applied sciences

MDPI

Article

Using Artificial Intelligence Techniques to Predict Intrinsic Compressibility Characteristic of Clay

Samuel J. Abbey *, Eyo U. Eyo. and Colin A. Booth

Faculty of Environment and Technology, University of the West of England, Bristol, BS16 1QY, UK; eyo.eyo@uwe.ac.uk (E.E.U.); colin.booth@uwe.ac.uk (C.A.B.)
* Correspondence: samuel.abbey@uwe.ac.uk

Abstract: Reconstituted clays have often provided the basis for the interpretation and modelling of the properties of natural clays. The term "intrinsic" was introduced to describe a clay remoulded or reconstituted at moisture content up to 1.5 times its liquid limit and consolidated one-dimensionally. In order to circumvent the difficulties of measuring an intrinsic constant called "intrinsic compressibility index" (C^*_c), a machine learning (ML) approach using traditional non-parametric tree-based and meta-heuristic ensembles was adopted in this study. Results indicated that tree-ensembles namely random decision forest (RDF) and boosted decision tree (BDT) performed better in C^*_c prediction (average R^2 of 0.84 and root mean square error, RMSE of 0.51) compared to stand-alone models. However, models' hyper parameters combined meta-heuristically, produced the highest accuracy (average R^2 of 0.90 and root mean square error, RMSE of 0.34). The greatest capacity to distinguish between positive and negative soil classes (average accuracy of 0.95, precision and recall of 0.86) were demonstrated by meta-ensembles in multinomial classification.

Keywords: machine learning; regression; big data; deep learning; reconstituted soil; compressibility index

Citation: Abbey, S.J.; Eyo, E.U.; Booth, C.A. Using Artificial Intelligence Techniques to Predict Intrinsic Compressibility Characteristic of Clay. *Appl. Sci.* **2022**, *12*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s):

Received: date
Accepted: date
Published: date

1. Introduction

Reconstituted clays can provide a frame of reference for an assessment of the influence of soil structure on the mechanical behaviour of intact clays. Hence, geotechnical engineers, and researchers have often relied on the mechanical properties (such as compressibility, expansion, and strength, etc.) of reconstituted or remoulded clays to interpret, extrapolate and establish the corresponding characteristics of clay subgrade materials. Based on this, several efforts have been made over the past 80 decades to formulate various constitutive models and theories to study the behaviour of reconstituted soils. One of the most popularly known reconstituted soil framework is the "critical state" soil mechanics which was developed between early and mid-20th century. Critical state soil mechanics has become so widely embraced as one of the logical concepts that can be applied to solve many engineering problems given its capacity to incorporate theories of plasticity, yielding, flow, etc for the modelling of soil behaviour [1–6].

A reconstituted clay could be defined as a clay that has been rigorously mixed at a moisture content that is equal to or greater than its liquid limit (LL). Burland [7] introduced the "intrinsic properties" concept of reconstituted clays to serve as a basis for the interpretation of the natural soil. The name "intrinsic" describes the properties of a clay that has been remoulded at a moisture content of between its liquid limit (LL) and 1.25 to 1.5 times its LL (without the need for air or oven drying) and then consolidated one-dimensionally. Figure 1 depicts the intrinsic or inherent compression curve for a given remoulded clay. The values e^*_{100} and e^*_{1000} are the intrinsic void ratios that correspond to effective state pressures (σ_v) 100 kPa and 1000 kPa respectively. It is important to note that

the asterisk is used to signify an intrinsic property. Figure 1 also indicates the normalisation of the values of intrinsic compressibility by the assignment of fixed values to e^*_{100} and e^*_{1000} through a parameter referred to as the void index, I_v (measure of the compactness of a given soil). Void index is mathematically expressed as in Equation (1), and could be seen in Figure 1, that a unique slope called the intrinsic compression line (ICL) has been achieved, and this slope represents the remoulded clay subgrade material itself rather than one that has undergone some post-depositional modifications through weathering, desiccation, unloading, etc.

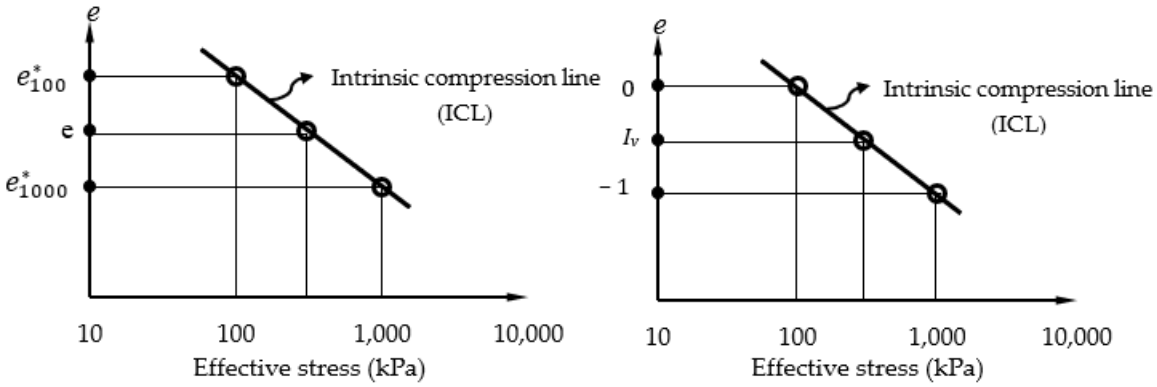

Figure 1. Plots of void ratio vs. effective stress showing intrinsic constants, ICL and void index.

In order to obtain an accurate measurement and assessment of intrinsic compressibility index of clays, very expensive and time-consuming experiments are often performed using a conventional oedometer or in some cases a modified form of the consolidometer. Moreover, when undertaken on clays with very high initial moisture contents, the procedure may even become inevitably cumbersome. The associated difficulties of measuring intrinsic compressibility index prompted a growing body of research in the use of statistically based estimates or correlation equations. Burland [7] demonstrated that soil's intrinsic constants can be empirically correlated with its Atterberg limits. Following this, several modifications to Burland's relationships has been made in the most recent past by various researchers [8–13]. Some of these studies have also attempted to extend these empirical correlations to cover the influence of factors such as initial moisture content, mineralogy, etc on the intrinsic compressibility of clays [14–17]. For instance, Xu and Yin [14] investigated the influence of different initial water contents on the compression behavior of three clays. They concluded that intrinsic compression indices tend to increase nonlinearly with increasing initial water contents. On the other hand, research carried out by Habibbeygi et al. [16], indicated that an inherent property such as clay mineralogy can have a considerable impact on the values of intrinsic constants of reconstituted clays.

Most of the presently used correlation models of predicting the intrinsic compressibility index of clays are composed essentially of relationships developed from linear regression techniques. Thus, the resulting analytical correlation equations only tend to determine unknown coefficients that affect the relationship of an intrinsic constant such as the compressibility index. Although these models may be effective in some instances however, they are mostly fraught with a lot of shortcomings that relate to the inherent nonlinearities and complexities of the interrelationships between soil variables. Hence, an application of artificial intelligence (AI) techniques through machine learning (ML) paradigms are proposed herein to solve the challenges of forecasting C^*_c.

This study uses a ML approach to intelligently model the intrinsic compressibility index of clays by adopting non-parametric tree-based ensemble learners (decision forest and boosted decision trees) and meta-heuristic ensembles or combinations of hyperparameters such as the voting and stacking ensembles. Predictions using stand-alone algorithms (multilinear regressors, Bayesian linear regressors, logistic regressors and artificial

neural networks) are also performed and compared with those of the ensemble learners. Furthermore, since various types of soils of different classes (classified according to the USCS) are utilised for the prediction, ML multiclass or multinomial classification is applied for the first time as a diagnostic test to determine the ability of the classifiers to adequately learn between the soil types and by so doing discriminate between positive and negative classes.

$$I_v = \frac{e - e^*_{100}}{e^*_{100} - e^*_{1000}} = \frac{e - e^*_{100}}{C^*_c} \quad (1)$$

2. Methodology

2.1. Database Generation and Pre-Processing

Very high-quality dataset of intrinsic compressibility index and basic soil's index parameters compiled from rigorous literature search are utilised for this study. Standardised methods of oedometer testing were adopted by the authors for data collection from the reconstituted soils [11,18–33]. Given the diverse nature and sets of data of intrinsic compressibility index (herein considered as the independent feature), it was necessary to transform and normalise these data into usable continuous variables to enable an improvement of the significance of findings, greater size effects, lesser threats to any causal inferences (i.e., the validity of statistical conclusion), and more reliable results. A two-step approach of data transformation was followed [34]. Step one involved a transformation of the intrinsic compressibility variables into a percentile rank resulting in uniformly distributed probabilities. Step two then applied the inverse-normal transformation to the results of step one into a variable comprising normally distributed z-scores. Figure 2 indicates the normally distributed dataset of the method with very low values of skewness (0.123457) and kurtosis (−0.09139) (Table 1) meaning therefore that the dataset is very reliable for use in the ML modelling.

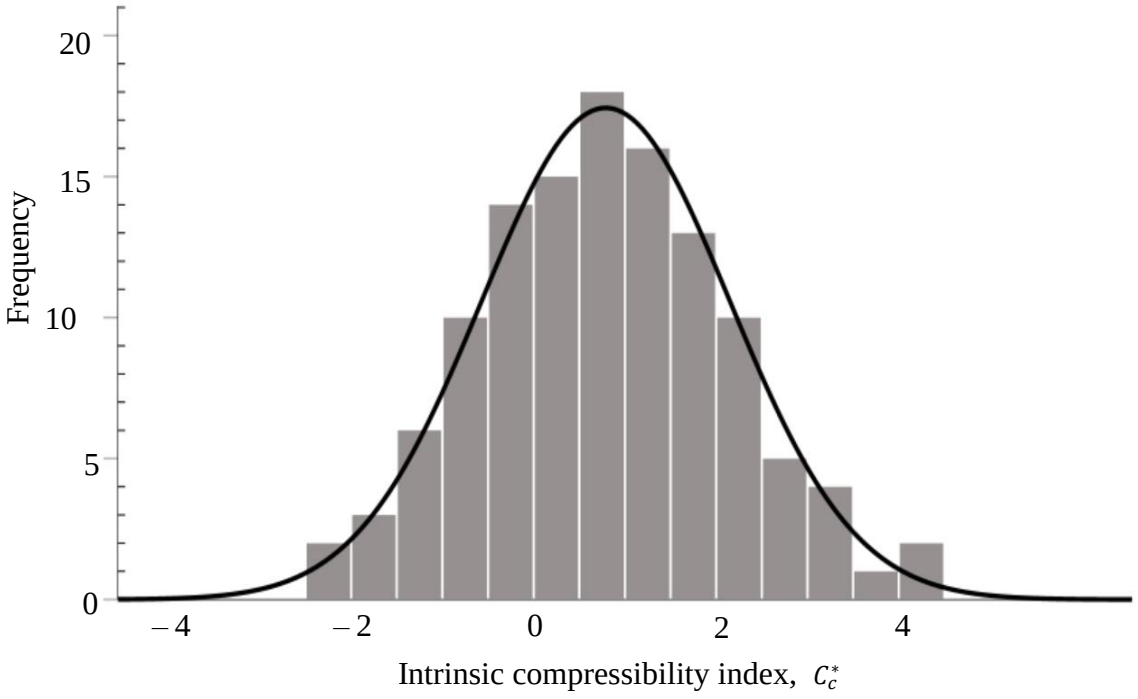

Figure 2. Distribution of intrinsic compressibility index.

Table 1. Statistics of intrinsic compressibility index (C_c^*.).

Mean	Standard Error	Standard Deviation	Kurtosis	Skewness	Min.	Max.	Range
0.789244	0.124782	1.36121	−0.0914	0.123457	0.08	8.5	8.42

On the other hand, a total of six explanatory features are used in this research namely, reconstituted soil intrinsic constants (e_L, e_P and e_{100}), Atterberg limits (Liquid limit (LL) and plasticity index (PI)) and a soil texture parameter, specific gravity (G). Table 2 depicts important statistical components of these features. It is observed that the range of values for the liquid limits (minimum and maximum of 22% and 560% respectively) and plasticity indices (minimum value of 5% and 508% respectively) suggest wide coverage of the soils of different plasticity properties. Frequency distribution of the independent features of the dataset as depicted in Figure 3 generally indicates non-uniform distribution for all the variables except for the specific gravity of the soils. The pattern of distribution which is noticed to be mostly right skewed indicates a very strong relationship among the features. Three classes of soils defined according to the USCS were captured in this study. This will be very useful for the multinomial classification prediction considered subsequently in this research to determine the ability of the models to learn between the different categories of soils used.

Table 2. Statistics of explanatory features.

Statistic	G	LL	PL	PI	e_L	e_p	e_{100}^*
Mean	2.71	87.62	31.11	56.50	2.37	0.84	1.51
Standard Error	0.01	8.375	1.196	7.82	0.22	0.03	0.10
Standard Deviation	0.10	91.36	13.05	85.25	2.43	0.36	1.08
Kurtosis	0.86	13.93	9.15	16.11	12.94	11.59	9.53
Skewness	-0.5	3.60	2.05	3.94	3.49	2.36	2.78
Range	0.56	538	96	503	13.68	2.75	6.65
Min.	2.37	22	12	5	0.59	0.31	0.45
Max.	2.93	560	108	508	14.27	3.06	7.10

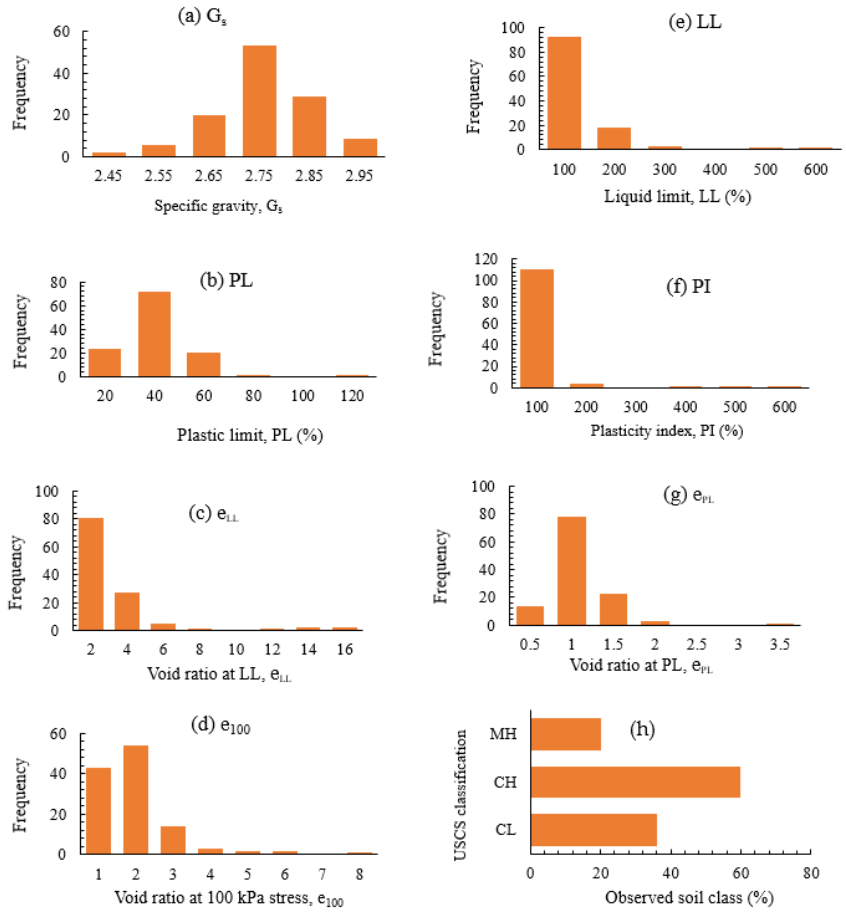

Figure 3. Distribution of ML features. (a) specific gravity; (b) Plastic limit; (c) Void ratio at LL; (d) Void ratio at 100kPa stress; (e) Liquid limit; (f) Plasticity index; (g) Void ratio at PL; (h) Soil class.

2.2. ML Cross-Validation

Using the cross-validation technique enables an assessment of both the variability of the dataset used and the reliability of the models utilised to train and test through the data. In this research, the dataset was first used to train and test the models by first splitting the data in the ratio 80:20. This means that 80 percent of the dataset were used for model training while the remaining 20 percent were utilised for testing. This method shall be referred simply as the train-(validation)-split (TVS) method. For the application of cross-validation, both the k-fold cross-validation (kFCV) and Monte Carlo cross-validation (MCCV) techniques were applied. Cross-validation employed in this way, enabled the sensitivity of the ML prediction to be tested especially when using the TVS method in order to ensure that overfitting of the dataset was avoided in the modelling process.

k-Fold and Monte Carlo Cross-Validation Techniques (kFCV)

In the k-fold cross-validation (kFCV) techniques, the training, testing, and validation are performed by splitting the *N*-dataset into *k* (*k* is typically set to 3, 5, or 10) mutually exclusive subset depending on the size of the data. The model is then trained on a collection of *k*–1 subset and the testing done on the remainder of the *k*th subset. This process is then iterated *k*-number of times and each time, a different subset would sequentially take up the role of the so-called "test set". The resulting *k*-test statistical predictions are then averaged to obtain a more realistic and representative output of the generalised performance of the model used. In this study, the values of *k* utilised are both 5-and 10 (i.e., five- and 10-fold CV) for the sake of comparison.

The Monte Carlo cross-validation (MCCV) is also an iterative technique and could be regarded loosely as a combination of the TVS and the kFCV methods but with some slight

variations. The MCCV technique involves splitting of the N dataset into n_t and n_v subsets by random sampling and without replacing the n_t subset data points. The n_t subset is then applied to train the model and validation performed on the n_v subset. It should be noted that unlike kFCV, there exist an $\binom{N}{n_t}$ unique training set, however, MCCV circumvents the need to run these many repetitions. It is very necessary to bear in mind that the choice of the number required for dataset splits (k and/or n_t) can influence the trade-offs between bias and variance. The larger the values of k and/or n_t, the higher the variance and the lower the biases. Moreover, overfitting could be the end result because larger training datasets tend to be more similar between iterations. An assessment of this phenomena shall be carried out subsequently during an analysis of the ML models.

3. Supervised ML Models

3.1. Tree-Based Ensembles and Decision Forest Classifier Models

Decision tree prediction models do resemble a natural tree plant with leaf nodes and decision branches functioning mostly by the aggregation of its separate parts into an ensemble as shown in Figure 4. A series of simple tests are performed for each instance by traversing a binary tree-data structure up until a decision node (or leaf node) is reached. There are different kinds of tree-based models however, in this study, the random decision forest and the boosted decision tree are considered.

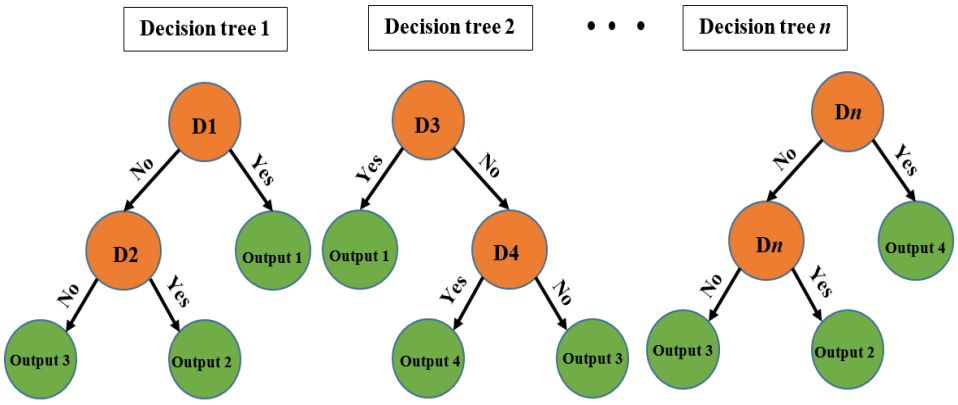

Figure 4. Typical decision trees.

3.1.1. Random Decision Forest (RDF) and Boosted Decision Trees (BDT)

The random decision forest regressor or classifier model is composed of an ensemble of various non-parametric decision trees. An individual tree in an RDF would output a Gaussian distribution as a prediction. The aggregation of trees is performed over the tree combination to search for the closest Gaussian distribution to the combined total distribution of all the trees in the decision model. The boosted decision tree model combines individual trees to reach a decision by using the technique of "boosting" in order to increase the accuracy of prediction. Boosting simply means that each succeeding tree is dependent on the preceding on. Hence, the algorithm or model learns by fitting each tree's residual that precedes it.

3.1.2. Meta-Heuristic Ensembles and Voting Ensemble (VE)

In order to further improve ML predictions, meta-ensembles are used. These are constructed by an aggregation of several models or hyperparameters and thus are also referred to as model of models. Some of the hyperparameters of the forgoing models were combined through averaging by the voting and stacking systems. The voting ensemble estimates the average predictive output of the sum of aggregated models through a majority voting exhibited by the model with the highest prediction confidence or the model that has the most popular output. A typical structure of VE is shown in Figure 5.

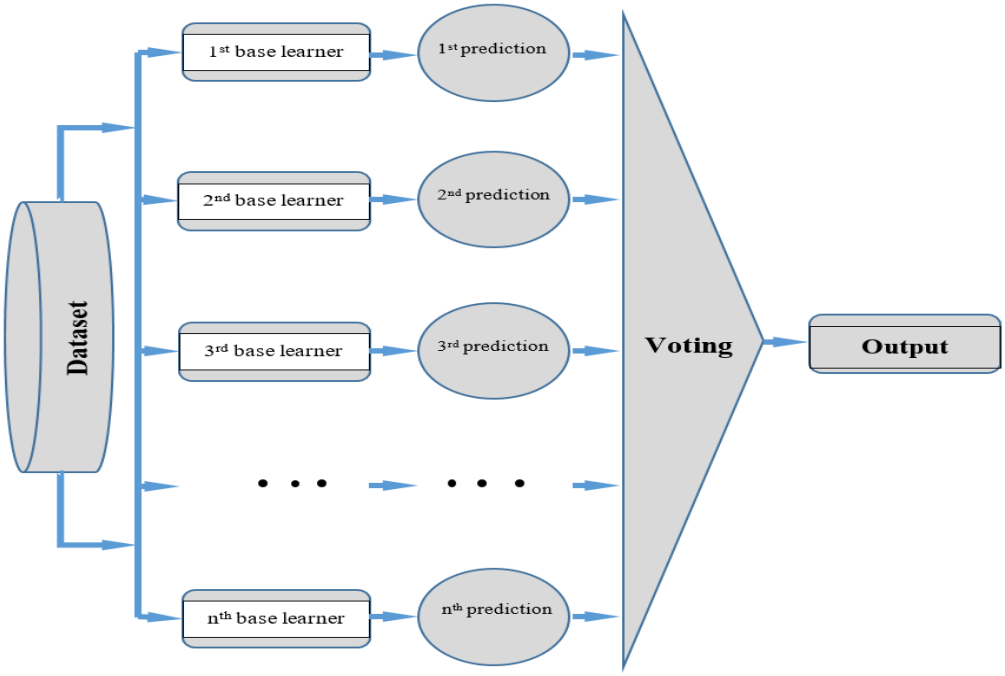

Figure 5. Typical structure of voting ensemble.

While the stacking ensemble (SE) is an adjunct of the averaging system where the hyperparameters learn and choose when to depend on themselves to enable a more generalised multistage prediction. Hence, the output of preceding models would become or serve as inputs for the subsequent models and predictions are being made as shown in Figure 6.

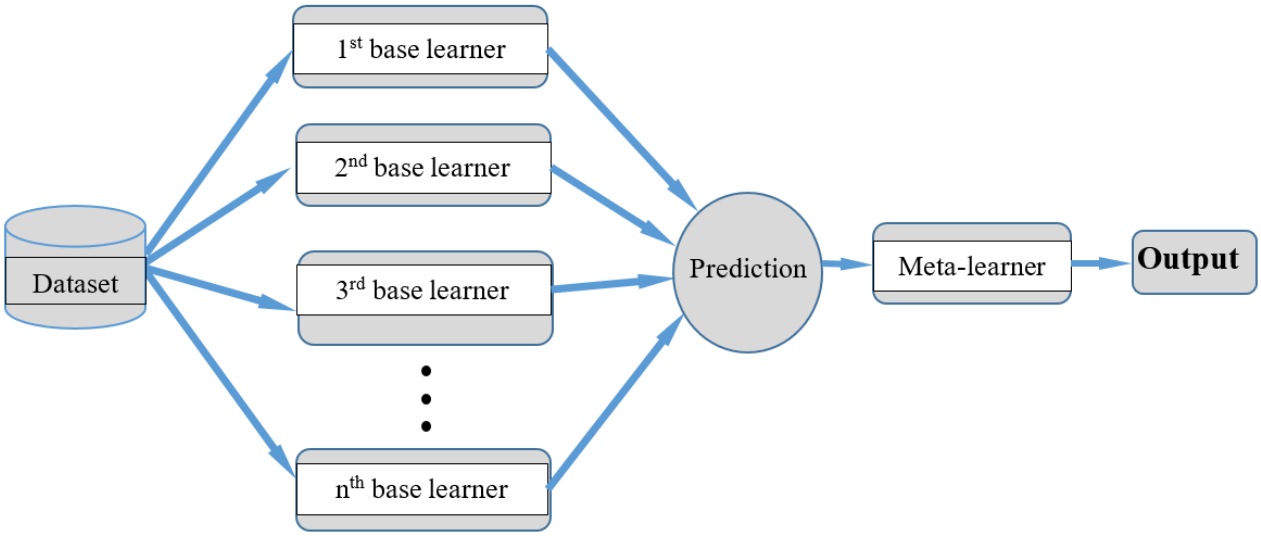

Figure 6. Typical structure of stacking ensemble.

3.2. Stand-Alone Algorithms

3.2.1. Linear Regression (REG)

Multiple linear regression is normally applied to enable estimates of certain unknowns (variables, parameters, or coefficients) by demonstrating how a change in one or more independent set of variables can affect a corresponding predictor variable. Mathematically, the general form of REG can be expressed as:

$$Y_n = \mu + \sum_{n=1}^{m} \alpha_n \cdot x_n \quad (2)$$

where Y = predictor variables and X_1, X_2, X_3,…, X_m represent the independent variables plus the error term that accounts for certain other unknown factors in the prediction.

3.2.2. Logistic Regression (LR)

Logistic regression normally arises from a conditional probability modelling that suggests that the outcome or predicted variable say $Y = 1$, given a set of input or predictor variables say X. Mathematically, the conditional probability (or the hypothesis function) is modelled by LR as:

$$P_\omega(y = \pm 1|x) = \frac{1}{1 + exp^{-y\omega^T x}} \quad (3)$$

where x represents the dataset, y represents class label and $\omega \epsilon \Re^n$ *is* the weight vector. For a binary classification problem with two-class training dataset $(x_i,y_i)_{i=1}$, $x_i \epsilon \Re^n$, $y_i \epsilon (1, -1)$, then LR tends to maximise the following regularised negative likelihood:

$$p\ (\omega) = C \sum_{i=1}^{l} \log\left(1 + exp^{-y\omega^T x_i}\right) + \frac{1}{2}\omega^T \omega \quad (4)$$

where $C > 0$ is regarded as a penalty parameter. It is important to mention that several optimisation techniques have been applied on a LR problem and some of which are documented in the literature [35,36].

For a multinomial classification problem, the conditional probability could be modelled using a maximum entropy as:

$$P_\omega(y|x) = \frac{\exp\ (\omega^T f(x, y)}{\sum_{y'} \exp\ (\omega^T f(x, y'))} \quad (5)$$

Here, the function vector is expressed $f(x, y) \epsilon \Re^n$

3.2.3. Bayesian Linear Regressor (BLR)

Just like the LR, Bayesian linear regressor represents a special case of REG that allows modelling to be performed within the "Bayes" theorem statistical inference.

Hence, for a given dataset say $D = (x_1, y_1),..., (x_n,y_n)$ where $x \epsilon \Re_d$, and $y \epsilon \Re$, then a BLR model can be expressed as:

Prior:

$$\omega \sim \mathcal{N}(0, \sigma_\omega^2 I_d)$$

Ω is the vector $(\omega_1,..., \omega_d)$T, making the previous distribution a multivariate Gaussian; and I_d is a d x d identity matrix.

Likelihood

$$Y_i \sim \mathcal{N}(\omega^T x_i, \sigma^2)$$

With the assumption that $Y_i \perp Y_j | \omega, i \neq j$

If we use the variance, $a = 1/\sigma^2$, and $b = 1/\sigma^2\omega$ then we assume that a and b are unknown.

We state the prior as:

$$p(\omega) \propto exp\left\{-\frac{b}{2}\omega^t\omega\right\} \quad (6)$$

Furthermore, the likelihood stated as:

$$p(D|\omega) \propto exp\left\{-\frac{a}{2}(y - A\omega)^T(y - A\omega)\right\} \quad (7)$$

where $y = (y_1,...,y_N)^T$ and A is a $n\ x\ d$ matrix

Then, the posterior is:

$$p(\omega|D) \propto p(D|\omega)p(\omega)$$

Which ultimately produces the expression:

$$p(\omega|D) \sim \mathcal{N}(\omega|\mu, \Lambda^{-1})$$

where the precision matrix $\boldsymbol{\Lambda}$ is:

$$\Lambda = aA^TA + bI_d$$

$$\mu = a\Lambda^{-1}A^Ty$$

For the predictive posterior:

$$p(y|x, D) = \int p(y|x, D, \omega)p(\omega|x, D)d\omega = \int p(y|x, \omega)p(\omega|D)d\omega$$

It is then possible to then get the following:

$$y|x, D \sim \mathcal{N}(\mu^Tx, \frac{1}{a} + x^T\Lambda^{-1}x) \quad (8)$$

3.2.4. Artificial Neural Networks (ANN)

The structure and architecture of neural networks are inspired by the network of input—processes (decisions)—output system such as that of the human nervous system. More technically, the information or data processing capability of ANN is represented as a network of input, hidden, and an output layers as depicted in Figure 7.

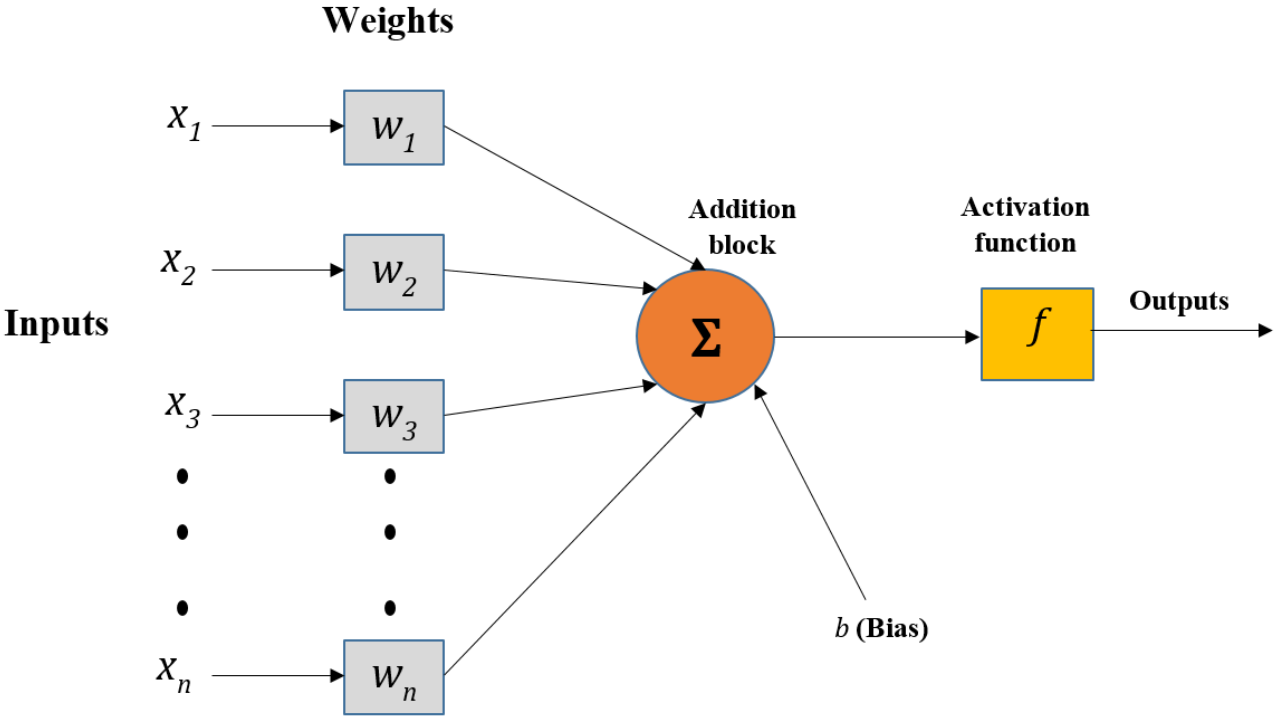

Figure 7. Architecture of artificial neural network.

Where x = inputs, w_{ij} = neurons' weight, b = bias and f = activation function that enables the inputs to be transformed into the output by inputs' (processing neuron) multiplied by the corresponding weights.

3.3. ML Model Implementation and Multiclass-Class Evaluation Metrics

The implementation and execution of dataset (training, testing, and evaluation) by the models was conducted on a cloud-based platform that supports Python programming including its associated libraries. The properties of ML algorithms adopted for optimal performance are presented in Table 3. The ML methodology followed in this study is summarised in Figure 8.

Table 3. Models' optimised parameter settings.

Stand-Alone Algorithms		
	Parameter	Option/value
REG	Regularisation wt. (L_2)	0.001
LR	Method	Ordinary least squares (OLS)
	Optimisation tolerance	1×10^{-7}
	Regularisation wt. (L_1 & L_2)	1.00
BLR	Regularisation wt. (L_2)	1.00
ANN	Normaliser	min-max
	No. of hidden nodes	100.00
	No. of iterative learning	100.00
	Hidden layer spec.	Full connection
Tree-ensembles		
	Parameter	Option/value
Boosted decision tree (BDT)	Constructed trees	100.00
	Tree-forming training instances	10.00
	Leaves/tree (max.)	20.00
	Tree depth (max.)	32.00
Random decision forest (RDF)	Constructed trees	8.00
	Method of resampling	Bagging
	Samples/leaf node (min.)	1.00
	Randomised splits/node	128.00

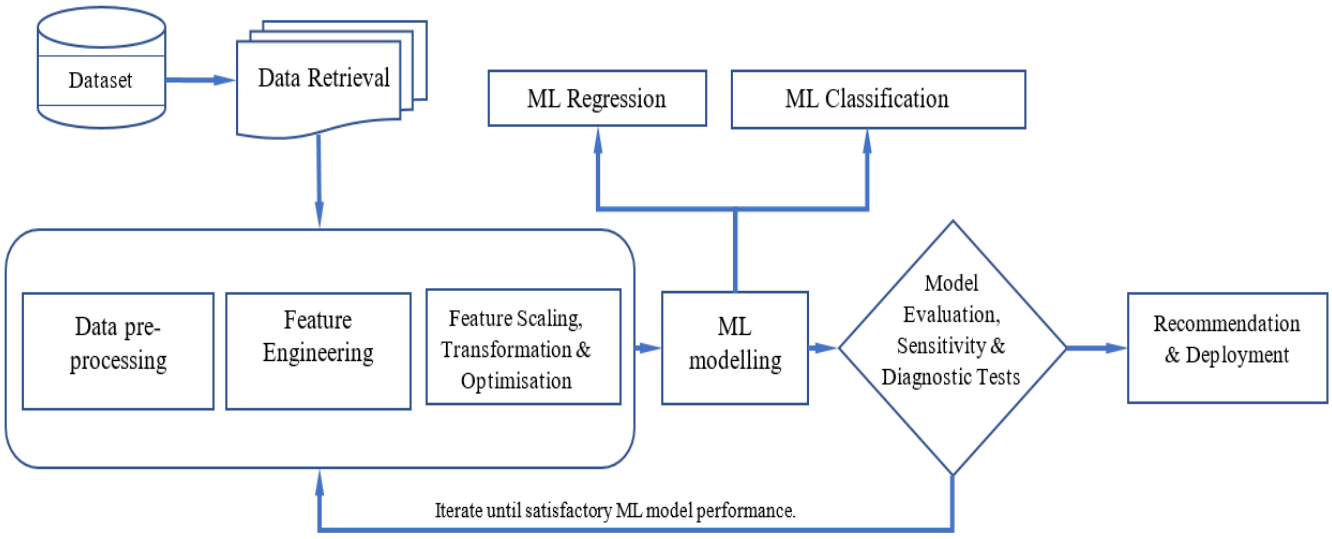

Figure 8. Summary of ML methodology.

In terms of the multiclass evaluation metrics, the indicators of performance considered in this study for ML regression are the coefficient of determination (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics. Detailed discussions for these frequently used regression metrics are given in the literature [37,38]. However, for ML multinomial classification, the following performance metrics shall be elucidated:

Accuracy—is simply an estimate of the average number of correct predictions in a ML classification problem. It can be expressed as:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (9)$$

where: TP= True Positive of prediction; TN= True Negative; FP= False Positive; FN= False Negative

Precision—is defined as the TP divided by the sum of the positively predicted outcomes. In order words, precision expresses the model's unit proportion that are positive as being actually positive. Precision is given as:

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

Recall—is defined as the TP divided by the sum of the positively predicted outcomes in which case, unlike precision, FN are the labels that have been classified as negative even though they are actually positive**.** Recall can be expressed as:

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

4. Results and Discussion

4.1. ML Regression

Table 4 indicates the statistical measures of performance of the ML regression algorithms used for prediction of intrinsic compressibility index of the soils. The TVS method was used for training, testing, and subsequent validation of the datasets. An indication of good performance is usually depicted by high values of coefficient of determination (R^2) with corresponding low mean error values given as the mean absolute error (MAE) and the root mean square error (RMSE). Among the traditional regression models, the tree ensembles (RDF and BDT) seem to outperform the stand-alone models (REG, ANN and BLR) as could be observed in Table 4. With regards the stand-alone algorithms, BLR does clearly produce the least accuracy when compared to REG and ANN. The outcome of REG and BLR on a non-linear problem such as that presented by this study may not be entirely surprising given the models' underlying assumptions of linearity thus, being unable to implicitly detect all the possible combinations or interactions between explanatory and predictor variables. Notwithstanding, the relatively less accuracy of prediction given by ANN is quite remarkable. Though, depending on a given regression problem, opinions are quite divided as to the behaviour of ANN as a result of its inherent structure comprising of 'black boxes' which may tend to either cause an over-estimation or not being able to explicitly learn several unobserved causal relationships during data training. ANN has a criterion that is not well-established to enable an interpretation of the weights and biases that exist in a connection matrix. In ANN's backpropagation method of instinctive search and optimisation of its weight matrix, successive upgrades or updates tend to all converge to local minimum error spaces rather than one that is more global.

Table 4. ML regression performance metrics.

Model	R^2	RMSE	MAE
		%	%
REG	0.67	0.73	0.54
ANN	0.59	0.81	0.61

BLR	0.19	1.14	1.02
BDT	0.82	0.53	0.44
RDF	0.86	0.48	0.38
SE	**0.93**	**0.34**	**0.29**
VE	**0.93**	**0.33**	**0.26**

Among the tree-based algorithms, RDF does appear to outperform (R^2 of 0.86 and RMSE of 0.48) though slightly, the BDT with an R^2 of approximately 0.82 and RMSE of about 0.53 both of which have incorporated the technique of 'bagging' or 'bootstrapping' in order to improve their performances. This method of boosting in addition to their innate architecture does improve the tree-ensembles' capacity to learn and predict the complexities of non-linear interactions between the input features especially when compared to the afore-mentioned stand-alone models. Nonetheless, tree-ensembles are also known to be ridden with some setbacks of their own. Since the tree-ensembles function through a progressive learning with subsequent 'tree construction' based on a previously fed training dataset, there may be an intuitively 'greedy' construction phase where a perceived best entity is most preferred and therefore selected without consideration of a successive aggregation of another entity which might give even more accuracy than the previous one. This phenomenon does result in a loss of information during training or testing because of the continuous splitting and partitioning process. Overall, as could be observed from Table 4, when all the above-mentioned models' hyperparameters are meta-heuristically combined into meta-ensembles (model of models) through the techniques of voting (VE) and stacking (SE), the accuracy of prediction appears to increase remarkably. Although, a closer examination would indicate that the method relying on voting does slightly produce better prediction (RMSE of 0.33 and MAE of 0.26) than model hybridisation through stacking (RMSE or 0.34 and MAE of 0.29). Further analyses and discussions regarding the sensitivity of the best performing meta-ensemble models are given in sections following.

4.2. Sensitivity Analysis of the Meta-Ensembles

4.2.1. Comparing between Cross-Validation Techniques

Table 5 and Figure 9 compare the performance of the hybrid meta-heuristic ensemble models across 3 cross validation methods. At first glance, it is observed that the method relying on the TVS seems to generally produce the highest values of R^2 (with corresponding lower RMSE and MAE scores) from Table 5. This phenomenon is also demonstrated by the less degree of deviation of the predicted curve from the ideal line in Figure 9. On the other hand, the iterative method of training, testing, and validation for which kFCV and MCCV depend, tend to produce slightly lowers values of R^2 compared to the TVS method. As mentioned previously, both kFCV and MCCV are used to improve the prediction of ML regression analyses while also serving as fine-tuning mechanisms to the TVS technique. Hence, the modelling with training and testing carried out using TVS when compared to kFCV and MCCV techniques seems to now indicate slight overfitting. Nonetheless, from Table 5 it is observed that the MCCV method does generally provide slightly much better accuracy than its kFCV counterpart. When also comparing between models, the prediction offered by the meta-ensembles using MCCV as a technique of cross validation, can be said to be more accurate than those of the stand-alone and tree ensembles which rely on the TVS method (Table 4). Although, when considered in terms of their Bias-Variance trade-offs, the MCCV is mostly deemed as having greater biases than the kFCV but with the former seeming to provide slightly more confidence in ML predictions given that it is more repeatable than the later due to its capacity to provide results with lower degrees of variance. Besides being characteristically prone to giving results with high variance, one other reason for KFCV's poor performance is attributable to the number of its partitioning being limited by the number of folds used. However, a much closer examination of Table 5 indicates that ML testing and validation relying on a smaller number of folds (for both MCCV and kFCV) does marginally produce more accuracy.

Table 5. ML regression performance metrics by method of dataset testing and validation.

Model				R^2	RMSE	MAE
	CV Method	**Set**	**No. of Folds**		**%**	**%**
SE	TVS	20	-	**0.93**	**0.34**	**0.29**
		10	-	0.67	0.66	0.62
	KFCV	-	5	**0.88**	**0.44**	**0.33**
		-	10	0.86	0.44	0.33
	MCCV	10	5	**0.89**	**0.40**	**0.31**
		20	5	0.88	0.46	0.35
		10	10	0.88	0.43	0.34
		20	10	0.88	0.46	0.36
VE	TVS	10	-	0.91	0.34	0.29
		20	-	**0.93**	**0.33**	**0.26**
	KFCV	-	5	**0.88**	**0.43**	**0.32**
		-	10	0.86	0.43	0.32
	MCCV	10	5	**0.90**	**0.40**	**0.32**
		20	5	0.88	0.45	0.35
		10	10	0.87	0.44	0.35
		20	10	0.88	0.46	0.36

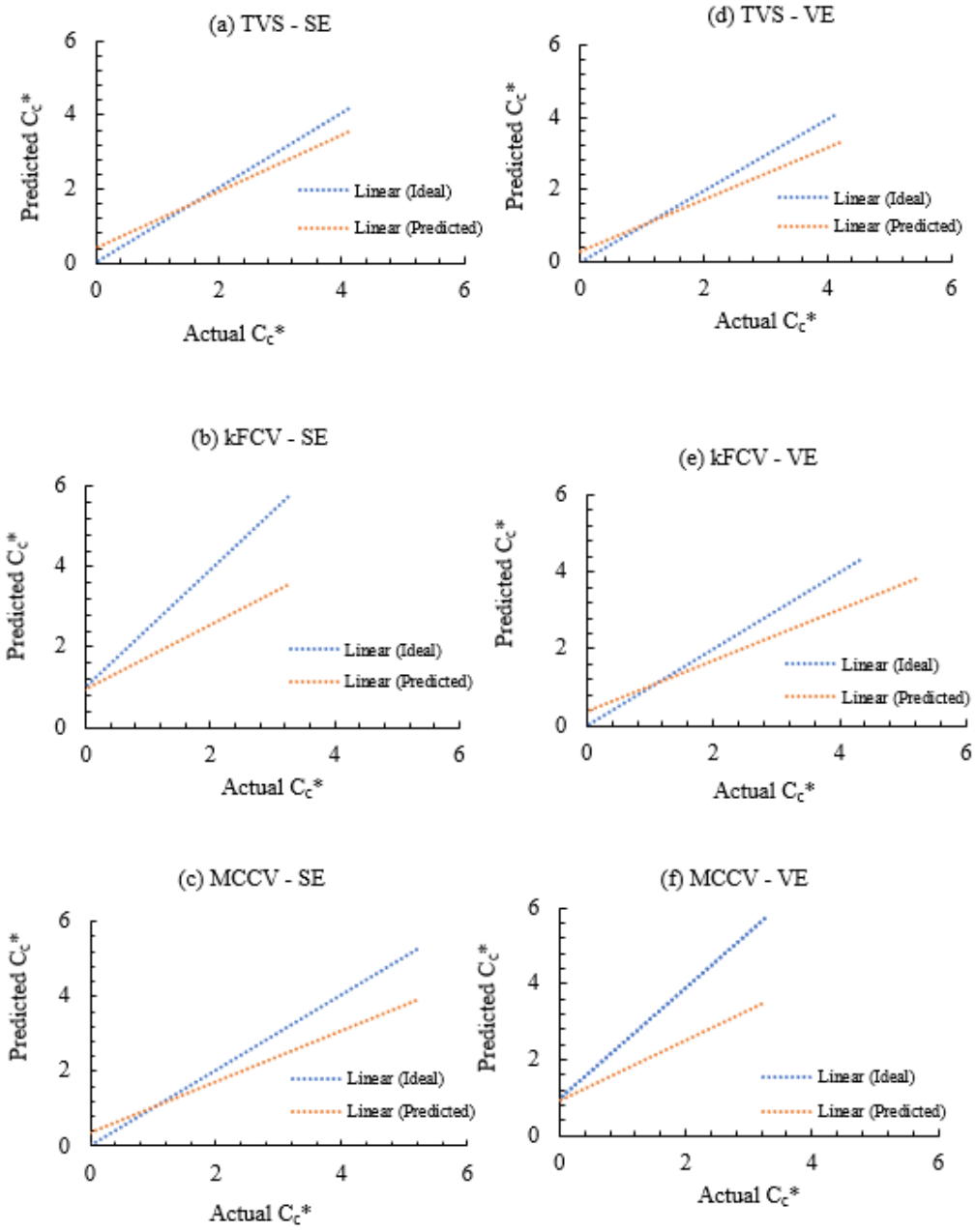

Figure 9. Ideal v actual intrinsic compressibility index.

It is important to bear in mind that depending on the situation, or the regression problem faced with, both MCCV and kFCV have been reported in the past as being able to provide better predictions than TVS. Nevertheless, within the context of this study, it is seen that the reverse seems to be the case. Again, judging from the method of cross validation used, model combination through voting does outperforms that done by stacking as Figure 9 indicates.

4.2.2. Model Residuals

Plot of the residuals of prediction does provide a means of validating ML models. By using residual plots, the observed errors of the best performing meta-heuristic models are assessed to ensure they are consistent with their corresponding stochastic errors (i.e., their randomness and unpredictability). Adopting this procedure is very necessary in this study given that none of the models used herein was able to produce 100% accuracy

hence, these models could be regarded as inherently possessing some slight degree of errors. Another reason for using the residual plot is to assess the claims made of supposed overfitting in using the TVS validation methods and to evaluate the closeness of prediction when utilising both iterative cross validation methods—kFCV and MCCV.

A good model must always show independence of the residuals by having random errors left when learning the dataset. Figure 10 compares the independence of the errors of the meta-heuristic models across the validation methods. It is very interesting to observe that the TVS method does show a less independence of the stochastic errors even though ML dataset training relying on it showed earlier that its performance in terms of the statistical measures was the highest. However, as indicated in Figure 10, the data points are less symmetric about the origin while there is a corresponding high density of points and some measures of trending around the zero line. Moreover, there appears to be a pattern of distribution that is linear along the horizontal axis. This is indeed a confirmation of the less accuracy of the models' prediction due to overfitting, when using the TVS method for training and testing of the dataset in this research. This also means that the models could not capture completely, the predictive information presented by the data hence, the reason there is a seepage of the data into the residuals.

Figure 10 also shows that the residuals of ML prediction given by the models with cross validation performed using both kFCV and MCCV techniques are quite independent. As could be observed, there are no forms of trending of the data points around the origin rather, there is much scatter and disordered patterns. Nevertheless, Figure 10 shows that that ML prediction of intrinsic compressibility with the cross validation carried out by using MCCV, provides the most accuracy given the symmetric and random distribution of residuals (no observable trends) about the origin for both voting and stacking models. Again, this diagnostic test has further validated the training, testing and validation of the meta-heuristic models performed using the MCCV as the best technique at least within the confines of this study.

4.2.3. Distribution of Residuals

Normal frequency distribution of the residuals does provide another ground for which a ML model effectiveness and authenticity can be assessed. Figure 11 indicates the normal distribution and histogram of the meta-heuristic models under three validation methods. Notice how both the models trained and validated by TVS and kFCV tend to be biased towards predicting values that are higher and lower, respectively, than the actual values of intrinsic compression index thus confirming their behaviour as demonstrated previously by their residual plots. On the other hand, ML prediction carried out with the models trained and validated using MCCV shows a much-balanced distribution as shown by the curve and indicating much better prediction. The corresponding histograms are also used to validate the accuracy of ML prediction. In this case, a very good model will tend to have its residuals peaking at zero but with few of the stochastic errors at its extremes while a low performing model will have its residual distribution spreading out but with fewer errors around zero. Using this theory, it could be observed that the best performing meta-ensemble models do have their training and validation done by adopting the MCCV method. It should also be mentioned that the voting technique across the three cross validations tend to slightly perform better than the technique of voting in general.

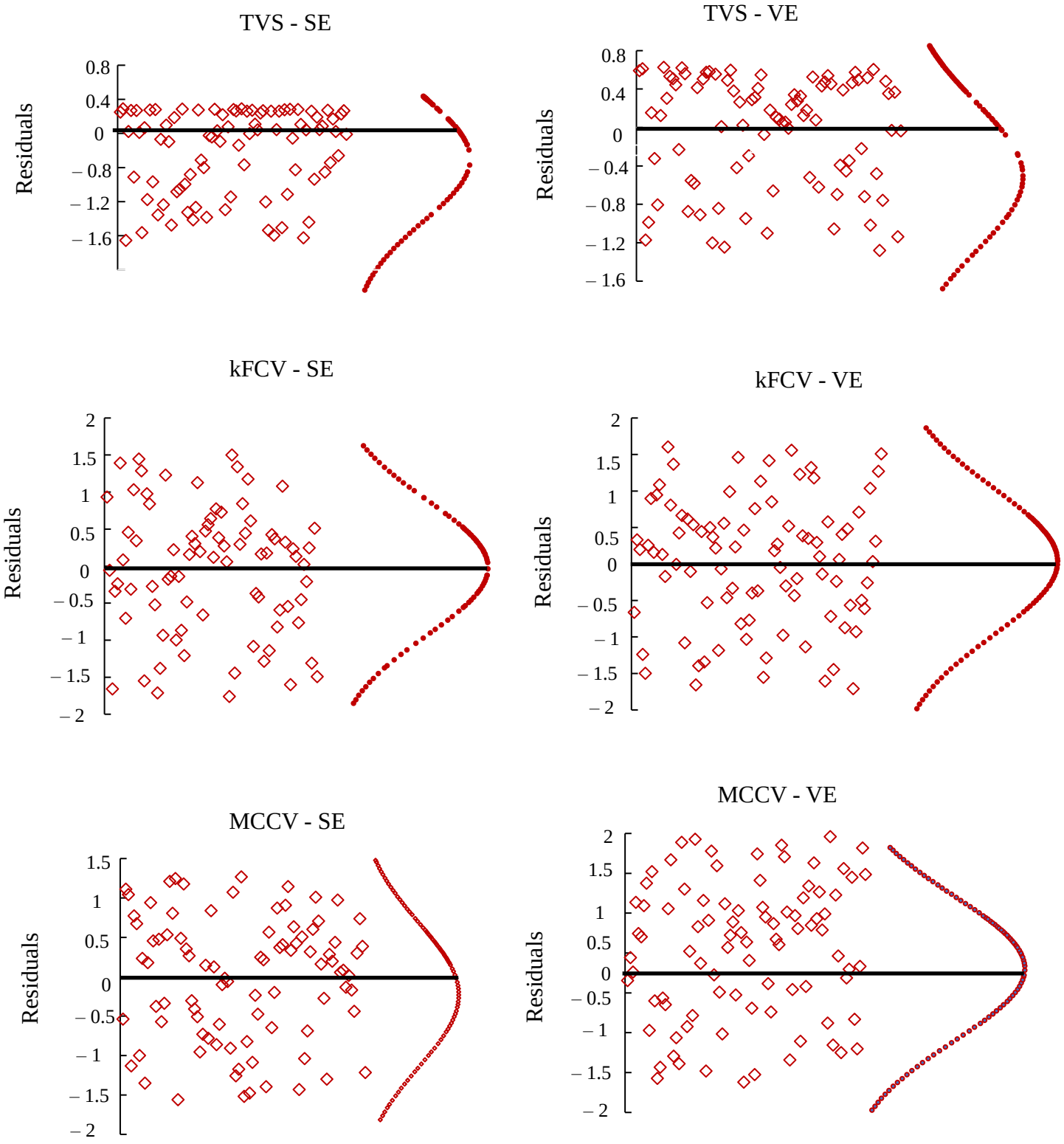

Figure 10. Residual plots.

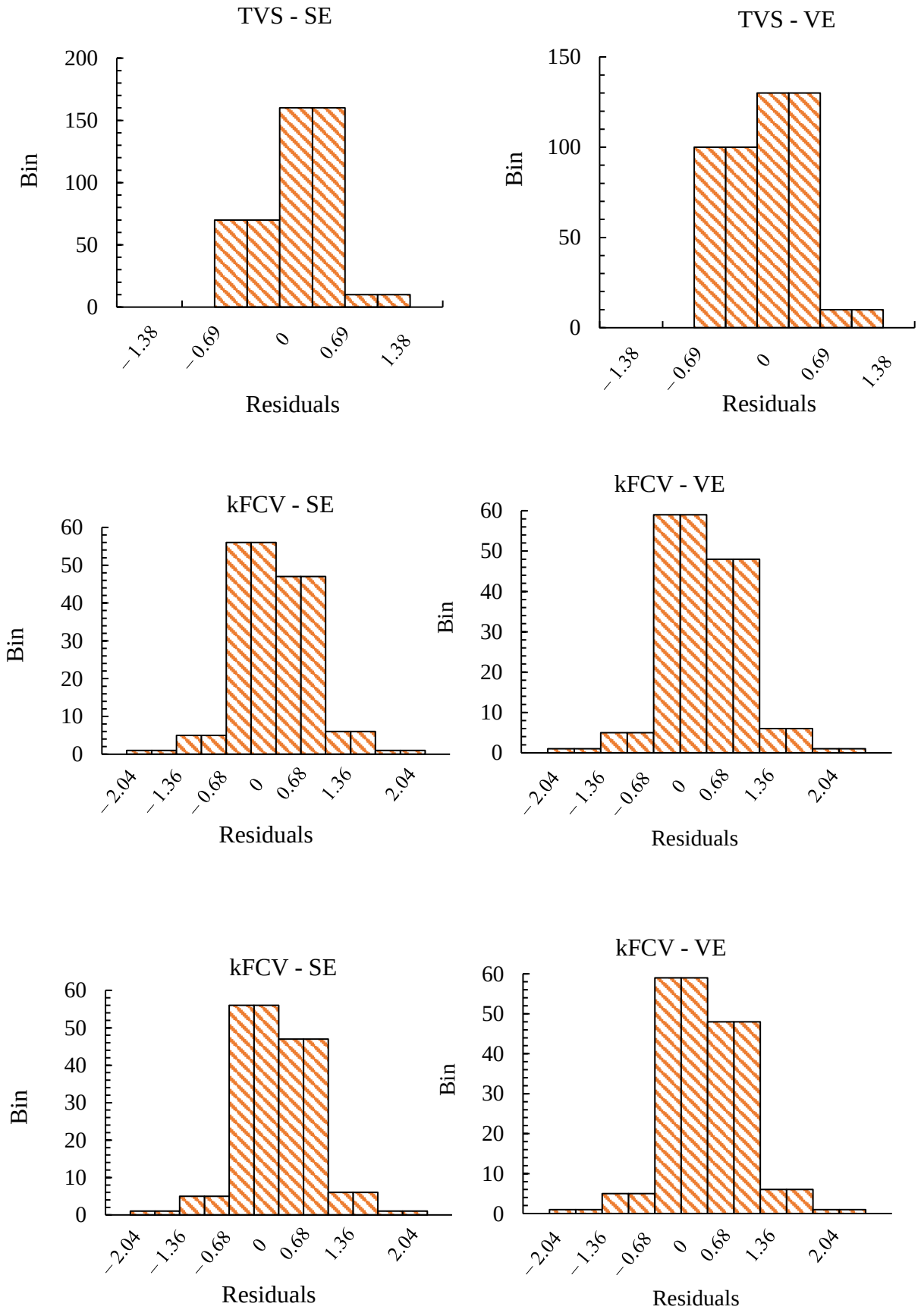

Figure 11. Distribution of residuals.

4.3. ML Classification

Different types of soils have been used for ML prediction from the forgoing. Hence, machine learning classification shall be applied as another form of diagnostic test in the forecast to ensure that the algorithms used properly learnt the different soil types and as such able to discriminate between each of the soil categories. By relying on their plasticity properties, this study utilises three different classes of soils defined according to the unified soil classification system (USCS) in the ML prediction. Hence, for the type of ML classification problem considered herein, the multiclass elements of the meta-models are used to predict the soil categories. The distribution of the soil classes in the dataset was given previously in Figure 3.

4.3.1. Receiver Operating Characteristic Curve (ROC) and Area under Curve (AUC)

To depict how well the meta-models are able to predict the probability of intrinsic compression index belonging to the different soil classes across some decision thresholds, the AUC-ROC are used. ROC plots the true positive rate (TPR) or sensitivity against false positive rate (FPR) or one less specificity given under various thresholds therefore separating "noise" from "signals" (Figure 12). Higher values on the horizontal axis, indicates higher number of the false positives compared to the true negatives. On the other hand, higher values on the vertical axis, indicate a higher number of true positives compared to false negatives. Hence, balancing between false positive and false negatives remains the choice of a chosen threshold. Meanwhile, AUC measures the actual capability of a model to differentiate between class labels. Hence, the higher the larger of AUC, the better the classifier at being able to distinguish between the positive and negative classes. When AUC equals 1, it therefore means that the model has been able to perfectly distinguish between the classes. However, a zero value would mean that the model has predicted all the positives as negatives and all the negatives as positives. When AUC is between 0.5 and 1, it means there is a higher chance that the model is capable of detecting more numbers of true positives and negatives than false negatives and positives. If AUC is exactly 0.5, then it means that the classifier is not able to differentiate between classes, in which case, it is just predicting constant or random class.

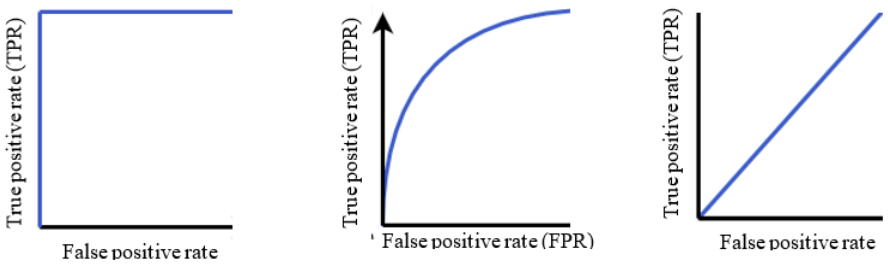

Figure 12. Typical ROC plots.

It could be observed from the ROC (Figures 13 and 14) that the meta-ensembles all demonstrate the ability to discriminate between positive and negative classes. This also means that using the meta-ensembles in the multinomial class prediction has enabled a reduction in both type 1 and 2 prediction errors.

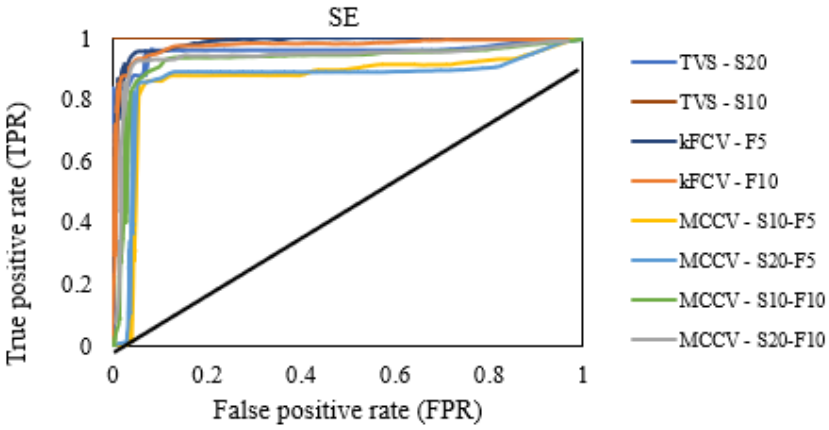

Figure 13. ROC plot for SE model.

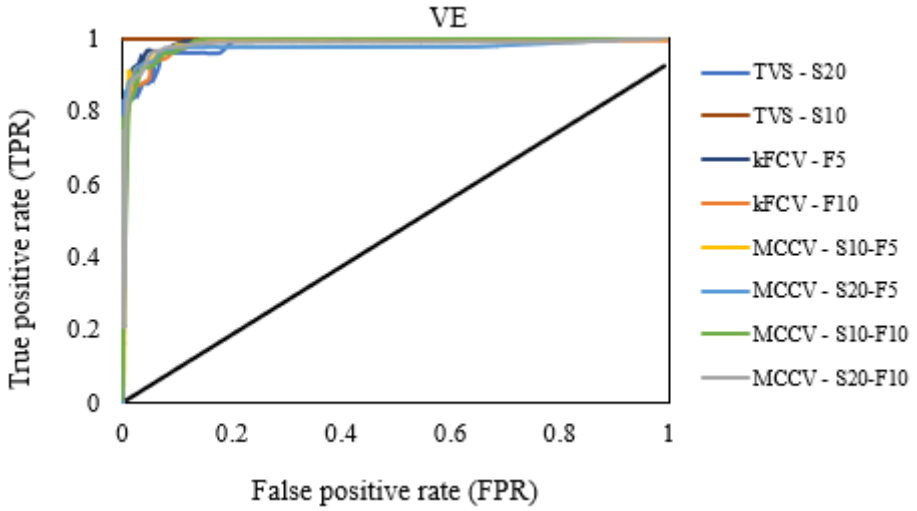

Figure 14. ROC plot for VE model.

The dataset training and testing performed using the TVS appears to show perfect classification with an alignment of ROC with the top horizontal and left vertical axes. This again indicates an overfitting of the dataset given that the application of iterative cross validation techniques (kFCV and MCCV) results in slightly lower sensitivity compared to the TVS method. A closer examination of the corresponding Tables 6 and 7 for both voting and stacking methods of model aggregation indicates that ML training and validation done by utilising the kFCV has the most sensitivity in distinguishing between the soil categories but with the 5-fold validation case having the highest overall AUC.

Table 6. AUC metrics for SE model.

Model	CV Method	Set	No. of Folds	AUC		
				Micro	Macro	Weighted
SE	TVS	20	-	0.957	0.869	0.975
	TVS	10	-	1.000	1.000	1.000
	KFCV	-	5	**0.989**	**0.982**	**0.992**
	KFCV	-	10	**0.982**	**0.986**	**0.992**
	MCCV	10	5	0.861	0.813	0.833
	MCCV	20	5	0.890	0.843	0.911
	MCCV	10	10	0.923	0.887	0.904
	MCCV	20	10	0.946	0.929	0.946

Table 7. AUC metrics for VE model.

Model	CV Method	Set	No. of Folds	AUC		
				Micro	Macro	Weighted
VE	TVS	20	-	0.986	0.986	0.996
	TVS	10	-	1.000	1.000	1.000
	KFCV	-	5	**0.992**	**0.984**	**0.993**
	KFCV	-	10	**0.991**	**0.996**	**0.998**
	MCCV	10	5	0.950	0.964	0.973
	MCCV	20	5	0.974	0.969	0.984
	MCCV	10	10	0.966	0.976	0.979
	MCCV	20	10	0.983	0.982	0.990

Further analysis of sensitivity as given by Table 8 indicates that the kFCV method has the best accuracy, recall rate and precision. When compared with the TVS method (with 20% of the dataset) originally used for testing of the algorithm, it could be observed that when using 5-fold cross validation multiclass predictions are higher by approximately 10–25% in terms of accuracy, rate of recall and precision. It is also interesting to observe that the kFCV method does also performs slightly better than MCCV on a multiclass problem such as that considered in this research. In this case, the disadvantage of using MCCV seems to weigh more heavily on multiclass prediction than on non-linear regression.

Table 8. ML multinomial classification performance metrics.

Model	CV Method	Set (S)	No. of Folds (F)	Accuracy		Recall			Precision		
				Overall	Weighted	Micro	Macro	Weighted	Micro	Macro	Weighted
SE	TVS	20	-	0.833	0.938	0.833	0.563	0.833	0.833	0.575	0.775
	TVS	10	-	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	KFCV	-	5	**0.924**	**0.972**	**0.924**	**0.758**	**0.924**	**0.924**	**0.787**	**0.908**
	KFCV	-	10	**0.933**	**0.964**	**0.933**	**0.856**	**0.933**	**0.933**	**0.855**	**0.923**
	MCCV	10	5	0.817	0.856	0.817	0.756	0.817	0.817	0.734	0.776
	MCCV	20	5	0.858	0.911	0.858	0.746	0.858	0.858	0.745	0.809
	MCCV	10	10	0.858	0.893	0.858	0.805	0.858	0.858	0.765	0.820
	MCCV	20	10	**0.908**	**0.946**	**0.908**	**0.806**	**0.908**	**0.908**	**0.803**	**0.875**
VE	TVS	20	-	0.875	0.957	0.875	0.625	0.875	0.875	0.683	0.850
	TVS	10	-	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	KFCV	-	5	**0.958**	**0.982**	**0.958**	**0.856**	**0.958**	**0.958**	**0.865**	**0.946**
	KFCV	-	10	**0.941**	**0.975**	**0.941**	**0.861**	**0.941**	**0.941**	**0.859**	**0.928**
	MCCV	10	5	0.917	0.936	0.917	0.865	0.917	0.917	0.883	0.922
	MCCV	20	5	**0.925**	**0.969**	**0.925**	**0.796**	**0.925**	**0.925**	**0.803**	**0.901**
	MCCV	10	10	0.900	0.935	0.900	0.824	0.900	0.900	0.816	0.877
	MCCV	20	10	**0.921**	**0.964**	**0.921**	**0.781**	**0.921**	**0.921**	**0.778**	**0.897**

Table 8 also indicate that the assignment of uniform weights to cater for any underrepresented class instances does clearly improves the individual multiclass accuracy compared to the overall accuracy metric. Further, it is also worth noting that the averaging of each class instances (i.e., micro-averaging), presents higher scores compared to the averaging that considers all equal class instances (i.e., macro-averaging) with respect to the most frequently occurring labels. Overall, by comparing the precision scores and recall rates, using micro-averaging does provide a higher sensitivity and thus better performance. However, it is needful to state that the application of the micro-averaging technique should be approached with care because unlike macro-averaging, it may not be suitable when dealing with an imbalanced class distribution given it does not average over larger group or class instances.

Generally, within the method of model aggregation used, it could also be observed that the voting method (VE) seems to generally perform better than its stacking counterpart across the validation methods used.

4.3.2. Comparing between Meta-Heuristic Ensembles and Traditional Classifiers

The superior performance of the meta-ensemble models (VE and SE) on the multiclass prediction of intrinsic compressibility of soils can be further validated by comparing them to other traditional ML multinomial classifiers namely, artificial neural networks (ANN), logistic regressor (LR), boosted decision trees BDT) and random decision forest (RDF). The sensitivity scores of the multiclass classification given in Table 9 indicates that

both the stand-alone and tree-ensemble traditional ML classifiers do have the capability of distinguishing between soil plasticity categories just like the meta-heuristic ensemble models. Notwithstanding, among the traditional classifiers, the tree ensembles have clearly outperformed the stand-alone algorithms. ML prediction given by ANN is generally the least accurate as could be observed from Table 9. Some of the elements of the structure and architectural make-up of neural networks that do sometimes contribute to its poor performance, especially when applied in ML prediction problems such as those considered in this study, were stated previously in the regression prediction. Nevertheless, it is quite apt to add that the stand-alone models (ANN and LR) are in many respects quite similar given their common roots in statistics. However, the functional form of expression used as indicated in their mathematical equations stated above is what differentiates these classifiers. That which is used in LR is parametric whereas the ANN operates semi- or non-parametric functions. This is a very important distinction because most of the contributions given by parameters of an LR can be sufficiently interpreted whereas, as stated previously in the discussion of ML regression, those of ANN whose contributions are from the weights and biases may not be easily interpreted. On the other hand, the relatively high prediction accuracy of the tree-based classifiers is partly because their behaviour in learning is often suggested by a prior set of rules and hence, they are referred to as 'white-boxes' as compared to the neural networks. Hence, they tend to predict outcomes of classifications by continually splitting their inputs based on a set of criteria which then leads to a maximisation of the separation between the dataset and a decrease in entropy. The tree-based classifiers do have their setbacks hence, an aggregation of models into multiple classifiers carried out by stacking or by voting does further increase the accuracy of prediction as seen in Table 9 with the dataset trained, tested, and validated by the kFCV method.

Table 9. Sensitivity metrics for traditional ML algorithms and meta-ensembles with dataset training and validation done under k-fold CV method.

Model	BDT		RDF		LR		ANN		VE		SE	
	10 CV	**5 CV**	**10 CV**	**5 CV**	**10 CV**	**5 CV**	**10 CV**	**5 CV**	**10 CV**	**5 CV**	**10 CV**	**5 CV**
Accuracy	0.898	0.907	0.898	0.866	0.747	0.731	0.721	0.690	**0.941**	**0.958**	**0.933**	**0.924**
Precision	0.869	0.805	0.836	0.772	0.531	0.523	0.523	0.455	**0.859**	**0.865**	**0.856**	**0.787**
Recall	0.878	0.836	0.842	0.715	0.634	0.570	0.585	0.520	**0.861**	**0.856**	**0.856**	**0.758**

4.4.3. Feature Importance

Indicators of the importance of independent features in ML prediction can also give an insight into the data used in the modelling as well as allow for an improvement of the efficacies of the models adopted. The relative significance and usefulness of the input variables used in the predictions of target features are depicted in Figure 15 for both the regression and multiclass classification ML predictions. The effect of each individual independent variables is herein assessed using the best prediction models. In terms of ML modelling and prediction of the intrinsic compressibility index of soils, it is clearly observed from Figure 15a that the feature of importance in this regard is the ratio of voids at 100 kPa of applied pressure. This is hardly surprising given the direct bearing of this measure quantity on C*c as previously demonstrated in the void index relationship. Among the other variables of indirect importance, the liquid limit including its corresponding void ratio prove to be the next most useful features in the forecast of C*c. It is very imperative to bear in mind that several strong linear relationships exist between the Atterberg limits and C*c as suggested by previously mentioned studies. On the other hand, when used for the first time in the prediction of C*c, the specific gravity or the texture element seems to bear the least influence at least within the remit of this research.

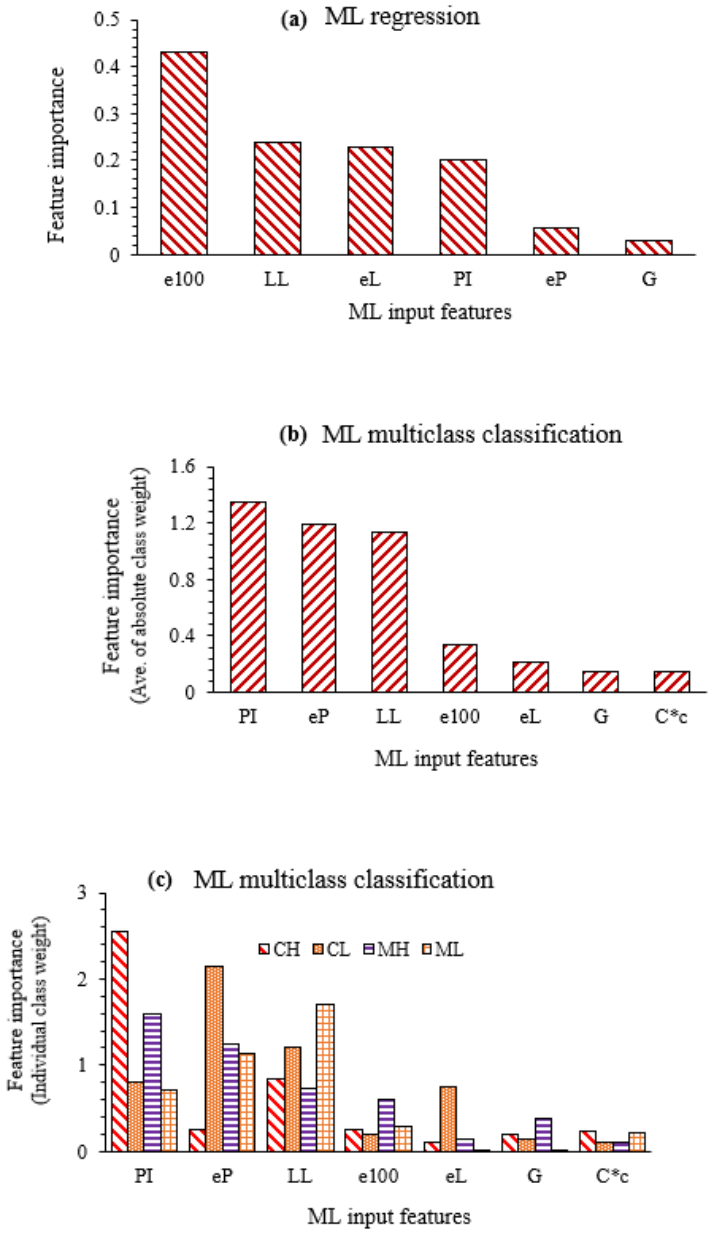

Figure 15. ML Feature importance.

In terms of multiclass classification and when the average of the absolute class weights of the variables are considered, Figure 15b indicates that the Atterberg limits of the soils are the most significant factors in the determination of soil classes with the plasticity index wielding or having the highest numerical importance. This is a confirmation of the established research and theory that uses mostly the soil plasticity index for the direct and indirect determination of soil classes especially if the classification is carried out based on a standard provided by the unified soil classification system (USCS). By virtue of individual class vectors, the Atterberg limits features of the soils do generally seem to carry the most influence on the multiclass prediction of the soil categories as observed in Figure 15c. It is also interesting to note the individual importance of the ratio of voids at PI on the prediction of the relatively least plastic soils when compared to the actual Atterberg limits (both PI and LL). Again, as could be observed, the soil textural characteristics does not seem to be a lot useful in the prediction of the soil classes not to mention the derived features C*c.

5. Study Significance and Implementation

The significance of machine learning to geotechnical engineering design cannot be over- emphasized. The concept of artificial intelligence as applied in this study to predict the intrinsic compressibility of soils can save time and cost during the initial planning and design stages of soil investigation. For example, various tedious laboratory experimentation and time consuming trials that also involve the determination of influencial factors on soil instrinsic compressibility can be circumvented by adopting the methods carrried out in this reserach. However, in order to practically apply the best performing algorithms from this study, it is recommended that the background coding and scripting be persisted and deployed on any organisation's computing resource and training or testing performed on new datasets of intrinsic compressibility of soils.

6. Conclusions

This study has used the machine learning (ML) approach to intelligently model the intrinsic compressibility index of soils by adopting non-parametric tree-based ensemble learners and meta-heuristic ensembles. Predictions using traditional stand-alone algorithms were also performed and compared with those of the ensemble learners. ML multiclass or multinomial classification was also applied to examine the ability of the classifiers to adequately learn between the soil types and by so doing distinguish between positive and negative classes. The following are the main conclusions from this study:

1. Among the traditional stand-alone ML regression models, BLR did produce the least accuracy of prediction of intrinsic compressibility index of soil when compared to REG and ANN. Although, a general consideration of the stand-alone ensembles such as REG and BLR would suggest their inability to implicitly detect the complexities of variable combinations due to their basic underlying assumption of linearity.
2. The tree ensembles (RDF and BDT) did generally outperform the stand-alone models with the RDF model having an R^2 of 0.86 and RMSE of 0.48 and the BDT model producing and an R^2 of 0.822 and RMSE of 0.53. The techniques of 'bagging' or 'bootstrapping' inherent in these tree-ensembles certainly enhanced their accuracy in that respect.
3. The technique of ensemble averaging with a meta-heuristic combination of models' hyperparameters through voting and stacking gave the best overall performance in the prediction of the intrinsic compressibility index of soils with an average R^2 of 0.9 and RMSE of 0.34.
4. Sensitive analysis and diagnostic tests used to examine the procedure and outcome of dataset training, testing and validation showed the Monte Carlo method of cross validation as giving the best results for ML prediction.
5. An analysis of the features with direct influence on the ML prediction revealed that the void ratio determined at effective stress of 100 kPa had the most significance followed by soil's Atterberg limits while specific gravity was the variable of least importance in the forecast of intrinsic compressibility index of soils C*c.

Author Contributions: Conceptualisation, E.E.U. and S.J.A.; methodology, E.E.U. and S.J.A.; validation, E.E.U. and S.J.A.; formal analysis, E.E.U. and S.J.A.; investigation, E.E.U. and S.J.A.; resources, S.J.A. and C.A.B.; data curation, E.E.U. and S.J.A.; writing—original draft preparation, E.E.U. and S.J.A.; writing—review and editing, E.E.U., S.J.A. and C.A.B.; visualisation, S.J.A. and C.A.B.; supervision, C.A.B.; project administration, C.A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gu, X.; Zhang, J.; Huang, X. DEM analysis of monotonic and cyclic behaviors of sand based on critical state soil mechanics framework. *Comput. Geotech.* **2020**, *128*, 103787. https://doi.org/10.1016/j.compgeo.2020.103787.
2. Roscoe, K.H.; Schofield, A.N.; Wroth, C.P. On the Yielding of Soils. *Geotechnique* **1958**, *8*, 22–53. https://doi.org/10.1680/geot.1958.8.1.22.
3. Yang, J.; Li, X.S. State-Dependent Strength of Sands from the Perspective of Unified Modeling. *J. Geotech. Geoenvironmental Eng.* **2004**, *130*, 186–198. https://doi.org/10.1061/(asce)1090-0241(2004)130:2(186).
4. Feng, L.; Zhang, S.; Jin, Z.; Zhang, M.; Sun, P.; Jia, J.; Chu, G.; Hu, W. The critical mechanics of the initiation of loess flow failure and implications for landslides. *Eng. Geol.* **2021**, *288*, 106165. https://doi.org/10.1016/j.enggeo.2021.106165.
5. Nguyen, H.B.K.; Rahman, M.; Fourie, A. The critical state behaviour of granular material in triaxial and direct simple shear condition: A DEM approach. *Comput. Geotech.* **2021**, *138*, 104325. https://doi.org/10.1016/j.compgeo.2021.104325.
6. de Bono, J.P.; McDowell, G.R. Micro mechanics of the critical state line at high stresses. *Comput. Geotech.* **2018**, *98*, 181–188. https://doi.org/10.1016/j.compgeo.2018.02.016.
7. Burland, J.B. On the compressibility and shear strength of natural clays. *Géotechnique* **1990**, *40*, 329–378. https://doi.org/10.1680/geot.1990.40.3.329.
8. Cerato, A.; Lutenegger, A.J. Determining Intrinsic Compressibility of Fine-Grained Soils. *J. Geotech. Geoenvironmental Eng.* **2004**, *130*, 872–877. https://doi.org/10.1061/(asce)1090-0241(2004)130:8(872).
9. Lee, C.; Hong, S.-J.; Kim, D.; Lee, W. Assessment of Compression Index of Busan and Incheon Clays with Sedimentation State. *Mar. Georesources Geotechnol.* **2015**, *33*, 23–32. https://doi.org/10.1080/1064119x.2013.764947.
10. Tsuchida, T. e–logσ_v' relationship for marine clays considering initial water content to evaluate soil structure. *Mar. Georesources Geotechnol.* **2017**, *35*, 104–119. https://doi.org/10.1080/1064119x.2015.1113577.
11. Zeng, L.-L.; Hong, Z.-S.; Cui, Y.-J. Determining the virgin compression lines of reconstituted clays at different initial water contents. *Can. Geotech. J.* **2015**, *52*, 1408–1415. https://doi.org/10.1139/cgj-2014-0172.
12. Al Haj, K.; Standing, J. Mechanical properties of two expansive clay soils from Sudan. *Geotechnique* **2015**, *65*, 258–273. https://doi.org/10.1680/geot.14.p.139.
13. Hong, Z.S.; Zeng, L.L.; Cui, Y.J.; Cai, Y.Q.; Lin, C. Compression Behaviour of Natural and Reconstituted Clays. *Geotechnique* **2012**, *62*, 291–301.
14. Xu, G.-Z.; Yin, J. Compression Behavior of Secondary Clay Minerals at High Initial Water Contents. *Mar. Georesources Geotechnol.* **2016**, *34*, 721–728. https://doi.org/10.1080/1064119x.2015.1080333.
15. Habibbeygi, F. Intrinsic compression characteristics of an expansive clay from western australia. *Int. J. Geomate* **2017**, *12*, 140–147. https://doi.org/10.21660/2017.29.20455.
16. Habibbeygi, F.; Nikraz, H.; Koul, B.K. Regression models for intrinsic constants of reconstituted clays. *Cogent Geosci.* **2018**, *4*, 1546978. https://doi.org/10.1080/23312041.2018.1546978.
17. Kootahi, K.; Moradi, G. Evaluation of compression index of marine fine-grained soils by the use of index tests. *Mar. Georesources Geotechnol.* **2017**, *35*, 548–570. https://doi.org/10.1080/1064119x.2016.1213775.
18. Cerato, A.B. *Influence of Specific Surface Area on Geotechnical Characteristics of Fine-Grained Soils*; University of Massachusetts: Amherst, MA, USA, 2001.
19. Onofrio, A.; Santucci de Magistris, F.; Olivares, L. Influence of Soil Structure on the Behavior of Two Natural Stiff Clays in the Pre-Failure Range. In Proceedings of the 2nd International Symposium on the Geotechnics of Hard Soils-Soft Rocks, Naples, Italy, 12–14 October 1998; pp. 497–505.
20. Yagi, N.; Yatabe, R.; Mukaitani, M. Compressive Characteristics of Cohesive Soils. In Proceedings of the International Symposium on Compression and Consolidation of Clayey Soils, IS-Hiroshima, Japan, 10–12 May 1995; Hiroshi, Y., Osamu, K., eds.; ~~Brookfield: Amsterdam, The Netherlands~~, A.A. Balkema, Rotterdam, 1995, 1995; pp. 221–226.
21. Gori, U.; Polidori, E.; Tonelli, G. The One-Dimensional Consolidation Effect on the Slightly Sensitive Soils. In Proceedings of the International Symposium on Compression and Consolidation of Clayey Soils, IS-Hiroshima, Japan, 10–12 May 1995; pp. 49–55.
22. Yamadera, A.; Miura, N.; Hino, T. Microscopic Consideration on Compression and Rebound Behaviors of Ariake Clay. In Proceedings of the International Symposium, Bothkennar, Drammen, Quebec & Ariake Clays, Yokosuka, Japan, 26–28 February 1997; pp. 175–184.
23. Nagaraj, T.S.; Srinivasa, B.R.; Murthy, S. A critical reappraisal of compression index equations. *Geotechnique* **1986**, *36*, 27–32.
24. Nagaraj, T.S.; Pandian, N.S.; Raju, N. Stress state-permeability fine-grained. *Geotechnique* **1993**, *43*, 333–336.
25. Sridharan, A.; Nagaraj, H.B. Compressibility behaviour of remoulded, fine-grained soils and correlation with index properties. *Can. Geotech. J.* **2000**, *37*, 712–722. https://doi.org/10.1139/t99-128.
26. Meriggi, R.; Paronuzzi, P.; Simeoni, L. Engineering geology characterization of lacustrine overconsolidated clays in an alpine area of Italy. *Can. Geotech. J.* **2000**, *37*, 1241–1251. https://doi.org/10.1139/t00-059.
27. Skempton, A.W.; Jones, O.T. Notes on the compressibility of clays. *Q. J. Geol. Soc.* **1944**, *100*, 119–135. https://doi.org/10.1144/gsl.jgs.1944.100.01-04.08.

28. Youssef, M.S.; Ramli, A.H.; Demery, M.E. Relationships Between Shear Strength, Consolidation, Liquid Limit, and Plastic Limit for Remoulded Clays. In Proceedings of the Sixth International Conference on Soil Mechanics and Foundation Engineering Montreal, QC, Canada, 8–15 September 1965; pp. 126–129.
29. Quigley, R.M.; Thompson, C.D. The Fabric of Anisotropically Consolidated Sensitive Marine Clay. *Can. Geotech. J.* **1966**, *3*, 61–73. https://doi.org/10.1139/t66-008.
30. Akagi, H.; Komiya, K. Constant Rate of Strain Consolidation Properties of Clayey Soil at High Temperature. Volume 1 of Compression and Consolidation of Clayey Soils. In Proceedings of the International Symposium on Compression and Con-solidation of Clayey Soils, IS-Hiroshima, Japan, 10–12 May 1995; pp. 3–8.
31. Raymond, G.P. Laboratory Consolidation of Some Normally Consolidated Soils. *Can. Geotech. J.* **1966**, *3*, 217–234. https://doi.org/10.1139/t66-026.
32. Locat, J.; Lefebure, G. The Compressibility and Sensitivity of an Artificially Sedimented Clay Soil: The Grande-Baleine Marine Clay. In Proceedings of the 2nd Canadian Conference on Marine Geotechnical Engineering, Halifax, Nova Scotia, Canada, 1-3 June 1982,
33. Jardine, R.J. Investigations of Pile-Soil Behaviour, with Special Reference to the Foundations of Offshore Structures. PhD thesis, The University of London, London, UK, 1985.
34. Templeton, G.F. A Two-Step Approach for Transforming Continuous Variables to Normal: Implications and Recommendations for IS Research. *Commun. Assoc. Inf. Syst.* **2011**, *28*, 41–58. https://doi.org/10.17705/1cais.02804.
35. Yu, H.-F.; Huang, F.-L.; Lin, C.-J. Dual coordinate descent methods for logistic regression and maximum entropy models. *Mach. Learn.* **2011**, *85*, 41–75. https://doi.org/10.1007/s10994-010-5221-8.
36. Huang, F.-L.; Hsieh, C.-J.; Chang, K.-W.; Lin, C.-J. Iterative scaling and coordinate descent methods for maximum entropy. *J. Mach. Learn. Res.* **2010**, *11*, 815–848. https://doi.org/10.3115/1667583.1667671.
37. Galwey, N. *Introduction to Mixed Modelling: Beyond Regression and Analysis of Variance*, 2nd ed.; Wiley: Chichester, UK, 2014.
38. Chatterjee, S.; Hadi, A. *Regression Analysis by Example*, 5th ed.; Wiley: Hoboken, NJ, USA, 2012.