JZUS

Application of machine learning to the identification of quick and highly sensitive clays from cone penetration tests[*]

Cristian GODOY[†1], Ivan DEPINA[2,3], Vikas THAKUR[1]

[1]*Department of Civil and Environmental Engineering, Norwegian University of Science and Technology,*
Trondheim 7031, Norway
[2]*Department of Infrastructure, SINTEF Community, Trondheim 7034, Norway*
[3]*Faculty of Civil Engineering, Architecture and Geodesy, University of Split, Split 21000, Croatia*
[†]E-mail: cristg@stud.ntnu.no

Abstract: Geotechnical classification is vital for site characterization and geotechnical design. Field tests such as the cone penetration test with pore water pressure measurement (CPTu) are widespread because they represent a faster and cheaper alternative for sample recovery and testing. However, classification schemes based on CPTu measurements are fairly generic because they represent a wide variety of soil conditions and, occasionally, they may fail when used in special soil types like sensitive or quick clays. Quick and highly sensitive clay soils in Norway have unique conditions that make them difficult to be identified through general classification charts. Therefore, new approaches to address this task are required. The following study applies machine learning methods such as logistic regression, Naive Bayes, and hidden Markov models to classify quick and highly sensitive clays at two sites in Norway based on normalized CPTu measurements. Results showed a considerable increase in the classification accuracy despite limited training sets.

Key words: Machine learning; Classification; Quick clays; Sensitive clays
https://doi.org/10.1631/jzus.A1900556   **CLC number:** TU19

1 Introduction

One of the primary concerns in the majority of construction projects in Norway is the presence of highly sensitive or quick clays, which significantly affects the feasibility of such projects. As cone penetration tests with pore water pressure measurement (CPTus) are widespread and present in almost every geotechnical exploration program in Norway, it would be convenient to determine whether a soil profile contains quick clays based on the CPTu test results.

The use of CPTu for soil classification is a common practice, particularly using the well-known classification charts found in (Lunne et al., 1997). However, a major challenge comes to light when the soil deposits comprise non-textbook soils, as in the case of quick or highly sensitive clays. In these cases, alternatives should be determined to maintain the convenience of using indirect field measurements without expending a large amount of resources.

In this context, the use of machine learning approaches is ideal as local data can be used to train a model to learn how the measured data characterize a certain kind of soil. With little information, results will not be satisfactory; however, as the exploration advances, the model will learn from the newly obtained data and adjust itself to provide better results.

ORCID: Cristian GODOY, https://orcid.org/0000-0001-8449-982X

This study investigates the potential of machine learning techniques to improve the identification of highly sensitive and quick clay soils using CPTu. All computations performed within this study use Python (van Rossum, 1995) as the programming environment. Machine learning algorithms used are logistic regression and Naive Bayes, as programmed in the scikit-learn library (Pedregosa et al., 2011), and the hidden Markov model (HMM), available in the hmmlearn library (hmmlearn, 2010).

The methodology followed considers the analysis of two CPTu datasets from previous studies at sites wherein highly sensitive and quick clays were encountered and wherein the layering (lithology) at each test location is known. The CPTu data were then used to classify the samples using well-known classification charts and machine learning methods. Finally, the results were compared against the actual layering, and performance measurements were computed to compare the different approaches.

2 Datasets

2.1 Norwegian geo-test sites dataset

The Norwegian geo-test sites (NGTS) dataset is a research consortium led by the Norwegian Geotechnical Institute (NGI), with the participation of the Norwegian University of Science and Technology and other organizations. Its main focus is to develop field laboratories for the testing, verification, and control of new methods and equipment for site investigations and foundation engineering (NGI, 2019). Within the NGTS framework, an important study subject is quick clays, for which the site at Tiller (Trondheim, Norway) was chosen. Fig. 1 presents the location of the CPTus, while Figs. 2 and 3 show the summary of the tests alongside the layering of the site. In this study, 31 CPTus were used (CPTu C18 was discarded due to high sleeve friction (f_s), which was not representative of the site).

The layering of the site consisted of 2.0 m of dry crust, followed by a clay layer up to a depth of 7.5 m on top of a quick clay layer of 12.5 m thick. The water table was 1.5 m from the surface (a hydrostatic condition was assumed). The terrain was flat, so the features described above were expected to have few variations over the study area. The lithology was based on previous studies conducted at the site,

including soundings and laboratory tests. Detailed information about this can be found in (L'Heureux et al., 2019).

2.2 Vegvesen dataset

This dataset consisted of seven CPTus that were part of the studies for the construction of a bridge on part of County Road 715 (Fv. 715) in Trøndelag County, Norway. The bridge foundations were planned to be placed in an area at high risk for quick clay slides. The soil layering at the site was not as regular as that at the NGTS site; the common sequence was a stiff upper layer followed by clay on top of a thick quick clay layer, and subsequent to a certain depth, there appeared clay or stiffer soil layers.

The location of the tests is shown in Fig. 4 (p.448), while the layering and water table depth are shown in Fig. 5 (p.449)(a hydrostatic condition was assumed for the groundwater). It is important to note that in this case, the layering was proposed by the authors based on the information from the site data report by Statens Vegvesen (2013), which included laboratory and field tests. The summary of the tests is shown in Figs. 6 and 7 (p.449).

3 Data processing

The CPTu data were received as raw files in ".cpt" format, comprising measurements of the depth, tip resistance (q_c), f_s, and pore pressure behind the cone (u_2). The tip resistance value was corrected from the effects of the pore pressure acting at the conical tip using the following formula: $q_t = q_c + (1 + r_a) \cdot u_2$, where r_a is the net area ratio dependent on the probe design and q_t is the corrected tip resistance. The normalized parameters were then computed according to the following equations.

For normalized cone resistance (Q_t),

$$Q_t = \frac{q_t - \sigma_{v0}}{\sigma'_{v0}}; \tag{1}$$

for normalized friction ratio (F_r),

$$F_r = \frac{f_s}{q_t - \sigma_{v0}}; \tag{2}$$

for pore pressure ratio (B_q),

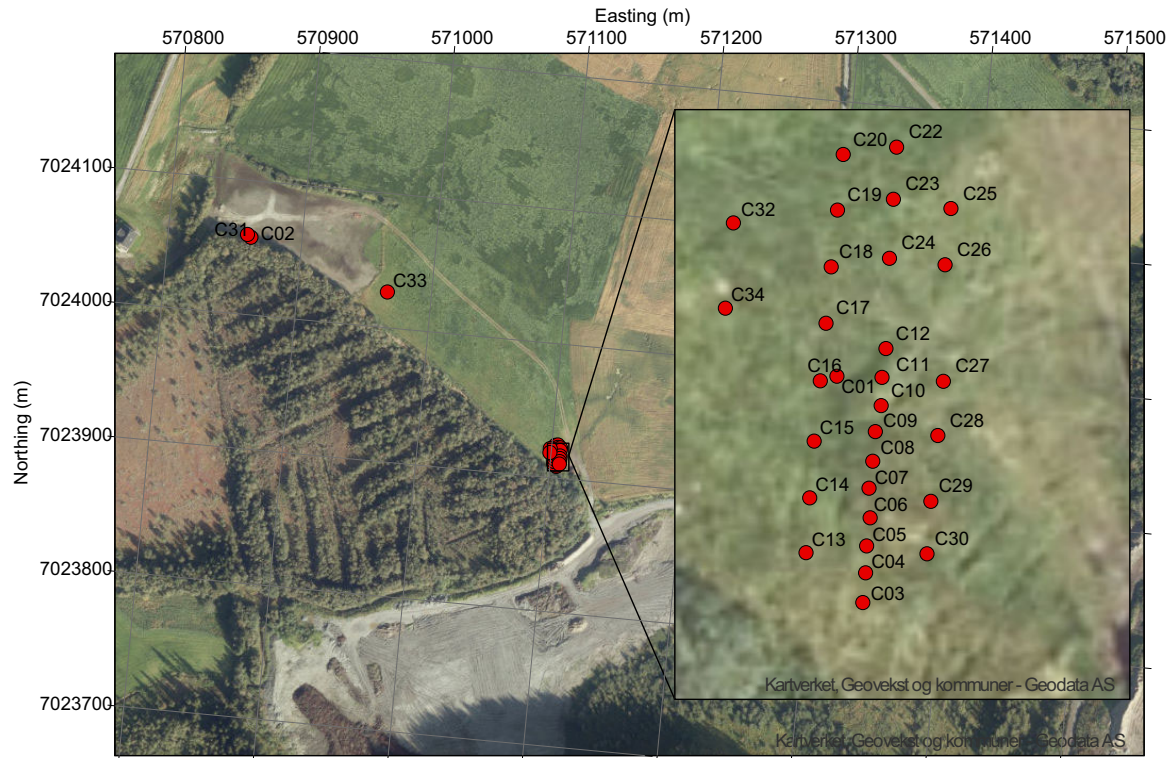$$B_q = \frac{u_2 - u_0}{q_t - \sigma_{v0}} = \frac{\Delta u_2}{q_t - \sigma_{v0}}; \tag{3}$$
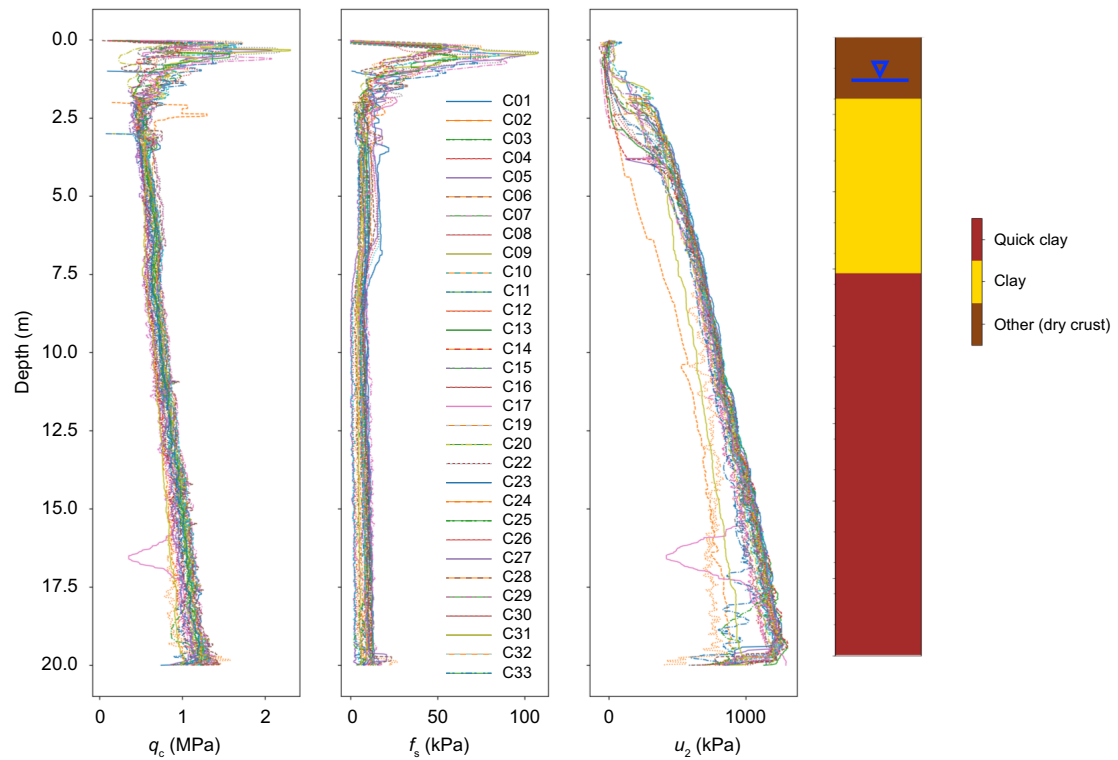
Fig. 1 CPTu layout at NGTS site

Fig. 2 CPTu summary and layering and water table depth at NGTS site. The plots show the q_c, f_s, and u_2 over the depth. References to color refer to the online version of this figure
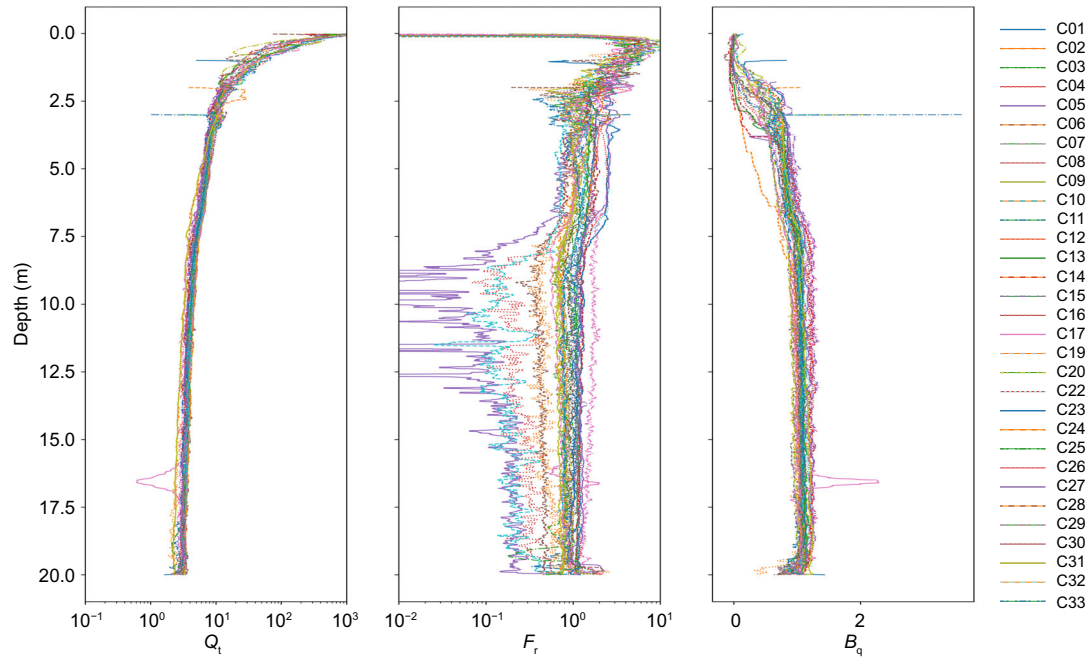
Fig. 3 Summary of CPTu at NGTS site in terms of normalized parameters. References to color refer to the online version of this figure
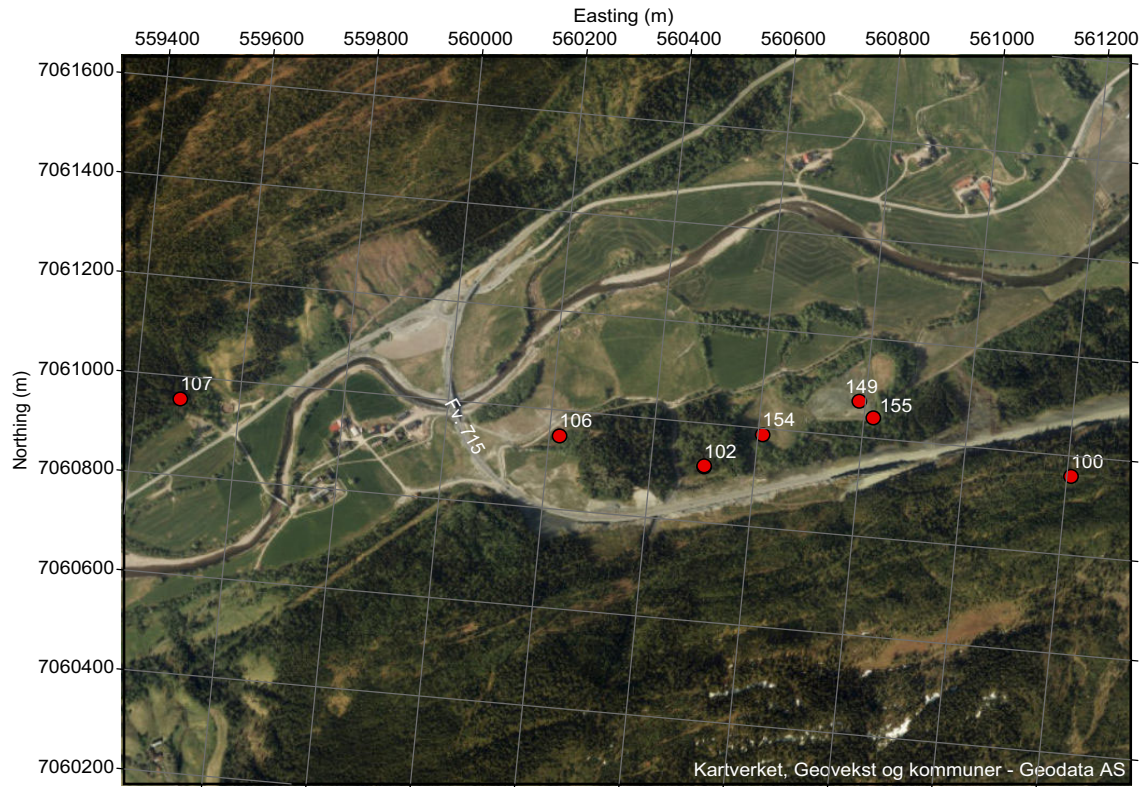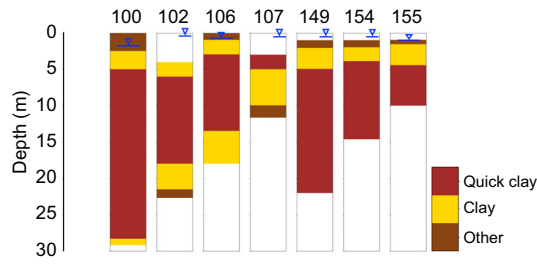

Fig. 4 CPTu layout at Vegvesen site

for normalized excess pore pressure (U_2),

$$U_2 = \frac{\Delta u_2}{\sigma'_{v0}}, \tag{4}$$

where σ_{v0} and σ'_{v0} are the in-situ total and effective stresses, respectively, u_0 is the in-situ pore pressure,

Fig. 5 Layering and water table depth at Vegvesen site. References to color refer to the online version of this figure

and $\Delta u_2 = u_2 - u_0$ is the excess u_2.

The data were inspected for abnormal measurements (e.g. negative f_s) and smoothed using a median statistical filter, as recommended by Wickremesinghe (1989), to remove unwanted spikes.

In the present study, for the machine learning classification, logarithmic transformations were performed over the normalized parameters to fit the data in the (0, 1) range. U_2 was preferred over B_q, as it is a better pore pressure parameter for soil type identification according to Schneider et al. (2008). The transformed normalized parameters, Q_t^{norm}, F_r^{norm}, and U_2^{norm}, are presented below:

$$Q_t^{\mathrm{norm}} = \frac{2 \cdot \lg(Q_t) + 1}{9}, \tag{5}$$

$$F_r^{\mathrm{norm}} = \frac{\lg(F_r) + 4.5}{6}, \tag{6}$$

$$U_2^{\mathrm{norm}} = \frac{\lg(U_2 + 40)}{2.2}. \tag{7}$$

To have a point with which to compare the machine learning approach, the classification was first performed using well-known charts that consider sensitive soils in their classification schemes. Charts used were those recommended by Robertson (1990, 2016), Eslami and Fellenius (1997), Schneider et al. (2008), and Gylland et al. (2017). The metric used to evaluate the accuracy was the accuracy score (A.S.), defined as follows:

$$\mathrm{A.S.} = \frac{\mathrm{Number\ of\ correctly\ classified\ samples}}{\mathrm{Total\ number\ of\ samples}}. \tag{8}$$

As this part of the work was focused on predicting the appearance of highly sensitive and quick clays from the CPTu measurements, only three soil classes were considered: sensitive, clayey, and other (coarser or stronger). The classification results were consequently adjusted to measure the classification accuracy.
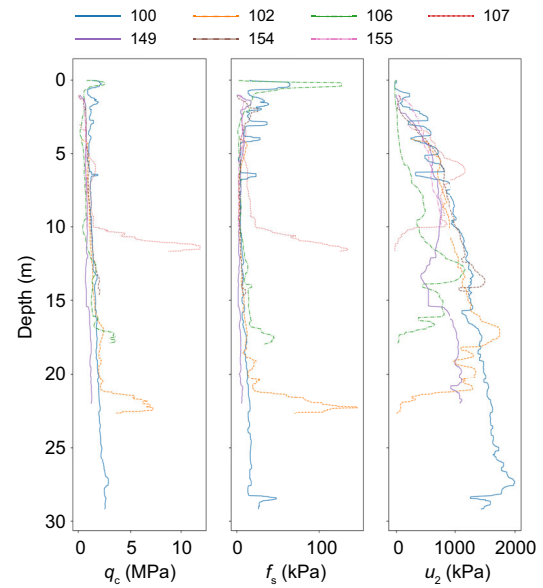
Fig. 6 Summary of CPTu at Vegvesen site in terms of the test measurements. References to color refer to the online version of this figure
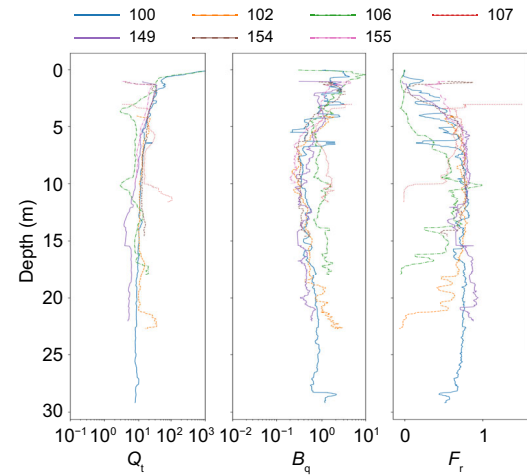
Fig. 7 Summary of CPTu at Vegvesen site in terms of the normalized parameters. References to color refer to the online version of this figure

3.1 Robertson (1990)

This classification chart is based on an extensive database of CPTus and uses the three normalized parameters introduced above to define soil behavior types. The soil behavior types are defined as: (1) sensitive and fine-grained soils; (2) organic soils and peat; (3) clays to silty clays; (4) silt mixtures; (5) sand mixtures; (6) sands; (7) gravelly sands to sands; (8) very stiff sands to clayey sands; (9) very stiff fine-grained soils.

The results of using this classification chart are shown in Figs. 8 and 9. For comparison, in this study,
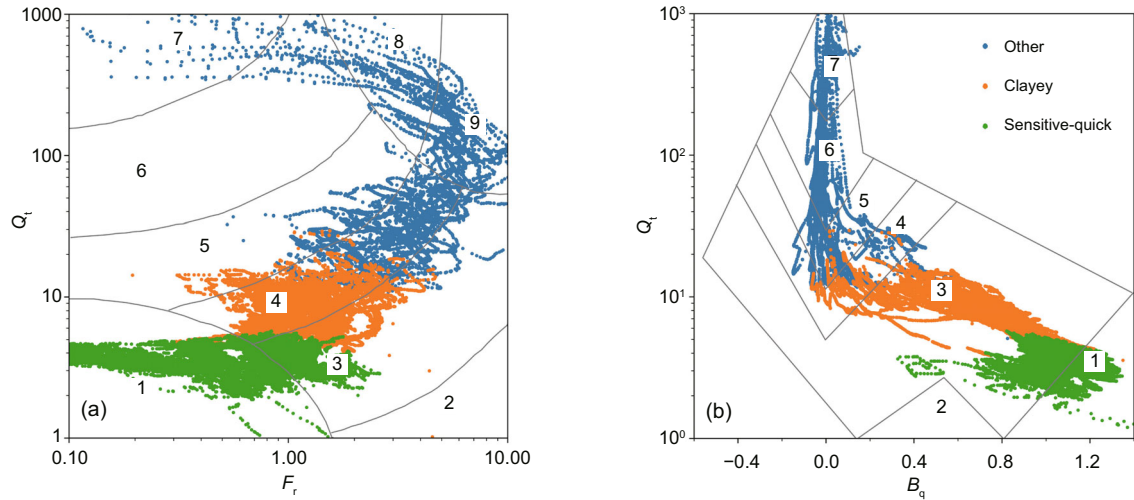
Fig. 8 Classification in (Robertson, 1990) showing the NGTS dataset: (a) Q_t-F_r; (b) Q_t-B_q. References to color refer to the online version of this figure
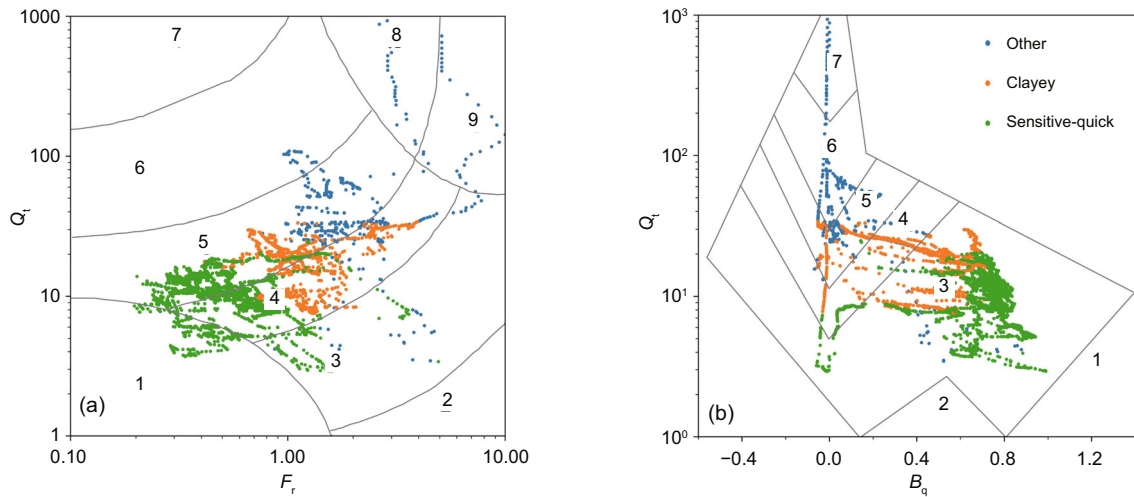

Fig. 9 Classification in (Robertson, 1990) showing the Vegvesen dataset: (a) Q_t-F_r; (b) Q_t-B_q. References to color refer to the online version of this figure

soil class 1 was considered as quick and sensitive clay, soil classes 3 and 4 as clayey, and soil classes 2, 5, 6, 7, 8, and 9 as other. The accuracy scores of this classification using the study datasets are shown below:

$$\begin{aligned} \text{NGTS}: \ &\text{A.S.} = 70\%, \ \text{for } Q_t\text{-}F_r \text{ chart}, \\ &\text{A.S.} = 53\%, \ \text{for } Q_t\text{-}B_q \text{ chart}; \\ \text{Vegvesen}: \ &\text{A.S.} = 28\%, \ \text{for } Q_t\text{-}F_r \text{ chart}, \\ &\text{A.S.} = 27\%, \ \text{for } Q_t\text{-}B_q \text{ chart}. \end{aligned}$$

From the chart, it is evident that the classification results for the Vegvesen dataset had a low accuracy score because of the high Q_t of the site's sensi-

tive clays compared with the zone defined by Robertson (1990). However, the results for the NGTS site showed better agreement with the chart, especially the Q_t-F_r plot.

3.2 Eslami and Fellenius (1997)

This classification chart was developed when investigating the use of cone penetration test (CPT) in pile design using data from 20 sites in five countries. In this case, the "effective" cone resistance and f_s values are used instead of the normalized ones. The effective cone resistance is defined as $q_E = q_t - u_2$. The chart defines five classes: (1) sensitive and

collapsible clay and/or silt; (2) clay and/or silt; (3) silty clay and/or clayey silt; (4) sandy silt and/or silty sand; (5) sand and/or sandy gravel.

The results of using chart in (Eslami and Fellenius, 1997) with the datasets in this study are shown in Figs. 10a and 10b. Better agreement was observed in the identification of sensitive soils for the Vegvesen dataset compared with Robertson (1990). In both datasets (though clearer in that of NGTS), it was possible to see a major overlap between the clayey and quick clay soils. The accuracy scores were 63% for NGTS and 74% for Vegvesen.

3.3 Schneider et al. (2008)

The work performed by Schneider et al. (2008) focused on improving the simple classification charts available at that time to consider the effects of undrained penetration on penetration resistance. The chart was plotted on a Q_t-U_2 space. The database used in this study included sensitive soils from Norway and Canada. The classification chart is divided into five different zones: (1a) silts and low-rigidity-index (I_r) clays, (1b) clays, (1c) sensitive clays, (2) essentially drained sands, and (3) transitional soils.

Figs. 10c and 10d show both datasets plotted on the Q_t-U_2 space. It was observed that the sensitive clays from the Vegvesen dataset showed a behavior closer to that predicted by the scheme, with an accuracy score of 75%. Meanwhile, in the NGTS dataset, there were more cases of "false positives" meaning that a large fraction of the clay layer was classified as sensitive when it was not; however, the accuracy score was still 75% as well.

3.4 Robertson (2016)

This chart updates that proposed by Robertson (1990) by using behavior-based descriptions and an updated normalized tip resistance Q_{tn}. Seven zones are defined in this classification system: (1) clay-like contractive sensitive (CCS); (2) clay-like contractive (CC); (3) clay-like dilative (CD); (4) transitional contractive (TC); (5) transitional dilative (TD); (6) sand-like contractive (SC); (7) sand-like dilative (SD).

Fig. 11 shows both datasets plotted on the Q_{tn}-F_r space. Once again, the NGTS dataset presented a large fraction of the clayey soil as sensitive, but

the other two soil classes seem to fall well within the correct groups. Meanwhile, the Vegvesen dataset showed a large fraction of sensitive soils classifying as TC or CC. The accuracy score was 75% for NGTS, while 52% for Vegvesen.

3.5 Gylland et al. (2017)

This work proposes a classification chart specifically focused on the identification of sensitive clays and is based on tests performed in Norway using parameters following the same philosophy as Robertson (1990) but with a different normalization: N_{mc}, B_{q1}, and R_{fu}, as shown below:

$$N_{mc} = \frac{q_t - \sigma_{v0}}{\sigma'_A + a}, \tag{9}$$

$$B_{q1} = \frac{\Delta u_1}{q_t - \sigma_{v0}}, \tag{10}$$

$$R_{fu} = \frac{f_s}{\Delta u_1}, \tag{11}$$

where in the reference stress $\sigma'_A = \sigma'^m_c + \sigma'^{(1-m)}_{v0}$, σ'_c is the effective pre-consolidation stress, a is the attraction, m is the SHANSEP-framework exponent (typically between 0.7 and 0.8 for Norwegian clays), and Δu_1 is the excess pore pressure at the tip of the cone (u_1).

The main drawback of this scheme is that it is necessary to know parameters that are not necessarily associated with the CPT itself, e.g. the attraction and pre-consolidation stress. Moreover, it requires the knowledge of u_1, which is not usually measured, preferring u_2. Therefore, in this case, it was necessary to use correlation involving the measured parameters to estimate those from the model.

Figs. 12 and 13 (p.453) show both datasets plotted on the chart proposed by Gylland et al. (2017). It was evident that the NGTS dataset showed better agreement than the Vegvesen dataset. The accuracy score here was only a binary classification score due to the nature of the classification proposed by the authors. The results are summarized below:

$$\begin{aligned} \text{NGTS}: \text{A.S.} &= 86\%, \quad \text{for } N_{mc}\text{-}B_{q1} \text{ chart,} \\ \text{A.S.} &= 87\%, \quad \text{for } N_{mc}\text{-}R_{fu} \text{ chart;} \\ \text{Vegvesen}: \text{A.S.} &= 39\%, \quad \text{for } N_{mc}\text{-}B_{q1} \text{ chart,} \\ \text{A.S.} &= 40\%, \quad \text{for } N_{mc}\text{-}R_{fu} \text{ chart.} \end{aligned}$$

The Vegvesen dataset was plotted almost completely out of the red shaded area defining sensitive clays, demonstrating a different behavior of the sensitive clays present in the area compared with those that were part of the dataset used by the authors, which included the NGTS site at Tiller.
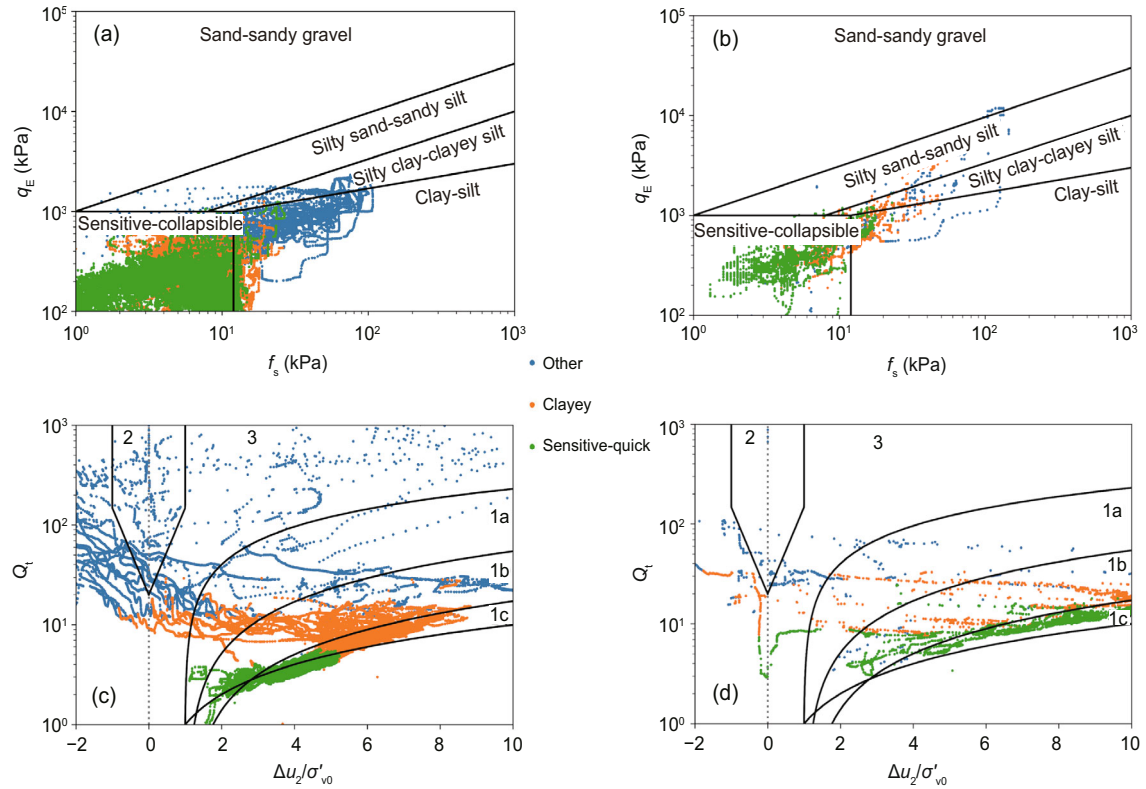

Fig. 10 Classification in (Eslami and Fellenius, 1997) showing the NGTS (a) and Vegvesen (b) datasets, and classification in (Schneider et al., 2008) showing the NGTS (c) and Vegvesen (d) datasets. References to color refer to the online version of this figure
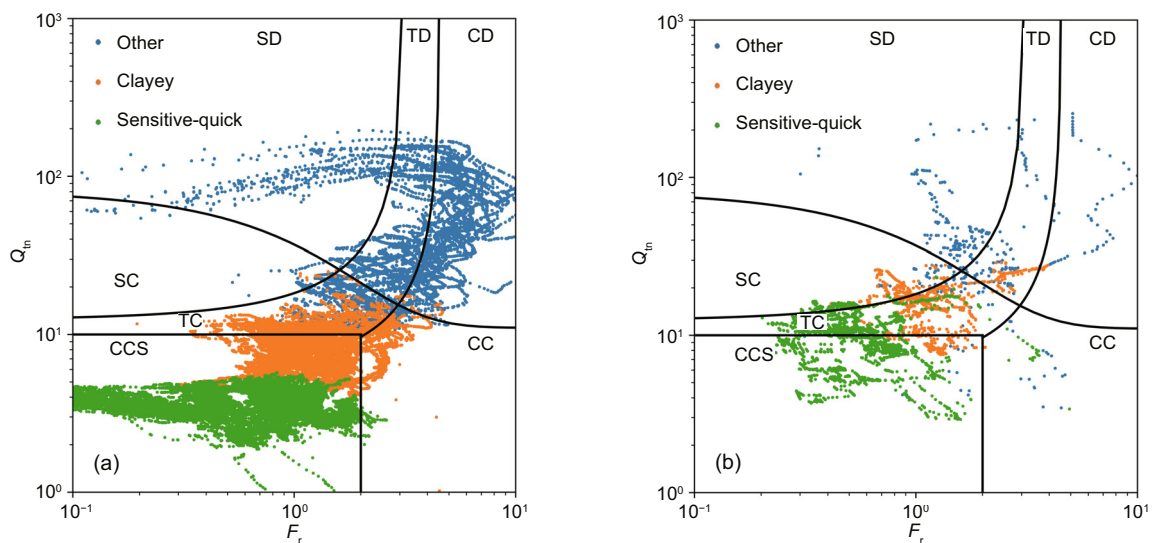

Fig. 11 Classification in (Robertson, 2016) showing the NGTS (a) and Vegvesen (b) datasets. References to color refer to the online version of this figure

4 Machine learning classification results

4.1 Machine learning classificators

The machine learning algorithms for the classification used in this work were logistic regression, Naive Bayes, and an HMM, as included in the Python libraries scikit-learn (Pedregosa et al., 2011) and hmmlearn (hmmlearn, 2010). Briefly, logistic regression uses a linear model to create a decision boundary that separates different classes, and the Naive Bayes approach uses a probabilistic framework based on Bayes' theorem, with simplification of the conditional independence of the data. Finally, an HMM uses Markov chains and a probabilistic framework to model the spatial correlation between measured data. The measurement of the model's performance was performed through the accuracy score introduced previously.

4.2 Results of the NGTS dataset

4.2.1 Logistic regression classifier

A logistic regression classificator was sequentially trained, and results of the predictions are shown in Fig. 14, using Q_t^{norm}-F_r^{norm}, Q_t^{norm}-U_2^{norm}, and the three parameters together as predictors.

Results showed a high accuracy score, even for the first estimation (using only one CPTu to train), with a sharp increase afterwards. It was observed that accuracy scores of at least 80% were reached upon using only four tests to train the classificator. It was also noted that the Q_t^{norm}-U_2^{norm} scheme showed higher accuracies with fewer data. Furthermore, the
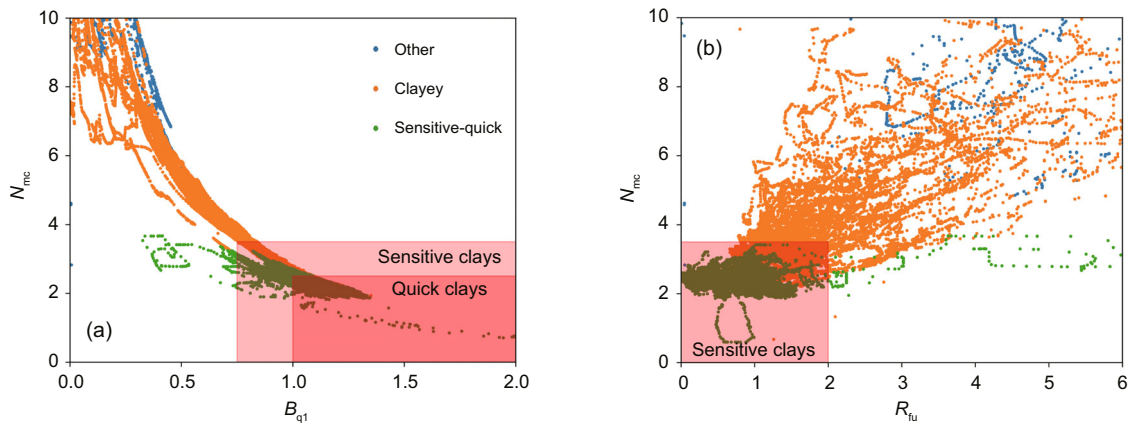

Fig. 12 Classification in (Gylland et al., 2017) showing the NGTS dataset: (a) N_{mc}-B_{q1}; (b) N_{mc}-R_{fu}. The red squares indicate the zones wherein sensitive and quick clay soils should be located. References to color refer to the online version of this figure

Fig. 13 Classification in (Gylland et al., 2017) showing the Vegvesen dataset: (a) N_{mc}-B_{q1}; (b) N_{mc}-R_{fu}. The red squares indicate the zones wherein sensitive and quick clay soils should be located. References to color refer to the online version of this figure
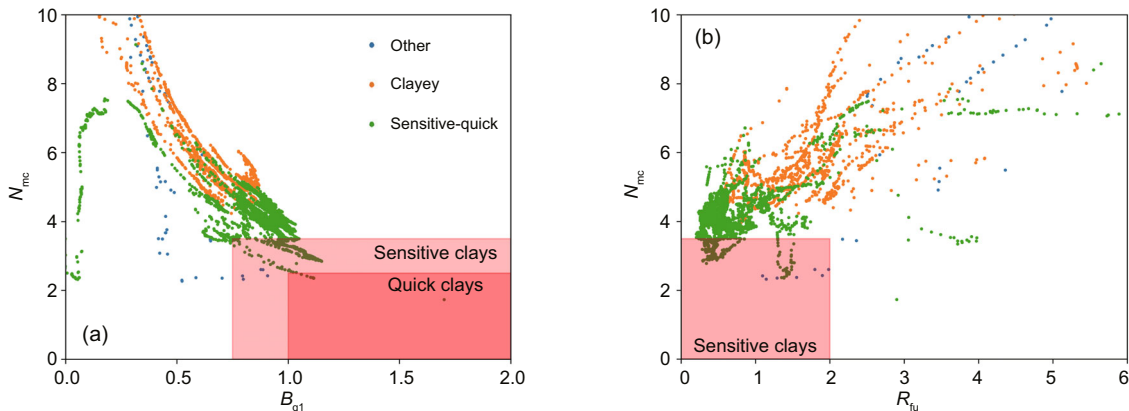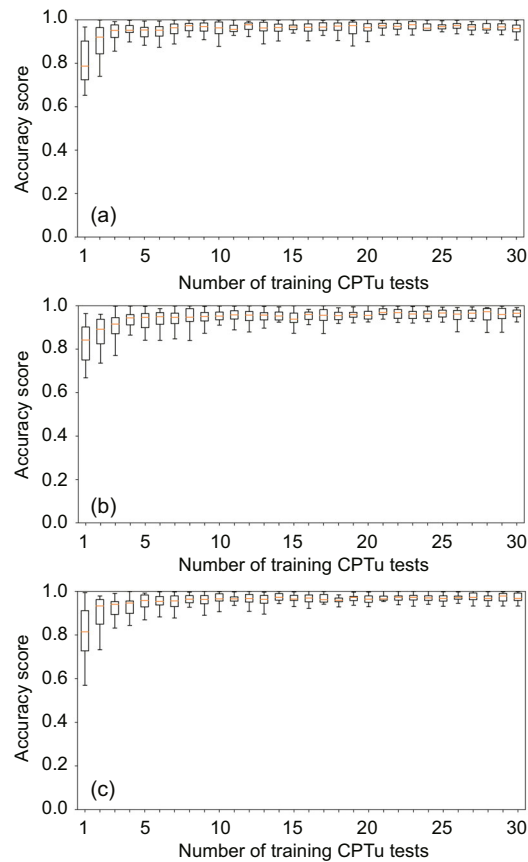
use of the three parameters to train the classificator demonstrated improved accuracies as well as less variability.

It is important to highlight that the NGTS site is a highly homogeneous site with a regular layering sequence and low dispersion of the measured parameters. These results should not be expected to occur in non-homogeneous sites.

4.2.2 Naive Bayes classifier

Results of using a Naive Bayes classificator on the NGTS dataset are shown in Fig. 15. It was observed that the results showed more scattering in the accuracy of the first estimation, especially when using F_r^{norm}, but quickly increased afterwards. The primary advantage of using the Naive Bayes approach compared with logistic regression is that the run time is around ten times less than that of the latter, which may make a major difference in large datasets.
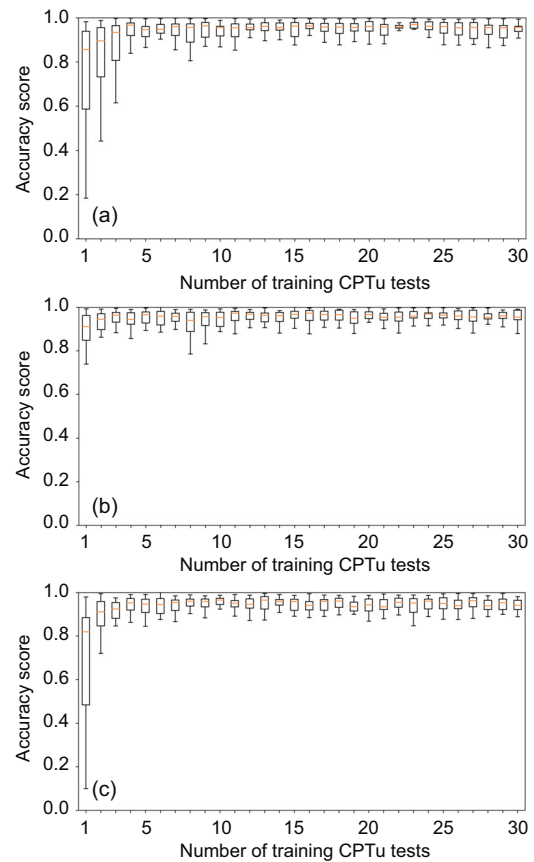
4.2.3 HMM

For this part, an HMM was trained in a semi-supervised manner. The model parameters (transition matrix, means, covariances, and starting probabilities) were estimated from the training data. Then, the model was allowed to update (optimize) the values of the transition matrix and covariances in the expectation-maximization stage, while the rest remained fixed. The Viterbi algorithm was used to determine the most likely sequence of states (soil classes). Sequential training was performed, but due to a restriction in the programmed code, it was only possible to use the CPTu test that defined the three soil classes. This restriction reduced the number of combinations available, but it was assumed that there were still enough to draw conclusions from. The results of the sequential training and classification are shown in Fig. 16.

It was observed that despite the prediction

Fig. 14 Results of a sequentially trained logistic regression classificator on the NGTS dataset using Q_t^{norm} and F_r^{norm} (a), Q_t^{norm} and U_2^{norm} (b), and the three parameters together (c) as predictors

Fig. 15 Results of a sequentially trained Naive Bayes classificator on the NGTS dataset using Q_t^{norm} and F_r^{norm} (a), Q_t^{norm} and U_2^{norm} (b), and the three parameters together (c) as predictors
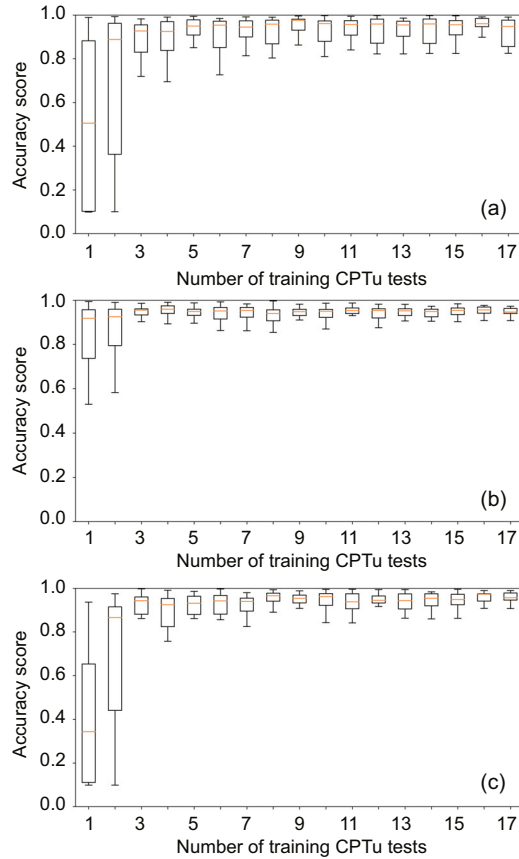
Fig. 16 Results of a sequentially trained HMM on the NGTS dataset using Q_t^{norm} and F_r^{norm} (a), Q_t^{norm} and U_2^{norm} (b), and the three parameters together (c) as predictors

Table 1 Comparison of the machine learning classification for the NGTS dataset in terms of the accuracy score

CPTu	Accuracy score		
	Logistic reg.	Naive Bayes	HMM
C01	96%	90%	76%
C02	89%	89%	91%
C03	98%	96%	90%
C04	95%	97%	96%
C05	94%	88%	46%
C06	97%	98%	97%
C07	96%	95%	95%
C08	99%	97%	88%
C09	99%	98%	96%
C10	96%	94%	95%
C11	98%	98%	96%
C12	95%	98%	99%
C13	96%	95%	90%
C14	95%	96%	98%
C15	93%	95%	95%
C16	94%	96%	99%
C17	96%	93%	50%
C19	97%	100%	97%
C20	90%	93%	94%
C22	97%	99%	77%
C23	96%	98%	99%
C24	96%	98%	99%
C25	94%	94%	98%
C26	96%	97%	98%
C27	95%	96%	98%
C28	97%	97%	91%
C29	95%	97%	88%
C30	96%	98%	92%
C31	86%	89%	92%
C32	97%	100%	99%
C33	96%	98%	99%
Median	96%	97%	96%
Mean error	4.7%	4.2%	8.8%

being less accurate with few data, it quickly increased, yielding results similar to those obtained using the other two methods. The advantage of using an HMM is that because it considers the likelihood of changing from one class (hidden state) to another, the predicted profiles do not have unrepresentative thin layers within one another. This can be seen in Fig. 17 for CPTus C04 and C10.

4.2.4 Site profiles

To visually compare the different classification methods, Figs. 17 and 18 display several selected CPTus with the actual site layering alongside the machine learning classification. The accuracy of each profile estimation is presented in Table 1.

For the profile estimation, only seven CPTus were used for the training to classify each test. The criteria used to select the training dataset were to sort the tests by name and use the seven closest to the CPTu to be classified.

The median accuracy scores were 96% for the logistic regression, 97% for the Naive Bayes, and 96% for the HMM. These values were quite similar to each other, but the mean error was considerably higher for the HMM (Table 1). A visual comparison showed that the HMM estimated profiles that were not close to reality, as in the case of C05, which may have been due to the training set chosen and could be improved via including more data in the training phase. For NGTS, the machine learning classification that performed best was the Naive Bayes, closely followed by the logistic regression. Table 2 shows the accuracy score for each soil type. In this case, the identification of quick and sensitive clays was higher than that of the others (which were still high, with accuracy scores over 80%). Table 2 also
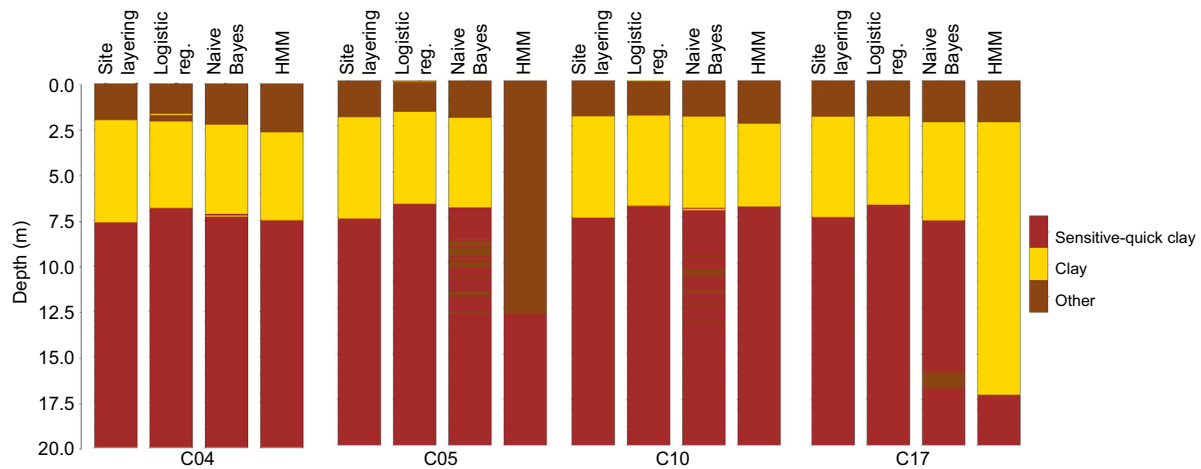
Fig. 17　Comparison of the machine learning classification for the NGTS dataset and the profile view of CPTus C04, C05, C10, and C17. References to color refer to the online version of this figure
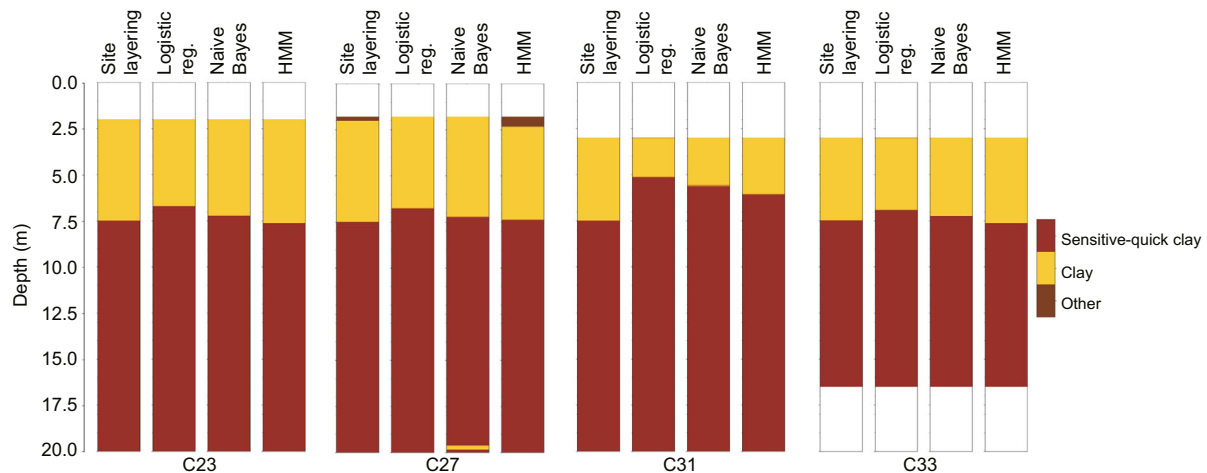

Fig. 18　Comparison of the machine learning classification for the NGTS dataset and the profile view of CPTus C23, C27, C31, and C33. References to color refer to the online version of this figure

Table 2　Comparison of the machine learning classification for the NGTS dataset in terms of the accuracy score for each soil class

Soil class	Accuracy score		
	Logistic reg.	Naive Bayes	HMM
Other	88%	94%	86%
Clay	91%	93%	84%
Sensitive-quick	100%	97%	84%

demonstrates the relatively worse performance of the HMM.

4.3　Results of the Vegvesen dataset

Since the Vegvesen dataset was much smaller than the NGTS dataset and the layering of the site was not as homogeneous, its machine learning classification was more challenging because in the process of splitting the dataset to train and test, a major share of the information was lost (it was not possible to use it to train). Thus, in this case, in addition to performing the sequential training and prediction, a "leave-one-group-out" cross-validation technique (Scikit-learn, 2019) was used to assess the performance of the classifiers. Here, a group was represented by a CPTu.

4.3.1　Logistic regression classifier

Results are shown in Fig. 19. More scattered behavior and less accuracy were observed in general compared with the NGTS site; however, given the fact that this dataset had fewer CPTus and the soil

layering was more complex, the results were good. Compared with the classification charts, the logistic regression classifier improved the results. This was even more evident when analyzing the cross-validation results shown in Table 3.
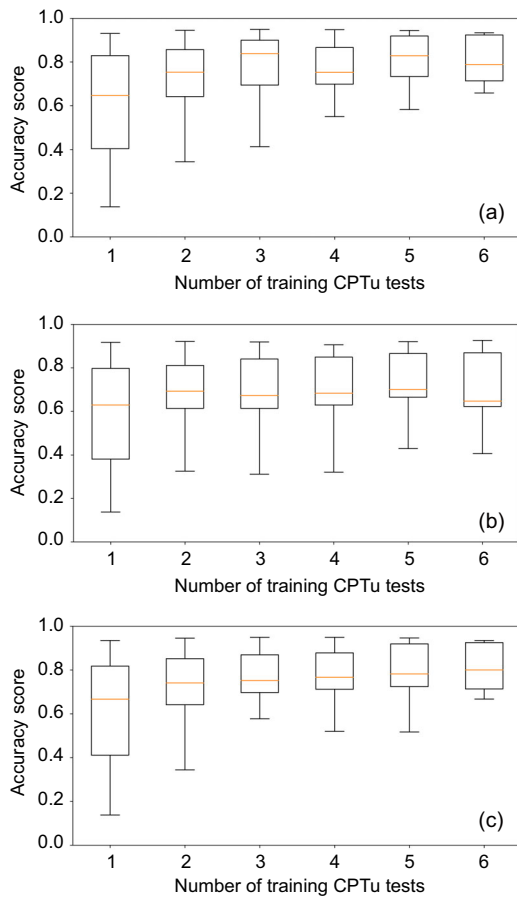
Fig. 19 Results of a sequentially trained logistic regression classificator on the Vegvesen dataset using Q_t^{norm} and F_r^{norm} (a), Q_t^{norm} and U_2^{norm} (b), and the three parameters together (c) as predictors

Table 3   Results of the leave-one-group-out cross-validation with logistic regression for the Vegvesen dataset

| "Left out"	Accuracy score		
CPTu	Q_t-F_r	Q_t-U_2	All
100	88%	87%	88%
102	71%	65%	71%
106	78%	69%	78%
107	66%	41%	67%
149	93%	85%	93%
154	92%	93%	93%
155	79%	62%	80%
Median	79%	69%	80%

4.3.2 Naive Bayes classifier

Results are shown in Fig. 20. As with the logistic regression, these results were more scattered and less accurate, but upon using the remaining six tests for the training, the accuracies increased considerably, as shown in Table 4. In the case of the Q_t-F_r classification, the results showed a median accuracy score of 91%, which was much more accurate than any of the classification charts.

4.3.3 HMM

The results of the sequential training and classification are shown in Fig. 21, while the results of the cross-validation are shown in Table 5.

The results showed high accuracies that were generally reached after using four tests to train the model.
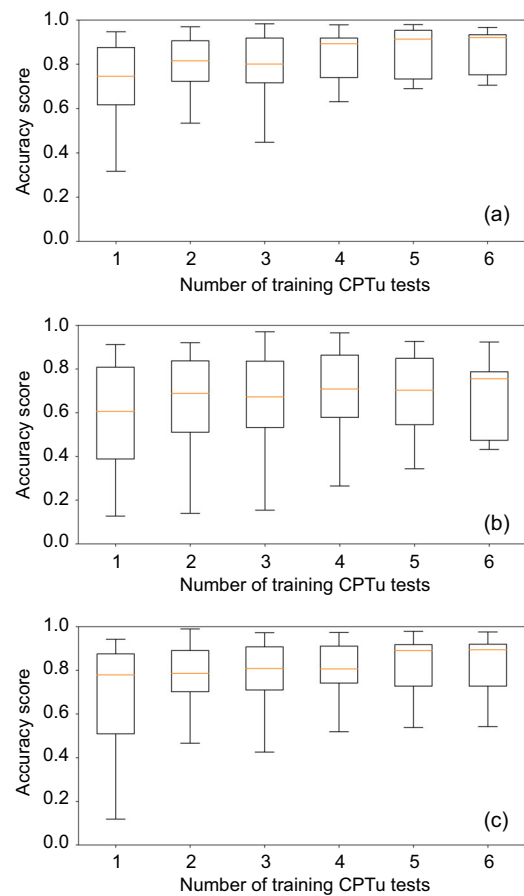
Fig. 20 Results of a sequentially trained Naive Bayes classificator on the Vegvesen dataset using Q_t^{norm} and F_r^{norm} (a), Q_t^{norm} and U_2^{norm} (b), and the three parameters together (c) as predictors

4.4 Site profiles

Results of the profiles estimated by the machine learning classification of the CPTus from Vegvesen dataset are shown in Figs. 22 and 23. The median accuracy scores were 79% for the logistic regression, 91% for the Naive Bayes, and 87% for the HMM. Here, as in the NGTS dataset, the Naive Bayes classification displayed the highest accuracy and the lowest mean error, as shown in Table 6.

Table 7 shows the accuracy scores for the classification of each type of soil considered. In this case,

there was a high accuracy for the highly sensitive and quick clay identification but a low accuracy for the other soil types. This may have been a reflection of the heterogeneity of the soil investigated compared with that of the NGTS dataset. Here, under the label "Other", there might have been a more varied range of soils.

5 Discussion

In general, soil classification charts meant for broad use fail to capture special soils like quick or

Table 4 Results of the leave-one-group-out cross-validation with Naive Bayes for the Vegvesen dataset

"Left out"	Accuracy score		
CPTu	Q_t-F_r	Q_t-U_2	All
100	92%	92%	90%
102	75%	79%	78%
106	71%	47%	54%
107	72%	43%	73%
149	97%	87%	98%
154	91%	67%	89%
155	93%	76%	92%
Median	91%	76%	89%

Table 5 Results of the leave-one-group-out cross-validation with the HMM for the Vegvesen dataset

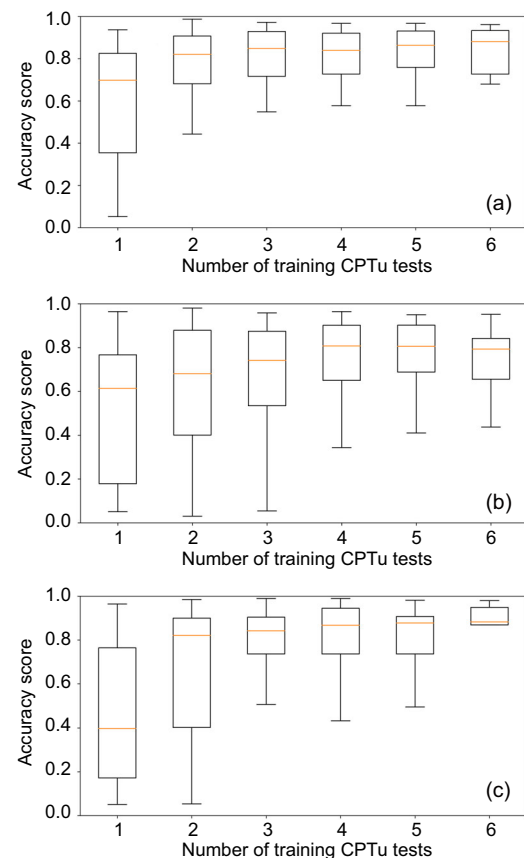| "Left out"	Accuracy score		
CPTu	Q_t-F_r	Q_t-U_2	All
100	93%	89%	91%
102	87%	85%	87%
106	70%	18%	30%
107	71%	81%	77%
149	67%	44%	48%
154	90%	68%	91%
155	95%	92%	96%
Median	87%	81%	87%

Table 6 Comparison of the machine learning classification for the Vegvesen dataset in terms of the accuracy score

CPTu	Accuracy score		
	Logistic reg.	Naive Bayes	HMM
100	88%	92%	93%
102	71%	75%	87%
106	78%	71%	70%
107	66%	72%	71%
149	93%	97%	67%
154	92%	91%	90%
155	79%	93%	95%
Median	79%	91%	87%
Mean error	19.0%	15.6%	18.1%

Fig. 21 Results of a sequentially trained HMM on the Vegvesen dataset using Q_t^{norm} and F_r^{norm} (a), Q_t^{norm} and U_2^{norm} (b), and the three parameters together (c) as predictors

Table 7 Comparison of the machine learning classification for the Vegvesen dataset in terms of the accuracy score for each soil class

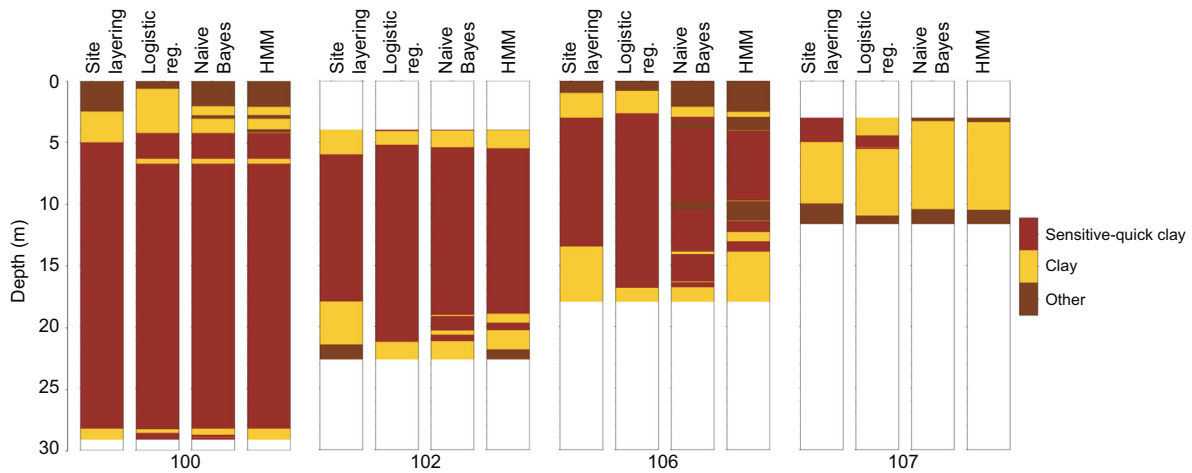| Soil class	Accuracy score		
	Logistic reg.	Naive Bayes	HMM
Other	24%	60%	77%
Clay	59%	65%	77%
Sensitive-quick	98%	95%	84%

Fig. 22 Comparison of the machine learning classification for the Vegvesen dataset and the profile view of CPTus 100, 102, 106, and 107. References to color refer to the online version of this figure
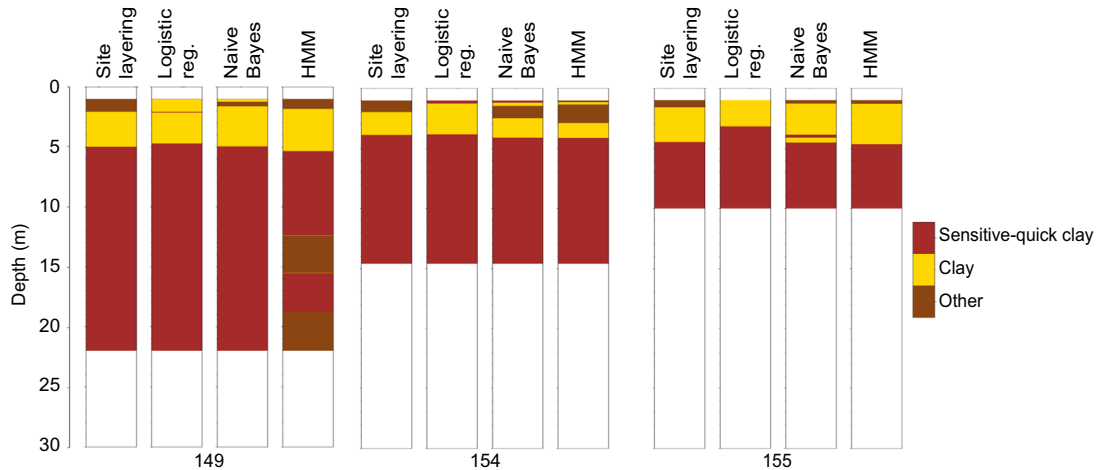

Fig. 23 Comparison of the machine learning classification for the Vegvesen dataset and the profile view of CPTus 149, 154, and 155. References to color refer to the online version of this figure

sensitive clays, especially Norwegian sensitive clays, which have distinctive features compared with other soils with similar behavior (e.g. Canadian quick clays). Furthermore, such classification charts use each data point individually without considering the spatial correlation between soils in the same layer.

This study used three machine learning approaches to study a methodology to improve the sensitive soil determination of CPTus: logistic regression, Naive Bayes, and HMM classifications.

The approaches were tested in two different datasets comprising soils of different characteristics. In the case of the NGTS site, wherein the soil layering was regular and the layer characteristics were homogeneous, the three methods displayed an excellent

performance, as measured by an accuracy score well above 90%. Although the three methods had a high accuracy, that of the HMM was slightly lower than the other two, particularly in two profiles, C05 and C17, as shown in Table 1 and Fig. 17. Advantages of an HMM are that it considers the spatial connection between the data and can be noticed in less presence of thin unrepresentative layers. The Vegvesen dataset had fewer data available to train the models, but the results were still quite favorable, with accuracy scores above 80%. The sequential training of the models graphically demonstrated how they learned from the data as it was incorporated. Both datasets showed that using only four tests to train the model significantly improved the accuracy of the classification. This could be helpful when performing site

investigations wherein the model can be trained as the data are retrieved from the soundings and laboratory tests or even when in-place visual classification is performed. Once the model is successfully trained, the necessity of laboratory tests might be reduced. This could work even in less homogeneous datasets, such as that from Vegvesen. Since this study primarily focused on comparing different classification approaches and discussing the advantages of following a machine learning classification approach, a simple performance measure was used: the accuracy score. It is recommended that a better performance measure, e.g. the balanced accuracy score or the F1 score (Scikit-learn, 2019), should be used for further research on this topic. It would also be interesting to explore other more advanced machine learning classification approaches, such as ordinal classifications or neural networks. Additionally, for further research on this topic, more data representing a broader range of soil properties are required, with clear identification of clays as sensitive, highly sensitive, or quick. Moreover, the authors encourage the incorporation of recent models aiming at identifying quick clays. A good example of such a study is in (Valsson, 2016).

6 Conclusions

In this study, a set of Python libraries were used to train three different machine learning algorithms: logistic regression, Naive Bayes, and an HMM. Results demonstrated the abilities of these methods to learn from the data, with good classification accuracies reached after only four training CPTus. However, these methods are not meant to be used as general classification solutions unless they are trained with a large dataset that includes different soil conditions to be detected. The major challenge here involved obtaining enough data to train the models as well as enough laboratory results to verify them.

In the future, it would be interesting to keep researching these methodologies and their applications in the geotechnical field since they have proven to yield good results that can aid engineers in optimizing both field and laboratory tests.

Contributors

Vikas THAKUR and Ivan DEPINA designed the research. Cristian GODOY processed the corresponding data and wrote the first draft of the manuscript. Ivan DEPINA helped to organize the manuscript. Cristian GODOY and Ivan DEPINA revised and edited the final version.

Conflict of interest

Cristian GODOY, Ivan DEPINA, and Vikas THAKUR declare that they have no conflict of interest.

References

Eslami A, Fellenius BH, 1997. Pile capacity by direct CPT and CPTu methods applied to 102 case histories. *Canadian Geotechnical Journal*, 34(6):886-904. https://doi.org/10.1139/t97-056

Gylland AS, Sandven R, Montafia A, et al., 2017. CPTU classification diagrams for identification of sensitive clays. *In*: Thakur V, L'Heureux JS, Locat A (Eds.), Landslides in Sensitive Clays. Springer, Cham, Switzerland, p.57-66. https://doi.org/10.1007/978-3-319-56487-6_5

hmmlearn, 2010. hmmlearn: Unsupervised Learning and Inference of Hidden Markov Models. hmmlearn. https://hmmlearn.readthedocs.io

L'Heureux J, Lindgård A, Emdal A, 2019. The Tiller-Flotten Research Site: Geotechnical Characterization of a Very Sensitive Clay Deposit. Technical Report No. 20160154-20-R. Norwegian Geotechnical Institute, Norway.

Lunne T, Robertson PK, Powell JJM, 1997. Cone Penetration Testing in Geotechnical Practice. Blackie Academic and Professional, London, UK.

NGI (Norwegian Geotechnical Institute), 2019. NGTS-Norwegian Geo-test Sites. NGI. https://www.ngi.no/eng/Projects/NGTS-Norwegian-Geo-Test-Sites

Pedregosa F, Varoquaux G, Gramfort A, et al., 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12:2825-2830.

Robertson PK, 1990. Soil classification using the cone penetration test. *Canadian Geotechnical Journal*, 27(1):151-158. https://doi.org/10.1139/t90-014

Robertson PK, 2016. Cone penetration test (CPT)-based soil behaviour type (SBT) classification system–an update. *Canadian Geotechnical Journal*, 53(12):1910-1927. https://doi.org/10.1139/cgj-2016-0044

Schneider JA, Randolph MF, Mayne PW, et al., 2008. Analysis of factors influencing soil classification using normalized piezocone tip resistance and pore pressure parameters. *Journal of Geotechnical and Geoenvironmental Engineering*, 134(11):1569-1586. https://doi.org/10.1061/(ASCE)1090-0241(2008)134:11(1569)

Scikit-learn, 2019. Scikit-learn User Guide. Release 0.21.2. Scikit-learn. https://scikit-learn.org/dev/versions.html

Statens Vegvesen, 2013. County Road 715 Keiseräs-Olsøy, Parcel: Leksvik Border-Olsøy. Data Report Nr. 2012039995-009/Ud925Ar09. Rissa, Norway (in Norwegian).

Valsson SM, 2016. Detecting quick clay with CPTu. Proceedings of the 17th Nordic Geotechnical Meeting: Challenges in Nordic Geotechnics.

van Rossum G, 1995. Python Reference Manual. Technical Report No. CS-R9525. Centrum Wiskunde & Informatica, Amsterdam, The Netherlands.

Wickremesinghe DS, 1989. Statistical Characterization of Soil Profiles Using in Situ Tests. PhD Thesis, University of British Columbia, Vancouver, Canada. https://doi.org/10.14288/1.0062495

中文概要

题 目：应用机器学习方法从静力触探结果中识别快黏土和高灵敏度黏土

目 的：研究机器学习技术在利用孔压静力触探测试（CPTu）识别高灵敏度黏土和快黏土的潜力。

创新点：1. 成功应用机器学习方法从CPTu结果中分类出高灵敏度黏土和快黏土，并将结果与不同地点的实际土层进行了比较。2. 通过对机器学习算法的多次训练确定了可以获得良好结果的最少CPTu个数。

方 法：1. 基于对两个位置已知和土层确定的CPTu数据集的分析，使用3种机器学习图像分类方法（逻辑回归、朴素贝叶斯和隐藏马尔科夫模型）将CPTu数据用于样本分类。2. 将结果与实际土层进行比较，识别高灵敏度黏土和快黏土，并从计算性能度量方面比较3个方法的优缺点。

结 论：仅采用4个CPTu训练样本便可获得基于逻辑回归、朴素贝叶斯和隐藏马尔科夫模型的识别高灵敏度黏土和快黏土的3个分类模型，且分类精度良好。

关键词：机器学习；分类；快黏土；高灵敏度黏土