-0.0001	0.6703	0.5541	66.67	0.6648	0.1565
HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
1HL & 5N	GD_K1	0.2559	0.2208	0.8324	0.1384	67.53	0.1389	0.6751	1.1123	1.0118	0.0412	0.5713	0.0067	0.6697	0.5700	91.67	0.6651	0.1609
1HL & 10N	GD_K2	0.3433	0.2864	0.7587	0.1851	43.85	0.1801	0.4152	0.6708	0.9635	0.0741	0.5200	-0.0685	0.8687	0.7647	58.33	0.5657	0.2159
1HL & 15N	GD_K3	0.3655	0.3029	0.8236	0.2020	35.55	0.1905	0.3371	0.6683	1.0304	0.0840	0.6310	0.0608	0.9187	0.8142	50.00	0.5406	0.2299
2HL & 5N	GD_K4	0.2783	0.2364	0.8978	0.1629	64.47	0.1487	0.6156	1.1725	0.9414	0.0487	0.6938	-0.0766	0.7170	0.6200	75.00	0.6415	0.1751
2HL & 10N	GD_K5	0.2549	0.1995	0.8565	0.1194	68.10	0.1255	0.6776	1.1597	1.0163	0.0409	0.6652	0.0262	0.6050	0.5678	83.33	0.6975	0.1603
2HL & 15N	GD_K6	0.2700	0.1925	0.8334	0.1319	64.29	0.1211	0.6382	1.0674	0.9831	0.0459	0.6409	-0.0306	0.5840	0.6015	83.33	0.7080	0.1698
3HL & 5N	**GD_K7**	**0.2050**	**0.1645**	**0.9574**	**0.1004**	**79.38**	**0.1035**	**0.7913**	**1.5055**	**0.9718**	**0.0264**	**0.8006**	**-0.0224**	**0.4991**	**0.4568**	**100.00**	**0.7505**	**0.1290**
3HL & 10N	GD_K8	0.2902	0.2469	0.8897	0.1734	68.81	0.1553	0.5820	1.1895	0.8970	0.0530	0.6016	-0.1462	0.7489	0.6465	50.00	0.6256	0.1825
3HL & 15N	GD_K9	0.2637	0.2419	0.8851	0.1660	72.33	0.1521	0.6549	1.2429	0.9258	0.0437	0.6230	-0.1174	0.7337	0.5875	50.00	0.6332	0.1659
4HL & 5N	GD_K10	0.2808	0.2292	0.8456	0.1527	63.05	0.1442	0.6086	1.0647	0.9569	0.0496	0.6271	-0.0664	0.6953	0.6256	66.67	0.6524	0.1766
4HL & 10N	GD_K11	0.2370	0.1948	0.9076	0.1366	76.14	0.1225	0.7213	1.3480	0.9388	0.0353	0.7072	-0.0899	0.5910	0.5279	75.00	0.7045	0.1491
4HL & 15N	GD_K12	0.2613	0.2161	0.8469	0.1516	71.19	0.1359	0.6612	1.1679	1.0813	0.0429	0.6325	0.1011	0.6556	0.5821	75.00	0.6722	0.1644
5HL & 5N	GD_K13	0.2347	0.1797	0.8557	0.1017	73.20	0.1130	0.7267	1.2295	1.0329	0.0346	0.6224	0.0329	0.5451	0.5228	100.00	0.7275	0.1476
5HL & 10N	GD_K14	0.2398	0.2032	0.9069	0.1399	72.53	0.1278	0.7146	1.3080	0.9607	0.0362	0.7009	-0.0463	0.6163	0.5342	83.33	0.6919	0.1508
5HL & 15N	GD_K15	0.2157	0.1993	0.9005	0.1349	78.98	0.1254	0.7690	1.3849	1.0455	0.0293	0.6953	0.0647	0.6046	0.4806	75.00	0.6977	0.1357

Table 19 (continued)

HL & N	Model ID	RMSE	MAE	R	MAPE	VAF	WMAPE	NS	PI	BF	NMBE	WI	MBE	LMI	RSR	a20	IOA	IOS
1HL & 5N	GDA_K1	0.2805	0.2295	0.8626	0.1536	62.45	0.1444	0.6096	1.0881	1.0329	0.0495	0.6893	0.0549	0.6962	0.6248	66.67	0.6519	0.1764
1HL & 10N	GDA_K2	0.2709	0.2067	0.8676	0.1374	67.68	0.1300	0.6358	1.1586	1.0585	0.0462	0.7029	0.0909	0.6271	0.6035	75.00	0.6865	0.1704
1HL & 15N	GDA_K3	0.3870	0.3464	0.8529	0.2370	25.79	0.2179	0.2565	0.5983	0.9858	0.0942	0.6013	0.0167	1.0508	0.8622	33.33	0.4746	0.2435
2HL & 5N	GDA_K4	0.2788	0.2504	0.8784	0.1704	63.04	0.1575	0.6143	1.1232	1.0324	0.0489	0.6725	0.0570	0.7597	0.6210	66.67	0.6202	0.1754
2HL & 10N	GDA_K5	0.2081	0.1698	0.8900	0.1107	78.67	0.1068	0.7851	1.3706	0.9954	0.0272	0.6906	-0.0177	0.5152	0.4635	91.67	0.7424	0.1309
2HL & 15N	GDA_K6	0.3114	0.2535	0.8493	0.1688	60.80	0.1594	0.5188	1.0180	1.0897	0.0610	0.6812	0.1340	0.7689	0.6937	58.33	0.6155	0.1959
3HL & 5N	GDA_K7	0.2328	0.1602	0.9556	0.1079	77.07	0.1008	0.7311	1.4512	0.9249	0.0341	0.7947	-0.0894	0.4861	0.5185	83.33	0.7570	0.1464
3HL & 10N	GDA_K8	0.1774	0.1575	0.9206	0.1072	84.61	0.0991	0.8438	1.5162	1.0265	0.0198	0.7154	0.0218	0.4777	0.3953	83.33	0.7611	0.1116
3HL & 15N	GDA_K9	0.2072	0.1774	0.9365	0.1200	86.78	0.1116	0.7869	1.5376	1.0881	0.0270	0.7611	0.1277	0.5383	0.4616	91.67	0.7309	0.1303
4HL & 5N	GDA_K10	0.2611	0.2224	0.8988	0.1471	74.46	0.1399	0.6618	1.2915	1.0832	0.0429	0.7076	0.1292	0.6746	0.5816	75.00	0.6627	0.1642
4HL & 10N	GDA_K11	0.2074	0.1695	0.9318	0.1070	83.42	0.1066	0.7864	1.4951	1.0637	0.0271	0.7715	0.0981	0.5142	0.4621	91.67	0.7429	0.1305
4HL & 15N	**GDA_K12**	**0.1399**	**0.0975**	**0.9685**	**0.0551**	**90.29**	**0.0613**	**0.9029**	**1.7009**	**0.9094**	**0.0123**	**0.8645**	**-0.0005**	**0.2959**	**0.3117**	**91.67**	**0.8521**	**0.0880**
5HL & 5N	GDA_K13	0.2750	0.2423	0.8191	0.1633	66.31	0.1524	0.6247	1.0590	1.0708	0.0476	0.5177	0.0880	0.7349	0.6126	75.00	0.6326	0.1730
5HL & 10N	GDA_K14	0.2360	0.1828	0.9124	0.1199	81.38	0.1150	0.7236	1.4102	0.9143	0.0350	0.6876	-0.1348	0.5545	0.5258	75.00	0.7227	0.1484
5HL & 15N	GDA_K15	0.2562	0.2077	0.9163	0.1340	67.85	0.1307	0.6742	1.2618	0.9695	0.0413	0.7374	-0.0294	0.6302	0.5708	83.33	0.6849	0.1612

*Bold values correspond to the best architectural model

Table 20 Database used in the present research

S. No	FG (%)	S (%)	SG	LL (%)	PI (%)	OMC (%)	MDD (g/cc)	K (10^-6) cm/sec	S. No	FG (%)	S (%)	SG	LL (%)	PI (%)	OMC (%)	MDD (g/cc)	K (10^-6) cm/sec
1	93.00	7.00	2.70	47.29	20.65	24.00	1.53	0.05	53	92.50	7.50	2.70	35.08	16.10	20.11	1.57	1.29
2	93.00	7.00	2.70	51.60	23.17	24.70	1.50	0.12	54	82.00	18.00	2.72	42.42	18.65	15.90	1.76	1.30
3	91.00	9.00	2.70	47.48	23.37	20.50	1.60	0.13	55	78.50	21.50	2.71	41.05	17.30	16.60	1.76	1.34
4	84.50	15.50	2.70	38.93	16.42	18.50	1.65	0.16	56	90.98	9.02	2.69	44.25	18.36	17.70	1.69	1.36
5	90.91	9.09	2.66	51.32	21.92	19.80	1.67	0.21	57	65.67	12.59	2.76	40.15	18.20	22.30	1.60	1.42
6	92.00	8.00	2.69	46.75	22.57	21.00	1.58	0.21	58	65.67	12.59	2.76	40.15	18.20	22.30	1.60	1.43
7	66.50	33.50	2.69	36.23	16.73	15.60	1.67	0.23	59	84.50	15.50	2.69	41.14	19.10	15.60	1.72	1.46
8	81.63	17.84	2.80	42.18	16.70	20.50	1.65	0.23	60	55.50	44.50	2.72	28.33	10.46	11.80	1.85	1.53
9	81.63	17.84	2.80	42.18	16.70	20.50	1.65	0.23	61	56.00	44.00	2.71	28.30	10.81	13.00	1.90	1.58
10	72.00	28.00	2.70	39.93	18.10	17.70	1.67	0.25	62	96.18	3.82	2.69	49.20	23.30	21.30	1.56	1.60
11	69.50	30.50	2.71	32.86	15.29	18.00	1.68	0.31	63	59.00	41.00	2.73	29.25	11.85	13.50	1.82	1.62
12	72.66	27.34	2.60	39.84	16.26	16.50	1.72	0.33	64	97.00	3.00	2.70	48.20	21.60	19.70	1.57	1.74
13	74.00	26.00	2.72	32.30	14.62	19.00	1.67	0.33	65	85.00	15.00	2.68	40.93	19.14	17.20	1.62	1.74
14	69.00	31.00	2.71	34.17	15.49	14.00	1.70	0.35	66	83.00	17.00	2.71	43.22	19.51	16.50	1.72	1.88
15	87.49	12.51	2.57	46.29	20.53	20.00	1.67	0.39	67	69.26	30.74	2.68	31.30	12.28	14.00	1.85	2.41
16	90.36	9.64	2.54	49.20	22.45	19.50	1.61	0.40	68	80.10	13.03	2.60	34.42	11.86	17.40	1.67	2.42
17	87.49	12.51	2.57	46.29	20.53	20.00	1.67	0.40	69	84.12	11.21	2.57	36.65	14.47	16.11	1.71	2.52
18	90.36	9.64	2.54	49.20	22.45	19.50	1.61	0.40	70	86.67	7.44	2.62	39.26	16.50	16.50	1.69	2.58
19	92.87	7.13	2.67	51.94	22.36	20.40	1.64	0.41	71	89.31	10.69	2.66	34.70	10.30	19.90	1.69	2.60
20	92.25	7.75	2.58	49.00	22.40	21.00	1.63	0.41	72	71.50	28.50	2.69	25.77	9.44	15.00	1.78	2.67
21	88.65	11.35	2.62	50.29	21.79	20.00	1.67	0.43	73	63.18	36.82	2.64	26.80	7.91	16.30	1.67	2.78
22	71.50	28.50	2.66	37.77	17.04	20.00	1.65	0.44	74	56.50	43.50	2.70	28.46	11.45	12.90	1.89	2.80
23	73.98	25.52	2.80	37.46	12.35	20.80	1.67	0.45	75	59.82	40.18	2.65	28.70	12.79	11.70	1.88	2.91
24	73.98	25.52	2.80	37.46	12.35	20.80	1.67	0.45	76	57.00	43.00	2.71	24.84	11.68	12.90	1.90	3.15
25	56.50	43.50	2.70	29.46	11.74	14.40	1.70	0.46	77	79.50	20.50	2.77	28.30	10.30	14.70	1.73	3.43
26	66.50	33.50	2.70	33.05	14.13	13.70	1.75	0.47	78	62.00	38.00	2.71	23.85	8.73	12.30	1.82	3.65
S. No	FG (%)	S (%)	SG	LL (%)	PI (%)	OMC (%)	MDD (g/cc)	K (10^-6) cm/sec	S. No	FG (%)	S (%)	SG	LL (%)	PI (%)	OMC (%)	MDD (g/cc)	K (10^-6) cm/sec
27	68.00	32.00	2.69	33.53	15.14	16.00	1.66	0.48	79	84.08	13.05	2.69	35.00	11.17	16.50	1.71	4.16
28	71.00	29.00	2.71	36.74	16.74	19.70	1.63	0.49	80	70.05	29.50	2.66	30.38	10.52	12.50	1.92	5.63
29	90.72	9.28	2.59	46.90	19.40	19.50	1.66	0.54	81	54.19	30.94	2.70	28.85	11.16	13.10	1.86	8.04
30	87.93	12.07	2.67	35.81	15.61	18.80	1.56	0.59	82	65.67	12.59	2.76	40.15	18.20	22.30	1.60	1.42
31	68.66	27.89	2.68	43.71	12.63	22.10	1.55	0.59	83	68.24	14.55	2.71	38.73	17.56	17.20	1.63	1.28
32	77.50	22.50	2.72	41.33	18.78	14.20	1.76	0.61	84	54.19	30.94	2.70	28.85	11.16	13.10	1.86	8.04
33	96.50	3.50	2.53	43.50	18.60	19.40	1.64	0.65	85	63.18	36.82	2.64	26.80	7.91	16.30	1.67	2.78

Table 20 (continued)

S. No	FG (%)	S (%)	SG	LL (%)	PI (%)	OMC (%)	MDD (g/cc)	K (10^-6) cm/sec	S. No	FG (%)	S (%)	SG	LL (%)	PI (%)	OMC (%)	MDD (g/cc)	K (10^-6) cm/sec
34	60.98	39.00	2.67	34.70	15.86	10.60	1.65	0.66	86	88.51	11.49	2.61	33.83	14.26	16.10	1.69	0.97
35	79.33	20.67	2.70	32.80	17.82	14.80	1.78	0.67	87	53.65	46.35	2.66	24.80	7.97	11.70	1.93	3.92
36	84.61	15.39	2.65	45.20	18.70	19.50	1.58	0.70	88	59.82	40.18	2.65	28.70	12.79	11.70	1.88	2.91
37	88.50	11.50	2.71	35.95	17.64	13.30	1.74	0.73	89	80.08	19.92	2.75	42.82	19.18	15.20	1.77	0.60
38	73.50	26.50	2.72	37.06	16.70	17.20	1.65	0.73	90	88.33	11.67	2.73	33.60	14.24	15.00	1.73	0.91
39	85.52	11.36	2.58	52.95	24.67	20.40	1.65	0.77	91	82.91	17.09	2.74	51.97	23.44	15.70	1.77	0.51
40	92.02	6.92	2.54	46.58	20.44	20.70	1.65	0.77	92	98.72	1.28	2.71	54.30	24.70	25.00	1.55	0.67
41	94.00	6.00	2.55	35.70	17.40	19.50	1.67	0.87	93	87.00	13.00	2.62	46.80	18.43	22.79	1.58	1.22
42	86.50	13.50	2.72	31.44	14.09	19.20	1.70	0.93	94	72.00	28.00	2.64	38.29	15.12	16.73	1.72	1.41
43	51.25	48.75	2.70	34.53	15.66	11.40	1.93	0.94	95	60.00	40.00	2.73	32.98	12.46	15.21	1.76	1.97
44	68.00	32.00	2.69	33.00	14.60	13.00	1.77	0.97	96	100.00	0.00	2.65	54.18	22.08	24.72	1.53	1.34
45	89.00	11.00	2.70	35.04	17.03	18.50	1.65	0.99	97	52.00	30.00	2.80	35.51	13.91	17.35	1.73	2.85
46	68.50	31.50	2.71	32.70	15.24	16.40	1.73	1.01	98	85.00	15.00	2.62	45.81	18.04	21.99	1.60	1.35
47	91.17	8.83	2.59	40.50	16.21	19.30	1.66	1.02	99	97.00	3.00	2.65	52.42	21.36	24.63	1.54	1.19
48	92.82	5.79	2.64	44.38	19.74	18.40	1.68	1.02	100	65.00	34.00	2.67	34.67	13.53	15.74	1.76	1.56
49	51.00	49.00	2.69	31.88	14.35	16.60	1.74	1.20	101	99.00	1.00	2.65	53.61	21.87	24.70	1.54	1.28
50	86.80	13.20	2.71	44.41	18.96	16.50	1.70	1.23	102	84.00	16.00	2.62	45.33	17.86	21.54	1.60	1.42
51	68.00	30.00	2.72	34.05	15.03	18.60	1.74	1.23	103	75.00	25.00	2.63	40.40	15.92	17.53	1.69	1.56
52	73.45	26.55	2.67	33.50	15.20	13.30	1.81	1.28	104	60.00	30.00	2.76	35.91	14.20	16.32	1.71	1.92

*Bold values correspond to the best architectural model

See Table 17, Table 18, Table 19 and Table 20.

Author contributions Jitendra Khatti: Main author, conceptualization, methodological development, modelling, design, manuscript preparation and finalization, validation, and analysis, manuscript finalization, detailed review, and editing; Kamaldeep Singh Grover: manuscript preparation and finalization, validation, and analysis, comprehensive analysis, detailed review, and editing.

Funding No funding was received to assist in the preparation of this manuscript.

Data availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

Albalasmeh A, Mohawesh O, Gharaibeh M, Deb S, Slaughter L, El Hanandeh A (2022) Artificial neural network optimization to predict saturated hydraulic conductivity in arid and semi-arid regions. CATENA 217:106459. https://doi.org/10.1016/j.catena.2022.106459

Al-Dosary NMN, Al-Sulaiman MA, Aboukarima AM (2019) Modelling the unsaturated hydraulic conductivity of a sandy loam soil using Gaussian process regression. Water SA, 45(1):121–130. https://hdl.handle.net/10520/EJC-13bdcd5372

Arshad M, Nazir MS, O'Kelly BC (2020) Evolution of hydraulic conductivity models for sandy soils. Proceedings of the Institution of Civil Engineers-Geotechnical Engineering 173(2):97–114. https://doi.org/10.1680/jgeen.18.00062

Asteris PG, Koopialipoor M, Armaghani DJ, Kotsonis EA, Lourenço PB (2021a) Prediction of cement-based mortars compressive strength using machine learning techniques. Neural Comput Appl 33(19):13089–13121. https://doi.org/10.1007/s00521-021-06004-8

Asteris PG, Lourenço PB, Hajihassani M, Adami CEN, Lemonis ME, Skentou AD, Marques R, Nguyen H, Rodrigues H, Varum H (2021b) Soft computing-based models for the prediction of masonry compressive strength. Eng Struct 248:113276. https://doi.org/10.1016/j.engstruct.2021.113276

Asteris PG, Skentou AD, Bardhan A, Samui P, Pilakoutas K (2021c) Predicting concrete compressive strength using hybrid ensembling of surrogate machine learning models. Cem Concr Res 145:106449. https://doi.org/10.1016/j.cemconres.2021.106449

Azarhoosh MJ, Koohmishi M (2023) Prediction of hydraulic conductivity of porous granular media by establishment of random forest algorithm. Constr Build Mater 366:130065. https://doi.org/10.1016/j.conbuildmat.2022.130065

Babaoglu Y, Simms P (2020) Improving hydraulic conductivity estimation for soft clayey soils, sediments, or tailings using predictors measured at high-void ratio. Journal of Geotechnical and Geoenvironmental Engineering 146(10):06020016. https://doi.org/10.1061/(ASCE)GT.1943-5606.0002344

Bahmed IT, Khatti J, Grover KS (2024) Hybrid soft computing models for predicting unconfined compressive strength of lime stabilized soil using strength property of virgin cohesive soil. Bull Eng Geol Env 83(1):46. https://doi.org/10.1007/s10064-023-03537-1

Bardhan A, Samui P, Ghosh K, Gandomi AH, Bhattacharyya S (2021) ELM-based adaptive neuro swarm intelligence techniques for predicting the California bearing ratio of soils in soaked conditions. Appl Soft Comput 110:107595. https://doi.org/10.1016/j.asoc.2021.107595

Báťková K, Matula S, Miháliková M, Hrúzová E, Abebrese DK, Kara RS, Almaz C (2023) Prediction of saturated hydraulic conductivity Ks of agricultural soil using pedotransfer functions. Soil Water Res 18(1)

Benson CH, Trast JM (1995) Hydraulic conductivity of thirteen compacted clays. Clays Clay Miner 43:669–681. https://doi.org/10.1346/CCMN.1995.0430603

Benson CH, Zhai H, Wang X (1994) Estimating hydraulic conductivity of compacted clay liners. Journal of Geotechnical Engineering 120(2):366–387. https://doi.org/10.1061/(ASCE)0733-9410(1994)120:2(366)

Bi J, Bennett KP (2003) Regression error characteristic curves. In Proceedings of the 20th international conference on machine learning (ICML-03) (pp 43–50)

Chan JYL, Leow SMH, Bea KT, Cheng WK, Phoong SW, Hong ZW, Chen YL (2022) Mitigating the multicollinearity problem and its machine learning approach: a review. Mathematics 10(8):1283. https://doi.org/10.3390/math10081283

Chandel A, Shankar V (2022) Evaluation of empirical relationships to estimate the hydraulic conductivity of borehole soil samples. ISH Journal of Hydraulic Engineering 28(4):368–377. https://doi.org/10.1080/09715010.2021.1902872

Chandel A, Sharma S, Shankar V (2022) Prediction of hydraulic conductivity of porous media using a statistical grain-size model. Water Supply 22(4):4176–4192. https://doi.org/10.2166/ws.2022.043

Chandel A, Shankar V, Kumar N (2023) Neural computing techniques to estimate the hydraulic conductivity of porous media. Water Supply. https://doi.org/10.2166/ws.2023.143

Chen L, Zhang X (2020) A model for predicting the hydraulic conductivity of warm saturated frozen soil. Build Environ 179:106939. https://doi.org/10.1016/j.buildenv.2020.106939

Daniel C, Khatti J, Grover KS (2024) Assessment of compressive strength of high-performance concrete using soft computing approaches. Comput Concr 33(1):55. https://doi.org/10.12989/cac.2024.33.1.055

Emberga TT, Opara AI, Onyekuru SO, Omenikolo AI, Bilar AA, Unegbu CC, Anuforo DN, Epuerie TE (2023) Estimates of aquifer hydraulic conductivity based on grain-size data and multiple regression techniques in Imo River Basin. Int J Energ Water Res:1–19. https://doi.org/10.1007/s42108-023-00244-1

Faloye OT, Ajayi AE, Ajiboye Y, Alatise MO, Ewulo BS, Adeosun SS, Babalola T, Horn R (2022) Unsaturated hydraulic conductivity prediction using artificial intelligence and multiple linear regression models in biochar amended sandy clay loam soil. J Soil Sci Plant Nutr 22(2):1589–1603. https://doi.org/10.1007/s42729-021-00756-x

Fatoba JO, Sanuade OA, Amosun JO, Hammed OS (2018) Prediction of hydraulic conductivity from Dar Zarrouk parameters using artificial neural network. Indian J Geosci 72(1):51–64

Gareth J, Daniela W, Trevor H, Robert T (2013) An introduction to statistical learning: with applications in R. Springer, New York

Ghorbani B, Arulrajah A, Narsilio G, Horpibulsuk S, Bo MW (2020) Development of genetic-based models for predicting the resilient modulus of cohesive pavement subgrade soils. Soils Found 60(2):398–412. https://doi.org/10.1016/j.sandf.2020.02.010

Golbraikh A, Tropsha A (2002) Beware of q2! J Mol Graph Model 20(4):269–276. https://doi.org/10.1016/S1093-3263(01)00123-1

Granata F, Di Nunno F, Modoni G (2022) Hybrid machine learning models for soil saturated conductivity prediction. Water 14(11):1729. https://doi.org/10.3390/w14111729

Gupta S, Lehmann P, Bonetti S, Papritz A, Or D (2021) Global prediction of soil saturated hydraulic conductivity using random forest in a covariate-based geoTransfer function (CoGTF)

framework. Journal of Advances in Modeling Earth Systems 13(4):e2020MS002242. https://doi.org/10.1029/2020MS002242
Hair JF, Celsi MW, Ortinau DJ, Bush RP (2017) Essentials of marketing research. McGraw-Hill/Irwin, New York
Hasanzadehshooiili H, Lakirouhani A, Medzvieckas J (2012) Superiority of artificial neural networks over statistical methods in prediction of the optimal length of rock bolts. J Civ Eng Manag 18(5):655–661. https://doi.org/10.3846/13923730.2012.724029
Hedayati-Azar A, Sadeghi H (2022) Semi-empirical modelling of hydraulic conductivity of clayey soils exposed to deionized and saline environments. J Contam Hydrol 249:104042. https://doi.org/10.1016/j.jconhyd.2022.104042
Hosseini Y, Sedghi R, Bairami S (2022) An evaluation of genetic algorithm method compared to geostatistical and neural network methods to estimate saturated soil hydraulic conductivity using soil texture. Iran Agric Res 36(1):91–104. https://doi.org/10.22099/iar.2017.4039
Hosseini S, Khatti J, Taiwo BO, Fissha Y, Grover KS, Ikeda H, Pushkarna M, Berhanu M, Ali M (2023) Assessment of the ground vibration during blasting in mining projects using different computational approaches. Sci Rep 13(1):18582. https://doi.org/10.1038/s41598-023-46064-5
Kashani MH, Ghorbani MA, Shahabi M, Naganna SR, Diop L (2020) Multiple AI model integration strategy—application to saturated hydraulic conductivity prediction from easily available soil properties. Soil and Tillage Research 196:104449. https://doi.org/10.1016/j.still.2019.104449
Khaja MA, Shah SR, Jha R (2023) Hydraulic conductivity estimation of sandy soils: a novel approach. ISH J Hydraul Eng:1–13. https://doi.org/10.1080/09715010.2023.2187712
Khalili-Maleki M, Poursorkhabi RV, Nadiri AA, Dabiri R (2022) Prediction of hydraulic conductivity based on the soil grain size using supervised committee machine artificial intelligence. Earth Sci Inf 15(4):2571–2583. https://doi.org/10.1007/s12145-022-00848-x
Khatti J, Grover KS (2021) Relationship between index properties and CBR of soil and prediction of CBR. In: Indian Geotechnical Conference. Springer Nature Singapore, Singapore, pp 171–185. https://doi.org/10.1007/978-981-19-6774-0_16
Khatti J, Grover KS (2023a) Prediction of compaction parameters for fine-grained soil: Critical comparison of the deep learning and standalone models. J Rock Mech Geotech Eng 15(11):3010–3038. https://doi.org/10.1016/j.jrmge.2022.12.034
Khatti J, Grover KS (2023b) CBR prediction of pavement materials in unsoaked condition using LSSVM, LSTM-RNN, and ANN approaches. Int J Pavement Res Technol:1–37. https://doi.org/10.1007/s42947-022-00268-6
Khatti J, Grover KS (2023c) Assessment of fine-grained soil compaction parameters using advanced soft computing techniques. Arab J Geosci 16(3):208. https://doi.org/10.1007/s12517-023-11268-6
Khatti J, Grover KS (2023d) Prediction of UCS of fine-grained soil based on machine learning part 2: comparison between hybrid relevance vector machine and Gaussian process regression. Multiscale Multidiscip Model Exp Des:1–41. https://doi.org/10.1007/s41939-023-00191-8
Khatti J, Grover KS (2023e) Estimation of intact rock uniaxial compressive strength using advanced machine learning. Trans Infrastruct Geotechnol:1–34. https://doi.org/10.1007/s40515-023-00357-4
Khatti J, Grover KS (2023f) A scientometrics review of soil properties prediction using soft computing approaches. Arch Comput Methods Eng:1–35. https://doi.org/10.1007/s11831-023-10024-z
Khatti J, Samadi H, Grover KS (2023) Estimation of settlement of pile group in clay using soft computing techniques. Geotech Geol Eng:1–32. https://doi.org/10.1007/s10706-023-02643-x
Khatti J, Grover KS, Kim HJ, Mawuntu KBA, Park TW (2024) Prediction of ultimate bearing capacity of shallow foundations on cohesionless soil using hybrid lstm and rvm approaches: an extended investigation of multicollinearity. Comput Geotech 165:105912. https://doi.org/10.1016/j.compgeo.2023.105912
Kim B, Roh G, Lee J, Yoon J, Lee J (2023) Characterizing the hydraulic conductivity of soil based on the moving average of precipitation and groundwater level using a regional database. AQUA-Water Infrastructure, Ecosystems and Society. https://doi.org/10.2166/aqua.2023.044
Kumar M, Samui P (2020) Reliability analysis of settlement of pile group in clay using LSSVM, GMDH, GPR. Geotech Geol Eng 38:6717–6730. https://doi.org/10.1007/s10706-020-01464-6
Li PN, Xu YS, Wang XW (2023) Estimation of hydraulic conductivity by the modified Kozeny-Carman equation considering the derivation principle of the original equation. J Hydrol 621:129658. https://doi.org/10.1016/j.jhydrol.2023.129658
Mady AY, Shein EV (2018) Support vector machine and nonlinear regression methods for estimating saturated hydraulic conductivity. Moscow University Soil Science Bulletin 73:129–133. https://doi.org/10.3103/S0147687418030079
Menard S (2002) Applied logistic regression analysis (No. 106). SAGE Publications, Thousand Oaks
Mentaschi L, Besio G, Cassola F, Mazzino A (2013) Problems in RMSE-based wave model validations. Ocean Model 72:53–58. https://doi.org/10.1016/j.ocemod.2013.08.003
Ming F, Chen L, Li D, Wei X (2020) Estimation of hydraulic conductivity of saturated frozen soil from the soil freezing characteristic curve. Sci Total Environ 698:134132. https://doi.org/10.1016/j.scitotenv.2019.134132
More SB, Deka PC (2018) Estimation of saturated hydraulic conductivity using fuzzy neural network in a semi-arid basin scale for murum soils of India. ISH Journal of Hydraulic Engineering 24(2):140–146. https://doi.org/10.1080/09715010.2017.1400408
More SB, Deka PC, Patil AP, Naganna SR (2022) Machine learning-based modeling of saturated hydraulic conductivity in soils of tropical semi-arid zone of India. Sādhanā 47(1):26. https://doi.org/10.1007/s12046-022-01805-6
Mufti S, Das A (2023) Modeling unsaturated hydraulic conductivity of granular soils using a combined discrete element and pore-network approach. Acta Geotech 18(2):651–672. https://doi.org/10.1007/s11440-022-01597-3
Mujtaba H, Shimobe S, Farooq K, Rehman ZU, Khalid U (2021) Relating gradational parameters with hydraulic conductivity of sandy soils: a renewed attempt. Arab J Geosci 14(18):1920. https://doi.org/10.1007/s12517-021-08281-y
Naganna SR, Deka PC (2019) Artificial intelligence approaches for spatial modeling of streambed hydraulic conductivity. Acta Geophys 67:891–903. https://doi.org/10.1007/s11600-019-00283-5
Nematolahi M, Jalali V, Hejazi Mehrizi M (2018) Predicting saturated hydraulic conductivity using particle swarm optimization and genetic algorithm. Arab J Geosci 11:1–11. https://doi.org/10.1007/s12517-018-3846-2
Peters A, Hohenbrink TL, Iden SC, Durner W (2021) A simple model to predict hydraulic conductivity in medium to dry soil from the water retention curve. Water Resources Research 57(5):e2020WR029211. https://doi.org/10.1029/2020WR029211
Peters A, Hohenbrink TL, Iden SC, van Genuchten MT, Durner W (2023a) Prediction of the absolute hydraulic conductivity function from soil water retention data. Hydrol Earth Syst Sci 27(7):1565–1582. https://doi.org/10.5194/hess-27-1565-2023
Peters A, Iden SC, Durner W (2023b) Full prediction of unsaturated hydraulic conductivity–comparison of four different capillary

bundle models. Hydrology and Earth System Sciences Discussions 2023:1–28. https://doi.org/10.5194/hess-2023-134
Pham K, Won J (2022) Enhancing the tree-boosting-based pedotransfer function for saturated hydraulic conductivity using data preprocessing and predictor importance using game theory. Geoderma 420:115864. https://doi.org/10.1016/j.geoderma.2022.115864
Piri H, Mobaraki M, Mir M (2023) Comparison and application of random forest, chaid and geostatistics models in predicting soil saturated hydraulic conductivity. Iranian J Ecohydrol 10(2):173–185
Qaderi K, Jalali V, Etminan S, Masoumi Shahr-babak M, Homaee M (2018) Estimating soil hydraulic conductivity using different data-driven models of ANN, GMDH and GMDH-HS. Paddy Water Environ, 16(4):823–833. https://doi.org/10.1007/s10333-018-0672-9
Rout S, Singh SP (2021) Prediction of compressibility and hydraulic conductivity of bentonitic mixtures. Proceedings of the Institution of Civil Engineers-Geotechnical Engineering 174(2):225–237. https://doi.org/10.1680/jgeen.19.00307
Ruan K, Fu XL (2022) A modified Kozeny-Carman equation for predicting saturated hydraulic conductivity of compacted bentonite in confined condition. Journal of Rock Mechanics and Geotechnical Engineering 14(3):984–993. https://doi.org/10.1016/j.jrmge.2021.08.010
Shan J, Yang Z, Kuang X, Li L, Liu J (2022) Comparison of seven Weibull distribution models for predicting relative hydraulic conductivity. Water Resources Research 58(5):e2021WR030683. https://doi.org/10.1029/2021WR030683
Shi XS, Yin J (2018) Estimation of hydraulic conductivity of saturated sand–marine clay mixtures with a homogenization approach. Int J Geomech 18(7):04018082. https://doi.org/10.1061/(ASCE)GM.1943-5622.0001190
Shrestha N (2020) Detecting multicollinearity in regression analysis. Am J Appl Math Stat 8(2):39–42
Sihag P (2018) Prediction of unsaturated hydraulic conductivity using fuzzy logic and artificial neural network. Modeling Earth Systems and Environment 4:189–198. https://doi.org/10.1007/s40808-018-0434-0
Sihag P, Esmaeilbeiki F, Singh B, Ebtehaj I, Bonakdari H (2019a) Modeling unsaturated hydraulic conductivity by hybrid soft computing techniques. Soft Comput 23:12897–12910. https://doi.org/10.1007/s00500-019-03847-1
Sihag P, Mohsenzadeh Karimi S, Angelaki A (2019b) Random forest, M5P and regression analysis to estimate the field unsaturated hydraulic conductivity. Appl Water Sci 9:1–9. https://doi.org/10.1007/s13201-019-1007-8
Sihag P, Tiwari NK, Ranjan S (2019c) Prediction of unsaturated hydraulic conductivity using adaptive neuro-fuzzy inference system (ANFIS). ISH J Hydraul Eng 25(2):132–142. https://doi.org/10.1080/09715010.2017.1381861
Sihag P, Singh B, Sepah Vand A, Mehdipour V (2020) Modeling the infiltration process with soft computing techniques. ISH Journal of Hydraulic Engineering 26(2):138–152. https://doi.org/10.1080/09715010.2018.1464408
Singh U, Sharma PK (2023) Comparison of saturated hydraulic conductivity estimated by surface NMR and empirical equations. J Hydrol 617:128929. https://doi.org/10.1016/j.jhydrol.2022.128929
Singh VK, Panda KC, Sagar A, Al-Ansari N, Duan HF, Paramaguru PK, Vishwakarma DK, Kumar A, Kumar D, Kashyap PS, Singh RM (2022) Novel Genetic Algorithm (GA) based hybrid machine learning-pedotransfer Function (ML-PTF) for prediction of spatial pattern of saturated hydraulic conductivity. Engineering Applications of Computational Fluid Mechanics 16(1):1082–1099. https://doi.org/10.1080/19942060.2022.2071994
Singh K, Singh B, Sihag P, Kumar V, Sharma KV (2023) Development and application of modeling techniques to estimate the unsaturated hydraulic conductivity. Modeling Earth Systems and Environment :1–15. https://doi.org/10.1007/s40808-023-01744-z
Smith GN (1986) Probability and statistics in civil engineering – An introduction. Collins, London
Tahmassebi A, Gandomi AH, Meyer-Baese A (2018) A Pareto front based evolutionary model for airfoil self-noise prediction. In: 2018 IEEE Congress on Evolutionary Computation (CEC). IEEE, pp 1–8. https://doi.org/10.1109/CEC.2018.8477987
Tan Y, Zhang P, Chen J, Shamet R, Nam BH, Pu H (2023) Predicting the hydraulic conductivity of compacted soil barriers in landfills using machine learning techniques. Waste Manage 157:357–366. https://doi.org/10.1016/j.wasman.2023.01.003
Tan Y, Chen J, Benson CH (2022) Predicting hydraulic conductivity of geosynthetic clay liners using a neural network algorithm. In Geo-Congress 2022, pp 21–28
Tao G, Chen Y, Zhu X, Xiao H, Chen Q (2022) Relationship between multifractal characteristics of particle-size distribution and hydraulic conductivity of unsaturated soils. Eur J Environ Civ Eng 26(8):3447–3458. https://doi.org/10.1080/19648189.2020.1802352
Teng J, Dong A, Yan H, Tong C, Zhang S (2023) Predicting the hydraulic conductivity of frozen coarse-grained soils. J Hydrol 617:129048. https://doi.org/10.1016/j.jhydrol.2022.129048
Thakur D, Chandel A, Shankar V (2022) Estimation of hydraulic conductivity of porous media using data-driven techniques. Water Practice & Technology 17(12):2625–2638. https://doi.org/10.2166/wpt.2022.151
Trejo-Alonso J, Quevedo A, Fuentes C, Chávez C (2020) Evaluation and development of pedotransfer functions for predicting saturated hydraulic conductivity for mexican soils. Agronomy 10(10):1516. https://doi.org/10.3390/agronomy10101516
Tseng CY, Ghadiri M, Kumar P, Meidani H (2023) Estimation of hydraulic conductivity in a watershed using sparse multi-source data via gaussian process regression and bayesian experimental design. Adv Water Resour:104489. https://doi.org/10.1016/j.advwatres.2023.104489
ur Rehman Z, Khalid U, Ijaz N, Mujtaba H, Haider A, Farooq K, Ijaz Z (2022) Machine learning-based intelligent modeling of hydraulic conductivity of sandy soils considering a wide range of grain sizes. Eng Geol 311:106899. https://doi.org/10.1016/j.enggeo.2022.106899
Veloso MF, Rodrigues LN, Fernandes Filho EI (2022) Evaluation of machine learning algorithms in the prediction of hydraulic conductivity and soil moisture at the Brazilian Savannah. Geoderma Reg:e00569. https://doi.org/10.1016/j.geodrs.2022.e00569
Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE (2006) Regression methods in biostatistics: linear, logistic, survival, and repeated measures models. https://doi.org/10.1007/b138825
Wang Y, Ma R, Zhu G (2023) Representation of the influence of soil structure on hydraulic conductivity prediction. J Hydrol 619:129330. https://doi.org/10.1016/j.jhydrol.2023.129330
Williams CG, Ojuri OO (2021) Predictive modelling of soils' hydraulic conductivity using artificial neural network and multiple linear regression. SN Applied Sciences 3:1–13. https://doi.org/10.1007/s42452-020-03974-7
Willmott CJ, Robeson SM, Matsuura K (2012) A refined index of model performance. Int J Climatol 32(13):2088–2094. https://doi.org/10.1002/joc.2419

Yan Y, Nakhli SAA, Jin J, Mills G, Willson CS, Legates DR, Manahiloh KN, Imhoff PT (2021) Predicting the impact of biochar on the saturated hydraulic conductivity of natural and engineered media. J Environ Manage 295:113143. https://doi.org/10.1016/j.jenvman.2021.113143

Zeitfogel H, Feigl M, Schulz K (2023) Soil information on a regional scale: Two machine learning based approaches for predicting saturated hydraulic conductivity. Geoderma 433:116418. https://doi.org/10.1016/j.geoderma.2023.116418

Zhang P, Tan Y, Chen J, Nam BH (2023) Using random forest algorithm to predict the hydraulic conductivity of compacted soil liners/covers. In Geo-Congress 2023, pp 193–200

Zhang L, Wang J (2023) Prediction of the soil saturated hydraulic conductivity in a mining area based on CT scanning technology. J Clean Prod 383:135364. https://doi.org/10.1016/j.jclepro.2022.135364