# Impact of CPT Feature Selection on Random Forest Performance for Soil Behavior Type Classification

**Ivanka Lee**

**Abstract** Accurate soil classification is essential for safe and economical geotechnical design such as in foundation selection, slope stabilization, and earthwork construction. Cone penetration testing (CPT) provides continuous resistance profiles, but conventional Soil Behavior Type (SBT) charts can be ambiguous near class boundaries. This study evaluates how progressive input feature sets affect a hyperparameter-optimized Random Forest classifier for normalized SBT (SBTn) prediction. A public dataset of 1,010,153 cleaned CPT datapoints from Austria and southern Germany was stratified, sampled, and divided into training and test subsets. Four feature sets were compared using accuracy, macro-averaged F1-score, and confusion matrices. The baseline set of depth, cone resistance qc, and sleeve friction fs achieved an accuracy of 0.9554 and a macro-averaged F1-score of 0.9397. Adding friction ratio Rf improved accuracy and macro-averaged F1-score metrics to 0.9648 and 0.9544 respectively. Adding the soil behavior type index Ic produced the best performing results with an accuracy of 0.9923 and a macro-averaged F1-score of 0.9845. However, Ic is calculated from the same normalized parameters used to assign SBTn labels. Feature sets that only use measured or simply derived quantities, such as depth, qc, fs, and Rf aremore informative for models trained on raw CPT measurements. Lastly, adding explicit overburden stresses provided no further improvement. The outcome of this study recommends a feature combination of depth, qc, fs, Rf, and Ic to assign Robertson SBTn classes automatically when Ic is already available. For the purposes of predicting soil classes from raw CPT measurements without Ic, this study recommends the feature combination of depth, qc, fs, and Rf.

## Introduction

Reliable soil classification is fundamental to geotechnical engineering because soil strength, stiffness, permeability, and deformation behavior govern the performance of foundations, embankments, retaining structures, and tunnels. Misclassification can produce unsafe or unnecessarily conservative designs, increase construction costs, and leave infrastructure
vulnerable to settlement, liquefaction, and slope failure (Lunne et al. 1997; Robertson 2009). The 2014 Oso landslide, shown in Figure 1, illustrates the scale of damage

that can accompany soil and slope instability (USGS 2014). Although CPT-based classification does not replace a complete site investigation, continuous subsurface profiles can reduce uncertainty and support more defensible engineering decisions.



**Figure 1. Oblique aerial view of the 2014 Oso landslide in Washington State, illustrating the scale of soil-slope failure and infrastructure exposure. (USGS 2014)**

In regions of Austria and southern Germany, the soil is primarily shaped by the Quaternary glaciation. This causes thick sequences of fine-grained lacustrine and fluvial sediment to be deposited in basins. Such basins include the Vienna Basin, Gastein Valley, and other Alpine foreland and intramontane basins (Oberhollenzer et al. 2021). As a result, the stratigraphy of the region is highly variable as it consists of interlayered clays, silts, sands, and gravels. Therefore, reliable and continuous soil classification is particularly valuable for engineering projects in these geological settings.

Traditional Approaches for Soil Classification. Laboratory-based systems, such as the Unified Soil Classification System and the European standard EN ISO 14688, classify soils according to grain-size distribution and plasticity characteristics (ASTM 2017). While these
methods are accurate when high-quality samples are available, such methods require
laboratory testing. Such work may be time-consuming and expensive, especially when continuous site characterization is necessary. Alternatively, in-situ methods, such as the Standard Penetration Test, are still widely used but offer relatively low resolution. Other methods such as geophysical techniques can cover large areas but generally lack the local mechanical detail needed for foundation design. Hence, the Cone Penetration Test (CPT) was chosen for this study as it is an informative and suitable in-situ tool for this study that produces near-continuous profiles of penetration resistance with depth (Lunne et al. 1997).

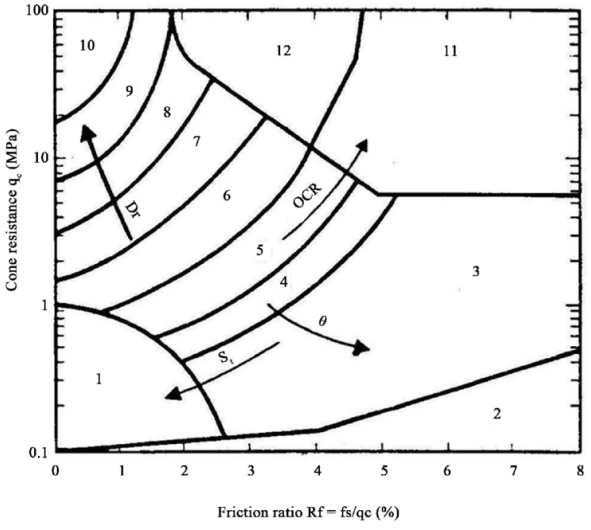



**Figure 2. Robertson SBT classification non-normalized chart (Bosch and Sotelo 2015; Robertson et al. 1986)**

While the original non-normalized SBT classification

ISSN 3106-7018 (Print) ISSN 3106-5791 (Online)

system had 12 classes, Robertson later introduced a normalized type, abbreviated as SBTn, which considers overburden stress and reduces the number of classes to nine as seen in Figure 3 (Robertson 1990; Robertson 2010; Robertson 2016).

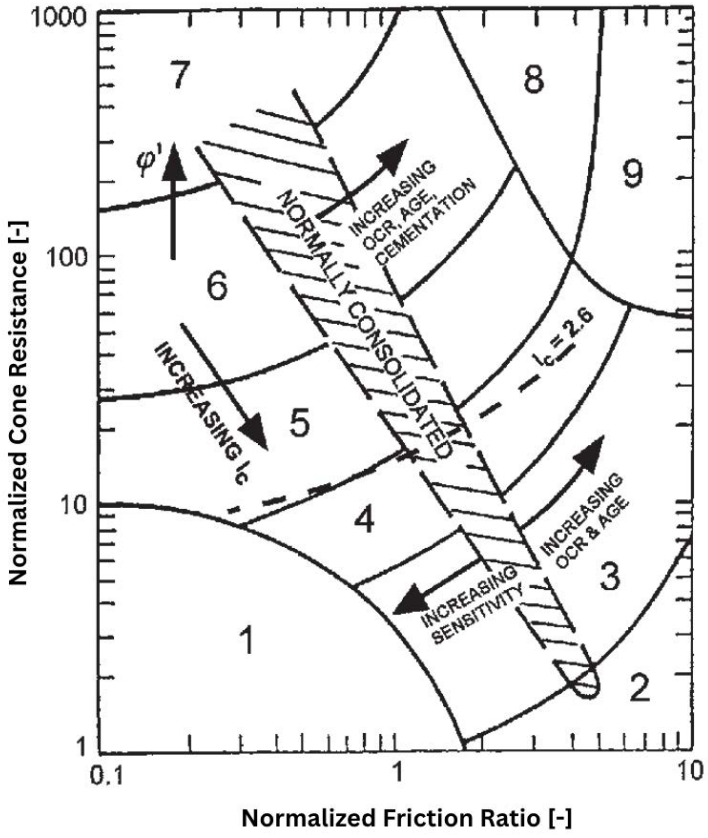



**Figure 3. Robertson SBT normalized chart (Robertson 1990)**

Normalized parameters improve consistency with depth and are now widely used in practice. The nine SBTn zone descriptions are the following: sensitive fine-grained, clay-organic soil, clays, silt mixtures, sand mixtures, sands, dense sand to gravelly sand, stiff sand to clayey sand, and stiff fine-grained. SBTn classes 8 and 9 are overconsolidated or cemented, meaning the soil is significantly stiffer and stronger than a normally consolidated soil of the same composition (Choo et al. 2018). A more detailed description and numerical assignment of the SBTn classes are shown in Table 1.

**Table 1. Normalized Soil Behavior Type (SBTn) classes (Robertson 2010, 2016)**

| SBTn Classes | SBT description |
|---|---|
| 1 | Sensitive fine-grained |
| 2 | Clay-organic soil |
| 3 | Clays: clay to silty clay |
| 4 | Silt mixtures: clayey silt & silty clay |
| 5 | Sand mixtures: silty sand to sandy silt |
| 6 | Sands: clean sands to silty sands |
| 7 | Dense sand to gravelly sand |
| 8 | Stiff sand to clayey sand* |
| 9 | Stiff fine-grained* |

*Overconsolidated or cemented

Electrical Cone Penetration Test and Measured Parameters. The electrical Cone Penetration Test hydraulically advances an instrumented electronic steel cylindrical cone vertically into the ground at a constant speed of 20 mm/s as seen in Figure 4 (Lunne et al., 1997; ASTM 2020).



**Figure 4. Electrical cone penetrometer positioned for field testing. (Kennedy 2022)**

A sensor continuously acquires data on the cone tip resistance qc and sleeve friction fs. qc is related to the strength and density of the soil, and a relatively large value generally indicates a denser material, such as sand and gravel. fs is the soil resistance measured along the cylindrical shaft of the cone (Rajapakse 2016).

From these two measurements, the friction ratio, Rf, is calculated in Equation 1:

| | |
|---|---|
| □!=□"<br>□# | × 100% |

**Equation 1. Rf, Friction ratio. (Robertson 1990)**

Friction ratio is particularly useful for distinguishing fine-grained from coarse-grained soils (Lunne et al. 1997; Robertson 2015). In the normalized SBTn framework, additional derived parameters are required for analysis. The soil behavior type index, Ic, is defined in Equation 2 (Robertson 2009, 2016), where Qtn is the normalized cone resistance and Fr is the normalized friction ratio.

□#= *(3.47−log□$%)&+ (log□'+1.22)&

**Equation 2. Ic, Soil Behavior Type index (Robertson and Wride 1998)**

Total vertical stress, σv, at a given depth, z, is estimated from the unit weight of the soil, γ, and depth as seen in Equation 3.

□(= □□

**Equation 3: Total Vertical Stress (Terzaghi 1943)**

Additionally, effective vertical stress, σ) * , is obtained by subtracting the pore-water pressure, u0, from the total vertical stress, σv as demonstrated in Equation 4.

□(*= □( −□+

**Equation 4. □(*, Effective vertical stress (Terzaghi 1936)**

## LITERATURE REVIEW

Random Forest has demonstrated consistently strong performance across multiple CPT-based soil classification studies. Rauter and Tschuchnigg (2021) reported accuracies of 0.97-0.99 for Robertson SBTn classes when training on more than 1000 soundings. Chala and Ray (2023) obtained 0.9923 overall accuracy with Random Forest on a collection of 232 CPT datasets. Fatehnia et al. (2024) achieved an F1-score of 0.896 when predicting borehole-derived soil types from raw CPT measurements. These results indicate that ensemble tree methods are well suited to the non-linear and noisy nature of CPT data.

Some deficiencies and recommendations for Future Work are presented in the literature. Rauter and Tschuchnigg (2021) found that the model trained on Austrian data performed poorly in Dutch CPT soundings, and it was not geographically transferable. They also mentioned that class imbalance affects minority soil types and proposed that more sophisticated neural networks, such as Deep Neural Networks with more complex neural architectures and improved handling of rare classes, be used. Chala and Ray (2023) found that Artificial Neural Network models had problems with minority classes and proposed adding more work on class-balanced training and systematic feature selection. Fatehnia et al. (2024) have presented the application of raw measurements but have not systematically compared progressive feature sets. Overall, the above studies indicate that controlled experiments are needed to isolate the effects of individual CPT parameters.

## METHODOLOGY

Dataset Description and Origin. This study used the publicly available CPT dataset published by Premstaller Geotechnik with Graz University of Technology in 2021 (Oberhollenzer et al.

2021). The dataset comprises 1,339 cone penetration tests (CPT, CPTu, SCPT, and SCPTu) executed by Premstaller Geotechnik ZT GmbH across Austria and southern Germany. The soundings cover a wide range of geological settings, including post-glacial basins and valleys such as the Vienna Basin, Gastein Valley, and other Alpine and foreland regions. Core drillings located within approximately 50 m of many of the soundings provide independent grain-size information that supports validation of the Robertson SBTn labels.

The following variables were extracted from the original database: depth (m), cone tip resistance qc (MPa), sleeve friction fs (kPa), friction ratio Rf (%), soil behavior type index Ic (–), total vertical stress σv (kPa), effective vertical stress σ′v (kPa), and the target variable SBTn (–).

ISSN 3106-7018 (Print) ISSN 3106-5791 (Online)

SBTn is the normalized Soil Behavior Type classification defined by Robertson (1990), which divides soils into nine behavior-based classes, listed in Table 1.

Data Preprocessing. Rows containing missing values were removed. Records with an invalid SBTn value of 0 were also excluded. After cleaning, the dataset contained 1,010,153 valid observations. The class distribution remained imbalanced: Classes 3, 5, and 6 were the most frequent, while Classes 1, 8, and 9 were relatively rare. To reduce computational demand while preserving class proportions, a stratified random sample of 20% of the cleaned data which contained 202,031 points was drawn. This sample was then split into 80% training and 20% testing subsets using stratified sampling so that the class distribution in both subsets matched the overall distribution. The processed dataset can be accessed in Appendix 1.

Feature Sets. Feature sets were created with each feature set having a different combination of parameters. Four progressive feature sets were defined to isolate the contribution of successive parameters. An additional parameter was added with each increase in set number as described in Table 2.

**Table 2. Sets and feature parameters in each set.**

| Set number | Feature sets, combinations of parameters |
|---|---|
| Set 1 | Depth, qc, fs |
| Set 2 | Depth, qc, fs, Rf |
| Set 3 | Depth, qc, fs, Rf, Ic |
| Set 4 | Depth, qc, fs, Rf, Ic, σv, σ′v |

Set 1 exists as a baseline that uses only quantities measured directly by the cone. Set 2 adds the friction ratio which is a derived parameter and part of traditional Robertson charts. Set 3 further includes the soil behavior type index Ic, which is a key derived parameter in the normalized SBTn framework. Set 4 adds total and effective vertical stresses, σv and σ′v, to test whether explicit stress information improves performance beyond what is already captured by depth and Ic.

Mathematical Formulation of Random Forest. Random Forest is an ensemble of decision trees (Breiman 2001). Given a training set of N observations, each tree is grown on a bootstrap sample drawn with replacement from the training data. At each node, rather than considering all available features, a random subset of m features is selected, and the best split among those m features is chosen. The impurity of a node is commonly measured by the Gini index calculated using Equation 5:

$$\text{Gini}(\square) = 1 - \sum \square, \&$$

**Equation 5. Gini index (Gini 1912).**

Where pk is the proportion of samples belonging to class k at node t. The decrease in impurity achieved by a candidate split is used to select the optimal threshold. After T trees have been grown, the final class prediction for a new observation x is obtained by majority vote as demonstrated in Equation 6:

$$\hat{y} = \text{mode}\{h1(x), h2(x)\ldots, hT(x)\}$$

Equation 6. Majority Voting Ensemble Method (Hastie et al. 2009).

In Equation 6, ŷ denotes the predicted value and hT(x) represents the prediction of the tree. This combination of bagging and random feature selection reduces correlation among trees and thereby lowers the variance of the ensemble without substantially increasing bias (Hastie et al. 2009).

Hyperparameter Optimization. For each feature set, hyperparameters were optimized using Randomized Search with 2-fold cross-validation. The search space included the number of trees, nestimators, ranging from 50 to 200, maximum tree depth; maxdepth, ranging from 10 to 25 or unrestricted; minimum samples required to split a node, minsamples_split ranging from 2 to 10; minimum samples required at a leaf node, minsamples_leaf, ranging from 1 to 4. The macro-averaged F1-score was adopted as the scoring metric.

Evaluation Metrics. Overall accuracy is defined as the proportion of correctly classified samples as shown in Equation 7:

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

**Equation 7. Overall Accuracy (Xiang et al 2025)**

TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. For multi-class problems, this generalizes the fraction of all test samples whose predicted label matches the true label.

Because SBTn classes in this dataset are not evenly distributed, accuracy alone should not be the only measure determining the overall performance. Thus, the F1- score, which is the harmonic mean of precision and recall, is used for measuring the performance within individual classes as seen in Equations 8, 9, and 10.

$$\text{Precision} = \frac{TP}{TP + FP}$$

**Equation 8. Precision (Hastie et al. 2009)**

$$\text{Recall} = \frac{TP}{TP + FN}$$

**Equation 9. Recall (Hastie et al. 2009)**

$$□1 = 2\ \text{Precision} \times \text{Recall}$$
$$\text{Precision} + \text{Recall}$$

**Equation 10. F1-score (Hastie et al. 2009).**

The macro-averaged F1-score is obtained by computing the F1-score independently for each of the nine SBTn classes and then taking the unweighted arithmetic mean of the classes. This metric gives equal importance to every class, regardless of how many samples it contains. Thus, the macro-averaged F1-score is more informative than accuracy when class frequencies differ substantially. Moreover, confusion matrices were also generated to examine class-wise performance as seen in Figures 6, 7, 8, and 9.

Computational Environment. All analyses were performed in Google Colab using the Python programming language and the scikit-learn machine-learning library. Local data preparation and exploratory analysis were carried out on a MacBook Air Model with an Apple M4 chip containing 10 cores with 4 performance cores, 6 efficiency cores, and 24 GB of unified memory.

The combination of local preprocessing and cloud-based model training allowed efficient handling of the large CPT dataset. The code for the analyses performed in Google Colab can be accessed in Appendix 2.

## RESULTS

This section presents the performance of the hyperparameter-optimized Random Forest classifier across the four progressive feature sets. The results show that an increase in input features improves the classification performance overall with the exception of Set 4. Overall, Set 3 obtained the highest accuracy and macro-averaged F1-score when Ic was included. The detailed metrics, comparison across feature sets, and class-wise performances are reported below.

Accuracy and Macro-averaged F1-score. The performance of the Random Forest model in the baseline Set 1 achieved an accuracy of 0.9554 and a macro-averaged F1-score of 0.9397.

Performance improved with Set 2 with the addition of Rf, reaching an accuracy of 0.9648 and a macro-averaged F1-score of 0.9544. The highest performing result was obtained by Set 3 with a further addition of Ic, producing an accuracy of 0.9923 and a macro-averaged F1-score of 0.9845. Notably, the model performance declined in Set 4 with the inclusion of σv and σ′v, yielding a lower accuracy of 0.9911 and macro-averaged F1-score of 0.9822. Table 3 summarizes the performance of the optimized Random Forest models across the four feature sets.

**Table 3. Accuracy and Macro-averaged F1-score for each feature set**

| Set | Features | Accuracy | Macro-averaged F1-score |
|---|---|---|---|
| 1 | Depth, qc, fs | 0.9554 | 0.9397 |
| 2 | Depth, qc, fs, Rf | 0.9648 | 0.9544 |
| 3 | Depth, qc, fs, Rf, Ic | 0.9923 | 0.9845 |
| 4 | Depth, qc, fs, Rf, Ic, σv, σ′v | 0.9911 | 0.9822 |

Figure 5 presents a bar-chart comparison of accuracy and macro-averaged F1-score across the four feature sets. The chart makes the progressive improvement from Set 1 to Set 3 visually apparent, as well as the slight decline observed in Set 4.

ISSN 3106-7018 (Print) ISSN 3106-5791 (Online)

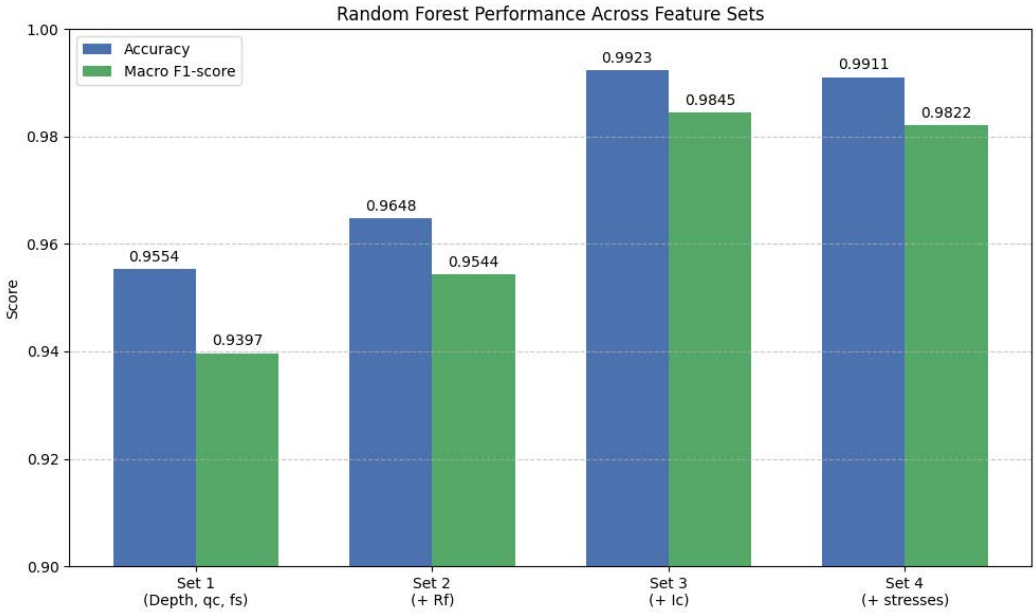



**Figure 5. Accuracy and macro-averaged F1-score across the four feature sets.**

Confusion Matrices. Confusion matrices were created to show the performance within each soil class in their respective sets. In each confusion matrix, the vertical rows represent the actual SBTn class, and the horizontal columns represent the predicted class. The diagonal cells that are highlighted in blue indicate the number of correct classifications. Off-diagonal cells show the number of times one class was confused with another. To note, due to the non-uniform distribution of soil classes, the numbers in the diagonal cells should not be viewed as an indicator of performance. Rather, the performance of each class should be indicated by comparing the number of correct classifications relative to the total number of classifications in that class. Confusion matrices for all four feature sets are shown in Figures 6-9.

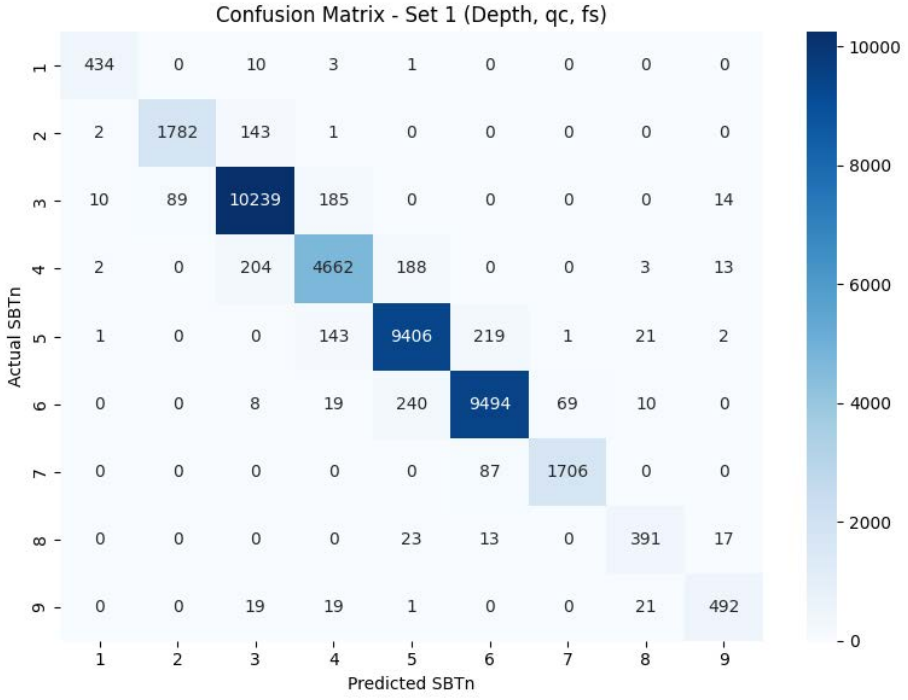



**Figure 6. Confusion matrix for Set 1: Depth, qc, fs.**

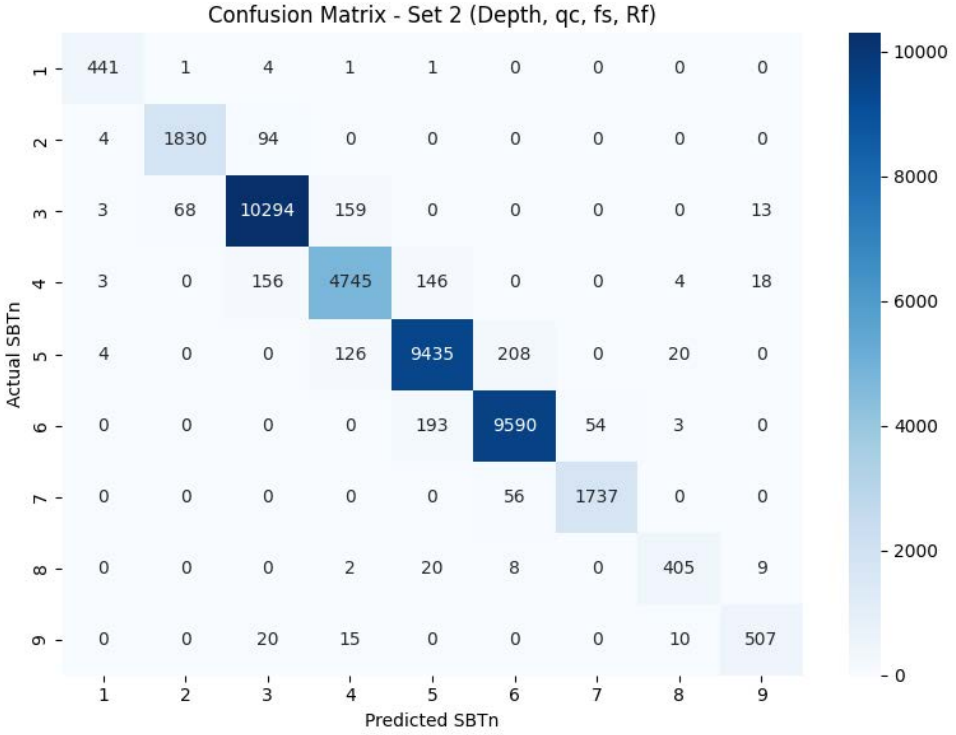



**Figure 7. Confusion matrix for Set 2: Depth, qc, fs, Rf.**

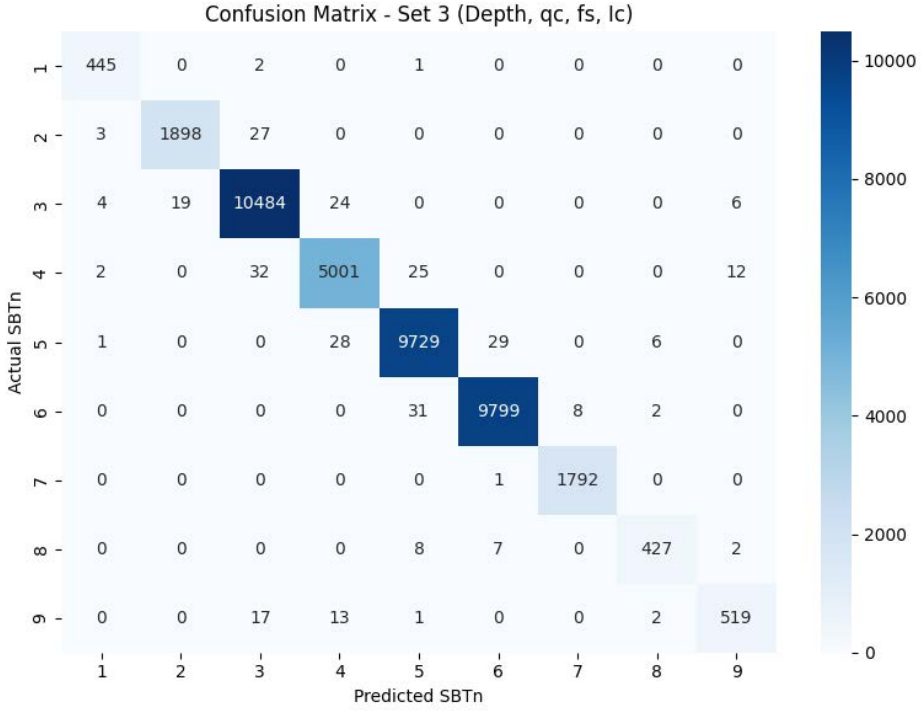



**Figure 8. Confusion matrix for Set 3: Depth, qc, fs, Rf, Ic.**

ISSN 3106-7018 (Print) ISSN 3106-5791 (Online)

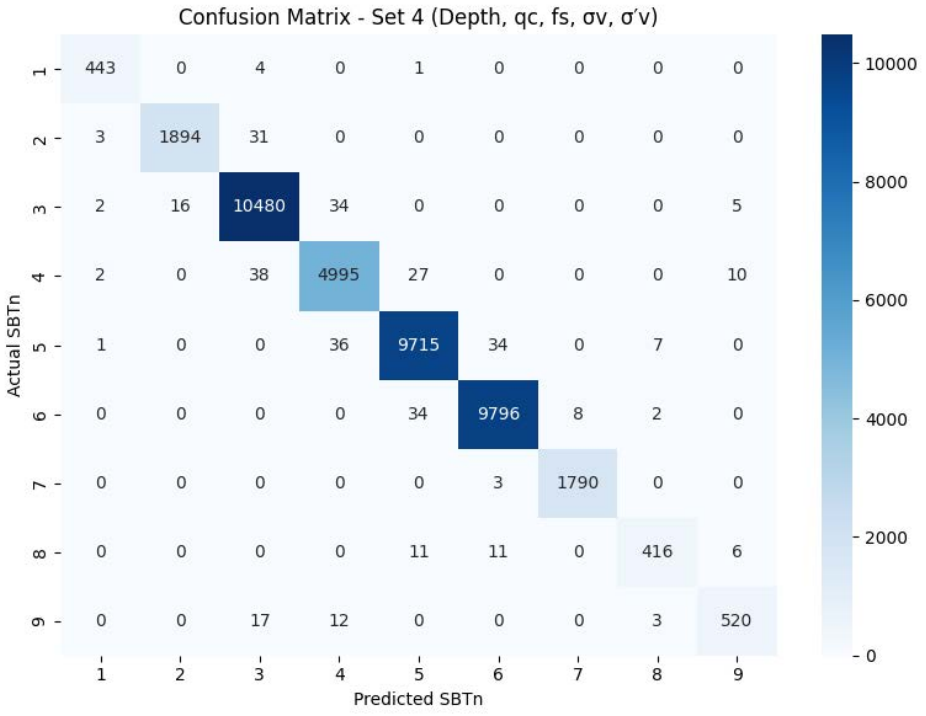



**Figure 9. Confusion matrix for Set 4: Depth, qc, fs, Rf, Ic, σv, σ′v.**

According to the inspection of the confusion matrices, most of the misclassifications were between adjacent SBTn classes, such as between Classes 3 and 4, or between Classes 5 and 6. Notable misclassifications were also observed between Classes 3 and 4 and Class 9 in addition to Classes 5 and 6 and Class 8 respectively.

## DISCUSSION

Influence of Feature Selection. As shown in the results, feature selection can improve the performance of Random Forest for CPT-based soil classification. The largest increase in accuracy was observed from the increase from 0.9648 in Set 2 to 0.9923 in Set 3. This rise in performance can be explained by the addition of Ic. Ic is derived from normalized CPT parameters and is central to the Robertson SBTn framework, providing the model with information that is already closely aligned with the target labels. The large performance improvement suggests that some of the apparent predictive power may be due to a close mathematical relationship between Ic and SBTn rather than independent learning from raw data.

Thus, Ic should be used with caution when training a model as it is only recommended when reproducing the Robertson SBTn chart and not for model learning.

The friction ratio Rf is defined as the ratio of sleeve friction to tip resistance. It is used to distinguish between fine-grained and coarse-grained behavior (Begemann 1965; Robertson et al. 1986). Therefore, adding Rf provides complementary information that is not fully contained in qc and fs separately. This could have caused the relatively small improvement that occurred between Set 1 and Set 2. Moreover, the soil behavior type index Ic is an all-in-one index that already incorporates normalized tip resistance and normalized friction ratio (Robertson and Wride 1998). Therefore, it processes a large part of the information that the Robertson chart itself uses to assign SBTn labels. After the addition of Ic, inclusion of the overburden stresses appears largely redundant for this classification task which is consistent with the slight decline observed in Set 4.

Comparison with Previous Studies. In previous studies such as Rauter and Tschuchnigg (2021), there was a reported 0.97 to 0.99 accuracy for SBTn classification using Random Forest on the same broader dataset family. The Set 3 and Set 4 results with 0.9923 and 0.9911 accuracy fall at the upper end of that range. Chala and Ray (2023) achieved 0.9923 accuracy with Random Forest on a smaller collection of 232 CPTs, matching the Set 3 result obtained here. The close

agreement across independent studies that used different hyperparameter-tuning strategies and different sample sizes suggests that Random Forest performance for SBTn prediction is robust once an adequate set of input features is provided.

Fatehnia et al. (2024) reported a lower F1-score of 0.896. However, the study's target was a borehole-derived soil type rather than a Robertson SBTn. Predicting an independent laboratory or borehole classification from CPT measurements is a fundamentally more difficult task, because the relationship between penetration resistance and grain-size composition is indirect and site-dependent. The higher metrics obtained in the present study and in the work of Rauter and Tschuchnigg (2021) and Chala and Ray (2023),

reflect the closer coupling between CPT parameters and behavior-type labels. This is not necessarily a superiority of one modelling approach over another.

Class-Wise Performance and Confusion Patterns. A confusion matrix is a table that crosses true class labels against predicted class labels. Diagonal entries count correct predictions while off-diagonal entries count specific types of error. In the present study, the matrices show that residual errors are concentrated between neighboring behavior classes. Such confusions can be explained by the transition from, for example, a clayey silt, Class 4, to a silty sand, Class 5.

Such transitions are gradual and CPT measurements near the boundary can be interpreted as belonging to either class. Minority classes, 1, 8, and 9, exhibit lower absolute counts of correct predictions due to how fewer samples of those classes were present in the test set. However, their per-class F1-scores remained acceptable under the macro-averaged metric, indicating that the model did not completely ignore rare classes.

Confusion Patterns in Class 3, 4, and 9. There were notable confusions between Class 3 and 9, and Class 4 and 9. For example, in Figure 8, Set 3, Class 9 soils were misclassified as Class 3 17 times and misclassified as Class 4 13 times respectively. To add, Class 3 soils were misclassified as Class 9 six times and Class 4 were misclassified as Class 9 12 times. However, due to how Class 3 and 4 are numerically distant from Class 9, these confusions cannot be explained by transitional errors observed previously. A possible explanation for the confusion in Class 9 is overconsolidation. This occurs when soil has experienced a higher pressure or load in the past than at the time of measurement (Choo et al. 2018). Class 3 and Class 9 soils both represent plastic, fine-grained materials. If a standard clay in Class 3 becomes heavily overconsolidated, the tip resistance may increase enough to cross into the adjacent high-strength fine-grained Class 9. Additionally, Class 4 silt-rich mixtures may undergo drying, suction, or cementation. Thus, Class 4 soils may show elevated shear strength and normalized resistance. The mechanical behavior may mimic the high stiffness of heavily overconsolidated fine soils in Class 9, explaining the confusion of the model. This can be visualized in Figure 3, in the introduction section, which represents Robertson's SBTn.

Reproducibility Recommendations. A reproducible analysis package should have precise cleaning rules, random seed, sampled row indices, train-test assignment, software version, optimized hyperparameters, and full class-wise results. The provided CSV in Appendix 1 shows that after cleaning, 1,010,153 complete nonzero-SBTn datapoints are still available, and a 20% sample contains about 202,031 datapoints. It is recommended to use a sample size greater than 20% of the datapoints to increase performance across all classes, specifically for minority class soils.

Minority Classes and Decision Consequences. Overall accuracy will still be relatively high if the classifier performs poorly in some soil classes because these classes have fewer samples. The macro-averaged F1-score partially addresses this problem by giving all classes equal weight, but a single macro-averaged F1-score value still fails to show the direction and engineering significance of individual errors. As shown in the provided confusion matrices, many errors occur between adjacent SBTn classes due to gradual soil changes and chart borders. However, the results of errors are not always symmetrical. Although the two similar intermediate classes have a small effect on the early screening profile, a thin weak layer or an unusually sensitive material may change the sampling decision and thus the design assumption. Future reports should thus provide per-class precision, recall, F1-score, and sample counts, and link error analysis to the intended engineering decision. Class weighting or balanced sampling can be used, but the untouched test data should maintain the natural class distribution so that the reported performance is representative.

Operational Quality Control. A field-ready classifier would require more than just a fitted model. It should consider the following controls to prevent an over-reliance on the automated results due to a high overall accuracy and to ensure operational quality control. The input unit should be automatically converted because qc, fs and stress are in different metric systems. A unit conversion error would otherwise occur in predictions. In addition, range checks should flag negative or physically unreasonable resistance values, unexpected depth reversals, missing intervals, and derived variables that are inconsistent with the source measurements.

Recommended Validation Sequence. The next validation stage is a grouped cross-validation using CPT sounding. All measurements from one sounding should be in the same training or validation fold. Hyperparameters should be selected before performing a complete sounding of a held-out area to form an untainted test set. A third stage can use data from another organization or device configuration to assess the sensitivity to equipment and processing differences. Results can be shown as several splits or confidence intervals instead of a single random seed. The same feature set should be applied to an independent target, such as carefully aligned borehole or laboratory classifications, and the difference between reproducing Robertson SBTn and predicting material properties quantified.

## Conclusion

This study systematically investigated how different combinations of input features affect the performance of a

ISSN 3106-7018 (Print) ISSN 3106-5791 (Online)

hyperparameter-optimized Random Forest classifier for CPT-based SBTn classification. A large public dataset from Austria and Southern Germany was used to compare four progressive feature sets resulting in the following notable findings. First, the basic feature set of depth, qc, and fs achieved an accuracy of 0.9554 and a macro-averaged F1-score of 0.9397.

Second, the addition of friction ratio increased the accuracy by 0.94 percentage points and the macro-averaged F1-score by 1.47 percentage points. Third, the addition of the soil behavior type index yielded the largest improvement, raising accuracy to 0.9923 and macro-averaged F1-score to 0.9845. Fourth, adding total and effective vertical stresses did not improve the performance further and decreased the performance of both accuracy and macro-averaged F1-score. Within classes, Classes 3 and 4 were often misclassified as Class 9 and vice versa. Less frequently, Classes 8 and 9 were also misclassified as each other.

Future work should explore model transferability in geological areas beyond Austria and southern Germany, investigate methods to enhance the performance of the minority class e.g., class weighting or synthetic oversampling, and compare Random Forest with newer ensemble models such as XGBoost and LightGBM under the same feature set as well as explore Deep Learning models. Expanding the analysis to include independent grain-size targets could provide a more rigorous test of predictive ability. In conclusion, a feature set of depth, qc, fs, Rf, and Ic, is recommended for workflows that aim to reproduce Robertson SBTn and already calculate Ic. However, for models that only train using raw CPT measurements, the feature set of depth, qc, fs, and Rf is recommended. For this dataset, adding total and effective vertical stress is not recommended.

## REFERENCES


[1] ASTM International. (2017). Standard practice for classification of soils for engineering purposes (Unified Soil Classification System). ASTM D2487-17. West Conshohocken, PA.

[2] ASTM International. (2020). Standard test method for electronic friction cone and piezocone penetration testing of soils. ASTM D5778-20. West Conshohocken, PA.

[3] Begemann, H. K. S. P. (1965). "The friction jacket cone as an aid in determining the soil profile." Proc., 6th Int. Conf. on Soil Mechanics and Foundation Engineering, Montreal, 1, 17-20.

[4] Bosch, D. K., and Sotelo, R. R. (2015). "Determination of Stratigraphy — Soil Types — Using Cone Penetration Test in Sedimentary Deposits in North-East of Argentina". Journal of Geoscience and Environment Protection, 03(06), 134-139.

[5] https://doi.org/10.4236/gep.2015.36019.

[6] Breiman, L. (2001). "Random forests." Machine Learning, 45(1), 5–32.

[7] https://doi.org/10.1023/A:1010933404324.

[8] Chala, A. T., and Ray, R. (2023). "Assessing the performance of machine learning algorithms for soil classification using cone penetration test data." Applied Sciences, 13(9), 5758.

[9] https://doi.org/10.3390/app13095758.

[10] Choo, H., Lee, W., Lee, C. (2018) "Overconsolidation and Cementation in Sands: Impacts on Geotechnical Properties and Evaluation Using Dilatometer Tests" Geotechnical Testing Journal, 41(5), 20170368. https://doi.org/10.1520/GTJ20170368

[11] Douglas, B. J., and Olsen, R. S. (1981). "Soil classification using electric cone penetrometer." Cone Penetration Testing and Experience, ASCE, 209–227.

[12] Fatehnia, M., Mahmoudabadi, V., and Amiri, S. (2024). "Utilizing machine learning for cone penetration test-based soil classification." Transportation Research Record: Journal of the Transportation Research Board, 2678(11), 1639–1648.

[13] https://doi.org/10.1177/03611981241245679.

[14] Gini, C. (1912). "Variabilità e mutabilità: Contributo allo studio delle distribuzioni e delle relazioni statistiche". Tipografia di Paolo Cuppini, Bologna, Italy.

[15] Goh, A. T. C., and Goh, S. H. (2007). "Support vector machines: Their use in geotechnical engineering as illustrated using seismic liquefaction data." Computers and Geotechnics, 34(5), 410–421. https://doi.org/10.1016/j.compgeo.2007.06.001.

[16] Hastie, T., Tibshirani, R., and Friedman, J. (2009). The Elements of Statistical Learning. 2nd ed. Springer, New York. https://doi.org/10.1007/978-0-387-84858-7.

[17] Kennedy, K. (2022). "Introduction to Cone Penetration Testing." Association of Geotechnical & Geoenvironmental Specialists. https://www.ags.org.uk/2022/09/introduction-to-cone-

[18] penetration-testing/.

[19] Kurup, P. U., and Griffin, E. P. (2006). "Prediction of soil composition from CPT data using general regression neural network." Journal of Computing in Civil Engineering, 20(4), 281–289. https://doi.org/10.1061/(ASCE)0887-3801(2006)20:4(281).

[20] Lunne, T., Robertson, P. K., and Powell, J. J. M. (1997). Cone Penetration Testing in

[21] Geotechnical Practice. Blackie Academic & Professional, London. ISBN 978-0-419-23750-1. Oberhollenzer, S., Premstaller, M., Marte, R., Tschuchnigg, F., Erharter, G. H., and Marcher, T.

[22] (2021). "Cone penetration test dataset Premstaller Geotechnik." Data in Brief, 34, 106618. https://doi.org/10.1016/j.dib.2020.106618.

[23] Abey Rajapakse, R. 2016. Pile Design and Construction Rules of Thumb. 1–30. 2nd ed. New York: Butterworth-Heinemann.

[24] Rauter, S., and Tschuchnigg, F. (2021). "CPT data interpretation employing different machine learning techniques." Geosciences, 11(7), 265. https://doi.org/10.3390/geosciences11070265. Reale, C., Gavin, K., Librić, L., and Jurić-Kaćunić, D. (2018). "Automatic classification of fine-grained soils using CPT measurements and artificial neural networks." Advanced Engineering Informatics, 36, 207–215. https://doi.

org/10.1016/j.aei.2018.04.003.
[25] Robertson, P. K. (1990). “Soil classification using the cone penetration test.” Canadian Geotechnical Journal, 27(1), 151–158. https://doi.org/10.1139/t90-014.
[26] Robertson, P. K. (2009). “Interpretation of cone penetration tests—A unified approach.” Canadian Geotechnical Journal, 46(11), 1337–1355. https://doi.org/10.1139/t09-065.
[27] Robertson, P. K. (2010). “Soil behaviour type from the CPT: An update.” Proc., 2nd Int. Symp. on Cone Penetration Testing, Huntington Beach, CA.
[28] Robertson, P. K. (2015). Guide to Cone Penetration Testing for Geotechnical Engineering. 6th ed., Gregg Drilling & Testing, Inc., Signal Hill, CA.
[29] Robertson, P. K. (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.
[30] Robertson, P. K., Campanella, R. G., Gillespie, D., and Greig, J. (1986). “Use of piezometer cone data.” Use of In Situ Tests in Geotechnical Engineering, GSP 6, ASCE, 1263–1280.
[31] Robertson, P. K., and C. E. Wride. (1998). "Evaluating cyclic liquefaction potential using the cone penetration test." Can. Geotech. J., 35 (3): 442–459.
[32] Terzaghi, K. (1936). "The shearing resistance of saturated soils and its graphical representation." In Vol. 1 of Proc., 1st Int. Conf. on Soil Mechanics and Foundation Engineering, 54–56. Cambridge, MA: Harvard Univ.
[33] Terzaghi, K. (1943). “Theoretical Soil Mechanics”. John Wiley & Sons, New York.
[34] U.S. Geological Survey (USGS). (2014). “2014 landslide in Washington State.”
[35] https://www.usgs.gov/media/images/2014-landslide-washington-state-7 (accessed Aug. 19, 2026).