scientific reports

OPEN
Sediment core analysis using artificial intelligence

Andrea Di Martino[1,4], Gianluca Carlini[2,4], Gastone Castellani[3], Daniel Remondini[2✉] & Alessandro Amorosi[1✉]

Subsurface stratigraphic modeling is crucial for a variety of environmental, societal, and economic challenges. However, the need for specific sedimentological skills in sediment core analysis may constitute a limitation. Methods based on Machine Learning and Deep Learning can play a central role in automatizing this time-consuming procedure. In this work, using a robust dataset of high-resolution digital images from continuous sediment cores of Holocene age that reflect a wide spectrum of continental to shallow-marine depositional environments, we outline a novel deep-learning-based approach to perform automatic semantic segmentation directly on core images, leveraging the power of convolutional neural networks. To optimize the interpretation process and maximize scientific value, we use six sedimentary facies associations as target classes in lieu of ineffective classification methods based uniquely on lithology. We propose an automated model that can rapidly characterize sediment cores, allowing immediate guidance for stratigraphic correlation and subsurface reconstructions.

Understanding subsurface stratigraphy is essential for a wide range of industrial and societal applications, including studies of global climate change[1,2], reservoir characterization[3–5], land subsidence calculations[6,7], and engineering geology[8]. When approaching the investigation of the subsurface, by nature inaccessible to direct observation, sediment cores are the fundamental source of information. Sedimentary facies, in particular, i.e. sediment bodies or packages of strata formed in specific depositional environments, bear unique physical and mechanical properties[9] that can be used effectively for subsurface stratigraphic modeling.

Recent studies have shown that building a detailed model of the shallow subsurface based on sedimentary facies properties can be an effective tool to: (i) assess patterns of active tectonic deformation[10], (ii) define site response to earthquakes[11], and (iii) predict earthquake damage risk[12].

Sediment facies analysis is the first step in most Earth and environmental research studies; nevertheless, high-resolution facies reconstructions require specific sedimentological expertise and training, usually involving a multidisciplinary research approach[9,13,14].

Recent advances in Artificial Intelligence research are setting new standards for many research fields, with automated methods based on Machine Learning (ML) and Deep Learning (DL) achieving state-of-the-art performance in solving complex problems. Among the principal applications of AI methods, we find Natural Language Processing (NLP)[15], Computer Vision (CV)[16], synthetic data generation[17], and more. In the last few years, AI methods have been increasingly applied to Earth and environmental research[18–25]. However, the proposed approaches did not fully exploit the potential of ML and DL systems and the usage of AI in geoscience. In a recent publication, Fleming et al.[26] pointed out the necessity for a deeper understanding of AI and automated algorithms to strengthen geosciences research policies[27–29].

An automatic approach was recently proposed to classify Holocene sediment facies using X-ray fluorescence (XRF) scanner data[30]. This approach, however, although effective on a local scale, can hardly be exported outside the study area: XRF data rely mostly on sediment composition, which can vary greatly from site to site in the same depositional environment simply as a function of sediment dispersal.

In this context, we propose a novel approach leveraging DL to perform automatic semantic segmentation of sediment cores digital images directly acquired in the field. Semantic segmentation consists in classifying each image pixel according to a specific set of categories, and Convolutional Neural Networks (CNNs) usually achieve state-of-the-art performance[31,32]. CNNs are a particular class of networks primarily used to efficiently analyze image data. We identified six target Holocene sedimentary classes from the Po Plain and the Adriatic

[1]Department of Biological, Geological and Environmental Sciences (BiGeA), University of Bologna, Piazza di Porta San Donato 1, 40126 Bologna, Italy. [2]Department of Physics and Astronomy, University of Bologna, 40127 Bologna, Italy. [3]Department of Medical and Surgical Sciences, University of Bologna, 40138 Bologna, Italy. [4]These authors contributed equally: Andrea Di Martino and Gianluca Carlini. ✉email: daniel.remondini@unibo.it; alessandro.amorosi@unibo.it

coastal plain of Marche, Abruzzo, and Apulia regions (Italy): Well-drained floodplain (WDF), Poorly-drained floodplain (PDF), Swamp (Sw), Peat layer (PL), Prodelta (P), and Fluvial sand (FS), deposits, with an additional background class. An expert sedimentologist manually annotated each core image, producing a final dataset of 82 non-overlapping, high-resolution digital images acquired from 32 continuous sediment cores with the associated segmentation masks. To perform the model validation, we divided the dataset into three mutually-exclusive subsets: training, validation, and test, containing 77, 11, and 12% of the data, respectively.

Our method can produce precise semantic segmentation and, thus, accurate facies interpretation, achieving high scores for the most used segmentation metrics. Our approach can drastically reduce the time and effort required to analyze core surveys; it can perform real-time predictions of high-resolution images on a regular computer and could be extended to mobile devices, making it suitable for on-field usage. This method does not need expensive data acquisition techniques or pre-processing, since it relies on images acquired with common digital cameras. Moreover, this approach is not necessarily limited to Holocene successions and could be adapted to different geologic conditions.

One of the major concerns with Deep Learning methods is the limited interpretability of model predictions. This is the reason why they are commonly called black-box methods[33–35]. In an attempt to better understand the results produced by our method, we visualize the regions of higher segmentation error and model prediction confidence, enhancing the results interpretability.

Results
Segmentation performance and visual evaluation

A deep CNN was trained to automatically produce the semantic segmentation masks of digital images acquired from continuous sedimentary cores. The images included six target classes corresponding to the observed sedimentary facies and a seventh background class. The model performance was evaluated on validation and test sets consisting of 11 and 12% of the total available data, respectively. We measured several standard segmentation metrics: the mean Intersection over Union (IoU), the F1-score, and the balanced accuracy. The results obtained for the validation and test data are shown in Table 1. The performance achieved on the validation dataset tends to be positively biased due to the model being fine-tuned on it, whereas the test performances are more rigorous in evaluating the capabilities of the model. The scores obtained with both datasets show no remarkable differences.

For a visual evaluation of the model performance, we produced the semantic segmentation mask of five full-resolution images from both the validation and test sets and compared them to the ground truths produced by the expert sedimentologist (Figs. 1 and 2). The detailed quantitative presentation of the results and errors is covered in Section "confusion matrix and model misclassifications" and shown Fig. 3. The five images were chosen as the most representative for the two datasets, showing all the target sedimentary facies. The visual performance on validation data shows high correlation between the model predictions and the ground truths. Figure 1A is one of the most complex images in the whole dataset, containing four target classes: Well-drained floodplain (WDF), Poorly-drained floodplain (PDF), Swamp (Sw), and Peat layer (PL). The model prediction accurately reproduces the sedimentologist segmentation mask, correctly classifying most sedimentary facies. Minor errors are present, mainly localized at facies transitions (WDF–PDF, PDF-Sw, and Sw-PL). The transition between Swamp and Peat layer is also visible in Fig. 1B and well classified by the model. In Fig. 1C, the model misclassified a portion of the Fluvial sand (FS) stratigraphic interval, while classifying the PDF correctly. Figure 1D shows the high capability of the model in classifying the Prodelta (P) target class, whose identification commonly requires specific sedimentological training, with an impressive overlap between the model prediction and the reference segmentation. The model performs well in Fluvial sand classification in Fig. 1E, while only minor classification errors are present for Sw and PDF.

There are no noticeable visual performance differences between the validation and test images. Furthermore, four out of the five images shown in Fig. 2 come from a set of sediment cores for which no images were present in the training set (red dots in Fig. 2). The usage of this set of images was intended to simulate the model application to external data acquired in the field, to validate its generalization capabilities. In Fig. 2A, WDF, PDF, and Sw are correctly classified, with satisfactory predictions also near facies transitions. Negligible errors are present, with a minor misclassification of PDF to Sw; Fig. 2B, C, D) shows an almost perfect model prediction for the four involved classes. The model confirms its robust prediction capabilities of sub-features in the sedimentary record, represented by PL. Consistent with the validation result, the Prodelta class is very well classified also in the test dataset. In Fig. 2E, WDF is correctly classified, while minor errors occur on Sw and PDF classification. However, in this case the model struggles to reproduce the sedimentologists sand classification.

Confusion Matrix and model misclassifications

For a deeper quantitative evaluation of the model performance, the confusion matrices for the validation and test sets were calculated (Fig. 3A–B). The confusion matrix is a table layout in which each row represents the instances

	Mean IoU	F1-score	Balanced accuracy
Validation set	0.884	0.936	0.905
Test set	0.853	0.916	0.861

Table 1. Model performance obtained on validation and test data, in terms of mean Intersection over Union, F1-score, and balanced accuracy.
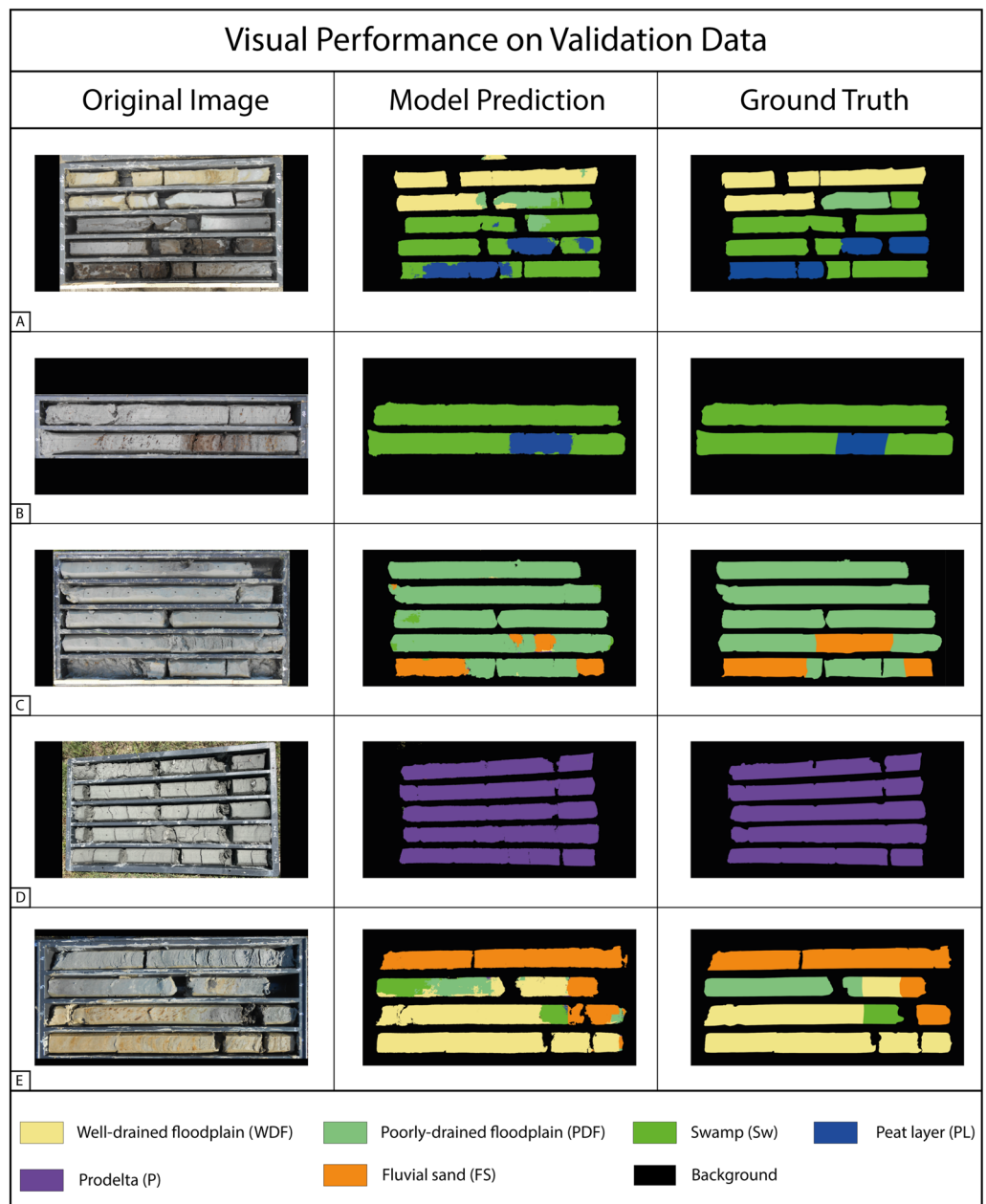
Figure 1. Visual performance of the model on five representative images of the validation dataset. The original full-resolution digital images, the model-produced segmentation masks, and the corresponding ground truths are shown in the left, central, and right columns, respectively.

in a ground truth class, while each column represents the class instances predicted by the model. In a semantic segmentation context, the instances are the pixels associated with each class. This matrix shows which classes have been correctly classified and which were confused with other classes during the model test. It is a standard estimator used in machine learning and statistics, since it provides more information about model performance than the standard metrics[36]. Most of the standard evaluation metrics can be derived from the confusion matrix, so it can be considered the most comprehensive method for performance evaluation in classification problems[37–40]. However, the confusion matrix becomes more complex to read with the growing number of classes.

We normalized the confusion matrix with respect to the number of true instance classes, i.e., to the rows of the matrix. It follows that the resulting matrix values are between zero and one, with one representing a perfect classification; a colormap was used for a more intuitive visual evaluation.

The confusion matrix for the validation data (Fig. 3A) shows good classification performance for all the classes. Excluding the background, the class with the highest classification accuracy is the Prodelta (0.963), followed by: Swamp (0.918), Well-drained floodplain (0.894), Poorly-drained floodplain (0.872), Fluvial sand (0.862), and Peat layer (0.838). The most significant classification errors occur for the PL, being misclassified as Sw (0.141), FS misclassified as PDF (0.080), PDF confused as Sw (0.068), and WDF confused as PDF (0.067).
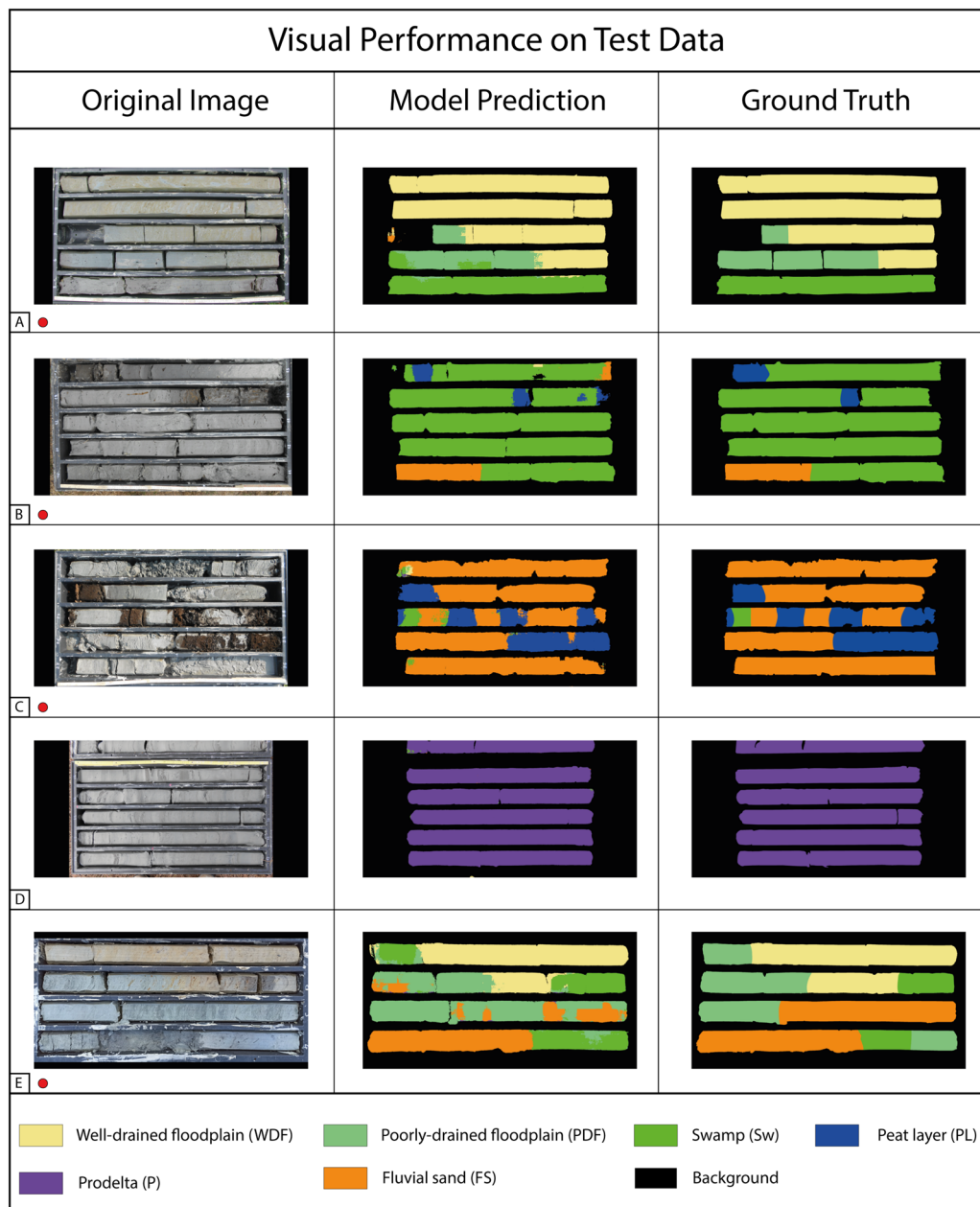
Figure 2. Visual performance of the model on five representative images of the test dataset. The original full-resolution digital images, the model produced segmentation masks, and the corresponding ground truths are shown in the left, central, and right columns, respectively. The red dots mark the images coming from the external set of sediment cores.

The highest classification accuracy for the test data, excluding the background, is achieved again for Prodelta (0.973), followed by: Well-drained floodplain (0.905), Swamp (0.901), Fluvial sand (0.823), Peat layer (0.750), and Poorly-drained floodplain (0.699). On average, the classification accuracies are slightly lower than the validation ones, in agreement with the metrics reported in Table 1. In this case, the most significative classification errors occur for the PDF being misclassified as Sw (0.191), followed by: PL misclassified as Sw (0.127) and FS (0.081), and FS confused as Sw (0.063).

The results reported in Table 1, the visual performance shown in Figs. 1 and 2, in combination with the confusion matrices of Fig. 3, show a robust classification result for all the target classes with only minor errors.

To better understand the limitations of the proposed CNN and the possible sources of error, we produced the error maps between model predictions and ground truths, along with the model prediction confidence (Fig. 4), for two representative cores from the validation and test datasets. The sediment core, ground truth, segmentation mask, and prediction are shown in Fig. 4A, B, C, F, G, H, respectively. In Fig. 4D–I, the model confidence is presented, with darker regions representing areas of lower confidence. Figure 4E–L reports the error between the prediction and the ground truth.
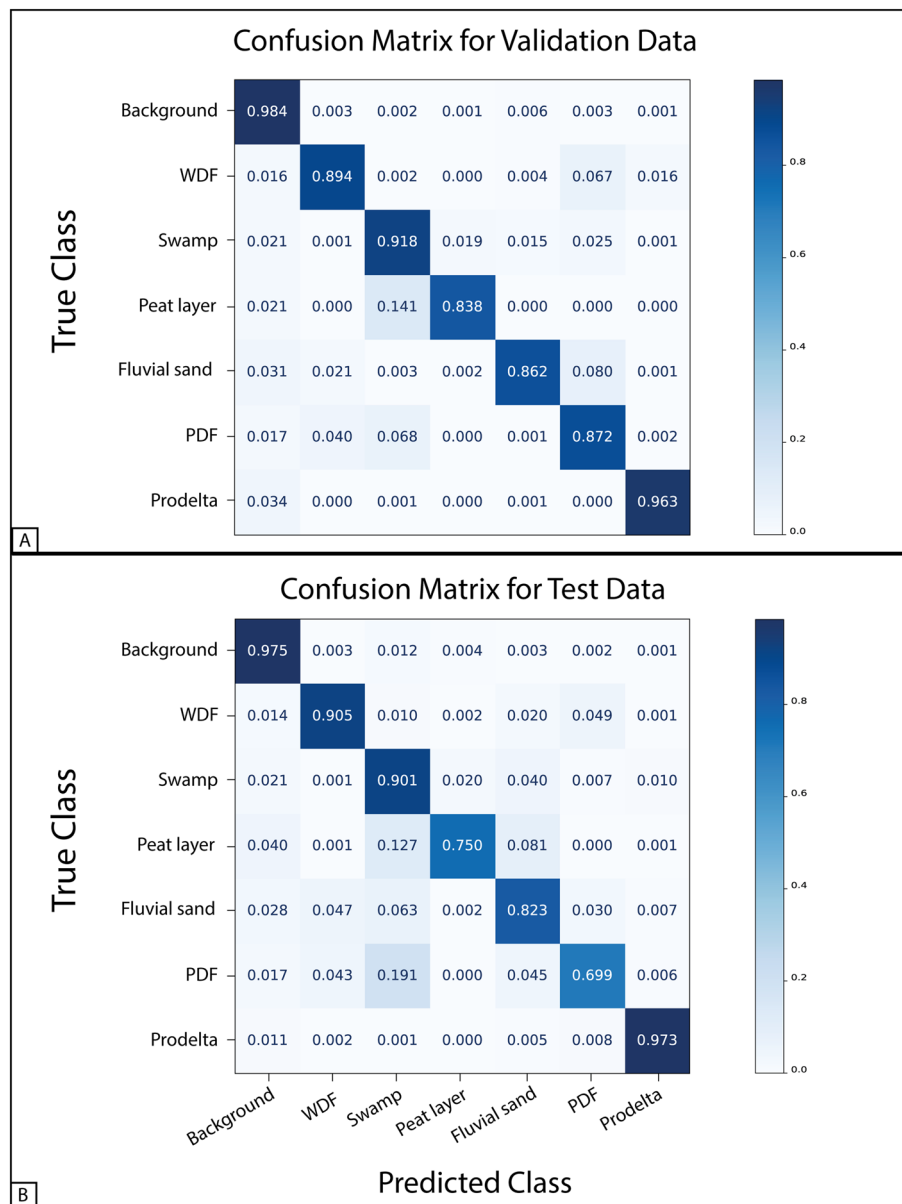
Figure 3. Confusion matrices for validation data (**A**) and test data (**B**). Each row of the matrices represents the instances in a ground truth class, while each column represents the class instances predicted by the model. The values were normalized with respect to the number of ground truth instances for each class. A colormap visually highlights the higher values with darker shades of blue.

The validation core (Fig. 4A) shows a prominent Peat layer as a sub-feature in the Swamp deposit, with no well-defined boundaries corresponding with the minimum confidence values, as shown in Fig. 4D. The model error is restricted to a smaller portion of the image, while no major errors are noticeable from other areas of the sedimentary core.

The test image (Fig. 4F) shows a possible real-case scenario with a poorly preserved core that was damaged during drilling operations. The high number of transition boundaries makes the prediction challenging; in Fig. 4I, a generally lower model confidence is visible, with minima located in correspondence of facies transitions and on the damaged parts of the core.

Discussion

A powerful method leveraging DL and CNNs is proposed to produce accurate sedimentary facies interpretations starting from standard digital images. It constitutes a fast, precise, and easy-to-deploy tool that could largely improve subsurface stratigraphic modeling, making subsurface facies analysis accessible to a wider range of scientists and professionals.

Using a convolutional neural network, we aimed to mimic the sedimentologist approach in facies classification; the model makes its prediction pixel-wise, but the decision process also considers local and global aspects of
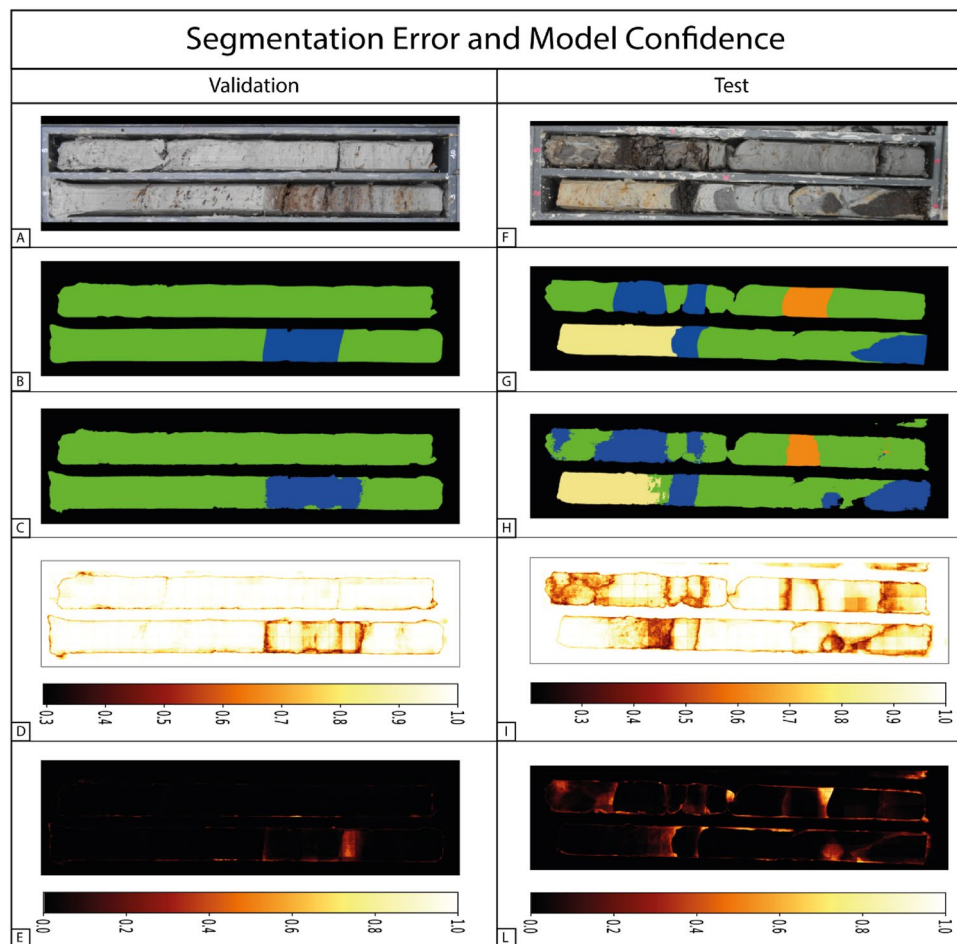
Figure 4. Segmentation error and model confidence. Core images (**A–F**), ground truth segmentation masks (**B–G**), model predictions (**C–H**), model confidence (**D–I**), and segmentation error (**E–L**) are shown for two validation and test representative cores. The model confidence represents the prediction probability associated with the predicted class. The segmentation error is the normalized categorical cross-entropy calculated between the prediction and the ground truth.

the image, such as the textural characteristics, color, and accessory materials, e.g., carbonate concretions, shells, and wood fragments[31,41]. The heterogeneity of the images used in the work and the data augmentation performed should make the model robust to different image lighting, contrast, orientation or quality.

The model performance obtained on the validation and test datasets shows robust generalization capabilities, with a strong agreement between the predicted classes and the ground truths identified by the sedimentologist. Furthermore, the scores obtained by the model in terms of mean IoU, F1-score, and balanced accuracy are notably high, considering the complexity of a multi-class semantic segmentation task. The numerical results are visually confirmed by the segmentation mask shown in Figs. 1, 2.

Swamp and Poorly-drained floodplain facies can have similar characteristics in terms of color and texture, thus can hardly be distinguished by visual inspection only. In such cases, the fossil content and geotechnical properties are fundamental tools for high-resolution facies analysis[9,42–44]. Our model generally performs well in discriminating these highly similar facies using visual information only, with minor errors. However, misclassification of Sw to PDF and vice-versa can hardly be considered real errors, because the distinction between these two classes is subtle and attributions could vary as a function of the sedimentologist expertise. Another typical misclassification error is due to the abundance of sand layers as sub-features within a clay deposit, such as in the case of WDF and PDF. These layers could be considered minor attributes by the sedimentologist, and thus ignored, or emphasized and interpreted as a fluvial facies. A further source of error is transition between sedimentary facies, which makes precise identification of their boundary challenging. The definition of a sharp boundary is a consequence and limitation of the semantic segmentation tasks; in some cases, however, the transition between two facies is not an abrupt change and should be considered more properly as a zone of transition. In this case, the model is not capable of fully reproducing the sedimentologist interpretation, as it may become highly subjective.

Given the black-box nature of neural networks, the information arising from the confidence and error maps can be used to highlight the criticisms of the proposed segmentation task. From Fig. 4, we acknowledge that overall the model correctly reproduces the sedimentologist interpretation.

For the validation core, the model error is restricted to a smaller portion of the image, and the overall stratigraphic interpretation does not change; moreover, the misclassified region corresponds to a facies transition, for which the identification of a sharp boundary is not straightforward, even for a sedimentologist. The test image shows a complex case in which the core was damaged during the drilling process. Also in this case, segmentation errors do not significantly impact the global stratigraphic interpretation. The damaged portions of the core are also well classified, though with a lower model confidence.

In summary, the sources of error can be grouped into three main categories: (i) visual overlap of sedimentary facies, (ii) occurrence of transitional facies boundaries, and (iii) subjectivity in sedimentologist interpretations.

There is growing interest in the application of AI methods to environmental and Earth sciences[18,19,25,45]. In this paper, we outline a novel approach to subsurface stratigraphy, performing semantic segmentation of Holocene sedimentary facies with convolutional neural networks. Previous subsurface studies have relied upon standard machine learning techniques, and for this reason they have been limited to simple data structures[20–22,30,46]. Our method leverages the power of deep learning models to produce accurate stratigraphic interpretations starting from digital images. This makes our approach easier to use and deploy in a wide range of geological applications.

During the last decades, a virtually continuous stratigraphic record of Holocene deposits in the Po Plain and beneath several coastal plains of Italy was acquired through core drilling. Robust high-resolution sequence-stratigraphic models that rely on information extracted from cores[47] demonstrated that the stratigraphic architecture of Holocene successions worldwide exhibits striking similarities in terms of sediment facies distribution. Sediment core analysis based on semantic segmentation of continuous core images, thus, represents a highly reproducible technique that is likely to be exported successfully to other coeval stratigraphic successions, constituting a reference framework for future CNN-based subsurface analysis. The research approach presented in this paper is naturally suited for in-situ analysis and could substantially reduce the time and effort needed for detailed sediment facies interpretation, making it a valuable tool for large-scale exploration and for a broad range of industrial applications. The future integration of imaging data with other data sources, such as geotechnical and compositional data, and their incorporation in an automated method based on machine learning can make a substantial contribution to the progress of geological research below the ground surface.

Materials and methods
Data acquisition and pre-processing
The dataset[48] used for this study consists of 82 digital images from 31 selected Holocene cored sedimentary successions of Italy (Po Plain and Adriatic coastal plains of Marche, Abruzzo, and Apulia regions), between December 2016 and July 2021 (Fig. 5A). Given the remarkable length (30–50 m) of sediment cores, a series of non-overlapping digital images were acquired every 5 m of recovered sediment, covering the whole core length. Digital images were obtained directly in the field using different devices, such as compact cameras and smartphones, with a broad range of resolutions, ranging from 1369×803 to 4605×2717 pixels. To make our method as general as possible, we did not enforce a strict image acquisition procedure in terms of camera settings and environmental conditions. Each image was resized to 3074×1538 pixels to obtain a homogeneous resolution. Whenever the aspect ratio of the target resolution was different from the aspect ratio of the image, this latter was padded with zero-valued pixels to maintain its original aspect ratio, i.e., to resize the image without distortions. No other pre-processing steps were carried out on the original images.

Using pre-existing, high-resolution stratigraphic reconstructions as a framework for facies interpretations[10,13,14], we produced the ground truth segmentation mask for every image (Fig. 5B), manually classifying the sediment core into seven classes: six target classes corresponding to the observed Holocene sedimentary facies, and a seventh background class. A segmentation mask is an image mapping each identified class to a corresponding unique RGB value. RGB masks are suited for visualization, but are not directly usable to train a segmentation model; thus, we quantized the original RGB masks, mapping each RGB triplet to a corresponding unique scalar value. There is a slight variation in the relative proportion of the six target classes, particularly for the Peat layer class, which is only locally observed, as it occurs at distinct stratigraphic intervals (Fig. 5).

The seven classes are as follows:

(i)   Well-drained floodplain deposits (WDF), typical of subaerially exposed environments, consist of stiff, varicolored light grey to green silt and clay, with yellowish and orange mottles due to Fe oxides. White carbonate concretions and other pedogenic features are common.
(ii) Poorly-drained floodplain deposits (PDF), characteristic of flood basins with fluctuating groundwater table, consist predominantly of soft, grey clay and silt, with scattered vegetal remains and a lack of body fossils.
(iii) Swamp deposits (Sw), typical of waterlogged environments, are dominated by grey to dark-grey clay, with abundant vegetal remains and wood fragments concentrated in discrete horizons or scattered along core sections.
(iv) Peat layers (PL) represent sub-features of swamp deposits, characterized by dark grey to black colors. They consist of wood fragments with subordinate clayey material and are typically organic-matter-rich.
(v)   Prodelta deposits (P) consist of homogenous, light grey clay formed at fluvial mouths, with common silt and sand intercalations, interpreted to represent flood layers. Plant debris and other organic matter are locally observed, whereas salt-water mollusks are common.
(vi) Fluvial sand (FS) includes a wide range of grain size fractions, from silty sand to very coarse sand, formed in fluvial/distributary channels or in adjacent areas (levees and crevasse splays).
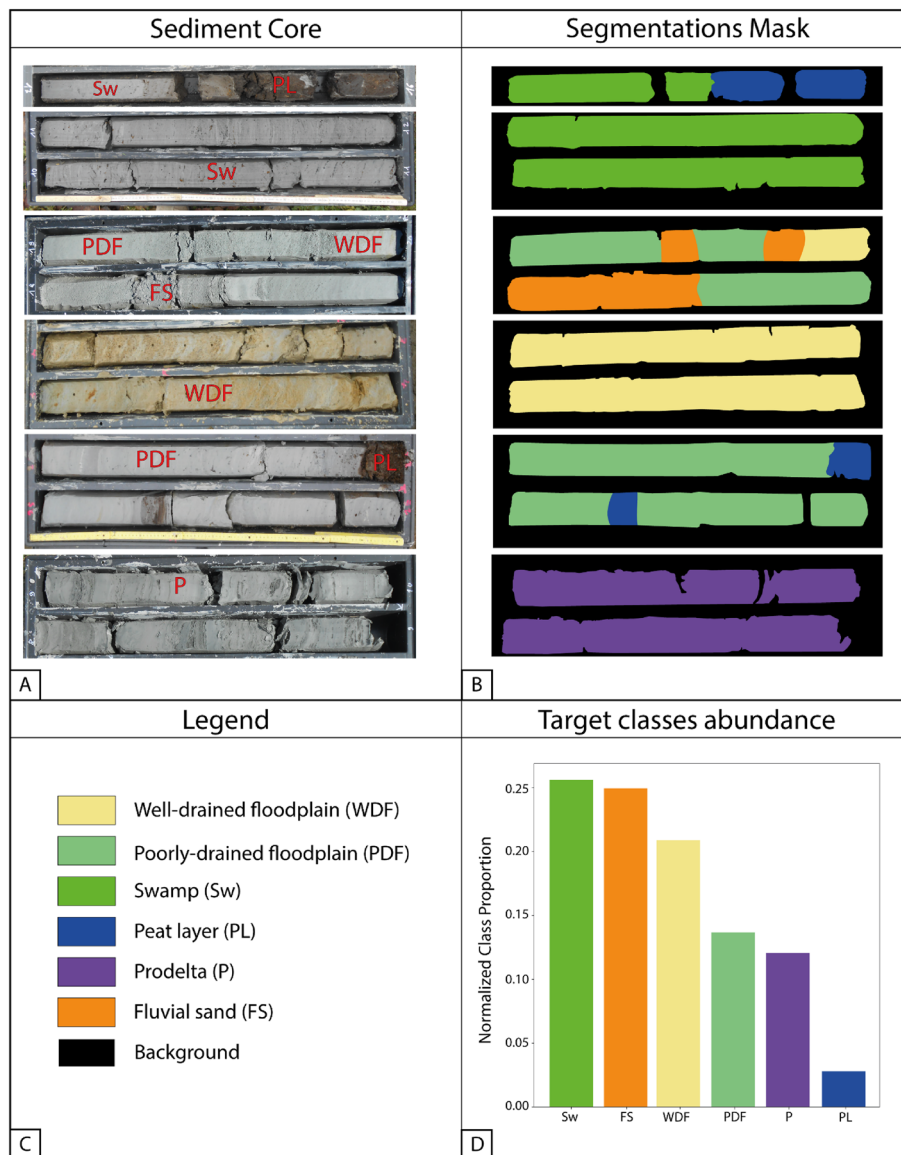(vii) Background, corresponding to the grey box containing the sediment core.

Figure 5. (**A**) Examples of digital images of continuous sediment cores with associated segmentation masks (**B**). (**C**) Target classes and background colors. (**D**) Relative target classes abundances.

The dataset was then divided into three, non-overlapping portions; 63 digital images were used for training, 9 for validations, and 10 for testing, corresponding to 76.83%, 10.97%, and 12.20% of the total, respectively. The data were stratified, so that each class was equally represented in every subset. To further validate the generalization capabilities of the model, one image in the validation dataset and six images in the test dataset were taken from sediment cores that were not present in the training dataset.

Due to computational limitations, full-resolution images could not be used for model training. Thus, following data subdivision, the images and corresponding masks were subdivided into (i) 1609 non-overlapping patches of resolution 384 × 384 pixels for model training; (ii) 250 patches for model validation; and (iii) 265 patches for model testing. Patch subdivision is also useful for increasing the number of available samples. The padding with zero-valued pixels during the resize operation could lead to the generation of patches with mostly zero-valued pixels at image borders. For this reason, patches with less than 5% non-zero pixels were automatically excluded from the data during the patch subdivision process.

Segmentation model

The model used for the image segmentation was a U-Net[49] with an EfficientNetB3 backbone[50], with weights pre-trained on ImageNet[51]. We chose the EfficientNetB3 as the backbone for the segmentation model because EfficientNets can achieve better performance than other popular model architectures[52], such as ResNets[53], while having a smaller number of parameters. Furthermore, to avoid overfitting, considering the limited number of training samples, we decided to employ a network with a relatively small number of parameters. Moreover, a lightweight model may also be used for real-time predictions without requiring a dedicated, powerful hardware.

The model was trained for 100 epochs using Adam[54] as optimizer, with a starting learning rate of 10^{-4} and a polynomial learning rate decay schedule. Data augmentation was used to improve the generalization capabilities of the model. The transformations used for data augmentation were: random rotation with a 360-degree range, random brightness variation, and random contrast variation. The loss function used for the model training was the categorical cross-entropy. During training, we monitored the mean Intersection over Union (IoU) as a measure of the model performance, and we saved the model weights achieving the highest mean IoU on the validation data. The IoU is a typical metric used in segmentation tasks; it measures the overlap between the predicted mask and the ground truth[55]. The IoU is zero when there is no overlap between the prediction and the ground truth, while it is equal to one for a perfect overlap.

We used the trained model to predict the image patches in the validation and test sets, and we computed the mean IoU, the F1-score, the balanced accuracy, and the confusion matrix to measure the prediction performance[56,57]. The convolutional neural network was built using the Tensorflow[58] python library, while the metrics and confusion matrix were calculated with the Scikit-learn[59] python library. In detail, the IoU, F1-score and accuracy are defined as follows:

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where

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

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

where TP is True Positive, FP is False positive, and FN is False Negative

$$\text{IoU} = \frac{TP}{TP + FP + FN}$$

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

where TN is True Negatives.

All the metrics were weighted with respect to the number of true instances for each class to take into account the classes unbalancing.

To predict a whole image, a sliding-window approach is used. A kernel of *patch-size* slides through the image with a stride of 96 pixels, corresponding to one-fourth of the patch dimension. The image portion identified by the kernel is given to the model to produce the prediction; then, the predicted regions are re-arranged to form the complete predicted image. With a stride smaller than the kernel dimension, the resulting predictions will overlap, with an overlapping portion depending on the stride size. The overlapping predictions are then averaged to produce the final predicted image, achieving a smoother result.

We produced visual maps for a thorough estimation of model errors and prediction confidence; we define as model prediction confidence the probability associated with the predicted class. The model confidence cannot be directly related to a measure of prediction uncertainty, since it could predict the wrong class, while being highly confident of its prediction. The error between the prediction and the ground truth is calculated as the normalized categorical cross-entropy, defined, for a single data point, as:

$$H(y, \hat{y}) = -\sum_{i=1}^{C} y_i \log(\hat{y}_i)$$

where C is the number of classes, y is the true probability distribution, and $\hat{y}$ is the predicted probability distribution from the model.

Data availability
All the data used in this study can be found at https://doi.org/10.6092/unibo/amsacta/7308 preserved in the repository AMSacta hosted by the University of Bologna and licensed under Creative Commons 4.0.

Code availability
All the code developed for this study can be found at GitHub—GianlucaCarlini/CoreNet: Semantic segmentation of continuous cores with CNNs.

References
1. Martinson, D. G. *et al.* Age dating and the orbital theory of the ice ages: Development of a high-resolution 0 to 300,000-year chronostratigraphy. *Quat. Res.* **27**, 1–29 (1987).

2. Mayewski, P. A. *et al.* Holocene climate variability. *Quat. Res.* **62**, 243–255 (2004).
3. Mitchum, R. M. Jr., Vail, P. R. & Thompson, S. I. Seismic stratigraphy and global changes of sea level, part 2: The depositional sequence as a basic unit for stratigraphic analysis. In *Seismic Stratigraphy—Applications to Hydrocarbon Exploration* Vol. 26 (ed. Payton, C. E.) (American Association of Petroleum Geologists, 1977).
4. Posamentier, H. W., Jervey, M. T. & Vail, P. R. Eustatic controls on clastic deposition I—conceptual framework. In *Sea-Level Changes: An Integrated Approach* Vol. 42 (eds Wilgus, C. K. *et al.*) (SEPM Special Publication, 1988).
5. Neal, J. & Abreu, V. Sequence stratigraphy hierarchy and the accommodation succession method. *Geology* **37**, 779–782 (2009).
6. Van Asselen, S., Stouthamer, E. & Van Asch, Th. W. J. Effects of peat compaction on delta evolution: A review on processes, responses, measuring and modeling. *Earth Sci. Rev.* **92**, 35–51 (2009).
7. Teatini, P., Ferronato, M., Gambolati, G. & Gonella, M. Groundwater pumping and land subsidence in the Emilia-Romagna coastland, Italy: Modeling the past occurrence and the future trend. *Water Resour. Res.* **42**, 200. https://doi.org/10.1029/2005WR004242 (2006).
8. Prins, L. T. & Andresen, K. J. A geotechnical stratigraphy for the shallow subsurface in the Southern Central Graben, North Sea. *Eng. Geol.* **286**, 106089 (2021).
9. Campo, B., Bruno, L. & Amorosi, A. Sedimentary facies characterization through CPTU profiles: An effective tool for subsurface investigation of modern alluvial and coastal plains. *Sedimentology* **70**(4), 1302–1327 (2023).
10. Amorosi, A., Bruno, L., Campo, B., Di Martino, A. & Sammartino, I. Patterns of geochemical variability across weakly developed paleosol profiles and their role as regional stratigraphic markers (Upper Pleistocene, Po Plain). *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **574**, 110413 (2021).
11. Kruiver, P. P. *et al.* Geological schematisation of the shallow subsurface of Groningen—for site response to earthquakes for the Groningen gas field (2015).
12. Tanabe, S., Ishihara, Y., Nakanishi, T., Stafleu, J. & Busschers, F. S. Distribution of Holocene marine mud and its relation to damage from the 1923 earthquake disaster in the Tokyo Metropolitan Area, Japan. *Geosciences* **11**, 272 (2021).
13. Campo, B. *et al.* Late Pleistocene to Holocene glacio-eustatic history as recorded in the Pescara paleovalley system (Central Italy, Adriatic basin). *Mar. Pet. Geol.* **145**, 105908 (2022).
14. Amorosi, A. *et al.* Late Quaternary sedimentary record of estuarine incised-valley filling and interfluve flooding: The Manfredonia paleovalley system (southern Italy). *Mar. Pet. Geol.* **147**, 105975 (2023).
15. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. http://arxiv.org/abs/1810.04805 (2019).
16. Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You Only Look Once: Unified, real-time object detection. http://arxiv.org/abs/1506.02640 (2016).
17. Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. High-resolution image synthesis with latent diffusion models. http://arxiv.org/abs/2112.10752 (2022).
18. Bolandi, V., Kadkhodaie, A. & Farzi, R. Analyzing organic richness of source rocks from well log data by using SVM and ANN classifiers: A case study from the Kazhdumi formation, the Persian Gulf basin, offshore Iran. *J. Pet. Sci. Eng.* **151**, 224–234 (2017).
19. Xiong, Y., Zuo, R. & Carranza, E. J. M. Mapping mineral prospectivity through big data analytics and a deep learning algorithm. *Ore Geol. Rev.* **102**, 811–817 (2018).
20. Kuwatani, T. *et al.* Machine-learning techniques for geochemical discrimination of 2011 Tohoku tsunami deposits. *Sci. Rep.* **4**, 7077 (2014).
21. Hadler, H. *et al.* Automated facies identification by Direct Push-based sensing methods (CPT, HPT) and multivariate linear discriminant analysis to decipher geomorphological changes and storm surge impact on a medieval coastal landscape. *Earth Surf. Process. Landf.* **46**, 3228–3251 (2021).
22. Jacq, K. *et al.* Sedimentary structure discrimination with hyperspectral imaging in sediment cores. *Sci. Total Environ.* **817**, 152018 (2022).
23. Kadow, C., Hall, D. M. & Ulbrich, U. Artificial intelligence reconstructs missing climate information. *Nat. Geosci.* **13**, 408–413 (2020).
24. Crnkovic-Friis, L. & Erlandson, M. *Geology Driven EUR Prediction Using Deep Learning* (OnePetro, 2015).
25. Wrona, T., Pan, I., Gawthorpe, R. L. & Fossen, H. Seismic facies analysis using machine learning. *Geophysics* **83**, O83–O95 (2018).
26. Fleming, S. W., Watson, J. R., Ellenson, A., Cannon, A. J. & Vesselinov, V. C. Machine learning in Earth and environmental science requires education and research policy reforms. *Nat. Geosci.* **14**, 878–880 (2021).
27. Bergen, K. J., Johnson, P. A., De Hoop, M. V. & Beroza, G. C. Machine learning for data-driven discovery in solid Earth geoscience. *Science* **363**, eaau0323 (2019).
28. Goldstein, E. B., Coco, G. & Plant, N. G. A review of machine learning applications to coastal sediment transport and morphodynamics. *Earth Sci. Rev.* **194**, 97–108 (2019).
29. Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A. & Kumar, V. Machine learning for the geosciences: Challenges and opportunities. *IEEE Trans. Knowl. Data Eng.* **31**, 1544–1554 (2019).
30. Lee, A.-S., Enters, D., Huang, J.-J.S., Liou, S. Y. H. & Zolitschka, B. An automatic sediment-facies classification approach using machine learning and feature engineering. *Commun. Earth Environ.* **3**, 1–9 (2022).
31. Li, Z., Liu, F., Yang, W., Peng, S. & Zhou, J. A survey of convolutional neural networks: Analysis, applications, and prospects. *IEEE Trans. Neural Netw. Learn. Syst.* **33**, 6999–7019 (2022).
32. Guo, Y., Liu, Y., Georgiou, T. & Lew, M. S. A review of semantic segmentation using deep neural networks. *Int. J. Multimed. Inf. Retr.* **7**, 87–93 (2018).
33. Ribeiro, M. T., Singh, S. & Guestrin, C. 'Why Should I Trust You?': Explaining the predictions of any classifier. http://arxiv.org/abs/1602.04938 (2016).
34. Selvaraju, R. R. *et al.* Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.* **128**, 336–359 (2020).
35. Vinogradova, K., Dibrov, A. & Myers, G. towards interpretable semantic segmentation via gradient-weighted class activation mapping (Student Abstract). *AAAI* **34**, 13943–13944 (2020).
36. Susmaga, R. Confusion matrix visualization. In *Intelligent Information Processing and Web Mining* (eds Kłopotek, M. A. *et al.*) 107–116 (Springer, 2004).
37. Tharwat, A. Classification assessment methods. *Appl. Comput. Inform.* **17**, 168–192 (2020).
38. Shaikh, S. A. Measures derived from a 2 x 2 table for an accuracy of a diagnostic test. *J. Biomet. Biostat.* **2**, 5 (2011).
39. Sokolova, M. & Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **45**, 427–437 (2009).
40. Powers, D. M. W. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. https://doi.org/10.48550/arXiv.2010.16061 (2020).
41. Gu, J. *et al.* Recent advances in convolutional neural networks. *Pattern Recogn.* **77**, 354–377 (2018).
42. Amorosi, A. & Marchi, N. High-resolution sequence stratigraphy from piezocone tests: An example from the Late Quaternary deposits of the southeastern Po Plain. *Sediment. Geol.* **128**, 67–81 (1999).
43. Scarponi, D., Kaufman, D., Amorosi, A. & Kowalewski, M. Sequence stratigraphy and the resolution of the fossil record. *Geology* **41**, 239–242 (2013).

44. Amorosi, A., Rossi, V., Scarponi, D., Vaiani, S. C. & Ghosh, A. Biosedimentary record of postglacial coastal dynamics: High-resolution sequence stratigraphy from the northern Tuscan coast (Italy). *Boreas* **43**, 939–954 (2014).
45. Wood, D. A. Enhancing lithofacies machine learning predictions with gamma-ray attributes for boreholes with limited diversity of recorded well logs. *Artif. Intell. Geosci.* **2**, 148–164 (2021).
46. Rapuc, W. *et al.* XRF and hyperspectral analyses as an automatic way to detect flood events in sediment cores. *Sediment. Geol.* **409**, 105776 (2020).
47. Amorosi, A. *et al.* Global sea-level control on local parasequence architecture from the Holocene record of the Po Plain, Italy. *Mar. Pet. Geol.* **87**, 99–111 (2017).
48. Di Martino, A. & Amorosi, A. Sedimentary facies analysis and segmentation (2023).
49. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015* (eds Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) 234–241 (Springer International Publishing, 2015). https://doi.org/10.1007/978-3-319-24574-4_28.
50. Tan, M. & Le, Q. V. EfficientNet: Rethinking model scaling for convolutional neural networks. https://doi.org/10.48550/arXiv.1905.11946 (2020).
51. Russakovsky, O. *et al.* ImageNet large scale visual recognition challenge. *Int. J. Comput. Vis.* **115**, 211–252 (2015).
52. Baheti, B., Innani, S., Gajre, S. & Talbar, S. Eff-unet: A novel architecture for semantic segmentation in unstructured environment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* 358–359 (2020).
53. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. https://doi.org/10.48550/arXiv.1512.03385 (2015).
54. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. https://doi.org/10.48550/arXiv.1412.6980 (2017).
55. Wang, Z., Wang, E. & Zhu, Y. Image segmentation evaluation: a survey of methods. *Artif. Intell. Rev.* **53**, 5637–5674 (2020).
56. Goutte, C. & Gaussier, E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *Advances in Information Retrieval* (eds Losada, D. E. & Fernández-Luna, J. M.) 345–359 (Springer, 2005).
57. Hay, A. M. The derivation of global estimates from a confusion matrix. *Int. J. Remote Sens.* **9**, 1395–1398 (1988).
58. Abadi, M. *et al.* Tensorflow: Large-scale machine learning on heterogeneous systems (2015).
59. Pedregosa, F. *et al.* Scikit-learn: Machine learning in python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

Acknowledgements

Author contributions

A.D.M. and G. Carlini equally contributed to the conceptualization of the work, data curation, design of the methodology, and original writing. A.A., G. Castellani, and D.R. jointly supervised the work contributing to funding acquisition, project administration, review, and editing of the original draft. Throughout all stages, the authors engaged in discussions regarding results and implications.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to D.R. or A.A.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.