

Article

Integrating Machine Learning in Geotechnical Engineering: A Novel Approach for Railway Track Layer Design Based on Cone Penetration Test Data

Matthieu Bernard

Department of Track Renewal, Infrabel, 85 Rue de France, 1060 Brussels, Belgium; matthieu.bernard@infrabel.be; Tel.: +32-2-525-21-57

Abstract: The cone penetration test (CPT) has emerged as a cost-effective and time-efficient method for assessing soil conditions relevant to railway track infrastructure. The geotechnical data obtained from the CPT serve as crucial input for asset managers in designing optimal sublayers and form layers for track renewal works. To properly assess the condition of soil layers, various soil behavior type charts and machine learning models based on CPT data have been published to help engineers classify soils into groups with similar properties. By understanding the properties of the soils, an optimal substructure can be designed to minimize extensive maintenance and reduce the risk of derailment. However, when analyzing multiple CPTs, the diversity and non-uniformity of subsoil characteristics pose challenges in designing a new optimal trackbed. This study presents an automated approach for recommending thicknesses of sublayers and form layers in railway tracks based on CPT data, employing machine learning algorithms. The proposed approach was tested using CPT data from the Belgian railway network and showed very good agreement with results from traditional soil investigation interpretations and layer design. A Random Forest classifier, fine-tuned through Bayesian optimization with a cross-validation technique and trained on 80% of the datasets, achieved an overall accuracy of 83% on the remaining 20%. Based on these results, we can conclude that the proposed model is highly effective at accurately designing sub-ballast layers using CPT data.



Citation: Bernard, M. Integrating Machine Learning in Geotechnical Engineering: A Novel Approach for Railway Track Layer Design Based on Cone Penetration Test Data. *Infrastructures* **2024**, *9*, 121. <https://doi.org/10.3390/infrastructures9080121>

Academic Editor: Giuseppe Cantisani

Received: 17 April 2024

Revised: 8 July 2024

Accepted: 22 July 2024

Published: 24 July 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: cone penetration test; trackbed; layer thickness; machine learning; random forest

1. Introduction

1.1. Geotechnical Engineering in Railway Track

Geotechnics plays a crucial role in the design and maintenance of railway infrastructure, directly impacting the stability, durability, and overall performance of the network. The sublayers beneath the ballast primarily serve to distribute loads, prevent settlements, and maintain the long-term integrity of the infrastructure. However, due to the complex interactions between the soil, the track superstructure (rails, sleepers, and ballast) and the loads exerted by trains, designing an optimal sublayer is an especially delicate task, relying on empirical engineering methods. These methods are mainly derived from field expertise accumulated over the years by asset managers and have led to significant advancements in sublayer design to extend the time between ballast maintenance, avoid subgrade mud pumping or ballast pockets, reduce the risks of derailment, and to accommodate increased wheel loads [1,2]. The most common technique to increase the stiffness and strength of the subgrade is soil reinforcement, sometimes referred to as ground treatment or stabilization. The solution considered in this paper is the replacement of soft soil with granular materials such as crushed rock or gravel. Bituminous sub-ballast layers or soil–cement stabilization techniques are not discussed in this paper. A general cross-section of a conventional railway track is shown in Figure 1. Even if each network manager has their own definitions of these layers and the materials used, as outlined in UIC norms [3], the cross-section typically includes, successively, the ballast, the sub-ballast, the form layer, and the foundation, which

is generally natural ground. When laying main line track, the UIC norms [3] recommend a sub-ballast layer with a minimum depth, determined based on the bearing capacity of the soil. In the Belgian standards, based on the UIC norms, a form layer is incorporated only when the soil's bearing capacity is low, and the stones used are of size 0/200. The sub-ballast layer, characterized by stones of size 0/32, is consistently present in singular assets such as level crossings and switches. Despite varying standards, each operator shares the common goal of designing optimal sublayers while adhering to their respective regulations [4].

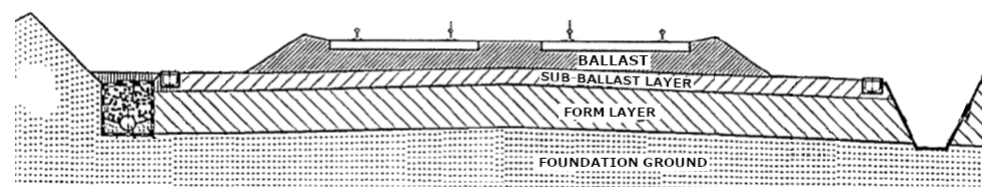


Figure 1. General cross-section of a conventional ballast railway track [3].

Nowadays, with the advent of dynamic penetration in situ tests that provide a direct and valuable method for soil characterization, the procedure for renewing the substructure of track assets, such as switches or level crossings, generally involves the utilization of lightweight measurement tools during field tests. Subsequently, geotechnical engineers manually interpret the data collected from these tests [5]. Geophysical alternative methods such as georadars or surface wave seismic methods, more specifically multichannel surface wave analysis (MASW), are establishing themselves as non-destructive techniques for fast, efficient track diagnostics [6,7]. These methods can provide information over longer linear distances and with greater density than standard geotechnical methods. Although these methods are of great interest as they enable a rapid and efficient assessment of the condition of railway infrastructures, they are not discussed further in this article.

While the cone penetration test (CPT) is a more efficient alternative to the laborious and costly core drilling methods, providing quicker and more reliable data, the manual interpretation of the soundings still requires a high level of expertise. On one hand, there is the individual analysis of soundings, which involves precisely identifying the different layers of the existing substructure (sublayer, form layer, and platform), i.e., the layer thickness and the mechanical properties. On the other hand, in case of soft soil, all the soundings from the studied site must be analyzed as a whole to propose a uniform soil reinforcement solution and avoid the negative effects of stiffness changes. Indeed, track structure stiffness variations can lead to differential track structure settlements and a higher concentration of track defects [8,9]. Typically, they are located near concrete structures, such as bridges or level crossings, known as transition zones. But other discontinuities, such as the extremities (transition between) of a reinforced section of the substructure, are also responsible for stiffness variations. Ideally, the track structure should be uniformly stiff.

In addition to meeting purely technical requirements, optimizing the thickness of the trackbed is an environmental necessity to use natural resources responsibly and avoid unnecessary stones mining. The green transition is challenging railway managers to be more responsive to environmental needs.

1.2. Artificial Intelligence

The automation of structural condition assessments for railway tracks is an evolving field, driven by the need to manage extensive data and reduce subjective interpretation. Machine learning models have demonstrated notable success in soil classification by analyzing properties like particle size and plasticity, typically determined through laboratory procedures. These models have been effective in categorizing various soil types, such as high-plasticity clay and low-plasticity clay [10,11]. However, traditional sample collection methods used in those papers, like soil drilling, are often less efficient compared with CPT, as explained previously.

Sastre [12] introduced a machine learning approach to classify soil types, such as fine-grained, silty, or clayey soils, fine sands, and gravels, based on penetrometric signal analysis. However, the accuracy of in situ soil classification using these models ranges between 50% and 60%, which is not optimal. The main reason lies in the fact that most samples used for training these models are predominantly composed of laboratory test data, making it difficult to generalize to more complex field situations. In contrast, using in situ samples for training has led researchers to achieve prediction accuracies of around 99% in soil classification based on Robertson's soil behavioral types [13,14], highlighting the importance of high-quality, field-based data samples.

Recently, there has been a growing interest in utilizing CPT data in conjunction with machine learning models for a wider range of geotechnical applications. This includes predicting soil shear wave velocity, demonstrating the versatility of this technique beyond mere soil classification [15]. Furthermore, open-access databases have been established to encourage the development of new methodologies for soil classification based on CPT data [16] or through the International Society for Soil Mechanics and Geotechnical Engineering.

Despite notable advancements in machine learning for soil classification using CPT data, current methodologies predominantly focus on analyzing individual soundings. This approach limits the scope of analysis, often failing to provide a comprehensive understanding of the soil conditions across the entire site under study. In the current literature, the more straightforward task of analyzing two-dimensional charts, which plots cone resistance, q_d , against depth, has been successfully automated by machine learning models. However, the responsibility of synthesizing these individual analyses into a cohesive, site-wide solution still predominantly falls to engineers. While global analyses, integrating multiple soundings for a holistic view of soil conditions, have been conducted in some fields [17,18], their application in railway engineering remains relatively unexplored. This gap highlights a significant opportunity for the advancement of machine learning applications in railway geotechnics, moving beyond individual soundings to a more integrated and comprehensive analysis of trackbed conditions.

This paper introduces a pioneering application of machine learning in the field of railway engineering, focusing on predicting the necessary soil reinforcement of sublayers directly from in situ CPT samples. By determining the depth of soil replacement in an automated manner, this novel approach facilitates the design strategies for renewing railway tracks. The applicability of the methodology considered in this paper is of course not limited to the Belgian network as it is actually a generic approach that can be implemented by each operator while adhering to their specific regulations.

1.3. Objectives of the Work

This study focuses on the specific application of artificial intelligence in designing sublayers for railways using light penetrometer surveys. The aim is to investigate how well-known artificial intelligence algorithms can be innovatively integrated to assist asset managers in the design of these sublayers based on multiple CPT soundings. The key benefits sought are:

- Enhanced decision-making support through analysis based on extensive datasets.
- Minimization of human errors in the interpretation of geotechnical data by automating the analysis process.
- Increased operational efficiency: faster analysis of survey data.
- Optimization of the lifespan of railway infrastructures.

2. Materials and Methods

2.1. Panda

In railway engineering, data from dynamic penetrometers (variable energy method) are increasingly being used to assess the quality of subgrades and platforms [5,19]. These tests aid track designers in making well-informed decisions for renewal projects. Therefore, the proposed geotechnical characterization was carried out using the PANDA dynamic

penetration test. This test allows for the examination of changes in stiffness by measuring the dynamic cone resistance, q_d , at various depths. As shown in Figure 2, the test involves driving a set of steel rods with a conical tip into the ground by hammering. The PANDA test is performed directly along the track axis between two sleepers to analyze the railway substructure. The outcomes of the test are methodically recorded and presented in the form of penetrograms, which are graphs depicting the evolution of cone resistance with depth. The testing commences at a depth of 0, corresponding to the uppermost surface of the railway sleeper. Notably, the test uses a 4 cm² cone, avoiding lateral friction measurements. Furthermore, to ensure minimal interference from lateral friction during the test, a preliminary excavation is conducted in the ballast using a crowbar. This preparatory step accounts for the observed zero soil resistance in the initial 50 cm of depth, which corresponds to the combined thickness of the sleepers (20 cm) and the ballast layer (30 cm). Thus, the analysis relies solely on cone resistance, q_d , and depth, with friction being negligible.

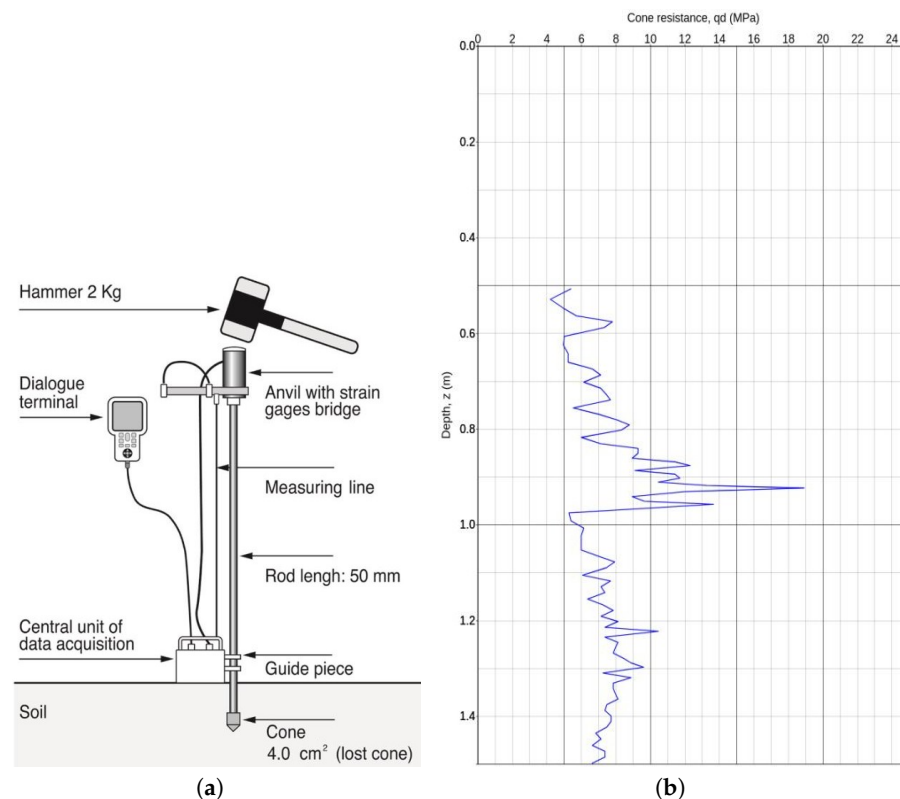


Figure 2. Example of a CPT test. (a) Illustration of the PANDA; (b) data of an individual sounding: cone resistance, q_d , as a function of depth.

2.2. Dataset Overview

The establishment of a database is a crucial step in the process. In this case, various sites had been studied since 2017, including level crossings, switches, and areas with geometric instabilities in the track. Each site includes multiple soundings. To streamline data encoding, the average resistance was calculated between depths of 50–60 cm, 60–70 cm, and so on, up to 140–150 cm for each sounding. For each site, recommendations for the substructure were made in accordance with the requirements applied to the Belgian rail network. Specifically, the recommendations include the thickness of the sublayer and the thickness of the form layer to be implemented during site renewal. The literature shows that the performance of critical zones with weak subgrade can be improved by increasing the granular layer thickness [20,21].

The features database structure is shown in Table 1, where each row contains the mean cone resistance value between two specified depths and the site identification parameters

related to this measure (site and sounding number). The label database structure is shown in Table 2, with the thickness of the sub-ballast and the form layer.

Table 1. Example of the features database: three sites, a, b, and c, with, respectively, 3, 4, and 3 soundings.

Site	Sounding	Mean Cone Resistance q_d [MPa]					
		50–60 cm	60–70 cm	70–80 cm	...	130–140 cm	140–150 cm
a	1	10.08	11.61	21.91	...	3.93	5.04
a	2	7.52	11.39	11.87	...	9.54	5.68
a	3	13.52	15.68	18.62	...	29.46	0.0
b	1	7.59	12.80	10.03	...	14.48	13.22
b	2	9.29	9.11	8.26	...	5.15	4.72
b	3	12.43	26.64	35.96	...	3.80	2.67
b	4	15.50	33.44	38.68	...	8.93	2.32
c	1	44.59	6.11	43.05	...	11.13	12.26
c	2	15.29	23.95	22.60	...	20.58	18.22
c	3	25.43	31.06	15.39	...	28.11	18.02

Table 2. Example of the label column: thickness of the sub-ballast layer and the form layer.

Site	Thickness Recommendation [cm]	
	Sub-Ballast Layer	Form Layer
a	20	0
b	10	40
c	40	0

In total, there are 2500 PANDA soundings, spread across 560 sites, averaging about 4 to 5 surveys per site.

2.3. Label Encoding

The challenge we faced involves classifying sets of soundings (sites) based on two distinct characteristics: the thickness of the sublayer and the thickness of the form layer (see Table 2). This presents as a multi-label classification problem. To tackle this complexity, each observed combination of sublayer and form layer thickness was encoded as a separate class. An alternative approach, such as classifier chaining, could have been considered, where a first model is trained to determine the sublayer thickness, and then a second model uses the output of the first to determine the form layer thickness. However, the combined approach was favored for its simplicity, ease of understanding, and the limited number of combinations in this case. Consequently, the classification challenge was redefined as a multi-class classification problem, with only 5 possible combinations of sublayer and form layer thicknesses, as shown in Table 3. To encode target labels with values ranging from 0 to $n_{classes} - 1$, the LabelEncoder from scikit-learn was used.

Table 3. Encoded target labels.

Label/Class	Thickness Recommendation [cm]	
	Sub-Ballast Layer	Form Layer
0	20	0
1	30	0
2	40	0
3	10	40
4	20	40

2.4. Cleaning Data of Outliers

To identify and eliminate outliers from the dataset, the z-score methodology was employed. Specifically, the mean and standard deviation of the dynamic cone tip resistance (q_d) were computed for each depth within every class. Subsequently, to assess whether a site associated with a particular class exhibited outlier behavior, the z-score of the cone resistance at each depth for the site was calculated relative to the mean cone resistance at that specific depth. If two z-scores exceeded a threshold of 3 (indicating that the site's q_d at a given depth was at least 3 standard deviations from the mean q_d for that depth within the class), the site was classified as an outlier. This approach served as a robust quality control measure, ensuring the integrity of the dataset by identifying and excluding data points that deviated significantly from the expected behavior within their respective classes.

2.5. Analysis of the Dataset

In order to visualise the cleaned dataset, one observed the variation of cone resistance with depth and the effect of the class group on this variation. Therefore, Figure 3 illustrates the variation of q_d with depth for each class group. One can readily observe that thicker subgrades are typically proposed for soils with poor resistance, while sites with sufficient resistance will receive thinner subgrades.

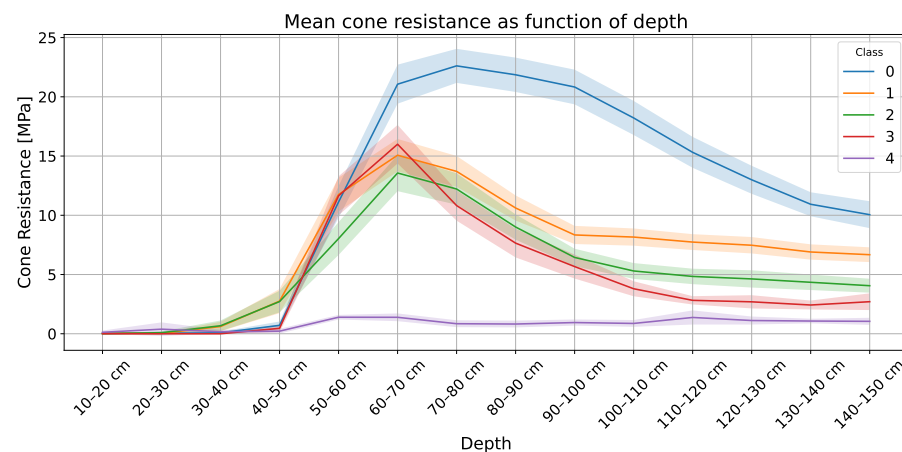


Figure 3. Mean values of q_d for each class group and for each depth.

Additionally, principal component analysis (PCA) was applied to the dataset to visualize each site individually and its position in reference to its neighbors, as shown in Figure 4. PCA is a linear dimensionality reduction technique where data are transformed onto a new coordinate system for easy identification of the largest variation in the data.

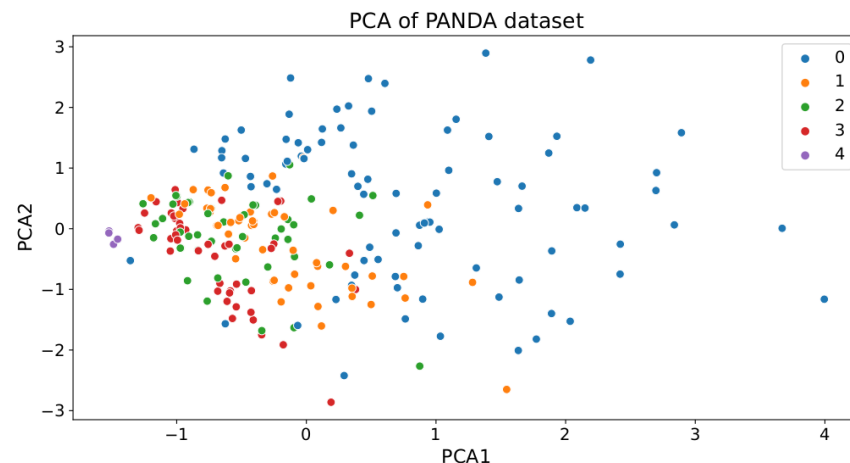


Figure 4. Principal component analysis of the PANDA dataset.

In this case, the relationship between points of the same classes is clearly perceptible, indicating the potential of standard machine learning models to successfully address this classification problem.

2.6. Pre-Processing

2.6.1. Padding

The distribution of the number of individual samples collected for each distinct site is shown in Figure 5. It highlights the fact that the model must deal with sites that have different numbers of samples, ranging from 2 to 13 per site.

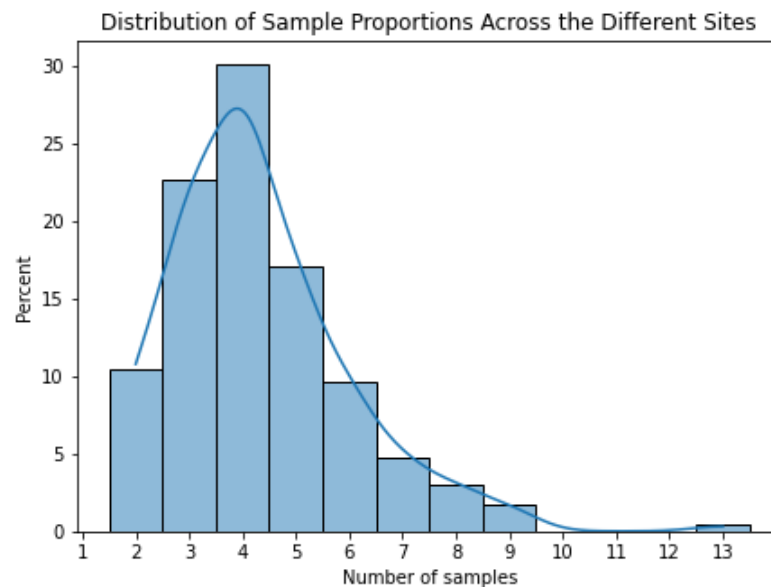


Figure 5. Distribution of sample proportions across the different sites.

The zero-padding method was used to make the data compatible with machine learning models that require fixed-size inputs. This technique involves adjusting all data sequences by adding zeros to the shorter sequences, thereby aligning the entire set with the longest sequence in the dataset.

2.6.2. Scaling

Scaling before applying machine learning is part of best practices. The main advantage of scaling is to avoid attributes in greater numeric ranges dominating those in smaller numeric ranges. By bringing together all the different types of variable units in the same order of magnitude, one eliminates the potential outlier measurements that would misrepresent the finding and bias the machine learning model. One used the RobustScaler of scikit-learn to scale both training and testing data. The RobustScaler removes the median and scales the data according to the quantile range (the range between the 1st quartile and the 3rd quartile).

2.6.3. Split Train Test

To assess model performance, a widely adopted practice involves dividing the dataset into an 80/20 split for training and testing, respectively. The model is constructed using the training dataset, and its performance is subsequently evaluated on the test dataset. To achieve this split, the scikit-learn library's train-test split function is employed. Notably, a crucial consideration is the utilization of stratified sampling, wherein each set aims to preserve a comparable percentage of samples from each target class as the entire dataset. This strategic approach ensures that minority classes are adequately represented in both the training and test datasets. In contrast, simple random sampling fails to capture the complete diversity of the population in the training and testing data.

2.6.4. Over Sampling

As is common in many datasets, the distribution of classes is uneven, potentially leading to model bias. The model might disproportionately favor the majority classes, overlooking the minority ones. To address this imbalance in class distribution, the SMOTE technique (Synthetic Minority Over-sampling Technique) was employed [22]. Its principle is to create new samples in the training dataset not merely through duplication but by combining features of neighboring samples. As illustrated in Figure 6, the class 4 (sublayer of 40 cm and form layer of 20 cm), comprising only 1.7% of the training dataset, is significantly underrepresented. This last class was oversampled in the training dataset using SMOTE, resulting in 69 samples, equaling the majority class 0.

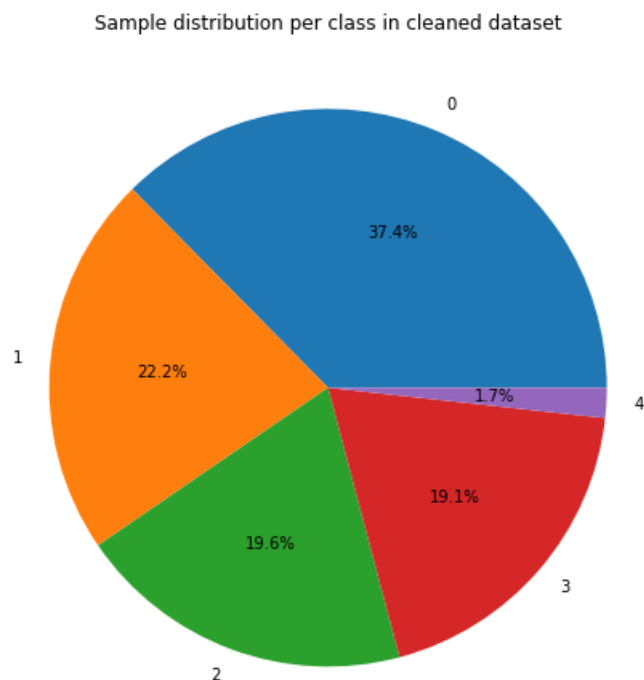


Figure 6. Imbalanced train dataset: the classes from Table 3 are not represented by the same number of samples (ranging from 37.4% of the total dataset for class 0 to 1.7% for class 4).

2.7. Classification Model

One opted for a Random Forest classifier for the classification problem, as it has already demonstrated good results in similar tasks [23]. This model is a commonly used machine learning algorithm that keeps the majority vote of multiple uncorrelated decision trees to increase the overall result. To create an uncorrelated forest of decision trees, only a random subset of the features is taken into consideration by the algorithm, which ensures low correlation among decision trees. This ensemble method combined with the feature randomness approach helps to reduce variance within a noisy dataset. Also, Random Forest makes it easy to evaluate features importance on the prediction, allowing if desired the dropping of features because they do not contribute enough to the prediction process, always in order to avoid overfitting.

2.8. Hyperparameter Optimization Using Cross-Validation

Random Forest algorithms have hyperparameters whose values need to be defined before training. These include the number of trees (`n_estimators`), the maximum depth of the trees (`max_depth`), and the number of features sampled (see sklearn documentation). Selecting the best combination of hyperparameters that deliver the best performance avoiding overfitting or underfitting is known as hyperparameter optimization. Among the various automatic optimization techniques, one has opted for a Bayesian Search combined with a v -fold cross validation.

The Bayes Search offers a solution by choosing candidate hyperparameters based on the performance of previously tried values, resulting in a more efficient hyperparameter tuning method than sweeping alternatives, e.g., Randomized Search or grid Search. One opted for the BayesSearchCV model of the scikit-optimize package. Concerning the v -fold cross-validation, it is a technique that consists in dividing the training set into v subsets of equal size. Sequentially, one subset is tested using the classifier trained on the remaining $v-1$ subsets. Thus, each instance of the whole training set is predicted once, so the cross-validation (CV) accuracy is the percentage of data that are correctly classified. With cross-validation, the validation set is no longer needed (which would reduce the number of samples that can be used for learning the model) while still providing model performance insights on unseen data and prevent the overfitting problem. In this paper, a 5-fold stratified repeated CV was applied within the BayesSearchCV during the training process.

Concretely, one specifies a range of values for each hyperparameter and selects a metric to optimize, and BayesSearchCV searches for a combination of hyperparameters that optimizes one's selected metric. In this paper, BayesSearchCV was set to optimize the F1 score while playing with the hyperparameters of the Random Forest classifier shown in Table 4. While accuracy is one of the most straightforward and popular metrics used in machine learning, it could be dangerously misleading when classes are imbalanced. The F1 score being a popular metric for imbalanced classification, one opted for this one. The F1 score can be interpreted as a harmonic mean of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal. The formula for calculating the F1 score is given by Equation (1):

$$F_1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (1)$$

where

$$\text{Precision} = \frac{\text{Truepositive}}{\text{Truepositive} + \text{Falsepositive}} \quad (2)$$

$$\text{Recall} = \frac{\text{Truepositive}}{\text{Truepositive} + \text{Falsenegative}} \quad (3)$$

Table 4. Hyperparameters of Random Forest classifier to optimize.

Hyperparameters	Boundaries	
	min	max
n_estimators	10	200
max_depth	1	20
min_samples_split	2	10
min_samples_leaf	1	10
max_features	0.1	1.0

2.9. Training, Validation, and Testing

After pre-processing the data, the building process of the machine learning model can be divided into three different steps. Firstly, the training step, where the machine learning algorithm learns from the training data. Secondly, the validation step, where the model is analyzed regarding generalization properties such as overfitting and underfitting. In this case, the validation is performed with the v -fold cross-validation technique. Thirdly, the testing step, where the model with the desired hyperparameter set is tested on unseen data from the test dataset. The results of the testing step are then used to generate the classification report and confusion matrix.

3. Results

The Random Forest classifier cross-validated F1 score is 87%, with the hyperparameter given in Table 5. On the test set, the model reached an accuracy of 83%. The confusion matrix for the classification of the test set can be seen in Figure 7, displaying the percentage of correct and incorrect predictions made by the classification model (precision). One observes that the highest prediction rate is for classes 0 and 4, which is quite obvious as they are extreme solutions, and thus both have only one neighbor (see Figure 3). On the other hand, classes 1, 2, and 3 have bigger difficulties in being differentiated as they are much closer to each other, almost intertwined (see Figure 4), and both have two neighbours. Either way, the classification results are acceptable with a minimal precision of 71% for class 2. Finally, the classification report compiles the state-of-the-art metrics for an imbalanced dataset: precision, recall, and F1 score (see Table 6).

Table 5. Best hyperparameters for Random Forest Classifier.

Hyperparameters	Best
n_estimators	200
max_depth	11
min_samples_split	2
min_samples_leaf	1
max_features	0.1

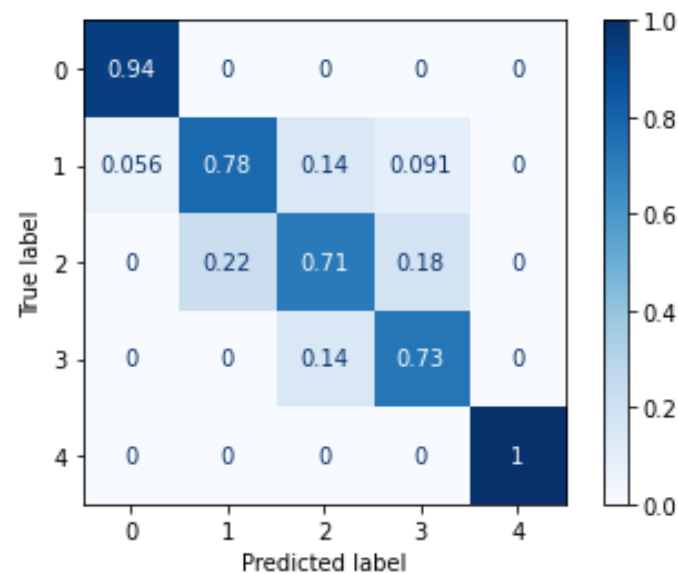


Figure 7. Random Forest classifier confusion matrix.

Table 6. Classification report.

Label	Metrics		
	Precision	Recall	F1 Score
0	0.94	1.00	0.97
1	0.78	0.70	0.74
2	0.71	0.56	0.63
3	0.73	0.89	0.80
4	1.00	1.00	1.00

4. Discussion

Despite optimizing the model's hyperparameters and implementing outlier removal techniques such as z-score normalization on the dataset, the model's performance remains capped by what we could call a glass ceiling. This barrier is primarily comprised of two factors: the Bayes error rate, which represents the minimum theoretical error the classifier can achieve with perfect knowledge of the data distribution, and the presence of noisy labels—errors or inconsistencies in the labels assigned to training examples. Noisy labels can stem from various sources including human annotation errors or ambiguities in class definitions.

While it is challenging to precisely quantify the individual contributions of each factor to the overall performance, one has observed that erroneous predictions, even if they are classified into the wrong class or, in other words, not as would have been suggested by the geotechnical engineer who analyzed the site, still result in a coherent classification and are generally more appropriate. This demonstrates that the higher the quality of the dataset is, the better is the model's performance. In this case, more accurate outlier detection could be achieved by manually removing sites with the help of PCA analyses, for example. These outliers may arise from inadequate interpretation of the penetrometer data or of technical constraints specific to the site and thus may not be representative of the proposed solution. For example, if the geotechnical engineer suggests a substructure depth of 40 cm, but environmental limitations necessitate a maximum depth of 20 cm, a substructure depth of 20 cm would be chosen.

Moreover, additional features such as the traffic speed, the tonnage, the humidity, the asset type (switch, level crossing, . . .), combined with feature selection, could potentially enhance the performance of the model.

5. Conclusions

This study employs a Random Forest classifier to support the design decisions regarding subgrades thicknesses for conventional track section or individual assets (such as switches or level crossings), utilizing in situ cone penetration testing data. This highlights the capability of a well-known machine learning model to integrate multiple soundings (ranging from 2 to 13) and conduct a meta-analysis of the trackbed condition in order to propose an optimal subgrade thickness.

To optimize the performance of the model, hyperparameter tuning was performed using Bayesian optimization. Additionally, cross-validation was employed to ensure the model generalization ability and optimal performance. The performance of the model was evaluated using performance metrics such as precision, recall, F1, and overall accuracy. The Random Forest model achieved an overall accuracy of 83% on test data.

The results demonstrate that the methodology adopted in this study can serve as a decision support tool for railway asset managers seeking to leverage historical data. The implemented pipeline of methodologies, encompassing data cleaning, resampling, and hyperparameter optimization, is generic in nature. While additional relevant features, such as traffic speed, UIC category, survey season (temperature, soil humidity, etc.), and geometric measurements of the track, can easily be added, depending on the preferences and requirements.

Future research could optimize subgrade thickness not only by focusing on CPT tests, but also by considering geoscope tests (granulometry), which would require neural networks to analyze the images. The results are expected to provide more comprehensive railway track diagnostics.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: This work was supported by the Belgian railway infrastructure manager (Infrabel).

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

Abbreviations

The following abbreviations are used in this manuscript:

CPT	Cone penetration test
UIC	International Union of Railways
PCA	Principal component analysis
SMOTE	Synthetic Minority Over-sampling Technique
CV	Cross-validation

References

- Li, D.; Wilk, S. Recent studies on railway-track substructure at TTCl. *Transp. Saf. Environ.* **2020**, *3*, 36–49. [\[CrossRef\]](#)
- Sauni, M.; Luomala, H.; Kolisoja, P.; Turunen, E. Investigating root causes of railway track geometry deterioration—A data mining approach. *Front. Built Environ.* **2020**, *6*, 122. [\[CrossRef\]](#)
- International Union of Railways (UIC). *Railway Infrastructure. Laying and Maintenance of Track. Maintaining and Improving Earthworks and Track Bed Layers*; Technical Report Ed1; UIC: Paris, France, 2023.
- Burrow, M.; Bowness, D.; Ghataora, G. A comparison of railway track foundation design methods. *Proc. Inst. Mech. Eng. Part J. Rail Rapid Transit* **2007**, *221*, 1–12. [\[CrossRef\]](#)
- Haddani, Y.; Breul, P.; Saussine, G.; Navarrete, M.A.B.; Ranvier, F.; Gourvès, R. Trackbed mechanical and physical characterization using PANDA®/geoendoscopy coupling. *Procedia Eng.* **2016**, *143*, 1201–1209. [\[CrossRef\]](#)
- Burzawa, A.; Bodet, L.; Dhemaied, A.; Dangeard, M.; Pasquet, S.; Vitale, Q.; Boisson-Gaboriau, J.; Cui, Y.J. Detecting mechanical property anomalies along railway earthworks by Bayesian appraisal of MASW data. *Constr. Build. Mater.* **2023**, *404*, 133224. [\[CrossRef\]](#)
- Obando Hernandez, E.; Hölscher, P.; Doornenbal, P.; Mas, C.; van 't Schip, J.; van Uiter, A. Characterization of Shallow Ground in Railway Embankments Using Surface Waves Measured by Dark Fiber Optics Sensors: A Case Study. *Sensors* **2023**, *23*, 9397. [\[CrossRef\]](#) [\[PubMed\]](#)
- Esvelde, C.; Esvelde, C. *Modern Railway Track*; MRT-Productions: Zaltbommel, The Netherlands, 2001; Volume 385.
- Paixao, A.; Fortunato, E.; Calçada, R. A contribution for integrated analysis of railway track performance at transition zones and other discontinuities. *Constr. Build. Mater.* **2016**, *111*, 699–709. [\[CrossRef\]](#)
- Aydın, Y.; Işıkdag, Ü.; Bekdaş, G.; Nigdeli, S.M.; Geem, Z.W. Use of machine learning techniques in soil classification. *Sustainability* **2023**, *15*, 2374. [\[CrossRef\]](#)
- Sally, N.O.N. Combining Hard and Soft Voting Machine Learning Algorithms for Soil Classification. In Proceedings of the Deep Learning Indaba, Accra, Ghana, 3 September 2023.
- Sastre, C.; Breul, P.; Benz Navarette, M.; Bacconnet, C. Automatic soil identification from penetrometric signal by using artificial intelligence techniques. *Can. Geotech. J.* **2021**, *58*, 1148–1158. [\[CrossRef\]](#)
- Chala, A.T.; Ray, R. Assessing the performance of machine learning algorithms for soil classification using cone penetration test data. *Appl. Sci.* **2023**, *13*, 5758. [\[CrossRef\]](#)
- Carvalho, L.O.; Ribeiro, D.B. A multiple model machine learning approach for soil classification from cone penetration test data. *Soils Rocks* **2021**, *44*, e2021072121. [\[CrossRef\]](#)
- Chala, A.T.; Ray, R.P. Machine Learning Techniques for Soil Characterization Using Cone Penetration Test Data. *Appl. Sci.* **2023**, *13*, 8286. [\[CrossRef\]](#)
- Oberhollenzer, S.; Premstaller, M.; Marte, R.; Tschuchnigg, F.; Erharder, G.H.; Marcher, T. Cone penetration test dataset Premstaller Geotechnik. *Data Brief* **2021**, *34*, 106618. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rauter, S.; Tschuchnigg, F. Identification of soil strata from in-situ test data using machine learning. In Proceedings of the International Conference of the International Association for Computer Methods and Advances in Geomechanics; Springer: Cham, Switzerland, 2022; pp. 37–44.
- Rogiers, B.; Mallants, D.; Batelaan, O.; Gedeon, M.; Huysmans, M.; Dassargues, A. Model-based classification of CPT data and automated lithostratigraphic mapping for high-resolution characterization of a heterogeneous sedimentary aquifer. *PLoS ONE* **2017**, *12*, e0176656. [\[CrossRef\]](#) [\[PubMed\]](#)
- Saussine, G.; Dhemaied, A.; Delforge, Q.; Benfeddoul, S. Statistical analysis of cone penetration resistance of railway ballast. *EPJ Web Conf.* **2017**, *140*, 16011. [\[CrossRef\]](#)
- Punetha, P.; Nimbalkar, S. An innovative rheological approach for predicting the behaviour of critical zones in a railway track. *Acta Geotech.* **2023**, *18*, 5457–5483. [\[CrossRef\]](#)
- AL-Abdullah, S.F.; Alani, Z.; Zaidan, M.; Aldahwi, S. An Approach Study of Reducing the Sub-Ballast Thickness of Railway Using Geotextiles. *GEOMATE J.* **2023**, *25*, 257–264. [\[CrossRef\]](#)

22. Lemaître, G.; Nogueira, F.; Aridas, C.K. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.* **2017**, *18*, 1–5.
23. Rauter, S.; Tschuchnigg, F. CPT data interpretation employing different machine learning techniques. *Geosciences* **2021**, *11*, 265. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.