



EXPLAINABLE INSIGHT INTO THE VISION-BASED CLASSIFICATION OF SOIL CORE SAMPLES FROM CLOSE-RANGE IMAGES

Andreas-Nizar Granitzer

Norwegian Geotechnical Institute, Norway. E-mail: andreas.nizar.granitzer@ngi.no

Johannes Beck

Helmut-Schmidt-University, Germany. E-mail: beckjo@hsu-hh.de

Johannes Leo, and Franz Tschuchnigg

Institute of Soil Mechanics, Foundation Engineering and Computational Geotechnics, Graz University of Technology, Austria. E-mail: leo@tugraz.at, franz.tschuchnigg@tugraz.at

Keywords: Core Sample, Ground Characterization, Vision Transformer, Computer Vision, Image Classification.

Abstract

The characterization of ground conditions typically involves the assessment of core sample images, representing a challenging task that requires expert knowledge and may be prone to human bias. This work proposes a computer vision (CV) approach for automated classification of soil core sample images. A labelled dataset comprising 3,607 squared core sample images serves as training and evaluation basis. Focus is placed on critical aspects in the model building and evaluation of class separability among main soil fractions. In addition, ongoing research streams to augment the CV pipeline and enhance the classification performance are briefly addressed.

1 Introduction

In many cases, the greatest technical and financial risks associated with building and civil engineering projects stem from ground conditions (Jaksa et al. 2005; Marzouk et al. 2024). To mitigate these risks and prevent structural distress, excessive conservatism, or unforeseen conditions that could lead to significant construction delays, geotechnical engineers conduct ground investigations. In the absence of sufficient knowledge about the local ground conditions, ground investigation programs typically involve intrusive procedures to extract and assess core samples. During core sampling, cylindrical sections of the ground are retrieved along vertical profiles using continuous sampling methods (ÖN ISO 22475-1). Core samples (see **Figure 1**) serve as the foundation for laboratory tests to determine geotechnical design parameters, and to enable the identification and classification of the ground conditions according to normative standards, such as ÖN EN ISO 14688-1 and DIN 4023.

Traditionally, a core sample is first visually examined in terms of its color, grain size distribution, and layering. This visual inspection, which provides insight into the main fraction (e.g., sand, gravel), is typically followed by a manual examination to resolve remaining uncertainties, identify secondary fractions, and assess the consistency of the core sample. In practice, the visual and manual inspection of core samples represent a time-consuming and costly procedure that may be prone to human bias (Johnson et al. 2023). Since the documentation of this logging process is typically supplemented with high-resolution images of the core samples (see **Figure 1**), Brinkgreve and Zekri (2024) emphasize the significant potential of Computer Vision (CV) techniques (Srivastava et al. 2021) to accelerate interpretation while ensuring reproducibility and transparency of core sample classification tasks. With respect to geotechnical engineering, however, the authors further concede that this research topic has not received adequate attention to date, particularly in the case of soil formations.

2 Background

From an algorithmic point of view, CV pipelines for automated image-based knowledge extraction regarding ground properties consist of four steps: (i) image acquisition, (ii) image segmentation, (iii) feature extraction, and (iv) image classification. Based on this, Srivastava et al. (2021) distinguishes two principal approaches to automated image understanding. The first approach involves pipelines with hand-crafted feature extraction using both, image processing techniques (Kornblith et al. 2018; Khandelwal 2025) and traditional Machine Learning methods for image classification, such as Support Vector Machines and Random Forests. The second approach relies on Deep Learning (DL); for example, see Soranzo et al. (2025).



Figure 1. Core sample image obtained from geotechnical site investigation, along with core sample imaging setup housed in a photo box of the BAW soil mechanics laboratory in Hamburg.

The development of Convolutional Neural Network (CNN) (Lecun et al. 1998) and Vision Transformer (ViT) (Dosovitskiy et al. 2020) variants as backbone architectures for CV applications has propelled the second approach based on DL for image-based knowledge extraction in cognate disciplines, especially in agricultural (Chatterjee et al. 2021; Jagetia et al. 2022) and petroleum engineering (Alzubaidi et al. 2021). According to Kameswari et al. (2023), CNNs excel at small datasets, whereas ViTs tend to yield improved results at large datasets. Details concerning the relative merits of either model architecture, including hybrid variants, however, are beyond the scope of this work and can be found elsewhere; for example, see Liu et al. (2021). Succinctly, the general trend towards DL-based pipelines can be attributed to their automated feature extraction (i.e., image features used for image classification tasks are learned by the model), automated image segmentation capabilities (e.g., to remove the background in core sample images), and enhanced performance in terms of generalization and accuracy.

Subsequently, a pre-trained ViT base model variant, known for its effectiveness in image classification tasks, is evaluated as backbone architecture for the downstream task of soil core sample image classification. The pretrained model weights are fine-tuned employing a human-annotated dataset comprising squared soil core sample images prepared by Katinah (2023).

3 Study Design

The experimental study design serves two primary objectives: firstly, to deploy a dataset that can be utilized as a valuable resource for generating end-to-end CV pipelines for automated core sample image classification; and secondly, to use this curated dataset for fine-tuning and evaluating a CV pipeline on this downstream task. The latter is prototyped using a serverless notebook environment with a T4 GPU (Bisong 2019).

3.1 Data acquisition

A total of 204 core sample images were retrieved from 19 boreholes with a diameter of 100 mm, associated with the exploration project "Schleuse Lüneburg" (Helfers et al. 2018). The boreholes reached depths ranging from 5.5 – 81 m. The core samples were obtained using a dry percussion drilling method with casing. The original core sample images were captured in the BAW soil mechanics laboratory in Hamburg using a Canon EOS 70D camera positioned approximately 110 cm above the samples. The imaging setup was housed in a photo box with dimensions of approximately 110×110 cm, illuminated by two REXROTH SL30 LED lamps; see **Figure 1**. Camera settings were adjusted based on the characteristics of the samples and have been documented accordingly. From the original high-resolution images ($3,648 \times 5,472$ pixels, 72 DPI; see **Figure 1**), a total of 3,607 square images (300×300 pixels, 72 DPI) were manually cropped using the GIMP image editor, as specified by Katinah (2023). Each cropped image attempts to contain only a uniform soil type, minimizing transitions between layers. Metadata describing the images was manually derived from human-annotated core drilling protocols. The categorical class attributes describing the main fraction (HB) and secondary fraction (NB) are assigned in accordance with the German standard DIN 4023; see **Figure 2**. For example, mS and T denote core sample images where the HB feature is classified as medium Sand (> 0.20 und ≤ 0.63 mm) and Clay (≤ 0.002 mm), respectively. At this regard, however, it should be mentioned that the results presented in this work are constrained to the prediction of the main fraction for brevity.

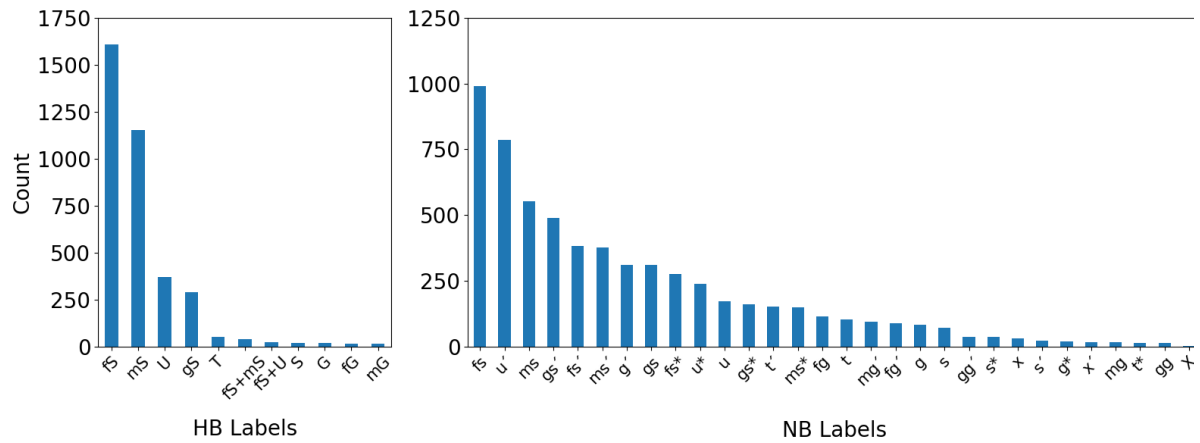


Figure 2. Class-wise distribution of main fraction and secondary fraction in curated dataset.

3.2 Data preparation and exploration

The initial data cleansing stage during the integration of the human-annotated metadata and soil core samples images involved a number of preprocessing steps concerning HB, including, but not restricted to, the removal of instances with duplicated, corrupted or zero-valued fields. Instances belonging to sparse class labels occurring fewer than three times are removed, as they could not be represented across the training, validation, and test splits.

Figure 2 shows the distribution of the unique class labels for the categorical features HB and NB in the final dataset, which is organized in a dictionary-type “DatasetDict” format (Wolf et al. 2019) with a 80% / 10% / 10% train / validation / test split. Obviously, both feature distributions are skewed. To mitigate the class imbalance problem (Sun et al. 2009), the samples are binned into six principal output classes for HB, as defined by DIN 4023: Gravel (G), coarse Sand (gS), medium Sand (mS), fine Sand (fS), Silt (U) and Clay (T). Samples with more than one main fraction, such as “fS+U”, are not considered to reduce ambiguity in the classification task.

3.3 Base model and hyperparameter selection

Based on a preliminary round to identify the most promising base model candidate in terms of computational efficiency and classification performance (Granitzer 2025), the Swin-T V2 (Liu et al. 2021) base model with approximately 27.6M parameters, pre-trained on ImageNet-1k (Deng et al. 2009) at a resolution of 256×256 pixels, is fine-tuned employing the transformers library (Wolf et al. 2019). Following standard procedures, the validation split is used to establish the termination criteria for the fine-tuning process, while the final evaluation is conducted on the test split (Steiner et al. 2021).

The hyperparameters are selected following best practice procedures using the automated hyperparameter optimization framework Optuna (Akiba et al. 2019), in combination with the Tree-structured Parzen Estimator algorithm (Watanabe, 2023). Based on this optimization, we employ the AdamW optimizer (Loshchilov and Hutter 2017) with a warmup ratio of 0.1 and early stopping (patience = 3), along with a batch size of 16 (Masters and Luschy 2018). Cross-Entropy Loss (Paszke et al. 2019) is used to address the multi-class classification problem involving six target labels. Image preprocessing included resizing, tensor conversion, and normalization. For the training split, on-the-fly data augmentation was applied to reduce memory consumption.

4 Results and Discussion

Key results concerning the performance and class separability are presented. Likewise, a demonstration case is included where we apply the fine-tuned model to one randomly selected core sample image per target label.

4.1 Performance assessment and demonstration case

Table 1 and **Figure 3** report the performance of the fine-tuned model in terms of Loss, Top-1 Accuracy, and aggregated F1-Scores (both Weighted and Macro), computed using the sklearn library (Pedregosa et al. 2012), and the confusion matrix. The aggregated F1-Score Macro (0.86) is particularly valuable in imbalanced settings, as it provides insight into the ability of a model to identify minority classes (Fan and Lin 2007). It should be noted that a comparable classification task is presented by Katinah (2023), who fine-tuned a VGG16-based CNN backbone (Simonyan and Zisserman 2014) for predicting fS, mS, gS, and U using the same raw dataset. In the present study, the Top-1 Accuracy improved from 82% to 89.2%, despite the inclusion of additional class labels (i.e., T and G), highlighting the relatively improved predictive capability of the model developed in this work.

Table 1. Performance metrics on the test set for fine-tuned model with extended HB target range (multi-class classification).

Base Model	Training Epochs	Loss	Top-1 Accuracy	F1-Score (Weighted)	F1-Score (Macro)
Swin-T V2	8	0.31	89.2%	0.89	0.86

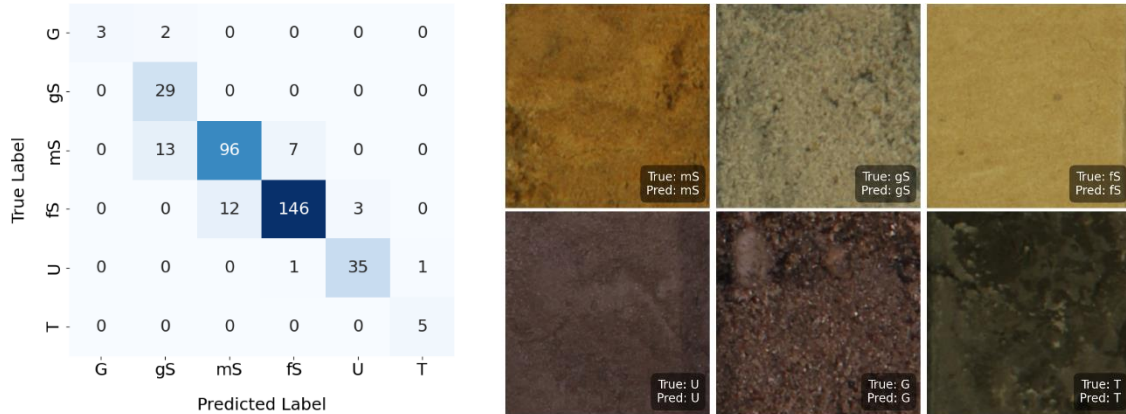


Figure 3. Confusion matrix evaluated on test set, alongside inferences for one randomly selected sample per target label.

4.2 Class separability

Figure 4 explores the ability of the fine-tuned model to learn discriminative patterns between target classes by aggregating patch embeddings into a single high-dimensional feature vector per image and visualizing them in 2D using t-SNE (van der Maaten and Hinton 2008). This approach captures a global semantic summary of each image. Clear separation in the t-SNE plot indicates effective feature learning, while overlap suggests class confusion. The results highlight the potential to maintain coherent clustering across the target labels, particularly among sand subclasses (fS, mS, gS), though distinguishing fine-grained classes (T, U) remains challenging. These findings reinforce empirical observations from previous studies (Soranzo et al. 2025; Katinah 2023) in an intuitive manner and suggest potential benefits from more balanced datasets and enhanced image quality.

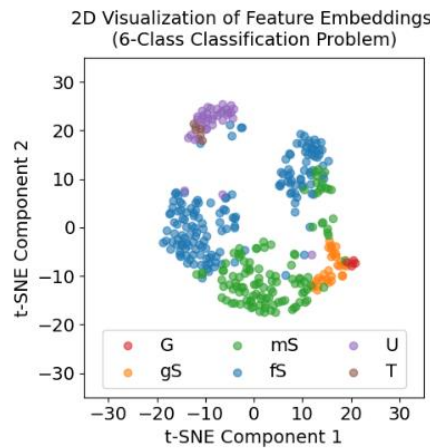


Figure 4. 2D visualization of feature embeddings evaluated on the test set.

5 Conclusions

This work highlights the significant potential of CV pipelines as intelligent tools for characterizing soil core sample images. A pretrained ViT-type model is fine-tuned using a curated dataset consisting of soil core sample images and human-annotated metadata. Empirical results demonstrate that the base model choice has a considerable impact on the classification performance. The prototype developed with the best-performing base model excels in the multi-class classification of six target labels representing the main fraction of soil core sample images. Efforts to extend the classification to include secondary fractions are ongoing. Future research will likely explore the use of image segmentation techniques to accelerate the core sample image preprocessing procedure and the integration of image enhancement techniques to further improve class separability.

6 Acknowledgements

Data used in this study was kindly provided by the Federal Waterways Engineering and Research Institute (BAW).

References

- Akiba, Takuya; Sano, Shotaro; Yanase, Toshihiko; Ohta, Takeru; Koyama, Masanori (2019): Optuna: A Next-generation Hyperparameter Optimization Framework. In *arXiv preprint*. Available online at <http://arxiv.org/pdf/1907.10902>.
- Alzubaidi, Fatimah; Mostaghimi, Peyman; Swietojanski, Pawel; Clark, Stuart R.; Armstrong, Ryan T. (2021): Automated lithology classification from drill core images using convolutional neural networks. In *Journal of Petroleum Science and Engineering* 197, p. 107933. DOI: 10.1016/j.petrol.2020.107933.
- Bisong, Ekaba (2019): Google Colaboratory. In Ekaba Bisong (Ed.): *Building Machine Learning and Deep Learning Models on Google Cloud Platform. A Comprehensive Guide for Beginners*. New York: Apress (Springer eBook Collection), pp. 59–64.
- Brinkgreve, Ronald; Zekri, Ashraf (2024): On the use of Machine Learning in Geotechnical Engineering. In Bundesanstalt für Wasserbau (BAW) (Ed.): *BAW Kolloquium. Numerik in der Geotechnik*. Karlsruhe, Nov 7-8, pp. 15–23.
- Chatterjee, Kajal; Obaidat, Mohammad S.; Samanta, Debabrata; Sadoun, Balqies; Islam, Hafizul; Chatterjee, Rajdeep (2021): Classification of Soil Images using Convolution Neural Networks. In : 2021 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI). Beijing, China, 15.10.2021 - 17.10.2021: IEEE, pp. 1–5.
- Deng, Jia; Dong, Wei; Socher, Richard; Li, Li-Jia; Li, Kai; Fei-Fei, Li (2009): ImageNet: A large-scale hierarchical image database. In : 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops). Miami, FL, 20.06.2009 - 25.06.2009: IEEE, pp. 248–255.
- Dosovitskiy, Alexey; Beyer, Lucas; Kolesnikov, Alexander; Weissenborn, Dirk; Zhai, Xiaohua; Unterthiner, Thomas et al. (2020): An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *arXiv preprint*. Available online at <http://arxiv.org/pdf/2010.11929>.
- ÖN EN 1997-2, 2022: Eurocode 7: Geotechnical design.
- Fan, Rong-En; Lin, Chih-Jen (2007): A Study on Threshold Selection for Multi-label Classification. In *National Taiwan University*. Available online at <https://api.semanticscholar.org/CorpusID:17029248>.
- ÖN ISO 22475-1, 2022: Geotechnical investigation and testing — Sampling methods and groundwater measurements.
- ÖN EN ISO 14688-1, 2020: Geotechnische Erkundung und Untersuchung - Benennung, Beschreibung und Klassifizierung von Boden.
- DIN 4023, 2023: Geotechnische Erkundung und Untersuchung - Zeichnerische Darstellung der Ergebnisse von Bohrungen und sonstigen direkten Aufschlüssen.
- Granitzer, Andreas-Nizar (2025): Data-centric core sample classification and constitutive model selection in geotechnics. Master's Thesis. University of Applied Sciences Kufstein, Kufstein.
- Helfers, Björn; Henke, Sascha; Kaya, Hatice (2018): Planung der Baugrube für eine neue Schleuse am Elbe-Seitenkanal in Lüneburg. In *Bautechnik* 95 (9), pp. 663–672. DOI: 10.1002/bate.201800040.
- Jagetia, Aaryan; Goenka, Umang; Kumari, Priyadarshini; Samuel, Mary (2022): Visual Transformer for Soil Classification. In *arXiv preprint*. Available online at <http://arxiv.org/pdf/2209.02950>.
- Jaksa, M. B.; Goldsworthy, J. S.; Fenton, G. A.; Kaggwa, W. S.; Griffiths, D. V.; Kuo, Y. L.; Poulos, H. G. (2005): Towards reliable and effective site investigations. In *Géotechnique* 55 (2), pp. 109–121. DOI: 10.1680/geot.55.2.109.59531.
- Johnson, Sam; Sagasan, Yasin; Stryk, Antoinette (2023): Extracting consistent geotechnical data from drill core imagery using Computer Vision at the Carrapateena deposit. In : Proceedings of the 4th AEGC. Geoscience - Breaking New Ground. 4th Australasian Exploration Geoscience Conference, March 13-18, pp. 1–7.
- Kameswari, Ch. Sita; J, Kavitha; Reddy, T. Srinivas; Chinthaguntla, Balaswamy; Jagatheesaperumal, Senthil Kumar; Gaftandzhieva, Silvia; Doneva, Rositsa (2023): An Overview of Vision Transformers for Image Processing: A Survey. In *IJACSA* 14 (8). DOI: 10.14569/IJACSA.2023.0140830.
- Katinah, Loay (2023): Bodenklassifizierung basierend auf Bilderkennung. Bachelorarbeit. Helmut-Schmidt-Universität Hamburg, Hamburg.
- Khandelwal, Neetika (2025): Image Processing in Python: Algorithms, Tools, and Methods You Should Know. <https://neptune.ai/blog/image-processing-python>. neptune.ai. Available online at <https://neptune.ai/blog/image-processing-python>, updated on 1/21/2025, checked on 3/24/2025.
- Kornblith, Simon; Shlens, Jonathon; Le, Quoc V. (2018): Do Better ImageNet Models Transfer Better? In *arXiv preprint*. DOI: 10.48550/arXiv.1805.08974.
- Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. (1998): Gradient-based learning applied to document recognition. In *Proc. IEEE* 86 (11), pp. 2278–2324. DOI: 10.1109/5.726791.
- Liu, Ze; Lin, Yutong; Cao, Yue; Hu, Han; Wei, Yixuan; Zhang, Zheng et al. (2021): Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *arXiv preprint*. DOI: 10.48550/arXiv.2103.14030.
- Loshchilov, Ilya; Hutter, Frank (2017): Decoupled Weight Decay Regularization. In *arXiv preprint*. Available online at <http://arxiv.org/pdf/1711.05101>.
- Marzouk, Islam; Granitzer, Andreas-Nizar; Rauter, Stefan; Tschuchnigg, Franz (2024): A Case Study on Advanced CPT Data Interpretation: From Stratification to Soil Parameters. In *Geotech Geol Eng* 42 (5), pp. 4087–4113. DOI: 10.1007/s10706-024-02774-9.
- Masters, Dominic; Luschi, Carlo (2018): Revisiting Small Batch Training for Deep Neural Networks. In *arXiv preprint*. Available online at <http://arxiv.org/pdf/1804.07612>.
-

- Paszke, Adam; Gross, Sam; Massa, Francisco; Lerer, Adam; Bradbury, James; Chanan, Gregory et al. (2019): PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *arXiv preprint*. DOI: 10.48550/arXiv.1912.01703.
- Pedregosa, Fabian; Varoquaux, Gaël; Gramfort, Alexandre; Michel, Vincent; Thirion, Bertrand; Grisel, Olivier et al. (2012): Scikit-learn: Machine Learning in Python. In *arXiv preprint*. DOI: 10.48550/arXiv.1201.0490.
- Simonyan, Karen; Zisserman, Andrew (2014): Very Deep Convolutional Networks for Large-Scale Image Recognition. In *arXiv preprint*. Available online at <http://arxiv.org/pdf/1409.1556>.
- Soranzo, Enrico; Guardiani, Carlotta; Wu, Wei (2025): Convolutional neural network prediction of the particle size distribution of soil from close-range images. In *Soils and Foundations* 65 (1), p. 101575. DOI: 10.1016/j.sandf.2025.101575.
- Srivastava, Pallavi; Shukla, Aasheesh; Bansal, Atul (2021): A comprehensive review on soil classification using deep learning and computer vision techniques. In *Multimed Tools Appl* 80 (10), pp. 14887–14914. DOI: 10.1007/s11042-021-10544-5.
- Steiner, Andreas; Kolesnikov, Alexander; Zhai, Xiaohua; Wightman, Ross; Uszkoreit, Jakob; Beyer, Lucas (2021): How to train your ViT? Data, Augmentation, and Regularization in Vision Transformers. In *arXiv preprint*. DOI: 10.48550/arXiv.2106.10270.
- Sun, Y.; Wong, A. K. C.; Kamel, M. S. (2009): Classification of imbalanced data: A review. In *Int. J. Patt. Recogn. Artif. Intell.* 23 (04), pp. 687–719. DOI: 10.1142/S0218001409007326.
- van der Maaten, Laurens; Hinton, Geoffrey (2008): Visualizing Data using t-SNE. In *Journal of Machine Learning Research* 9 (86), pp. 2579–2605.
- Wolf, Thomas; Debut, Lysandre; Sanh, Victor; Chaumond, Julien; Delangue, Clement; Moi, Anthony et al. (2019): HuggingFace's Transformers: State-of-the-art Natural Language Processing. In *arXiv preprint*. Available online at <http://arxiv.org/pdf/1910.03771>.