ORIGINAL PAPER

Estimation of California bearing ratio for hill highways using advanced hybrid artificial neural network algorithms

Ishwor Thapa[1] · Sufyan Ghani[1]

Abstract

California bearing ratio (CBR) is one of the important parameters that is used to express the strength of the pavement subgrade of railways, roadways, and airport runways. CBR is usually determined in the laboratory in soaked and unsoaked conditions, which is an exhaustive and time-consuming process. Therefore, to sidestep the operation of conducting actual laboratory tests, this study presents the development of efficient hybrid soft computing techniques, by hybridizing artificial neural network (ANN) with nature-inspired optimization algorithm, namely, gradient-based optimization (GBO), firefly algorithm (FF), cultural algorithms (CA), grey wolf optimization (GWO), genetic algorithm (GA), particle swarm optimization (PSO), Harris Hawk optimization (HHO), teaching learning-based optimization (TLBO), Whale optimization algorithm (WOA) and invasive weed optimization (IWO). For this purpose, a data set was prepared from the experimental results of soaked CBR of soil samples collected from an ongoing Nepal's Mid-Hill Highway project. Based on the detailed comparative study one explicit model is proposed to estimate the CBR of soils in soaked conditions. The predictive accuracy of the proposed models was evaluated via several statistical and graphical parameters. Separate statistical indices were employed to evaluate the generalization capabilities of the developed models. In addition, in the end, the best predictive model was determined using a novel tool called order analysis. The results of the study reveal that the proposed artificial neural network coupled with the gradient-based optimizer (ANN–GBO) model attained the most accurate prediction ($R^2 = 0.997$ and $R^2 = 0.956$, during the training and testing phase) in predicting the soaked CBR. Based on the accuracies attained, the proposed ANN–GBO model has very potential to be an alternate solution to estimate the CBR value in different phases of civil engineering projects.

Keywords California bearing ratio · Artificial intelligence · Statistical parameters · Gradient-based optimizer · Order analysis

1 Introduction

In the realm of transportation infrastructure, such as railways, roads, and airport runways, the proper construction of a sub-base layer is essential. This sub-base layer often serves as the primary load-bearing foundation, with the upper layer of the subsoil being prepared as a sub-base or subgrade. Meeting various engineering and technical criteria, including settlement requirements, subgrade reaction, bearing capacity, and swelling properties, is imperative for these layers.

✉ Sufyan Ghani
Sufyan.ghani@sharda.ac.in

Ishwor Thapa
2022822566.ishwor@dr.sharda.ac.in

1 Department of Civil Engineering, Sharda University, Greater Noida, India

Consequently, the assessment of such layers is of paramount importance in geotechnical engineering, particularly for road transportation network. It is the most commonly used mode of transportation that allows the moving of goods and passengers from one location to another conveniently. Strong pavement needs to be designed and built to develop a successful road network (Koti Marg and Puram 2019). One common method employed to evaluate the shear strength and stiffness modulus of a subgrade is the California bearing ratio (CBR). This method indirectly measures the strength of the subgrade material by comparing it to the strength of a standard crushed rock sample (Bardhan et al. 2021a, b). CBR is defined as the ratio of the force per unit area needed to penetrate a soil mass to that required for the standard material. To determine this ratio, a standardized circular piston with a velocity of 1.27 mm/min is utilized to penetrate the soil sample, a process carried out either in a laboratory setting or in the field. Typically, laboratory CBR tests are conducted on compacted

soil samples at their optimum moisture content (OMC). Field CBR tests, on the other hand, are performed at various levels: the natural ground surface, the prepared subgrade level, or on a level surface within a test pit at construction sites. While laboratory tests can be executed on both unsoaked and soaked compacted soil samples with OMC, they can also be conducted on natural soils. The results obtained from these tests are then compared with standard values to estimate the soil sample's strength. This information is highly significant in geotechnical engineering and its associated structures, such as highways, bridge abutments and piers, highway embankments, and more.

However, obtaining accurate CBR values for design purposes often presents challenges to engineers, practitioners, and industry professionals. Inadequate soil investigation results are a common issue, and the CBR test itself is labor-intensive and time-consuming. Furthermore, inaccuracies in laboratory results can arise due to sample disturbances, testing negligence, and inadequate testing facilities (Taskiran 2010; Yildirim and Gunaydin 2011). Consequently, researchers have devised approximation models that predict CBR values by considering soil index properties as the primary governing parameters. Numerous studies have elucidated the influence of soil types and their fundamental characteristics on CBR, revealing various correlations between CBR and other soil parameters. According to previous research, soil's CBR value is influenced by its plasticity properties, which include its w_L, w_P, and I_P gradational parameters, which comprise the fine, sand, and gravel content, as well as compaction criteria, such as maximum dry density and optimal moisture content.

Studies in the past used regression techniques as well as soft computing techniques such as multivariate adaptive regression splines (MARS), support vector machines (SVM), gene-expression programming (GEP), group method of data handling (GMDH), extreme learning machines (ELM) and artificial neural network (ANN) to predict the CBR value of a given soil sample. Various soft computing techniques were employed in the development of the prediction model for the soaked CBR value of different soil types. Table 1 shows the complete range of geotechnical factors as well as the number of data sets utilized to create computer models for estimating CBR. In comparison with statistical analysis, the new soft computing techniques show a stronger connection with experimental values.

Kiran and Ravi (2008) compared the performance of soft computing techniques, such as ANN and fuzzy logic, with statistical analysis methods in modelling and predicting experimental data. The findings demonstrated that soft computing methods had a stronger connection with experimental values, demonstrating their advantage in capturing complicated relationships and managing noisy data. In analysing

experimental data, Soltanali et al. (2021) contrasted the effectiveness of soft computing techniques, such as SVM and evolutionary algorithms, with more conventional statistical analysis techniques. The results showed that soft computing techniques performed better than statistical methods in terms of precision and prognostication ability, emphasising their efficiency in handling complicated and nonlinear patterns found in experimental data. Huang et al. (2015) looked into the efficiency of soft computing methods for modelling and simulating experimental data, including fuzzy systems and evolutionary algorithms. The outcomes showed that when compared to conventional statistical analysis techniques, soft computing techniques offered better approximations and forecasts of the experimental data. According to the study's findings, soft computing techniques are better suited to handle ambiguous and imprecise data, improving correlation with experimental results.

A literature analysis of numerous works on the prediction of CBR values for various types of soils using various computer methodologies is included in Table 1. The type of soil, number of data sets used in the study, computational strategy, and coefficient of determination (R^2/R) attained in the prediction models are all listed for each entry in the table.

There are some noticeable gaps in the research despite the table's insightful analysis of how different computational techniques and soil types perform when predicting CBR values. The research listed in the table mostly concentrate on sand, granular, fine-grained, and mixed soil types. Other soil types, such as expansive soils, organic soils, and clayey soils, are not well-represented. In addition, it can be observed that the majority of studies use traditional machine learning techniques, such as multiple linear regression (MLR), artificial neural networks (ANN), and support vector machines (SVM). There is a need for exploring more advanced machine learning algorithms, as well as the integration of metaheuristic optimization techniques, to potentially improve CBR predictions. Some studies have relatively small data sets, which might limit the generalizability of their models. Many studies do not explicitly consider temporal or regional variations in their models. Soil properties can vary significantly with time and location. In addition, there is a need for more studies that not only predict CBR values but also validate the accuracy of these predictions against real-world field data. In addition, comparative studies that assess the strengths and weaknesses of different computational approaches in different soil types could provide valuable insights.

Addressing these literature gaps would contribute to the advancement of CBR prediction techniques, making them more reliable and applicable to a wider range of soil types and conditions, thereby benefiting geotechnical engineering and infrastructure design. In view of the above, the present study provides a detailed comparative study of ten hybrid computational techniques for the prediction of CBR value of sub-base

Table 1 List of literature review to forecast the CBR of various soil types

Literature study	Type of soil	Number of data sets	Computational approach	R^2/R
Alam et al. (2020)	Fine-grained soil	20	GEP	0.94
			ANN	0.96
Bardhan et al. (2021a, b)	Mixed soil samples	312	MARS	0.9
			GP	0.88
			SVM	0.87
Erzin and Turkoz (2016)	Sandy soil	61	MLR	0.81
			ANN	0.98
Katte et al. (2019)	Subgrade soil	33	MLR	0.84
Kurnaz and Kaya (2020)	Mixed soil samples	158	GMDH	0.97
Tenpe and Patel (2020)	Mixed soil samples	389	GEP	0.82
			ANN	0.89
Taha et al. (2019)	Granular soil	218	ANN	0.97
Tenpe and Patel (2020)	Mixed soil samples	389	GEP	0.78
			SVM	0.8
Taskiran (2010)	Fine-grained soil	151	ANN	0.91
			GEP	0.92
Yildirim and Gunaydin (2011)	Granular soil	124	SLR	0.86
			MLR	0.88
			ANN	0.93
Varghese et al. (2013)	Fine-grained soil	112	MLR	0.83
			ANN	0.85

soil under soaked condition. A large database of 100 observations spanning a wide variety of geotechnical characteristics has been employed to train test and validate the developed models. The predictive accuracy of the proposed models was evaluated via several statistical and graphical parameters. Separate statistical indices were employed to evaluate the generalization capabilities of the developed models. In addition, in the end, the best predictive model was determined using a novel tool called order analysis.

1.1 Significance of research and contribution

The primary contribution of this study resides in the development of a precise and efficient model for the assessment of soil's California bearing ratio (CBR) without the need for experimental investigations. The research utilized 100 in situ samples of sub-base soil, encompassing a wide range of index and soil engineering properties. Laboratory tests, conducted following the Bureau of Indian Standards (BIS) (IS 2720-16) requirements, generated a vast data set of geotechnical characteristics. The research outcome has the potential to instigate a paradigm shift in geotechnical engineering practices by harnessing hybrid computational models based on artificial intelligence (AI) to predict the soaked CBR value of sub-base soil. This achievement holds the promise of significantly elevating construction standards, improving project timelines, and curbing costs, thereby making substantial contributions to the advancement of sustainable and resilient infrastructure development.

2 Machine learning algorithms

2.1 Artificial neural networks

ANNs are computational models inspired by the structure and functionality of the human brain. ANNs consist of interconnected nodes, called artificial neurons, which process and transmit information. These networks learn from examples and adjust their weights to make predictions or classify

data. ANN is widely used in various fields, including image recognition, natural language processing, and financial forecasting. Artificial neurons, the interconnected nodes that makeup ANNs, process and transfer information (Ghani et al. 2022b). These networks learn from examples and adjust their weights to make predictions or classify data. ANN is widely used in various fields, including image recognition, natural language processing, and financial forecasting.

Among the various AI methods, ANN emerged as the most widely used and preferred, accounting for 52% of the studies (Baghbani et al. 2022). Other methods were also used to a lesser extent, including fuzzy inference system (FIS), adaptive neuro-fuzzy inference system (ANFIS), support vector machine (SVM), long short-term memory (LSTM), convolutional neural network (CNN), residual neural network (ResNet), and generative adversarial network (GAN). Furthermore, in the past few years, a comparative and parametric study of AI-based models for risk assessment against soil liquefaction for high-intensity earthquakes for fine clayed soil has been studied (Ghani et al. 2022b). Ghani and Kumari (2023) explain that by leveraging artificial intelligence, engineers can improve their understanding of soil behaviour and make more accurate predictions. CBR of soil by comparing the models based on the least square support vector machine (LSSVM), long–short-term memory (LSTM), and artificial neural network (ANN) approach has been studied (Khatti and Grover 2023a). Two artificial intelligence approaches, artificial neural network (ANN) and random forest (RF) regression, were used to forecast the strength and CBR characteristics of stabilized CG mixes. The ANN models outperformed the RF regression models, achieving high correlation coefficient values of 0.993, 0.995, and 0.997 for UCS, unsoaked CBR, and soaked CBR, respectively. The mean absolute error values were 45.98 kPa, 1.41%, and 1.18% for UCS, unsoaked CBR, and soaked CBR, respectively (Amin et al. 2022). Vamsi Krishna et al. (2023) chemically processed black cotton soil with stabilizing additives, such as fly ash, lime, and cement to enhance granule connections and lessen expansibility and contractility. The primary objective is to assess the effects of these stabilizers on the CBR and Unconfined Compressive Strength (UCS) of the soil, thereby determining their suitability for road construction. Using additional soil parameters as inputs, the researchers effectively create an ANN-based model to predict CBR and UCS. A high coefficient of determination and low root mean square error, mean absolute error, and relative root mean square error values show that the results show that the ANN technique is suitable for expanding soil stabilization. Khatti and Grover (2023b) identified the best model for predicting the unsoaked California bearing ratio (CBRu) of soil as the LSTM model MD 14, achieving high accuracy and performance. It also highlights the importance of a nonlinear approach, the need for a substantial database for

artificial neural networks, and the influence of gravel content and maximum dry density on CBRu prediction. Using a large data set of 1011 in situ soil samples from a highway project, Verma et al. (2023) investigated various computer techniques such as kernel ridges regression, K-nearest neighbour, and Gaussian process regression to predict soaking CBR values of soils. explored novel computational methods, such as kernel ridges regression, K-nearest neighbour, and Gaussian process regression to predict soaked CBR values of soils, utilizing a vast data set of 1011 in situ soil samples from a highway project. The study highlights key influencing factors and demonstrates the efficacy of the GPR model, particularly when combined with K-Fold data division while emphasizing the impact of soil geological location on predictive accuracy. Bardhan et al. (2021a, b) introduced four efficient soft computing techniques for estimating California Bearing Ratio (CBR) in soaked conditions, with the MARS-L model demonstrating the highest accuracy ($R^2 = 0.9686$ and RMSE $= 0.0359$). These models offer a promising alternative to time-consuming laboratory tests for assessing subgrade strength in civil engineering projects. Samui et al. (2021) also introduced innovative ELM-based ANSI models, notably ELM–MPSO, for accurately predicting soil CBR in railway subgrade layers, outperforming traditional methods and offering robust solutions for geotechnical engineering challenges. Khatti and Grover (2023c) used an ANN model that outperformed other AI approaches in predicting soil's California bearing ratio (CBR), with a correlation coefficient of 0.9736. The ANN model, identified as the best, shows promise for accurate CBR predictions, reducing the need for time-consuming experimental procedures. Similarly, Khatti and Grover (2023c)research findings highlight that Model 21, a GA-optimized Laplacian kernel-based SRVM model, demonstrates superior predictive accuracy for fine-grained soil's soaked CBR, making it a robust and high-performing choice with minimal prediction error. Kassa and Wubineh (2023) used machine learning models, including random forest, decision tree, linear regression, and artificial neural network, were employed to predict soil CBR values based on seven laboratory-derived predictors in a study of 252 soil samples categorized using AASHTO M 145. Random forest outperformed other algorithms, demonstrating superior accuracy in CBR estimation, as evidenced by smaller errors and a higher R^2 value. Nagaraju et al. (2023) used a novel ELM–CSO approach to predict CBR values for laterite soil in the West Godavari area, demonstrating superior performance over standard ELM, with high correlation coefficients and accuracy, making it a promising tool for real-time engineering predictions in road construction. Kim et al. (2023) employed artificial intelligence, specifically the GMDH-type neural network algorithm, to predict California bearing ratio (CBR) values for both coarse- and

Fig. 1 ANN model network diagram
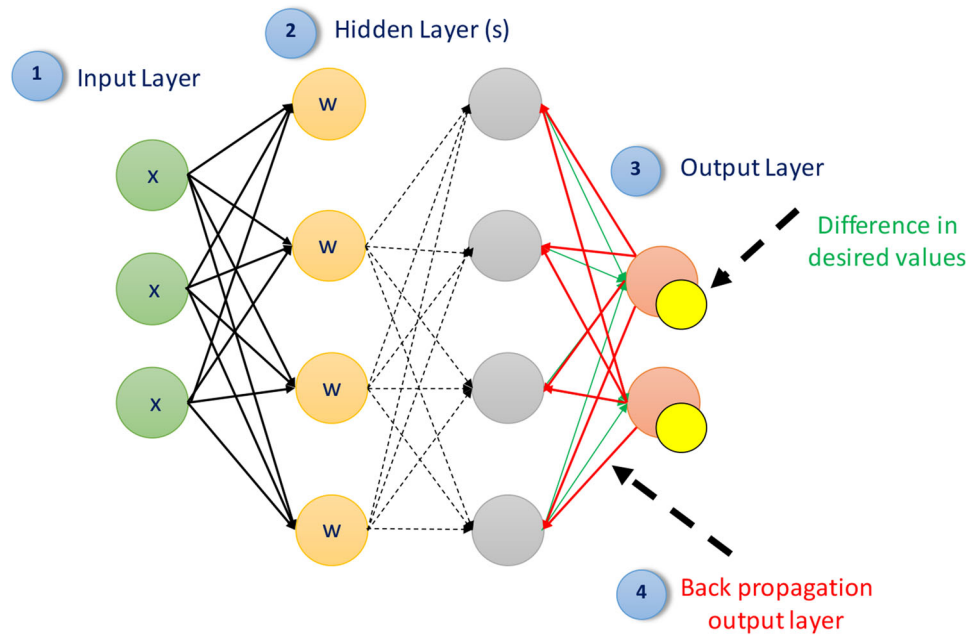
fine-grained soils. Results indicated superior predictive performance of the GMDH-type NN models compared to other regression methods, achieving high R^2 values of 0.938 for coarse-grained and 0.829 for fine-grained soils using specific input variables. The implications of this study are valuable for practising engineers, as it provides insights into effective modelling techniques and generates essential data to address soil mechanics problems using ANN for foundation and pavement design in geotechnical and engineering geology applications. Figure 1 depicts the full procedures used to construct the ANN model.

2.1.1 Nature-inspired metaheuristic optimization algorithm

The GWO algorithm is inspired by a pack of grey wolves' hierarchical structure and predator–prey dynamics. It is made up of alphas (male and female leaders), betas (alphas' assistants), omegas (the lowest rank), and deltas (dominant over omegas). GWO is a popular engineering tool for optimizing ANN (Bardhan et al. 2023). TLBO simulates the interaction between teachers and students in a classroom setting. TLBO is divided into two phases: teacher and learner. The quality of the teacher determines the performance of the learners. Each student strives to imitate the teacher and enhance their performance (Bui et al. 2022). FF algorithm is a population-based metaheuristic approach to optimization issues. Light variety and enticing formulas are the two most significant aspects of FF. These lights are used by fireflies to find mates for mating, detect prey, and generate awareness or terror among the swarm (Ghani et al. 2022a). The GA is a search method that is based on Darwin's theory of natural selection. It starts with a population of random solutions called chromosomes.

These chromosomes are binary strings that indicate potential problem solutions. In each generation, the chromosomes are evaluated using fitness functions, and new offspring are created through crossover and mutation. This method finds the best set of chromosomes as generations go by, reflecting the best or nearly the best solution to the issue. GA has been successfully used by researchers to solve challenging problems, including civil engineering problems (Ghani et al. 2022a). PSO is a popular optimization approach that has been used to solve a variety of technical challenges. It behaves similar to a flock of birds or a school of fish, with particles representing the population. Each particle is assigned a place and velocity at random, and their positions are revised based on their individual experience and the best position determined so far globally. To arrive at the best option, the algorithm iteratively evaluates the fitness of each particle's position using a cost function approach (Ghani et al. 2022a). GBO seeks an optimal design by utilizing function gradient information. The gradients of the objective function and the restrictions for a given point in the design space are calculated as the first stage in the numerical search process. The gradients are computed using a finite-difference approximation of the derivative (first-forward finite-difference techniques). Once the gradient has been computed, there are numerous ways to find the minimum. Sequential quadratic approaches such as the modified method of feasible directions (MMFD) can be utilized for confined problems, whereas quasi-Newton methods using a line search procedure are successful for unconstrained issues (Dababneh et al. 2018). Weeds are commonly characterized as undesirable plants that develop in agricultural land. Weeds are unhelpful and take up so much space in the field that they outweigh the plants grown for

regular use. As a result, a common agronomical idea is "The Weeds Always Win." A lot of times, weeds generate a lot of seeds, which are then transported by the wind or other natural forces. They are also highly adaptable and may thrive in challenging environments. These distinctive characteristics of weed growth offer a way to create optimization tactics. This frequent occurrence in agriculture served as inspiration for the IWO approach, which is based on the spread of invasive weeds (Kumar et al. 2020).

The WOA algorithm for problem-solving and optimization was inspired by the humpback whale's bubble-net attacking mechanism. As a result, the WOA is widely employed in a variety of industries, including management, energy, image processing, and machine vision. Near the ocean's surface, humpback whales frequently pursue krill or small fish for food. Because they move slowly, they have developed a unique hunting technique called foam feeding. Whales can make a spiral path with a decreasing radius by luring schools of fish to the surface and then catching them (Zhou et al. 2022). CA are based on theories developed in sociology and archaeology that attempt to model cultural change. According to such theories, cultural evolution can be viewed as an inheritance process that operates on two levels one a micro-evolutionary level, which consists of the genetic material that an offspring inherits from its parents, and the other a macro-evolutionary level, which consists of the knowledge acquired by individuals through generations. Once encoded and saved, this knowledge is utilized to steer the conduct of individuals who belong to a specific group. Population (Coello and Becerra 2004). HHO is inspired by the cooperative conduct and chasing manner of Harris' hawks in nature known as surprise pounce. In this clever approach, several hawks pounce on a victim from different directions in an attempt to catch it off guard. Harris hawks can exhibit a variety of pursuit patterns based on the dynamic nature of events and the prey's fleeing movements. To design an optimization algorithm, this work mathematically duplicates such dynamic patterns and behaviours (Prakash et al. 2023).

2.1.2 Hybridization of machine learning model

The learning parameters of ANN are optimized in the current study using ten well-known metaheuristic optimization techniques to estimate the soaked CBR of sub-base soil. ANN learning parameters include input weights, hidden neuron biases, output weights, and output bias. After the ANN model has been initialized, the optimization is done to maximize the parameters, which are the biases and weight to increase the learning of the parameters (Ceryan and Samui 2020). Before maximizing the learning parameters of ML models in this regard, the population size (m), maximum number of iterations (it), lower bound, upper bound, and other parameters other than the number of neurons (N) in the hidden

layer of ANN are set up. Then, using the aforementioned optimization process, the ANN's weights and biases, are optimized and finally, ten hybrid models named ANN–GBO, ANN–FF, ANN–CA, ANN–GWO, ANN–GA, ANN–PSO, ANN–HHO ANN–TLBO, ANN–WOA and ANN–IWO are created using this method. The lowest root mean square error is used to calculate the optimized values of learning parameters (RMSE). Even though all ten optimization algorithms maintain the same values for the above-mentioned parameters the optimized learning parameter values vary. In addition, keep in mind that the aforementioned parameters (n, N, it, lb, ub) are crucial to the optimization process and must be correctly tweaked during the hybridization process. Figure 2 depicts the full hybridization process, including the procedures used to construct the ANN model in conjunction with the optimization technique.

2.1.3 Data collection and data preparation for computational analysis

The soil samples taken on-site from the Mid-Hill Road project portion in Gandaki Province (GP), Nepal, were used to create a database for the analysis, as shown in Fig. 3. For quality assurance/control purposes, 100 soil samples in total were examined at the geotechnical engineering lab at Sharda University, India. The *% Gravel(G), % Sand(S), % Fine Content (M-C)*, w_L, w_P, I_P, S_L, *MDD, OMC*, and CBR at 2.5 mm were determined as useful geotechnical parameters. The Bureau of Indian Standards (IS 2720-16) was used to assess these parameters in controlled conditions. Figure 4 provides an overview of the predominant soil profile at the site, where these soil samples were collected. In addition, Fig. 4 showcases the laboratory testing equipment employed for conducting the soaked CBR tests on the soil samples (Fig. 5).

2.2 Statistical visualization and correlation analysis

One of the most used methods for determining the relationship between parameters is Pearson's correlation (R). R is a number between -1 and $+1$, where 1 denotes a strong correlation between the parameters and 0 (zero) denotes no correlation. The positive or negative sign indicates that the associated parameters are simultaneously increasing or decreasing. The correlation matrix created for all geotechnical indicators is shown in Fig. 6. CBR is negatively connected with *%S* and *OMC*, but positively with *%(M–C)*, w_L, w_P, I_P, and *MDD*, as shown in Fig. 6. *%G,%S, %(M–C), w_L,w_P, I_P, S_L, OMC* and *MDD* are the input variables for the CBR prediction model.

This positive correlation coefficient (0.82) suggests that as the percentage of gravel in the soil increases, the value of CBR also tends to increase. CBR is a measure of soil's

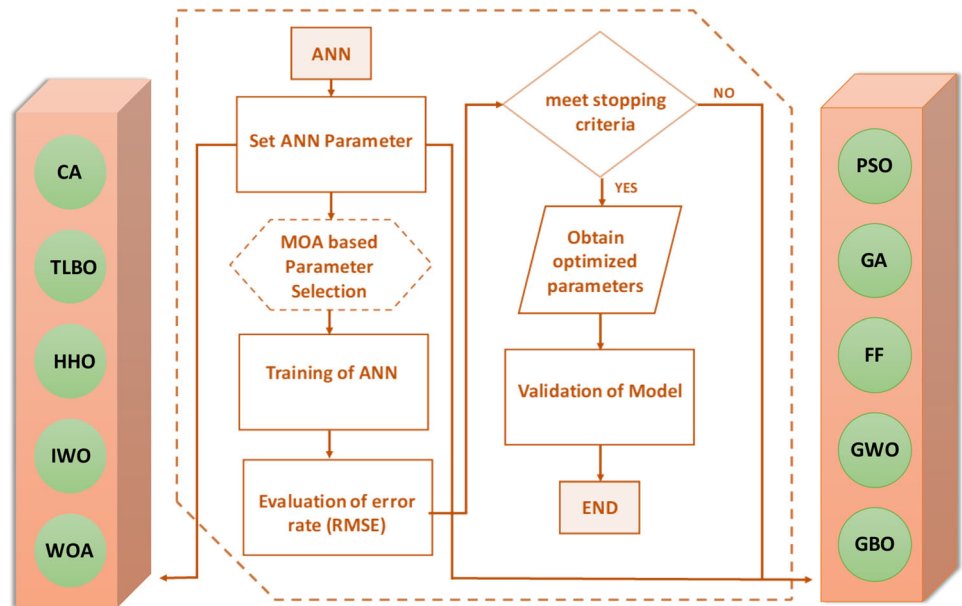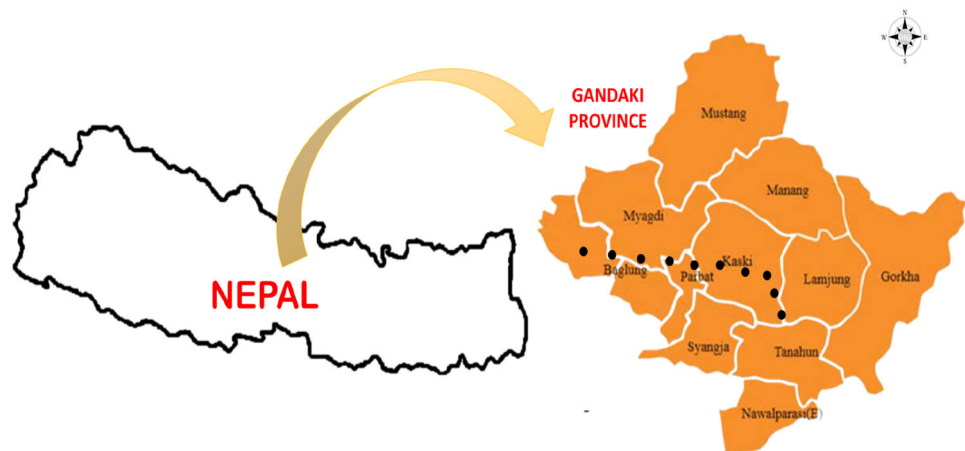**Figure. 2:** Flow chart for the hybridization process

Fig. 3 Study area Gandaki Province of Nepal

bearing capacity, so higher gravel content may improve the soil's strength and load-bearing capacity. The negative correlation coefficient (− 0.78) between %S content and CBR indicates an inverse relationship. As the percentage of sand increases, the CBR value decreases and vice versa. Sands have a low cohesion value, so a soil with higher sand content has a lower CBR value. The positive correlation coefficient (0.8) for w_L, (0.78) and (0.67) w_P indicates that soils with higher plastic limits also tend to have higher CBR values. Both w_L and w_P are related to the plasticity of soils, and this correlation suggests that soils with more plasticity tend to have higher CBR values. The positive correlation coefficient (0.17) suggests that as the *MDD* of the soil increases, the *CBR* value also tends to increase. The negative correlation coefficient (− 0.64) indicates that as the OMC of the soil increases, the CBR tends to decrease. This means that higher moisture content in soils decreases the CBR value of soil.

Figure 7 shows the effect of the nine independent variables on the CBR value of the soil. This plot shows that three of the independent variables % *S*, S_L (%), and *OMC*(%) are negatively correlated with the soaked CBR value and % *G*, % (M–C), w_L(%), w_P(%), I_P(%) and *MDD* (kg/m³) are positively correlated. Since sand particles have wider spaces between them, they drain better and retain less moisture. As a result, having a higher percentage of sand in the soil diminishes its ability to retain water, lowering the soaked CBR value (Koti Marg and Puram 2012). S_L (%) is the moisture content at which the soil considerably shrinks. Soils with a larger shrinkage limit have less water content, which results in lower wet CBR values. The *OMC* is the moisture content at which the soil compacts the most. Soils with higher *OMC* values contain more water, resulting in lower soaked CBR values.
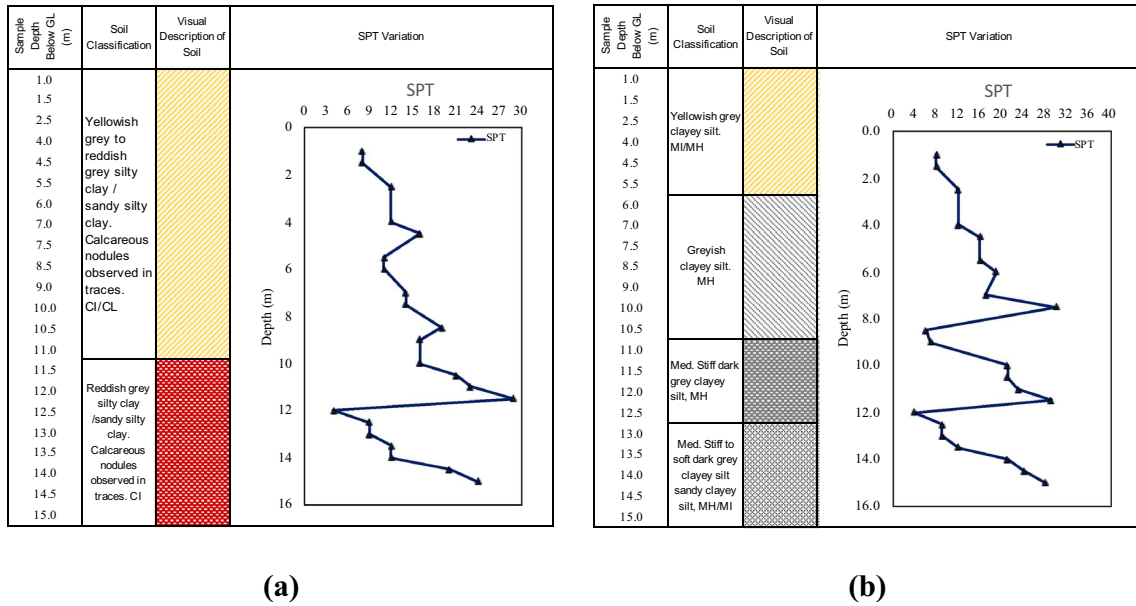
Fig. 4 Soil profile of the mid-hill road project portion in Gandaki Province (GP), Nepal
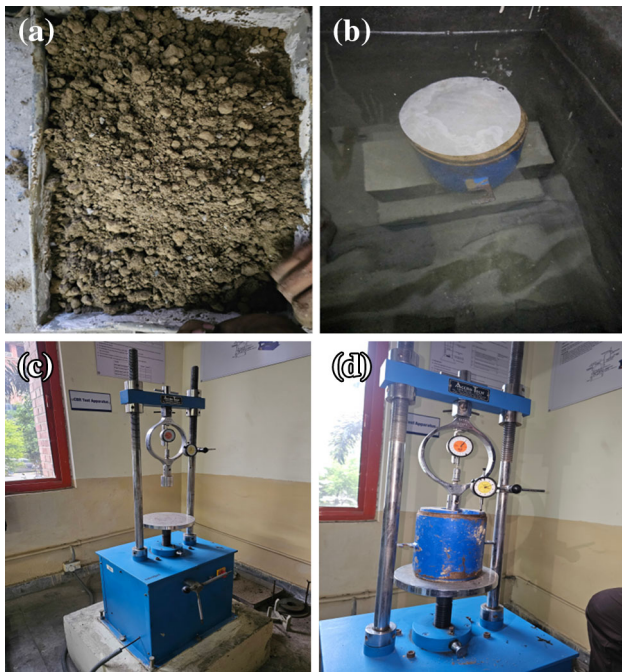

Fig. 5 Figure illustrating the experimental setup **a** soil sample, **b** soaked CBR sample, **c** CBR machine, **d** testing setup for CBR

Similarly, gravel particles improve soil stability and strength, hence increasing the soaking CBR value. Fine particles such as silt and clay can fill spaces between coarser particles, improving soil compaction and soaking CBR value. The amount of water in the soil is critical for its compactness and strength within specific limits, a larger water content can enhance the soaked CBR value. The I_P denotes the moisture content range in which the soil remains plastic. Soils with higher I_P values (that are more plastic) have higher wet CBR values. The Maximum *MDD* of soil reflects its maximum compaction. Higher *MDD* values, as a result of better compaction, contribute to higher soaking CBR values(Koti Marg and Puram 2012).

2.3 Sensitivity analysis of the soaked CBR based on experimental database

Sensitivity analysis (SA) is used to examine whether changes in the input parameters have an impact on the model's output. This can give feedback on which input parameters are the most important, and by deleting the less important ones, the input space can be shrunk, lowering the complexity of the model and cutting down on the amount of time needed for training. The popular cosine amplitude method (CAM) (Asteris et al. 2021) was chosen and employed to carry out the SA. The strength is checked from the range of 0–1 the values nearer to 1 are regarded as higher sensitive input variables. In CAM, data pairs are utilized to form a data array, $= x_1, x_2, x_3, \ldots, x_i, \ldots, x_n$. The variable xi in the array, x, is a length vector of n as

$$xi = x_{i1}, x_{i2}, x_{i3}, \ldots, x_{im} \tag{1}$$

The relationship between S_{ij} (strength of the relation) and data sets of x_i and x_j is presented by the following equation:

$$S_{ij} = \frac{\sum_{k=1}^{m} x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^{m} x_{ik}^2 \sum_{k=1}^{m} x_{jk}^2}} \tag{2}$$

Fig. 6 Geotechnical correlation matrix for sub-base soil parameters

	% Gravel	% Sand	%Fines (Silt-Clay)	W_L (%)	W_P (%)	I_P (%)	S_L (%)	MDD (Kg/m³)	OMC (%)	CBR (%)
% Gravel	1									
% Sand	-0.87	1.00								
%Fines (Silt-Clay)	0.14	-0.61	1.00							
W_L (%)	0.67	-0.63	0.20	1.00						
W_P (%)	0.63	-0.61	0.21	0.75	1.00					
I_P (%)	0.58	-0.53	0.14	0.91	0.52	1.00				
S_L (%)	-0.21	0.10	0.14	-0.35	-0.36	-0.31	1.00			
MDD (Kg/m²)	0.15	-0.17	0.09	0.10	0.07	0.09	-0.06	1.00		
OMC (%)	-0.52	0.53	-0.23	-0.51	-0.54	-0.39	0.28	-0.08	1.00	
CBR (%)	0.82	-0.78	0.24	0.80	0.78	0.67	-0.47	0.17	-0.64	1.00

Fig. 7 Effect of independent variables on CBR values of soil
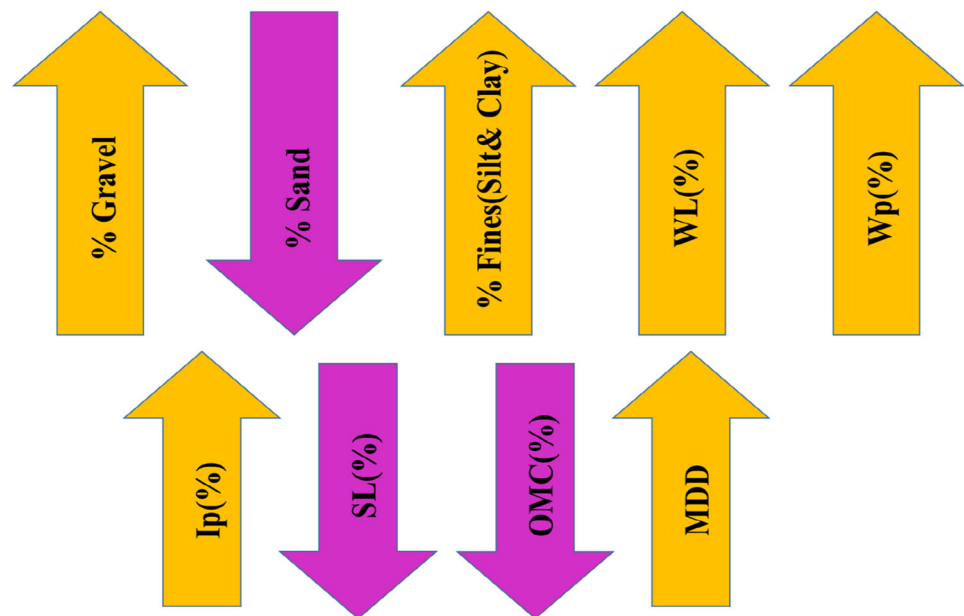
The S_{ij} values between the soaked CBR and the input parameters are shown in Fig. 8. This analysis reveals some surprising results as the *%G* has the greatest influence of 0.9610 followed by w_L (%) 0.9604. Interestingly, the *MDD* (kg/m³) comes 4th with a value of 0.9150, followed by the I_P*(%), %(M–C), OMC (%), S_L(%)* and *% S,* respectively.

2.4 Data division for computational analysis

Table 2 provides statistical information for each factor, including the maximum, minimum, mean, mode, median, standard deviation (SD), and variance. The highest and lowest values that were really measured are the maximum and minimum values. The mode is the value that occurs the most frequently, while the mean is the average value. When the data set is sorted, the median is the midway value. The variance is the square of the standard deviation, which measures how far the data set deviates from the mean. We can learn more about the distribution and variability of the data set by examining descriptive statistics. Making informed judgements and inferences regarding the data set can be helped by this. Statistics such as the maximum, minimum, mean, mode, median, SD, and variance are provided for each factor. The maximum and minimum values, respectively, are the highest and lowest numbers reported in the data set. The table's descriptive statistics can be used to draw conclusions about the data set and gain a better understanding of the variability and distribution of the data.
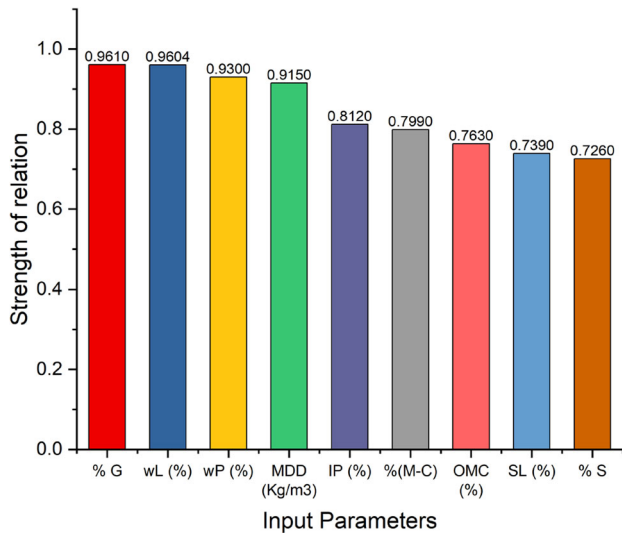
Fig. 8 Sensitivity analysis of soaked CBR with input parameters

The process of partitioning data sets into training (TR) and testing (TS) subsets is known as data division. In this work, the model was trained using about 70% of the whole data set and tested using the remaining 30%. The fundamental problem with machine learning modelling is that unless a model is evaluated on an unidentified, independent data set, it cannot be said how well it works or will perform. A constructed model that is built on TR data and has 100% accuracy or 0 error may not generalize to unobserved data and, as a result, may over-predict or under-predict the results which make the model untrustworthy. Since the goal of machine learning is a generalization, the model's accuracy and dependability can only be accepted in light of its performance during testing with data points that were not used during training. The descriptive statistics for the training and testing data sets are displayed in Table 2.

Similarly, Fig. 9 shows the illustrations of computational analysis applied in the present study. The descriptive statistic values for all sub-base soil properties are shown in Table 2. The obtained database, as can be seen, spans a wide variety of CBR values ranging from 36.26 to 10.38. The gravel content ranges from 1.36% to 68.02%, the sand percentage goes from 3.93% to 86.95%, and the fine content ranges from 0.05% to 55.49%. Similarly, the soil consistency result for the selected database shows that the w_L ranges from 3.52% to 65.26%, the w_P ranges from 0% to 30.01%, and the I_P ranges from 3.33% to 47.17%. Fine-grained soil MDD and OMC range from 2992.71 to 1905.66 kg/m^3 and 29.63–11.46%, respectively.

3 Results and discussion

3.1 Predictive performance of machine learning models

In this section, the predictive performance of ten machine learning models has been presented and elaborately discussed. Furthermore, graphical representations of the results that compare the predicted values with the laboratory-calculated values for both the training and testing data have been presented. By assessing the predictive performance of these models, the best and most accurate computational algorithm can be selected for future predictions eradicating the drawbacks associated with laboratory experiments.

3.2 Graphical realization of results

Figures 10 and 11 present a comparison between the CBR values of sub-base soil determined by laboratory-based tests and the values predicted by developed machine learning models. These machine-learning models are based on intelligent computational algorithms inspired by the structure and functioning of a human brain. Their main objective is to predict the CBR values of soil samples based on fundamental soil parameters. To achieve this prediction, the machine learning models were trained using a data set containing information about soil properties, such as $\%G$, $\%S$, $\%M–C$, w_L, w_P, I_P, S_L, MDD, and OMC, along with corresponding experimental CBR values.

Table 2 Values of descriptive statistics for different sub-base soil parameters

| Statistical values	% G	% S	% (M–C)	w_L (%)	w_P (%)	I_P (%)	S_L (%)	MDD (kg/m^3)	OMC (%)	CBR (%)
Max	68.02	86.85	55.49	65.26	30.01	47.17	9.71	2992.71	29.63	36.26
Min	1.36	3.93	0.05	3.52	0.00	3.33	1.68	1905.66	11.46	10.38
Mean	30.24	53.94	15.82	25.72	13.30	8.94	5.19	1960.10	15.45	21.69
Mode	47.00	60.00	22.00	0.00	0.00	0.00	6.37	0.00	0.00	0.00
Median	22.74	60.00	13.84	21.93	15.61	5.71	5.58	2000.91	15.74	26.52
SD	17.57	21.90	10.81	13.56	9.59	10.39	2.28	324.21	5.18	10.26
Variance	308.68	479.53	116.86	183.85	91.97	107.91	5.19	105,110.7	26.87	105.31
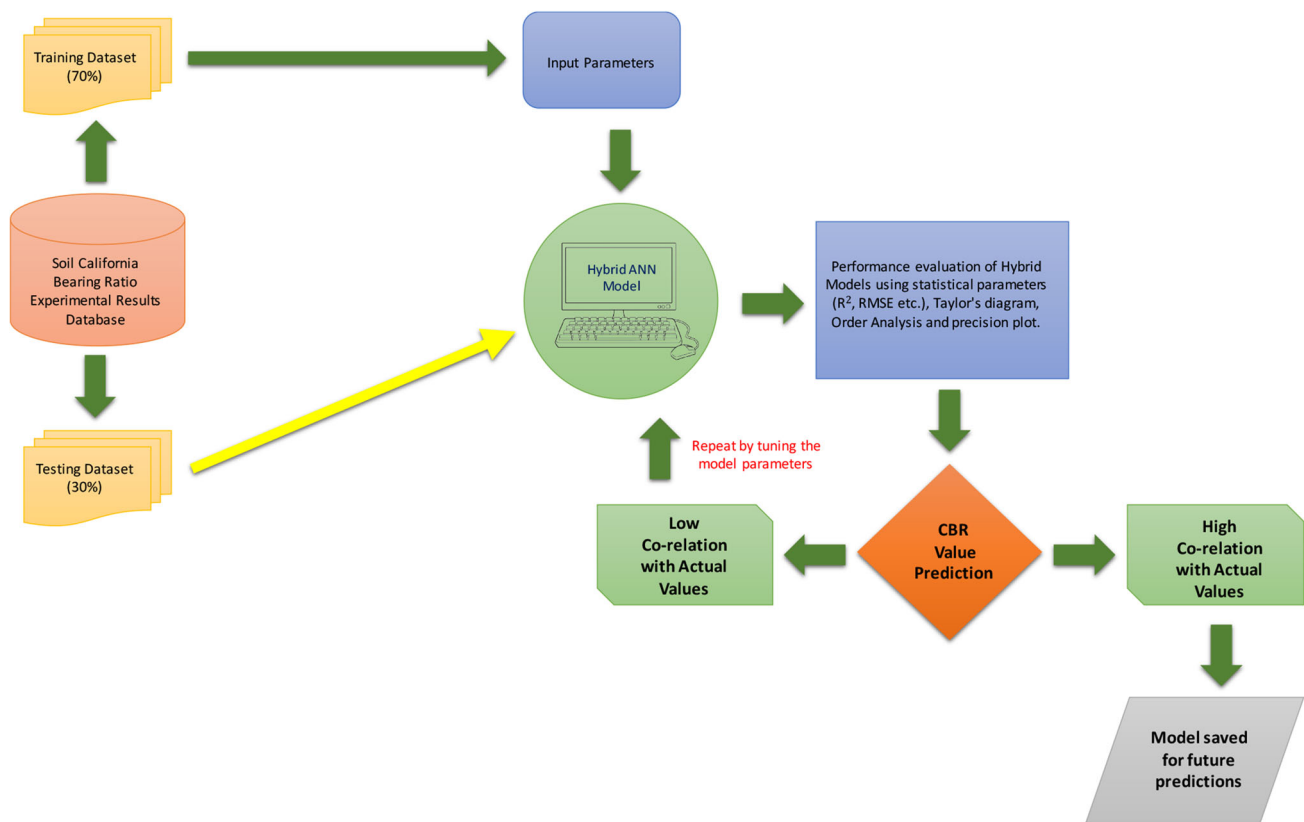
Fig. 9 Illustrations of computational analysis applied in the present study

The predicted CBR values estimated by the machine learning models are then compared to the actual CBR values obtained from laboratory-based tests. This comparison serves to assess the accuracy of the machine learning models in their ability to predict CBR values based on the input soil property. If the predicted values closely match the actual CBR values from laboratory tests, it indicates that the models have learned the underlying patterns in the data and are performing well. On the other hand, substantial differences between predicted and observed values would imply that the models may require further development or fine-tuning.

The laboratory-based values are plotted with a black colour pattern line. As shown in Fig. 10 for the training data ANN–GBO predicted CBR value with red colour and ANN–FF predicted CBR value with blue colour following the same pattern as the laboratory-based value for both the sub-base soil. Similarly, other models ANN–GWO, ANN–GA, ANN–PSO, ANN–TLBO, ANN–WOA, and ANN–IWO follow a near about laboratory pattern. Whereas ANN–CA and ANN–HHO models do not follow the laboratory pattern.

Similarly, Fig. 11 represents the actual and predicted CBR values obtained from the developed ten machine learning models for the testing data sets. The values from the

lab are represented with a black pattern line. For the testing data, as shown in Fig. 11 the projected CBR values for sub-base soil from ANN–GBO and ANN–FF are similar to the laboratory-based values in both red and blue color. Other models, including ANN–GWO, ANN–GA, ANN–PSO, ANN–TLBO, ANN–WOA, and ANN–IWO, also closely resemble the laboratory pattern. The ANN–CA and ANN–HHO models, however, do not adhere to the laboratory pattern.

Figures 12 and 13 show the comparative bar diagram for the ten computational models for training and testing data for soaked sub-base CBR concerning the ideal statistical realization value. As seen in the bar diagram of performance measurement indicators R^2, adj.R^2, R, MAE, MAPE, RMSE, VAF, I_P, IOA and IOS values for the ANN–GBO model are very near to the ideal values indicating the effectiveness of the GBO algorithm in optimizing ANN model.

3.3 Statistical realization of results

The statistical realization of results for the developed computational models to estimate soaked CBR for the sub-base is presented in this sub-section. The statistical features for the computational models are presented in Tables 5 and 6, for
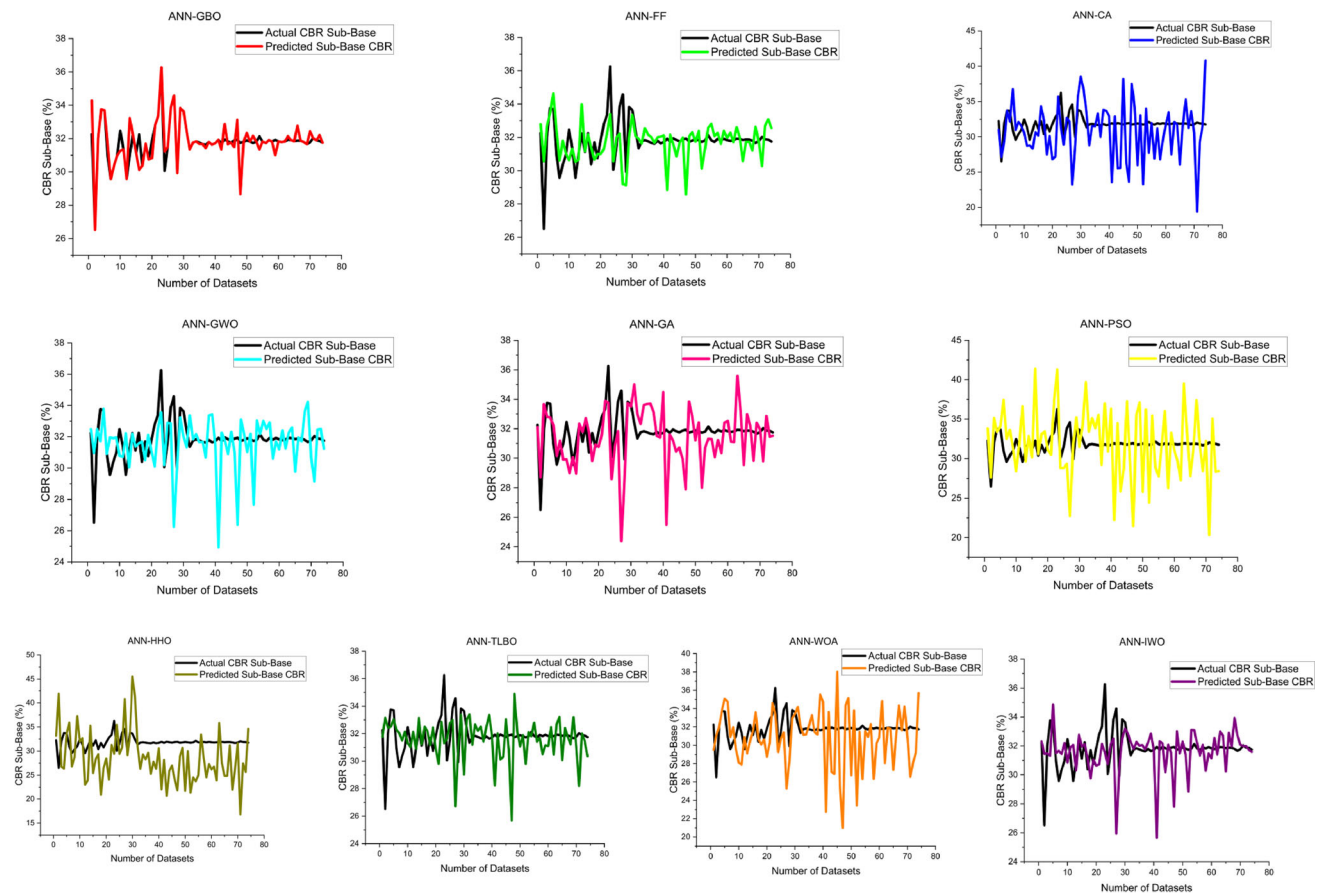
Fig. 10 Actual vs predicted soaked CBR values for sub-base training data for different ML

training and testing data, respectively, for the soaked CBR estimations. The abilities of the constructed models for the training and testing stages are shown here. It should be underlined that the training subset performance is used to define the goodness of fit of the developed models, while the testing data set is used to evaluate their generalization potential. The accuracy of each model was evaluated using a variety of statistical performance measures. R^2, adjusted R^2, adjusted R^2, coefficient of correlation (R), mean absolute error (MAE), mean absolute percentage error (MAPE), root-mean-square error (RMSE), variance accounted for (VAF), performance index (IP), Willmott's index of agreement (IOA), and index of scattering (IOS) are some of the commonly used performance measurement indicators (Bardhan et al. 2023; Ghani et al. 2023; Tenpe and Patel 2020).

Furthermore, Figs. 12 and 13 depict the error of the created models during the prediction stage in the training and testing stages, respectively. The best-selected model must be error-free or have the least feasible error for the practical implementation of these models. According to the error representation, the ANN–GBO model represents the best

error-free relationship in forecasting the soaking CBR with the maximum accuracy and the fewest computational errors (Figs. 14 and 15).

According to Table 3, it is seen that the developed ANN–GBO achieved the highest R^2 and lowest RMSE values of 0.997 and 0.033, respectively, during the training phase for soaked CBR prediction. On the contrary, the results of Table 4 exhibit that the developed ANN–FF ($R^2 = 0.956$ and RMSE = 0.105) and ANN–GBO ($R^2 = 0.978$ and RMSE = 0.109) models were found to be the top two models during the testing phase of soaked CBR estimation. According to the overall results of the CBR estimation in the training phase, the ANN–GBO was determined to be the best-fitted model with $R^2 = 0.997$ and RMSE = 0.033, followed by ANN–FF ($R^2 = 0.988$ and RMSE = 0.066). The developed ANN–HHO model was the least performing model, with $R^2 = 0.726$ (lowest among other developed models) and RMSE = 0.333 (highest among other developed models) in the training phase as well as $R^2 = 0.745$ (lowest among other developed models) and RMSE = 0.295 (highest among other developed models) in the testing phase. These findings demonstrate the
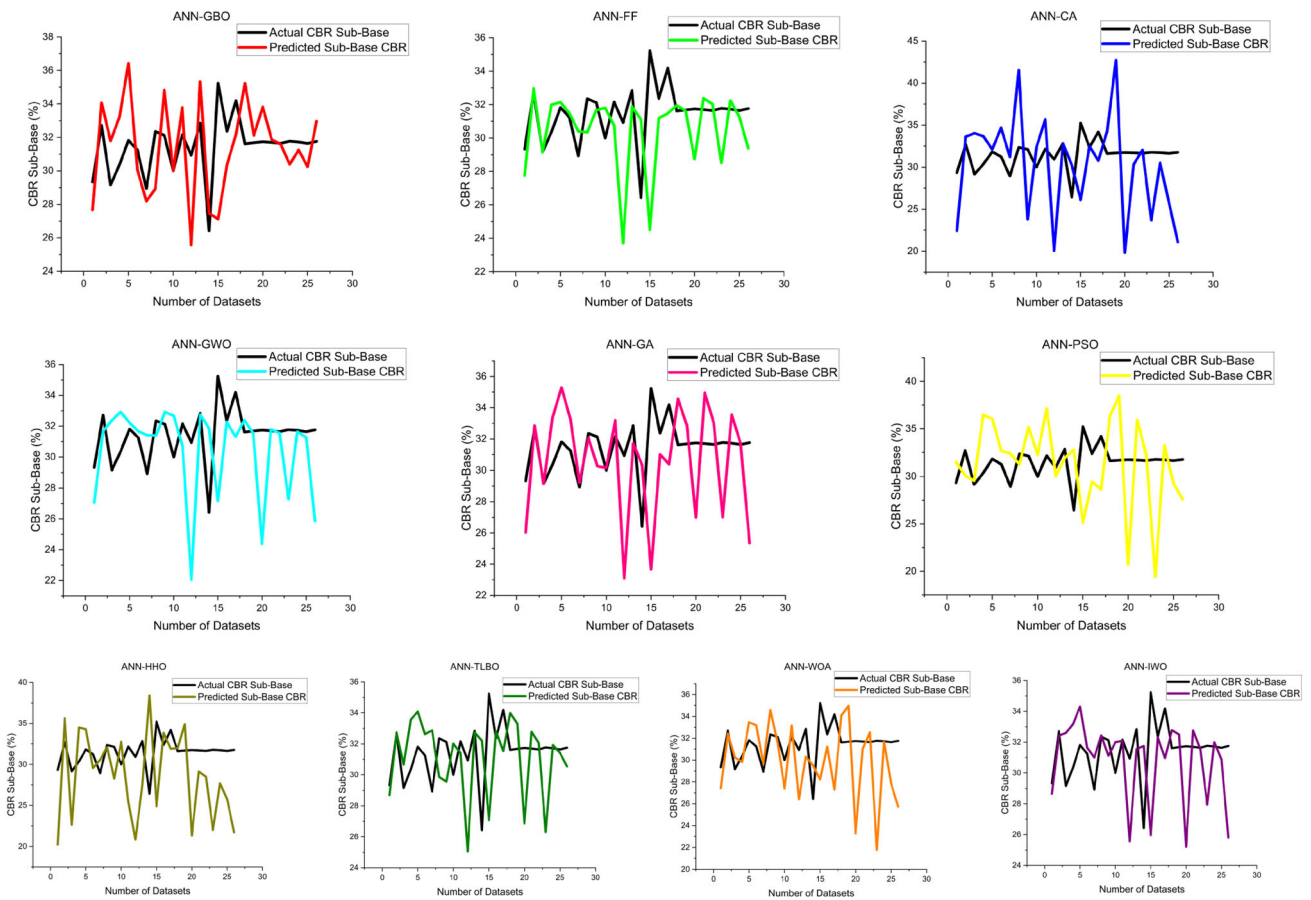
Fig. 11 Actual vs predicted soaked CBR values for sub-base testing data for different ML

good predictive performance of the suggested ANN–GBO and ANN–FF models in soaked CBR predictions.

3.4 Taylor's diagram

The Taylor diagram is used to quickly assess the accuracy of a model in terms of the coefficient of correlation, ratio of standard deviations, and RMSE index (Taylor 2001). Taylor's diagram representation is a modern-day technique to identify the precision of the machine learning model and justify the reliability of these approaches (Ghani et al. 2022a). In general, a point inside a Taylor diagram indicates a model. The position of the point should line up with the reference point for an ideal model. The Taylor diagrams for the computational models developed for soaked CBR predictions for

the training and testing phase are shown in Fig. 16a, respectively. As can be seen, ANN–GBO has been lined up with the reference line in both the training and testing phases, while ANN–HHO is very far away from the reference line in both the training and testing phase. The reliability of results for CBR prediction for ANN–GBO is significantly confirmed by Taylor's diagram.

Figure 17a, b shows the % change in the standard deviation of the ten computational models for the TR data set of which ANN–GBO has the lowest % change in standard deviation with a value of 0.204%, whereas the ANN–HHO has highest 16.1% change in standard deviation Similarly, ANN–PSO has − 3.27% which is below the reference line which is justified the above Taylor's diagram in Fig. 16a, where ANN–PSO is plotted outside the reference line.
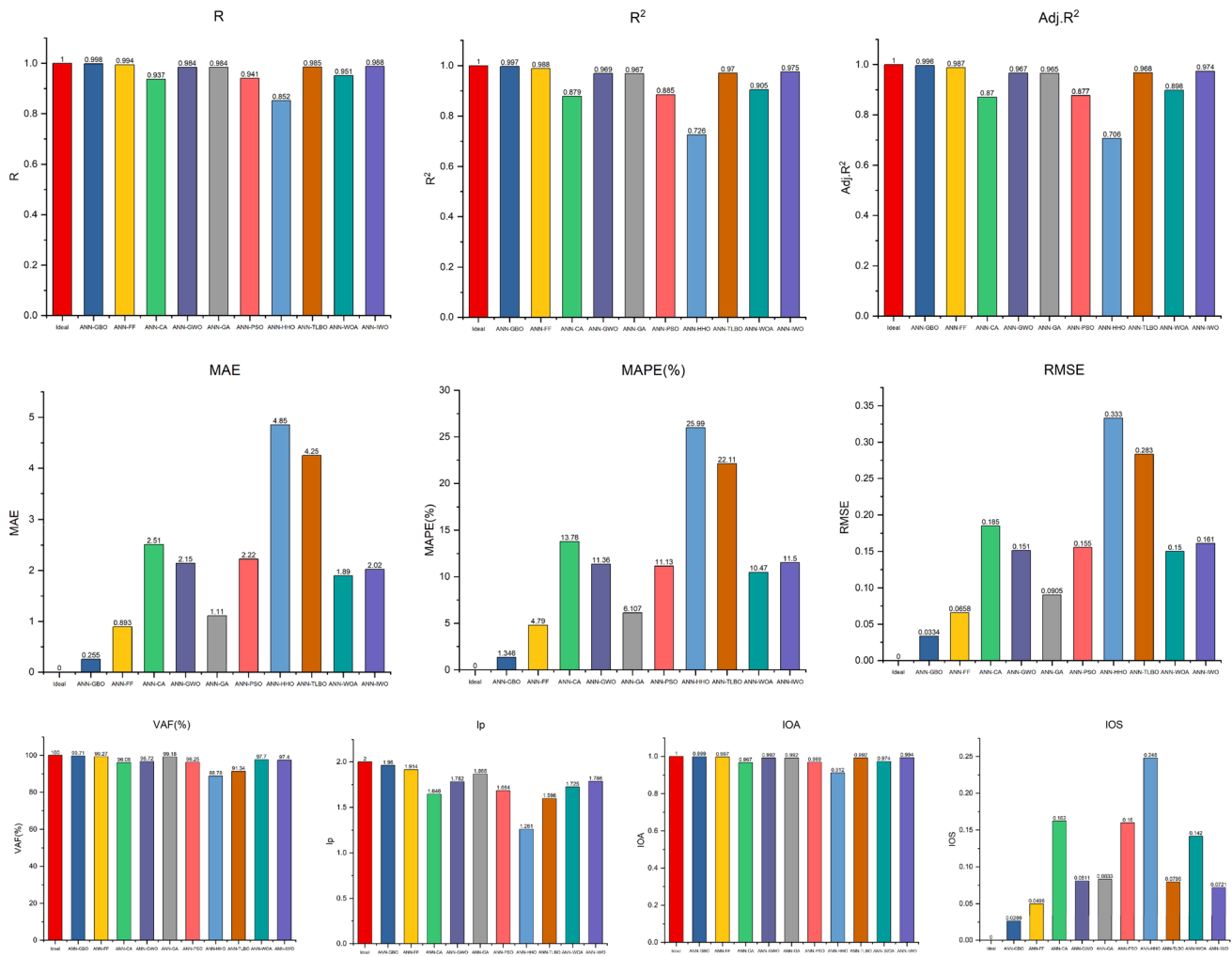
Fig. 12 Statistical realization diagram for soaked sub-base CBR values for TR data for different ML

Similarly, In Fig. 18a, b, the % change in the standard deviation is illustrated for ten computational models applied to the TS data set. Notably, the ANN–GBO model exhibits the lowest percentage change in standard deviation, standing at a mere 0.798%. Conversely, the ANN–HHO model demonstrates the highest percentage change in standard deviation, reaching a substantial 15.4%. Furthermore, the ANN–PSO and ANN–CA model displays a negative percentage change of − 0.62% and − 1.5%, respectively. This observation is particularly significant as it falls below the reference line. This finding substantiates the graphical representation in Fig. 16b, depicted in Taylor's diagram, where the ANN–PSO and ANN–CA model is situated outside the reference line.

3.5 Order analysis (OA)

Several graphical and statistical indices were created to assess the effectiveness of the best machine learning model,

and conclusions were drawn from the outcomes. However, this is a traditional strategy that could lead to inaccurate results. To validate the claims based on index parameters, "Order Analysis" can be carried out (Tables 5, 6). This is a very effective and well-organized method for quickly assessing a model's overall performance. According to OA, the model with the highest value in particular performance indices receives a maximum score of s (equivalent to the entire number of relevant models), a minimum score is given to the model with the lowest value, and the scores for the remaining intermediate models are given in ascending order. The OA of the training and testing stages for the 10 computational models are displayed in Tables 5 and 6, respectively. After obtaining the score from the OA for the training and testing stages the scores are summed to find the total score of OA which is shown in Table 7. Based on the total scores obtained in Table 7, the ANN–GBO model emerges as the top performer with a score of 198, showcasing its exceptional
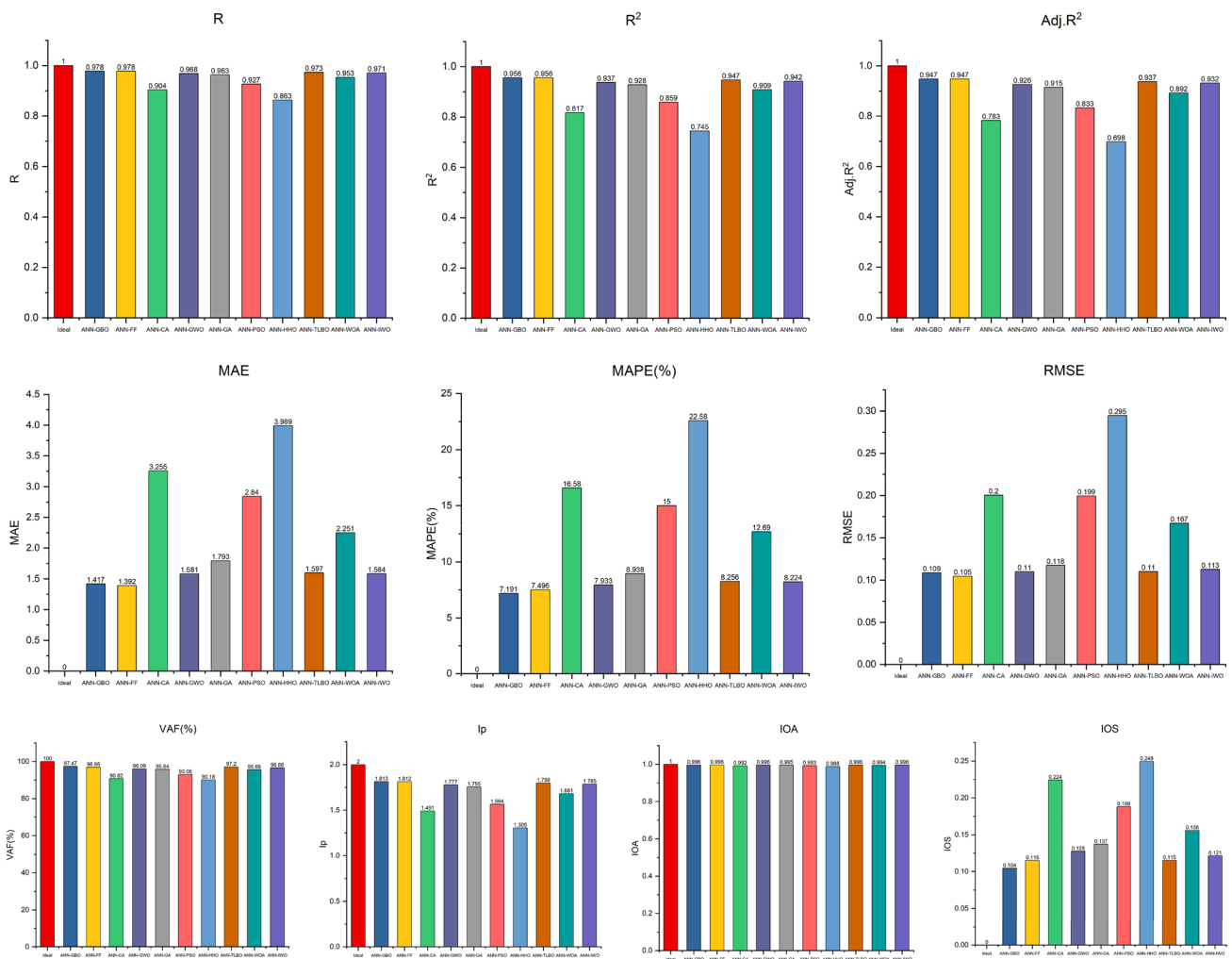
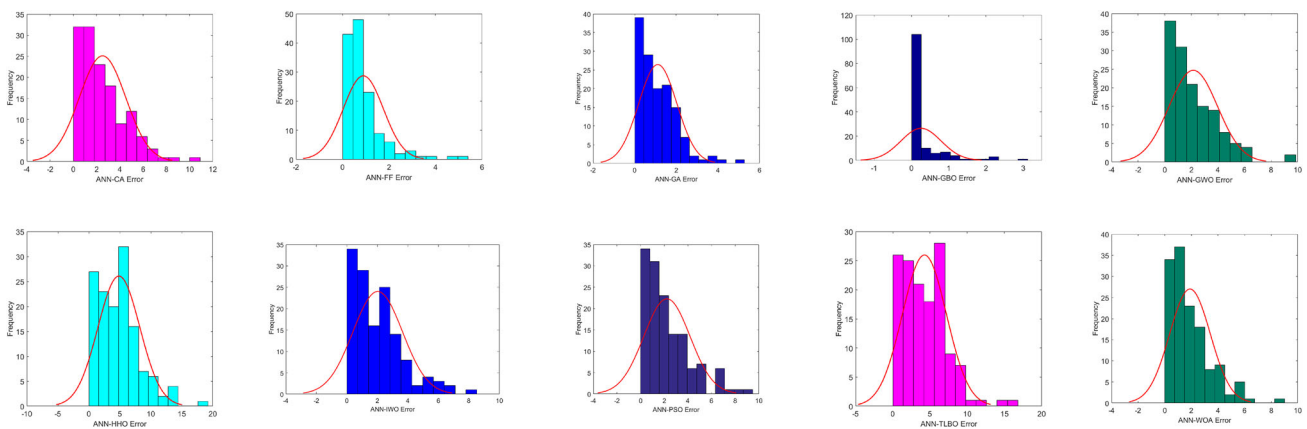**Fig. 13** Statistical realization diagram for soaked sub-base CBR values for TS data for different ML

Fig. 14 Error for predicted soaked sub-base CBR values for training data for different ML
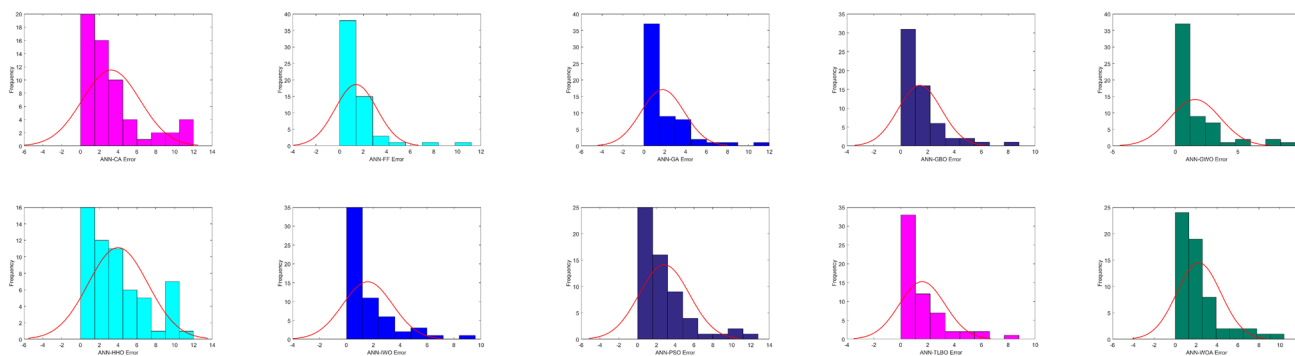
Fig. 15 Error for predicted soaked sub-base CBR values for testing data for different ML

effectiveness and versatility. The ANN–FF model secures the second position with a score of 186, highlighting its robustness and reliability. The IWO model ranks third with a score of 145, demonstrating its competence in addressing the problem. The GWO model grabs the fourth spot with a score of 130, while the GA model lands in the fifth position with a score of 127. The TLBO, WOA, PSO, CA, and HHA models claim the sixth to tenth positions, respectively. These rankings provide valuable insights into the performance of each model, although their suitability may vary depending on the specific problem and data set.

4 Discussion

In the preceding subsections, ten different machine learning techniques were utilized to assess the prediction of soaking CBR values for the sub-base. Through Pearson correlation tests, it was determined that certain input parameters—namely, *%S, %(M–C), w_L, I_P, MDD*, and *OMC*—held significance in predicting the soil's soaked CBR value. Furthermore, nine different input parameters were employed to develop computational models, and the performance of these models was evaluated using statistical measures. After conducting a comprehensive analysis of the models' performance on training data across various accuracy parameters, the ANN–GBO model emerged as the dominant performer. This particular model consistently displayed exceptional results in key metrics, including R, R^2, MAE, MAPE, RMSE, VAF, I_p, and IOA, surpassing other models. The ANN–GBO model exhibited a remarkable R value of 0.998, signifying a superior fit to testing data, coupled with a notably low MAE of 0.255, indicating minimal prediction errors. Furthermore, the model's impressive R^2 value of 0.997 underscored its ability to explain a significant portion of the variance in the dependent variable. A high VAF of 99.71% further validated the model's capacity to account for variance in the data.

In addition, the ANN–GBO model excelled in terms of MAPE (1.35%) and RMSE (0.033), showcasing its precision

in predicting values with minimal error. The model's remarkable agreement with observed data was further shown by its I_p of 1.96 and IOA of 0.999, indicating its appropriateness for both modelling and prediction applications. It became clear that the ANN–FF model consistently outperformed its competitors after a thorough review of several accuracy factors for the performance of various models on testing data. The ANN–FF model consistently demonstrated superior outcomes in key metrics, such as R, R^2, MAE, MAPE, RMSE, VAF, I_p, and IOA. Impressive R and R^2 values of 0.978 and 0.956, respectively, showcased the model's remarkable data fitting capabilities and ability to explain variance in testing data. Moreover, the ANN–FF model's low MAE of 1.392 and MAPE of 7.19% highlighted its accuracy in predicting values with minimal errors. Notably, its RMSE of 0.105 emphasized its precision in minimizing root mean squared errors. The model achieved a VAF of 97.47%, reflecting its capability to account for variance in testing data, along with an I_p of 1.81 and an IOA of 0.996, affirming its strong agreement with observed values.

5 Summary and conclusion

The California bearing ratio (CBR) is an important statistic for determining the thickness of subgrade layers in many civil engineering projects. Laboratory testing is often performed on samples of compacted soil that have been soaked, a process that takes a significant 4 days and costs a lot of money. Consequently, this study was initiated to replace the time-consuming work of conducting physical laboratory tests with the implementation of efficient machine-learning models. These models are designed to predict CBR values by leveraging the existing experimental database, offering a more streamlined and cost-effective alternative. It is worth noting that achieving precise and dependable estimations of the soaked California bearing ratio (CBR) can lead to significant time and cost savings compared to conducting traditional

Table 3 Training data performance parameters

	Ideal index	ANN–GBO	ANN–FF	ANN–CA	ANN–GWO	ANN–GA	ANN–PSO	ANN–HHO	ANN–TLBO	ANN–WOA	ANN–IWO
R	1	0.998	0.994	0.937	0.984	0.984	0.941	0.852	0.985	0.951	0.988
R^2	1	0.997	0.988	0.879	0.969	0.967	0.885	0.726	0.970	0.905	0.975
RMSE	0	0.033	0.066	0.185	0.151	0.090	0.155	0.333	0.283	0.150	0.161
Adj.$R2$	1	0.996	0.987	0.870	0.967	0.965	0.877	0.706	0.968	0.898	0.974
MAE	0	0.255	0.893	2.512	2.145	1.111	2.224	4.850	4.251	1.894	2.022
I_p	2	1.96	1.91	1.65	1.78	1.87	1.68	1.26	1.60	1.73	1.79
IOA	1	0.999	0.997	0.967	0.992	0.992	0.969	0.912	0.992	0.974	0.994
IOS	0	0.027	0.050	0.162	0.081	0.083	0.160	0.248	0.080	0.142	0.072
MAPE (%)	0	1.35%	4.79%	13.78%	11.36%	6.11%	11.13%	25.99%	22.11%	10.47%	11.50%
VAF (%)	100	99.71	99.27	96.08	96.72	99.18	96.25	88.78	91.34	97.70	97.40

Table 4 Testing data performance parameters

|	Ideal index	ANN–GBO	ANN–FF	ANN–CA	ANN–GWO	ANN–GA	ANN–PSO	ANN–HHO	ANN–TLBO	ANN–WOA	ANN–IWO
R	1	0.978	0.978	0.904	0.968	0.963	0.927	0.863	0.973	0.953	0.971
R^2	1	0.956	0.956	0.817	0.937	0.928	0.859	0.745	0.947	0.909	0.942
RMSE	0	0.109	0.105	0.200	0.110	0.118	0.199	0.295	0.110	0.167	0.113
Adj.R^2	1	0.947	0.947	0.783	0.926	0.915	0.833	0.698	0.937	0.892	0.932
MAE	0	1.417	1.392	3.255	1.581	1.793	2.840	3.989	1.597	2.251	1.584
I_p	2	1.81	1.81	1.49	1.78	1.76	1.56	1.30	1.80	1.68	1.79
IOA	1	0.996	0.996	0.992	0.996	0.995	0.993	0.988	0.996	0.994	0.996
IOS	0	0.104	0.115	0.224	0.128	0.137	0.188	0.249	0.115	0.156	0.121
MAPE (%)	0	7.19%	7.50%	16.58%	7.93%	8.94%	15.00%	22.58%	8.26%	12.69%	8.22%
VAF (%)	100	97.47	96.95	90.82	96.09	95.84	93.06	90.18	97.20	95.66	96.66
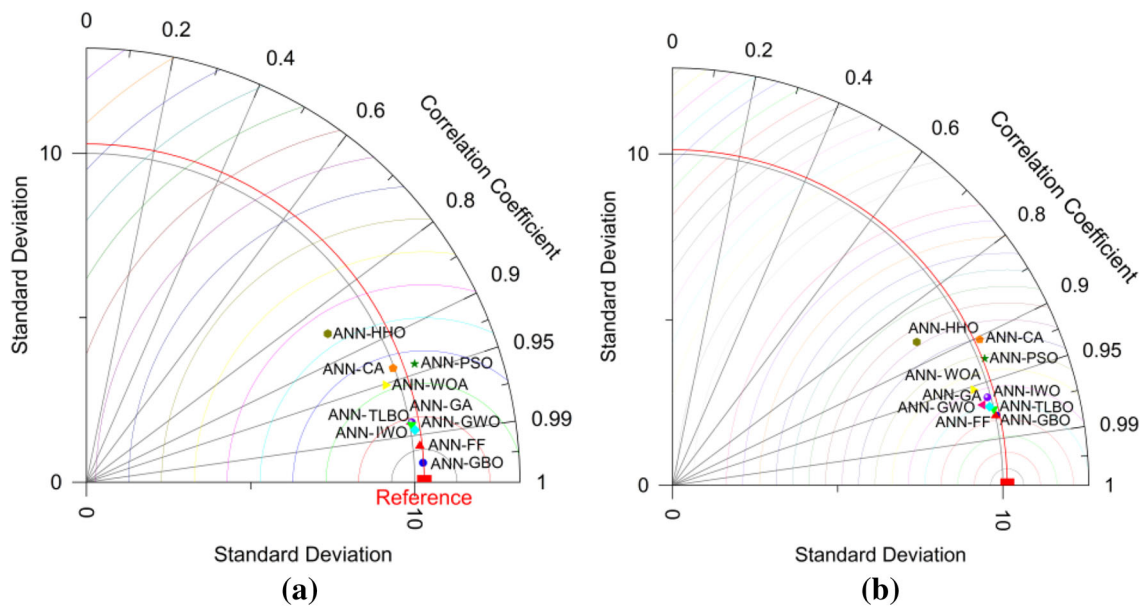
Fig. 16 Taylor's diagram TR (**a**) and TS (**b**) for soaked CBR
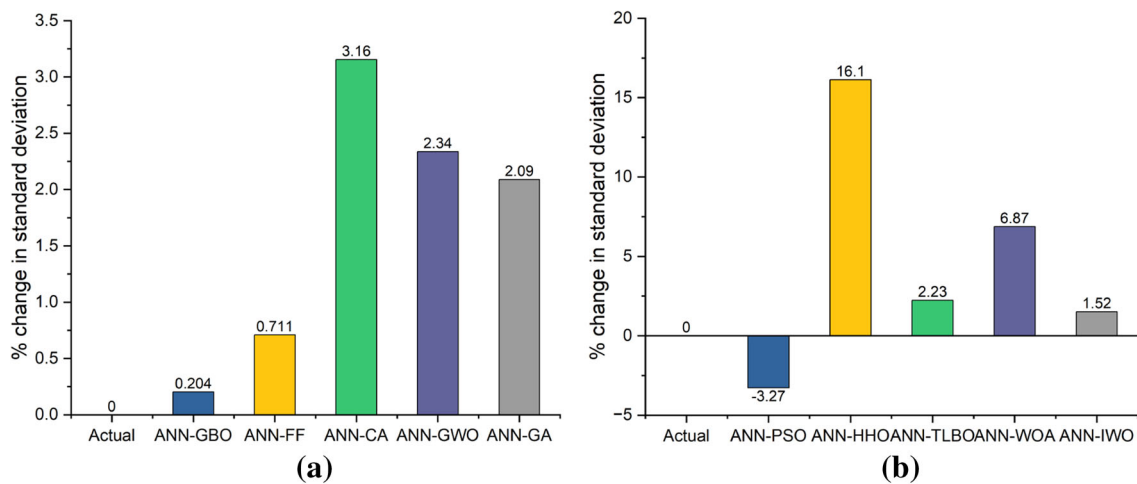

Fig. 17 % change in standard deviation for TR (**a**) and TR (**b**) for soaked CBR

laboratory tests. To achieve this goal, soil samples were collected from the deposit of fine-grained soils in a project currently underway to construct the Mid-Hill Highway in Nepal Post the experimental analysis, an effective machine-learning system was developed using the laboratory obtained CBR values as its foundation. Our method integrates artificial neural networks (ANN) and nature-inspired optimisation approaches in a novel way. Gradient-based optimisation (GBO), firefly algorithm (FF), cultural algorithms (CA), grey wolf optimisation (GWO), genetic algorithm (GA), particle swarm optimisation (PSO), Harris Hawk optimisation (HHO), teaching learning-based optimisation (TLBO), whale optimisation algorithm (WOA), and invasive weed optimisation (IWO) were the algorithms we used in the

ANN-based optimisation modelling. For the development of the CBR prediction models, we utilized a comprehensive data set comprising 100 experimental results. To ensure the robustness of the proposed models, a detailed comparative study was employed. The test results demonstrate that the proposed ANN–GBO model achieved the highest level of accuracy, with an $R^2 = 0.997$, VAF = 99.71, RMSE = 0.033 in the training stage and $R^2 = 0.956$, VAF = 97.47, RMSE = 0.109 in the testing phase. In addition, the results of the ANN–GBO model are significantly better than those obtained by other ANN-based models and are primarily characterized by a faster convergence rate and higher accuracy. The model not only adeptly fit the data and minimized errors but also demonstrated accurate predictive capabilities.
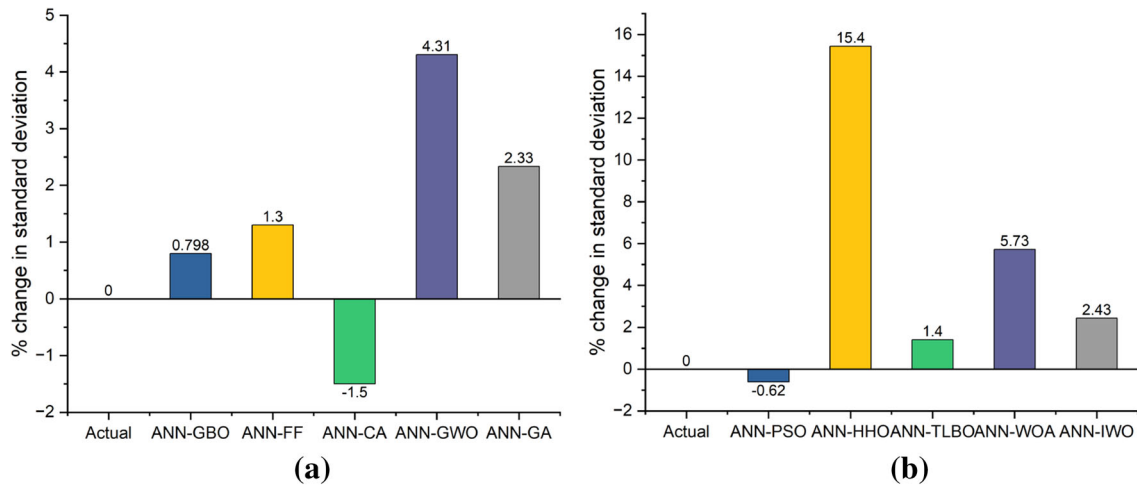
Fig. 18 % change in standard deviation for TS (**a**) and TS (**b**) for soaked CBR

In addition, the results of the present study strongly indicate that the GBO algorithm excels as the optimal choice for precise prediction and data fitting among all the models scrutinized. Considering the statistical and graphical results, the proposed ANN–GBO model provides a new alternative to estimate the CBR of soaked soils using only the basic soil properties. Researchers and practitioners operating in

the domain of highway and soil engineering should seriously consider adopting this algorithm for predictive tasks. Its exceptional performance across a spectrum of accuracy metrics and consistent outperformance of alternative models make it a compelling choice for predicting the CBR of soil in different engineering projects.

Table 5 Order analysis training MLM

	ANN–GBO	ANN–FF	ANN–CA	ANN–GWO	ANN–GA	ANN–PSO	ANN–HHO	ANN–TLBO	ANN–WOA	ANN–IWO
R	10	10	3	7	6	4	2	9	5	8
R^2	10	10	3	7	6	4	2	9	5	8
Adj.R^2	10	10	3	7	6	4	2	9	5	8
MAE	9	10	2	6	5	3	1	7	4	8
MAPE (%)	10	9	2	8	5	3	1	6	4	7
RMSE	9	10	3	8	6	4	2	8	5	7
VAF (%)	10	8	2	6	3	4	1	9	5	7
I_p	10	10	3	7	6	4	2	9	5	8
IOA	10	10	6	10	9	7	5	10	8	10
IOS	10	9	2	6	5	3	1	8	4	7
Total score	98	96	29	72	57	40	19	84	50	78

Table 6 Order analysis testing MLM

	ANN–GBO	ANN–FF	ANN–CA	ANN–GWO	ANN–GA	ANN–PSO	ANN–HHO	ANN–TLBO	ANN–WOA	ANN–IWO
R	10	9	3	6	6	5	2	7	4	8
R^2	10	9	3	6	6	5	2	7	4	8
Adj.R^2	10	9	3	6	6	5	1	2	7	8
MAE	10	9	3	5	8	4	1	2	7	6
MAPE (%)	10	9	3	5	8	6	1	2	7	4
RMSE	10	9	3	6	8	5	1	2	7	4
VAF (%)	10	9	3	5	8	4	1	2	7	6
I_p	10	9	3	6	8	4	1	2	5	7
IOA	10	9	4	7	7	5	3	7	6	8
IOS	10	9	2	6	5	3	1	7	4	8
Total score	100	90	30	58	70	46	14	40	58	67

Table 7 Total score of order analysis of training and testing MLM

	ANN–GBO	ANN–FF	ANN–CA	ANN–GWO	ANN–GA	ANN–PSO	ANN–HHO	ANN–TLBO	ANN–WOA	ANN–IWO
Training Score	100	90	30	58	70	46	14	40	58	67
Testing score	98	96	29	72	57	40	19	84	50	78
Total score	198	186	59	130	127	86	33	124	108	145

Declarations

Appendix

Following are the steps provided for the supplementary code: See Table 8

Table 8 Normalized input parameters and experimental CBR training data set

%G	%S	% (M–C)	W_p	W_L	I_p	S_L	MDD (kg/m^3)	OMC (%)
0.534	0.417	0.44	0.751	0.737	0.666	0.493	0.448	0.719
0.946	0.202	0.336	0.194	0	0.581	0.358	1	0.441
0.616	0.309	0.514	0.498	0.813	0.364	0.723	0.326	0.269
0.575	0.288	0.586	0.441	0.614	0.434	0.539	0	0.138
0.754	0.324	0.353	0.851	0.407	0.976	0.826	0.221	0.312
0.58	0	1	0.468	0.709	0.401	0.684	0.526	0.21
0.615	0.329	0.486	0.13	0.758	0.032	0.485	0.602	0.29
0.461	0.283	0.71	0.465	0.787	0.348	0.519	0.658	0.287
0.87	0.257	0.333	0.217	0.764	0.116	0.623	0.307	0.827
0.743	0.389	0.27	0.33	0.818	0.193	0.787	0.308	0.723
0.746	0.309	0.382	0.1	0.601	0.103	0.235	0.5	1
0.565	0.363	0.488	0.197	0.858	0.035	0.718	0.643	0.068
0.368	0.345	0.714	0.101	0.549	0.137	0.381	0.463	0.263
0.793	0.178	0.525	0.557	0.687	0.503	0.849	0.486	0.656
0.46	0.383	0.565	0.175	0.737	0.091	0.122	0.559	0.411
0.54	0.102	0.892	0.412	0.807	0.282	0.73	0.385	0.077
0.719	0.407	0.268	0.275	0.489	0.349	0.405	0.622	0.443
0.732	0.492	0.132	0.318	0.644	0.293	1	0.516	0.187
0.647	0.357	0.414	0.271	0.742	0.183	0.675	0.62	0.282
0.439	0.324	0.673	0.448	0.808	0.317	0.719	0.665	0.556
0.32	0.33	0.785	0.436	0.776	0.326	0.702	0.225	0.466
0.777	0.342	0.303	0.799	0.657	0.764	0.426	0.522	0.392
0.718	0.342	0.364	0.985	1	0.731	0.759	0.156	0.366
0.705	0.209	0.569	0.066	0.463	0.157	0.48	0.555	0.661
0.911	0.439	0.026	0.304	0.626	0.291	0.593	0.416	0.567
0.919	0.412	0.058	0.331	0.581	0.346	0.492	0.268	0.637
0.501	0.276	0.679	0.162	0.15	0.453	0.695	0.215	0.774
0.471	0.194	0.829	0.241	0.806	0.112	0.757	0.838	0.608
0.961	0.273	0.217	0.386	0.828	0.243	0.662	0.157	0.527
0.58	0.072	0.894	1	0.603	1	0.396	0.733	0.548
1	0.081	0.457	0.52	0.534	0.565	0.402	0.581	0.611
0.955	0.278	0.217	0.577	0.807	0.447	0.753	0.704	0.4
0.559	0.445	0.375	0.137	0.748	0.045	0.065	0.748	0.402
0.575	0.591	0.146	0.37	0.728	0.291	0.034	0.411	0.575
0.549	0.516	0.281	0.322	0.682	0.272	0.025	0.312	0.49
0.59	0.477	0.298	0.399	0.717	0.327	0.025	0.315	0.474
0.457	0.393	0.553	0.173	0.607	0.171	0.015	0.197	0.54
0.516	0.657	0.111	0.505	0.883	0.327	0.073	0.836	0.407
0.285	0.81	0.122	0.788	0.68	0.739	0.014	0.191	0.39
0.878	0.371	0.159	0.312	0.566	0.336	0.035	0.417	0.368
0.074	1	0.061	0.264	0.62	0.254	0.045	0.529	0.757
0.542	0.588	0.185	0.328	0.807	0.199	0.023	0.288	0.288
0.061	0.884	0.242	0.31	0.781	0.197	0.057	0.656	0.246

Table 8 (continued)

%G	%S	% (M–C)	W_p	W_L	I_p	S_L	MDD (kg/m^3)	OMC (%)
0.209	0.589	0.519	0.216	0.539	0.259	0.074	0.847	0.379
0.524	0.582	0.212	0.72	0.776	0.61	0.023	0.284	0.187
0.395	0.759	0.085	0.141	0.825	0.001	0.046	0.537	0.74
0.1	0.992	0.046	0.016	0.631	0	0.06	0.691	0.291
0.628	0.655	0.001	0.82	0.855	0.659	0.013	0.185	0.584
0.791	0.514	0.039	0.402	0.718	0.329	0.044	0.519	0.327
0.136	0.868	0.189	0.318	0.823	0.178	0.023	0.293	0.521
0.715	0.549	0.066	0.327	0.687	0.275	0.052	0.609	0.343
0.133	0.885	0.167	0.241	0.712	0.173	0.04	0.472	0.746
0.34	0.595	0.379	0.47	0.941	0.255	0.048	0.555	0.359
0.287	0.867	0.037	0.305	0.951	0.084	0.041	0.487	0.628
0.223	0.93	0.011	0.603	0.684	0.552	0.04	0.473	0.382
0.161	0.735	0.357	0.667	0.433	0.776	0.063	0.726	0.723
0.618	0.587	0.109	0.229	0.873	0.057	0.056	0.652	0.327
0.283	0.705	0.278	0.2	0.569	0.222	0.041	0.484	0.494
0.641	0.534	0.163	0.138	0.628	0.123	0.049	0.571	0.505
0.615	0.492	0.25	0.113	0.47	0.2	0.018	0.237	0
0.308	0.733	0.211	0.539	0.632	0.521	0	0.039	0.248
0.395	0.634	0.268	0.074	0.638	0.053	0.036	0.434	0.333
0.938	0.408	0.045	0.447	0.819	0.309	0.061	0.704	0.714
0.267	0.706	0.292	0.621	0.688	0.566	0.029	0.359	0.586
0.39	0.379	0.641	0	0.442	0.104	0.04	0.479	0.474
0.267	0.604	0.44	0.435	0.768	0.33	0.017	0.22	0.514
0.713	0.351	0.355	0.356	0.895	0.17	0.037	0.439	0.37
0	0.708	0.56	0.699	0.869	0.529	0.038	0.452	0.382
0.285	0.895	0	0.906	0.671	0.862	0.006	0.104	0.611
0.766	0.33	0.333	0.163	0.378	0.308	0.059	0.679	0.55
0.506	0.3	0.55	0.282	0.454	0.192	0.832	0.59	0
0.695	0.235	0.432	0.717	0	0.909	0.399	0.348	0.413
0.352	0.238	0.793	0.507	0.893	0.29	0.284	0.712	0.248
1	0.185	0.169	0.823	0.334	0.903	0.64	0.753	0.413
0.956	0.282	0.094	0.78	0.006	0.984	0.178	0.538	0.506
0.807	0.162	0.404	0.318	0.362	0.272	0.373	0.32	0.011
0.817	0.171	0.382	0.212	0.017	0.282	0.667	0.713	0.053
0.441	0	1	0.685	0.784	0.553	0.389	0.49	0.201
0.425	0.356	0.565	1	0.774	0.944	0.649	0.637	0.006
0.855	0.357	0.106	0.588	0.848	0.408	0.683	0.63	0.51
0.498	0.04	0.888	0.433	0.693	0.28	0.663	0.233	0.088
0.52	0.288	0.55	0.254	0.463	0.152	1	0.606	0.144
0.762	0.382	0.173	0.708	0.344	0.759	0.544	0.543	0.247
0.788	0.017	0.608	0.451	0.569	0.352	0.651	1	0.455
0.592	0.159	0.636	0	0.069	0	0.581	0	0.346
0.983	0.2	0.17	0.388	0.453	0.321	0.856	0.307	0.4

Table 8 (continued)

%G	%S	% (M–C)	W_p	W_L	I_p	S_L	MDD (kg/m^3)	OMC (%)
0.729	0.35	0.25	0.594	0.657	0.492	0.658	0.325	0.606
0.637	0.485	0.176	0.859	0.205	1	0.046	0.519	0.381
0.74	0.207	0.418	0.854	1	0.673	0.022	0.287	0.362
0.066	1	0.132	0.551	0.216	0.618	0.04	0.465	0.388
0.692	0.393	0.235	0.469	0.605	0.36	0.021	0.277	0.542
0.53	0.469	0.31	0.555	0.214	0.623	0.036	0.429	0.157
0.303	0.821	0.106	0.149	0.272	0.101	0.025	0.315	0.174
0.64	0.436	0.234	0.412	0.411	0.368	0.028	0.345	0.266
0.264	0.61	0.415	0.615	0.245	0.684	0.055	0.606	0.412
0.228	0.762	0.26	0.6	0.304	0.642	0.054	0.599	1
0	0.914	0.311	0.581	0.705	0.457	0	0.083	0.072
0.586	0.553	0.144	0.459	0.602	0.349	0.068	0.728	0.439
0.448	0.783	0	0.486	0.523	0.413	0.063	0.682	0.123
0.407	0.125	0.878	0.686	0.333	0.736	0.053	0.584	0.177

1) Table 2 lists the parameter's minimum and maximum values.
2) Place the normalized data in an Excel file named "Input" with the sheet name "M" as follows:
   % G
   % S
   %(M-C)
   w_L (%)
   w_P (%)
   I_P (%)
   S_L (%)
   MDD (Kg/m3)
   OMC (%)
3) Run the. m file named "*CBR*"
4) The results, i.e., CBR, will be saved in an Excel file named
   "Soaked CBR"
5) **MATLAB Program:**

```
clc;
clear;
close all;
data=xlsread('Inputs', 'M');
Inputs=data';
%%
load('ANN-GBO-Model');
%%
Predict= net_f(Inputs);
xlswrite('Soaked CBR', Predict)
```

References

Alam SK, Mondal A, Shiuly A (2020) Prediction of CBR value of fine grained soils of Bengal Basin by genetic expression programming, artificial neural network and krigging method. J Geol Soc India 95(2):190–196. https://doi.org/10.1007/s12594-020-1409-0

Amin MN, Iqbal M, Ashfaq M, Salami BA, Khan K, Faraz MI, Alabdullah AA, Jalal FE (2022) Prediction of strength and CBR characteristics of chemically stabilized coal gangue: ANN and random forest tree approach. Materials. https://doi.org/10.3390/ma15124330

Asteris PG, Skentou AD, Bardhan A, Samui P, Pilakoutas K (2021) 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

Baghbani A, Choudhury T, Costa S, Reiner J (2022) Application of artificial intelligence in geotechnical engineering: A state-of-the-art review. Earth-Sci Rev 228:103991. https://doi.org/10.1016/j.earscirev.2022.103991

Bardhan A, Gokceoglu C, Burman A, Samui P, Asteris PG (2021a) Efficient computational techniques for predicting the California bearing ratio of soil in soaked conditions. Eng Geol 291:106239. https://doi.org/10.1016/j.enggeo.2021.106239

Bardhan A, Samui P, Ghosh K, Gandomi AH, Bhattacharyya S (2021b) 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

Bardhan A, Singh RK, Ghani S, Konstantakatos G, Asteris PG (2023) Modelling soil compaction parameters using an enhanced hybrid intelligence paradigm of ANFIS and improved grey wolf optimiser. Mathematics 11(14):3064. https://doi.org/10.3390/math11143064

Bui QAT, Al-Ansari N, Le HV, Prakash I, Pham BT (2022) Hybrid model: teaching learning-based optimization of artificial neural network (TLBO-ANN) for the prediction of soil permeability coefficient. Math Probl Eng. https://doi.org/10.1155/2022/8938836

Ceryan N, Samui P (2020) Application of soft computing methods in predicting uniaxial compressive strength of the volcanic rocks with different weathering degree. Arab J Geosci 13(7):288. https://doi.org/10.1007/s12517-020-5273-4

Coello Coello CA, Becerra RL (2004) Efficient evolutionary optimization through the use of a cultural algorithm. Eng Optim 36(2):219–236. https://doi.org/10.1080/03052150410001647966

Dababneh O, Kipouros T, Whidborne JF (2018) Application of an efficient gradient-based optimization strategy for aircraft wing structures. Aerospace. https://doi.org/10.3390/aerospace5010003

Erzin Y, Turkoz D (2016) Use of neural networks for the prediction of the CBR value of some Aegean sands. Neural Comput Appl 27(5):1415–1426. https://doi.org/10.1007/s00521-015-1943-7

Ghani S, Kumari S (2023) Plasticity-based liquefaction prediction using support vector machine and adaptive neuro-fuzzy inference system. In: Muthukkumaran-Kasinathan KS, Ayothiraman R (eds) Soil dynamics, earthquake and computational geotechnical engineering. Springer Nature, Singapore, pp 515–527

Ghani S, Kumari S, Ahmad S (2022a) Prediction of the seismic effect on liquefaction behavior of fine-grained soils using artificial intelligence-based hybridized modeling. Arab J Sci Eng 47(4):5411–5441. https://doi.org/10.1007/s13369-022-06697-6

Ghani S, Kumari S, Jaiswal S, Sawant VA (2022b) Comparative and parametric study of AI-based models for risk assessment against soil liquefaction for high-intensity earthquakes. Arab J Geosci 15(14):1262. https://doi.org/10.1007/s12517-022-10534-3

Ghani S, Kumari S, Choudhary AK (2023) Geocell mattress reinforcement for bottom ash: a comprehensive study of load-settlement

characteristics. Iran J Sci Technol Trans Civ Eng. https://doi.org/10.1007/s40996-023-01205-8

Huang M, Ma Y, Wan J, Chen X (2015) A sensor-software based on a genetic algorithm-based neural fuzzy system for modeling and simulating a wastewater treatment process. Appl Soft Comput 27:1–10. https://doi.org/10.1016/j.asoc.2014.10.034

Kassa SM, Wubineh BZ (2023) Use of machine learning to predict california bearing ratio of soils. Adv Civ Eng 2023:1–11. https://doi.org/10.1155/2023/8198648

Katte VY, Mfoyet SM, Manefouet B, Wouatong ASL, Bezeng LA (2019) Correlation of California bearing ratio (CBR) value with soil properties of road subgrade soil. Geotech Geol Eng 37(1):217–234. https://doi.org/10.1007/s10706-018-0604-x

Khatti J, Grover KS (2023a) CBR prediction of pavement materials in unsoaked condition using LSSVM, LSTM-RNN, and ANN approaches. Int J Pavement Res Technol. https://doi.org/10.1007/s42947-022-00268-6

Khatti J, Grover KS (2023b) Prediction of soaked CBR of fine-grained soils using soft computing techniques. Multiscale Multidiscip Model Exp Des 6(1):97–121. https://doi.org/10.1007/s41939-022-00131-y

Khatti J, Grover KS (2023c) Relationship between index properties and CBR of soil and prediction of CBR. Springer Nature, Singapore, pp 171–185. https://doi.org/10.1007/978-981-19-6774-0_16

Kim M, Ordu S, Arslan O, Ko J (2023) Prediction of California bearing ratio (CBR) for coarse- and fine-grained soils using the GMDH-model. Geomech Eng 33(2):183–194

Koti Marg K, Puram R (2012) IRC: 37-2012 iii guidelines for the design of flexible pavements Indian roads congress

Koti Marg K, Puram R (2019) Guidelines for the design of flexible pavements Indian roads congress

Kumar D, Gandhi BGR, Bhattacharjya RK (2020) Introduction to invasive weed optimization method. Springer, Cham, pp 203–214. https://doi.org/10.1007/978-3-030-26458-1_12

Kurnaz TF, Kaya Y (2020) The performance comparison of the soft computing methods on the prediction of soil compaction parameters. Arab J Geosci 13(4):159. https://doi.org/10.1007/s12517-020-5171-9

Nagaraju TV, Bahrami A, Prasad ChD, Mantena S, Biswal M, Islam MdR (2023) Predicting California bearing ratio of lateritic soils using hybrid machine learning technique. Buildings 13(1):255. https://doi.org/10.3390/buildings13010255

Prakash S, Kumar S, Rai B (2023) A new technique based on the gorilla troop optimization coupled with artificial neural network for predicting the compressive strength of ultrahigh performance concrete. Asian J Civ Eng. https://doi.org/10.1007/s42107-023-00822-y

Raj Kiran N, Ravi V (2008) Software reliability prediction by soft computing techniques. J Syst Softw 81(4):576–583. https://doi.org/10.1016/j.jss.2007.05.005

Soltanali H, Rohani A, Abbaspour-Fard MH, Farinha JT (2021) A comparative study of statistical and soft computing techniques for reliability prediction of automotive manufacturing. Appl Soft Comput 98:106738. https://doi.org/10.1016/j.asoc.2020.106738

Taha S, Gabr A, El-Badawy S (2019) Regression and neural network models for california bearing ratio prediction of typical granular materials in Egypt. Arab J Sci Eng 44(10):8691–8705. https://doi.org/10.1007/s13369-019-03803-z

Taskiran T (2010) Prediction of California bearing ratio (CBR) of fine grained soils by AI methods. Adv Eng Softw 41(6):886–892. https://doi.org/10.1016/j.advengsoft.2010.01.003

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 106(D7):7183–7192. https://doi.org/10.1029/2000JD900719

Tenpe AR, Patel A (2020) Utilization of support vector models and gene expression programming for soil strength modeling. Arab J

Sci Eng 45(5):4301–4319. https://doi.org/10.1007/s13369-020-04441-6

Vamsi Krishna SH, Sai Santosh B, Sai Prasanth BHS (2023) Prediction of UCS and CBR of a stabilized Black-cotton soil using artificial intelligence approach: ANN. Mater Today Proc. https://doi.org/10.1016/j.matpr.2023.05.097

Varghese VK, Babu SS, Bijukumar R, Cyrus S, Abraham BM (2013) Artificial neural networks: a solution to the ambiguity in prediction of engineering properties of fine-grained soils. Geotech Geol Eng 31(4):1187–1205. https://doi.org/10.1007/s10706-013-9643-5

Verma G, Kumar B, Kumar C, Ray A, Khandelwal M (2023) Application of KRR, K-NN and GPR algorithms for predicting the soaked CBR of fine-grained plastic soils. Arab J Sci Eng. https://doi.org/10.1007/s13369-023-07962-y

Yildirim B, Gunaydin O (2011) Estimation of California bearing ratio by using soft computing systems. Expert Syst Appl 38(5):6381–6391. https://doi.org/10.1016/j.eswa.2010.12.054

Zhou J, Zhu S, Qiu Y, Armaghani DJ, Zhou A, Yong W (2022) Predicting tunnel squeezing using support vector machine optimized by whale optimization algorithm. Acta Geotech 17(4):1343–1366. https://doi.org/10.1007/s11440-022-01450-7