Geomechanics and Engineering, *Vol. 30, No. 3 (2022) 247-258*
https://doi.org/10.12989/gae.2022.30.3.247

Study of oversampling algorithms for soil classifications by field velocity resistivity probe

Jong-Sub Lee[1a], Junghee Park[1b], Jongchan Kim[2c] and Hyung-Koo Yoon*[3]

[1]*School of Civil, Environmental and Architectural Engineering, Korea University, 145, Anam-ro, Seongbuk-gu, Seoul 02841, Republic of Korea*
[2]*Department of Civil and Environmental Engineering, University of California at Berkeley, Berkeley, CA 94720-1710, USA*
[3]*Department of Construction and Disaster Prevention Engineering, Daejeon University, Daejeon 34520, Republic of Korea*

(*Received January 13, 2022, Revised July 1, 2022, Accepted July 5, 2022*)

Abstract. A field velocity resistivity probe (FVRP) can measure compressional waves, shear waves and electrical resistivity in boreholes. The objective of this study is to perform the soil classification through a machine learning technique through elastic wave velocity and electrical resistivity measured by FVRP. Field and laboratory tests are performed, and the measured values are used as input variables to classify silt sand, sand, silty clay, and clay-sand mixture layers. The accuracy of k-nearest neighbors (KNN), naive Bayes (NB), random forest (RF), and support vector machine (SVM), selected to perform classification and optimize the hyperparameters, is evaluated. The accuracies are calculated as 0.76, 0.91, 0.94, and 0.88 for KNN, NB, RF, and SVM algorithms, respectively. To increase the amount of data at each soil layer, the synthetic minority oversampling technique (SMOTE) and conditional tabular generative adversarial network (CTGAN) are applied to overcome imbalance in the dataset. The CTGAN provides improved accuracy in the KNN, NB, RF and SVM algorithms. The results demonstrate that the measured values by FVRP can classify soil layers through three kinds of data with machine learning algorithms.

Keywords: classification; conditional tabular generative adversarial network (CTGAN); field velocity resistivity probe (FVRP); machine learning; synthetic minority oversampling technique (SMOTE)

1. Introduction

The field velocity resistivity probe (FVRP) was developed to measure compressional waves, shear waves, and electrical resistivity through penetration tests in soil, and the design consideration, measurement system, and measured data were already published in the Journal of Soil Dynamics and Earthquake Engineering (Yoon and Lee 2010). The first paper focused on introducing the new tool for obtaining compressional waves, shear waves, and electrical resistivity and verifying the reliability of measured values. Then, the authors attempted to expand the application of the FVRP to predict soil behavior of pore pressure parameter B through measured elastic waves, which are compressional and shear waves, and that study was published in the same journal (Lee and Yoon 2018). In this study, we tried to perform soil classification with the measured values including compressional waves, shear waves, and electrical resistivity by FVRP through machine learning algorithms.

Soil classification helps geotechnical engineers better understand the physical and chemical characteristics of soil materials (Colmenares *et al*. 2018, Al-Bared *et al*. 2019). Many studies were performed to develop the traditional classification system considering spatial delineations and representations in order to provide a more objective tool appropriate for soil, environmental, and ecosystem fields. Only shear wave velocity has been intermittently used as the input parameter to classify soil. Luzi *et al*. (2011) suggested soil categories to make hazard maps through the shear wave velocity of the upper 30 m, and the soil fundamental frequency is also used as an additional input parameter to increase reliability. Forte *et al*. (2019) provided seismic soil classification of a specific area in Italy based on averaged shear wave velocities of the upper 30 m and surface geological conditions. However, the suggested classification is focused on only the elastic response spectra as a dynamic property, without specific classification of various soils. The reason why compressional waves, shear waves, and electrical resistivity were used for soil classification is that each factor can be used independently to reflect the soil characteristics (Chen *et al*. 2018). Although there are studies using all three factors, they were usually used to understand the trend and behavior of target samples (Lee *et al*. 2017).

Compressional waves, shear waves, and electrical resistivity depend on various geotechnical properties, including elastic characterization, pore water, and particle distribution. Among the affecting factors, elastic waves are mainly affected in terms of bulk modulus because it is a representative parameter to address volume change during wave propagation. Thus, Gassmann (1951) theoretically defined elastic wave propagation in a porous medium with porosity and bulk moduli of material, skeleton, grain, and

*Corresponding author, Associate Professor
E-mail: hyungkoo@dju.ac.kr
[a]Professor
[b]Research Professor
[c]Post-doctoral fellow

ISSN: 2005-307X (Print), 2092-6219 (Online)

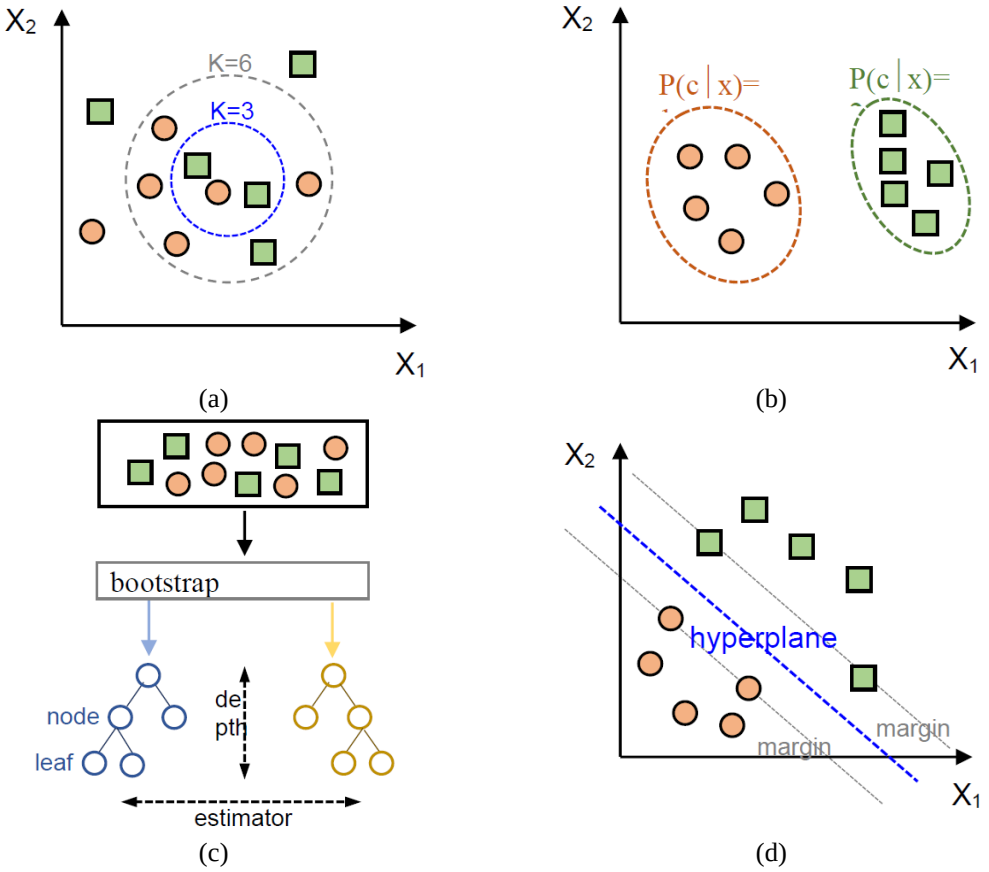

Fig. 1 Architecture of each algorithm: (a) *k*-nearest neighbors (KNN); (b) naive Bayes (NB); (c) random forest (RF); (d) support vector machine (SVM)

fluid. On the other hand, electrical resistivity, which is the reciprocal of electrical conductivity, mainly depends on soil particles, electrolytes, specific surface, and porosity, related to current flow (Song *et al*. 2019). Both elastic waves and electrical resistivity are influenced by porosity, which is the ratio between pore size and total volume, and the theoretical relationship between them was also identified in terms of porosity (Lee and Yoon 2015). Porosity is also affected by bulk density, soil aggregation, soil particles, and soil packing (Fereidooni 2018). Thus, it is possible to classify soil layers through compressional waves, shear waves, and electrical resistivity, which are functions of porosity. Using this concept, this study attempted to perform classification through the three measured factors.

It is necessary to extract a reasonable range of compressional waves, shear waves, and electrical resistivity reflecting each layer to perform classification, and machine learning is a highly versatile method (Liu *et al*. 2020, Bai *et al*. 2021). Machine learning, including random forest (Gambill *et al*. 2016), support vector machine (Kovačević *et al*. 2010), and *k*-nearest neighbor (Heung *et al*. 2016), has been applied to classify soil through soil texture, percent organic material, water storage, pH, nitrogen, and potassium. Each method shows the error ratio and important variables for enhancing reliability under the given input parameters and selected machine learning algorithm. Thus, this study also applied machine learning techniques to examine whether the three factors measured by FVRP are applicable to soil classification.

In this study, four kinds of machine learning algorithms—*k*-nearest neighbors (KNN), naive Bayes (NB), random forest (RF), and support vector machine (SVM)—were applied, and the theoretical concepts of these methods are addressed in the background theory section. After explaining the data collection through FVRP, the accuracy and confusion matrix are demonstrated to compare the resolution. Then, an explanation of the synthetic minority oversampling technique (SMOTE) and conditional tabular generative adversarial network (CTGAN) to overcome the limitation of the amount of data is addressed in the discussion section. Finally, the possibility of classification and reliability is discussed.

2. Theoretical framework

Four types of machine learning algorithms, optimized for classification, were used to classify compression wave velocity, shear wave velocity, and electrical resistivity measured by FVRP according to each soil layer, and the features of each algorithm are briefly addressed as follows.

2.1 K-Nearest Neighbors (KNN)

KNN can classify data through a number of characteristics at the nearest position based on a specific *k*

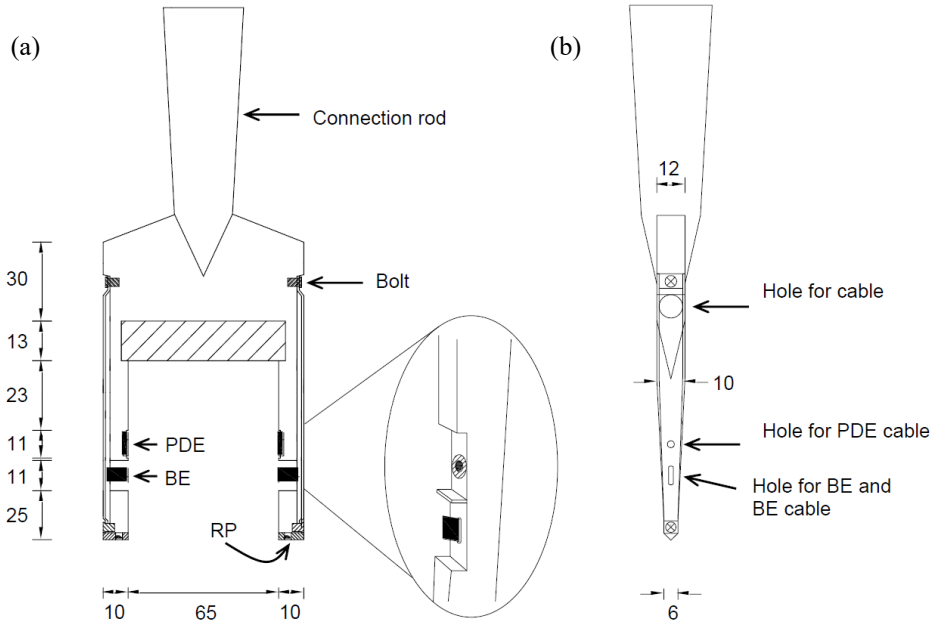

Fig. 2 Schematic drawing of the FVRP: (a) plane view; (b) side view. The units are mm. The BE, PDE and RP denote the bender elements (for S-waves), the piezoelectric disk elements (for P-waves), and the resistivity probe, respectively. The Fig. 2 is referred by Yoon and Lee (2010)

value, and this method is used for classification in various fields (Seong *et al*. 2018). The feature vector and distance are determined based on the *k* value when new data are added at a clustered area, as shown in Fig. 1(a), and new data are classified into certain sections considering the majority. Thus, the *k* value is a crucial hyperparameter that greatly affects the accuracy, since it determines the majority of KNN. Even though Euclidean and Manhattan distances are selected, Minkowski distance is generally used, with the advantage of a comprehensive value in a normalized vector field.

2.2 Naive Bayes (NB)

NB is a conditional independence method that classifies data through the relationship between prior and posterior probability (Feng *et al*. 2018). The posterior probability is calculated through Eq. (1), assuming that the values of each feature have an independent relationship. The classification is performed through an independent contribution, called naive assumption, to determine a corresponding classification range.

$$P(\mathrm{c}|x)=\frac{P(x|c)P(c)}{P(x)} \tag{1}$$

where $P(c|x)$ denotes the posterior probability, through which new data are classified, as shown in Fig. 1(b). $P(c)$ and $P(x)$ show the class prior and predictor prior probabilities. $P(x|c)$ is likelihood, and becomes zero when new data are unclassified in the corresponding section. This leads to an error, resulting in classifying the opposite characteristic, and is called zero frequency. Laplace smoothing is performed to prevent zero frequency by adding a constant value to the denominator and numerator in Eq. (1), thus, the value of Laplace smoothing is regarded as an important hyperparameter in NB (Raizada and Lee 2013).

2.3 Random Forest (RF)

RF is an algorithm based on decision trees and uses a hierarchy of various trees to make decisions, as shown in Fig. 1(c). The dataset was configured through bootstrap aggregation, which allows duplication to randomly select the data, unlike the decision tree algorithm. Thus, RF makes each tree correlated and consequently improves the generalization of the result. In addition, the ensemble model is applied to increase the efficiency of parallel computation, because learning of all trees is independent. The number of trees (estimator) and the allowable depth (depth) are important influence factors in the accuracy of RF, and the reliability is also influenced by min-samples-leaf and min-samples-node, which represent the minimum amount of data in each leaf and node.

2.4 Support Vector Machine (SVM)

SVM is a probabilistic binary linear classifier that determines which category the input data belongs to. The learning data is placed in a certain hyperplane to find the maximum margin between two constructed categories, as shown in Fig. 1(d). The testing data are mapped to the categories they belong to in the determined hyperplane. The input data must be plotted in a region appropriate for a large width with the hyperplane to improve reliability, thus the hyperplane should be properly constructed. The nonlinear hyperplane can be set through the radial basis function (Rbf) kernel when it is difficult to completely separate the input data linearly. Parameter C is also used to determine

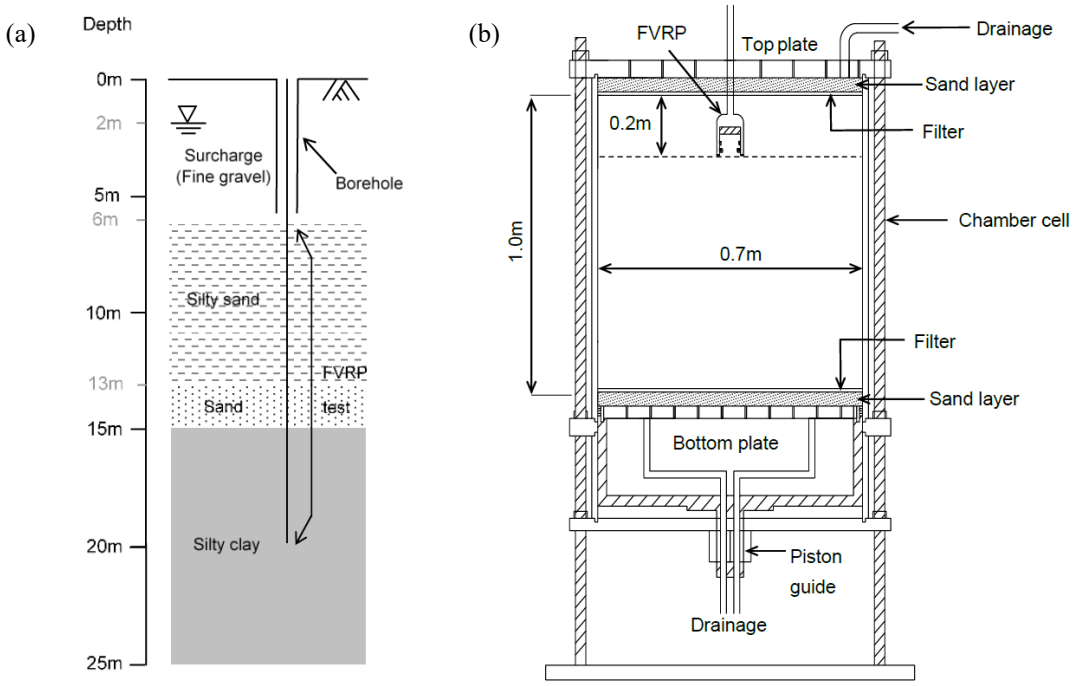

Fig. 3 FVRP penetration test in (a) field (at Kwang Yang); (b) laboratory. This Figure is referred by Yoon and Lee (2010)

the hyperplane to consider how many data samples will be allowed. Additionally, gamma, which is related to the standard deviation of the Gaussian function, is a crucial hyperparameter to determine the degree to which each piece of data affects the entire dataset.

3. Data collection

The field velocity resistivity probe (FVRP), which can measure the compressional waves, shear waves, and electrical resistivity of the ground through penetration experiments as shown in Fig. 2, was used to gather the data. Compressional waves, shear waves, and electrical resistivity can be obtained from the piezo disk element, bender element, and electrical resistivity probe, respectively, installed in the FVRP. The sampling rate of the FVRP depends on the performance of the penetration rig, and the FVRP has been employed for acquisition at a depth of 30 m with intervals of at least 10 cm. Thus, the equipment has the advantage of providing subsurface characterizations with high resolution. The data measured by FVRP have already been verified in previously published papers, and the shape, measurement system, and experimental method will be replaced with references (Yoon and Lee 2010, Lee and Yoon 2018). Three kinds of properties were measured in four soil types through laboratory and field tests as shown in Fig. 3. The field tests were performed on the southern coast of the Korean peninsula at Kwang Yang, where surcharge layers formed, improving the soft ground. The layers were classified through SPT N-value, split barrel sampler, and soil properties, and the testing site was classified as silty sand, sand, and silty clay layers at a depth of 5-12, 13-15, and 15-21 m, respectively, as reported in Yoon and Lee (2010). In silty sand and silty clay layers, the ranges of compressional waves, shear waves, and electrical resistivity were 1461-1623 m/s, 100-184 m/s, and 1.08-1.91 Ω·m, and 1468-1627 m/s, 94-159 m/s, and 1.01-1.40 Ω·m, respectively. The measured ranges of compressional waves, shear waves, and electrical resistivity were 1493-1646 m/s, 113-194 m/s, and 1.23-2.17 Ω·m in the sand layer.

Even though the compression wave velocity had similar ranges in silty sand and silty clay, the silty sand showed higher shear wave velocity and electrical resistivity because the sand particles makes soil dense with the contact effect (Byun *et al*. 2019). In the sand layer, the values of the three properties were relatively high due to denser packing with stiff particles (Lee *et al*. 2017). A laboratory experiment was carried out with a clay-sand mixture in which the consolidation process was completed after mixing the clay and sand. At the first step of the consolidation process, stabilization was kept with a vertical load of 50 kPa for 2 days, then the vertical load was gradually increased to 200 kPa for 30 days. After completing the consolidation process, the FVRP was penetrated into the specimen to characterize the clay-sand mixture. The measured values are shown in Fig. 4, with ranges of 1510-1675 m/s, 118-137 m/s, and 6.39-7.14 Ω·m for compressional waves, shear waves, and electrical resistivity, respectively. The number of pieces of data was 66, 15, 56, and 30 for silty sand, sand, silty clay, and clay-sand mixture, respectively. Although the amount of data for learning was small, the study evaluated the possibility of classification only focusing on expanding the application of FVRP.

4. Optimization method

A flowchart of this study is demonstrated in Fig. 5. The KNN, NB, RF, and SVM algorithms were applied to

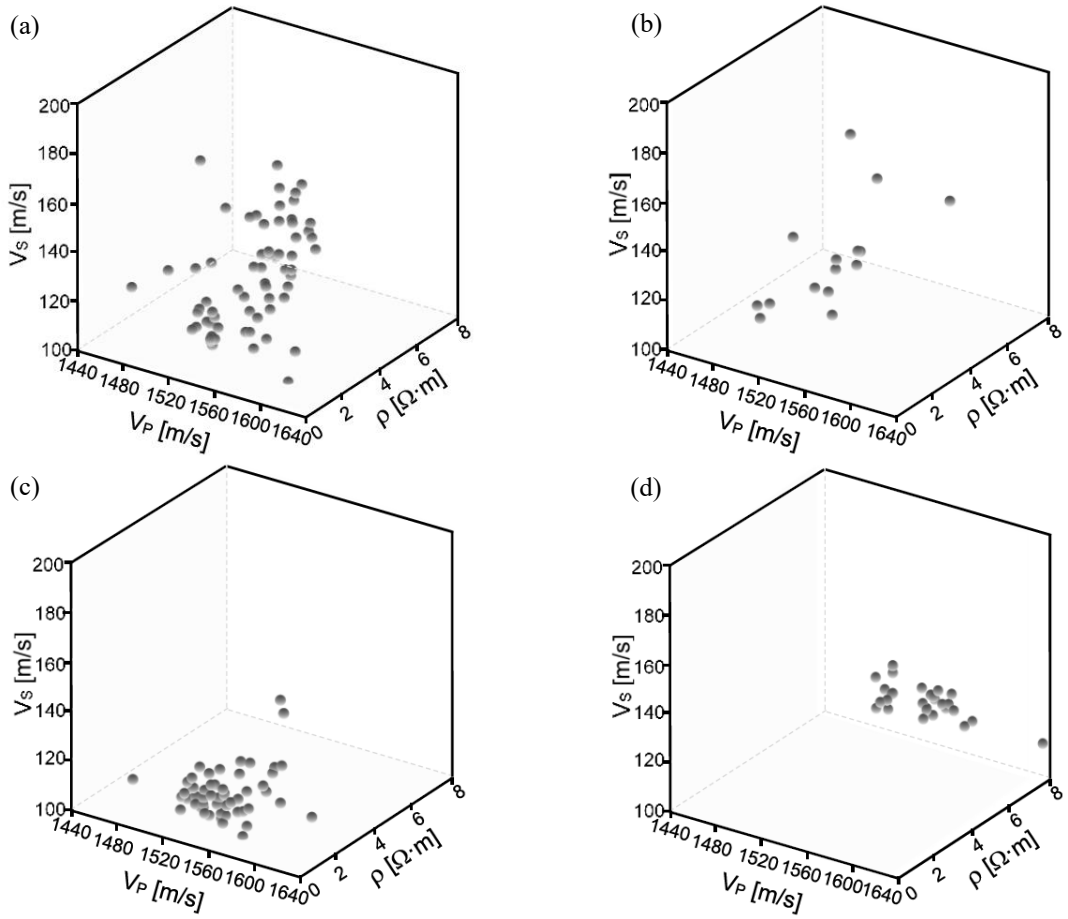

Fig. 4 Profiles of measured compressional wave velocity (VP), shear wave velocity (VS), and electrical resistivity (ρ) through field velocity resistivity probe (FVRP): (a) silty sand; (b) sand; (c) silty clay; (d) clay–sand mixture

classify the measured values of compressional waves, shear waves, and electrical resistivity. Various ranges of hyperparameters were applied to find the best performance of each algorithm with reference to previous studies (Kovačević *et al*. 2010, Gambill *et al*. 2016, Heung *et al*. 2016, Yoon 2020, Lee and Yoon 2021, Min and Yoon 2021). The *k* values in applied in KNN were 1 to 10 in 1 steps based on Minkowski distance, and the Laplace smoothing in NB was selected as 0.0001, 0.001, 0.01. 0.1, 1, and 10 assuming Gaussian distribution for calculating the probability of continuous input variables. The estimator and depth related to RF were determined to be 1, 2, 5, 10, 20, and 50 and 1, 2, 5, 10, and 20, respectively. Entropy was used as a criterion to measure the degree of disorder of input variables in RF. Finally, the C and gamma values of SVM were set equally to 0.1, 1, 10, and 100 with the Rbf kernel to handle the nonlinear relationship of the data. The test size was also given various conditions from 0.1 to 0.4 in 0.1 steps to assess the reliability of the selected algorithm and hyperparameter. All algorithms used a standard scaler among scaling techniques to prevent overflow and underflow of data and reduce the number of independent variables. This can improve stability and convergence speed in the optimization process. To perform the machine learning process, the Pandas package (version 2.8.1) was used with the Python language (version 3.9).

5. Results

The reliability of optimization methods was expressed as accuracy according to true negative, false negative, false positive, and true positive rates and the total dataset, and the values are plotted in Figs. 6-9 for the KNN, NB, RF, and SVM, algorithms respectively. The dotted line in the bar graph indicates the highest accuracy. In the KNN algorithm, excellent accuracy of 0.76 was calculated a k value of 5 and a test size of 0.2. NB showed very high accuracy of 0.91 at two Laplace smoothing values of 0.0001 and 0.001 with the same test size as KNN. For RF, various levels of accuracy were derived according to the test size, and high performance was recorded at 0.88, 0.94, 0.84, and 0.85 when the test size was 0.1, 0.2, 0.3, and 0.4, respectively. Therefore, depth was found to be 10 and 20 under 0.2 and 50 test size and estimator, respectively, to provide optimal performance in RF. Even though both min-samples-leaf and min-samples-node were set to 1, 2, 5, and 10, there was no change in accuracy. Thus, min-samples-leaf and min-samples-node were fixed at 1 and 2 to perform the RF algorithm. Finally, SVM also showed different accuracy according to the test size, and a high accuracy value of 0.88 was derived for test sizes of 0.2 and 0.3. It can be seen that gamma and C values correspond to 0.1, 1 and 10, 100, as shown in Fig. 9. The confusion matrix, which is a useful

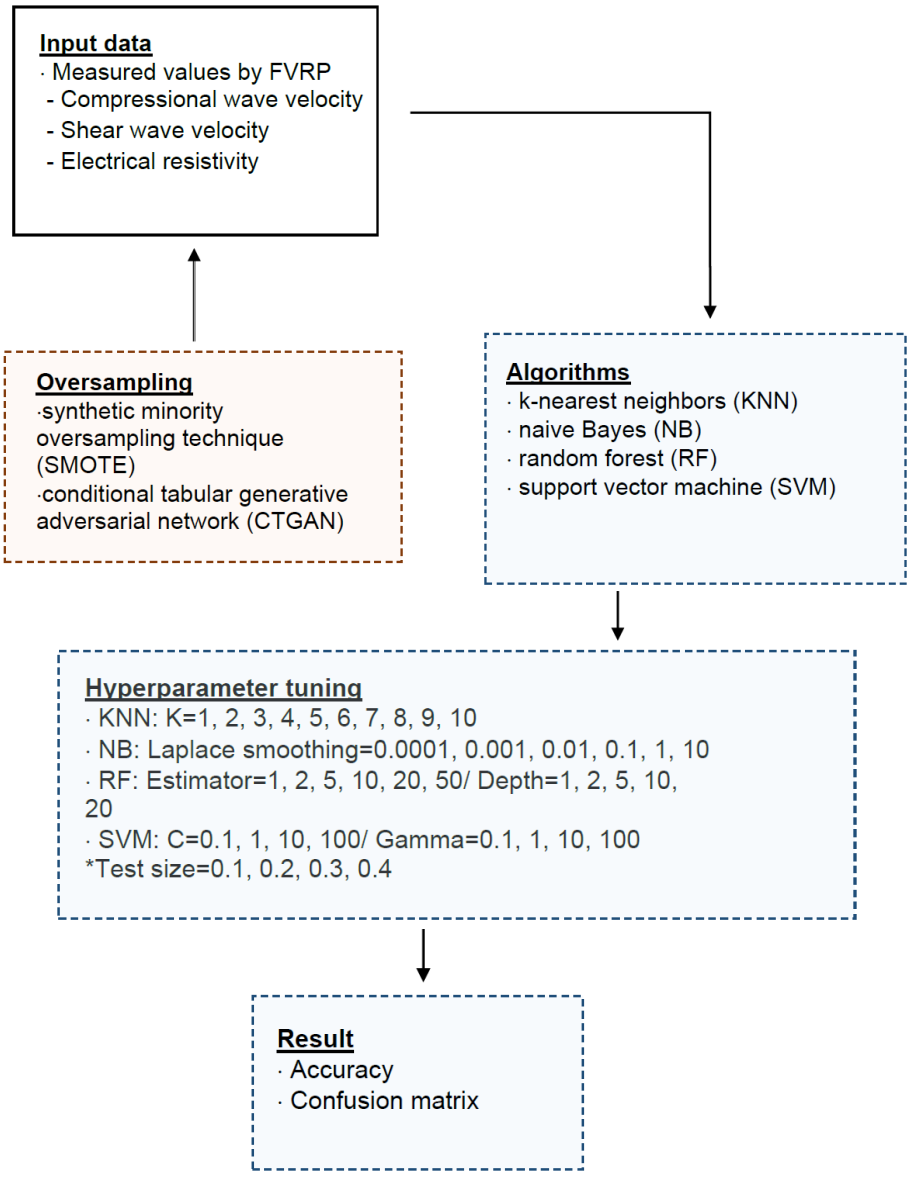

Fig. 5 Flowchart of this study for machine learning

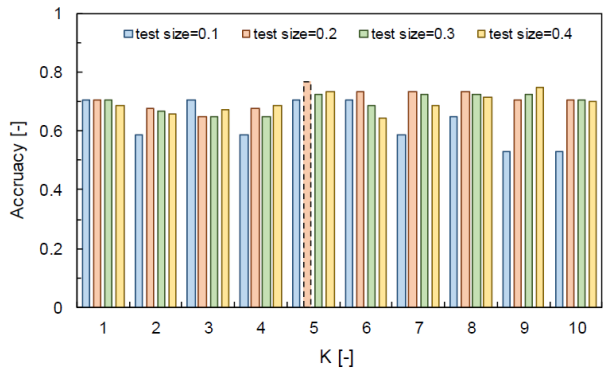

Fig. 6 Results of accuracy based on k-nearest neighbors (KNN). The dotted line shows the highest accuracy

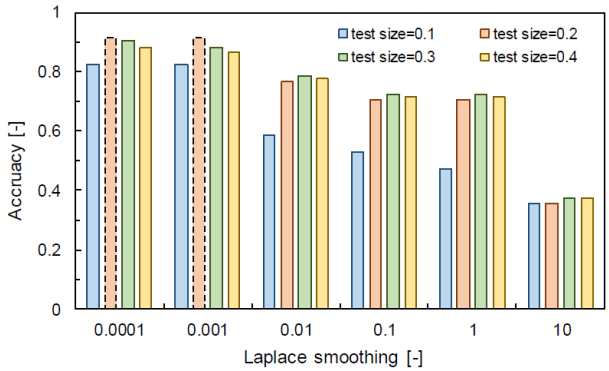

Fig. 7 Results of accuracy based on naive Bayes (NB). The dotted line shows the highest accuracy

method to visualize results in N classes, is plotted in Fig. 10 as a 3×3 square matrix based on optimal hyperparameter.

The clay-sand mixture shows a value of 1 for all algorithms, which is excellent performance. Even though silty clay has slightly lower reliability with KNN, the overall classification is satisfactory. Silty sand also showed relatively low accuracy with KNN, however, it showed an accuracy of 0.9 or higher with the other algorithms.

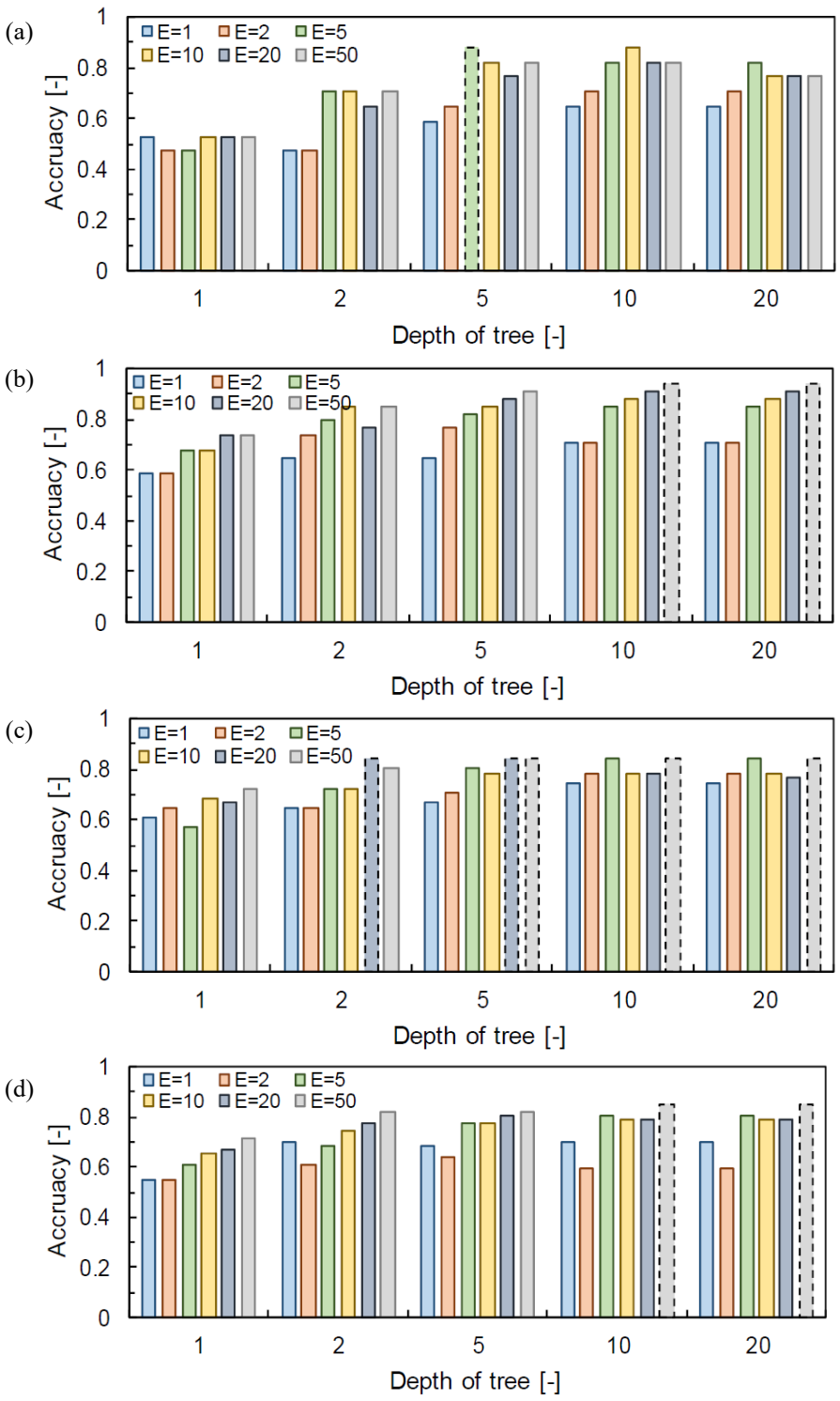

Fig. 8 Results of accuracy based on random forest (RF) with test sizes of: (a) 0.1, (b) 0.2, (c) 0.3, and (d) 0.4. The dotted line shows the highest accuracy

Misclassification was shown in the sand layer by all algorithms. The reason for this is that the number of data items in the sand layer was 15, and a relatively small training dataset was constructed. This makes it difficult to properly reflect the characteristics of compressional waves, shear waves, and electrical resistivity. Although the number of items of clay-sand mixture data was also small at 30, the characterization was easy because the range of measured electrical resistivity values was larger than that of other layers.

6. Discussion

To overcome the imbalanced distribution of the amount of data in silty sand, sand, silty clay, and clay-sand mixture,

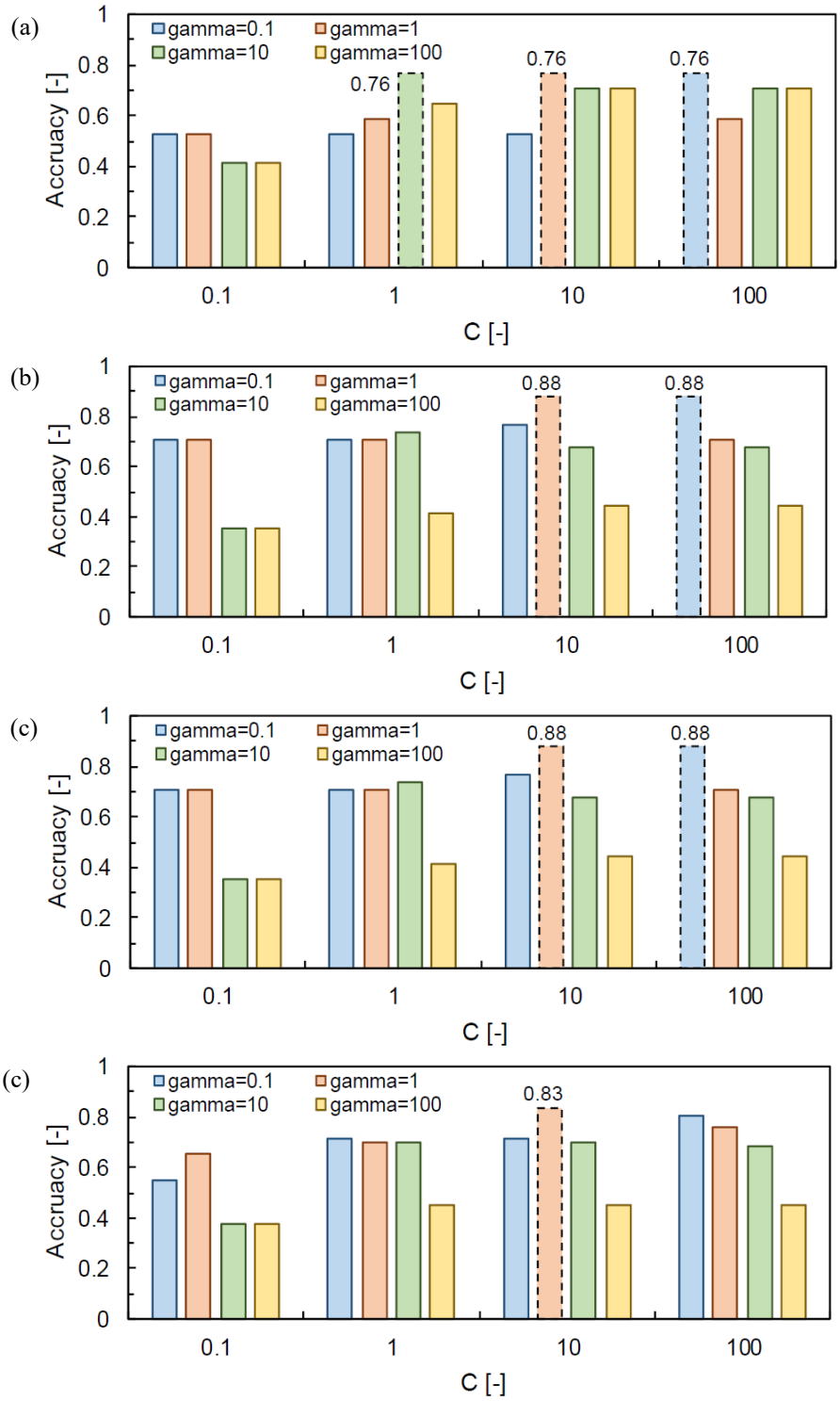

Fig. 9 Results of accuracy based on support vector machine (SVM) with test sizes of: (a) 0.1, (b) 0.2, (c) 0.3, and (d) 0.4. The dotted line shows the highest accuracy

the synthetic minority oversampling technique (SMOTE) and conditional tabular generative adversarial network (CTGAN) were applied.

SMOTE can provide the oversampling of positive or minority classes assuming that a minority data are distributed in a linearly separable space (Puri and Gupta 2021). The new data are generated corresponding to the desirable amount of data for the positive class based on *k*-nearest neighbor values. The synthetic data can be created from randomly selected values from a minority class, and this is repeated until the amount of data of the minor class is equal to that of the positive class.

Thus, all numbers of data items of sand, silty clay, and the clay-sand mixture were oversampled to 66 to match the data count of the positive class in silty sand. The *k*-nearest neighbor values were set from 1 to 14 in consideration of

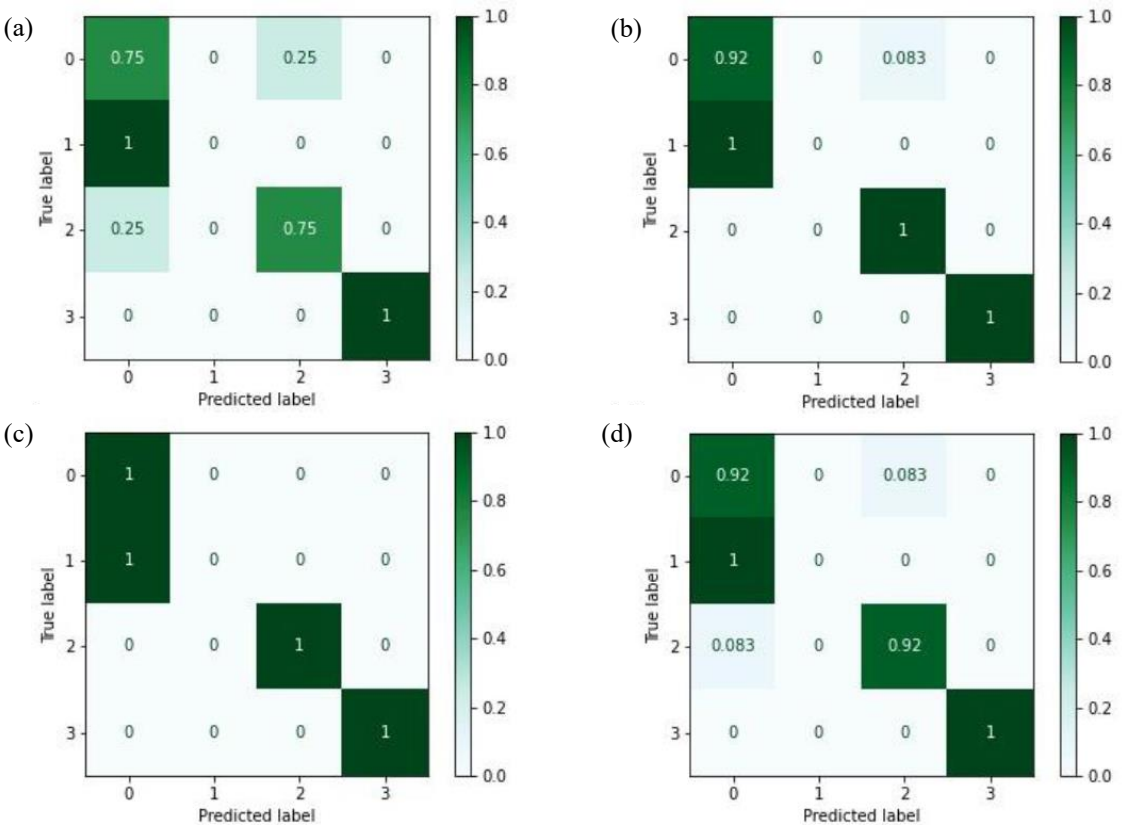

Fig. 10 Confusion matrices based on optimal hyperparameters as shown in Table 1: (a) KNN; (b) NB; (c) RF; (d) SVM. Numbers 0, 1, 2, and 3 denote silty sand, sand, silty clay, and clay–sand mixture, respectively

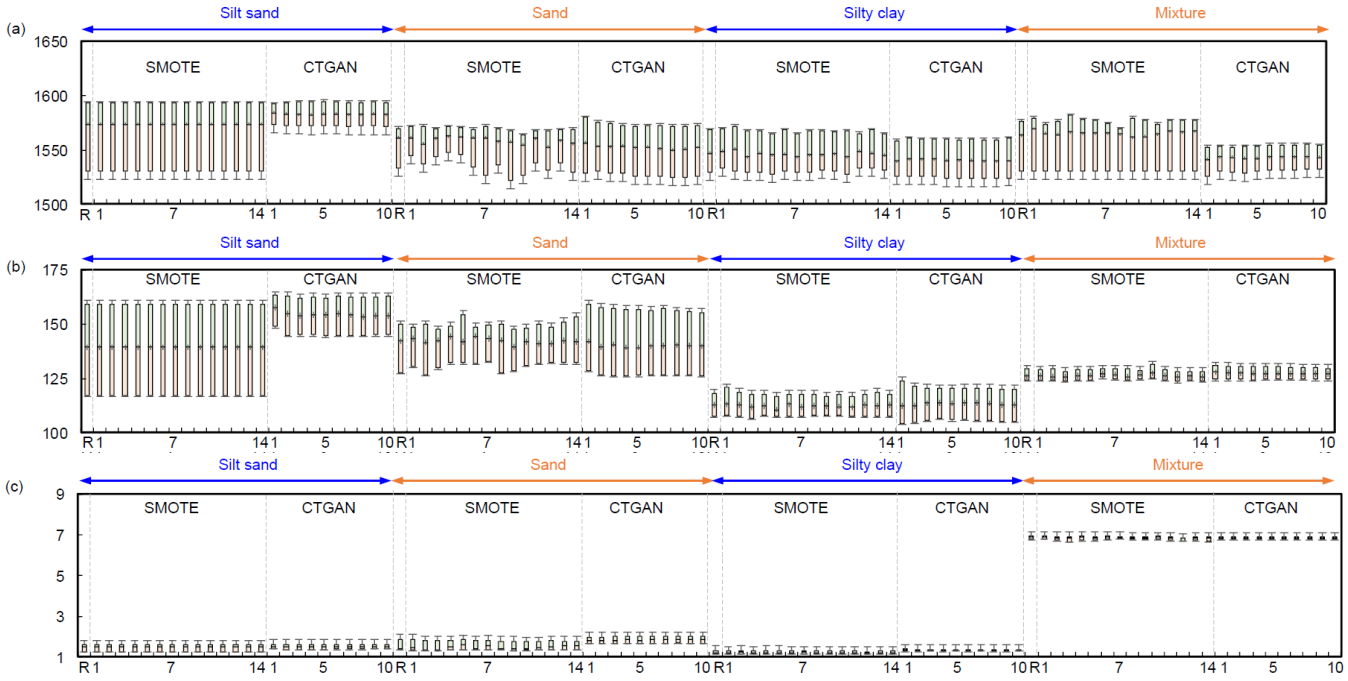

Fig. 11 Distributions of data based on raw, SMOTE and CTGAN with box plot: (a) compressional wave; (b) shear wave; (c) electrical resistivity. R denotes the raw data. The numbers 1 to 14 in SMOTE mean *k* values, and 1 to 10 in CTGAN means the multiple of the data

the data distributions of sand, which had the minimum number of 15 data points; then, oversampling was performed.

Generative adversarial network (GAN) is a method of oversampling the amount of data after learning its distribution, and it consists of a generator for creating data

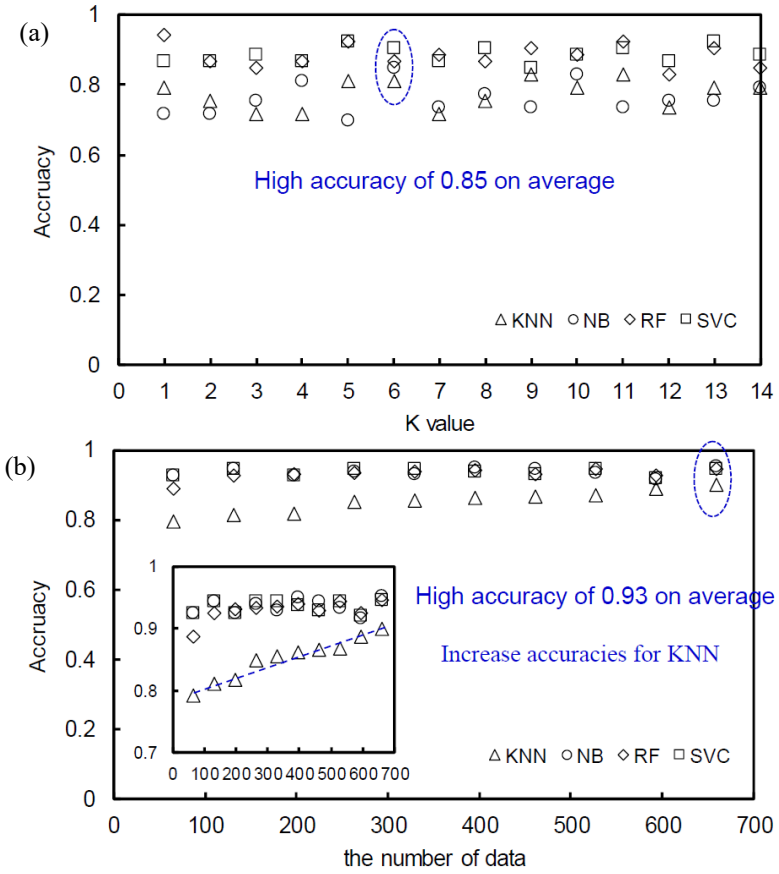

Fig. 12 Calculated accuracy with oversampling data deduced by (a) SMOTE; (b) CTGAN

and a discriminator for verifying them (Engelmann and Lessmann 2021). Even though GAN is widely used for generating images, they have limitations for applications in structured data with independent features because they follow a non-Gaussian distribution. To improve this, CTGAN has been proposed and applied to structured data to comply with the column-specific variational Gaussian mixture (VGM) of generated data. The number of CTGAN-based oversampling increased from 1 to 10 times based on the silty sand, showing the maximum amount of data. Therefore, the number of data in each layer was oversampled from 66 (1 time) to 660 (10 times). The oversampled data were trained through KNN, NB, RF, and SVM algorithms with the same hyperparameter ranges, as shown in Fig. 5. In this study, the Imblearn package (version 2.1.0) with Python language (version 3.9) was used to perform SMOTE and CTGAN.

The distribution of data created by SMOTE and CTGAN is shown in Fig. 11 as a box plot, and the R shows the original data. Although the range of the generated data was slightly different from the original data, the median values were almost similar. In the case of silty sand created by SMOTE, the same distributions according to the *K* value were shown because the silty sand is a positive class. However, various distributions increasing the amount of data were oversampled in all soil layers, regardless of class. Note that the data oversampled by SMOTE had a large variation in the maximum, minimum, and interquartile range with growing K values; however, the data derived by CTGAN showed almost constant boxes and whiskers regardless of the number of data points. The reason for this result is that SMOTE is performed by focusing on the amount of data in the positive class according to the *K* value. On the other hand, CTGAN reflects the unique characteristics of raw data and provides data within a certain range; there seems to be a difference in the data distribution oversampled by the two methods.

Fig. 12 shows the accuracy result of applying machine learning algorithms of KNN, NB, RF and SVM through the oversampled data by SMOTE and CTGAN. In the data generated by SMOTE, the accuracy estimated by each algorithm exhibited high fluctuations without a constant tendency, increasing the *k* value. When the *k* value was 6, the averaged accuracy of all algorithms was 0.85, which was relatively high. However, when KNN, NB, RF, and SVM algorithms were applied to the data derived from CTGAN with 600 data values, a high accuracy of 0.93, on average, was found. In addition, it was found that the accuracy of the KNN algorithm gradually increased as the amount of data oversampled by CTGAN increased. Therefore, the results derived from CTGAN show excellent accuracy with increasing the amount of data, which proves that CTGAN provides high-quality data.

The highest accuracy values in each algorithm deduced by original data and oversampled data were compared and are plotted in Fig. 13. The data generated by SMOTE showed higher accuracy in KNN, RF, and SVM algorithms than raw data; however, the NB algorithm exhibited a lower

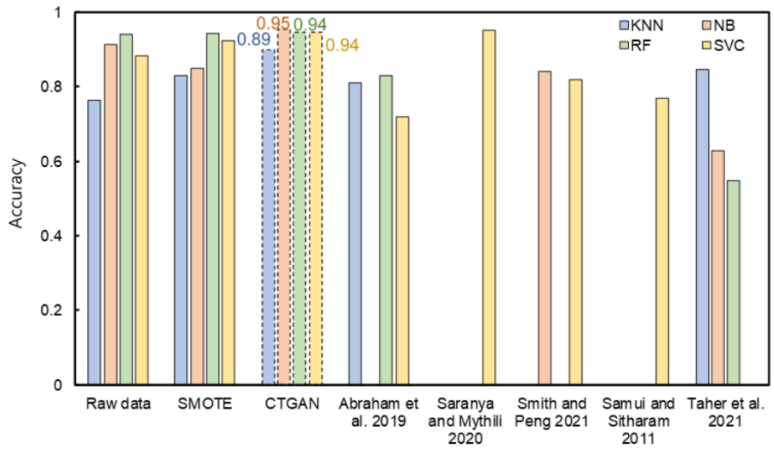

Fig. 13 Comparisons of accuracies derived from previous research results of soil classification. The dotted line shows the highest accuracy

accuracy. The data oversampled by CTGAN showed higher accuracy than raw data in all algorithms, and accuracy values were also demonstrated to be higher than those achieved by SMOTE. The oversampling method through CTGAN is essential to improve the soil classification accuracy of data measured by an FVRP. Accuracy results achieved in a previous study based on machine learning for soil classification are also displayed in Fig. 13 to verify the reliability of results in this study. Among various studies, the results (Smith and Peng 2009, Samui and Sitharam 2011, Abraham *et al*. 2019, Saranya and Mythili 2020, Taher *et al*. 2021) using the same algorithm as the corresponding study are shown. As a result of comparison, only the RF algorithm showed a high accuracy when the raw data were used to perform soil classification; however, previous studies demonstrated better accuracy with the rest of the algorithms. When oversampling was performed by CTGAN, accuracy was improved in all algorithms; thus, more reliable results could be obtained than other research results. The direct comparison has a limitation because the input parameters of previous studies were selected as various characteristics of soil, including geotechnical properties, chemical properties and hydrologic soil groups, rather than compressional waves, shear waves, and electrical resistivity measured by the FVRP. However, it shows that excellent performance in soil classification can be attained when oversampling is performed by CTGAN to the measured values by FVRP. This study shows that the oversampling method of CTGAN can be accurately applied to structural data and that soil classification is possible through FVRP, which has a limitation of obtaining data only through the penetration test.

7. Conclusions

Machine learning techniques were applied to estimate the possibility of evaluating soil classification by measuring compressional waves, shear waves, and electrical resistivity with an FVRP. The detailed conclusions are as follows:

Four algorithms, representing optimal algorithms in machine learning, were used to classify soil strata: *k*-nearest neighbors (KNN), naive Bayes (NB), random forest (RF), and support vector machine (SVM).

The synthetic minority oversampling technique (SMOTE) and conditional tabular generative adversarial network (CTGAN) were applied to overcome the limitations of an insufficient amount of data in FVRP.

Even though only RF provided the highest accuracy in SMOTE, the oversampled data through CTGAN increase the reliability of soil classification performed by each algorithm with compressional wave, shear wave and electrical resistivity of FVRP.

Acknowledgments

This work is supported by the Korea Agency for Infrastructure Technology Advancement (KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 21CTAP-C164152-01).

References

Abraham, S., Huynh, C. and Vu, H. (2020), "Classification of soils into hydrologic groups using machine learning", *Data*, **5**(1), 2. https://doi.org/10.3390/data5010002.

Al-Bared, M.A., Harahap, I.S., Marto, A., Abad, S.V.A.N.K. and Ali, M.O. (2019), "Undrained shear strength and microstructural characterization of treated soft soil with recycled materials", *Geomech. Eng.*, **18**(4), 427-437. https://doi.org/10.12989/gae.2019.18.4.427.

Bai, X.D., Cheng, W.C., Ong, D.E. and Li, G. (2021), "Evaluation of geological conditions and clogging of tunneling using machine learning", *Geomech. Eng.*, **25**(1), 59-73. https://doi.org/10.12989/gae.2021.25.1.059.

Byun, Y.H., Hong, W.T. and Yoon, H.K. (2019), "Characterization of cementation factor of unconsolidated granular materials through time domain reflectometry with variable saturated conditions", *Mater.*, **12**(8), 1340. https://doi.org/10.3390/ma12081340.

Chen, Y., Irfan, M., Uchimura, T., Cheng, G. and Nie, W. (2018), "Elastic wave velocity monitoring as an emerging technique for rainfall-induced landslide prediction", *Landslid.*, **15**(6), 1155-1172. https://doi.org/10.1007/s10346-017-0943-3.

Colmenares, J., Dávila, J., Vega, J. and Shin, J. (2018), "Tunnelling on terrace soil deposits: Characterization and experiences on the Bogotá-Villavicencio road", *Geomech. Eng.*, **15**(3), 899-910. https://doi.org/10.12989/gae.2018.15.3.899.

Engelmann, J. and Lessmann, S. (2021), "Conditional Wasserstein GAN-based oversampling of tabular data for imbalanced learning", *Exp. Syst. Appl.*, **174**, 114582. https://doi.org/10.1016/j.eswa.2021.114582.

Feng, X., Li, S., Yuan, C., Zeng, P. and Sun, Y. (2018), "Prediction of slope stability using naive Bayes classifier", *KSCE J. Civil Eng.*, **22**(3), 941-950. https://doi.org/10.1007/s12205-018-1337-3.

Fereidooni, D. (2018), "Assessing the effects of mineral content and porosity on ultrasonic wave velocity", *Geomech. Eng.*, **14**(4), 399-406. https://doi.org/10.12989/gae.2018.14.4.399.

Forte, G., Chioccarelli, E., De Falco, M., Cito, P., Santo, A. and Iervolino, I. (2019), "Seismic soil classification of Italy based on surface geology and shear-wave velocity measurements", *Soil Dyn. Earthq. Eng.*, **122**, 79-93. https://doi.org/10.1016/j.soildyn.2019.04.002.

Gambill, D.R., Wall, W.A., Fulton, A.J. and Howard, H.R. (2016), "Predicting USCS soil classification from soil property variables using Random Forest", *J. Terramech.*, **65**, 85-92. https://doi.org/10.1016/j.jterra.2016.03.006.

Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E. and Schmidt, M.G. (2016), "An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping", *Geoderma*, **265**, 62-77. https://doi.org/10.1016/j.geoderma.2015.11.014.

Kovačević, M., Bajat, B. and Gajić, B. (2010), "Soil type classification and estimation of soil properties using support vector machines", *Geoderma*, **154**(3-4), 340-347. https://doi.org/10.1016/j.geoderma.2009.11.005.

Lee, J.S. and Yoon, H.K. (2015), "Theoretical relationship between elastic wave velocity and electrical resistivity", *J. Appl. Geophys.*, **116**, 51-61. https://doi.org/10.1016/j.jappgeo.2015.02.025.

Lee, J.S. and Yoon, H.K. (2018), "Application example: Field Velocity Resistivity Probe (FVRP) for predicting pore pressure parameter B", *Soil Dyn. Earthq. Eng.*, **107**, 214-217. https://doi.org/10.1016/j.soildyn.2018.01.039

Lee, J.S., Byun, Y.H. and Yoon, H.K. (2017), "Study of Activation Energy in Soil through Elastic Wave Velocity and Electrical Resistivity", *Vadose Zone J.*, **16**(6), 1-9. https://doi.org/10.2136/vzj2016.08.0073

Lee, S.J. and Yoon, H.K. (2021), "Discontinuity predictions of porosity and hydraulic conductivity based on electrical resistivity in slopes through deep learning algorithms", *Sensor.*, **21**(4), 1412. https://doi.org/10.3390/s21041412.

Liu, L.L., Yang, C. and Wang, X.M. (2020), "Landslide susceptibility assessment using feature selection-based machine learning models", *Geomech. Eng.*, **25**, 1-16. https://doi.org/10.12989/gae.2021.25.1.001.

Luzi, L., Puglia, R., Pacor, F., Gallipoli, M.R., Bindi, D. and Mucciarelli, M. (2011), "Proposal for a soil classification based on parameters alternative or complementary to Vs 30", *Bull. Earthq. Eng.*, **9**(6), 1877-1898. https://doi.org/10.1007/s10518-011-9274-2.

Min, D.H. and Yoon, H.K. (2021), "Suggestion for a new deterministic model coupled with machine learning techniques for landslide susceptibility mapping", *Sci. Rep.*, **11**(1), 1-24. https://doi.org/10.1038/s41598-021-86137-x.

Puri, A. and Gupta, M.K. (2021), "Knowledge discovery from noisy imbalanced and incomplete binary class data", *Exp. Syst. Appl.*, **181**, 115179. https://doi.org/10.1016/j.eswa.2021.115179.

Raizada, R.D. and Lee, Y.S. (2013), "Smoothness without smoothing: why Gaussian naive Bayes is not naive for multi-subject searchlight studies", *PloS one*, **8**(7), e69566. https://doi.org/10.1371/journal.pone.0069566.

Samui, P. and Sitharam, T.G. (2011), "Machine learning modelling for predicting soil liquefaction susceptibility", *Nat. Hazard. Earth Syst. Sci.*, **11**(1), 1-9. https://doi.org/10.5194/nhess-11-1-2011.

Saranya, N. and Mythili, A. (2020), "Classification of soil and crop suggestion using machine learning techniques", *IJERT*, **9**(02), 671-673.

Seong, H., Son, H. and Kim, C. (2018), "A comparative study of machine learning classification for color-based safety vest detection on construction-site images", *KSCE J. Civil Eng.*, **22**(11), 4254-4262. https://doi.org/10.1007/s12205-017-1730-3.

Smith, D. and Peng, W. (2009), "Machine learning approaches for soil classification in a multi-agent deficit irrigation control system", *2009 IEEE International Conference on Industrial Technology*, 1-6.

Song, J.U., Lee, J.S. and Yoon, H.K. (2019), "Application of electrical conductivity method for adsorption of lead ions by rice husk ash", *Measure.*, **144**, 126-134. https://doi.org/10.1016/j.measurement.2019.04.094.

Taher, K.I., Abdulazeez, A.M. and Zebari, D.A. (2021), "Data mining classification algorithms for analyzing soil data", *Asian J. Res. Comput.*, 17-28. https://doi.org/10.9734/ajrcos/2021/v8i230196.

Yoon, H.K. (2020), "Relationship between aspect ratio and crack density in porous-cracked rocks using experimental and optimization methods", *Appl. Sci.*, **10**(20), 7147. https://doi.org/10.3390/app10207147.

Yoon, H.K. and Lee, J.S. (2010), "Field velocity resistivity probe for estimating stiffness and void ratio", *Soil Dyn. Earthq. Eng.*, **30**(12), 1540-1549. https://doi.org/10.1016/j.soildyn.2010.07.008.

IC