Archives of Acoustics Vol. ***, No. ***, https://doi.org/10.24423/archacoust.2026.4351

Online First June 3, 2026

Technical Note

Semi-Supervised Learning for Sediment Classification Using Convolutional Neural Networks with Digital Elevation Model and Backscatter Data

Paweł Nadachowski[(1)], Zbigniew Łubniewski[(1)*], Karolina Trzcińska[(2)], Radosław Wróblewski[(3)], Maria Rucińska[(2)], Jarosław Tęgowski[(2)]

[(1)] *Faculty of Electronics, Telecommunications and Informatics, Department of Geoinformatics, Gdansk University of Technology*
Gdansk, Poland

[(2)] *Faculty of Oceanography and Geography, Department of Geophysics, University of Gdansk*
Gdynia, Poland

[(3)] *Maritime Institute, Gdynia Maritime University*
Gdansk, Poland

[*]Corresponding Author: lubniew@eti.pg.edu.pl

Received October 2, 2025; accepted February 3, 2026; version of record June 3, 2026; published issue XXXX.

Accurate sediment classification is crucial for advancing marine research, environmental monitoring, and sustainable seabed use. However, acquiring large amounts of labeled data in such settings is often challenging, expensive, and time-consuming. To address this limitation, a semi-supervised learning framework has been proposed that leverages convolutional neural networks for sediment classification using both labeled and unlabeled data. The approach utilizes pseudo-labeling, where confident predictions on unlabeled samples are iteratively incorporated into training to enhance model generalization. The method is applied and evaluated on a dataset that includes multi-modal inputs such as a digital elevation model and multibeam sonar backscatter data. Experimental results indicate that semi-supervised learning with convolutional neural networks can achieve high classification accuracy in scenarios characterized by limited labeled data and a large volume of unlabeled data. This approach highlights the potential of deep learning combined with semi-supervised strategies for efficient underwater environment classification.

Keywords: convolutional neural network (CNN), deep learning, digital elevation model (DEM), multibeam sonar backscatter, pseudo label, residual neural network (ResNet), sediment classification, semi-supervised learning (SSL).

1. Introduction

Supervised seabed sediment classification requires a large amount of ground truth data, which is very expensive to obtain. This is especially troublesome if there is a need to use deep learning methods, such as convolutional neural networks (CNNs) (Krizhevsky *et al.*, 2012), which typically require large-scale annotated datasets to achieve optimal generalization. In marine and geological environments, collecting accurately labeled data is particularly challenging due to logistical constraints, high costs of field surveys, and the need for expert interpretation of sediment types. Consequently, there is a growing interest in exploring data-efficient learning strategies to reduce dependency on large amounts of ground truth data (Qin *et al.*, 2021; Zhao *et al.*, 2023).

To alleviate this problem, this study examines how semi-supervised learning (SSL) (Hady, Schwenker, 2013) methods can improve model performance by leveraging both labeled and unlabeled data. One such SSL technique is the pseudo-labeling (Lee, 2013) method, which offers a simple yet effective approach to extend the training dataset. In pseudo-labeling, a supervised model initially trained on a small subset of labeled data is used to predict labels for the unlabeled samples. Predictions with high confidence are then treated as ground truth (pseudo-labels) and used for further training, effectively augmenting the training dataset with automatically annotated examples.

By incorporating pseudo-labeled samples, CNNs can be trained to better generalize from sparse labeled data, capturing more complex patterns and improving classification accuracy on unseen samples. This method is especially relevant in sediment classification tasks, where expert-labeled examples are scarce and expensive to obtain. The pseudo-labeling framework also allows for the use of different confidence thresholds to control the quality of the generated pseudo-labels. This study evaluates the impact of different confidence thresholds on model performance, aiming to identify the optimal balance between label quality and data coverage. Results demonstrate that, when applied carefully, pseudo-labeling can achieve high classification accuracy while minimizing the need for exhaustive manual annotation.

2. Study site and methods

The experiments were conducted at a single study site, where multiple processing and analysis steps were carried out using different techniques. Each step was aimed at optimizing the classification process and improving overall performance. The applied approach, which included data acquisition, ground-truth labeling, model selection and SSL implementation, ultimately led to high classification accuracy.

2.1. Study site

The study area encompasses a section of the seabed, covering nearly 10 km, at depths ranging from 3.5 m to 22 m. It is located less than one kilometer offshore, north of the village of Rowy, in the southern Baltic Sea, in the Polish Territorial Sea, within the Natura 2000 site (code PLB990002, Birds Directive), partially overlapping with the Słowiński National Park. The study site location is presented in Fig. 1.

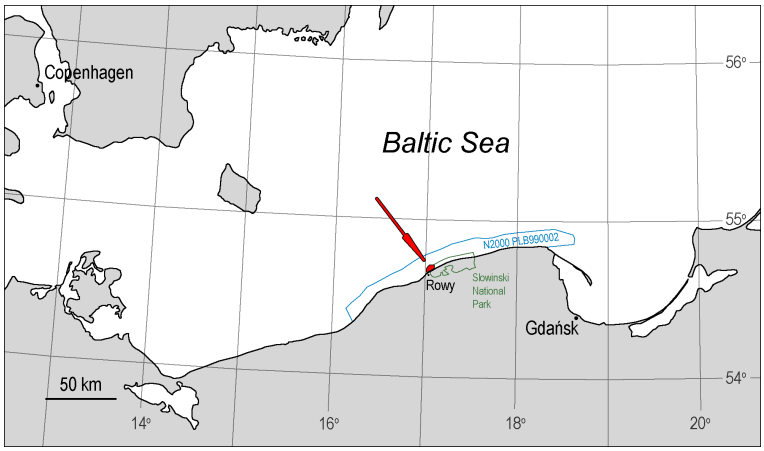

Fig. 1. Study site location.

The seabed structure and morphology in this region are primarily shaped by glacial processes that occurred during the Pleistocene, followed by deglaciation and subsequent developments associated with the evolution of the Baltic Sea during the Holocene. The seabed sediment composition includes Pleistocene glacial and fluvioglacial sediments, as well as Holocene marine deposits.

The predominant geomorphological feature of the area is a rugged relief formed by irregular, elevated outcrops of cohesive sediments, often with surface cobbles and boulders, interspersed with accumulations of sand and gravel.

The eastern and southeastern portions of the area are characterized by relatively level, gently undulating sandy surfaces. Similar sand accumulations are also present in patches across the central and western parts of the study area.

2.2. Data acquisition

Bathymetric and backscatter data were recorded between 2018 and 2020 using a Teledyne Reson SeaBat 7125 multibeam echosounder (MBES) mounted on the R/V Oceanograf vessel belonging to the University of Gdańsk. The MBES recording was performed at a frequency of 400 kHz using 512 beams with a width of 1° along-track and 0.5° across-track. Registration was performed in the QPS Qinsy software, combining information from MBES, sound velocity profiles in the water column, locations, and rotations of the measuring unit. A Global Navigation Satellite System/Inertial Navigation System supported by real time kinematic (RTK) corrections from the NAWGEO network (ASG-EUPOS[1]) provided real-time positioning with a few cm accuracy. A Valeport miniSVP Sound Velocity Profiler was used to record the speed of sound in the water column. The Octans gyrocompass and motion sensor were used to record the roll, pitch, heave, and heading of the vessel during measurements.

2.3. Ground-truth labeling

In the study area, ground-truth samples were obtained by collecting surface sediments using a Van Veen grab sampler and by recording video of the seabed with an underwater camera. The locations of the samples were recorded using GPS during measurement campaigns in 2018–2020. The sample locations were selected to take into account the diversity of sediments based on knowledge of the study area, bathymetric and backscatter maps, while maintaining, to the extent possible, an even spatial distribution of samples across the mapped area. However, due to the limited survey duration, the data set is not very large and contains 73 samples. The sediment samples were analyzed using granulometric methods in the laboratory, comprising Folk and Ward parameters as well as the Wentworth classification system. The underwater camera recordings were analyzed by an expert who assessed the type of sediment at the bottom, which was particularly important in places where it was not possible to collect samples with a grab, e.g., boulder fields. After analyzing all samples and recordings in the study area, generalizations were made and the sediments were divided into four classes: boulders (B), sandy gravel or gravelly sand (SG-GS), sand (S), very fine sand (VFS). Samples of recordings for each sediment type are presented in Fig. 2 and the visualization of bathymetry and backscatter data with the locations of different sediment types is shown in Fig. 3.

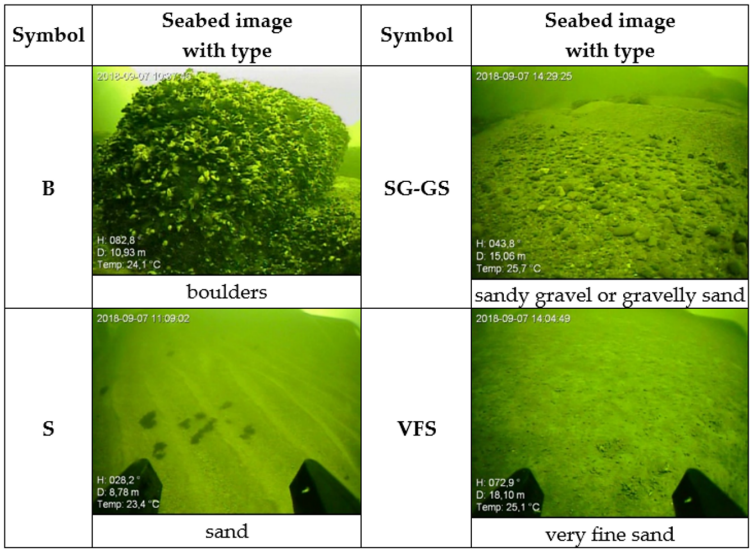

Fig. 2. Samples of recordings for each sediment type.

[1] https://www.asgeupos.pl

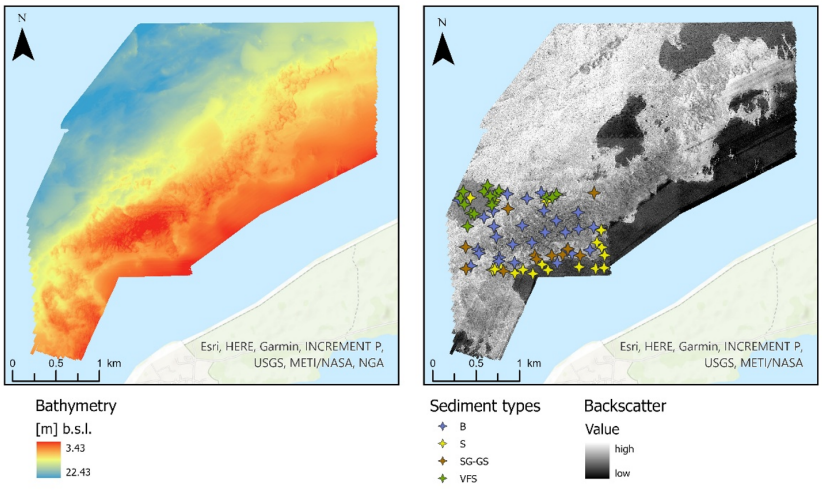

Fig. 3. Visualization of bathymetry and backscatter data with sediments.

2.4. Data preprocessing

Following the collection of bathymetric and backscatter data, it was processed to be ready for analysis. The MBES datasets were processed in QPS Qimera and Fledermaus Geocoder Toolbox (FMGT). In Qimera software, bathymetric data was cleaned and mosaicked, then exported to Geotiff with a resolution of 0.5 m × 0.5 m pixel size in the Universal Transverse Mercator (zone 33N) projected coordinate system. Similarly, the backscatter mosaic exported from FMGT was saved after manual removing of outliers as Geotiff. The backscatter mosaicking method used beam time series (snippets) and angle varying gain (AVG) correction within the Geocoder engine (Fonseca, Calder, 2005). The AVG method, commonly used to correct MBES angular dependency and normalize seafloor backscatter, was applied with default settings: 'flat' mode for noise reduction, 'blend' mosaicking for managing swath overlaps, and a 300-ping processing window (Schimel *et al.*, 2018; Parnum, Gavrilov, 2011).

Since convolutional neural networks were employed in this study, the preprocessed data was converted into a structured grid format suitable for the model input. Both labeled and unlabeled data were prepared in this way. Regarding the labeled dataset, a 16 × 16 grid was generated for each ground-truth point, capturing neighboring values from both the DEM and multibeam sonar backscatter channels. Next, the datasets were divided into two groups based on the year of acquisition: 31 samples from 2018 and 42 samples from 2019. In each group the data were imbalanced, with some classes significantly underrepresented. To address this issue, the imbalanced-learn (Lemaître *et al.*, 2017) library was used to perform oversampling of the minority classes to match the size of the majority class.

For the unlabeled dataset, 3000 random points were generated using QGIS tools, ensuring that they did not overlap with the area of the labeled data. For each of these points, a 16 × 16 grid was also created, incorporating neighboring values from both the DEM and multibeam sonar backscatter channels.

2.5. Convolutional neural network

CNNs are well-suited for solving problems in the field of computer vision. In this study, the input data consist of gridded representations of the digital elevation model (DEM) and multibeam sonar backscatter values, which can be treated as image-like inputs. Therefore, CNN-based architectures are a natural choice for addressing this classification task.

A specific type of convolutional network is the residual network (ResNet) (He *et al.*, 2016) which was employed in this study. ResNet architecture is known for its use of residual connections, which help mitigate the vanishing gradient problem and enable the effective training of deep neural networks. ResNet architecture has many variants and ResNet-18 was selected for this research. This particular architecture consists of 18 layers (17 convolutional and 1 fully connected) and is relatively lightweight, which makes it more suitable for training

on datasets with a limited number of labeled (ground-truth) samples. Its reduced complexity helps prevent overfitting while maintaining sufficient depth to capture spatial patterns in the input data. Visualization of ResNet-18 architecture is presented in Fig. 4.

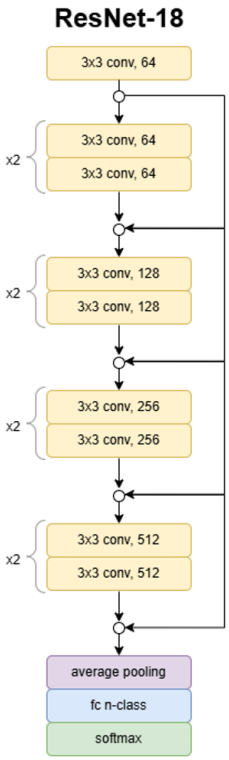

FIG. 4. Visualization of the ResNet-18 architecture.

2.6. SEMI-SUPERVISED LEARNING

The SSL is a technique that can be used when a small labeled dataset is available alongside a large volume of unlabeled data. A diagram of the required steps for the SSL is presented in Fig. 5.

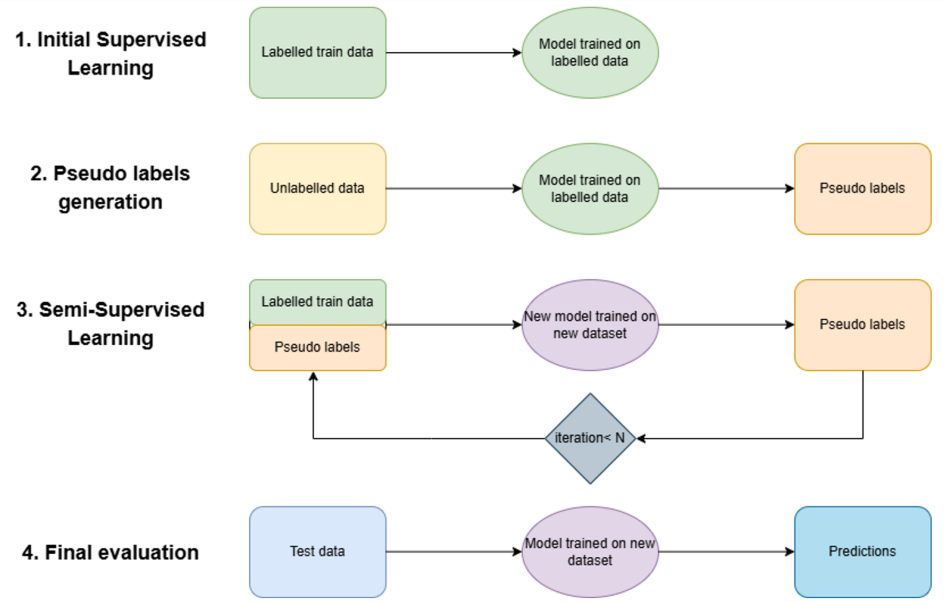

FIG. 5. Schema of required steps for SSL.

The first step is to train a ResNet-18 instance on a small labeled dataset. The labeled data was divided into training and test datasets based on the year of acquisition: samples from 2018 were used for training, and samples from 2019 for testing. In this study, the ResNet-18 model from the PyTorch (Paszke *et al.*, 2019) library's torchvision module was used. The hyperparameters were selected using a grid search method. The final values were: a learning rate of 1×10^{-7}, a batch size of 8, the CrossEntropyLoss as a loss function, and the Adam (Kingma, Ba, 2017) as an optimization algorithm.

After training the model on the supervised data, it can be used in the second step, where pseudo-label generation takes place. To generate pseudo-labels, a confidence threshold is selected, and all predictions on the unlabeled dataset with probabilities greater than or equal to this threshold are used as pseudo-labels. In this study, several confidence threshold values ranging from 0.6 to 0.9 were evaluated.

In the next step, the pseudo-labels are incorporated into the labeled dataset in order to train a new instance of ResNet-18 from scratch. This process can be repeated iteratively to expand and refine the dataset. The iterative approach requires a termination condition, which can be defined as an integer indicating the number of iterations to perform. In this study, the termination condition was set to 14 iterations, which, combined with the initial supervised training, results in 15 total training iterations, each consisting of 100 epochs.

After each training iteration, the model's accuracy on the test dataset is evaluated. The model with the best accuracy is saved as checkpoint, and if a subsequent model performs worse, the saved model is loaded.

When the termination condition is met, the final model performance is evaluated on the test dataset.

3. Results

Throughout the study, various confidence thresholds were applied to each instance of the SSL approach, and their performance was subsequently evaluated on the test dataset.

3.1. Comparison of different confidence thresholds

In the experiment, confidence thresholds ranging from 0.6 to 0.9 were tested for SSL. For each threshold value, a new instance of the model was trained from scratch. Each instance yielded different accuracy results and changes in dataset size across iterations. Charts with results and the dataset size change for each instance are presented in Fig. 6.

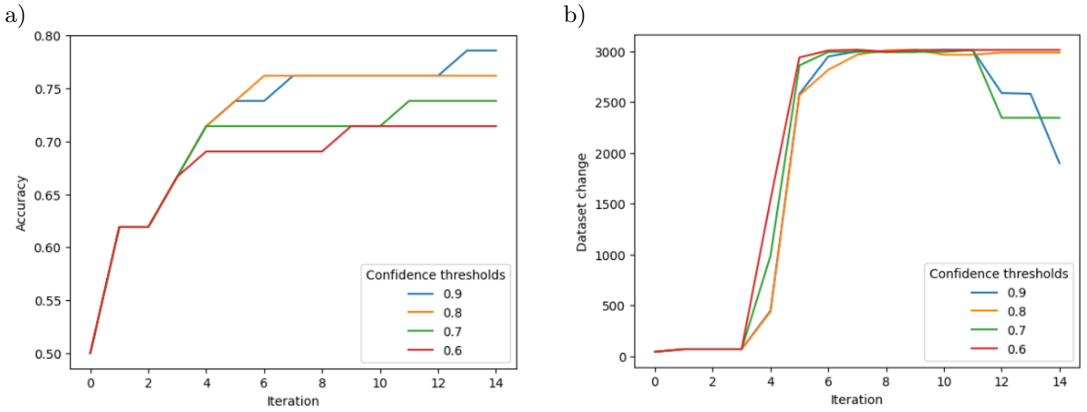

Fig. 6. Comparison of accuracy (a) and dataset size change (b) throughout each iteration of training for different confidence thresholds.

As shown in the charts, the overall accuracy on the test dataset ranges from 71.4 % to 78.6 %. The best performance was observed at a confidence threshold of 90 %, which includes only the most confident predictions as pseudo-labels. In comparison with other instance the results indicate that higher confidence thresholds are associated with improved model performance, likely due to the inclusion of more reliable pseudo-labels during training. Additionally, the dataset initially received very few pseudo-labels during the first three iterations of each

of the model instances. This was followed by a rapid increase to approximately 3 000 samples from the unlabeled dataset. In some cases, including the best-performing model instance, it discarded certain pseudo-labeled samples, likely to reduce noise and preserve label quality, thereby facilitating additional gains in performance.

3.2. Best model instance

Table 1 presents detailed information regarding the best-performing instance of the ResNet-18 model, obtained after completing all iterations of the SSL process with a confidence threshold set at 90 %. The table includes the confusion matrix with prediction information for each type of sediment and performance metrics such as: producer's accuracy (recall), user's accuracy (precision), omission error, commission error, overall accuracy, kappa, and $F1$-score.

Table 1. Confusion matrix for the best ResNet-18 instance evaluated on test dataset.

	Reference				**Sum**
	Boulders	Sand	Sandy gravel	Very fine sand	
Boulders	8	3	0	0	**11**
Sand	0	6	0	0	**6**
Sandy gravel	2	1	6	0	**9**
Very fine sand	1	2	0	13	**16**
Sum	**11**	**12**	**6**	**13**	**42**
Producer's accuracy	0.727	0.500	1.000	1.00	–
User's accuracy	0.727	1.000	0.667	0.813	–
Omission error	0.273	0.500	0.000	0.000	–
Commission error	0.273	0.000	0.333	0.187	–
Overall accuracy	0.786	–	–	–	–
Kappa	0.711	–	–	–	–
$F1$-score (weighted)	0.757	–	–	–	–

The confusion matrix reveals that the most challenging types of sediments to accurately predict is the category of boulders and sand classes. Specifically, boulders are frequently misclassified as sandy gravel and very fine sand, while the sand class itself is often confused with all of the other classes. This overlap makes sand the most difficult sediment type for the model to correctly identify. In contrast, sandy gravel and very fine sand were classified with relatively high accuracy, indicating better model performance on these classes. Other evaluation metrics further demonstrate the model's effectiveness, with an overall accuracy of 0.786 and a weighted $F1$-score of 0.757, reflecting solid predictive capability of the model.

4. Discussion

From this study, it can be concluded that selecting an appropriate confidence threshold during SSL has a noticeable impact on the final results. Each of the thresholds tested – from 60 % to 90 % led to different outcomes by the end of the SSL sessions. Although the differences in accuracy are not very large, ranging from 71.4 % to 78.6 %, the confidence threshold adjustment can still be significant, given that the same model architecture and hyperparameters were used consistently across all SSL experiments.

In terms of prediction details for the model using the optimal 90 % confidence threshold, several misclassification patterns were observed. Specifically, boulders were frequently misclassified as either sandy gravel or very fine sand. Additionally, the sand class posed a challenge, being commonly confused with all other sediment types. This class overlap made sand the most difficult sediment type for the model to accurately identify. In contrast, sandy gravel and very fine sand were classified with relatively high accuracy, indicating that the model performed better in these categories.

Misclassification of boulders as gravel may occur due to altered surface characteristics caused by red algae covering the boulders. Their presence increases the acoustic roughness and backscatter intensity of individual

boulders, which may be mistakenly interpreted as clusters of smaller rock fragments or finer sediments. Consequently, these features can be incorrectly classified as gravel accumulations rather than single boulders.

The misinterpretation of sandy seabed areas results from the specific nature of sand accumulations and their high susceptibility to seabed-shaping processes. In some locations, sand deposits form flat, even surfaces, while in others they form areas with megaripples or distinctly undulating relief. This variability in the morphology of sandy substrates can lead to divergent or inaccurate interpretations.

The analysis of sandy gravel and very fine sand yielded very good results. These areas are generally homogeneous, with limited morphological variability. Unlike sandy seabeds, which are often shaped by dynamic seabed processes. This consistency explains the relatively high accuracy achieved in the interpretation of gravelly-sand areas.

Future work could focus on some improvements to the performance of SSL for sediment classification. One promising direction is the incorporation of other SSL strategies, such as co-training (Chen *et al.*, 2024) or graph based SSL (Song *et al.*, 2021) which may also lead to more reliable results. Additionally exploring more sophisticated model architectures, such as Vision Transformers (Dosovitskiy *et al.*, 2021) or newer CNN architectures like EfficientNet (Tan, Le, 2019), may also offer performance gains. Finally, expanding the dataset with more diverse and balanced samples would help improve generalization of the models.

5. Conclusion

Sediment classification is an important topic of marine environmental research, which enables a better understanding of geological processes, habitat mapping, and the impact of human activities on underwater areas. The results obtained in this study confirm that the use of SSL methods is effective under conditions of limited availability of labeled data and a large volume of unlabeled data. By leveraging both annotated and unannotated samples, SSL approaches significantly reduce the dependence on labor-intensive and costly manual labeling, while maintaining competitive classification performance. A key conclusion that came out from this study is that the effectiveness of these methods depends on the confidence thresholds used to select pseudo-labeled samples, which directly influence the quality and reliability of the training process. The presented method increase the possibilities for practical application of deep learning in seafloor classification tasks and other domains where the acquisition of labeled samples is expensive, time-consuming, or logistically challenging. The findings highlight the potential of SSL as a cost-effective solution for advancing automated interpretation of seabed data.

Fundings

The work was supported by the National Science Center, Poland, project STREAMBAL No. 2021/41/B/ST10/01086.

Conflict of interest

The authors declare that there are no known competing financial interests or personal relationships that could have influenced the work described in this paper.

Authors' contributions

All authors contributed to the conceptualization of the study and participated in the analysis and interpretation of the results. Paweł Nadachowski was responsible for the study design. Maria Rucińska carried out data collection and pre-processing. The original draft of the manuscript was prepared by Paweł Nadachowski, Karolina Trzcińska, and Radosław Wróblewski. Revision of the manuscript was undertaken by Zbigniew Łubniewski, Jarosław Tęgowski, and Maria Rucińska. All authors agree to be accountable for all aspects of the work. All authors reviewed and approved the final manuscript.

References

1. Chen M., Du Y., Zhang Y., Qian S., Wang C. (2024), Semi-supervised learning with multi-head co-training, *arXiv*, http://arxiv.org/abs/2107.04795.
2. Dosovitskiy A. *et al.* (2021), An image is worth 16x16 words: Transformers for image recognition at scale, *arXiv*, http://arxiv.org/abs/2010.11929.
3. Fonseca L., Calder B. (2005), Geocoder: An efficient backscatter map constructor, [in:] *Proceedings of the U.S. Hydro 2005 Conference*.
4. Hady M.F.A., Schwenker F. (2013), Semi-supervised learning, [in:] *Handbook on Neural Information Processing*, Bianchini M., Maggini M., Jain L. [Eds], **49**: 215–239, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-36657-4˙7.
5. He K., Zhang X., Ren S., Sun J. (2016), Deep residual learning for image recognition, [in:] *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, https://doi.org/10.1109/CVPR.2016.90.
6. Kingma D.P., Ba J. (2017), Adam: A method for stochastic optimization, *arXiv*, http://arxiv.org/abs/1412.6980.
7. Krizhevsky A., Sutskever I., Hinton G.E. (2012), ImageNet classification with deep convolutional neural networks, [in:] *Proceedings of the 26th International Conference on Neural Information Processing Systems*.
8. Lee D.H. (2013), Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, [in:] *Workshop on Challenges in Representation Learning*.
9. Lemaître G., Nogueira F., Aridas C.K. (2017), Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning, *Journal of Machine Learning Research*, **18**(17): 1–5.
10. Parnum I.M., Gavrilov A.N. (2011), High-frequency multibeam echo-sounder measurements of seafloor backscatter in shallow water: Part 2 – Mosaic production, analysis and classification, *Underwater Technology*, **30**(1): 13–26, https://doi.org/10.3723/ut.30.013.
11. Paszke A. *et al.* (2019), PyTorch: An imperative style, high-performance deep learning library, [in:] *Proceedings of the 33rd International Conference on Neural Information Processing Systems*.
12. Qin X., Luo X., Wu Z., Shang J. (2021), Optimizing the Sediment Classification of Small Side-Scan Sonar Images Based on Deep Learning, *IEEE Access*, **9**: 29416–29428, https://doi.org/10.1109/ACCESS.2021.3052206.
13. Schimel A.C.G. *et al.* (2018), Multibeam sonar backscatter data processing, *Marine Geophysical Research*, **39**: 121–137, https://doi.org/10.1007/s11001-018-9341-z.
14. Song Z., Yang X., Xu Z., King I. (2021), Graph-based semi-supervised learning: A comprehensive review, *arXiv*, http://arxiv.org/abs/2102.13303.
15. Tan M., Le Q. (2019), EfficientNet: Rethinking model scaling for convolutional neural networks, [in:] *Proceedings of the 36th International Conference on Machine Learning*, **97**: 6105–6114.
16. Zhao Y., Zhu K., Zhao T., Zheng L., Deng X. (2023), Small-sample seabed sediment classification based on deep learning, *Remote Sensing*, **15**(8): 2178, https://doi.org/10.3390/rs15082178.