A comparative analysis of hybrid computational models constructed with swarm intelligence algorithms for estimating soil compression index

Abidhan Bardhan[1], Navid Kardani[2], Abdel Kareem Alzo`ubi[3], Pijush Samui[1], Amir H Gandomi[4], Candan Gokceoglu[5, *]

[1] Department of Civil Engineering, National Institute of Technology (NIT) Patna, Patna, Bihar - 800005, India. E-mail: abidhan@nitp.ac.in; pijush@nitp.ac.in

[2] Discipline of Civil and Infrastructure Engineering, School of Engineering, Royal Melbourne Institute of Technology (RMIT), GPO Box 2476, Melbourne, VIC 3001, Australia, E-mail: navid.kardani@rmit.edu.au

[3] Department of Civil Engineering, Abu Dhabi University, Abu Dhabi 59911, UAE. E-mail: abdelkareem@gmail.com

[4] Faculty of Engineering and Information Technology, University of Technology Sydney, Australia, E-mail: a.h.gandomi@gmail.com

[5] Department of Geological Engineering, Hacettepe University, 06800, Beytepe, Ankara, Turkey. E-mail: cgokce@hacettepe.edu.tr; candan.gokceoglu@gmail.com

*** Corresponding author:** Prof. Candan Gokceoglu, Hacettepe University, Department of Geological Engineering, 06800, Beytepe, Ankara, Turkey. E-mail: cgokce@hacettepe.edu.tr

Abstract: The determination of the compression index (C_c) of clay through oedometer tests is time-consuming and expensive. To replace the practice of conducting laboratory oedometer tests, this study presents a comparative analysis of hybrid machine learning models for estimating the soil C_c based on actual laboratory test data. Ten swarm intelligence algorithms, namely particle swarm optimization, ant colony optimization, artificial bee colony, grey wolf optimizer, moth flame optimizer, whale optimization algorithm, salp swarm algorithm, Harris hawks optimization, slime mould algorithm, and marine predator algorithm, were used to optimize the learning parameters of an artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS). Subsequently, 20 hybrid ANN and ANFIS models were constructed to obtain the best prediction model. A sum of 700 oedometer test data was acquired from an Indian Railways project to construct and validate the hybrid models. Besides, 30 new oedometer experiments were performed for external validation of the developed hybrid models. Experimental outcomes show that the proposed ANFIS and PSO hybrid model (ANFIS-PSO) attained the most accurate prediction of soil C_c, which is much superior to the developed hybrid ANN and ANFIS models. Based on the experimental results, the proposed ANFIS-PSO model demonstrates high potential as a robust alternative to the actual oedometer test to assist geotechnical engineers in the introductory stage of civil engineering projects.

Keywords: *Soil consolidation parameter; Soil texture; Indian Railways; Meta-heuristic optimization; Hybrid machine learning; Swarm intelligence.*

Statements and Declarations:

Competing Interests: None declared.

Funding: No external funding has been received.

Acknowledgement: The authors are very thankful to the Dedicated Freight Corridor of India Limited (DFCCIL) New Delhi, India and officials of Larsen and Toubro Construction, DFCCIL CTP-3(R) Project Site, Ahmedabad, India to provide experimental data and their kind support during the course of this study.

1. Introduction

The accessibility and development of a nation's infrastructural facilities have a substantial impact on its growth and advancement. These facilities are important drivers of economic growth and aid in the reducing poverty in the country [1]. As a result of the rapid rise of industrialization and urbanization, there is a constant demand for engineering and construction research and innovation [2]. The development of infrastructure facilities, particularly in the field of transportation and related engineering, has seen a spectacular change in recent years. Generally, these facilities are built on the natural ground surface and require substantial investments. Also, it is pertinent to mention that the safety of an engineering structure cannot be ensured without considering the stability of the foundation, which is an element of a structure that transfers loads from the structure into the ground. Hence, a detailed analysis of the safety and serviceability of structures, such as safe bearing capacity, swelling potential, and settlement of soils, is required. For these reasons, estimating the settlement beneath any foundations is a critical aspect of the foundation design in geotechnical engineering discipline [3–5].

In geotechnical engineering, stresses on soils due to structural loads lead to an increment in stress in the soil layers, causing settlement. The major responsibility of geotechnical engineers is to compute the amount of expected settlement for the safety and reliability of the foundation [6]. Compressibility characteristics of soil deposits under the applied load play a key role in settlement analysis. In general, the compressibility of soil is defined as the volume loss under pressure caused by pore water drainage [7]. The compression index (C_c), coefficient of consolidation, and coefficient of compressibility are essential parameters of soil compressibility characteristics; among which, C_c is frequently utilized for direct settlement estimation beneath the foundation. The determination of compressibility parameters, such as C_c, is usually obtained from the consolidation test performed on an undisturbed soil sample collected from sites. Subsequently, the amount of settlement of soil/sub-soils is calculated based on the values of C_c [7].

An oedometer test is performed in the laboratory to determine the soil C_c, but requires a cumbersome and time-consuming operation [8,9]. Generally, because oedometer tests take at least 7 days to complete, the estimation of C_c would be possible only after a week. Besides, specific laboratory equipment, experienced engineers, and highly-skilled lab technicians are required to attain reliable results [7]. In addition, high-quality and undisturbed clay specimens are required for the oedometer tests. Therefore, to circumvent these difficulties, numerous attempts have been made to estimate C_c by utilizing basic physical properties of soils, such as liquid limit (LL), initial void ratio (e_0), void ratio (e), natural moisture/water content (NMC/NWC), specific gravity (Gs), plasticity index (I_P), etc. [10–14]. Numerous empirical models have been developed based on conventional statistical analysis (i.e., simple and multiple linear regressions); however, a close examination of the existing models reveals that their applicability is limited, highlighting significant modelling drawbacks [7]. On the other hand, soils are vastly heterogenic in nature and possess nonlinear stress-strain relationships under different loading conditions [9]. Traditional empirical formulations fall short of describing the predictive relationship between C_c and physical properties of soil. Hence, smart and intelligent approaches are required to estimate the values of soil C_c based on soil parameters that are normally determined when the samples are brought into the laboratory.

Recent studies have resorted to advanced soft computing techniques as potential alternatives to predict C_c of soils. Soft computing techniques, offering competence in nonlinear modelling, capture the structure of a model by learning from existing data and simulating complicated processes [9,15–19]. In geotechnical engineering, the use of various soft computing techniques is increasing, such as adaptive neuro-fuzzy inference system (ANFIS),

artificial neural network (ANN), backpropagation multi-layer perceptron neural network (BP-MLP), radial basis function neural network (RBF-NN), Bayesian regularization neural network (BRNN), Levenberg-Marquardt ANN (LM-ANN), gene expression programming (GEP), multi-expression programming (MEP), multi-variate adaptive regression splines (MARS), genetic programming (GP), multi-gene genetic programming (MGGP), extreme learning machine (ELM), Gaussian process regression (GPR), random forest (RF), regression tree (RT), support vector machine/regression (SVM/SVR), and so on [1,7,9,20–35][36].

In geotechnical engineering practice, determining the C_c of soils is an essential criterion for analysing the settlement of soil and sub-soil layers. Over the last decade, multiple machine learning (ML) models, including ANFIS [35], ANN/BP-MLP/RBF-NN/BRNN/LM-ANN [1,7,9,21,29,31,34,35,37–39], GP/GEP/MEP/MGGP [7,20,33,39–41], GPR [9], RF [9,41], and SVM/SVR [9,34,35], have been employed to build prediction models for soil C_c based on existing experimental datasets. Several researchers [1,7,9,21,29,31,34,35,37–39,42] have attained varying levels of accuracy using ANN-based models, ranging from 0.6464 to 0.9610 in the training phase and 0.5625 to 0.9580 in the testing phase, based on R^2 (determination coefficient) value. Bardhan et al. [41], Benbouras et al. [39], Mohammadzadeh et al. [7,33,40], and Bourouis et al. [20] used GP-based models (i.e., GP/MEP/MGGP/GEP) to predict soil C_c and obtained varied accuracy levels in the range of 0.0574 to 0.9966 (based on R^2 value). In addition, regression-based ML models (such as GPR, SVM/SVR, etc.) and tree-based models (e.g. RF and RT) have also been utilized for predicting soil C_c [34,35,42,43]. The details of earlier studies on soil C_c prediction are given in Table 1, which includes the types of employed models, number of the dataset used, and accuracies attained in training and testing phases.

It is worth noting that the majority of these predictive models were developed using conventional ML (CML) models, including ANFIS, ANN, and regression-based ML models, as shown in Table 1. In many cases, higher predicted accuracies between 0.6780 and 0.9890 based on R^2 value were attained [7,9,21,31,32,35,37]. However, because the models were constructed and corroborated with small datasets (the cumulative number of samples was less than 250), these studies cannot be considered particularly credible (see Table 1 for related studies). Also, the predicted accuracy of CML models decreased with larger datasets (the number of datasets in the range of 700 to 947), and prediction performance ranged between 0.7600 and 0.8850 based on R^2 value. Herein, the model's performance obtained in the testing phase is examined and reported. The lack of a diverse set of input parameters, complicated data types, and inappropriate training of ML algorithms could all contribute to the decrease in accuracy. Overfitting is also a major concern for the success rate of CML methods, which could be one reason for reduced accuracy. Due to their inability to reach the exact global optimum, certain CML models, such as ANN, yield poor outcomes and are more likely to get stuck in local minima, resulting in erroneous results [41,44,45].

To address these challenges, researchers have employed many hybrid models that combine optimization algorithms (OAs) and CML techniques to identify the actual global optimum rather than local minima [44,45]. Integration of OAs and CML techniques balances the exploration and exploitation (E&E) processes and generates optimum learning parameters, which are then used to improve the performance of CML models. Multiple OAs, namely particle swarm optimization (PSO) [41,45], biogeography-based optimization (BBO) [46], grey wolf optimizer (GWO) [47], artificial bee colony (ABC) [48], imperialist competitive algorithm (ICA) [45], genetic algorithm (GA) [48], and so on [26], have been successfully employed to optimize the learning parameters of CML algorithms. A study performed by Ojha et al. [49] showed an increasing trend of integrating feed-forward neural networks and meta-heuristic optimizer in complex process modelling. Moreover, recent literature exposes

the that hybrid ANN and ANFIS models are highly suitable for predicting the desired output. Successful application of several ANN and ANFIS-based hybrid models, namely ANN-PSO, ANFIS-PSO, ANN-BBO, ANFIS-BBO, ANN-GA, ANFIS-GA, ANN-ICA, and ANN-GWO, has been demonstrated in predicting fly-rock distance resulting from blasting [45], axial load-carrying capacity of concrete-filled steel tubes [46], compressive strength of normal and high-performance concretes [47], heating and cooling loads on buildings [48], mapping the spatial distribution of soil texture fractions [50], spatial modelling of soil electrical conductivity [51]. soil electrical conductivity [52], soil moisture [53], and so on.

Recently, Bui et al. [9] proposed a new approach by integrating the MLP neural nets and PSO (PSO-MLP neural nets) to estimate C_c of soils, utilizing 154 tests data collected from a high-rise building project in Vietnam. The authors obtained an accuracy of 95.70% (R^2 = 0.9570) in the training phase, which dropped to 88.40% (R^2 = 0.8840) in the testing phase with the same input parameters. Samui et al. [42] integrated artificial bee colony (ABC) and the Levenberg-Marquardt (LM) algorithm of ANN and developed two hybrid models, namely ABC-LM-ANN and ABC-ANN, to estimate the C_c of soils for a housing construction project. Bourouis et al. [20] applied NN-PSO for predicting the secondary C_c of fine-grained soils. Based on the results, the authors concluded that the employed NN-PSO could attain a good agreement between the observed and predicted values. Recently, Bardhan et al. [41] used ANN and ELM-based hybrid models, namely hybrid models of ANN and PSO (ANN-PSO), ANN and equilibrium optimizer (ANN-EO), ANN and Harris hawks optimization (ANN-HHO), ANN and slime mould algorithm (ANN-SMA), ANN and marine predators algorithm (ANN-MPA), ELM and PSO (ELM-PSO), ELM and EO (ELM-EO), ELM and HHO (ELM-HHO), ELM and SMA (ELM-SMA), and ELM and MPA (ELM-MPA), to predict the C_c of soils. The details of the above-mentioned studies on soil C_c prediction are also presented in Table 1.

A detailed review of the literature reveals that none of the previous studies employed hybrid models to predict C_c of soils, except Bui et al. [9], Samui et al. [42], Bourouis et al. [20], and Bardhan et al. [41]. In addition, in-depth assessment of hybrid ANN and ANFIS models constructed with a specific group of OAs has not been investigated for predicting the C_c of soils. Also, no comprehensive evaluation of hybrid models has been performed to address the stochastic nature of OAs in hybrid modelling. Thus, the current study is an attempt to fill this gap in the literature for predicting C_c of soils utilizing hybrid ANN and ANFIS models optimized with multiple swarm intelligence (SI) algorithms. Among the developed SI algorithms, ten OAs, namely PSO, ant colony optimization (ACO), ABC, GWO, moth-flame optimization (MFO), ant-lion optimizer (ALO), whale optimisation algorithm (WOA), salp swarm algorithm (SSA), SMA, and MPA, were selected and employed to optimize the learning parameters of ANN and ANFIS. Accordingly, 20 hybrid models, including ANN-PSO, ANN-ACO, ANN-ABC, ANN-GWO, ANN-MFO, ANN-ALO, ANN-WOA, ANN-SSA, ANN-SMA, ANN-MPA, ANFIS-PSO, ANFIS-ACO, ANFIS-ABC, ANFIS-GWO, ANFIS-MFO, ANFIS-ALO, ANFIS-WOA, ANFIS-SSA, ANFIS-SMA, and ANFIS-MPA, were constructed to perform a comparative assessment of hybrid ANN and ANFIS models for predicting the C_c of soils.

To develop and validate the proposed hybrid models, a sum of 700 oedometer test data, including 12 influencing factors, was acquired from an Indian Railways (IR) dedicated freight corridor (DFC) project. In addition, 30 new experiments were executed at the geotechnical engineering laboratory of the National Institute of Technology Patna (NIT Patna) for the external validation of the developed models in predicting soil C_c. Per the

authors' knowledge, this study is the pioneer work to use both hybrid ANN and ANFIS models to estimate soil C_c, which is one of the most critical real-time challenges in the field of geotechnical engineering.

Table 1. A summary of earlier studies.

Earlier studies reference	Dataset used	Model employed	R^2 (TR/TS)	Soil parameters used as input variables (full from given in the table footnote)
Conventional models				
Alam et al. [38]	391	ANN	0.7600/0.7200	In-11, In-13, In-19, In-20
Bardhan et al. [41]	695	GP	0.9388/0.9305	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
	695	RF	0.9152/0.9124	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
	695	GBM	0.8805/0.8679	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
Benbouras et al. [39]	373	ANN	0.6464/0.5625	In-7, In-11, In-13, In-14 – In-16
	373	GP	0.6058/0.0574	In-11, In-13, In-14 – In-16
Bourouis et al. [20]	203	MGGP	0.9966**	In-19, In-20, In-25
Bui et al. [9]	154	BP-MLP	0.9350/0.8620	In-1 – In-3, In-5, In-7, In-9, In-11 – In-14, In-23, In-24
	154	RBF-NN	0.8420/0.6780	In-1 – In-3, In-5, In-7, In-9, In-11 – In-14, In-23, In-24
	154	RF	0.9900/0.8040	In-1 – In-3, In-5, In-7, In-9, In-11 – In-14, In-23, In-24
	154	GPR	0.9040/0.7970	In-1 – In-3, In-5, In-7, In-9, In-11 – In-14, In-23, In-24
	154	SVR	0.8810/0.7770	In-1 – In-3, In-5, In-7, In-9, In-11 – In-14, In-23, In-24
Pham et al. [35]	189	ANFIS	0.9980/0.9570	In-1 – In-13
	189	ANN	0.9080/0.9390	In-1 – In-13
	189	SVM	0.9800/0.9890	In-1 – In-13
Pham et al. [21]	189	ANN	0.9010*	In-1 – In-9
Kashefipour and Daryaee [31]	137	ANN	0.6700/0.7000	In-7, In-11, In-13, In-17, In-18
Kolay et al. [1]	700	ANN	-/0.7600	In-2, In-4 – In-7, In-11, In-13, In-19, In-26 – In-28, In-33
Kurnaz et al. [32]	246	ANN	0.8926/0.8973	In-11, In-13, In-19, In-20
Kurnaz and Kaya [34]	351	BRNN	0.9153/0.8887	In-11, In-13, In-19, In-20
	351	ELM	0.8642/0.8898	In-11, In-13, In-19, In-20
	351	SVM	0.9147/0.8761	In-11, In-13, In-19, In-20
Majdi et al. [37]	150	ANN	0.9610/0.9580	In-2, In-6, In-11, In-13, In-19 – In-22
Mohammadzadeh et al. [7]	108	ANN	0.8705/0.8593	In-11, In-12, In-20
	108	MEP	0.8742/0.8118	In-11, In-12, In-20
Mohammadzadeh et al. [33]	108	MGGP	0.8560/0.8400	In-11, In-12, In-20
Mohammadzadeh et al. [40]	108	GEP	0.8231/0.8603	In-11, In-12, In-20
Park and Lee [29]	947	ANN	0.8960/0.8850	In-6, In-7, In-11, In-13, In-19, In-30 – In-32
Samui et al. [42]	441	LM-ANN	0.8730/0.8110	In-1 – In-3, In-5, In-7, In-9, In-11 – In-14, In-23, In-24
	441	RT	0.9550/0.8100	In-1 – In-3, In-5, In-7, In-9, In-11 – In-14, In-23, In-24
Hybrid models				
Bardhan et al. [41]	695	ANN-PSO	0.7454/0.6907	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
	695	ANN-EO	0.9504/0.9426	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
	695	ANN-HHO	0.6651/0.6361	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
	695	ANN-SMA	0.8995/0.9010	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
	695	ANN-MPA	0.9530/0.9352	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
	695	ELM-PSO	0.9543/0.9412	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
	695	ELM-EO	0.9720/0.9468	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
	695	ELM-HHO	0.9736/0.8749	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
	695	ELM-SMA	0.8232/0.9461	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
	695	ELM-MPA	0.9702/0.9475	In-1, In-2, In-4 – In-6, In-11 – In-13, In-26 – In-29
Bourouis et al. [20]	203	NN-PSO	0.9980**	In-19, In-20, In-25
Bui et al. [9]	154	PSO-MLP	0.9570/0.8840	In-1 – In-3, In-5, In-7, In-9, In-11 – In-14, In-23, In-24
Samui et al. [42]	441	ABC-LM-ANN	0.8600/0.8400	In-1 – In-3, In-5, In-7, In-9, In-11 – In-14, In-23, In-24
	441	ABC-ANN	0.8210/0.7920	In-1 – In-3, In-5, In-7, In-9, In-11 – In-14, In-23, In-24

Note: In-1 = Depth of sample; In-2 = clay content; In-3 = moisture content; In-4 = bulk density; In-5 = dry density; In-6 = specific gravity; In-7 = void ratio; In-8 = porosity; In-9 = liquidity index; In-10 = degree of saturation; In-11 = liquid limit; In-12 = plastic limit; In-13 = plasticity index; In-14 = wet density; In-15 = water content; In-16 = fine content; In-17 = primary soil water content; In-18 = relative density; In-19 = natural water/moisture content; In-20 = initial void ratio; In-21 = natural unit weight; In-22 = shear wave velocity; In-23 = sand percentage; In-24 = loam percentage; In-25 = vertical stress; In-26 = gravel content; In-27 = sand content; In-28 = silt content; In-29 = free swell index; In-30 = weight percentage of grain size larger than 0.075 mm; In-31 = weight percentage of grain size between sieve 0.075 mm and 0.005 mm; In-32 = weight percentage of grain size less than 0.005 mm; and In-33 = Pre-consolidation pressure.
* Reported average value; ** overall/partly/performance reported; TR = training phase; TS = testing phase

The rest of this paper is structured as follows. Section 2 describes the literature review and the research significance. The study area is described in section 3, which is followed by a discussion on the background of meta-heuristic OAs and brief overview of the employed OAs in section 4. The theoretical background of ANN, ANFIS, and optimization procedure are presented in section 5. Section 6 presents the descriptive statistics of the collected dataset, and section 7 discusses the data processing, analysis, and performance parameters. Section 8 reports the experimental results, comparative analysis, external validation, and monotonicity analysis, which is followed by a summary and conclusion in the final section.

2. Literature review and research significance

Over the decades, various attempts have been undertaken to establish empirical relationships between C_c and various soil parameters [10–14]. These empirical equations are widely used by practitioners to save the time of actual oedometer tests. Nonetheless, as stated above, these empirical models possess notable modelling drawbacks since soils exhibit extremely complicated behaviour due to their non-linear stress-strain relationships and elastoplastic performance under diverse loading conditions. Thus, such empirical equations cannot be considered very suitable for determining soil C_c utilizing index properties of soils [7,9].

Several conventional/standalone ML techniques for estimating soil C_c have been developed in the last decade. For instance, Alam et al. [38] used ANN to model the C_c soil using 391 datasets with LL, I_P, NWC, and e_0. Bardhan et al. [41] used GP, RF, and GBM to relate C_c of fine-grained soils utilizing 695 experimental results and attained accuracy in the range of 86.79% – 93.88%, based on R^2 value. The authors [41] used depth of soil sample, particle sizes, bulk and dry density, Gs. LL, PL, I_P, and free swell index as the input parameters. Benbouras et al. [39] applied ANN and GP for 373 datasets and Bourouis et al. [20] used MGGP for 203 datasets. Bui et al. [9] used a total of 154 datasets featuring 12 influencing factors and developed BP-MLP, RBF-NN, RF, GPR, and SVR models for soil C_c. The authors achieved accuracy levels ranging from 67.80% – 99%. Using 189 consolidated test data, Pham et al. [21,35] employed ANN, ANFIS, and SVM to estimate soil C_c and reported greater accuracy levels in both the training and testing phases. Kashefipour and Daryaee [31] investigated fine-grained soils and found that the ANN predicted C_c better than the Rendon-Herrero formula. Kolay et al. [1], Kurnaz et al. [32], Majdi et al. [37], and Park and Lee [29] employed ANNs to relate C_c of various soils depending on their physical properties. Kurnaz and Kaya [34] compared the ELM, SVM, and BRNN approaches and found that the BRNN method predicts C_c better than the ELM and SVM methods. Mohammadzadeh et al. [7,33,40] used ANN and GP-based approaches (such as MEP, MGGP, and GEP) for relating soil C_c based on LL, PL, and e_0. Samui et al. [42] used a total of 441 datasets and developed LM-ANN and RF models. The authors achieved 87.30% – 95.50% and 81.1% – 81% accuracy in the training and testing phases, respectively, based on the findings.

On the other hand, Bardhan et al. [41] used a combination of Principal component analysis (PCA)-based hybrid ANN and ELM models to develop a high-performance soil C_c prediction model. To train and validate the hybrid ANNs and ELMs, Bardhan et al. [41] used a total of 695 consolidation test results featuring different soil properties. The authors concluded that the ELM-MEO is a promising alternative for predicting soil C_c based on the findings. Bourouis et al. [20] and Bui et al. [9] used PSO to develop hybrid NNs. Samui et al. [42] selected a large-scale real-life urban project (Hai Phong city, Vietnam) as a case study and produced two hybrid models, namely ABC-LM-ANN and ABC-ANN. Based on the findings, the authors [42] showed that the developed ABC-LM-ANN attained 86% and 84% (R^2 = 0.86 and 0.84) accuracy in the training and testing phases, respectively. The aforementioned discussion has been summarized in Table 1.

In recent years, hybrid computational modelling has exploded in popularity among researchers from various engineering fields. However, the use of hybrid models constructed with a specific group of OAs has not been examined for estimating the C_c of soils. In addition, the accuracy of hybrid models has not been thoroughly assessed to estimate the desired outcomes, including soil C_c. Taking this into consideration, this work was driven by a desire to fill gaps in the literature. A comparative analysis of hybrid ANN and ANFIS models constructed with swarm intelligence algorithms was performed followed by experimental validation. Moreover, two distinct combinations of soil parameters (based on grain size distribution analysis) were investigated to ensure the

effectiveness of silt and clay particles on soil C_c. At the end, the most efficient hybrid model was identified for predicting the C_c of soils.

3. Study area

To meet the volume of freight demand in India, the Ministry of IR has planned to build over a 10,122-km DFC route under the Golden Quadrilateral Corridor (GQC) [54]. There are currently 6 DFCs under the GQC, and the proposed rail networks will join the four largest cities of Chennai, Delhi, Kolkata, and Mumbai, along with two diagonals North-South DFC (Delhi-Chennai) and East-West DFC (Kolkata-Mumbai). Presently, two DFCs, the western DFC (WDFC) and eastern DFC (EDFC), are under construction. In this work, the Iqbalgarh-Vadodara section, a 340-km long (final length may be changed at the time of completing the project) portion of WDFC, was selected as the study area. Fig. 1 shows the geographical location of the study area along with the route layout of the Iqbalgarh-Vadodara section. The entire Iqbalgarh-Vadodara section will pass through the state of Gujarat, India (see Fig. 1).

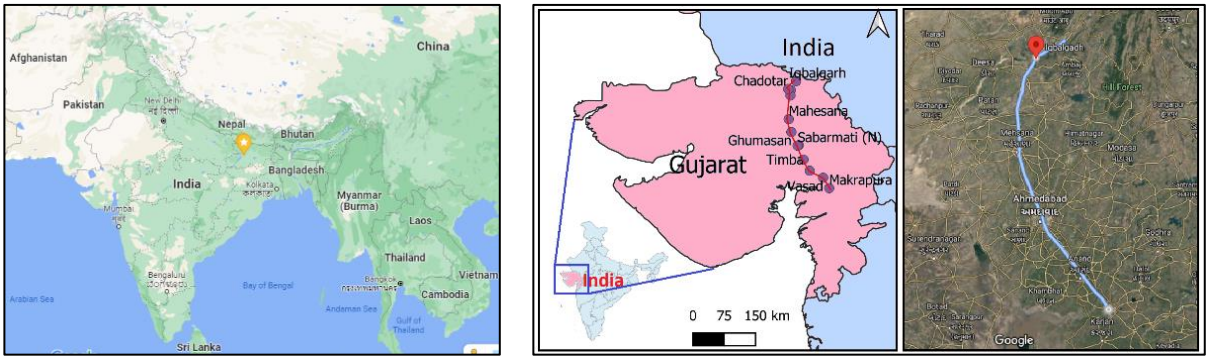

Fig. 1. Geographical layout of the study area.

WDFC is a 1506-km long (final length may be changed at the time of completing the project) railway freight corridor that will begin in Dadri in the state of Uttar Pradesh and will terminate in Jawaharlal Nehru Port near Navi Mumbai, India. As mentioned, the Iqbalgarh-Vadodara section, which is currently under construction, was selected for this study. A sum of 700 oedometer test results was acquired from the mentioned section for predicting soil C_c. Besides consolidation test data, sub-soil data, grain size analysis results, plasticity characteristics, and other basic soil parameters were also collected.

4. Overview of optimization algorithms

The rising complexity and difficulty of real-world issues has necessitated the use of optimization techniques, particularly meta-heuristic OAs, in recent decades. These methods are primarily stochastic and are utilized to estimate optimal solutions to a variety of linear and non-linear optimization problems. By reducing or maximizing the objective function of a problem, the optimization method identifies the optimal decision variables. Because of their simplicity and ease of implementation, meta-heuristic OAs have been devised and used as competing alternative solvers for the problem at-hand [55,56].

In general, there are two sorts of meta-heuristic OAs: population-based (p-based) and single solution-based. As the name implies, the p-based OA evolves a set of solutions in each iteration of the optimization process, whereas the single solution-based OA processes only one solution during the course of optimization. The majority of p-based meta-heuristic algorithms are centred on natural occurrences. These algorithms begin the optimization

process by creating a population of individuals, each of which represents a potential solution to the optimization problem. The population will iteratively evolve by replacing the present population with a freshly created population using stochastic operators. The optimization process is carried out until the stopping criteria are met.

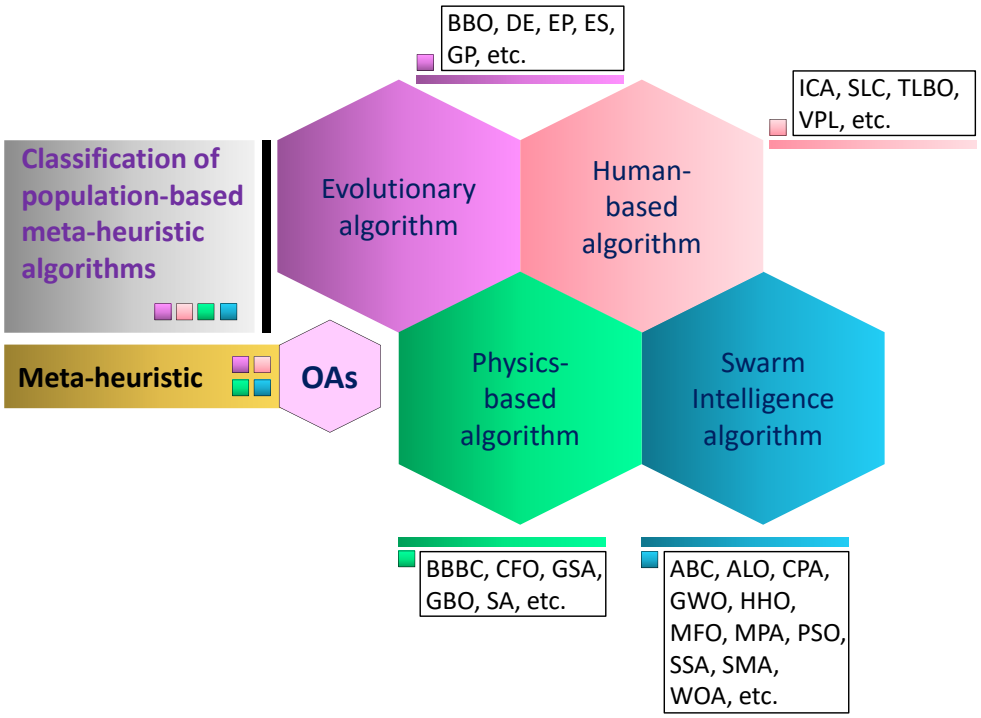

Fig. 2. Classification of meta-heuristic OAs.

The p-based meta-heuristic OAs can be divided into four categories based on their inspiration: (a) evolutionary algorithms (EAs), (b) human-based algorithms, (c) physics-based algorithms, and (d) swarm intelligence (SI) algorithms. Among these groups, the EAs mimic biological evolutionary behaviours, such as recombination, mutation, and selection. The GA [57] is the most common EA, as it models Darwin's theory of evolution. The BBO [58], differential evolution (DE) [59], evolutionary programming (EP) [60], evolution strategy (ES) [61], and GP [62] are examples of EAs. The second category of OA is human-based approaches, which are inspired by human collaboration and community behaviour. The ICA [63], which is inspired by human socio-political growth, is one of the most widely utilized algorithms in this group and includes soccer league competition (SLC) [64], teaching-learning-based optimization (TLBO) [65], and volleyball premier league (VPL) [66]. The physical laws motivate physics-based algorithms, such as Big-Bang Big-Crunch (BBBC) [67], central force optimization (CFO) [68], gravitational search algorithm (GSA) [69], gradient based optimizer (GBO) [70], and simulated annealing (SA) [71]. The SI algorithms are the final category of p-based meta-heuristic algorithms that mimic the social behaviours of organisms living in swarms, flocks, and herds. Examples of SI algorithms include ABC [72], ALO [73], GWO [74], HHO [75], MFO [76], MPA [77], PSO [78], SSA [79], SMA [80], WOA [81], and so on . Fig. 2 displays the different groups of meta-heuristic OAs as discussed above.

In the present study, ten different SI algorithms, namely PSO, ACO, ABC, GWO, MFO, WOA, SSA, HHO, SMA, and MPA, have been used to construct hybrid ANN and ANFIS models. In the following sub-sections, a short discussion on these OAs is presented.

4.1. Particle swarm optimization (PSO)

Kennedy and Eberhart [78] presented PSO as a member of the swarm-based community in 1995, which was motivated by the flocking and schooling habits of fish and birds. The main objective of PSO is to discover global-optimal solutions in a multidimensional environment. PSO starts by implementing the random speeds and locations of objects, which each adjusts its position to pick the appropriate status in a multidimensional environment depending on its speed, personal best position, and global best position. The best position obtained by individual particles is proven to be the ideal global status, while the preferred option attained by the particle is found to be the ideal personal position. The position of the particle becomes modified based on its best personal position and the orientation of the best global location. Meanwhile, the speeds of the objects are adjusted depending on the discrepancy between their best personal position and the best global place. Through a combination of E&E, the particles converge together around the optima. The acceleration coefficients, c_1 (cognitive coefficient) and c_2 (social coefficient) with fixed values of 1 and 2, respectively, are dependent on the issue and demonstrate the confidence level of an element compared to its personal and global status. The detailed working principle of PSO can be obtained from previous studies [9,48].

4.2. Ant colony optimization (ACO)

ACO is a meta-heuristic method for solving NP-hard problems that, as the name implies, mimics the genuine and natural behaviour of ants in nature [82]. When ants discover food, they leave chemical trails called pheromones on the path back to their colony, which aids other inhabitants of the colony to locate food resources. Other members are more inclined to pursue the trail rather than searching for food at random, then also spread pheromones upon locating the food supply. The pheromones dissipate over time, and the trail vanishes, whereby the longer the route, the more difficult it is to maintain the route to the food. As a result, the strength of the pheromone deposits on the shortest route is progressively increased to the degree that is regulated with evaporation. This rise and reduction in pheromone quantity on a certain route entice the ants to follow that path. This natural behaviour is exploited in the ACO algorithm by taking into account a set of parameters that represent the continually adjusting pheromone values dependent on the merit of the obtained solutions. ACO's basic approach is to discover a method that proceeds by building an incremental solution according to the algorithm's specified route, whereby the pheromone's probabilistic solution choice determines the route.

4.3. Artificial bee colony (ABC)

Karaboga and Basturk [72] created the ABC optimization algorithm with the social structure of bees in mind. In ABC, every bee is seen as a basic component that are combined to create a bee colony. The resulting colony displays a coherent and sophisticated behaviour and provides an interconnected system to find and investigate flowers' nectar. A colony is comprised of three types of bees: scouts, employed workers, and observer bees, each with a specific function. Scout bees are in charge of discovering new sources and will search the surrounding area at random. When they find to a new food source, they remember it. When scout bees return to the hive, they perform a waggle dance with the other bees to communicate the food supply information, then some scout bees are recruited to seek the source. Employed bees are responsible for reusing old sources of food. Onlooker bees wait for other bees to communicate information with them via the waggle dance and then choose a source based on the bees' exploratory response efficiency.

4.4. Grey wolf optimizer (GWO)

GWO [74] is a nature-inspired optimization algorithm that mimics the strict plural hierarchy of grey wolves whose main behaviour is hunting. GWO includes a couple of males and females, whereby the best solution is recognised as the α (alpha) group that makes main decisions like hunting. The β (beta) wolves are the second level, which participate in making a decision and following the alpha wolves. The beta can be females, help in flock adjustment, and are the best candidates for substitution of a dead or old alpha wolf. The third level of wolves is δ (delta) wolves, which are responsible for sentinels and scouts and employ in the hunt. The final group of individuals, ω (omega), is recognized as the weakest level and is responsible for watching younger wolves. Muro et al. [83] presented grey wolf hunting method of three levels: (a) identifying, pursuing, and closing in on the target; (b) surrounding the target; and (c) rushing the target. These two various social behaviours are regarded in the GWO algorithm. In the modelling step of the mentioned algorithm, α (alpha) is the fittest solution, then β (beta), δ (delta), and ω (omega) are suitable solutions in the next steps. The original study of Mirjalili et al. [74] can be referred for detailed descriptions of GWO.

4.5. Moth flame optimization (MFO)

MFO is a nature-inspired meta-heuristic algorithm used to solve optimization problems [76], which is inspired by the travelling behaviour of moths at night with their backs to the moon. Throughout their flight, moths are frequently caught by artificial lights like a flame, lamp, or bulb. To solve this challenge, moths use a spiral approach to approach the flame, known as the optimization technique of moths. They can traverse in 1D, 3D, and hyper-plane dimensional space by varying locations. As a population-based approach, MFO uses 3-tuple to estimate the global optimum of the issue at-hand. The first step is to initialize the population, followed by the primary function, which transfers the moths across the search area, and finally the halting the function yields true if the ending condition is fulfilled and yields false otherwise. Because the moth must find fame in the search area, the moth's location concerning fame is critical.

4.6. Whale optimization algorithm (WOA)

WOA is a newly developed optimization algorithm that mimics the natural hunting behaviour of humpback whales [81]. Based on Watkins and Schevill [84], the most remarkable aspect of humpback whales is their unique hunting strategy, known as the bubble-net feeding method. The little fish near the surface are known as humpback whales. As shown in Figure 4, the hunting process is finished when several unique bubbles are formed along a nine-shaped or a circular route. The behaviour mentioned above was investigated in 2011 and prior to that, based on surface observations. Nonetheless, Goldbogen et al. [85] performed a distinct study using tag sensors. In the novel inquiry approach, they gathered 300 tag-derived bubble-net feeding episodes of nine distinct humpback whales and developed two new bubble motion plans, namely upward spirals and double loops. Specifically, humpback whales generate bubbles in a spiral form all around prey then swim toward the top in the preceding movement plan. The novel movement plan consists of three distinct steps: coral loop, lobtail, and capture loop, about which further information and behaviour descriptions can be found in the literature [81,85]. It is vital to emphasise that bubble-net feeding is a distinct activity seen exclusively in humpback whales.

4.7. Salp swarm algorithm (SSA)

Mirjalili et al. [79] proposed SSA by mimicking translucent water invertebrates called salps, or salp chain, that forage in a swarming fashion. The algorithm is built on a leader-follower connection, where the leader adjusts its position depending on the position of the best food. The mathematical model of SSA is divided into three sections: salps are allocated in random places in the first phase; the salp with the lowest fitness value and closest to the food supply is designated as the leader in the next step, while the others are designated as followers; then the position is updated in the following step. The leader shifts its position in relation to the best global solution and explores better options. The follower salps update their location in rank towards the leader, and the iterations continue until the termination condition, or the maximum number of iterations is met.

4.8. Harris hawks optimization (HHO)

HHO is a unique SI-based optimization approach presented by Heidari et al. [75] that relates the hunting behaviour of Harris hawks to mathematical computer systems. A community of Harris hawks assaults the prey (typically rabbits) in numerous directions and adopts various dynamic and sophisticated strategies adapting to the prey's fleeing pattern, resulting in bewildered and tired prey. The algorithm is divided into three steps. The exploring phase is the initial step of waiting, searching, and discovering, in which the birds represent potential solutions to the chosen challenge. The second step is the change from exploration to exploitation, which is determined by the type and energy of the prey. During the exploitation process, the third step, the identified prey is assaulted by surrounding and laid besieged from numerous directions. Depending on the prey's energy determined in the second stage, the besiege could be mild or severe.

4.9. Slime mould algorithm (SMA)

SMA is a recently-developed nature-inspired meta-heuristic OA [80] that considers the mathematical simulating modelling of slime mould propagation waves while making the best path for connecting foods. Owing to their unique characteristic and pattern, slime mould, a eukaryotic organism found in nature, simultaneously uses several food sources in order to create a venous network for their connection. Slime mould can grow up to lengths greater than 900 cm^2 if sufficient food exists in the environment. When a vein obtains a source of food, the bio-oscillator creates a spreading wave, which enhances the cytoplasmic flow into the vein, causing the vein to become thicker by increasing the speed of cytoplasm flow. Regarding these positive and negative responses, the slime could form the optimal path for food connection in a relatively greater way. Thus, slime mould has also been modelled mathematically and used in path networks and graph theory, in which the procedure of creating positive and negative responses through the wave propagation is simulated. Slime mould also could adjust their patterns of dynamic search based on the provenience of foodstuff quality. This procedure is implemented in various engineering optimization problems. The two major levels in the slime mould algorithm are: (a) obtaining food following that the behaviour of slime to acquire food according to its odour in the air; and (b) warp foodstuff in which the slime behaviour in performing contraction of its venous configuration. Detailed information and the working principle of SMA can be found in the original work of Li et al. [80].

4.10. Marine predators algorithm (MPA)

In MPA, the optimization process can be divided into three major stages based on the velocity ratio as well as imitating patterns of prey and predator [77]. These stages depend on the velocity of movements of prey to flee from predators: high-velocity ratio, unit velocity ratio, as well as a low-velocity ratio. Every phase in MPA is outlined and allocated a certain iteration interval and are characterized by the laws that govern predator and prey movement. Throughout phase 1, the prey moves quicker than the predator, resulting in a greater velocity ratio, which generally happens during the first iteration when experimentation is more crucial. Even though the velocity ratio is more than 10, the optimum option for the predator in this scenario is to remain motionless. In phase 2, both the predator and prey travel simultaneously to find their food, thus this phase is often referred to as the unit velocity ratio. The move from E&E happens during this stage, which is considered the intermediate phase of optimization. Therefore, E&E occur during this period, whereby half of the population is assigned for exploration and the remaining population is designated for exploitation. Significantly, the prey is in charge of exploitation, while the predator is in charge of exploration. Per the unit velocity ratio ($v \approx 1$), if the prey exhibits Lévy motion, the Brownian movement would be optimal for the predator to attack the prey. The low-velocity ratio is visible in phase 3 because the predator travels faster than the prey to attack it. This low-velocity ratio ($v = 0.1$) demonstrates the predator's excellent exploitation capabilities, wherein Lévy is the optimal tactic. The prey moves in Brownian motion during the first stage of the movement, then in Lévy motion during phase two. Each phase receives one-third of the iterations, resulting in better-optimized outcomes than switching or repeating the approach.

5. Theoretical background of ANN and ANFIS

This section describes the theoretical background of ANN and ANFIS used in this research work. This is followed by the methodological development of the proposed hybrid ANN and ANFIS models for predicting C_c of soils.

5.1. Artificial neural network (ANN)

As a computational approach that is inspired by the structure of the human brain, ANN is comprised of small and simple processing units known as artificial neurons or nodes. By harnessing this structure, ANN has become an effective mathematical tool for different purposes, such as pattern recognition and function approximation. ANN is made up of three layers in which neurons are located (see Fig. 3). The first layer and last layer of ANNs are called the input and output layers in which the number of neurons is equal to the number of the input and output variables of the problem, respectively. Between these two layers, there are one or more layers known as hidden layers, which are the computational engine of the network. Two essential parameters in the ANN are weights and biases. Interconnected relationships between neurons of a layer are indicated by weights, and the degree of freedom for the networks is determined by biases. With the exception of the input nodes, each node uses a nonlinear activation function term as a transfer function in order to determine the output of that node, providing a set of inputs.

Then, these outputs are employed as input for the next node, and so on, until a suitable solution is found to the original problem. The most predominant types of activation functions are the sigmoid, linear, and hyperbolic tangent functions. In order to compute the error of comparing the actual outcome, i.e., the target of the problem, and the predicted outcome, i.e., the network's outcome, a backpropagation algorithm is employed. This error is

then propagated back one layer at a time across the ANN structure, and the weights are changed based on their contribution to the error. Successful applications of ANNs can be found in the literature [19,48,86–92].

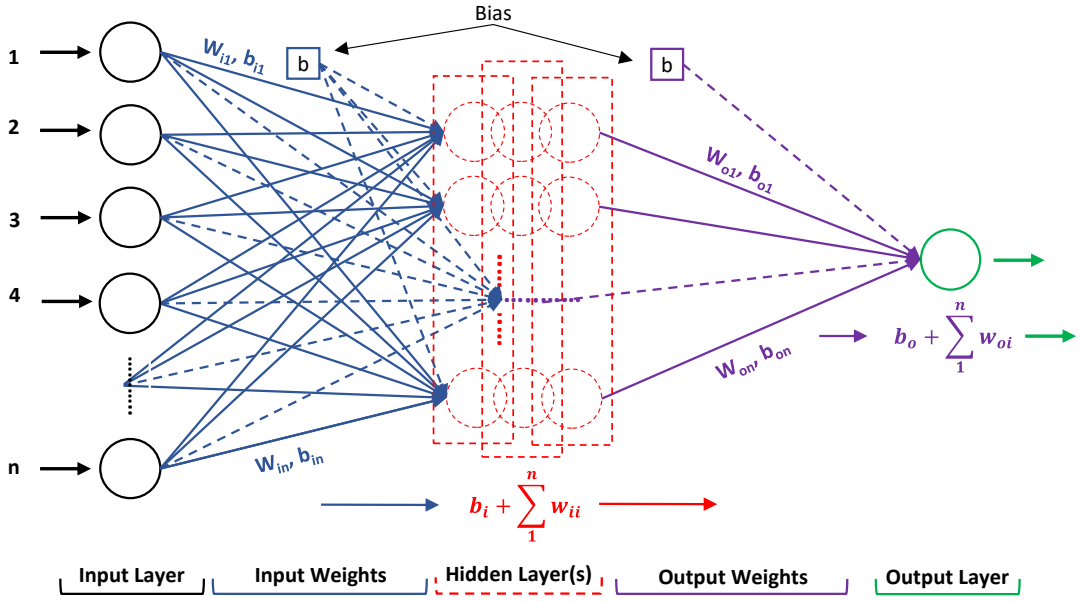

Fig. 3. Basic structure of ANN.

5.2. Adaptive neuro-fuzzy inference system (ANFIS)

ANFIS is a combination of ANN and a fuzzy inference system (FIS) introduced by Jang and Lee [93] for the purpose of surmounting the impediments of ANN and FIS. The ANFIS is based on fuzzy logic and rules, which are produced during the training process. As demonstrated in Fig. 4, FIS is comprised of five layers, whereby nodes in layers 1 and 5 are typical of the inputs and output, respectively. In the hidden layers, membership functions (MFs) and rules are represented as fixed and flexible nodes. The relation between the input and output is specified by 'if-then fuzzy' rules. The model then contains two fuzzy rules based on the form of 'Takagi and Sugeno' that can be represented as follows:

$$\textbf{Rule 1}: if\ x_1\ is\ A_1\ and\ x_2\ is\ B_1, then\ f_1 = p_1x_1 + q_1x_2 + r_1 \quad (1)$$

$$\textbf{Rule 2}: if\ x_1\ is\ A_2\ and\ x_2\ is\ B_2, then\ f_2 = p_2x_1 + q_2x_2 + r_2 \quad (2)$$

where x_1 and x_2 are the two input variables; f is the output variable; A_1, B_1, A_2 and B_2 stand for the linguistic labels; and p_1, q_1, r_1 and p_2, q_2, r_2 are the consequent and antecedent parameters. A brief discussion of each layer is presented below, while further details of ANFIS can be obtained from the literature [47,86].

Fuzzifying layer: Layer 1, in which neurons are considered as adaptive nodes that comprise premise parameters.

Implication layer: Layer 2, in which the neurons are labelled as Π_1 and Π_2 (see Fig. 4). In this layer, the output nodes (i.e., W_1 and W_1), which indicate the firing strength of a rule, are formed based on incoming signals.

Normalizing layer: Layer 3, in which every neuron (labelled as N_1 and N_2) is a fixed neuron. The output is achieved according to the ratio of the ith rule's firing strength over the summation of firing strength of all rules.

Defuzzifying layer: Layer 4, in which the neurons are categorized as adaptive neurons that contain consequence parameters.

Combining layer: Layer 5, contains a single neuron adding up all the inputs.

5.3. Methodological development of hybrid ANN and ANFIS models

Recent studies in the engineering field have focused on augmenting the performance of CML algorithms through OAs, such as PSO, GA, GWO, ABC, etc. [45,94]. The amalgamation of CML and OAs aids in the search for the exact global minimum by optimizing the learning parameters of CML algorithms, which in turn produces more consistent results, especially in the testing phase [41,48]. The learning parameters of ANN include input weights, hidden biases, output biases, and output weights [86], while those of ANFIS are consequent and antecedent (C&A) parameters [48]. In this study, ten swarm intelligence OAs were used to optimize the learning parameters of ANN and ANFIS, and twenty hybrid models, namely ANN-PSO, ANN-ACO, ANN-ABC, ANN-GWO, ANN-MFO, ANN-WOA, ANN-SSA, ANN-HHO, ANN-SMA, ANN-MPA, ANFIS-PSO, ANFIS-ACO, ANFIS-ABC, ANFIS-GWO, ANFIS-MFO, ANFIS-WOA, ANFIS-SSA, ANFIS-HHO, ANFIS-SMA, and ANFIS-MPA, were constructed to predict soil C_c. The optimization procedure includes the following steps: (a) initialization of ANN/ANFIS and selection of their hyper-parameters, i.e., activation function and the number of hidden neurons (N_H) for ANN, and the number of FIS parameters for ANFIS (N_{FIS}); (b) generation of learning parameters; (c) incorporation of OAs; (d) set terminating criteria; (e) training of algorithm using training dataset and generation of learning parameters through OAs; (f) final training of ANN/ANFIS; (g) check fitness value; and (h) selection of optimum values of optimized learning parameters based on the fitness value. After the development of optimized learning parameters, the constructed hybrid models were employed to predict the testing dataset. Note that, apart from the hyper-parameters of ANN and ANFIS, the deterministic parameters of OAs, such as number of swarm/particle size (N_S), iteration count (k), inertia weight (w), upper and lower bounds (ub and lb), E&E parameters, and other parameters also play a vital role in constructing the hybrid models [45]. Therefore, they must be carefully calibrated during the optimization process [41,94]. The entire aforementioned hybridisation process is presented in Fig. 5, showing the steps for constructing hybrid ANN and ANFIS models in predicting soil C_c.

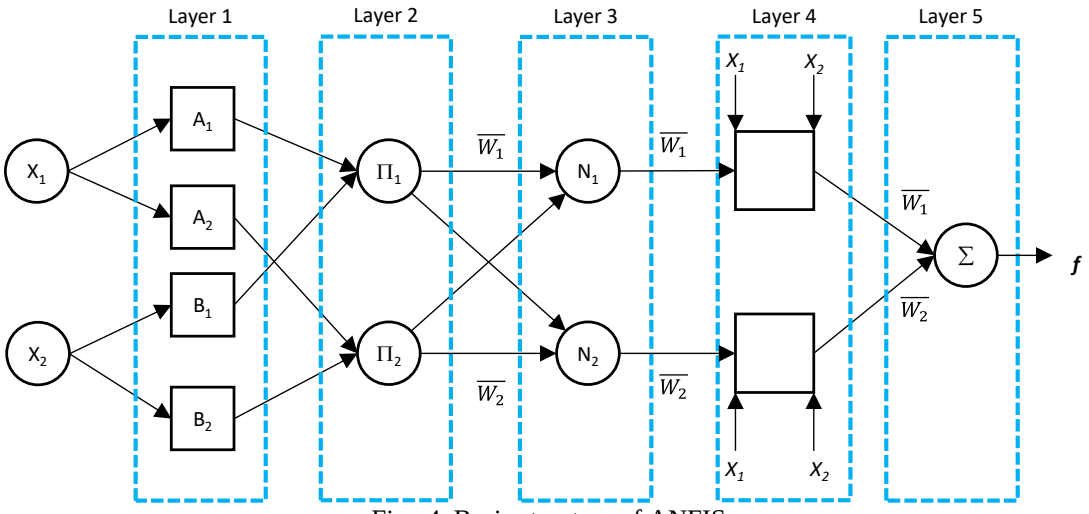

Fig. 4. Basic structure of ANFIS.

6. Descriptive statistics of the collected dataset

As stated above, a sum of 700 oedometer test data was acquired from a DFC project and used for estimating soil C_c. At the construction site, the investigation process was carried out at a depth of 0.5-32.75 m, and necessary soil samples were collected and tested by the executing agency. In total, there were 215 boreholes, from which samples of 6 different types of soils were collected and tested. Therefore, the collected dataset consists of oedometer test data of 6 different soils, namely CI (inorganic clay with intermediate plasticity), CL (inorganic clay with low plasticity), ML (low plasticity silt), ML-CL (clayey silt with low plasticity), SC (clayey sand), and SM-SC (silty sand with clayey sand) as per the Indian Standard Soil Classification System (ISSCS). Apart from the oedometer test results, i.e., the values of soil C_c, the depth of soil samples (D), details of grain size analysis, basic soil properties, and plasticity characteristics of soils, were also collected. The information extracted from the details of grain size analysis includes gravel content (G), content of coarse sand (CS), medium sand (MS), fine sand (FS), total sand (S), silt (M), clay (C), and total silt and clay (M&C). The basic soil parameters include bulk density (BD), dry density (DD), and specific gravity (Gs) of soils, whereas the plasticity characteristics of soils are PL and LL of soils. These parameters were used as the input information to predict the C_c of soils. Descriptive statistics of all the variables used in this study are given in Table 2. The frequency histograms as well as data distribution of the input soil parameters and soil C_c are shown in Fig. 6, from which the nature of experimental dataset i.e., type of distribution, skewness, etc., can be observed. In addition, to better highlight the diversity of soil C_c, the minimum, average, and maximum values of soil C_c are presented in Table 3 separately for all six soil types.

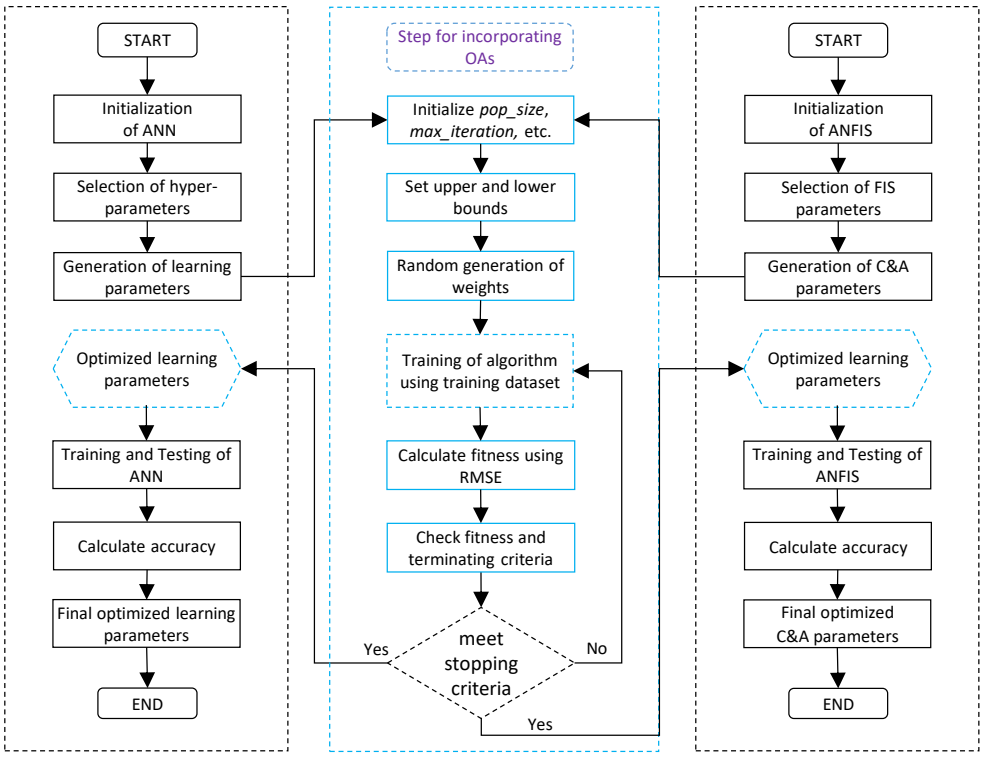

Fig. 5. Flow chart showing the process of hybrid ANN and ANFIS modelling.

From the descriptive statistics presented in Table 2, it can be observed that the present database contains test data of soil samples collected from depths ranging 0.50 m – 32.75 m. The G and S contents vary 0% – 18% and 3% – 68%, respectively, while the M&C content in the soil range between 28% and 97%. The content of CS, MS, and FS lies in the range of 0% – 11%, 0% – 25%, and 1% – 62%, respectively. Moreover, the LL and PL of soils

range 22% – 49% and 12% – 26%, respectively. The soil C_c is scattered in the range of 0.0700 to 0.1676, which shows that most of the samples have large M&C content. The details of C_c of each soil are presented in Table 3, revealing that, except ML soil, all other soils have a wide range of C_c values and, hence, can be considered significant for modelling the soil C_c. Also, the distribution of different soil parameters shown in Fig. 6 indicates that the parameter D. G, CS, and MS are not normally distributed and they are skewed to the left. The S, M, M&C, BD, DD, and Gs are some of the soil parameters that are normally distributed. The output parameter, C_c, is found to be skewed to the right and hence, not normally distributed. The skewness of each soil parameters can be obtained from Table 2.

Table 2. Descriptive statistics of the collected dataset.

Parameters and description		Statistical particulars						
		Min.	Avg.	Max.	Stnd. error	Stnd. dev.	Kurtosis	Skewness
D	Depth of soil samples (m)	0.50	7.48	32.75	0.21	5.55	1.48	1.18
G	Gravel content (%)	0.00	3.42	18.00	0.14	3.67	1.58	1.32
CS	Coarse sand content (%)	0.00	2.15	11.00	0.07	1.89	2.35	1.21
MS	Medium sand content (%)	0.00	3.90	25.00	0.13	3.41	7.18	2.21
FS	Fine sand content (%)	1.00	26.40	62.00	0.47	12.45	-0.42	0.45
S	Total sand content (%)	3.00	32.45	68.00	0.46	12.24	-0.49	0.13
M	Silt content (%)	19.00	48.01	75.00	0.36	9.65	0.15	0.26
C	Clay content (%)	5.00	16.11	37.00	0.19	4.99	0.05	0.45
M&C	Silt and clay content (%)	28.00	64.13	97.00	0.46	12.28	-0.39	0.03
BD	Bulk density (gm/cm^3)	1.61	1.80	1.99	0.00	0.07	-0.30	0.15
DD	Dry density (gm/cm^3)	1.46	1.59	1.71	0.00	0.05	0.12	-0.06
Gs	Sp. Gravity (-)	2.64	2.67	2.70	0.00	0.01	0.07	-0.08
LL	Liquid limit (%)	22.00	32.86	49.00	0.17	4.39	0.57	0.69
PL	Plastic limit (%)	12.00	18.84	26.00	0.07	1.76	1.97	-0.55
C_c	Compression index (-)	0.0700	0.1208	0.1676	0.0007	0.0198	0.1321	-0.4492

Table 3. Range of C_c for different soil types.

Particulars	Different soil types as per ISSCS					
	CI	CL	ML	ML-CL	SC	SM-SC
Min.	0.0700	0.0700	0.1008	0.0940	0.0857	0.0904
Avg.	0.1238	0.1201	0.1008	0.1025	0.1171	0.1186
Max.	0.1622	0.1676	0.1008	0.1097	0.1656	0.1490

7. Data processing and analysis

This section presents the data processing and analysis procedure for predicting soil C_c. It is worth noting that soils are made up of small particles, which were produced by the weathering of large rocks of various mineralogy into smaller rock fragments and finally soil particles. Thus, soils are categorized as heterogeneous materials and show nonlinear elastoplastic behaviour due to the presence of different size particles, such as gravel, sand, silt, and clay. The plasticity of soils is greatly affected by the mineralogy, size, and shape of their particles. Soils with a large content of G and S possess no plasticity because of their large particles, while soils containing high proportions of clay manifest a large plasticity. In addition, the proportions of different soil particles affect the consolidation behaviour of soils. Therefore, to precisely explain the mapping relationship between soil parameters and C_c, this study investigated two different combinations of input parameters, i.e., Case-1 and Case-2, to assess the influence of different soil particles, such as S, M, and C particles, in modelling the C_c of soils. Specifically, Case-1 included 9 input parameters, while Case-2 considered 12 input parameters. The details of two different input combinations are presented in Table 4. Note that the contents of CS, MS, FS, M, and C were not considered separately in Case-1; instead, S and M&C contents were included for predicting soil C_c.

Table 4. Details of input parameters of two different combinations.

Combination	Input parameters considered	Input data size
Case-1	D, G, S, M&C, BD, DD, G_S, LL, and PL	9 × 700
Case-2	D, G, CS, MS, FS, M, C, BD, DD, G_S, LL, and PL	12 × 700

To construct and validate the hybrid ANN and ANFIS models, the main dataset was randomly separated into training and testing subsets. In the field of ML, the most critical step for any type of problem is considered to be the processing of data. Generally, the normalization of data is a pre-processing task that is performed in the primary stage to nullify the dimensional effect of the input variables. Thus, prior to the model development, the entire dataset was standardized between 0 and 1, utilizing $\{(x - x_{min})/(x_{max} - x_{min})\}$, in which, x_{min} and x_{max} represent the minimum and maximum values of the parameter under consideration (i.e., x), respectively.

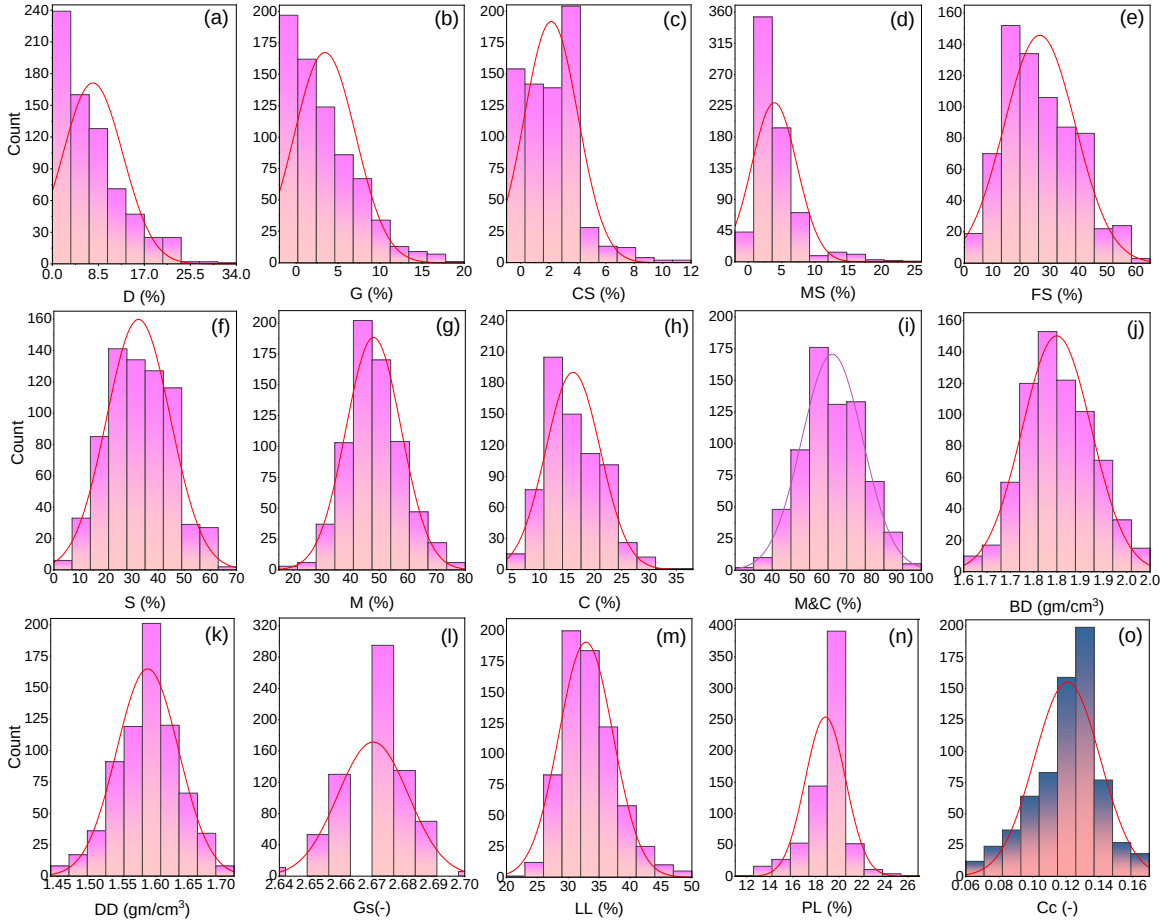

Fig. 6. Frequency histograms of input (a - n) and output (o) variables.

In this work, 80% of the total dataset was randomly selected for the training subset, while the remaining 20% was used as the testing subset. It is pertinent to mention here that, although there are no set standards or criteria for the number of samples to be used in a prediction model, it is of the researchers' choice; however, a model built with a large dataset can be deemed more reliable than one built with a small dataset. In addition, a model that has been validated with a large dataset is more trustworthy. Thus, in this work, 560 and 140 samples were used for the training and testing subsets to develop and validate the hybrid ANN and ANFIS models, respectively.

In the subsequent step, eight widely-used performance parameters, including Nash-Sutcliffe efficiency (NS), R^2, performance index (PI), Willmott`s index of agreement (WI), mean absolute error (MAE), root mean square error to observation's standard deviation ratio (RSR), root mean square error (RMSE), and weighted mean absolute percentage error (WMAPE), were used to assess the performance of the proposed hybrid models. The expressions of the above-mentioned indices are as follows:

$$NS = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - y_{mean})^2} \tag{3}$$

$$R^2 = \frac{\sum_{i=1}^{n}(y_i - y_{mean})^2 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - y_{mean})^2} \tag{4}$$

$$PI = adj.R^2 + 0.01VAF - RMSE \tag{5}$$

$$WI = 1 - \left[\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}\{|\hat{y}_i - y_{mean}| + |y_i - y_{mean}|\}^2}\right] \tag{6}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(\hat{y}_i - y_i)| \tag{7}$$

$$RSR = \frac{RMSE}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{mean})^2}} \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

$$WMAPE = \frac{\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times y_i}{\sum_{i=1}^{n} y_i} \tag{10}$$

where y_i and $\hat{y}_i$ are the actual and estimated i^{th} values; n is the number of data samples under consideration (irrespective of training and testing samples); and y_{mean} indicates the mean of the experimental values. Fig. 7 illustrates the entire steps of data analysis and computational modelling.

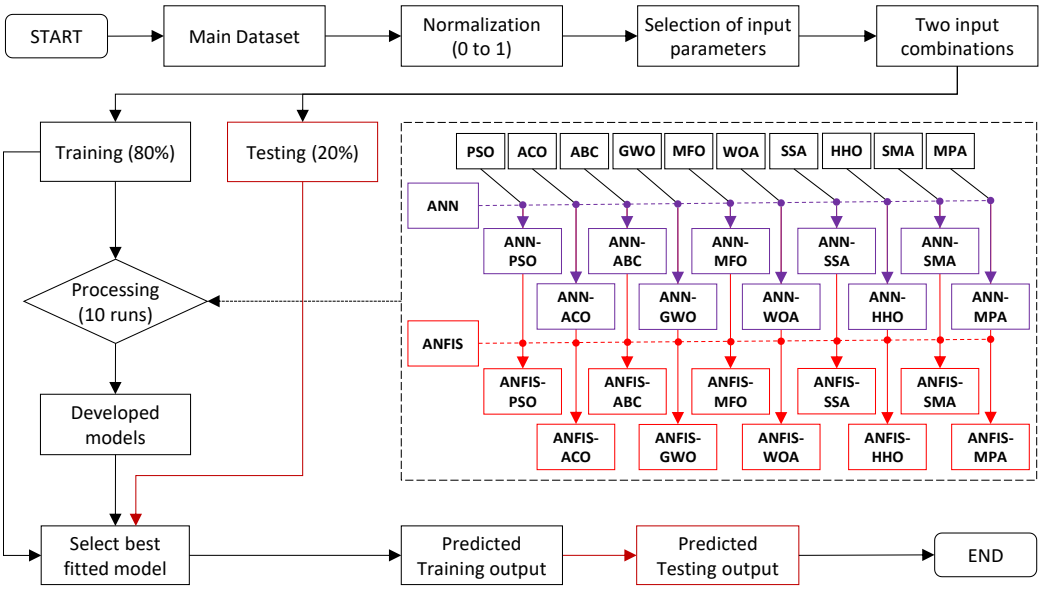

Fig. 7. Steps of data processing, analysis and computational modelling.

8. Results and discussions

The results of the developed hybrid models in predicting soil C_c are presented and discussed in this section. To construct and validate the hybrid ANN and ANFIS models, the main dataset was divided into training and testing subsets, as previously described. For predicting the C_c of soils, two alternative combinations of input parameters (i.e., Case-1 and Case-2) were explored in this study. Note that the training subset was used to construct the hybrid ANN and ANFIS models, whilst the testing subset was utilized to corroborate their predictive abilities for internal validation of the constructed models. In addition, the generalization capabilities of the constructed hybrid models were evaluated using a completely new dataset, referred to as external validation. It is important to note that a model with a better predictive accuracy achieved during the validation phase should be accepted with greater conviction; therefore, both internal and external validation were performed. Firstly, the testing dataset was used for internal validation, while separate oedometer tests were performed for external validation. Thirty new experiments were conducted in the geotechnical laboratory at NIT Patna, India, to generate separate datasets, which were then used for external validation of the developed models.

In the following sub-sections, a detailed comparative assessment of the outcomes of the developed hybrid models for the training dataset is presented, followed by validation of the prediction models. Before reporting the outcomes of proposed models, the parametric configurations of the OAs for constructing optimum hybrid models are presented and discussed.

8.1. Parametric configurations of the OAs

As mentioned earlier, to construct optimum hybrid models, it is compulsory to pre-specify the hyper-parameters of ANN and ANFIS. In this work, N_H in the hidden layer, ranging from 5 to 20, was investigated for the hybrid ANN models. Moreover, the Levenberg-Marquardt backpropagation and tan-sigmoid activation functions were examined to obtain the optimum structure of hybrid ANNs. The values of Ns and k were set as 50 and 200, respectively, and kept constant across all analyses. The performance of hybrid ANNs in the validation phase was employed to evaluate the appropriateness of N_H, the type of activation function, and values of other parameters of the OAs. Following a trial-and-error approach, the most appropriate values of N_H were obtained as 10 and kept constant for other hybrid ANN models. The most effective choice of hyper-parameters and other deterministic parameters of the optimised ANN models are given in Table 5 and Table 6 for Case-1 and Case-2 input combinations, respectively. Based on the optimum N_H value, i.e., N_H = 10, the total number of optimized weights and biases (O_{w+b}) was calculated as 111 (i.e., 9×10+10+10+1) for Case-1 and 141 (i.e., 12×10+10+10+1) for Case-2 of C_c prediction.

Analogous to hybrid ANN models, the value of N_{FIS} was set to 10 for constructing the hybrid ANFIS models. Subsequently, PSO, ACO, ABC, GWO, MFO, WOA, SSA, HHO, SMA, and MPA were incorporated to optimize the C&A parameters of ANFIS. After a preliminary experimental process, the Gaussian membership function was used in the input layer, while a linear membership function was used in the output layer. Thus, a sum of 190 C&A membership functions of ANFIS were optimized for the nine-dimensional input space, i.e., for Case-1 of C_c prediction (refer Table 5 for other parameters). Note that, these optimised parameters were chosen after trial-and-error tests regarding both input space dimensionality and the performance of the testing dataset.

It may be noted that the parametric configurations of the hybrid models discussed above are only for the first input combination, i.e., Case-1. Also note that, although the procedure for developing the ANN and ANFIS-based

hybrid models is the same, the values of optimized learning parameters of the developed models are different in each case. Nonetheless, the optimum configuration of hyper-parameters of ANN and ANFIS and deterministic parameters of OAs obtained during the modelling of the first input combination were used to construct the predictive models for the second input combination, i.e., for Case-2 with 12 input parameters (as listed in Table 4). This was done to make a fair comparison and assess the effect of different-sized particles (i.e., CS, MS, FS, M, and C) on predicting the C_c of soils. Details of the developed hybrid models along with their parametric configurations are given in Table 6 for the Case-2 input combination of soil C_c prediction.

Table 5. Configuration of OAs of optimum hybrid ANN and ANFIS models (for Case-1).

Parameters	**PSO**	**ACO**	**ABC**	**GWO**	**MFO**	**WOA**	**SSA**	**HHO**	**SMA**	**MPA**
N_H	10	10	10	10	10	10	10	10	10	10
N_{FIS}	10	10	10	10	10	10	10	10	10	10
N_S	50	50	50	50	50	50	50	50	50	50
k	200	200	200	200	200	200	200	200	200	200
w	0.40	-	-	-	-	-	-	-	-	-
ub, lb	+1, -1	+1, -1	+1, -1	+1, -1	+1, -1	+1, -1	+1, -1	+1, -1	+1, -1	+1, -1
C_1	1	-	-	-	-	-	-	-	-	-
C_2	2	-	-	-	-	-	-	-	-	-
z	-	-	-	-	-	-	-	-	0.2	-
No. of O_{w+b}	111	111	111	111	111	111	111	111	111	111
No. of C&A	190	190	190	190	190	190	190	190	190	190

Table 6. Configuration of OAs of optimum hybrid ANN and ANFIS models (for Case-2).

Parameters	**PSO**	**ACO**	**ABC**	**GWO**	**MFO**	**WOA**	**SSA**	**HHO**	**SMA**	**MPA**
N_H	10	10	10	10	10	10	10	10	10	10
N_{FIS}	10	10	10	10	10	10	10	10	10	10
N_S	50	50	50	50	50	50	50	50	50	50
k	200	200	200	200	200	200	200	200	200	200
w	0.40	-	-	-	-	-	-	-	-	-
ub, lb	+1, -1	+1, -1	+1, -1	+1, -1	+1, -1	+1, -1	+1, -1	+1, -1	+1, -1	+1, -1
C_1	1	-	-	-	-	-	-	-	-	-
C_2	2	-	-	-	-	-	-	-	-	-
z	-	-	-	-	-	-	-	-	0.2	-
No. of O_{w+b}	141	141	141	141	141	141	141	141	141	141
No. of C&A	250	250	250	250	250	250	250	250	250	250

It is mandatory to mention here that the stochastic nature of OAs often prevents them from observing the characteristics predicted by the hybrid model directly. In general, a stochastic model employs random variables to generate the intended output, which results in a wide variety of outcomes under different parametric configurations. Therefore, it requires multiple runs to guarantee that the best solution will be obtained. Taking this point into consideration, each hybrid model was tested 10 times and the model with the best performance was chosen based on the results acquired during the testing phase. Table 7 lists the error details of the constructed models for 10 run times. Herein, the best, worst, mean and the standard deviation of RMSE values for the training subset were obtained after performing 10 runs of each hybrid model. As can be seen, the ANN-GWO and ANFIS-PSO achieved a very high accuracy for inferring values of soil C_c with low average values of 0.0565 and 0.0410 in the Case-1 input combination, whereas the low average values of RMSE for the same models were 0.0494 and 0.0349, respectively, in the Case-2 input combination. In both cases, the ANN-GWO and ANFIS-PSO models attained a higher prediction accuracy, which can be evident by lowest RMSE values of 0.0527 and 0.0353 in Case-1 and 0.0465 and 0.0325 in Case-2 of soil C_c prediction, respectively.

647

648

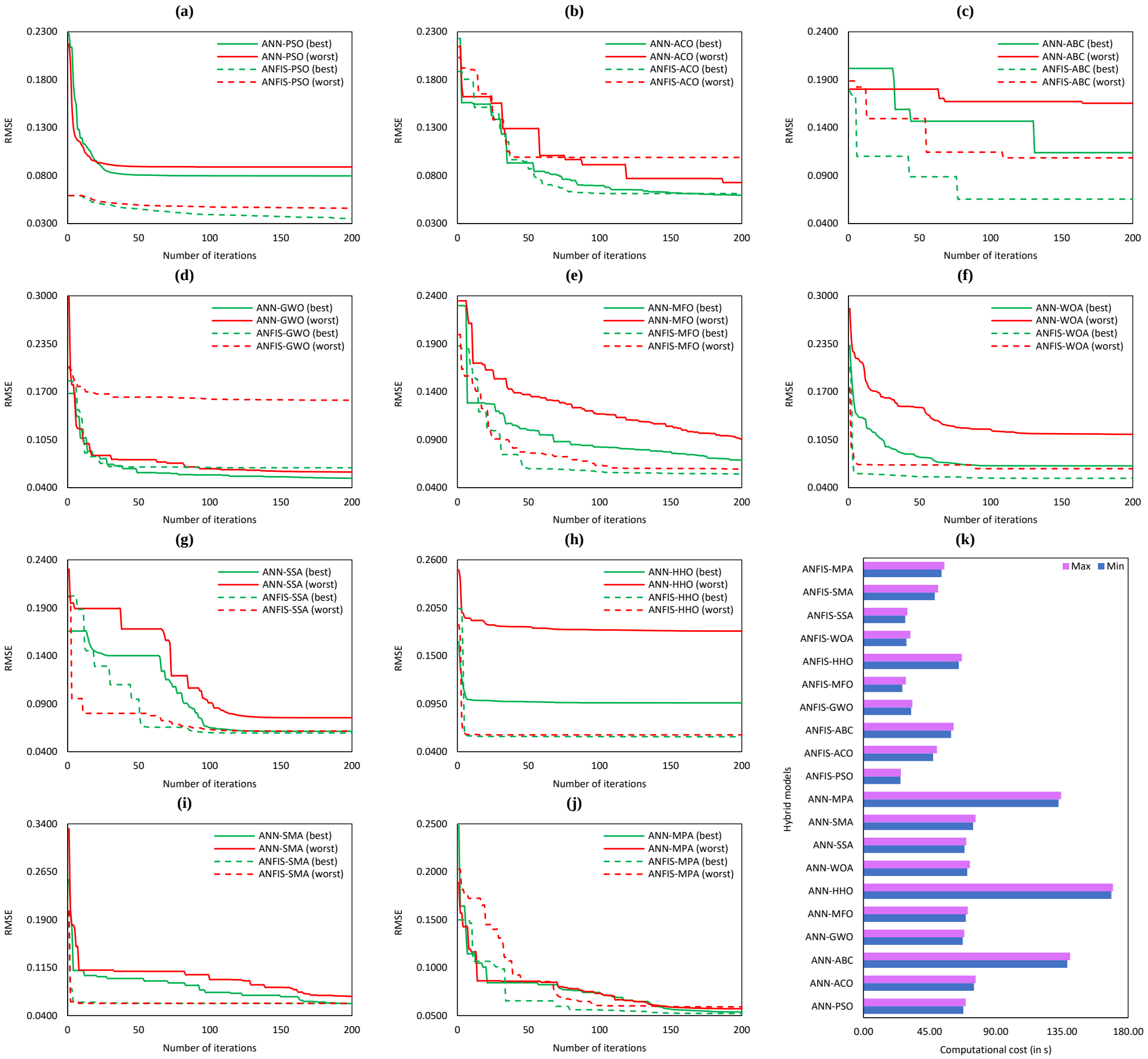

Fig. 8. (a-j) Convergence behaviour and (k) computational cost of the hybrid ANN models (Case-1).

649

650

651

652

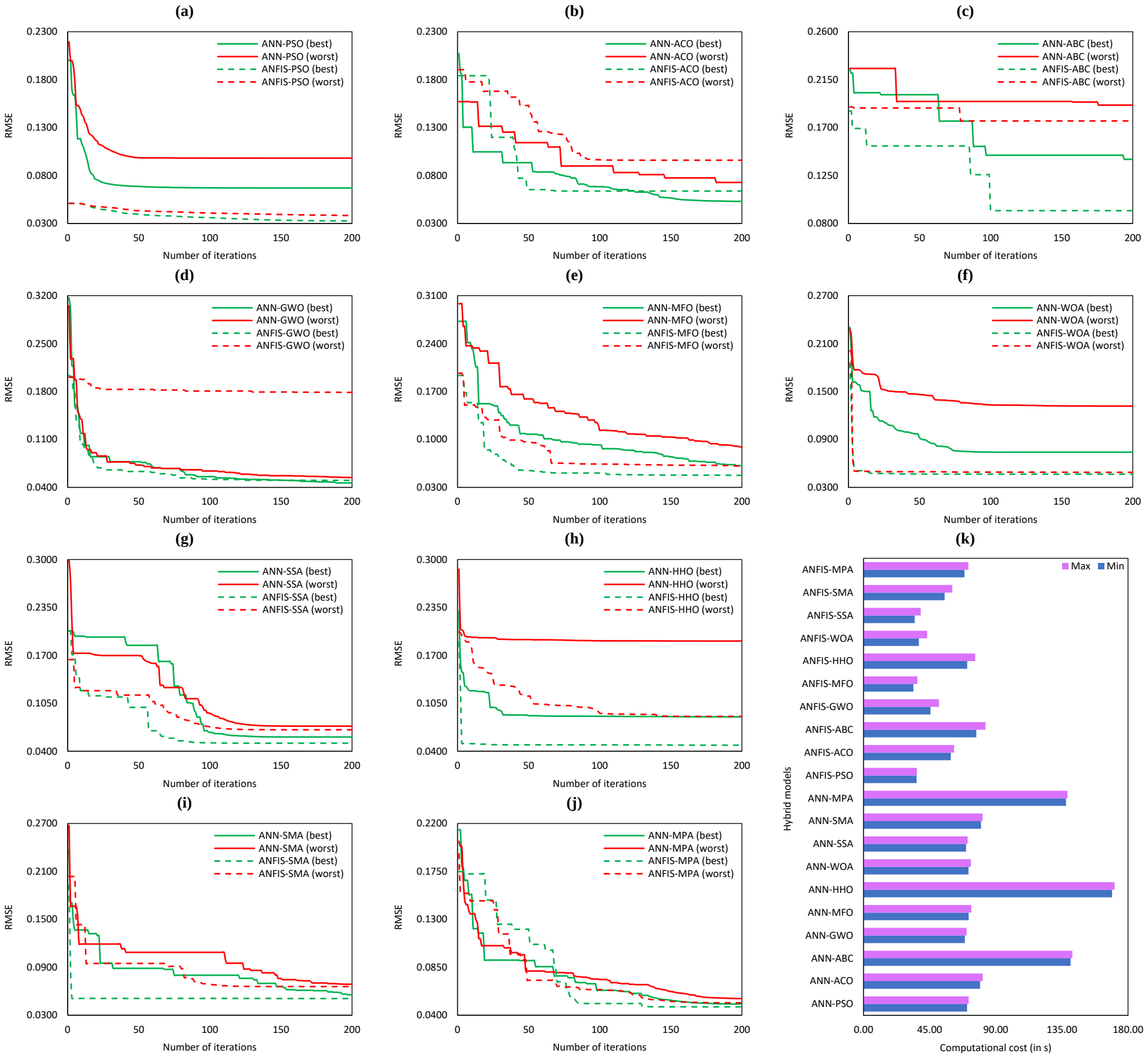

Fig. 9. (a-j) Convergence behaviour and (k) computational cost of the hybrid ANN models (Case-2).

653

654

It is also worth noting that the convergence behaviour of any OAs is critical in assessing its performance as it reveals the ability of OAs to escape from local minima and reach the faster solution. The convergence curves of the developed hybrid models for Case-1 and Case-2 are presented in Fig. 8(a-j) and Fig. 9(a-j), respectively. Herein, the best and worst convergence behaviours of all the developed models are presented. The details of computational costs are provided in Fig. 8(k) and Fig. 9(k) for the Case-1 and Case-2 input combinations, respectively. In both phases, the ANFIS-PSO achieved the desired convergence faster than other developed hybrid models. The minimum and maximum computational times were recorded as 25.180488 and 26.046925 s in Case-1 and 36.141250 and 36.345923 s in Case-2, respectively. It can also be observed that the hybrid ANFIS models achieved faster convergence than the hybrid ANN models in all cases (best hybrid ANFIS models are shown in green dotted line), indicating higher generalization ability. However, the computational times of the ANN-HHO, ANN-MPA, and ANN-ABC models were found to be on the higher side. On the other hand, the developed ANFIS-PSO and ANFIS-MFO models were found to have lower computing costs. Note that, all hybrid models were developed in MATLAB 2015a version with i5-8400 CPU @ 2.80 GHz, 8 GB RAM, and the computational costs were recorded in a non-parallel computing environment.

Table 7. Error details (RMSE value) of the hybrid ANN and ANFIS models after 10 runs.

Models	Particulars	PSO	ACO	ABC	GWO	MFO	WOA	SSA	HHO	SMA	MPA
ANN	Best	0.0797	0.0597	0.1140	**0.0527**	0.0687	0.0695	0.0612	0.0960	0.0589	0.0539
(Case-1)	Worst	0.0890	0.0729	**0.1655**	0.0611	0.0908	0.1123	0.0754	0.1783	0.0701	0.0587
	Mean	0.0818	0.0627	0.1484	**0.0565**	0.0749	0.0839	0.0658	0.1379	0.0634	0.0568
	Stnd. dev.	0.0031	0.0041	0.0184	0.0026	0.0069	0.0141	0.0044	0.0309	0.0033	**0.0016**
ANN	Best	0.0670	0.0530	0.1402	**0.0465**	0.0616	0.0740	0.0593	0.0867	0.0555	0.0505
(Case-2)	Worst	0.0981	0.0728	**0.1911**	0.0543	0.0883	0.1317	0.0741	0.1895	0.0688	0.0556
	Mean	0.0792	0.0570	0.1614	**0.0494**	0.0719	0.0910	0.0634	0.1333	0.0601	0.0522
	Stnd. dev.	0.0104	0.0065	0.0157	0.0025	0.0093	0.0171	0.0047	0.0340	0.0044	**0.0015**
ANFIS	Best	**0.0353**	0.0613	0.0655	0.0669	0.0541	0.0526	0.0595	0.0571	0.0591	0.0523
(Case-1)	Worst	0.0461	0.0990	**0.1085**	0.1585	0.0592	0.0657	0.0615	0.0592	0.0592	0.0592
	Mean	**0.0410**	0.0805	0.0892	0.0785	0.0573	0.0571	0.0603	0.0586	0.0592	0.0582
	Stnd. dev.	0.0037	0.0115	0.0136	0.0282	0.0018	0.0034	0.0007	0.0007	**0.0000**	0.0021
ANFIS	Best	**0.0325**	0.0638	0.0920	0.0502	0.0476	0.0465	0.0510	0.0483	0.0508	0.0478
(Case-2)	Worst	0.0382	0.0960	0.1762	**0.1787**	0.0602	0.0488	0.0692	0.0873	0.0658	0.0516
	Mean	**0.0349**	0.0826	0.1288	0.0756	0.0508	0.0476	0.0573	0.0541	0.0524	0.0503
	Stnd. dev.	0.0018	0.0088	0.0252	0.0373	0.0038	0.0009	0.0059	0.0117	0.0047	**0.0011**

8.2. Performance of the developed models

The predictive outcomes of the developed hybrid ANN and ANFIS models for estimating soil C_c are presented in Table 8, Table 9, Table 10, and Table 11 for both Case-1 and Case-2 input combinations. Herein, the models' performance in predicting the training outputs is reported. It should be noted that each model's performance with the training subset was used to express the goodness of fit of the constructed models. Based on the experimental results, ANN-GWO and ANFIS-PSO attained the highest R^2 and lowest RMSE values in both the Case-1 and Case-2 input combinations of soil C_c prediction. Among the developed hybrid ANN models, ANN-GWO attained the most desired accuracy with R^2 = 0.9326 and RMSE = 0.0527 in Case-1 and R^2 = 0.9475 and RMSE = 0.0465 in Case-2 of C_c prediction. On the other hand, ANFIS-PSO achieved the best prediction performance with R^2 = 0.9698 and RMSE = 0.0353 in Case-1 and R^2 = 0.9745 and RMSE = 0.0325 in Case-2 of C_c prediction. These findings demonstrate that the proposed hybrid models, ANN-GWO and ANFIS-PSO, have a good predictive performance in both cases of C_c prediction.

In addition, MAE and WMAPE values of the developed ANFIS-PSO model in the training phase were determined to be 0.0217 and 0.0396 in Case-1 and 0.0225 and 0.0412 in Case-2, respectively. For the ANN-GWO model, these performance indices were 0.0369 and 0.0676 in Case-1 and 0.0309 and 0.0563 in Case-2,

respectively. The values of R^2 are higher than 0.90 in both the Case-1 and Case-2 input combinations, which indicates that the proposed ANN-GWO and ANFIS-PSO obtained a good fit to the collected dataset. Overall, the developed ANN-GWO and ANFIS-PSO attained the most accurate prediction accuracy in both the Case-1 and Case-2 input combinations of C_c prediction, as evidenced by the score analysis; in both cases, the total score was determined to be 80.

Among the developed ANNs, ANN-MPA (total score = 72, R^2 = 0.9296 and RMSE = 0.0539) and ANN-SMA (total score = 63, R^2 = 0.9159 and RMSE = 0.0589) showed to be the second and third best models in Case-1, respectively, while ANN-MPA (total score = 72, R^2 = 0.9382 and RMSE = 0.0505) and ANN-ACO (total score = 64, R^2 = 0.9319 and RMSE = 0.0530) are the second and third best models in Case-2 of C_c prediction, respectively. Among the developed ANFIS models, ANFIS-MPA (total score = 71, R^2 = 0.9342 and RMSE = 0.0523) and ANFIS-WOA (total score = 65, R^2 = 0.9330 and RMSE = 0.0526) are the second and third best models in Case-1, respectively, while ANFIS-WOA (total score = 72, R^2 = 0.9478 and RMSE = 0.0465) and ANFIS-MFO (total score = 62, R^2 = 0.6452 and RMSE = 0.0476) are the second and third best models in Case-2 of C_c prediction. Observably, ANN-ABC achieved the lowest predictive performance with total score of 8 in both cases, whereas ANFIS-ABC was found to be least accurate model among the developed ANFIs models with total scores of 12 and 8 in Case-1 and Case-2 of soil C_c prediction, respectively. Fig. 10 and Fig. 11 show the experimental and predicted outcomes of the top three hybrid ANN (ANN-GWO, ANN-MPA and ANN-SMA/ANN-ACO) and ANFIS models (ANFIS-PSO, ANFIS-MPA/ANFIS-WOA, and ANFIS-WOA/ANFIS-MFO) based on total score value for Case-1 and Case-2 input combinations, respectively. It is worth noting that majority of the samples used in the training phase have a deviation less than ± 10% (points between the two red dotted lines in Fig. 10 and Fig. 11), implying that the developed models can reliably predict soil C_c.

Table 8. Performance of hybrid ANN models (training phase for Case-1).

Models/Particulars		NS	R^2	PI	WI	MAE	RSR	RMSE	WMAPE	Total score
ANN-PSO	Value	0.8468	0.8525	1.6166	0.9604	0.0557	0.3913	0.0797	0.1038	24
	Score	3	3	3	3	3	3	3	3	
ANN-ACO	Value	0.9143	0.9143	1.7672	0.9767	0.0441	0.2928	0.0597	0.0808	57
	Score	7	7	7	7	8	7	7	7	
ANN-ABC	Value	0.6871	0.6863	1.2520	0.9025	0.0843	0.5594	0.1140	0.1564	8
	Score	1	1	1	1	1	1	1	1	
ANN-GWO	Value	**0.9330**	**0.9326**	**1.8113**	**0.9824**	**0.0369**	**0.2588**	**0.0527**	**0.0676**	**80**
	Score	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	
ANN-MFO	Value	0.8863	0.8860	1.7009	0.9697	0.0523	0.3372	0.0687	0.0980	38
	Score	5	5	5	5	4	5	5	4	
ANN-WOA	Value	0.8838	0.8830	1.6946	0.9679	0.0506	0.3409	0.0695	0.0926	34
	Score	4	4	4	4	5	4	4	5	
ANN-SSA	Value	0.9098	0.9091	1.7556	0.9759	0.0456	0.3004	0.0612	0.0836	48
	Score	6	6	6	6	6	6	6	6	
ANN-HHO	Value	0.7783	0.7781	1.4560	0.9322	0.0720	0.4709	0.0960	0.1334	16
	Score	2	2	2	2	2	2	2	2	
ANN-SMA	Value	0.9164	0.9159	1.7715	0.9776	0.0443	0.2891	0.0589	0.0806	63
	Score	8	8	8	8	7	8	8	8	
ANN-MPA	Value	0.9301	0.9296	1.8042	0.9816	0.0390	0.2643	0.0539	0.0712	72
	Score	9	9	9	9	9	9	9	9	

Table 9. Performance of hybrid ANFIS models (training phase for Case-1).

Models/Particulars		NS	R^2	PI	WI	MAE	RSR	RMSE	WMAPE	Total score
ANFIS-PSO	Value	**0.9700**	**0.9698**	**1.9039**	**0.9924**	**0.0217**	**0.1732**	**0.0353**	**0.0396**	**80**
	Score	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	
ANFIS-ACO	Value	0.9095	0.9257	1.7722	0.9726	0.0426	0.3009	0.0613	0.0768	34
	Score	3	6	5	3	5	3	3	6	
ANFIS-ABC	Value	0.8967	0.8975	1.7279	0.9718	0.0485	0.3214	0.0655	0.0894	12
	Score	2	1	1	2	1	2	2	1	
ANFIS-GWO	Value	0.8923	0.9074	1.7305	0.9670	0.0465	0.3282	0.0669	0.0833	12
	Score	1	2	2	1	2	1	1	2	
ANFIS-MFO	Value	0.9295	0.9315	1.8056	0.9823	0.0395	0.2656	0.0541	0.0725	56
	Score	7	7	7	7	7	7	7	7	
ANFIS-WOA	Value	0.9333	0.9330	1.8121	0.9827	0.0374	0.2582	0.0526	0.0685	65
	Score	8	8	8	9	8	8	8	8	
ANFIS-SSA	Value	0.9148	0.9145	1.7678	0.9769	0.0433	0.2919	0.0595	0.0790	30
	Score	4	3	3	4	4	4	4	4	
ANFIS-HHO	Value	0.9215	0.9213	1.7840	0.9796	0.0421	0.2801	0.0571	0.0770	46
	Score	6	5	6	6	6	6	6	5	
ANFIS-SMA	Value	0.9158	0.9152	1.7699	0.9775	0.0436	0.2902	0.0591	0.0798	34
	Score	5	4	4	5	3	5	5	3	
ANFIS-MPA	Value	0.9342	0.9342	1.8147	0.9824	0.0348	0.2566	0.0523	0.0631	71
	Score	9	9	9	8	9	9	9	9	

Table 10. Performance of hybrid ANN models (training phase for Case-2).

Models/Particulars		NS	R^2	PI	WI	MAE	RSR	RMSE	WMAPE	Total score
ANN-PSO	Value	0.8920	0.8952	1.7176	0.9725	0.0486	0.3287	0.0670	0.0907	32
	Score	4	4	4	4	4	4	4	4	
ANN-ACO	Value	0.9323	0.9319	1.8092	0.9823	0.0381	0.2602	0.0530	0.0698	64
	Score	8	8	8	8	8	8	8	8	
ANN-ABC	Value	0.5265	0.6499	1.0410	0.8860	0.1107	0.6881	0.1402	0.2073	8
	Score	1	1	1	1	1	1	1	1	
ANN-GWO	Value	**0.9479**	**0.9475**	**1.8474**	**0.9865**	**0.0309**	**0.2283**	**0.0465**	**0.0563**	**80**
	Score	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	
ANN-MFO	Value	0.9086	0.9081	1.7524	0.9757	0.0459	0.3024	0.0616	0.0844	40
	Score	5	5	5	5	5	5	5	5	
ANN-WOA	Value	0.8681	0.8683	1.6594	0.9627	0.0524	0.3632	0.0740	0.0954	24
	Score	3	3	3	3	3	3	3	3	
ANN-SSA	Value	0.9152	0.9151	1.7685	0.9779	0.0417	0.2912	0.0593	0.0769	48
	Score	6	6	6	6	6	6	6	6	
ANN-HHO	Value	0.8192	0.8222	1.5537	0.9480	0.0625	0.4252	0.0867	0.1166	16
	Score	2	2	2	2	2	2	2	2	
ANN-SMA	Value	0.9259	0.9256	1.7939	0.9801	0.0393	0.2723	0.0555	0.0703	56
	Score	7	7	7	7	7	7	7	7	
ANN-MPA	Value	0.9386	0.9382	1.8245	0.9840	0.0343	0.2478	0.0505	0.0623	72
	Score	9	9	9	9	9	9	9	9	

Table 11. Performance of hybrid ANFIS models (training phase for Case-2).

Models/Particulars		NS	R^2	PI	WI	MAE	RSR	RMSE	WMAPE	Total score
ANFIS-PSO	Value	**0.9746**	**0.9745**	**1.9159**	**0.9936**	**0.0225**	**0.1595**	**0.0325**	**0.0412**	**80**
	Score	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	
ANFIS-ACO	Value	0.9020	0.9294	1.7653	0.9692	0.0424	0.3130	0.0638	0.0751	16
	Score	2	2	2	2	2	2	2	2	
ANFIS-ABC	Value	0.7961	0.8309	1.5485	0.9487	0.0671	0.4515	0.0920	0.1231	8
	Score	1	1	1	1	1	1	1	1	
ANFIS-GWO	Value	0.9394	0.9391	1.8266	0.9840	0.0335	0.2462	0.0502	0.0605	40
	Score	5	5	5	5	5	5	5	5	
ANFIS-MFO	Value	0.9455	0.9452	1.8416	0.9857	0.0325	0.2334	0.0476	0.0588	62
	Score	8	8	8	8	7	8	8	7	
ANFIS-WOA	Value	0.9479	0.9478	1.8477	0.9867	0.0312	0.2282	0.0465	0.0568	72
	Score	9	9	9	9	9	9	9	9	
ANFIS-SSA	Value	0.9375	0.9372	1.8219	0.9834	0.0343	0.2501	0.0510	0.0621	24
	Score	3	3	3	3	3	3	3	3	
ANFIS-HHO	Value	0.9439	0.9436	1.8376	0.9852	0.0331	0.2369	0.0483	0.0599	48
	Score	6	6	6	6	6	6	6	6	
ANFIS-SMA	Value	0.9377	0.9374	1.8226	0.9837	0.0335	0.2495	0.0508	0.0608	32
	Score	4	4	4	4	4	4	4	4	
ANFIS-MPA	Value	0.9449	0.9446	1.8400	0.9856	0.0325	0.2347	0.0478	0.0587	58
	Score	7	7	7	7	8	7	7	8	

719

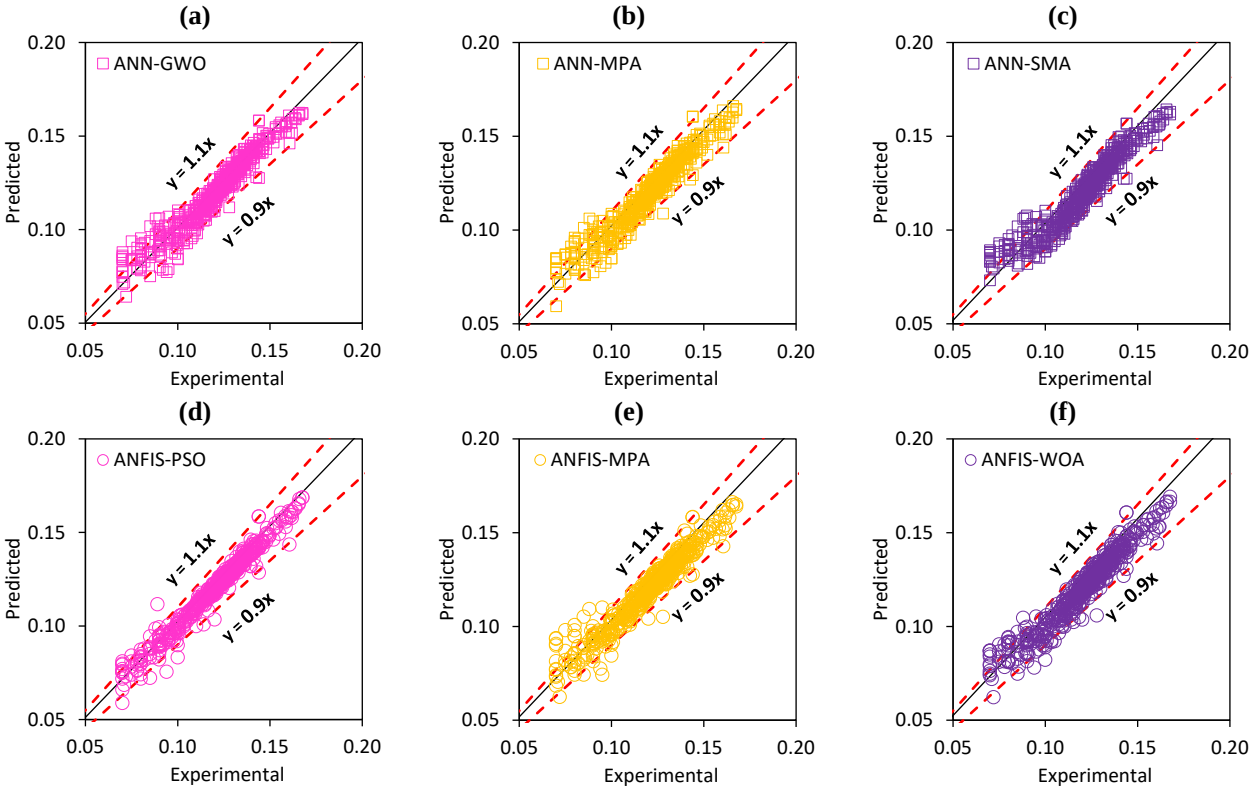

Fig. 10. Illustration of results for top three: (a-c) hybrid ANNs and (d-f) hybrid ANFIS models (Case-1 of training phase).

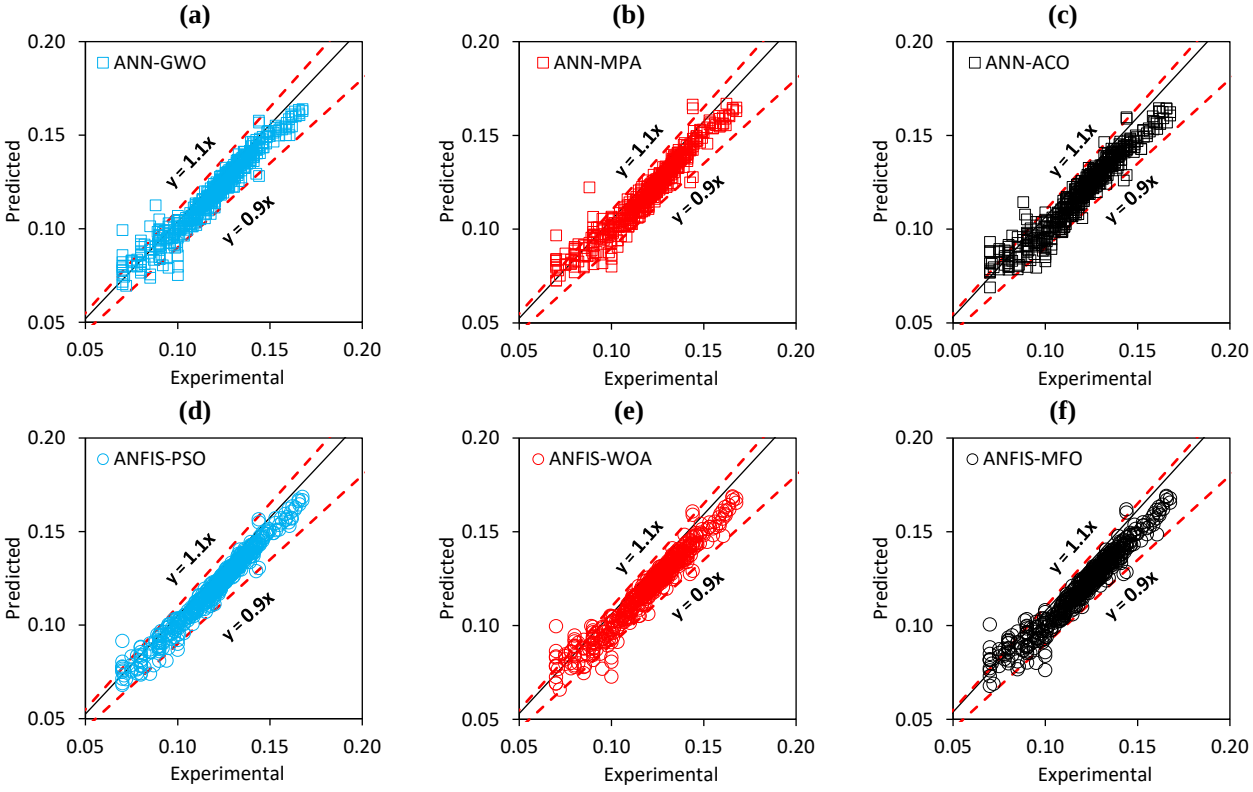

Fig. 11. Illustration of results for top three: (a-c) hybrid ANNs and (d-f) hybrid ANFIS models (Case-2 of training phase).

720
721

8.3. Validation of models

It is worth noting that a model with a better predictive accuracy in the validation phase is likely to be more accurate and robust. Furthermore, a higher predictive accuracy gained during the training phase does not always imply a better predictive model; therefore, proper model validation is always required. Thus, checking the correctness of any predictive model for a completely new dataset is practical, because a predictive model validated with additional datasets gathered from other experiments (executed in diverse conditions/environments) can be regarded practically significant. Considering these points as a reference, both internal and external validation were performed on the constructed hybrid models. As stated above, internal validation was carried out using the unused dataset, i.e., the testing dataset, whereas external validation was conducted via separate oedometer test results. A total of 30 additional experiments were carried out in the geotechnical laboratory at NIT Patna to generate separate datasets, which were then utilized to validate the developed models. In the following sub-sections, the generalization ability and robustness of the constructed hybrid models is presented and discussed.

8.3.1. Internal validation

As previously stated, internal validation was conducted on the unused dataset, i.e., the testing dataset (a subset of the main dataset), and the prediction results of the proposed models are presented in Table 12, Table 13, Table 14, and Table 15. Herein, the models' performances in predicting the testing output are reported and analysed. It is pertinent to mention that the performance of models with the testing dataset was used to validate their predictive ability. The performance indices showing high accuracy can be observed for both input combinations based on R^2 and RMSE criteria. Among the developed hybrid ANNs, ANN-GWO attained the best outcomes with R^2 = 0.9563, RMSE = 0.0447, and MAE = 0.0324 in Case-1 and R^2 = 0.9689, RMSE = 0.0368, and MAE = 0.0276 in Case-2 of C_c prediction. On the other hand, among the developed hybrid ANFIS models, ANFIS-PSO was determined to be the best performing model with R^2 = 0.9667, RMSE = 0.0382, MAE = 0.0265 in Case-1 and R^2 = 0.9789, RMSE = 0.0303, and MAE = 0.0212 in Case-2 of C_c prediction. Based on the results of score analysis, ANN-GWO and ANFIS-PSO secured the highest total scores of 80 in both cases, indicating superior predictive performance in the internal validation phase.

According to the findings of the score analysis, ANN-MPA is the second-best model in both cases among the hybrid ANNs developed in this study. The performance of the developed ANN-MPA model was recorded with a total score = 72, R^2 = 0.9519 and RMSE = 0.0468 in Case-1 and total score = 72, R^2 = 0.9644 and RMSE = 0.0401 in the Case-2 input combination. On the other hand, ANFIS-MPA was found to be the second-best model in both cases with a total score = 70, R^2 = 0.9575 and RMSE = 0.0431 in Case-1 and total score = 72, R^2 = 0.9734 and RMSE = 0.0354 in the Case-2 input combination. In both cases, MPA's performance in constructing hybrid ANN and ANFIS models was found to be quite excellent. However, ANN-ABC (total score of 8) and ANFIS-ABC (total scores of 9 in Case-1 and 9 in Case-2) achieved poor prediction performance in both cases of C_c prediction.

Table 12. Performance of hybrid ANN models (internal validation phase for Case-1).

Models/Particulars		NS	R^2	PI	WI	MAE	RSR	RMSE	WMAPE	Total score
ANN-PSO	Value	0.8182	0.8489	1.5727	0.9574	0.0609	0.4263	0.0852	0.1115	19
	Score	2	3	2	2	3	2	2	3	
ANN-ACO	Value	0.9350	0.9374	1.8173	0.9837	0.0387	0.2550	0.0510	0.0704	54
	Score	7	7	7	7	6	7	7	6	
ANN-ABC	Value	0.7111	0.7560	1.3574	0.9268	0.0822	0.5375	0.1075	0.1484	8
	Score	1	1	1	1	1	1	1	1	
ANN-GWO	Value	**0.9500**	**0.9563**	**1.8594**	**0.9880**	**0.0324**	**0.2237**	**0.0447**	**0.0592**	**80**
	Score	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	
ANN-MFO	Value	0.8955	0.9105	1.7360	0.9752	0.0484	0.3233	0.0646	0.0889	32
	Score	4	4	4	4	4	4	4	4	
ANN-WOA	Value	0.9311	0.9331	1.8085	0.9823	0.0386	0.2626	0.0525	0.0696	44
	Score	5	5	5	5	7	5	5	7	
ANN-SSA	Value	0.9355	0.9403	1.8214	0.9842	0.0371	0.2540	0.0508	0.0683	64
	Score	8	8	8	8	8	8	8	8	
ANN-HHO	Value	0.8384	0.8439	1.5911	0.9578	0.0641	0.4020	0.0804	0.1180	21
	Score	3	2	3	3	2	3	3	2	
ANN-SMA	Value	0.9331	0.9365	1.8150	0.9832	0.0405	0.2586	0.0517	0.0729	46
	Score	6	6	6	6	5	6	6	5	
ANN-MPA	Value	0.9453	0.9519	1.8488	0.9868	0.0345	0.2339	0.0468	0.0639	72
	Score	9	9	9	9	9	9	9	9	

Table 13. Performance of hybrid ANFIS models (internal validation phase for Case-1).

Models/Particulars		NS	R^2	PI	WI	MAE	RSR	RMSE	WMAPE	Total score
ANFIS-PSO	Value	**0.9635**	**0.9667**	**1.8906**	**0.9911**	**0.0265**	**0.1911**	**0.0382**	**0.0477**	**80**
	Score	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	
ANFIS-ACO	Value	0.9444	0.9497	1.8438	0.9865	0.0313	0.2359	0.0472	0.0572	65
	Score	8	7	8	8	9	8	8	9	
ANFIS-ABC	Value	0.9238	0.9284	1.7928	0.9811	0.0428	0.2760	0.0552	0.0781	9
	Score	1	1	1	2	1	1	1	1	
ANFIS-GWO	Value	0.9276	0.9372	1.8095	0.9794	0.0403	0.2691	0.0538	0.0718	15
	Score	2	2	2	1	2	2	2	2	
ANFIS-MFO	Value	0.9409	0.9488	1.8376	0.9859	0.0359	0.2431	0.0486	0.0658	48
	Score	6	6	6	6	6	6	6	6	
ANFIS-WOA	Value	0.9428	0.9502	1.8429	0.9863	0.0346	0.2393	0.0478	0.0633	57
	Score	7	8	7	7	7	7	7	7	
ANFIS-SSA	Value	0.9370	0.9408	1.8259	0.9842	0.0386	0.2509	0.0502	0.0705	35
	Score	5	4	4	4	4	5	5	4	
ANFIS-HHO	Value	0.9369	0.9426	1.8272	0.9845	0.0369	0.2512	0.0502	0.0676	37
	Score	4	5	5	5	5	4	4	5	
ANFIS-SMA	Value	0.9341	0.9405	1.8214	0.9839	0.0389	0.2567	0.0513	0.0713	24
	Score	3	3	3	3	3	3	3	3	
ANFIS-MPA	Value	0.9536	0.9575	1.8675	0.9885	0.0325	0.2153	0.0431	0.0592	70
	Score	9	9	9	9	8	9	9	8	

Table 14. Performance of hybrid ANN models (internal validation phase for Case-2).

Models/Particulars		NS	R^2	PI	WI	MAE	RSR	RMSE	WMAPE	Total score
ANN-PSO	Value	0.9047	0.9112	1.7490	0.9764	0.0466	0.3087	0.0617	0.0856	24
	Score	3	3	3	3	3	3	3	3	
ANN-ACO	Value	0.9536	0.9568	1.8647	0.9887	0.0312	0.2154	0.0431	0.0560	64
	Score	8	8	8	8	8	8	8	8	
ANN-ABC	Value	0.5887	0.6729	1.1132	0.8995	0.1006	0.6413	0.1282	0.1800	8
	Score	1	1	1	1	1	1	1	1	
ANN-GWO	Value	**0.9661**	**0.9689**	**1.8966**	**0.9917**	**0.0276**	**0.1842**	**0.0368**	**0.0496**	**80**
	Score	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	
ANN-MFO	Value	0.9213	0.9311	1.7964	0.9807	0.0425	0.2806	0.0561	0.0774	40
	Score	5	5	5	5	5	5	5	5	
ANN-WOA	Value	0.9164	0.9169	1.7701	0.9780	0.0430	0.2891	0.0578	0.0779	32
	Score	4	4	4	4	4	4	4	4	
ANN-SSA	Value	0.9445	0.9493	1.8432	0.9865	0.0360	0.2356	0.0471	0.0651	48
	Score	6	6	6	6	6	6	6	6	
ANN-HHO	Value	0.8167	0.8276	1.5576	0.9474	0.0642	0.4281	0.0856	0.1152	16
	Score	2	2	2	2	2	2	2	2	
ANN-SMA	Value	0.9507	0.9530	1.8574	0.9876	0.0330	0.2221	0.0444	0.0593	56
	Score	7	7	7	7	7	7	7	7	
ANN-MPA	Value	0.9597	0.9644	1.8817	0.9903	0.0305	0.2007	0.0401	0.0553	72
	Score	9	9	9	9	9	9	9	9	

Table 15. Performance of hybrid ANFIS models (internal validation phase for Case-2).

Models/Particulars		NS	R^2	PI	WI	MAE	RSR	RMSE	WMAPE	Total score
ANFIS-PSO	Value	**0.9771**	**0.9789**	**1.9246**	**0.9944**	**0.0212**	**0.1514**	**0.0303**	**0.0387**	**80**
	Score	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	
ANFIS-ACO	Value	0.9461	0.9662	1.8663	0.9841	0.0347	0.2323	0.0464	0.0616	16
	Score	2	2	2	2	2	2	2	2	
ANFIS-ABC	Value	0.8384	0.8943	1.6617	0.9640	0.0639	0.4020	0.0804	0.1142	8
	Score	1	1	1	1	1	1	1	1	
ANFIS-GWO	Value	0.9657	0.9678	1.8952	0.9915	0.0284	0.1852	0.0370	0.0520	30
	Score	4	3	3	4	4	4	4	4	
ANFIS-MFO	Value	0.9667	0.9687	1.8976	0.9917	0.0280	0.1826	0.0365	0.0512	50
	Score	6	6	7	6	6	6	6	7	
ANFIS-WOA	Value	0.9643	0.9703	1.8967	0.9914	0.0288	0.1890	0.0378	0.0527	31
	Score	3	8	5	3	3	3	3	3	
ANFIS SSA	Value	0.9676	0.9694	1.9003	0.9919	0.0280	0.1799	0.0360	0.0512	60
	Score	8	7	8	8	7	8	8	6	
ANFIS-HHO	Value	0.9672	0.9684	1.8974	0.9918	0.0275	0.1812	0.0362	0.0503	54
	Score	7	4	6	7	8	7	7	8	
ANFIS-SMA	Value	0.9657	0.9684	1.8962	0.9915	0.0283	0.1851	0.0370	0.0519	39
	Score	5	5	4	5	5	5	5	5	
ANFIS-MPA	Value	0.9687	0.9734	1.9074	0.9924	0.0273	0.1770	0.0354	0.0499	72
	Score	9	9	9	9	9	9	9	9	

To better demonstrate the prediction outcomes of the developed models, the illustrations of experimental and predicted values are presented in Fig. 12 and Fig. 13 for Case-1 and Case-2 input combinations, respectively. Herein, the scatterplots of only the top three hybrid ANN (ANN-GWO, ANN-MPA and ANN-SSA/ANN-ACO) and ANFIS (ANFIS-PSO, ANFIS-MPA and ANFIS-ACO) models (based on total score value) are presented. As can be seen, the majority of the predicted datasets have less than ± 10% deviation, implying the superiority and robustness of the developed models. This means that the points between the dotted lines have a ratio of actual to predicted values between 0.9 and 1.1. In addition, consistent behaviour can be observed in Fig. 12 and Fig. 13, with predictions closely fitting to experimental values over the entire range of soil C_c.

8.3.2. External validation

It may be noted that a model with a higher predicted accuracy during the validation phase should be regarded as robust and adopted with greater confidence. Yet, this statement can be deemed partially correct due to the fact that a model constructed and validated with the same source dataset (generally, training and testing datasets are generated from a single source dataset) can perform well in both training and testing phases; however, the significant performance level of such models must be ensured for a completely new dataset. This means that evaluating the generalisation potential of a predictive model using only training and testing datasets is unreliable unless the model's performance is validated using a completely new dataset. This not only ensures robustness a predictive model, but also demonstrates its generalisation capabilities in predicting the intended output(s) for future use.

787

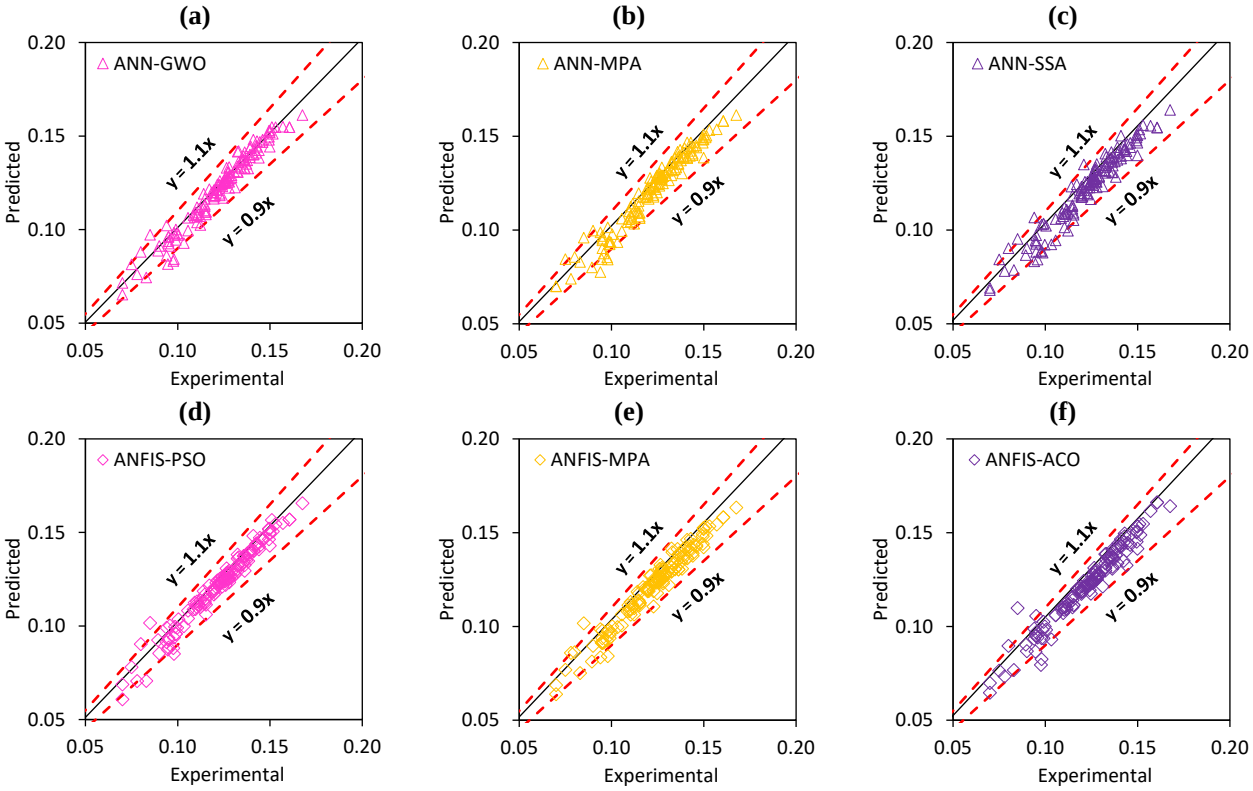

Fig. 12. Illustration of results for top three: (a-c) hybrid ANNs and (d-f) hybrid ANFIS models (Case-1 of internal validation).

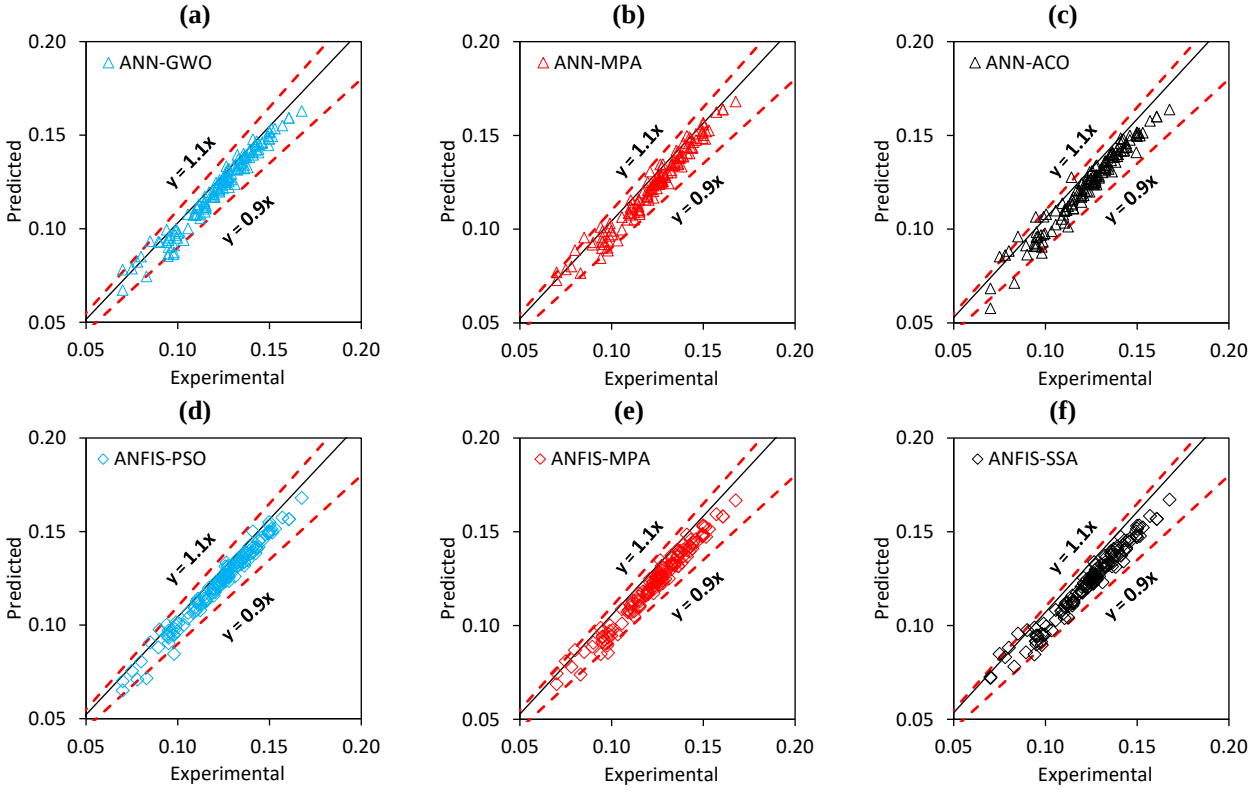

Fig. 13. Illustration of results top three: (a-c) hybrid ANNs and (d-f) hybrid ANFIS models (Case-2 of internal validation).

788
789

Using the preceding point as a reference, this study evaluated the performance of the constructed hybrid models on a completely new dataset, which is referred to as external validation. Thirty new oedometer tests were performed at the geotechnical laboratory of NIT Patna for this purpose, and the results were used for external validation of the established models in estimating soil C_c. The descriptive statistics of the new experimental datasets are given in Table 16. As can be observed from Table 16, the soils have a G content between 0% and 8%, while the S content in the soils ranges 15% to 40%. The contents of CS, MS, and FS are scattered in the ranges of 0% - 8%, 2% - 14%, and 5% - 31%, respectively, which indicates that most of the soils have fewer proportions of G, CS, and FS contents. The content of M&C lies between 57% and 84%, in which the content of M lies in the range of 23% to 72%. In addition, LL ranges between 29% and 49%, while PL indicates that the plasticity of soils varies from 12% to 22%. It is also seen that the newly performed laboratory tests have a wide range of C_c values (between 0.0150 and 0.1152) with a high standard deviation of 0.0191, indicating that the current database covers a wide range of soil samples and can be considered very reliable for external validation of the developed hybrid models. Fig. 14 shows photographs of experimental setups for a) grain size analysis, b) plasticity analysis, and c) samples, sample set up, and oedometer testing. It is important to mention that, all the laboratory experiments were performed in accordance with the applicable Indian Standards code of practice.

Table 16. Descriptive statistics of 30 new experimental datasets.

Parameters and description		Statistical properties						
		Min.	Avg.	Max.	Stnd. error	Stnd. dev.	Kurtosis	Skewness
D	Depth of soil samples (m)	3.00	5.90	9.00	0.39	2.16	-0.95	0.05
G	Gravel content (%)	0.00	2.43	8.00	0.41	2.24	-0.19	0.70
CS	Coarse sand content (%)	0.00	2.60	8.00	0.41	2.22	-0.44	0.73
MS	Medium sand content (%)	2.00	6.37	14.00	0.54	2.98	0.49	0.78
FS	Fine sand content (%)	5.00	17.80	31.00	1.22	6.66	-0.76	0.28
S	Total sand content (%)	15.00	26.77	40.00	1.30	7.12	-1.16	0.09
M	Silt content (%)	23.00	56.07	72.00	1.86	10.21	2.30	-0.89
C	Clay content (%)	9.00	14.73	39.00	1.05	5.77	10.38	2.74
M&C	Silt and clay content (%)	57.00	70.80	84.00	1.42	7.79	-1.10	0.04
BD	Bulk density (gm/cm^3)	1.80	1.89	1.96	0.01	0.05	-0.78	-0.33
DD	Dry density (gm/cm^3)	1.60	1.66	1.84	0.01	0.04	11.26	2.70
Gs	Sp. Gravity (-)	2.66	2.68	2.69	0.00	0.01	-0.24	-0.16
LL	Liquid limit (%)	29.00	33.80	49.00	0.89	4.85	2.83	1.75
PL	Plastic limit (%)	12.00	16.17	22.00	0.43	2.34	0.28	0.62
C_c	Compression index (-)	0.0150	0.0809	0.1152	0.0035	0.0191	3.9633	-1.1399

The prediction performances of all the hybrid models are presented in Table 17, Table 18, Table 19, and Table 20. As can be observed, the ANN-GWO and ANFIS-PSO models attained the best prediction in all performance matrices. Based on the R^2, RMSE and MAE criteria, ANN-GWO exhibited the best prediction performance among the constructed hybrid ANN models with $R^2 = 0.8640$, RMSE = 0.0923, and MAE = 0.0666 in Case-1 and $R^2 = 0.8709$, RMSE = 0.0760, and MAE = 0.0594 in the Case-2 input combination. Based on the score analysis, ANN-SMA and ANN-WOA were found to be the second-best models with $R^2 = 0.8175$, RMSE = 0.1065, and MAE = 0.0848 in Case-1 and $R^2 = 0.7976$, RMSE = 0.0892, and MAE = 0.0729 in Case-2 of soil C_c prediction. The values of total score were recorded as 66 and 65 in Case-1 and Case-2, respectively.

Analogous to hybrid ANNs, experimental outcomes exhibit that ANFIS-PSO outperformed other models in both cases with $R^2 = 0.9133$, RMSE = 0.0583, and MAE = 0.0417 in Case-1 and $R^2 = 0.9286$, RMSE = 0.0519, and MAE = 0.0433 in Case-2 of C_c prediction. The results of score analysis indicate that ANFIS-SSA (total score = 69) and ANFIS-SMA (total score = 67) are the second-best models in Case-1 and Case-2 of soil C_c prediction, respectively. However, ANFIS-ABC (total score = 14, $R^2 = 0.6697$ and RMSE = 0.1247) and ANFIS-ACO (total score = 10, $R^2 = 0.8737$ and RMSE = 0.1452) are the worst performing hybrid ANFIS models.

To further demonstrate the prediction outputs of the hybrid models, illustrations of actual and estimated values for Case-1 and Case-2 input combinations are presented in Fig. 15 in the form of scatter (on the left side of the figure) and line plots. Herein, the performance of only the top-performing models, i.e., ANN-GWO and ANFIS-PSO, is presented. The closer the line graphs show the difference between the experimental and estimated values, the more accurate the estimation. As can be observed, the ANFIS-PSO achieved the highest level of accuracy in both cases of input combinations.

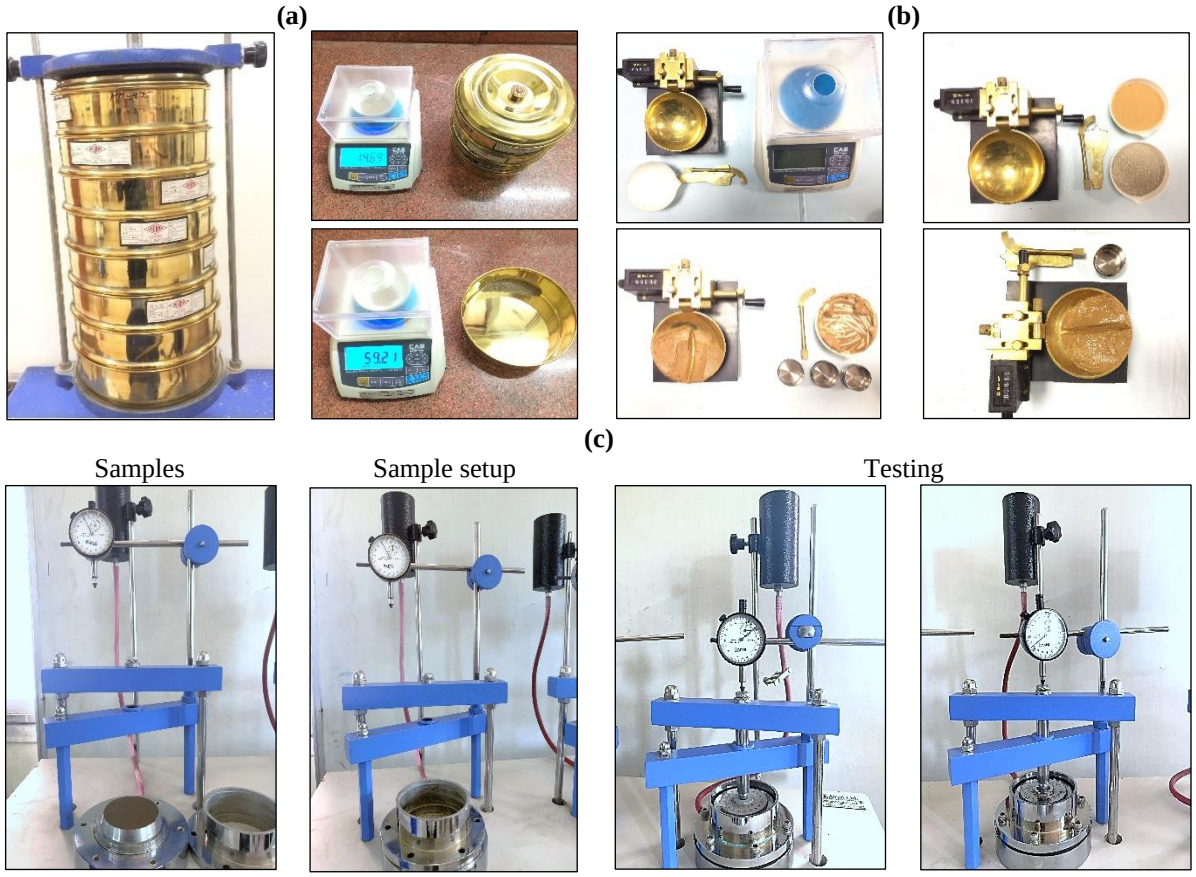

Fig. 14. Experiment set up at NIT Patna Laboratory: a) sieve analysis; b) PL test, and c) consolidation test.

Table 17. Performance of hybrid ANN models (external validation phase for Case-1).

Models/Particulars		NS	R^2	PI	WI	MAE	RSR	RMSE	WMAPE	Total score
ANN-PSO	Value	0.5423	0.5512	0.7702	0.8273	0.0818	0.6765	0.1299	0.4639	44
	Score	5	4	4	5	8	5	5	8	
ANN-ACO	Value	0.6522	0.6930	1.1054	0.8669	0.0814	0.5898	0.1133	0.4719	64
	Score	8	8	8	8	9	8	8	7	
ANN-ABC	Value	-1.2043	0.3088	0.0125	0.5007	0.2400	1.4847	0.2851	1.3870	8
	Score	1	1	1	1	1	1	1	1	
ANN-GWO	Value	**0.7692**	**0.8640**	**1.4885**	**0.9126**	**0.0666**	**0.4804**	**0.0923**	**0.3763**	**80**
	Score	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	
ANN-MFO	Value	0.2357	0.3350	0.1278	0.7428	0.1378	0.8743	0.1679	0.9308	16
	Score	2	2	2	2	2	2	2	2	
ANN-WOA	Value	0.5711	0.5957	0.8822	0.8643	0.1022	0.6549	0.1258	0.6942	44
	Score	6	5	5	7	5	6	6	4	
ANN-SSA	Value	0.4933	0.6296	0.9546	0.8347	0.1039	0.7118	0.1367	0.5931	39
	Score	4	6	6	6	4	4	4	5	
ANN-HHO	Value	0.4738	0.5178	0.6766	0.8211	0.1156	0.7254	0.1393	0.7375	25
	Score	3	3	3	4	3	3	3	3	
ANN-SMA	Value	0.6925	0.8175	1.3474	0.8758	0.0848	0.5546	0.1065	0.4876	66
	Score	9	9	9	9	6	9	9	6	
ANN-MPA	Value	0.5859	0.6786	1.0220	0.8195	0.0826	0.6435	0.1236	0.4136	54
	Score	7	7	7	3	7	7	7	9	

Table 18. Performance of hybrid ANFIS models (external validation phase for Case-1).

Models/Particulars		NS	R^2	PI	WI	MAE	RSR	RMSE	WMAPE	Total score
ANFIS-PSO	Value	**0.9079**	**0.9133**	**1.7291**	**0.9757**	**0.0417**	**0.3035**	**0.0583**	**0.3393**	**80**
	Score	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	
ANFIS-ACO	Value	0.2202	0.8691	1.4539	0.7911	0.1492	0.8831	0.1696	0.8898	16
	Score	1	8	2	1	1	1	1	1	
ANFIS-ABC	Value	0.5784	0.6697	1.0392	0.8853	0.0917	0.6493	0.1247	0.6836	14
	Score	2	1	1	2	2	2	2	2	
ANFIS-GWO	Value	0.7731	0.8246	1.4770	0.9343	0.0752	0.4764	0.0915	0.5061	35
	Score	5	2	3	5	5	5	5	5	
ANFIS-MFO	Value	0.7532	0.8512	1.5357	0.9294	0.0790	0.4968	0.0954	0.5118	27
	Score	3	4	5	3	3	3	3	3	
ANFIS-WOA	Value	0.7642	0.8444	1.5219	0.9322	0.0780	0.4856	0.0933	0.5103	31
	Score	4	3	4	4	4	4	4	4	
ANFIS-SSA	Value	0.8631	0.8650	1.5980	0.9626	0.0505	0.3700	0.0711	0.3875	69
	Score	9	7	8	9	9	9	9	9	
ANFIS-HHO	Value	0.8223	0.8741	1.6030	0.9477	0.0699	0.4215	0.0810	0.4662	54
	Score	6	9	9	6	6	6	6	6	
ANFIS-SMA	Value	0.8330	0.8618	1.5788	0.9513	0.0661	0.4087	0.0785	0.4560	59
	Score	8	5	6	8	8	8	8	8	
ANFIS-MPA	Value	0.8322	0.8620	1.5794	0.9511	0.0664	0.4096	0.0787	0.4578	55
	Score	7	6	7	7	7	7	7	7	

Table 19. Performance of hybrid ANN models (external validation phase for Case-2).

Models/Particulars		NS	R^2	PI	WI	MAE	RSR	RMSE	WMAPE	Total score
ANN-PSO	Value	0.5536	0.7225	1.0987	0.8129	0.1003	0.6681	0.1283	0.5397	27
	Score	3	5	3	2	2	3	3	6	
ANN-ACO	Value	0.6932	0.7900	1.3031	0.8784	0.0745	0.5539	0.1064	0.3944	54
	Score	6	7	7	5	8	6	6	9	
ANN-ABC	Value	-0.0361	0.4242	0.1700	0.7505	0.1615	1.0179	0.1955	0.9632	8
	Score	1	1	1	1	1	1	1	1	
ANN-GWO	Value	**0.8436**	**0.8709**	**1.5865**	**0.9499**	**0.0594**	**0.3955**	**0.0760**	**0.3833**	**80**
	Score	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	
ANN-MFO	Value	0.6432	0.6963	1.1233	0.9028	0.0928	0.5973	0.1147	0.6155	35
	Score	5	3	4	6	4	5	5	3	
ANN-WOA	Value	0.7844	0.7976	1.4041	0.9435	0.0729	0.4644	0.0892	0.5539	65
	Score	9	8	8	9	9	9	9	4	
ANN-SSA	Value	0.5871	0.7080	1.1453	0.8639	0.0964	0.6426	0.1234	0.6156	30
	Score	4	4	5	4	3	4	4	2	
ANN-HHO	Value	0.5428	0.5507	0.7693	0.8317	0.0890	0.6761	0.1299	0.5524	23
	Score	2	2	2	3	5	2	2	5	
ANN-SMA	Value	0.7292	0.8434	1.4734	0.9082	0.0793	0.5203	0.0999	0.4414	61
	Score	7	9	9	8	6	7	7	8	
ANN-MPA	Value	0.7347	0.7589	1.2911	0.9068	0.0750	0.5151	0.0989	0.4850	57
	Score	8	6	6	7	7	8	8	7	

Table 20. Performance of hybrid ANFIS models (external validation phase for Case-2).

Models/Particulars		NS	R^2	PI	WI	MAE	RSR	RMSE	WMAPE	Total score
ANFIS-PSO	Value	**0.9270**	**0.9286**	**1.7728**	**0.9804**	**0.0433**	**0.2703**	**0.0519**	**0.3323**	**80**
	Score	**10**	**10**	**10**	**10**	**10**	**10**	**10**	**10**	
ANFIS-ACO	Value	0.4288	0.8737	1.5294	0.8488	0.1262	0.7558	0.1452	0.7711	10
	Score	1	2	2	1	1	1	1	1	
ANFIS-ABC	Value	0.5841	0.6848	1.0800	0.8872	0.0866	0.6449	0.1239	0.6387	14
	Score	2	1	1	2	2	2	2	2	
ANFIS-GWO	Value	0.8149	0.8993	1.6652	0.9477	0.0690	0.4303	0.0826	0.4289	24
	Score	3	3	3	3	3	3	3	3	
ANFIS-MFO	Value	0.8875	0.9143	1.7218	0.9683	0.0543	0.3355	0.0644	0.3504	41
	Score	5	5	5	5	5	5	5	6	
ANFIS-WOA	Value	0.8659	0.9048	1.6942	0.9627	0.0603	0.3662	0.0703	0.4015	32
	Score	4	4	4	4	4	4	4	4	
ANFIS-SSA	Value	0.9130	0.9195	1.7455	0.9775	0.0465	0.2949	0.0566	0.3651	62
	Score	7	9	9	9	9	7	7	5	
ANFIS-HHO	Value	0.8978	0.9193	1.7383	0.9715	0.0524	0.3196	0.0614	0.3446	51
	Score	6	8	6	6	6	6	6	7	
ANFIS-SMA	Value	0.9138	0.9187	1.7437	0.9765	0.0472	0.2937	0.0564	0.3353	67
	Score	9	7	8	8	8	9	9	9	
ANFIS-MPA	Value	0.9134	0.9187	1.7436	0.9764	0.0473	0.2943	0.0565	0.3355	59
	Score	8	6	7	7	7	8	8	8	

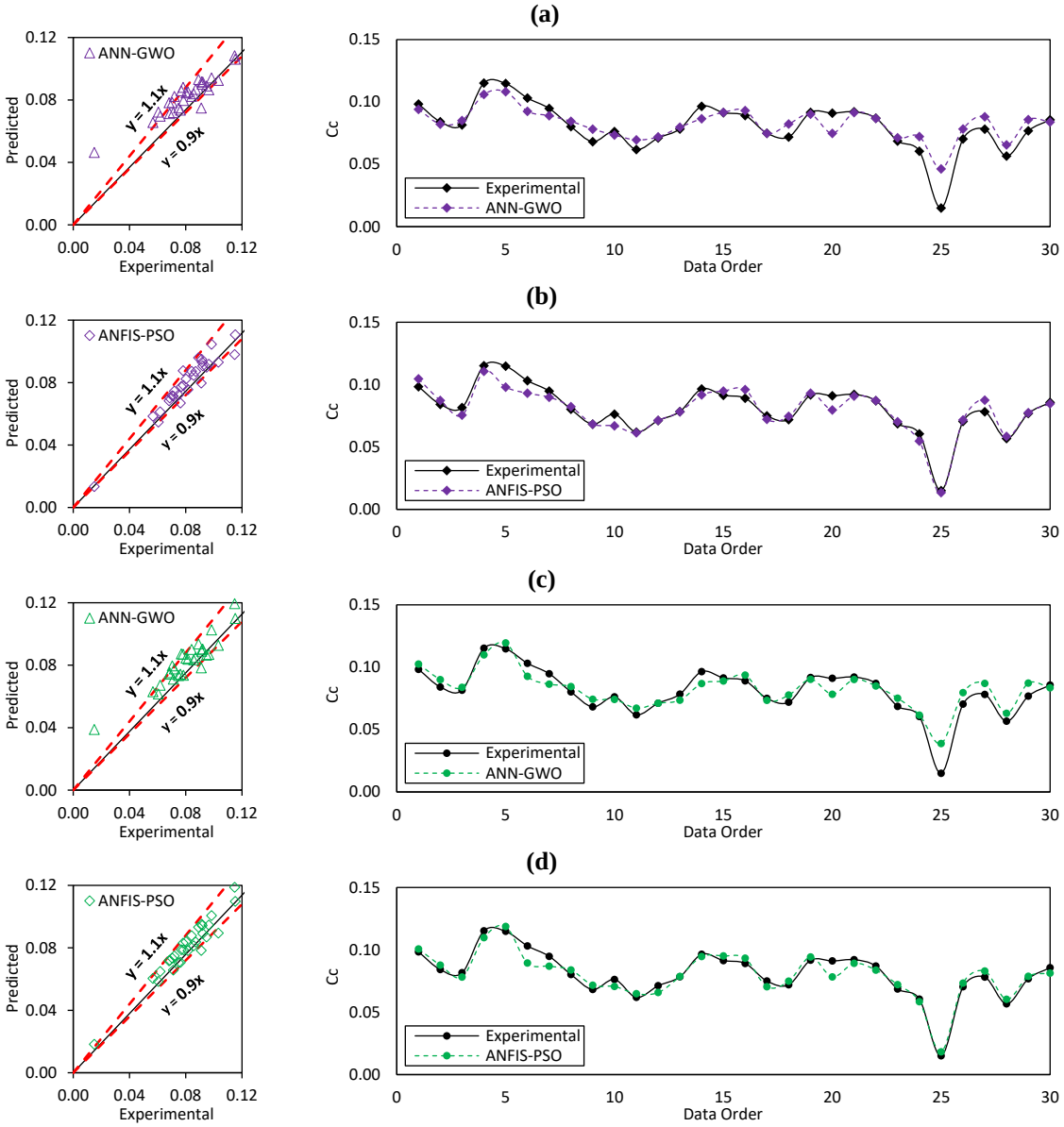

Fig. 15. Illustration of experimental vs. predicted results for external validation phase: (a-b) best two models for Case-1 and (c-b) best two models for Case-2.

8.4. Uncertainty analysis

This sub-section describes the quantitative assessment of the developed hybrid models in estimating the C_c of soils. Previously, Bardhan et al. [41] used the width of confidence bound (WCB) approach to estimate the uncertainty of prediction models. In this study, the quantitative evaluation of the developed models was evaluated via uncertainty analysis (UA). This assessment was carried out for all data in both the internal and external validation phases, which includes 140 and 30 experimental data, respectively. Thus, these data can be considered reliable to compare the trustworthiness of the developed models. To perform UA, the MAE and the standard deviation of absolute error (SDAE) of the predictions were computed initially. The standard error (SE), lower and upper bounds (LB and UB), and the margin of error (MOE) were evaluated at a 95% confidence interval to determine WCB. Note that the WCB value reflects an error range in which approximately 95% of the entire data is located.

The results of the UA are given in Table 21 (internal validation Case-1), Table 22 (internal validation Case-2), Table 23 (external validation Case-1), and Table 24 (external validation Case-1) with the details of several indices, including MAE, SDAE, SE, MOE, LB, UB, and WCB values. The index values obtained for UA can be used to assess the performance of the hybrid ANN and ANFIS models. Specifically, the stronger the model certainty, the lower the WCB value in these indices, indicating that a model with less error can predict the intended output with greater confidence. According to the UA results, ANN-GWO and ANFIS-PSO achieved the lowest WCB in both cases of C_c prediction among the developed hybrid ANN and ANFIS models, respectively. In the internal validation phase, the WCB values for these two models were 0.0102 and 0.0092 in Case-1 and 0.0082 and 0.0072 in Case-2 of C_c prediction, respectively. On the other hand, in the external validation phase, the WCB values for ANN-GWO and ANFIS-PSO were 0.0478 and 0.0304 in Case-1 and 0.0354 and 0.0212 in Case-2 of C_c prediction, respectively. These results clearly demonstrate that the developed ANN-GWO and ANFIS-PSO are very reliable in predicting soil C_c, which can also be observed in the details of other performance parameters presented above.

All of the developed models were ranked in groups based on the WCB value. Subsequently, all of the hybrid ANN and ANFIS models were ranked together to see which one performed best in Case-1 and Case-2 of soil C_c prediction. As can be observed, the ANFIS-PSO is the most precise predictive model in both phases. Furthermore, compared to other techniques employed in the study, the lower values of MAE and SDAE display the more reliable ANN-PSO model. Moreover, in the internal validation phase, the ANFIS-ABC for Case-1 and ANN-ABC for Case-2 display the maximum uncertainty, while in the external validation phase, the ANN-ABC for Case-1 and ANFIS-ABC for Case-2 demonstrate the maximum uncertainty. However, the models were ranked according to the MAE value when the WCB value was the same.

Table 21. Results of UA (internal validation phase for Case-1).

Models	MAE	SDAE	SE	MOE	LB	UB	WCB	Sub-rank	Overall rank
ANN-PSO	0.0609	0.0597	0.0050	0.0100	0.0509	0.0709	0.0200	9	18
ANN-ACO	0.0387	0.0332	0.0028	0.0055	0.0332	0.0442	0.0110	4	9
ANN-ABC	0.0822	0.0693	0.0059	0.0116	0.0706	0.0938	0.0232	10	19
ANN-GWO	0.0324	0.0308	0.0026	0.0051	0.0273	0.0375	0.0102	1	3
ANN-MFO	0.0484	0.0429	0.0036	0.0072	0.0412	0.0556	0.0144	7	16
ANN-WOA	0.0386	0.0356	0.0030	0.0059	0.0327	0.0445	0.0118	6	14
ANN-SSA	0.0371	0.0347	0.0029	0.0058	0.0313	0.0429	0.0116	5	12
ANN-HHO	0.0641	0.0484	0.0041	0.0081	0.0560	0.0722	0.0162	8	17
ANN-SMA	0.0405	0.0322	0.0027	0.0054	0.0351	0.0459	0.0108	3	6
ANN-MPA	0.0345	0.0316	0.0027	0.0053	0.0292	0.0398	0.0106	2	4
ANFIS-PSO	0.0265	0.0275	0.0023	0.0046	0.0219	0.0311	0.0092	1	1
ANFIS-ACO	0.0313	0.0353	0.0030	0.0059	0.0254	0.0372	0.0118	8	13
ANFIS-ABC	0.4663	0.2273	0.0192	0.0380	0.4283	0.5043	0.0760	10	20
ANFIS-GWO	0.0403	0.0356	0.0030	0.0059	0.0344	0.0462	0.0118	9	15
ANFIS-MFO	0.0359	0.0327	0.0028	0.0055	0.0304	0.0414	0.0110	5	8
ANFIS-WOA	0.0346	0.0330	0.0028	0.0055	0.0291	0.0401	0.0110	4	7
ANFIS-SSA	0.0386	0.0321	0.0027	0.0054	0.0332	0.0440	0.0108	3	5
ANFIS-HHO	0.0369	0.0341	0.0029	0.0057	0.0312	0.0426	0.0114	7	11
ANFIS-SMA	0.0389	0.0335	0.0028	0.0056	0.0333	0.0445	0.0112	6	10
ANFIS-MPA	0.0325	0.0283	0.0024	0.0047	0.0278	0.0372	0.0094	2	2

Table 22. Results of UA (internal validation phase for Case-2).

Models	MAE	SDAE	SE	MOE	LB	UB	WCB	Sub-rank	Overall rank
ANN-PSO	0.0466	0.0405	0.0034	0.0068	0.0398	0.0534	0.0136	8	17
ANN-ACO	0.0312	0.0296	0.0025	0.0049	0.0263	0.0361	0.0098	3	11
ANN-ABC	0.1006	0.0795	0.0067	0.0133	0.0873	0.1139	0.0266	10	20
ANN-GWO	0.0276	0.0244	0.0021	0.0041	0.0235	0.0317	0.0082	1	8
ANN-MFO	0.0425	0.0366	0.0031	0.0061	0.0364	0.0486	0.0122	6	15
ANN-WOA	0.0430	0.0386	0.0033	0.0065	0.0365	0.0495	0.0130	7	16
ANN-SSA	0.0360	0.0304	0.0026	0.0051	0.0309	0.0411	0.0102	5	13
ANN-HHO	0.0642	0.0566	0.0048	0.0095	0.0547	0.0737	0.0190	9	19
ANN-SMA	0.0330	0.0297	0.0025	0.0050	0.0280	0.0380	0.0100	4	12
ANN-MPA	0.0305	0.0260	0.0022	0.0043	0.0262	0.0348	0.0086	2	10
ANFIS-PSO	0.0212	0.0216	0.0018	0.0036	0.0176	0.0248	0.0072	1	1
ANFIS-ACO	0.0347	0.0309	0.0026	0.0052	0.0295	0.0399	0.0104	9	14
ANFIS-ABC	0.0639	0.0488	0.0041	0.0082	0.0557	0.0721	0.0164	10	18
ANFIS-GWO	0.0284	0.0238	0.0020	0.0040	0.0244	0.0324	0.0080	7	7
ANFIS-MFO	0.0280	0.0234	0.0020	0.0039	0.0241	0.0319	0.0078	4	4
ANFIS-WOA	0.0288	0.0245	0.0021	0.0041	0.0247	0.0329	0.0082	8	9
ANFIS-SSA	0.0280	0.0226	0.0019	0.0038	0.0242	0.0318	0.0076	3	3
ANFIS-HHO	0.0275	0.0237	0.0020	0.0040	0.0235	0.0315	0.0080	5	5
ANFIS-SMA	0.0283	0.0238	0.0020	0.0040	0.0243	0.0323	0.0080	6	6
ANFIS-MPA	0.0273	0.0225	0.0019	0.0038	0.0235	0.0311	0.0076	2	2

Table 23. Results of UA (external validation phase for Case-1).

Models	MAE	SDAE	SE	MOE	LB	UB	WCB	Sub-rank	Overall rank
ANN-PSO	0.0818	0.1010	0.0184	0.0377	0.0441	0.1195	0.0754	9	17
ANN-ACO	0.0814	0.0788	0.0144	0.0294	0.0520	0.1108	0.0588	5	13
ANN-ABC	0.2400	0.1539	0.0281	0.0575	0.1825	0.2975	0.1150	10	20
ANN-GWO	0.0666	0.0639	0.0117	0.0239	0.0427	0.0905	0.0478	1	9
ANN-MFO	0.1378	0.0959	0.0175	0.0358	0.1020	0.1736	0.0716	8	16
ANN-WOA	0.1022	0.0733	0.0134	0.0274	0.0748	0.1296	0.0548	3	11
ANN-SSA	0.1039	0.0889	0.0162	0.0332	0.0707	0.1371	0.0664	6	14
ANN-HHO	0.1156	0.0777	0.0142	0.0290	0.0866	0.1446	0.0580	4	12
ANN-SMA	0.0848	0.0644	0.0118	0.0240	0.0608	0.1088	0.0480	2	10
ANN-MPA	0.0826	0.0920	0.0168	0.0344	0.0482	0.1170	0.0688	7	15
ANFIS-PSO	0.0417	0.0408	0.0074	0.0152	0.0265	0.0569	0.0304	1	1
ANFIS-ACO	0.1492	0.0806	0.0147	0.0301	0.1191	0.1793	0.0602	9	18
ANFIS-ABC	0.0917	0.0845	0.0154	0.0316	0.0601	0.1233	0.0632	10	19
ANFIS-GWO	0.0752	0.0522	0.0095	0.0195	0.0557	0.0947	0.0390	7	7
ANFIS-MFO	0.0790	0.0536	0.0098	0.0200	0.0590	0.0990	0.0400	8	8
ANFIS-WOA	0.0780	0.0510	0.0093	0.0190	0.0590	0.0970	0.0380	6	6
ANFIS-SSA	0.0505	0.0501	0.0091	0.0187	0.0318	0.0692	0.0374	5	5
ANFIS-HHO	0.0699	0.0409	0.0075	0.0153	0.0546	0.0852	0.0306	2	2
ANFIS-SMA	0.0661	0.0423	0.0077	0.0158	0.0503	0.0819	0.0316	3	3
ANFIS-MPA	0.0664	0.0422	0.0077	0.0158	0.0506	0.0822	0.0316	4	4

Table 24. Results of UA (external validation phase for Case-2).

Models	MAE	SDAE	SE	MOE	LB	UB	WCB	Sub-rank	Overall rank
ANN-PSO	0.1003	0.0801	0.0146	0.0299	0.0704	0.1302	0.0598	8	17
ANN-ACO	0.0745	0.0759	0.0139	0.0283	0.0462	0.1028	0.0566	6	15
ANN-ABC	0.1615	0.1102	0.0201	0.0411	0.1204	0.2026	0.0822	10	19
ANN-GWO	0.0594	0.0473	0.0086	0.0177	0.0417	0.0771	0.0354	1	9
ANN-MFO	0.0928	0.0674	0.0123	0.0252	0.0676	0.1180	0.0504	5	13
ANN-WOA	0.0729	0.0513	0.0094	0.0192	0.0537	0.0921	0.0384	2	10
ANN-SSA	0.0964	0.0771	0.0141	0.0288	0.0676	0.1252	0.0576	7	16
ANN-HHO	0.0890	0.0946	0.0173	0.0353	0.0537	0.1243	0.0706	9	18
ANN-SMA	0.0793	0.0608	0.0111	0.0227	0.0566	0.1020	0.0454	3	11
ANN-MPA	0.0750	0.0645	0.0118	0.0241	0.0509	0.0991	0.0482	4	12
ANFIS-PSO	0.0433	0.0285	0.0052	0.0106	0.0327	0.0539	0.0212	1	1
ANFIS-ACO	0.1262	0.0716	0.0131	0.0267	0.0995	0.1529	0.0534	9	14
ANFIS-ABC	0.0866	0.0885	0.0162	0.0330	0.0536	0.1196	0.0660	10	20
ANFIS-GWO	0.0690	0.0455	0.0083	0.0170	0.0520	0.0860	0.0340	8	8
ANFIS-MFO	0.0543	0.0346	0.0063	0.0129	0.0414	0.0672	0.0258	6	6
ANFIS-WOA	0.0603	0.0362	0.0066	0.0135	0.0468	0.0738	0.0270	7	7
ANFIS-SSA	0.0465	0.0323	0.0059	0.0121	0.0344	0.0586	0.0242	5	5
ANFIS-HHO	0.0524	0.0319	0.0058	0.0119	0.0405	0.0643	0.0238	4	4
ANFIS-SMA	0.0472	0.0309	0.0056	0.0115	0.0357	0.0587	0.0230	2	2
ANFIS-MPA	0.0473	0.0309	0.0056	0.0115	0.0358	0.0588	0.0230	3	3

8.5. Monotonicity analysis

This study also investigated the feasibility of different soil particles through monotonicity analysis. Herein, the best performing model, i.e., ANFIS-PSO, for predicting the C_c of soils is analysed. The purpose of this analysis is to see how well the proposed hybrid model handles the overfitting problem using the concept of feasibility modelling of input parameters. For this purpose, a simulated dataset was produced for seven input parameters, namely S, M&C, C, M, FS, MS, and CS, with one input parameter changing linearly while the others stayed fixed. However, the principle of grain size analysis was followed, which states that the total proportions of S, M&C, and G contents or C, M, FS, MS, CS, and G contents cannot exceed 100%. Therefore, to assure the trend between the input parameter and soil C_c, a linear increase in one parameter was countered by a linear decrease in another, and vice-versa. To generate the simulated dataset, S and M&C, C and M, FS and MS, and CS and FS were altered as per the details given in Table 25, and the dataset was generated accordingly. In each scenario, the variable range of input parameters is also indicated in Table 25. As stated above, the summation of G, S and M&C; M, FS, MS, CS, and G were kept exactly 100% in all circumstances to imitate the practical case. It is also worth noting that the values of the other input parameters, as indicated in Table 25, were preserved at their mean values.

Table 25. Dataset generation for monotonicity analysis.

Variable input parameter				Input parameters with contact mean value	Fig. reference
Param.	**Range**	**Increased by**	**Set**		
S	20.00 – 34.00	+1.00	A	D = 7.48, G = 3.42, BD = 1.80, DD = 1.59, Gs = 2.67, LL = 32.86, PL = 18.84.	Fig. 16(a)
M&C	69.10 – 55.10	-1.00			Fig. 16(b)
S	30.00 – 44.00	+1.00	B	D = 7.48, G = 3.42, BD = 1.80, DD = 1.59, Gs = 2.67, LL = 32.86, PL = 18.84.	Fig. 16(c)
M&C	59.10 – 45.10	-1.00			Fig. 16(d)
C	15.00 – 29.00	+1.00	C	D = 7.48, G = 3.42, CS = 2.15, MS = 3.90, FS = 26.40, BD = 1.80, DD = 1.59, Gs = 2.67, LL = 32.86, PL = 18.84.	Fig. 16(e)
M	49.13 – 35.13	-1.00			Fig. 16(f)
C	34.13 – 20.13	-1.00	D	D = 7.48, G = 3.42, CS = 2.15, MS = 3.90, FS = 26.40, BD = 1.80, DD = 1.59, Gs = 2.67, LL = 32.86, PL = 18.84.	Fig. 16(g)
M	30.00 – 44.00	+1.00			Fig. 16(h)
FS	10.00 – 24.00	+1.00	E	D = 7.48, G = 3.42, CS = 2.15, M = 48.01, C = 16.11, BD = 1.80, DD = 1.59, Gs = 2.67, LL = 32.86, PL = 18.84.	Fig. 16(i)
MS	6.31 – 20.31	-1.00			Fig. 16(j)
CS	3.00 – 17.00	+1.00	F	D = 7.48, G = 3.42, MS = 3.90, M = 48.01, C = 16.11, BD = 1.80, DD = 1.59, Gs = 2.67, LL = 32.86, PL = 18.84.	Fig. 16(k)
FS	25.56 – 11.56	-1.00			Fig. 16(l)

Fig. 16 displays all of the trends with smooth curves and reveals that soil C_c decreased as the content of S increases, while the completely opposite occurred for M&C content in both cases (Fig. 16(a - d)). In addition, Fig. 16(e, g) shows the expected increase in C_c with increasing C content, whereas, Fig. 16(f, h) illustrates the opposite trend, i.e., higher C_c with decreasing M content. The estimated C_c increased as the FS content grew, but declined as the content of MS increased (see Fig. 16(i,j)). Similarly, the content of CS and FS exhibited (see Fig. 16(k, l)) opposite trends as expected. It is observed that the estimated C_c increases with the increase of C or M&C contents, whereas the predicted value decreases as the sand content increases. This occurrence implies that soil C_c is influenced more by M and C contents. The impact of different-sized sand particles can also be observed. It is worth noting that the ANFIS-PSO model was used to ensure expected trends using simulated datasets, but real-life analysis may provide different results.

It is also significant to mention that, in the majority of cases, the sub-soils of railways, highways, buildings, bridge foundations, etc. require preliminary treatment to improve their bearing capacity. In general, practitioners use soil stabilisation, soil mixing, and soil compaction to increase the quality of the subgrade material. Thus, the notion of the monotonicity analysis described here will allow engineers to change the soil qualities by modifying the mix of different-sized particles based on construction site requirements. Furthermore, the impact of different particle sizes on soil C_c at various soil particle proportions can be assessed.

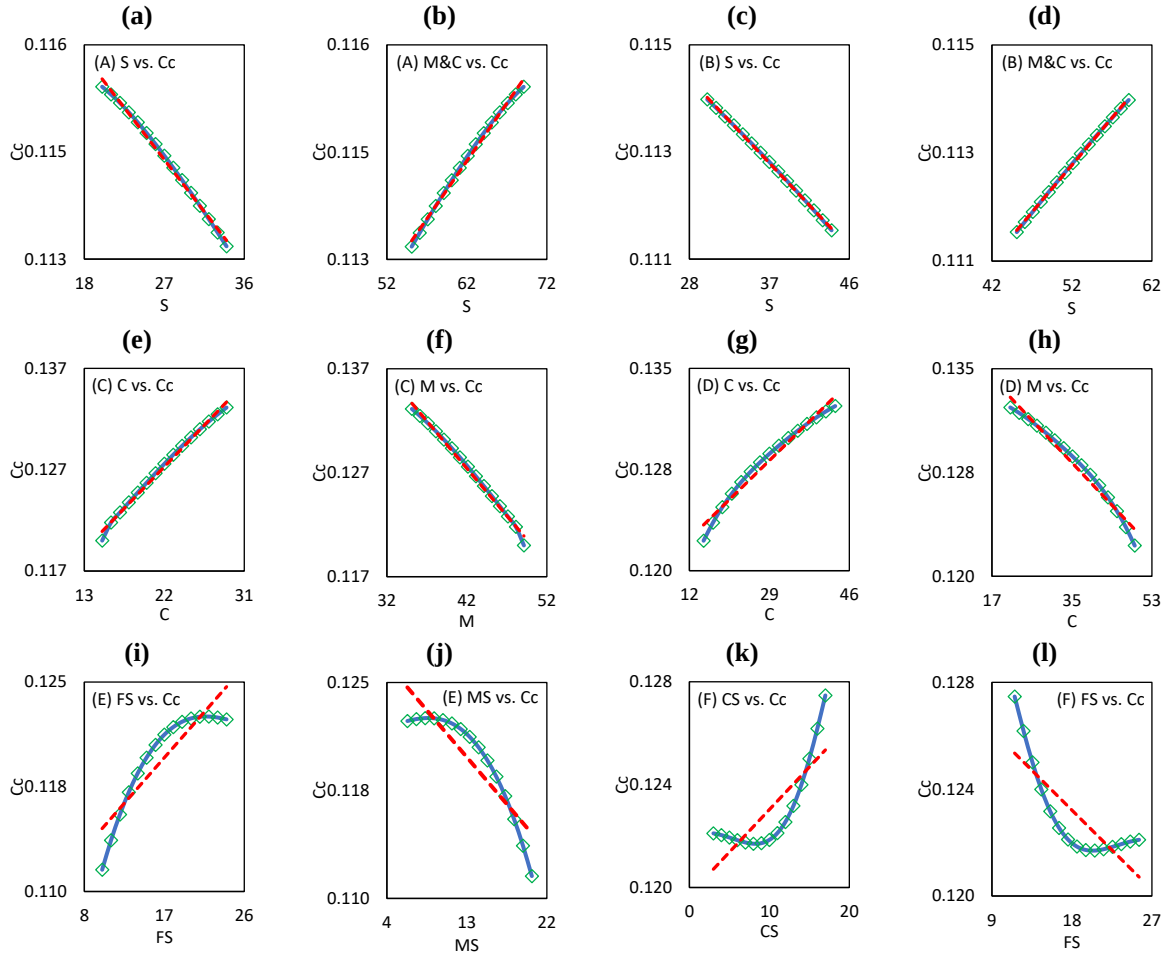

Fig. 16. Illustration of monotonicity analysis using ANFIS-PSO model.

8.6. Discussion

The above sub-sections describe the outcomes and generalization capabilities of the hybrid ANN and ANFIS models in predicting soil C_c. After the models were developed, multiple performance indices were determined and analysed. In this study, two different combinations of input parameters i.e., Case-1 and Case-2, were considered to investigate the mapping relationship between soil parameters and C_c. All the hybrid ANN and ANFIS models were trained for both input combinations and validated thoroughly. The obtained results suggest that all developed models can capture the relationship between soil parameters and C_c in both training and validation phases; however, the prediction performance of the proposed ANN-GWO and ANFIS-PSO models was found to be significantly better than that of other developed models. In addition, separate laboratory experiments were conducted to determine the capability of the developed hybrid models for future application. It is also seen that the proposed ANN-GWO and ANFIS-PSO are the most accurate models in predicting soil C_c

The overall status of the best performing model with the highest prediction accuracy in both Case-1 and Case-2 input combinations is presented in Table 26. For this purpose, the RMSE and WCB criteria were examined. As can be seen, ANFIS-PSO performed better than ANN-GWO in every phase. Compared to the Case-1 input combination, the proposed ANN-GWO and ANFIS-PSO models constructed with the second input combination (i.e., Case-2) achieved a greater prediction accuracy in both the internal and external validation phases. These findings demonstrate that both the models attained better performance when the contents of CS, FS, MS, M, and

C were considered separately in predicting the C_c of soils. In the internal validation phase, ANFIS-PSO achieved a very high accuracy for inferring values of soil C_c, which is expressed by the relatively low values of RMSE (0.0382 in Case-1 and 0.0303 in Case-2) and WCB (0.0092 in Case-1 and 0.0072 in Case-2). Comparatively, in the external validation phase, RMSE values were 0.0583 in Case-1 and 0.0519 in Case-2, and WCM values were 0.0304 in Case-1 and 0.0212 in Case-2. The developed ANN-GWO also attained a higher prediction accuracy in Case-2 of soil C_c prediction (refer Table 26 for more details). However, the overall performance of the constructed ANFIS-PSO supports its superior generalization and robustness at all levels.

Table 26. Summarized performance of the best performing model.

RMSE and WCB values in different phases	Case-1 best	Case-2 best	Final best
RMSE (internal validation)	ANN-GWO = 0.0447 ANFIS-PSO = 0.0382	**ANN-GWO = 0.0368** **ANFIS-PSO = 0.0303**	ANFIS-PSO
RMSE (external validation)	ANN-GWO = 0.0923 ANFIS-PSO = 0.0583	**ANN-GWO = 0.0760** **ANFIS-PSO = 0.0519**	ANFIS-PSO
WCB value (internal validation)	ANN-GWO = 0.0102 ANFIS-PSO = 0.0092	**ANN-GWO = 0.0082** **ANFIS-PSO = 0.0072**	ANFIS-PSO
WCB value (external validation)	ANN-GWO = 0.0478 ANFIS-PSO = 0.0304	**ANN-GWO = 0.0354** **ANFIS-PSO = 0.0212**	ANFIS-PSO

Furthermore, the validity of the proposed ANFIS-PSO model and sensitivity of the input variables on the predicted results, especially the impact of CS, MS, FS, M, and C contents, were analysed. The monotonicity analysis demonstrates the overall behaviour of the proposed ANFIS-PSO model with regards to its expected behaviour in terms of soil C_c and other soil parameters, indicating that overfitting did not occur in either of the two input combinations for predicting soil C_c.

9. Conclusion

In the present work, a comparative assessment of hybrid ANN and ANFIS models constructed with ten SI algorithms for estimating soil C_c has been presented. To this end, a sum of 700 oedometer test results was employed to create and validate the models, and 30 new laboratory experiments were conducted for external validation of the proposed hybrid models. Based on the experimental outcomes, the ANN-GWO and ANFIS-PSO were found to be the best hybrid models in the groups of hybrid ANN and ANFIS models. However, experimental results exhibit that the proposed ANFIS-PSO obtained the most accurate prediction across all stages (i.e., training, internal validation, and external validation phases) among the developed hybrid models. Based on the R^2 and RMSE criteria, the proposed ANFIS-PSO outperformed the other hybrid models with $R^2 = 0.9789$ and RMSE = 0.0303 in the internal validation phase and $R^2 = 0.9286$ and RMSE = 0.0519 in the external validation phase. The validation of the ANFIS-PSO model with a completely new dataset demonstrates its robustness and generalizability in estimating soil C_c. The performance obtained in the external validation phase indicates that the proposed ANFIS-PSO can be used to estimate soil C_c in the future. This is one of the major contributions of this study.

On the contrary, based on grain size distribution analyses, two different combinations of basic soil parameters (i.e., Case-1 and Case-2) were explored in this study to ensure the effectiveness of silt and clay particles on soil C_c. Experimental outcomes of the two best models, i.e., ANFIS-PSO and ANN-GWO, exhibit that the Case-2 input combination offers better prediction performance than the Case-1 input combination for C_c prediction. Therefore, the use of grain size distribution with independent contents of CS, MS, FS, M, and C appears to be highly effective for modelling the C_c of soils The impact of different-sized particles on soil C_c was also

investigated via a monotonicity analysis considering different combinations of input parameters. This will allow researchers and practitioners to customize the soil $C_{c.}$ by adjusting the mix proportions of S, CS, MS, FS, M&C, M, and C to meet their specific needs.

The present work proposes a high-performance hybrid model to replace the oedometer tests and provides a comprehensive literature review of previous studies, including the results of the utilized ML models. Details of the input soil parameters used in earlier studies to estimate the C_c of soils have also been included, which will benefit researchers in their future studies. The main advantages of the proposed ANFIS-PSO model include: a) faster convergence, b) lower computational time, c) use of real-life datasets collected from a large project of IR, d) higher prediction accuracy in the validation phases, and e) high degree of reliability. With fewer FIS parameters and a reduced computational cost, the ANFIS-PSO model achieved significant prediction accuracy. highlighting the effectiveness of the said model. This is another major advantage of the proposed hybrid approach. However, the confined search space of the OAs (i.e., fixed values of different deterministic parameters) and required multiple runs can be seen as limitations of this study. Therefore, more research should be conducted to expand the use of hybrid ANN and ANFIS models for forecasting the desired output in different engineering disciplines.

Nonetheless, the future direction of this work may include: a) a detailed assessment of the accuracy of other hybrid models, such as hybrid ELMs, hybrid SVMs, etc. via actual data from various areas of geotechnical engineering; b) evaluation of the ANFIS-PSO model's superiority over other hybrid ANFIS models built using evolutionary and physics-based OAs; c) implementation of advanced and enhanced of meta-heuristic algorithms for constructing hybrid ANFIS models; d) soil sample preparation based on the simulated dataset and validation of results through laboratory experiments; e) large-scale investigation and validation of the developed hybrid models; and e) study of the impact of soil particles on C_c, especially silt and clay contents for other types of soils. Nevertheless, considering the effect of different-sized soil particles, the current research proposes a novel promising strategy for estimating soil C_c that eliminates the undesired oedometer tests for the first time.

Reference

[1] P.K. Kolay, A.B. Rosmina, Y. Shirley, Prediction of compression index for tropical soil by using Artificial Neural Network (ANN), in: 13th Int. Conf. IACMAG, 2011: pp. 542–547.

[2] Y. Huang, C. Chen, D. Su, S. Wu, Comparison of leading-industrialisation and crossing-industrialisation economic growth patterns in the context of sustainable development: Lessons from China and India, Sustain. Dev. 28 (2020) 1077–1085.

[3] M. Mohammed, A. Sharafati, N. Al-Ansari, Z.M. Yaseen, Shallow foundation settlement quantification: application of hybridized adaptive neuro-Fuzzy inference system model, Adv. Civ. Eng. 2020 (2020).

[4] A.S. Osman, M.D. Bolton, A new approach to the estimation of undrained settlement of shallow foundations on soft clay, Eng. Pract. Perform. Soft Depos. IS-OSAKA. (2004) 93–98.

[5] H. Shahir, A. Pak, Estimating liquefaction-induced settlement of shallow foundations by numerical approach, Comput. Geotech. 37 (2010) 267–279.

[6] R.W. Day, Foundation engineering handbook: design and construction with the 2009 international building code, McGraw-Hill Education, 2010.

[7] D. Mohammadzadeh S., J. Bolouri Bazaz, A.H. Alavi, An evolutionary computational approach for

formulation of compression index of fine-grained soils, Eng. Appl. Artif. Intell. 33 (2014) 58–68. https://doi.org/10.1016/j.engappai.2014.03.012.

[8] J.D. Nelson, D.K. Reichler, J.M. Cumbers, Parameters for heave prediction by oedometer tests, in: Unsaturated Soils 2006, 2006: pp. 951–961.

[9] D. Tien Bui, V.H. Nhu, N.D. Hoang, Prediction of soil compression coefficient for urban housing project using novel integration machine learning approach of swarm intelligence and Multi-layer Perceptron Neural Network, Adv. Eng. Informatics. 38 (2018) 593–604. https://doi.org/10.1016/j.aei.2018.09.005.

[10] T.S. Nagaraj, B.R.S. Murthy, Prediction of the preconsolidation pressure and recompression index of soils, Geotech. Test. J. 8 (1985) 199–202.

[11] Y. Nishida, A brief note on compression index of soil, J. Soil Mech. Found. Div. 82 (1956) 1–14.

[12] S.D. Koppula, Statistical estimation of compression index, Geotech. Test. J. 4 (1981) 68–73.

[13] K. Terzaghi, R.B. Peck, G. Mesri, Soil Mechanics in Engineering Practice, John Wiley & Sons, Inc., New York. (1967).

[14] A.W. Skempton, O.T. Jones, Notes on the compressibility of clays, Q. J. Geol. Soc. 100 (1944) 119–135.

[15] G. Kumar, S. Jain, U.P. Singh, Stock Market Forecasting Using Computational Intelligence: A Survey, Springer Netherlands, 2021. https://doi.org/10.1007/s11831-020-09413-5.

[16] P. Zhang, Z.Y. Yin, Y.F. Jin, State-of-the-Art Review of Machine Learning Applications in Constitutive Modeling of Soils, Arch. Comput. Methods Eng. 28 (2021) 3661–3686. https://doi.org/10.1007/s11831-020-09524-z.

[17] P.G. Asteris, F.I.M. Rizal, M. Koopialipoor, P.C. Roussis, M. Ferentinou, D.J. Armaghani, B. Gordan, Slope Stability Classification under Seismic Conditions Using Several Tree-Based Intelligent Techniques, Appl. Sci. 12 (2022) 1753.

[18] K.C. Onyelowe, F.E. Jalal, M. Iqbal, Z.U. Rehman, K. Ibe, Intelligent modeling of unconfined compressive strength (UCS) of hybrid cement-modified unsaturated soil with nanostructured quarry fines inclusion, Innov. Infrastruct. Solut. 7 (2022) 1–18.

[19] U.M.R. Paturi, S. Cheruku, N.S. Reddy, The Role of Artificial Neural Networks in Prediction of Mechanical and Tribological Properties of Composites—A Comprehensive Review, Springer Netherlands, 2022. https://doi.org/10.1007/s11831-021-09691-7.

[20] M. el A. Bourouis, A. Zadjaoui, A. Djedid, Contribution of two artificial intelligence techniques in predicting the secondary compression index of fine-grained soils, Innov. Infrastruct. Solut. 5 (2020) 1–11. https://doi.org/10.1007/s41062-020-00348-1.

[21] B.T. Pham, M.D. Nguyen, H.-B. Ly, T.A. Pham, V. Hoang, H. Van Le, T.-T. Le, H.Q. Nguyen, G.L. Bui, Development of artificial neural networks for prediction of compression coefficient of soft soil, in: CIGOS 2019, Innov. Sustain. Infrastruct., Springer, 2020: pp. 1167–1172.

[22] W. Zhang, Y. Zhang, A.T.C. Goh, Multivariate adaptive regression splines for inverse analysis of soil and wall properties in braced excavation, Tunn. Undergr. Sp. Technol. 64 (2017) 24–33.

[23] W. Zhang, R. Zhang, A.T.C. Goh, Multivariate adaptive regression splines approach to estimate lateral wall deflection profiles caused by braced excavations in clays, Geotech. Geol. Eng. 36 (2018) 1349–1363.

[24] W. Zhang, R. Zhang, C. Wu, A.T.C. Goh, S. Lacasse, Z. Liu, H. Liu, State-of-the-art review of soft computing applications in underground excavations, Geosci. Front. 11 (2020) 1095–1106.

[25] L. Wang, W. Zhang, F. Chen, Bayesian approach for predicting soil-water characteristic curve from particle-size distribution data, Energies. 12 (2019) 2992.
[26] S. Emamgholizadeh, B. Mohammadi, New hybrid nature-based algorithm to integration support vector machine for prediction of soil cation exchange capacity, Soft Comput. 25 (2021) 13451–13464.
[27] M. Emadi, R. Taghizadeh-Mehrjardi, A. Cherati, M. Danesh, A. Mosavi, T. Scholten, Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran, Remote Sens. 12 (2020) 2234.
[28] A. Mosavi, F. Sajedi-Hosseini, B. Choubin, F. Taromideh, G. Rahi, A.A. Dineva, Susceptibility mapping of soil water erosion using machine learning models, Water. 12 (2020) 1995.
[29] H. Il Park, S.R. Lee, Evaluation of the compression index of soils using an artificial neural network, Comput. Geotech. 38 (2011) 472–481.
[30] X.C. Shi, Y.F. Guo, Application of genetic arithmetic and support vector machine in prediction of compression index of clay, Appl. Mech. Mater. 438–439 (2013) 1167–1170. https://doi.org/10.4028/www.scientific.net/AMM.438-439.1167.
[31] S.M. Kashefipour, M. Daryaee, Modeling the compression index for fine soils using an intelligent method, J. Biodivers. Environ. Sci. 5 (2014) 197–204.
[32] T.F. Kurnaz, U. Dagdeviren, M. Yildiz, O. Ozkan, Prediction of compressibility parameters of the soils using artificial neural network, Springerplus. 5 (2016). https://doi.org/10.1186/s40064-016-3494-5.
[33] D. Mohammadzadeh S, J. Bolouri Bazaz, S.H. Vafaee Jani Yazd, A.H. Alavi, Deriving an intelligent model for soil compression index utilizing multi-gene genetic programming, Environ. Earth Sci. 75 (2016) 1–11. https://doi.org/10.1007/s12665-015-4889-2.
[34] T. Fikret Kurnaz, Y. Kaya, The comparison of the performance of ELM, BRNN, and SVM methods for the prediction of compression index of clays, Arab. J. Geosci. 11 (2018). https://doi.org/10.1007/s12517-018-4143-9.
[35] B.T. Pham, M.D. Nguyen, D. Van Dao, I. Prakash, H.B. Ly, T.T. Le, L.S. Ho, K.T. Nguyen, T.Q. Ngo, V. Hoang, L.H. Son, H.T.T. Ngo, H.T. Tran, N.M. Do, H. Van Le, H.L. Ho, D. Tien Bui, Development of artificial intelligence models for the prediction of Compression Coefficient of soil: An application of Monte Carlo sensitivity analysis, Sci. Total Environ. 679 (2019) 172–184. https://doi.org/10.1016/j.scitotenv.2019.05.061.
[36] R. Taghizadeh-Mehrjardi, K. Schmidt, N. Toomanian, B. Heung, T. Behrens, A. Mosavi, S.S. Band, A. Amirian-Chakan, A. Fathabadi, T. Scholten, Improving the spatial prediction of soil salinity in arid regions using wavelet transformation and support vector regression models, Geoderma. 383 (2021) 114793.
[37] A. Alizadeh Majdi, R. Dabiri, N. Ganjian, A. Ghalandarzadeh, Determination of the Soil Compression Index (Cc) in Clayey Soils Using Shear Wave Velocity (Case Study: Tabriz City), Iran. J. Sci. Technol. - Trans. Civ. Eng. 43 (2019) 577–588. https://doi.org/10.1007/s40996-018-0209-x.
[38] S. Alam, S. Khuntia, C. Patra, Prediction of compression index of clay using artificial neural network, in: Int. Conf. Ind. Eng. Sci. Appl. Durgapur, 2014.
[39] M.A. Benbouras, R. Kettab Mitiche, H. Zedira, A.I. Petrisor, N. Mezouar, F. Debiche, A new approach to predict the compression index using artificial intelligence methods, Mar. Georesources Geotechnol. 37

(2019) 704–720. https://doi.org/10.1080/1064119X.2018.1484533.
[40] S. Danial Mohammadzadeh, S.F. Kazemi, A. Mosavi, E. Nasseralshariati, J.H.M. Tah, Prediction of compression index of fine-grained soils using a gene expression programming model, Infrastructures. 4 (2019) 1–12. https://doi.org/10.3390/infrastructures4020026.
[41] A. Bardhan, A. GuhaRay, S. Gupta, B. Pradhan, C. Gokceoglu, A novel integrated approach of ELM and modified equilibrium optimizer for predicting soil compression index of subgrade layer of Dedicated Freight Corridor, Transp. Geotech. 32 (2022) 100678.
[42] P. Samui, N.D. Hoang, V.H. Nhu, M.L. Nguyen, P.T.T. Ngo, D.T. Bui, A new approach of hybrid bee colony optimized neural computing to estimate the soil compression coefficient for a housing construction project, Appl. Sci. 9 (2019) 1–18. https://doi.org/10.3390/app9224912.
[43] B.T. Pham, L.H. Son, T.A. Hoang, D.M. Nguyen, D. Tien Bui, Prediction of shear strength of soft soil using machine learning methods, Catena. 166 (2018) 181–191. https://doi.org/10.1016/j.catena.2018.04.004.
[44] C. Xu, B. Gordan, M. Koopialipoor, D.J. Armaghani, M.M. Tahir, X. Zhang, Improving performance of retaining walls under dynamic conditions developing an optimized ANN based on ant colony optimization technique, IEEE Access. 7 (2019) 94692–94700.
[45] M. Koopialipoor, A. Fallah, D.J. Armaghani, A. Azizi, E.T. Mohamad, Three hybrid intelligent models in estimating flyrock distance resulting from blasting, Eng. Comput. 35 (2019) 243–256. https://doi.org/10.1007/s00366-018-0596-4.
[46] H.-B. Ly, B.T. Pham, L.M. Le, T.-T. Le, V.M. Le, P.G. Asteris, Estimation of axial load-carrying capacity of concrete-filled steel tubes using surrogate models, Neural Comput. Appl. 33 (2021) 3437–3458.
[47] E.M. Golafshani, A. Behnood, M. Arashpour, Predicting the compressive strength of normal and High-Performance Concretes using ANN and ANFIS hybridized with Grey Wolf Optimizer, Constr. Build. Mater. 232 (2020) 117266. https://doi.org/10.1016/j.conbuildmat.2019.117266.
[48] L.T. Le, H. Nguyen, J. Dou, J. Zhou, A comparative study of PSO-ANN, GA-ANN, ICA-ANN, and ABC-ANN in estimating the heating load of buildings' energy efficiency for smart city planning, Appl. Sci. 9 (2019). https://doi.org/10.3390/app9132630.
[49] V.K. Ojha, A. Abraham, V. Snášel, Metaheuristic design of feedforward neural networks: A review of two decades of research, Eng. Appl. Artif. Intell. 60 (2017) 97–116.
[50] R. Taghizadeh-Mehrjardi, M. Emadi, A. Cherati, B. Heung, A. Mosavi, T. Scholten, Bio-inspired hybridization of artificial neural networks: An application for mapping the spatial distribution of soil texture fractions, Remote Sens. 13 (2021) 1025.
[51] M.A. Ghorbani, R.C. Deo, M.H. Kashani, M. Shahabi, S. Ghorbani, Artificial intelligence-based fast and efficient hybrid approach for spatial modelling of soil electrical conductivity, Soil Tillage Res. 186 (2019) 152–164.
[52] A. Mosavi, S. Samadianfard, S. Darbandi, N. Nabipour, S.N. Qasem, E. Salwana, S.S. Band, Predicting soil electrical conductivity using multi-layer perceptron integrated with grey wolf optimizer, J. Geochemical Explor. 220 (2021) 106639.
[53] R. Moazenzadeh, B. Mohammadi, M.J.S. Safari, K. Chau, Soil moisture estimation using novel bio-inspired soft computing approaches, Eng. Appl. Comput. Fluid Mech. 16 (2022) 826–840.

[54] G. Raghurama, A. Vermaa, Dedicated Freight Corridor: Current Challenges, (2018).
[55] C. Batur Şahin, L. Abualigah, A novel deep learning-based feature selection model for improving the static analysis of vulnerability detection, Neural Comput. Appl. 33 (2021) 14049–14067. https://doi.org/10.1007/s00521-021-06047-x.
[56] M. Abd Elaziz, A. Dahou, L. Abualigah, L. Yu, M. Alshinwan, A.M. Khasawneh, S. Lu, Advanced metaheuristic optimization techniques in applications of deep neural networks: a review, Neural Comput. Appl. 33 (2021) 14079–14099. https://doi.org/10.1007/s00521-021-05960-5.
[57] D.E. Goldberg, J.H. Holland, Genetic algorithms and machine learning, (1988).
[58] D. Simon, Biogeography-based optimization, IEEE Trans. Evol. Comput. 12 (2008) 702–713.
[59] R. Storn, K. Price, Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces, J. Glob. Optim. 11 (1997) 341–359.
[60] X. Yao, Y. Liu, G. Lin, Evolutionary programming made faster, IEEE Trans. Evol. Comput. 3 (1999) 82–102.
[61] I. Rechenberg, Evolution strategy: Nature's way of optimization, in: Optim. Methods Appl. Possibilities Limitations, Springer, 1989: pp. 106–126.
[62] J.R. Koza, J.R. Koza, Genetic programming: on the programming of computers by means of natural selection, MIT press, 1992.
[63] E. Atashpaz-Gargari, C. Lucas, Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition, in: 2007 IEEE Congr. Evol. Comput., Ieee, 2007: pp. 4661–4667.
[64] N. Moosavian, B.K. Roodsari, Soccer league competition algorithm: A novel meta-heuristic algorithm for optimal design of water distribution networks, Swarm Evol. Comput. 17 (2014) 14–24.
[65] R.V. Rao, V.J. Savsani, D.P. Vakharia, Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems, Comput. Des. 43 (2011) 303–315.
[66] R. Moghdani, K. Salimifard, Volleyball premier league algorithm, Appl. Soft Comput. 64 (2018) 161–185.
[67] O.K. Erol, I. Eksin, A new optimization method: big bang–big crunch, Adv. Eng. Softw. 37 (2006) 106–111.
[68] R.A. Formato, Central force optimization, Prog Electromagn Res. 77 (2007) 425–491.
[69] E. Rashedi, H. Nezamabadi-Pour, S. Saryazdi, GSA: a gravitational search algorithm, Inf. Sci. (Ny). 179 (2009) 2232–2248.
[70] I. Ahmadianfar, O. Bozorg-Haddad, X. Chu, Gradient-based optimizer: A new metaheuristic optimization algorithm, Inf. Sci. (Ny). 540 (2020) 131–159.
[71] S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, Optimization by simulated annealing, Science (80-.). 220 (1983) 671–680.
[72] D. Karaboga, B. Basturk, A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm, J. Glob. Optim. 39 (2007) 459–471. https://doi.org/10.1007/s10898-007-9149-x.
[73] S. Mirjalili, The ant lion optimizer, Adv. Eng. Softw. 83 (2015) 80–98. https://doi.org/10.1016/j.advengsoft.2015.01.010.
[74] S. Mirjalili, S. Mohammad, A. Lewis, Advances in Engineering Software Grey Wolf Optimizer, Adv.

Eng. Softw. 69 (2014) 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007.
[75] A.A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, H. Chen, Harris hawks optimization: Algorithm and applications, Futur. Gener. Comput. Syst. 97 (2019) 849–872. https://doi.org/10.1016/j.future.2019.02.028.
[76] S. Mirjalili, Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm, Knowledge-Based Syst. 89 (2015) 228–249. https://doi.org/10.1016/j.knosys.2015.07.006.
[77] A. Faramarzi, M. Heidarinejad, S. Mirjalili, A.H. Gandomi, Marine Predators Algorithm: A nature-inspired metaheuristic, Expert Syst. Appl. 152 (2020) 113377. https://doi.org/10.1016/j.eswa.2020.113377.
[78] J. Kennedy, R. Eberhart, Particle Swarm Optimization, Proc. ICNN'95-International Conf. Neural Networks. 4 (1995) 1942–1948. https://doi.org/10.1109/TST.2016.7442504.
[79] S. Mirjalili, A.H. Gandomi, S.Z. Mirjalili, S. Saremi, H. Faris, S.M. Mirjalili, Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems, Adv. Eng. Softw. 114 (2017) 163–191. https://doi.org/10.1016/j.advengsoft.2017.07.002.
[80] S. Li, H. Chen, M. Wang, A.A. Heidari, S. Mirjalili, Slime mould algorithm: A new method for stochastic optimization, Futur. Gener. Comput. Syst. 111 (2020) 300–323. https://doi.org/10.1016/j.future.2020.03.055.
[81] S. Mirjalili, A. Lewis, The Whale Optimization Algorithm, Adv. Eng. Softw. 95 (2016) 51–67. https://doi.org/10.1016/j.advengsoft.2016.01.008.
[82] M. Dorigo, G. Di Caro, Ant colony optimization: a new meta-heuristic, in: Proc. 1999 Congr. Evol. Comput. (Cat. No. 99TH8406), IEEE, 1999: pp. 1470–1477.
[83] C. Muro, R. Escobedo, L. Spector, R.P. Coppinger, Wolf-pack (Canis lupus) hunting strategies emerge from simple rules in computational simulations, Behav. Processes. 88 (2011) 192–197.
[84] W.A. Watkins, W.E. Schevill, Aerial observation of feeding behavior in four baleen whales: Eubalaena glacialis, Balaenoptera borealis, Megaptera novaeangliae, and Balaenoptera physalus, J. Mammal. 60 (1979) 155–163.
[85] J.A. Goldbogen, A.S. Friedlaender, J. Calambokidis, M.F. McKenna, M. Simon, D.P. Nowacek, Integrative approaches to the study of baleen whale diving behavior, feeding performance, and foraging ecology, Bioscience. 63 (2013) 90–100.
[86] D.J. Armaghani, P.G. Asteris, A comparative study of ANN and ANFIS models for the prediction of cement-based mortar materials compressive strength, Springer London, 2021. https://doi.org/10.1007/s00521-020-05244-4.
[87] A.N. Bhatt, N. Shrivastava, Application of Artificial Neural Network for Internal Combustion Engines: A State of the Art Review, Arch. Comput. Methods Eng. 29 (2022) 897–919. https://doi.org/10.1007/s11831-021-09596-5.
[88] P.G. Asteris, M.E. Lemonis, T.-T. Le, K.D. Tsavdaridis, Evaluation of the ultimate eccentric load of rectangular CFSTs using advanced neural network modeling, Eng. Struct. 248 (2021) 113297.
[89] P.G. Asteris, P.B. Lourenço, M. Hajihassani, C.-E.N. Adami, M.E. Lemonis, A.D. Skentou, R. Marques, H. Nguyen, H. Rodrigues, H. Varum, Soft computing-based models for the prediction of masonry compressive strength, Eng. Struct. 248 (2021) 113276.

[90] P.G. Asteris, P.B. Lourenço, P.C. Roussis, C.E. Adami, D.J. Armaghani, L. Cavaleri, C.E. Chalioris, M. Hajihassani, M.E. Lemonis, A.S. Mohammed, Revealing the nature of metakaolin-based concrete materials using artificial intelligence techniques, Constr. Build. Mater. 322 (2022) 126500.

[91] M.I. Khan, M.H. Sutanto, K. Khan, M. Iqbal, M. Bin Napiah, S.E. Zoorob, J.J. Klemeš, A. Bokhari, W. Rafiq, Effective use of recycled waste PET in cementitious grouts for developing sustainable semi-flexible pavement surfacing using artificial neural network, J. Clean. Prod. (2022) 130840.

[92] P. Zhang, Z.Y. Yin, Y.F. Jin, Machine Learning-Based Modelling of Soil Properties for Geotechnical Design: Review, Tool Development and Comparison, Arch. Comput. Methods Eng. 29 (2022) 1229–1245. https://doi.org/10.1007/s11831-021-09615-5.

[93] J.-S. Jang, ANFIS: adaptive-network-based fuzzy inference system, IEEE Trans. Syst. Man. Cybern. 23 (1993) 665–685.

[94] M. Parsajoo, D.J. Armaghani, P.G. Asteris, A precise neuro-fuzzy model enhanced by artificial bee colony techniques for assessment of rock brittleness index, Neural Comput. Appl. 34 (2022) 3263–3281.