

ORIGINAL RESEARCH

Open Access



Classification of geogrid reinforcement in aggregate using machine learning techniques

Samuel Olamide Aregbesola¹ and Yong-Hoon Byun^{1*} 

*Correspondence:
yhbyun@knu.ac.kr

¹ Department of Agricultural
Civil Engineering, Kyungpook
National University, 80 Daehak-ro,
Buk-gu, Daegu 41566, South
Korea

Abstract

The present study proposes a novel ML methodology for differentiating between unstabilized aggregate specimens and those stabilized with triangular and rectangular aperture geogrids. This study utilizes the compiled experimental results obtained from stabilized and unstabilized specimens under repeated loading into a balanced, moderate-sized database. The efficacy of five ML models, including tree-ensemble and single-learning algorithms, in accurately identifying each specimen class was explored. Shapley's additive explanation was used to understand the intricacies of the models and determine global feature importance ranking of the input variables. All the models could identify the unstabilized specimen with an accuracy of at least 0.9. The tree-ensemble models outperformed the single-learning models when all three classes (unstabilized specimens and specimens stabilized by triangular and rectangular aperture geogrids) were considered, with the light gradient boosting machine showing the best performance—an accuracy of 0.94 and an area under the curve score of 0.98. According to Shapley's additive explanation, the resilient modulus and confining pressure were identified as the most important features across all models. Therefore, the proposed ML methodology may be effectively used to determine the type and presence of geogrid reinforcement in aggregates, based on a few aggregate material properties and performance under repeated loading.

Keywords: Aggregate, Classification, Feature importance, Geogrid reinforcement, Machine learning

Introduction

Geogrids play a prominent role in flexible pavement construction, particularly in maintaining subgrade restraint and stabilizing the base course. These applications either reduce the base course thickness or extend the pavement life. When used to build pavement working platforms, geogrids are installed at the base–subgrade interface to enhance the load-bearing capacity of the pavement foundation. For base stabilization, geogrids can be positioned at the center of the base course or at the base–subgrade interface [1–4]. The successful implementation of both methods has afforded reduction in permanent deformation in the base course [5].

Geogrids are widely recognized for their primary function of enhancing horizontal stiffness in mechanically stabilized base layers, a process known as lateral confinement.

Lateral confinement prevents excessive base course deformations and improves the bearing capacity under vehicular loading due to the interlocking between the geogrid and aggregate particles. Various geogrid products, characterized by differences in properties such as aperture shapes and sizes, rib shape, and tensile strength, affect the interlocking mechanisms [6]. Previous studies have reported that optimal geogrid stabilization and interlock can be achieved when the geogrid aperture size matches the aggregate grain size [7, 8]. Tutumluer et al. [9] demonstrated the difference between the shear forces in case of triangular and rectangular geogrid aperture shapes using aggregate image-aided discrete element method simulation. Byun et al. [10] used shear wave transducers to successfully quantify different horizontal stiffness profiles of aggregate specimens enhanced by rectangular and triangular geogrid apertures. However, without coring and sampling in the field, information regarding the geogrid reinforcement type and presence may not be readily available. This can potentially affect the assessment of pavement conditions and selection of appropriate maintenance techniques.

Numerous machine learning (ML) studies in geotechnical engineering focus on supervised learning. In supervised learning, labeled datasets are used to predict or classify outcomes, where most of the related studies target regression problems [11–14]. Nevertheless, some studies have explored classification problems. For instance, Eyo et al. [15] demonstrated that meta-ensemble models could classify three different binder combinations previously used in predicting the unconfined compressive strength of reinforced soils. Ensemble classification algorithms have been used in slope stability analysis to determine landslide susceptibility based on field data [16, 17]. Furthermore, Ozsagir et al. [18] achieved a high prediction accuracy in assessing the liquefaction potential of fine-grained soils using a decision tree model, while Jas and Dodagoudar [19] explored the prediction of the liquefaction potential of soils using extreme gradient boosting-Shapley additive explanations (SHAP). Soil classification based on physical and chemical properties has also been recently investigated [20–23]. However, a major limitation of some of these studies is their insufficient datasets for training ML models as well as the interpretability of the models [19]. Thus, more studies focusing on applying ML classification algorithms and interpreting these models are required.

This study aims to investigate the efficacy of ML classification algorithms in distinguishing between unstabilized aggregate specimens and those stabilized with triangular and rectangular aperture geogrids. To resolve the problem of data insufficiency noted in previous studies, a relatively moderate dataset with 4500 experimental observations is employed herein to train five ML models, including single-learning and tree-ensemble algorithms. The performances of these models are compared using appropriate ML classification metrics. Furthermore, the global feature importance across all algorithms is examined to enhance the interpretability of the applied ML methods. Finally, the engineering implications of the current study are discussed.

Methodology

Database generation

The database comprises numerical data obtained from repeated-load triaxial experimental testing for the resilient modulus characterization of crushed aggregates. The crushed aggregates were obtained from a quarry located in North Carolina. Sieve

analysis results indicate that the material was well-graded, with a mean diameter of 8.4 mm and a maximum particle size of 25 mm. A detailed description of the indexes and compaction properties of the aggregates can be found in previous studies [10, 24]. The experimental program utilized an unstabilized specimen and two specimens stabilized with triangular and rectangular aperture geogrids. The two geogrids were designed for base course stabilization in flexible pavements, and the dimensions are also detailed in previous studies [10, 24]. Five confining pressures, 100 load repetitions of varying deviatoric stresses, and 15 sequences were considered during the experimental program, following the procedure outlined in AASHTO T307 [25].

Figure 1a illustrates the typical haversine pulse obtained from each loading cycle of the repeated-load triaxial test, representing a standard wheel loading involving a 100 ms loading period and a 900 ms rest period. Furthermore, recoverable and plastic strains were recorded for each cycle and subsequently used to calculate the resilient modulus and permanent deformation. Herein, the accumulated permanent strain (ϵ_p) (Fig. 1b) was considered as an input feature alongside the resilient modulus (M_R), number of cycles (N), confining pressure (σ_c), and deviatoric stress (σ_d). Notably, previous studies demonstrated that the deviatoric stress and number of cycles are major factors influencing the accumulated permanent strain [26, 27]. Finally, each datum

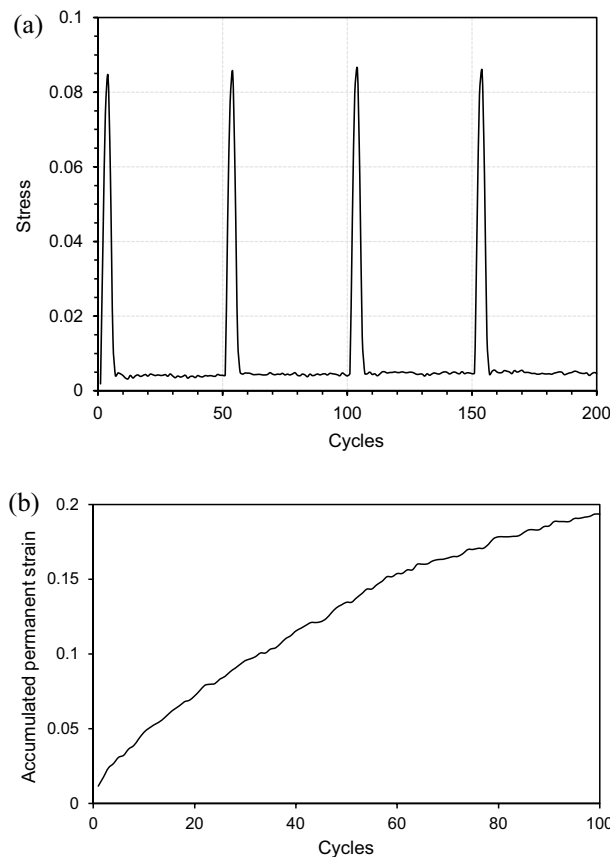


Fig. 1 Typical responses of repeated-load triaxial tests: **a** axial stress at different cycles and **(b)** accumulated permanent strain for unstabilized specimen (ninth sequence)

was assigned to one of three classes: rectangular geogrid, triangular geogrid, and unstabilized.

Data preprocessing and exploratory data analysis

Data preprocessing is an important step that allows the examination and assurance of database integrity. Herein, no missing values or duplicate data points were found. Upon further investigation, the database was found to comprise 4500 observations, with the statistics summarized in Table 1. All classes were equally represented in the database used for this study, with 1500 data points for each class. This equal distribution was due to the consistent application of N , σ_c , and σ_d for each specimen, resulting in varied M_R and ε_z values. A common issue with multiclass ML models is class imbalance, often leading to biased models [28, 29]. The class balance in this study helps avoid this problem and aids reliable performance evaluation.

The distribution of values for each input feature across the three classes was presented along the diagonal of the plot shown in Fig. 2. The input features N , σ_c , and σ_d exhibited the same distribution across all the classes. However, the unstabilized specimen showed higher M_R values than the triangular and rectangular stabilized specimens (Fig. 3) because the two geogrid-stabilized specimens had slightly lower densities [10]. When examining the relation between M_R and N , overlapping M_R values were observed between the triangular and rectangular specimens. For the relation between ε_z and N , the ε_z for the triangular specimen was higher than those of the other classes. Owing to this general trend of overlapping M_R and ε_z values, linear ML models might face challenges in accurately separating the classes.

The dataset was partitioned into training and testing sets with a ratio of 0.8, as recommended in the literature [30, 31]. Subsequently, the values of the input features must be scaled using a standard scaler so that they have a mean of 0 and a standard deviation of 1. This standard scaling can be mathematically represented as

$$z = \frac{x - \mu}{s} \quad (1)$$

where x is the feature value, μ is the mean of the feature value, and s is the corresponding standard deviation.

Table 1 Relevant statistics of independent features in the database

Variable	Min	Max	Mean
N	1.00	100.00	50.50
σ_c (kPa)	20.70	137.90	73.08
σ_d (kPa)	20.70	275.80	103.40
M_R (MPa)	75.57	117.26	85.30
ε_z (%)	0.00	0.30	0.05

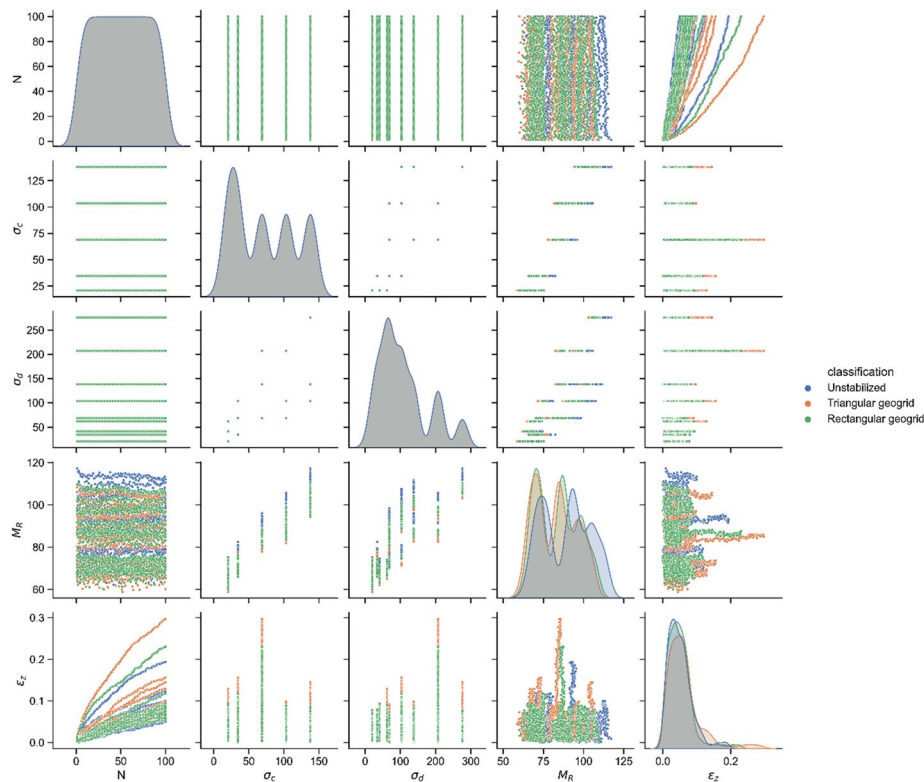


Fig. 2 Pairwise relation between variables in the dataset

Machine learning algorithms

Logistic regression

Logistic regression (LR) employs a logistic function to establish the relation between a binary target variable and a set of predictor variables [22]. Given a binary target variable represented as Y and a vector of predictor variables denoted as $X = (X_1, X_2, \dots, X_p)$, LR formulates the probability of Y equaling 1 as follows:

$$P(Y = 1|X) = \frac{1}{1 + e^{-Z}} \tag{2}$$

where Z is the linear predictor defined by $Z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$, and $\beta_0, \beta_1, \beta_2, \dots, \beta_p$ are the regression coefficients.

The key assumptions of LR include the absence of influential outliers and multicollinearity, the independence of errors, and linearity in the logit. The regression coefficients are typically estimated using maximum likelihood estimation techniques, such as iterative reweighted least squares and the Newton–Raphson method. Although LR is well suited for binary classification problems, a one-vs.-rest (OvR) approach can be applied to multiclass problems. The OvR approach involves training a separate binary LR model for each class, treating that class as positive while all other classes as negative.

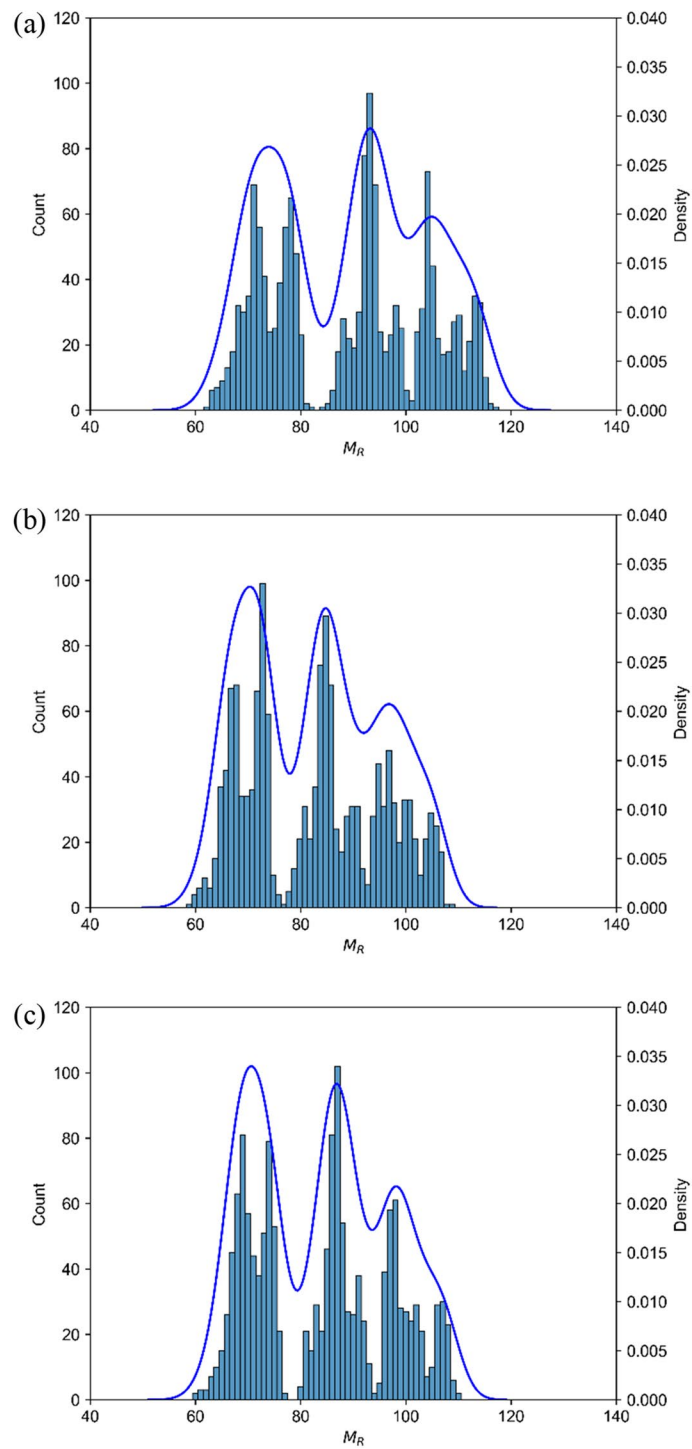


Fig. 3 Resilient modulus distribution for all classes: (a) unstabilized, (b) triangular geogrid, and (c) rectangular geogrid

Support vector machines

Support vector machines (SVMs) are used to find an optimal hyperplane in a high-dimensional feature space that maximizes the separation of data points between different classes. This is achieved using a kernel function, such as a linear, polynomial, or radial basis function (RBF), suitable for solving complex nonlinear problems [32, 33]. Using the RBF, the SVM can be expressed as

$$f(X_1, X_2) = \exp(-\gamma \|X_1 - X_2\|^2) \quad (3)$$

where X_1 and X_2 represent the input vectors and γ defines the influence of a single training point on nearby data points.

SVMs aim to identify support vectors, which are the data points closest to the decision boundary. These support vectors are crucial for defining the margin used to optimize the hyperplane. Generally, SVMs effectively handle high-dimensional data and are less susceptible to overfitting. However, they can be sensitive to hyperparameter tuning and may become computationally expensive when handling large datasets.

Tree-ensemble models

This study considered three tree-based/ensemble algorithms: random forest (RF), extra trees classifier (ETC), and light gradient boosting machine (LGBM). RF creates decision trees by training each tree using a random subset of data and features. The outcome is determined by combining the predictions from all individual trees [34]. Meanwhile, ETC shares similarities with RF but introduces additional randomness in the tree-building process. Random thresholds are used instead of identifying the optimal split of features based on a quality criterion, such as Gini impurity or entropy [35]. LGBM employs a unique method called gradient-based one-side sampling, which reduces the number of data instances required to build each tree while maintaining accuracy. This method considerably reduces the training time without compromising performance.

Generally, tree-ensemble methods combine multiple decision trees to create a robust and accurate predictive model. The primary difference between the methods lies in the way the decision trees are constructed. These methods are known for their interpretability based on embedded feature importance, ability to generalize when using new datasets, and computational efficiency, particularly concerning the tree-ensemble methods selected in this study. The fundamental equation for tree-ensemble methods is given by

$$F(x) = \sum_{m=1}^M \gamma_m h_m(x) \quad (4)$$

where $F(x)$ is the prediction, γ_m denotes the weight assigned to the m th tree, and $h_m(x)$ represents the prediction from the m th tree for input x .

Machine learning implementation

The ML process typically comprises training, validation, and testing phases. During training, the model attempts to learn about the patterns and correlations that exist in the data. Optimization algorithms are used during model training to reduce errors. In addition, each ML algorithm includes hyperparameters that can be combined to improve

model performance. Techniques, such as grid search, help to assess the different model hyperparameters and choose the best combination using the validation set [36]. Finally, the model is evaluated on the test set using suitable predefined metrics. Model evaluation helps to assess the generalization of the model and determine its performance on unseen data.

Herein, ML model training and validation were conducted using Python programming language. After separating the datasets into training and testing sets, the input features were scaled for both sets. A parameter grid containing a range of hyperparameters for each model was defined. Only significant hyperparameters were considered to expedite the training process. A grid search algorithm [36] was used to identify the optimal combination of hyperparameters for each model, as outlined in Table 2. Subsequently, each model was trained and tested using the metrics defined in Sub Sect. “Performance evaluation metrics”. The *k*-fold cross-validation method was also adopted in this study. This method divides the entire training set into *k* subsets or folds. The model was then trained using all but one of the subsets and validated using the remaining subset. This process was repeated iteratively *k* times. A recommended tenfold cross-validation was used in this study [37, 38].

Performance evaluation metrics

Herein, the capability of each model to identify and correctly classify specimens was evaluated using the metrics detailed below. The first metric is accuracy, which is the ratio of correct predictions to the total number of predictions, mathematically expressed as

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

where *TP* refers to the true positive, i.e., the model correctly predicts a positive class; an example is when a model correctly classifies an unstabilized specimen as unstabilized.

Table 2 Selected hyperparameters for machine learning models

Model	Hyperparameters	Description	Value
Random forest (RF)	max_depth	max_depth	None
	min_samples_split	Minimum number of samples required to split a node in the tree	2
	n_estimators	Minimum number of samples required to split a node in the tree	300
Support vector machine (SVM)	C	Penalty for misclassifications	10
	kernel	Type of kernel	rbf
Logistic regression (LR)	C	Penalty for misclassifications	1
	solver	Algorithm used to train model	newton-cg
Light gradient boosting machine (LGBM)	n_estimators	Number of trees	500
	max_depth	Maximum depth of trees	20
Extra trees classifier (ETC)	max_depth	Maximum depth of trees	20
	min_samples_leaf	Number of samples required to be in a leaf node	1
	min_samples_split	Minimum number of samples required to split a node in the tree	2
	n_estimators	Number of trees	500

Conversely, false positives (*FPs*) are instances where the model incorrectly predicts a negative class as positive. In this case, an example would be classifying the unstabilized specimen as rectangular. True negatives (*TNs*) occur when the model correctly identifies the negative class. An example is when a model correctly classifies an unstabilized specimen as not being a specimen stabilized by triangular or rectangular aperture geogrid. Finally, false negatives (*FNs*) arise when the model inaccurately predicts a positive class as negative.

Notably, in the case of unbalanced classes, relying solely on accuracy to assess the performance of a model can be misleading. The next metric is thus precision, which gauges the *FP* rate of a model. A higher precision score implies a lower rate of *FPs* and vice versa. Precision should be computed as

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

Recall is another important metric that evaluates the *FN* rate of a model. A higher recall value indicates a low *FN* rate. Recall is calculated as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

Finally, the F1-score is the harmonic mean of precision and recall, balancing these two metrics. The F1-score is mathematically expressed as

$$F1 - score = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (8)$$

Receiver operating characteristic curve

An effective method for visualizing the performance of ML models is using an “area under the curve–receiver operating characteristics” (AUC–ROC) curve. In this study, the AUC–ROC curve is used to assess the discriminatory capacity of the models. Discriminatory capacity, as the term implies, is the ability of a model to accurately distinguish between classes. The ROC curve represents the probability distribution of the outcomes, while the AUC quantifies the degree of class separability. More specifically, the ROC curve plots the *TP* rate (TPR) against the *FP* rate (FPR), with the TPR on the *y*-axis and the FPR on the *x*-axis. Typically, the AUC values fall within the range of 0–1. High AUC values, particularly those approaching 1, are desirable as they reflect the capacity of a model to accurately predict classes. By contrast, models with an AUC close to 0 exhibit poor class separation. In the case of an AUC of 0.5, the model cannot discriminate between the classes and essentially resorts to random guessing.

Shapley additive explanations

ML models are often considered black-box models because of their complex nature. Explainable artificial intelligence (XAI) has been developed to bridge the gap between the intricacies of artificial intelligence algorithms and human comprehension. An essential component of XAI is SHAP, which employs game theory to compute Shapley values [39]. In this study, SHAP feature importance was used to determine the rank of features

across all models. The global feature importance was obtained by calculating the arithmetic mean of the absolute Shapley values for each feature. The features with high absolute Shapley values were deemed the most important. Notably, the kernel SHAP and tree SHAP [40] were used to determine the importance of the single-learning and tree-based models, respectively. Generally, the feature importance is given by

$$I_j = \frac{1}{n} \sum_{i=1}^n |\phi_j^{(i)}| \tag{9}$$

where n represents the total number of instances or samples in the dataset and ϕ_j denotes the Shapley value of an observation j , given by

$$\phi_j = \sum_{S \subseteq \{1,2,\dots,p\} \setminus \{j\}} \frac{|S|!(p - |S| - 1)!}{p!} [f(x_S \cup \{j\}) - f(x_S)] \tag{10}$$

where S denotes a subset of features excluding j , $|S|$ represents the number of features in the subset S , and p denotes the total number of features. $f(x_S \cup \{j\})$ is the prediction output when j is included in S , while $f(x_S)$ is the prediction output when j is excluded from S .

Results and discussion

Performance evaluation of ML models

The performances of the multiclassification ML algorithms used to predict the geogrid stabilization category are plotted in Fig. 4. The tree-based learning models (LGBM, ETC, and RF) performed better than the single-learning models (LR and SVM). LR exhibited the worst performance among all the models, with an average precision, recall, and F1-score of 0.71, 0.70, and 0.70, respectively. The major reason for the poor performance of the LR model might be the assumption of a linear relation between the input variables and the log-odds of the target variable by the algorithm [41]. Feature engineering, such as creating interaction terms, could potentially enhance the performance of the LR model. However, this might increase the risk of overfitting and degrade performance when using new datasets. Meanwhile, the SVM model performed considerably better than the LR model, with a precision, recall, and F1-score of 0.85. This improved

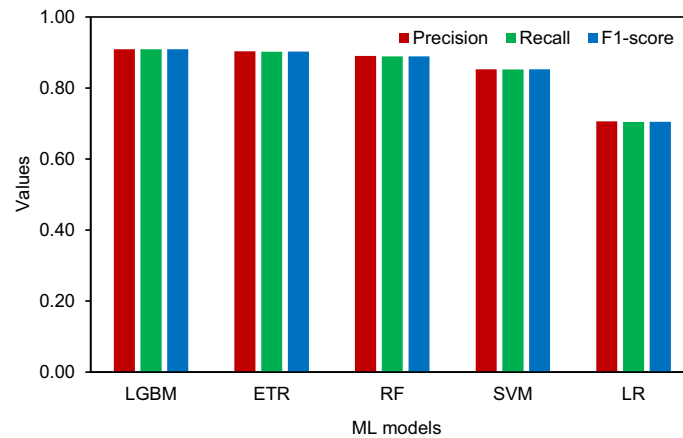


Fig. 4 Average precision, recall and F1-score for the five models

performance could be attributed to the ability of SVM to transform the initial feature space into a higher dimension where classes are more easily separated using a suitable kernel function. The radial basis function kernel used in this study was especially effective in handling nonlinear and complex multiclass decision boundaries [32]. Among the five ML algorithms implemented in this study, the LGBM model showed the best performance, with a precision, recall, and F1-score of 0.91. The RF and ETC models also exhibited good performances, with precision, recall, and F1-scores of 0.89 and 0.90, respectively. Notably, the three tree-based models did not assume a linear relation between features and targets, leading to an improved performance. Furthermore, tree-based models combine the prediction of multiple base models to make a final prediction. By aggregating predictions, tree-based models tend to produce more robust and generalized models. Finally, the tree-based models have been shown to outperform even deep-learning models when using tabular datasets with < 10,000 training examples [42].

Prediction uncertainty

The confusion matrix was used to illustrate the prediction uncertainties for each multiclass model (Fig. 5). Overall, the unstabilized specimens were better classified than the specimens stabilized by triangular and rectangular aperture geogrids. Each model correctly identified unstabilized specimens at least 90% of the time, with the LGBM model making accurate predictions in 94% of such instances. This could be attributed to the fact that unstabilized specimens possessed higher and more distinguishable M_R than the specimens stabilized by triangular and rectangular aperture geogrids, which exhibited overlapping M_R values. Note that a previous study demonstrated that the M_R obtained from repeated-load triaxial tests could not discern the effect of the geogrid on the stabilization of aggregates [10]. The performance of the LR model was considerably poor when predicting specimens stabilized by triangular and rectangular aperture geogrids, correctly predicting the specimens only 59 and 62% of the time, respectively. The reasons for the poor performance were explored in Sect. “Performance evaluation of ML models”. The SVM model accurately predicted the specimens stabilized by triangular and rectangular aperture geogrids ~80% of the time. By contrast, the tree-based models exhibited less uncertainty than the single-learning models, making accurate predictions at least 85% of the time for RF and 88% of the time for both LGBM and ETC.

Sensitivity analysis

The average AUC and accuracy values of each model across all classes are summarized in Table 3. The models were found to demonstrate a strong capacity to distinguish positives from negatives. Both the SVM and tree-based models presented AUC values of > 0.96, while the LR model exhibited an AUC value of 0.87. However, this relatively high value for the LR model was considerably affected by the predictions of the unstabilized class. Figure 6 shows the ROC curves for the LGBM model considering each classification. The unstabilized specimens and the specimens stabilized by triangular and rectangular aperture geogrids exhibit AUC values of 0.99, 0.98, and 0.97, respectively. These results emphasize the potential of applying these algorithms to geotechnical classification problems.

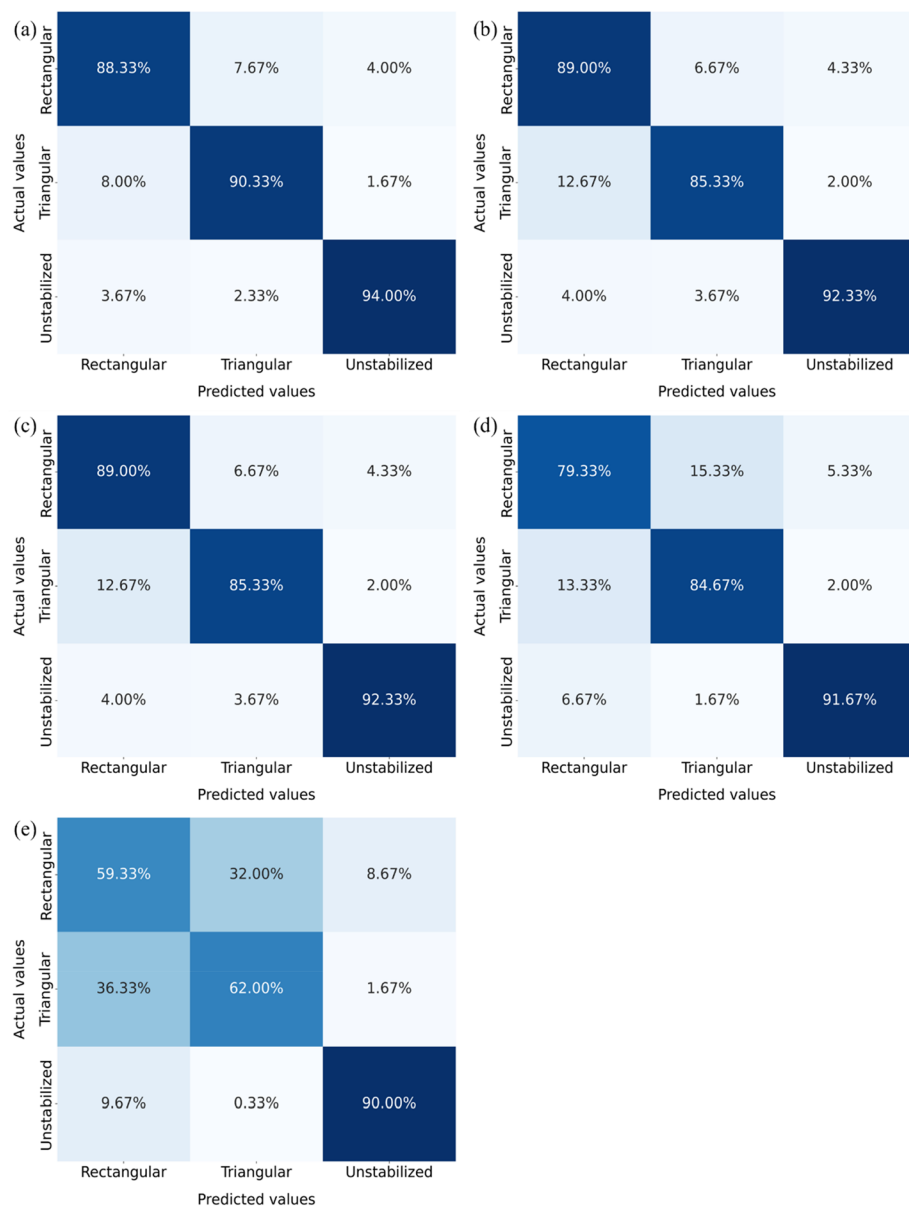


Fig. 5 Confusion matrix for (a) light gradient boosting (b) extra trees classifier (c) random forest (d) support vector machine, and (e) logistic regression

Table 3 Performance of five models on test dataset

Model	Area under the curve (AUC) value	Accuracy
LGBM	0.98	0.91
ETC	0.98	0.90
RF	0.97	0.89
SVM	0.96	0.85
LR	0.87	0.70

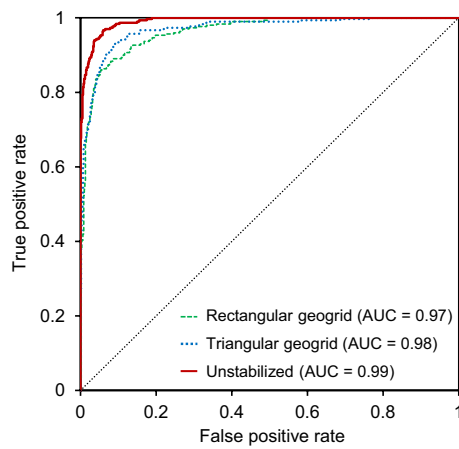


Fig. 6 Receiver operating characteristic curve for light gradient boosting machine

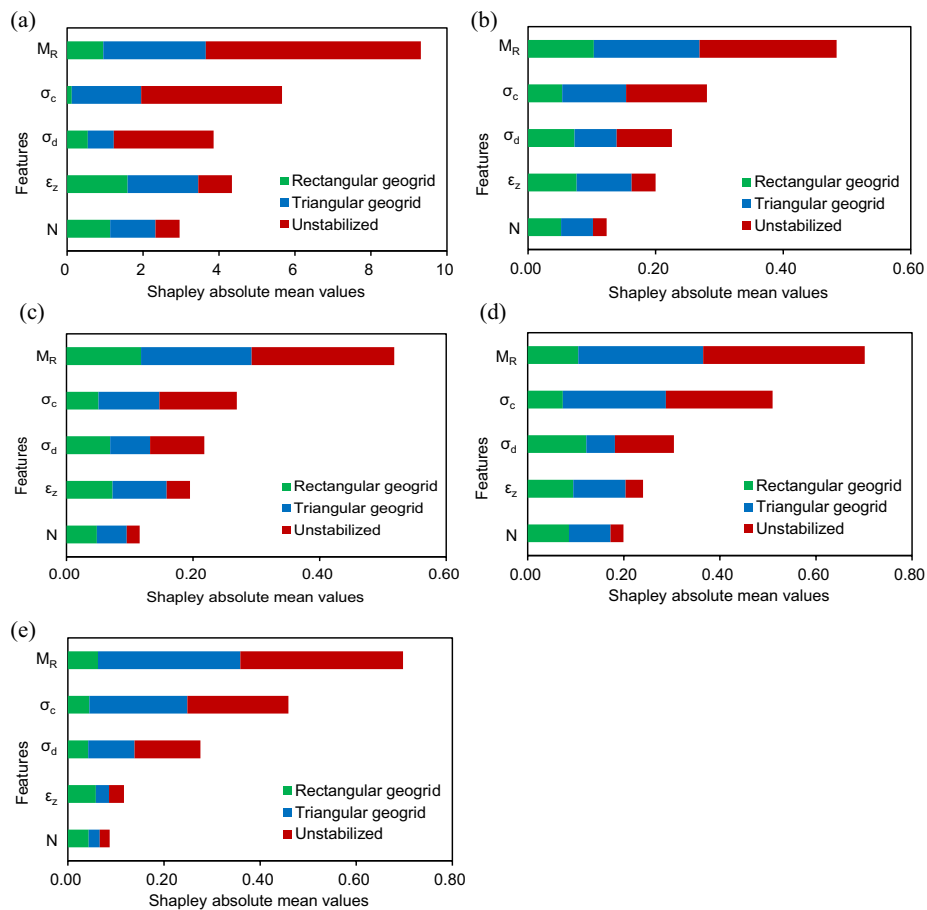


Fig. 7 Shapley multioutput absolute mean values (a) light gradient boosting, (b) extra trees classifier, (c) random forest, (d) support vector machine, and (e) logistic regression

Feature importance

The Shapley multiclass output of each model using mean absolute values is illustrated in Fig. 7. Herein, 100 simulations were selected for each ML model to expedite the feature

importance algorithm and ensure consistent ranking of the models. First, the mean absolute Shapley values for the LGBM model were found to be considerably higher than those for the ETC, RF, SVC, and LR models. This indicates a wider value range for the LGBM model compared with the narrow range displayed by the other models. In practice, the wider range of mean absolute Shapley values indicates that the LGBM model is better at distinguishing between important and unimportant features for making predictions. M_R emerged as the most important feature across all the models. The final rank of feature importance is presented in Table 4, where σ_c , σ_d , ϵ_z , and N complete the ranking in the specified order. For the LGBM model, ϵ_z and σ_d emerged as the third- and fourth-ranked features, respectively, whereas the ETC, RF, and single-learning models showed a ranking pattern similar to the final ranking. Generally, the feature ranking for all the models was remarkably consistent and almost identical. Figure 7 also reveals the contribution of each feature to the prediction of each class across all the models. For each model, M_R considerably contributes in predicting the class of the unstabilized specimen, which could be attributed to the higher and easily separable M_R values for the unstabilized specimen. In addition, the specimen stabilized by the triangular aperture geogrid was effectively explained by M_R . Furthermore, σ_c predicted unstabilized specimens more than the specimens stabilized by triangular and rectangular aperture geogrids. However, for the specimens stabilized by rectangular geogrid, ϵ_z and σ_d considerably contributed to predicting the class.

Engineering implications

Unbound aggregate base layers play a crucial role in pavement systems, providing structural support, distributing loads to the subgrade, and facilitating drainage. The resilient and permanent deformation behaviors of the base course are the two key aspects of the pavement design and performance analysis. Accordingly, the resilient modulus and rutting, representing recoverable and permanent deformation under repeated traffic loading, respectively, were evaluated. The resilient modulus and permanent deformation can be influenced by various factors, such as deviatoric stress and the number of load cycles. For pavement rehabilitation, the resilient modulus and permanent deformation can also be assessed using nondestructive field-testing methods, such as using falling weight deflectometers. These deflectometers are typically used to estimate elastic-layer moduli based on the impulse load and measured deflection [43]. However, detecting geogrid insertion into the base course can be challenging when using a specific nondestructive field-testing method because of the lesser

Table 4 Feature importance rank using Shapley absolute mean values

Model	N	σ_c	σ_d	M_R	ϵ_z
LGBM	5	2	4	1	3
ETC	5	2	4	1	3
RF	5	2	3	1	4
SVM	5	2	3	1	4
LR	5	2	3	1	4
Rank	5	2	3	1	4

thickness of the geogrid than that of the base course. Furthermore, identifying the geogrid type without coring and sampling can be quite difficult in the context of pavement rehabilitation. However, using the five features, either measured from field tests or computed from numerical analyses, the ML technique can identify the type and presence of geogrid reinforcement in aggregates. Therefore, the ML classification of geogrid-stabilized aggregates proposed in this study may be a promising approach for assessing pavement conditions and help further selecting adequate maintenance techniques.

Conclusions

This study proposed a novel ML technique to distinguish between unstabilized aggregate specimens and the specimens stabilized with triangular and rectangular aperture geogrids. Five ML models were employed using an experimental database, comprising five input variables and 4500 data points equally distributed among the three classes. In addition, the feature importance across all the models was examined to improve the interpretability of the ML models. All models could correctly identify the unstabilized specimen with a minimum accuracy of 0.90. This result could be attributed to the higher, more distinguishable resilient modulus values of the unstabilized specimen than those of the stabilized specimens. The single-learning algorithms, especially the LR model, produced less accurate predictions than the tree-ensemble algorithms for the geogrid-reinforced specimens. Generally, the LGBM model exhibited the highest accuracy, with an overall score of 0.91, which could be attributed to the robustness of the algorithm and its high nonlinear mapping capabilities. The results from the sensitivity analysis showed that the SVM and tree-based models correctly identified the positive class, with a minimum AUC value of 0.96. The LGBM model exhibited the highest AUC value of 0.98, while LR showed the lowest AUC value (0.87), considerably affected by the accuracy scores obtained from the prediction of the unstabilized specimen. Eventually, the resilient modulus was identified as the most important variable in classification, followed by confining pressure, deviatoric stress, permanent strain, and the number of cycles in that order. The ML approach proposed in this study may efficiently assess pavement conditions by providing information regarding the type and presence of geogrid reinforcement in aggregates without coring and sampling.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. NRF-2020R1C1C1008925; NRF-2021R1A5A1032433). The authors also would like to acknowledge Prof. Erol Tutumluer of the University of Illinois Urbana-Champaign for his initial support in the laboratory experiments.

Author Contributions

SOA was involved in the acquisition of data and images, drafting of the manuscript, and providing the revised content of the article. YB contributed by defining the area of study, acquiring data, drafting the manuscript, revising the article content, and giving final approval of the version to be submitted. All authors have read and approved the final manuscript.

Data Availability

Data may be made available on request.

Declarations

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

Received: 16 August 2023 Accepted: 22 January 2024

Published online: 12 February 2024

References

1. Mekkawy MM, White DJ, Suleiman MT, Jahren CT (2011) Mechanically reinforced granular shoulders on soft sub-grade: laboratory and full scale studies. *Geotext Geomembr* 29:149–160
2. Sun X, Han J, Kwon J, Parsons RL, Wayne MH (2015) Radial stresses and resilient deformations of geogrid-stabilized unpaved roads under cyclic plate loading tests. *Geotext Geomembr* 43:440–449
3. Abu-Farsakh M, Hanandeh S, Mohammad L, Chen Q (2016) Performance of geosynthetic reinforced/stabilized paved roads built over soft soil under cyclic plate loads. *Geotext Geomembr* 44:845–853
4. Suku L, Prabhu SS, Sivakumar Babu GL (2017) Effect of geogrid-reinforcement in granular bases under repeated loading. *Geotext Geomembr* 45:377–389
5. Moghaddas-Nejad F, Small JC (1996) Effect of geogrid reinforcement in model track tests on pavements. *J Transp Eng* 122:468–474
6. Giroud JP, Han J (2004) Design method for geogrid-reinforced unpaved roads. I. Development of design method. *J Geotechn Geoenviron Eng* 130:775–786
7. Holtz RD, Christopher BR, Berg RR (2008) Geosynthetics design & construction guidelines—reference manual, FHWA-HNI-07-092. United States. National highway institution. Federal highway administration. U.S. Department of Transportation, Washington, D.C.
8. M. Mulabdić, K. Minažek, J. Kaluder, Geogrids - what is important, 5th International Conference on Road and Rail Infrastructure. 5 (2019) 215–221.
9. Tutumluer E, Huang H, Bian X (2012) Geogrid-aggregate interlock mechanism investigated through aggregate imaging-based discrete element modeling approach. *Int J Geomech* 12:391–398
10. Byun YH, Tutumluer E, Feng B, Kim JH, Wayne MH (2019) Horizontal stiffness evaluation of geogrid-stabilized aggregate using shear wave transducers. *Geotext Geomembr* 47:177–186
11. Feng J, Zhang H, Gao K, Liao Y, Gao W, Wu G (2022) Efficient creep prediction of recycled aggregate concrete via machine learning algorithms. *Constr Build Mater* 360:129497
12. Chen Y, Li F, Zhou S, Zhang X, Zhang S, Zhang Q, Su Y (2023) Bayesian optimization based random forest and extreme gradient boosting for the pavement density prediction in GPR detection. *Constr Build Mater* 387:131564
13. Javad Azarhoosh M, Koohmishi M (2023) Prediction of hydraulic conductivity of porous granular media by establishment of random forest algorithm. *Constr Build Mater* 366:130065
14. Hilloulin B, Tran VQ (2023) Interpretable machine learning model for autogenous shrinkage prediction of low-carbon cementitious materials. *Constr Build Mater* 396:132343
15. Eyo EU, Abbey SJ (2021) Machine learning regression and classification algorithms utilised for strength prediction of OPC/by-product materials improved soils. *Constr Build Mater* 284:122817
16. Pyakurel A, Dahal BK, Gautam D (2023) Does machine learning adequately predict earthquake induced landslides? *Soil Dyn Earthq Eng* 171:107994
17. Wengang Z, Hanlong L, Lin W, Xing Z, Yanmei Z (2023) Prediction of slope stability using ensemble learning techniques. In: Wengang Zhang, Hanlong Liu, Lin Wang, Xing Zhu, Yanmei Zhang (eds) *Application of machine learning in slope stability assessment*. Springer Nature Singapore, Singapore
18. Ozsagir M, Erden C, Bol E, Sert S, Özocak A (2022) Machine learning approaches for prediction of fine-grained soils liquefaction. *Comput Geotech* 152:105014
19. Jas K, Dodagoudar GR (2023) Explainable machine learning model for liquefaction potential assessment of soils using XGBoost-SHAP. *Soil Dyn Earthq Eng* 165:107662
20. Kim Y, Yun TS (2021) How to classify sand types: a deep learning approach. *Eng Geol* 288:106142
21. Pham BT, Nguyen MD, Nguyen-Thoi T, Ho LS, Koopialipour M, Kim Quoc N, Armaghani DJ, Van Le H (2021) A novel approach for classification of soils based on laboratory tests using adaboost, tree and ANN modeling. *Transp Geotechn* 27:100508
22. Eyo E, Abbey S (2022) Multiclass stand-alone and ensemble machine learning algorithms utilised to classify soils based on their physico-chemical characteristics. *J Rock Mech Geotechn Eng* 14:603–615
23. Li L, Iskander M (2022) Use of machine learning for classification of sand particles. *Acta Geotech* 17:4739–4759
24. Byun YH, Tutumluer E (2019) Local stiffness characteristic of geogrid-stabilized aggregate in relation to accumulated permanent deformation behavior. *Geotext Geomembr* 47:402–407
25. AASHTO T307, Standard method of test for determining the resilient modulus of soils and aggregate materials, (1999).
26. Gu C, Zhan Y, Wang J, Cai Y, Cao Z, Zhang Q (2020) Resilient and permanent deformation of unsaturated unbound granular materials under cyclic loading by the large-scale triaxial tests. *Acta Geotech* 15:3343–3356
27. Won J, Tutumluer E, Byun Y-H (2023) Predicting permanent strain accumulation of unbound aggregates using machine learning algorithms. *Transp Geotech*. <https://doi.org/10.1016/j.trgeo.2023.101060>
28. Elrahman SMA, Abraham A (2013) A review of class imbalance problem. *J Netw Innov Comput* 1:332–340
29. Liu F, Ye Z, Wang L (2022) Deep transfer learning-based vehicle classification by asphalt pavement vibration. *Constr Build Mater* 342:127997
30. Joshi AV (2020) *Machine learning and artificial intelligence*. Springer International Publishing, Cham
31. Zeng Z, Zhu Z, Yao W, Wang Z, Wang C, Wei Y, Wei Z, Guan X (2022) Accurate prediction of concrete compressive strength based on explainable features using deep learning. *Constr Build Mater* 329:127082
32. J.A.K. Suykens, Nonlinear modelling and support vector machines, IMTC 2001. Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference. Rediscovering Measurement in the Age of Informatics (Cat. No.01CH 37188). 1 (2001) 287–294.

33. Zhao Y, Zhang K, Zhang Y, Luo Y, Wang S (2022) Prediction of air voids of asphalt layers by intelligent algorithm. *Constr Build Mater* 317:125908
34. Breiman L (2001) Random forests. *Mach Learn* 45:5–32
35. Geurts P, Ernst D, Wehenkel L (2006) Extremely randomized trees. *Mach Learn* 63:3–42
36. P.B. Liashchynskiy, P. Liashchynskiy, Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS, arXiv preprint arXiv:1912.06059 (2019).
37. A. Košir, A. Odić, M. Tkalčić, How to improve the statistical power of the 10-fold cross validation scheme in recommender systems, ACM International Conference Proceeding Series. (2013) 3–6.
38. Vakharia V, Gujar R (2019) Prediction of compressive strength and portland cement composition using cross-validation and feature ranking techniques. *Constr Build Mater* 225:292–301
39. SM Lundberg, PG Allen, S-I Lee. A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30 (2017). https://scholar.google.com/scholar_lookup?arxiv_id=1705.07874
40. S.M. Lundberg, G.G. Erion, S.-I. Lee, Consistent Individualized Feature Attribution for Tree Ensembles, arXiv preprint arXiv:1802.03888 (2018).
41. Sainani KL (2014) Logistic regression. *PM R* 6:1157–1162
42. Grinsztajn L, Oyallon E, Varoquaux G (2022) Why do tree-based models still outperform deep learning on tabular data? *Adv Neural Inf Process Syst* 35:507–520
43. Duddu SR, Chennarapu H (2022) Quality control of compaction with lightweight deflectometer (LWD) device: a state-of-art. *Int J Geo-Eng* 13(1):6

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.