



Automated identification of fossil benthic foraminifera from the Peruvian margin using convolutional neural networks

Sikandar Hayat¹, Meryem Mojtahid¹, Mary Elliot¹, Jorge Cardich², Emmanuelle Geslin¹,
Thibault de Garidel-Thoron³, Matthieu Carré⁴, and Christine Barras¹

¹Univ Angers, Nantes Université, Le Mans Université, CNRS, Laboratoire de Planétologie et Géosciences,
LPG UMR 6112 – Université d'Angers, 49000 Angers, France

²CIDIS-Facultad de Ciencias e Ingeniería, Universidad Peruana Cayetano Heredia, Lima, Peru

³Aix-Marseille Université, CNRS, IRD, INRAE, CEREGE, Aix-en-Provence 13545, France

⁴IPSL/Laboratoire d'Océanographie et du Climat: Expérimentations et Approches Numériques –
CNRS-IRD-MNHN-Sorbonne Universités, Paris, France

Correspondence: Sikandar Hayat (sikandar.hayat@univ.angers.fr)

Received: 5 October 2025 – Revised: 5 January 2026 – Accepted: 6 January 2026 – Published: 26 January 2026

Abstract. Benthic foraminifera tests preserved in marine sediments are well-established proxies for bottom-water dynamics, yet their minute size and high diversity demand laborious manual identification of hundreds of individuals to reconstruct subtle faunal shifts and are prone to observer-dependent taxonomic inconsistencies. The recent advances in image acquisition hardware and image identification software have made it possible to acquire and identify large image datasets quickly. Here, we trained convolutional neural networks (CNNs) to identify benthic foraminifera morphospecies from 31 samples from two sedimentary cores from offshore Peru, spanning the past 18 000 years. Our best-performing model achieves 92 % overall classification accuracy, 93.4 % precision, and 92.4 % recall, enabling high-temporal-resolution reconstructions of benthic foraminifera assemblages along the Peruvian margin. Automated outputs closely matched manual results across 31 samples, from counts and relative abundances to diversity indices, multivariate assemblage patterns, and dissolved oxygen estimates, indicating the suitability of automated identification for paleo-ecological applications. The highest-performing CNN model (trained on a dataset of 5860 images) from this study can be adapted to analyse benthic foraminifera from equivalent depths of the Peruvian margin, providing high-resolution insights into the eastern tropical Pacific oxygen minimum zone (OMZ). In addition to offering a scalable, objective alternative for high-temporal-resolution analysis of benthic foraminifera, this study also highlights the current limitations of automated workflows.

1 Introduction

Foraminifera, unicellular protists abundant from the coast to the deep ocean, are key microfossils utilized to study past ocean conditions. Fossil evidence places the first appearance of foraminifera in the early Cambrian, yet molecular data point to a Neoproterozoic origin, hidden by their absent or fragile early tests (Pawlowski et al., 2003). Their small size, high sensitivity to the ambient environment, rapid evolution, and high diversity make them an excellent biostrati-

graphic and paleoceanographic tool (e.g. Sen Gupta, 1999; Schiebel and Hemleben, 2017). Foraminiferal assemblages are typically analysed using a binocular microscope, with each sample often requiring the picking (isolating the specimens), the counting, and the identification of over 300 individuals to ensure statistically robust and reliable results (Patterson and Fishbein, 1989; Schönfeld et al., 2012). This process is time-consuming, requiring months to years of training, and is prone to inter-observer variability (e.g. Al-Sabouni et al., 2018; Mitra et al., 2019). The complexity is

further increased by the fact that there are around 50 000 species of foraminifera, with about 9000 still existing today (Hayward et al., 2025; Murray, 2008), of which about 50 are planktonic (Brummer and Kučera, 2022). Mitra et al. (2019) found that taxonomists' ability to identify species accurately depends largely on their prior experience with those specific species. Yet, greater overall experience in taxonomy does not always lead to improved precision in species identification (Austen et al., 2016; Al-Sabouni et al., 2018). Furthermore, this identification process becomes particularly challenging for high-temporal-resolution studies of sediment cores, where a greater number of samples is needed to capture finer timescales. Achieving precise and consistent data across numerous samples complicates the analysis and increases the time required for such studies.

Recent advances in automated identification techniques, including convolutional neural networks (CNNs: the artificial neural networks trained on image datasets to find patterns and utilize this information to classify a new dataset; Marret, 2023) and deep-learning algorithms, have revolutionized microfossil studies by improving the speed and accuracy of species identification (e.g. Dollfus and Beaufort, 1999; Beaufort and Dollfus, 2004; Mitra et al., 2019). Moreover, the development of automated imaging systems has enabled the rapid acquisition of thousands of images with minimal human intervention. Such methods have been successfully applied to a variety of fossil groups, including coccoliths (Beaufort et al., 2014), radiolarians (Tetard et al., 2020), planktonic foraminifera (Mitra et al., 2019; Hsiang et al., 2019; Johansen et al., 2021; Marchant et al., 2020; Carvalho et al., 2020), benthic foraminifera (Johansen et al., 2021; Marchant et al., 2020; Carvalho et al., 2020; Plavetić et al., 2025), fusulinids (Pires De Lima et al., 2020), graptolites (Niu and Xu, 2022), pollen grains (Gimenez et al., 2024), and diatoms (Godbillot et al., 2024). There has been a growing interest in the automated identification of foraminifera in recent years, driven by their small size, easy extraction, morphological diversity, abundance, and crucial role in paleontology and stratigraphy (Hsiang et al., 2019; Mitra et al., 2019; Pires De Lima et al., 2020; Carvalho et al., 2020; Marchant et al., 2020; Plavetić et al., 2025). User-friendly, open-source software such as ParticleTrieur (Marchant et al., 2020; <https://github.com/microfossil/particle-trieur>, last access: 5 August 2025), which requires neither high computing power nor specialized training, has opened new avenues for paleontologists to incorporate CNNs into their workflow. By merging machine learning techniques with micropaleontology, vast datasets can be processed quickly and at low cost. However, despite their transformative potential, CNNs still rely on extensive training with large datasets of labelled images. So far, regarding foraminifera, their successful application has been more common for planktonic forms as they comprise fewer species and show fewer variations in test shapes, chamber arrangements, wall compositions, and aperture morphologies compared to benthic foraminifera (Schiebel et al.,

2018; Gooday et al., 2008; Debenay, 2012; Murray, 2008). For example, Mitra et al. (2019) used CNNs to identify 1437 specimens of planktonic foraminifera ($> 250 \mu\text{m}$, six species), achieving precision (defined as the percentage of images labelled as class X that are truly X) comparable to and recall (defined as the percentage of actual class- X images that the model identified correctly) higher than that of taxonomists. A substantially larger dataset of modern planktonic foraminifera, mostly from core top sediments, comprising 27 737 training and 6903 test images (35 species) of high quality, was used by Hsiang et al. (2019), achieving a validation accuracy of 87 %. However, the study only retained specimens with high on-screen agreement among experts, e.g. at least 75 % agreement among four taxonomists for 24 569 specimens; the dataset overrepresents canonical specimens that are easier to identify from 2D photographs. The higher morphological diversity of benthic foraminifera implies that CNN-based approaches developed for planktonic foraminifera cannot be directly transferred to benthic assemblages without methodological adaptation.

Research on species-level automated identification of benthic foraminifera using CNNs remains limited; to the best of our knowledge, there are four peer-reviewed studies. Their morphological complexity and interspecies similarities make identification far more challenging, requiring significantly larger and more diverse image datasets. Marchant et al. (2020) successfully applied ResNet50 to identify a set of fossil benthic foraminifera species (12 species) from a north-eastern Pacific core and achieved comparable relative abundance trends among manual and automated identification. Kahanamoku-Meyer et al. (2024) compiled 10 827 benthic foraminifera images from the Santa Barbara Basin spanning the past 800 years, covering 77 species, though only 28 had more than 10 specimens. The data were split according to a ratio of 80 / 10 / 10 into training, validation, and test sets. A ResNet50 transfer-learning classifier reached 80.6 % species-level and 85.6 % genus-level validation accuracy compared to the on-screen image identification by six undergraduate students, but rare species showed very high misclassification. Yayan et al. (2024) reported the best accuracy of 90 % (ResNet50) for benthic foraminifera identification (nine species), yet the cited image source is Endless Forams, which is a repository of modern planktonic foraminifera images (Hsiang et al., 2019). Moreover, train and test image sets were generated by randomly splitting images from the same folder, a procedure that risks data leakage and can inflate performance estimates. Most recently, Plavetić et al. (2025) applied the YOLO (You Only Look Once) object detection model to images of picked benthic foraminifera and raw sediment from Skagerrak fjords, distinguishing 29 species with a mean average precision of 78.8 %–90.3 % across models.

In this study, we trained a CNN to identify benthic foraminiferal morphospecies from the oxygen minimum zone (OMZ) off the coast of Peru. This region is crucial for studying past climatic and oceanographic variations as the

highly productive Peruvian upwelling system preserves these variations in laminated sediments, enabling high-temporal-resolution paleo-oxygenation and paleo-productivity reconstructions (Salvatteci et al., 2016). To do so, we built upon the methodology of Marchant et al. (2020). However, since each ocean basin has distinct species compositions, sediment characteristics, and intraspecific variations, a separate CNN model was necessary for this region. We trained three CNN models using ~ 100 , ~ 200 , and ~ 400 specimens per species, with additional refinement by adding more images of the most abundant and morphologically variable species. We then compared manual (under binocular) and automated identification across 31 samples (from ~ 18 kyr BP to the late Holocene), evaluating how classification accuracy varies. By analysing the impact of dataset size and species-specific adjustments on classification performance, this study provides insights into the strengths, limitations, and potential improvements in automated foraminiferal identification while providing an open-source reference dataset for future automated benthic foraminifera identification in the region. Downcore, time-resolved reconstructions are beyond the scope of this methodological contribution and will be presented in a separate publication.

2 Materials and methods

2.1 Sampling site

Foraminifera samples were collected from two gravity marine sediment cores retrieved off the coast of Peru during the Galathea-3 expedition (Salvatteci et al., 2014, 2019) (Fig. 1a, b). Cores G10 (5.22 m length; 14.23° S, 76.4° W; 312 m water depth) and G14 (5.25 m length; 14.38° S, 76.42° W; 390 m water depth) were taken from the current core of the OMZ (200–400 m water depth), which extends from 50 to 500 m (Fuenzalida et al., 2009; Salvatteci et al., 2016). Based on the age models of Salvatteci et al. (2016), core G14 covers the time interval from 25.2 to 13.4 kyr BP, while core G10 covers 10.3 to 0.4 kyr BP.

2.2 Sample preparation

Sediment samples were wet sieved using a $125\ \mu\text{m}$ sieve, and the $> 125\ \mu\text{m}$ fraction was retained for analysis (Fig. 1c). The $> 125\ \mu\text{m}$ fraction was used for all samples to ensure methodological consistency as, in some cases, only this fraction remained available after prior analyses. Previous studies from the Peruvian margin used a range of size fractions, including $> 63\ \mu\text{m}$ (e.g. Oberhänsli et al., 1990; Mallon, 2012; Erdem and Schönfeld, 2017), $> 125\ \mu\text{m}$ (Heinze and Wefer, 1992; Oberhänsli et al., 1990), and $> 150\ \mu\text{m}$ (Malmgren and Funnell, 1990). Although a direct sieve size comparison is not available for Peru, this choice aligns with OMZ studies elsewhere that reported broadly comparable community structure across sieve sizes (e.g. the Arabian Sea (63–

125 vs. $> 125\ \mu\text{m}$; Caulle et al., 2014), the Gulf of Alaska (> 63 , 63–125, $> 125\ \mu\text{m}$; Sharon et al., 2021), the California margin (> 150 vs. $> 63\ \mu\text{m}$; Palmer et al., 2020)). Smaller size classes increase misidentification rates (Fenton et al., 2018) and disproportionately include propagules and juveniles lacking diagnostic features, thereby obscuring species-level patterns (Hermelin, 1986; Lo Giudice Cappelli and Austin, 2019).

The foraminiferal residues were then dried in an oven at 40°C . Samples were split to achieve a count of at least 300 specimens unless the total number of foraminifera was less than 300 in the sample (species counts are available at <https://www.seanoe.org/data/00976/108821>, last access: 3 October 2025). Taxonomic identification of 31 samples (16 samples from core G10 and 15 from core G14) was performed under a binocular microscope (Leica MZ16) before the image acquisition. In 16 of these samples, foraminifera were manually picked and placed onto Plummer cells with a black background for taxonomic determination. In the remaining 15 samples, foraminifera were directly identified and counted under the microscope. The taxonomic identification of benthic foraminifera (Fig. 2) was based on standard references for the eastern Pacific Ocean (Erdem and Schönfeld, 2017; Mallon, 2012; Resig, 1981; Smith, 1963; Cushman and McCulloch, 1942).

2.3 Image acquisition

For image acquisition, foraminifera were photographed either directly on the Plummer cells (for 16 samples) or within the $> 125\ \mu\text{m}$ residue spread on a plate (for 15 samples). The latter approach eliminates the need for the time-consuming task of manual picking. Various background plates (ranging from grey to black) were tested, and we ultimately chose a 3D printed black plate with the least reflective surface, along with black background Plummer cells. The photography setup consists of a 3D printer with a camera attached (Fig. 1d). We used a $4\times$ telecentric lens with a resolution of $3.45\ \mu\text{m}^2$ per pixel. The plate (or Plummer cell) was illuminated by a ring light at a 30° angle. Due to the camera's limited depth of field ($90\ \mu\text{m}$), images were captured at multiple depths and were later stacked using the image stacking software HeliconSoft (Helicon Soft Ltd.). A slight overlap ($300\ \mu\text{m}$) was maintained between images to minimize the loss of foraminifera at the edges. The images were subsequently stitched together with a custom image processing algorithm using Python to create seamless, high-resolution composite images (code will be uploaded to GitHub). After stitching, we re-cut the composite image into its original tile sizes, each tile with its coordinates, which were used in the next steps for duplicate removal.

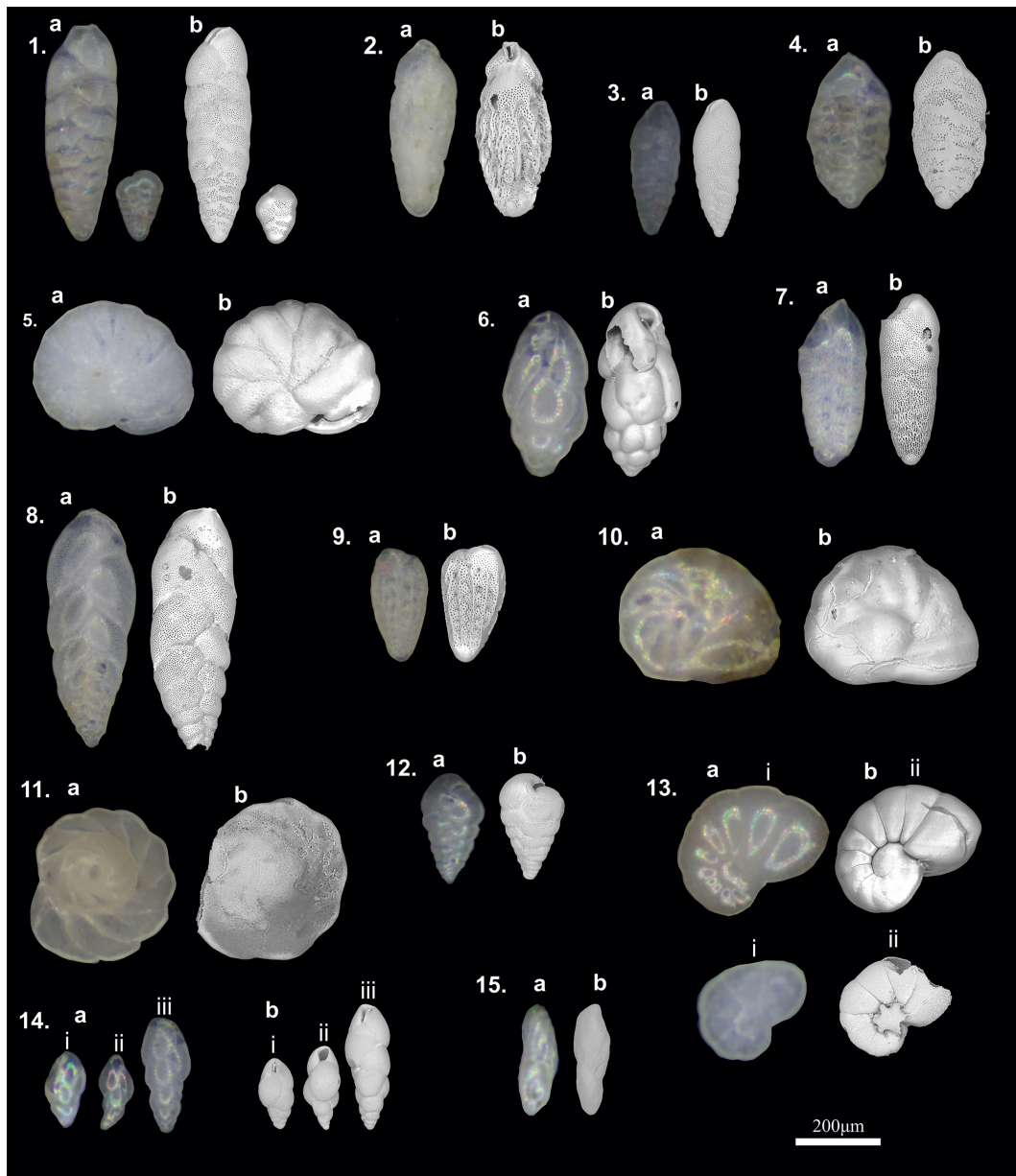


Figure 2. Foraminifera images obtained from automation setup (a in each pair) and scanning electron microscope (b in each pair). Each pair represents one class for the CNN models. The images in each pair are not of the same foraminifera specimen. (1) *Bolivina humilis*, (2) *Bolivina plicata*, (3) *Pseudoparrella pacifica*, (4) *Bolivina advena*, (5) *Cassidulina limbata*, (6) *Ebuliminella curta*, (7) *Bolivina seminuda*, (8) *Loxotomum pseudobeyrichii*, (9) *Bolivina costata*, (10) *Epistominella obesa*, (11) *Epistominella pacifica*, (12) *Suggrunda porosa*, (13) *Nonionella* spp. (i) *Nonionella auris*, (ii) *Nonionella stella*, (14) *Fursenkoina* spp. (i) *Fursenkoina fusiformis*, (ii) *Fursenkoina* sp., (iii) *Fursenkoina texturata*, (15) *Buliminella elegantissima*.

cated to training and 20 % to validation. Three CNN models (M-1, M-2, and M-3) were trained on all the major species (*Bolivina humilis*, *B. costata*, *B. seminuda*, *B. plicata*, *B. advena*, *Ebuliminella curta*, *Fursenkoina* spp., *Epistominella obesa*, *Pseudoparrella pacifica*, *Suggrunda porosa*, and *Cassidulina limbata*) and less abundant species (*Buliminella elegantissima*, *Bolivina pacifica*, *Loxotomum pseudobeyrichii*, and *Nonionella* spp.) present in our samples from across the

core using increasing dataset sizes. Here, we define major species as those comprising $\geq 1\%$ of individuals in our manual counts (across all samples combined) and less abundant species as those comprising $< 1\%$. For each major species, the number of specimens is ~ 100 in M-1, ~ 200 in M-2, and ~ 400 in M-3. The less abundant species are added to each model in low numbers (see Table 1) based on the available number of photos and to keep the training data repre-

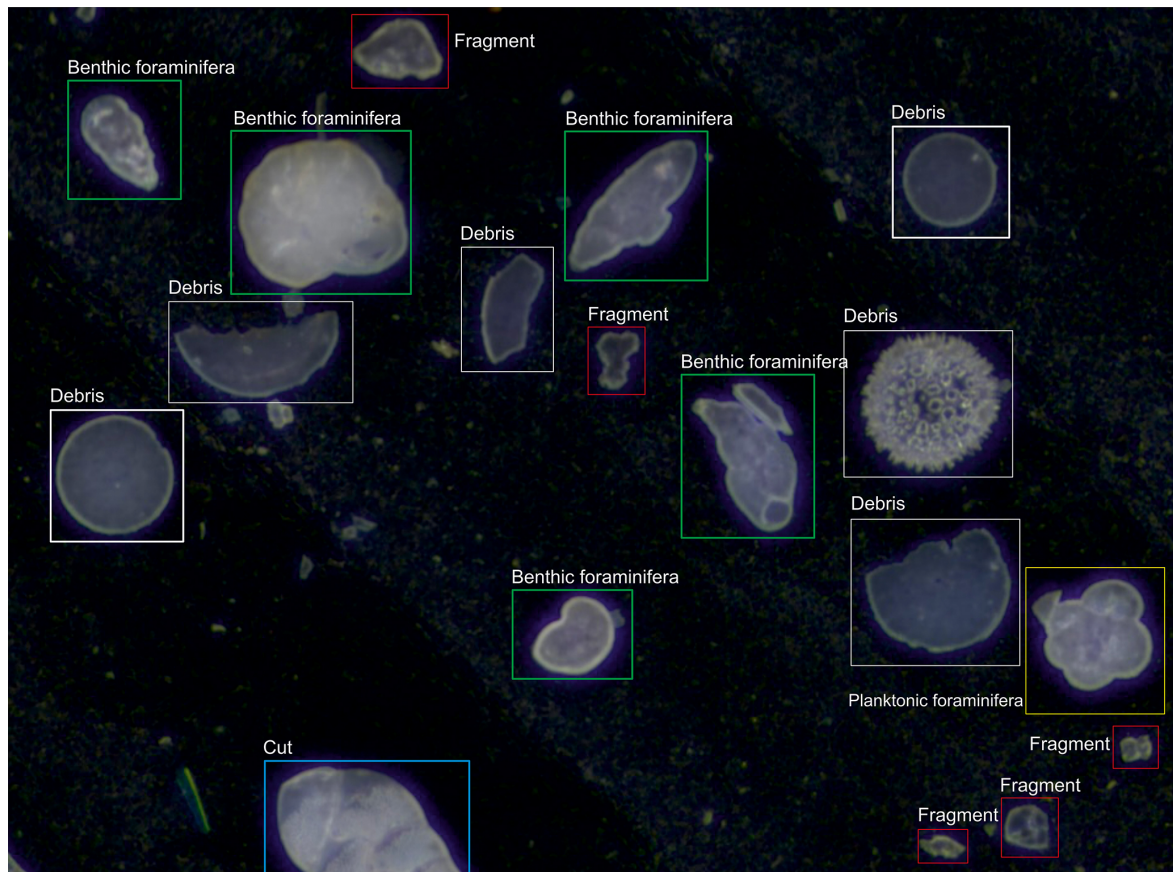


Figure 3. Segmentation of different classes in CVAT. The five classes, including benthic foraminifera, debris, cut, fragment, and planktonic foraminifera, are bounded by rectangles of different colours.

sentative of actual samples. However, the species with fewer than 10 images were not included as this would force the model to learn a decision boundary from very few examples, which tends to produce unstable predictions and can degrade the performance of other classes. The number of specimens of each species in different models is listed in Table 1. Species with a very close morphological resemblance, e.g. *Fursenkoina fusiformis*, *F. texturata*, and *Fursenkoina* sp., were assigned to a single class, i.e. *Fursenkoina* spp. After training, each model generated a confusion matrix (i.e. a matrix that compares the model's predicted labels to the true labels) (Fig. 4) and identified mislabelled species (Fig. 5), along with providing performance metrics. These outputs helped refine model accuracy and recall by correctly labelling the misidentified species and adding more images for species with low classification accuracy and recall. We reran M-3 five times to see the fluctuation in performance. Finally, to further improve performance metrics, a fourth model (M-4) was developed by incorporating additional specimens of the most abundant species with high morphological variability (*Bolivina humilis*, *Fursenkoina* spp., and *Suggrunda porosa*). The images used to train the model were different

from the ones the model was applied on to avoid circular-reasoning bias.

2.6 Model application

Before applying the model for species identification, duplicate images were removed in two steps. First, we applied a custom algorithm (which will be released on GitHub) to the foraminifera crops to flag duplicates based on the overlap in coordinates between image pairs. Next, the images were uploaded to ParticleTrieur, where remaining duplicates were removed using a 97 % similarity threshold based on the feature vector (kNN (*k* nearest neighbour)-based clustering technique). Finally, the best-performing pretrained CNN model (M-4) was applied to determine species counts. The automated counts were compared to manual taxonomic counts conducted by a micropaleontologist following the standard counting procedures. The CNN identification threshold for model application was set at 50 %, e.g. if a specimen had 50 % similarity to *B. humilis*, 30 % similarity to *B. seminuda*, and 20 % similarity to *B. pacifica*, it would be identified as *B. humilis*. Raising the threshold to 70 %–80 % increased the proportion of specimens relegated to an “unsure” category

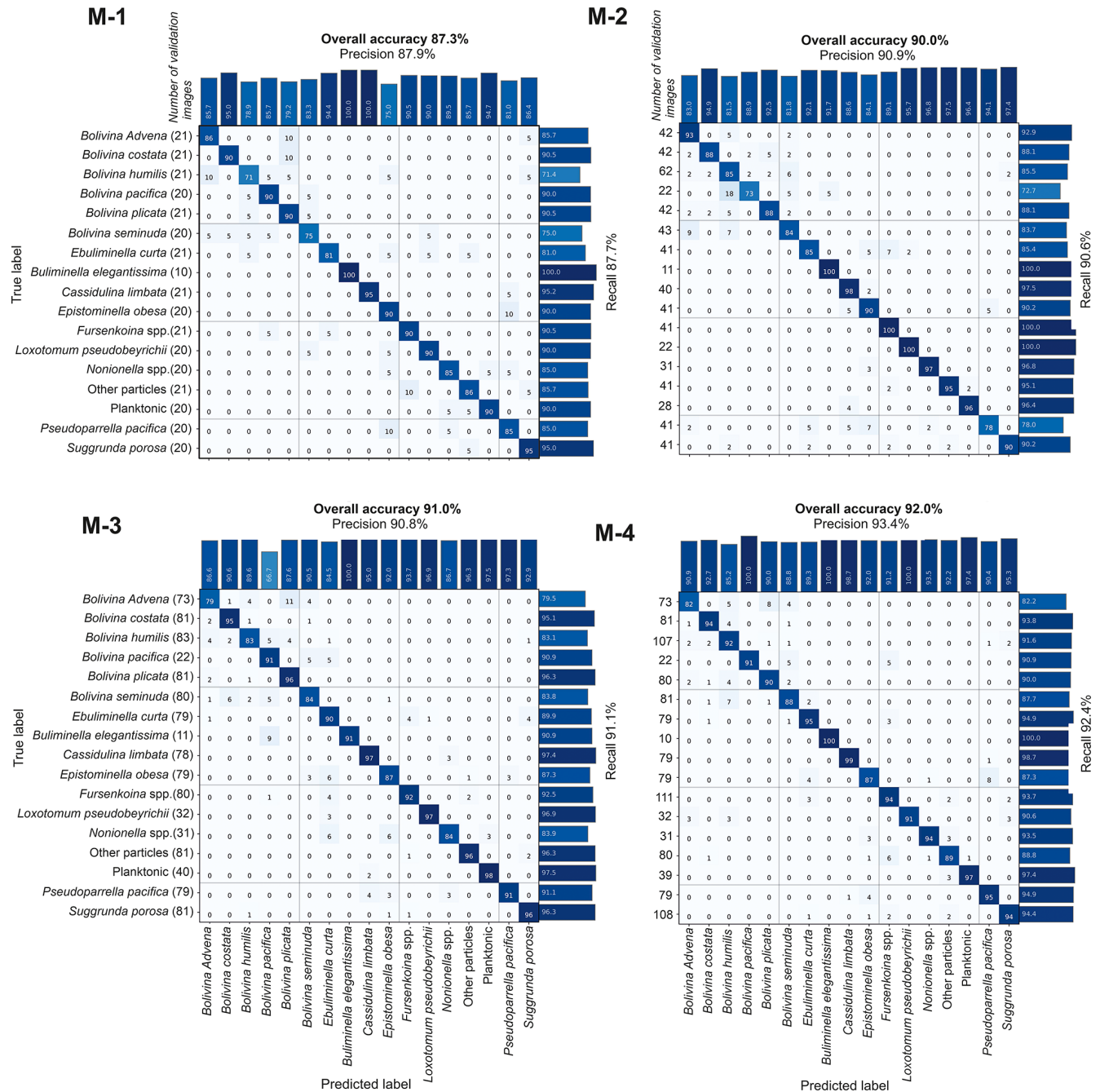


Figure 4. Confusion matrix of four CNN models (M-1 to M-4). The dataset contained 1690 images for M-1, 3155 for M-2, 5455 for M-3, and 5860 for M-4, with 80 % used for training and 20 % for validation. Species names and the number of individuals in the validation set are listed on the left. Each cell indicates the percentage of images from the true class (rows) that were predicted as the corresponding class (columns). Perfect classification is indicated by 100 % values along the diagonal, while off-diagonal values reflect misclassifications, indicating inter-class confusion.

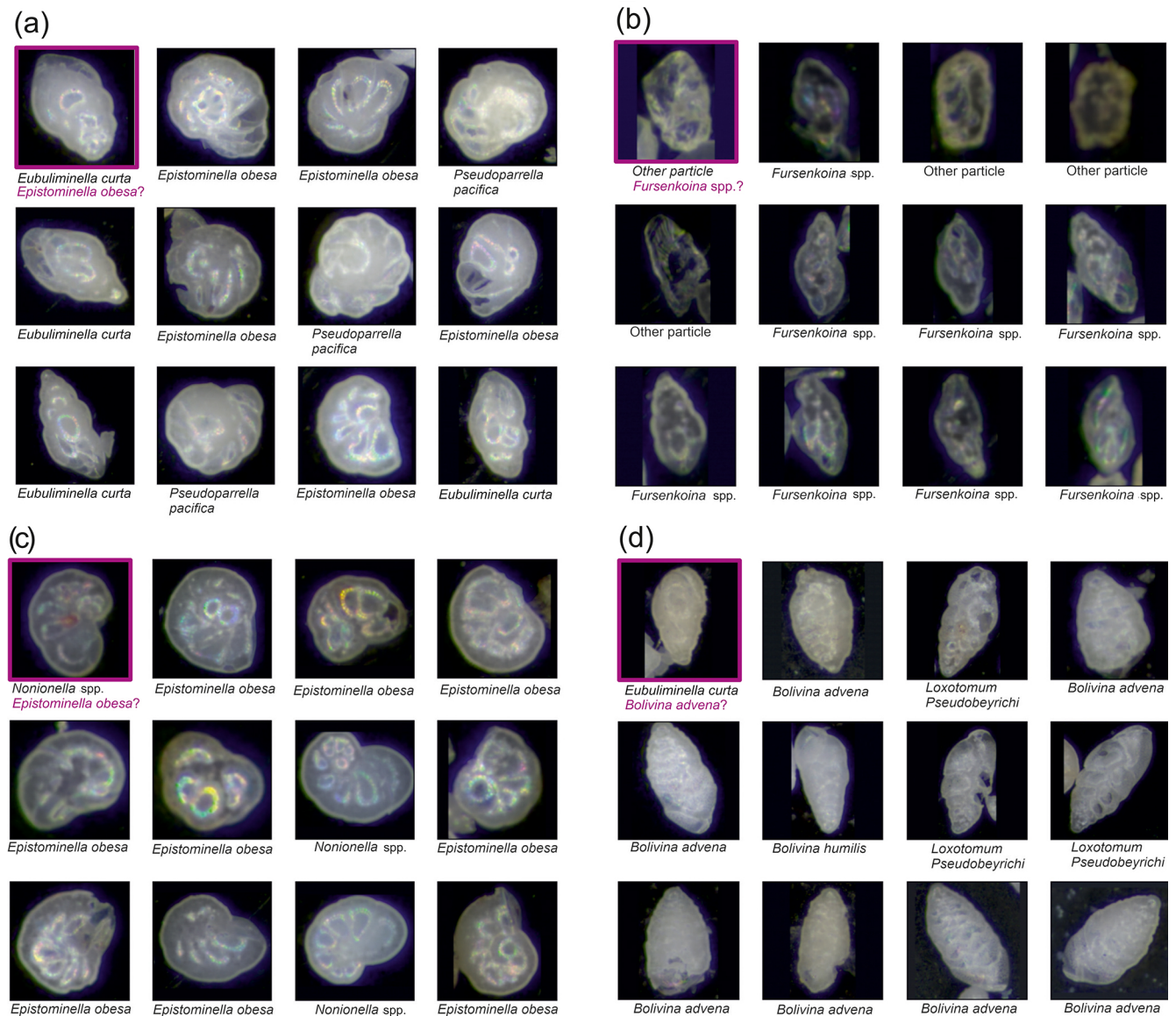


Figure 5. Example of the images that create confusion for the model. The image in the top-left corner (purple border) of each composite image is mislabelled according to the model, and it suggests a different name (in purple) based on the surrounding training images. **(a)** *Eubuliminella curta* (top-left corner) is suggested to be *Epistominella obesa* by the model; it also shows other *Epistominella obesa* specimens for comparison. **(b)** The model suggests that the “other particle” is *Fursenkoina* spp. **(c)** *Nonionella* spp. is confused with *Epistominella obesa*. **(d)** *Eubuliminella curta* is confused with *Bolivina advena*.

(e.g. a best match of 70 % would remain “unsure” under an 80 % threshold setting). A total of 15 % of specimens were labelled as unsure when the threshold was 80 % while only 3 % were labelled as such when the threshold was 50 %.

3 Results and discussion

3.1 CNN training

Our analyses showed that training-data size affects model performance. Increasing the per-species training images

from ~ 100 to ~ 400 raised the overall accuracy from 87.3 % to 91.0 % and the recall from 87.7 % to 91.1 % (Fig. 4, M-1 and M-3). We ran M-3 five times to assess run-to-run variability. Accuracy ranged from 91 % to 92.0 %, precision ranged from 90.8 % to 92.0 %, and recall ranged from 91.4 % to 91.8 %. We then expanded the training set by adding additional specimens of the most abundant and morphologically variable species in our samples, *Bolivina humilis* (125), *Fursenkoina* spp. (155), and *Suggrunda porosa* (135). This enhanced the performance in terms of precision (93.4 %), recall (92 %), and accuracy (92 %) (Fig. 4, M-4), consistently

with the trend reported by Zhong et al. (2017) for planktonic foraminifera. The results highlight the need for a training set that accurately reflects the distribution and morphological variation of foraminifera in the samples to prevent over- or under-representation of species.

The recall and precision metrics of our best-performing model (M-4) (92 % and 93.4 %, respectively) surpass those of earlier studies employing machine learning and image processing techniques for foraminiferal classification. For instance, Hsiang et al. (2019) reported an accuracy of 89 % for planktonic foraminifera, while Marchant et al. (2020) achieved 89 % accuracy and 80.7 % recall for benthic species and 90.7 % accuracy with 77.6 % recall for planktonic species. Also, the previous works reported recall rates below 80 % for several species (e.g. Marchant et al., 2020; Hsiang et al., 2019), whereas our model M-4 achieved recall exceeding 90 % for all but three species (*B. advena*, *B. seminuda*, and *E. obesa*).

Additionally, we tested the impact of different orientations of each species on classification accuracy. While we attempted to include images of species in different orientations, we noticed that most species tend to land on the plate in a distinctive position, and artificially (with a brush) altering their orientation led to misclassification. For example, *Bolivina humilis* constantly lies flat with its aperture on the side, with only a few specimens positioned vertically. Moreover, species that require 3D views for accurate identification and that represent similar ecological conditions, such as *Fursenkoina texturata*, *Fursenkoina* sp., and *F. fusiformis*, were lumped together as *Fursenkoina* spp. (high photodetritus input; low oxygen; Das et al., 2017) due to the difficulty in distinguishing them in 2D images and co-occurrence in samples. Interestingly, while humans often misidentify species within the same genus due to morphological similarities, machine learning models tend to make misclassifications that are not always phylogenetically conservative (Hsiang et al., 2019; Pires De Lima et al., 2020). For example, the model confused different species like *Epistominella obesa* and *Ebuliminella curta* (Fig. 5a), as well as *Nonionella* spp. and *Epistominella obesa* (Fig. 5c), due to similarities in light reflection patterns and outline. This suggests that CNNs focus primarily on visual features without incorporating the contextual or phylogenetic knowledge that human experts use, indicating a fundamental difference in classification behaviour between human experts and machine learning models. Some species, especially *Fursenkoina* spp., were confused with the “other particles” class due to their fragile and translucent tests that often show partial breakage (Fig. 5b). Additionally, some specimens that were altered or had surface deposits were misclassified, which can be attributed to the lack of their representation in the training data (Fig. 5d), highlighting the challenges of capturing fine morphological details in automated identification systems.

3.2 Model application

In this study, we compared automated and manual counting of benthic foraminifera from 31 samples to evaluate model performance. Automated and manual counts show a strong correlation for most species and samples (Figs. 6 and 7). Notably, our model successfully differentiated between *Bolivina* species despite their significant morphological similarities (Marchant et al., 2020). However, despite the strong positive correlations between AI and manual counts (Fig. 7), some differences between the two methods can be noticed. For each species, we tested whether the distribution of paired differences between automated and manual counts across samples was normal using the Shapiro–Wilk test; the assumption was not met ($p < 0.05$). We therefore applied the Wilcoxon signed-rank test (paired), which indicated significant differences for all species except *Epistominella obesa* and *Fursenkoina* spp. The direction of the effect, with automated counts that were generally higher, is consistent with slight overcounting by the automated workflow (Fig. 6c), likely due to misidentification of fragments as foraminifera. However, *Bolivina humilis*, the most dominant species in our samples, exhibits the highest discrepancy between automated and manual counts, with the human count being significantly higher (Fig. 6c). This is probably because humans tend to assign a species identity that is most common in samples to ambiguous individuals (Hsiang et al., 2019), a bias known as “pull towards the common” that does not influence CNNs (Hsiang and Hull, 2022). Historically, *Bolivina humilis* was considered to be a variety of *Bolivina seminuda* (*Bolivina seminuda* var. *humilis*) by Cushman and McCulloch (1942) before being recognized as a separate species. Although these species exhibit morphological differences, such as a sub-cylindrical cross-section in *B. seminuda* vs. an oval cross-section for *B. humilis*, they can be difficult to differentiate in a 2D view (García-Gallardo et al., 2021). Overall, our model performed well in identifying both species (*B. seminuda* and *B. humilis*), with a high correlation between automated and manual counts.

Most misclassifications occurred in samples where foraminifera were crowded (touching each other) or where abundant diatoms were attached to the foraminifera tests. To minimize such errors, proper spreading of sediment residues on the imaging plate is necessary, and a quick preliminary check under a binocular microscope to ensure adequate spacing between particles is recommended. Some specimens of *Bolivina advena* and *Bolivina plicata* were confused with each other, probably due to their similar surface texture and overall morphology, especially in juveniles. A similar issue occurred with *Loxotomum pseudobeyrichii* as its early chambers resemble those of *Bolivina humilis*, leading to misidentification as *B. humilis*. Additionally, *Bolivina pacifica* consistently had higher automated identification counts (Fig. 6), likely due to its close morphological similarity with *Bolivina seminuda* (Tetard et al., 2024).

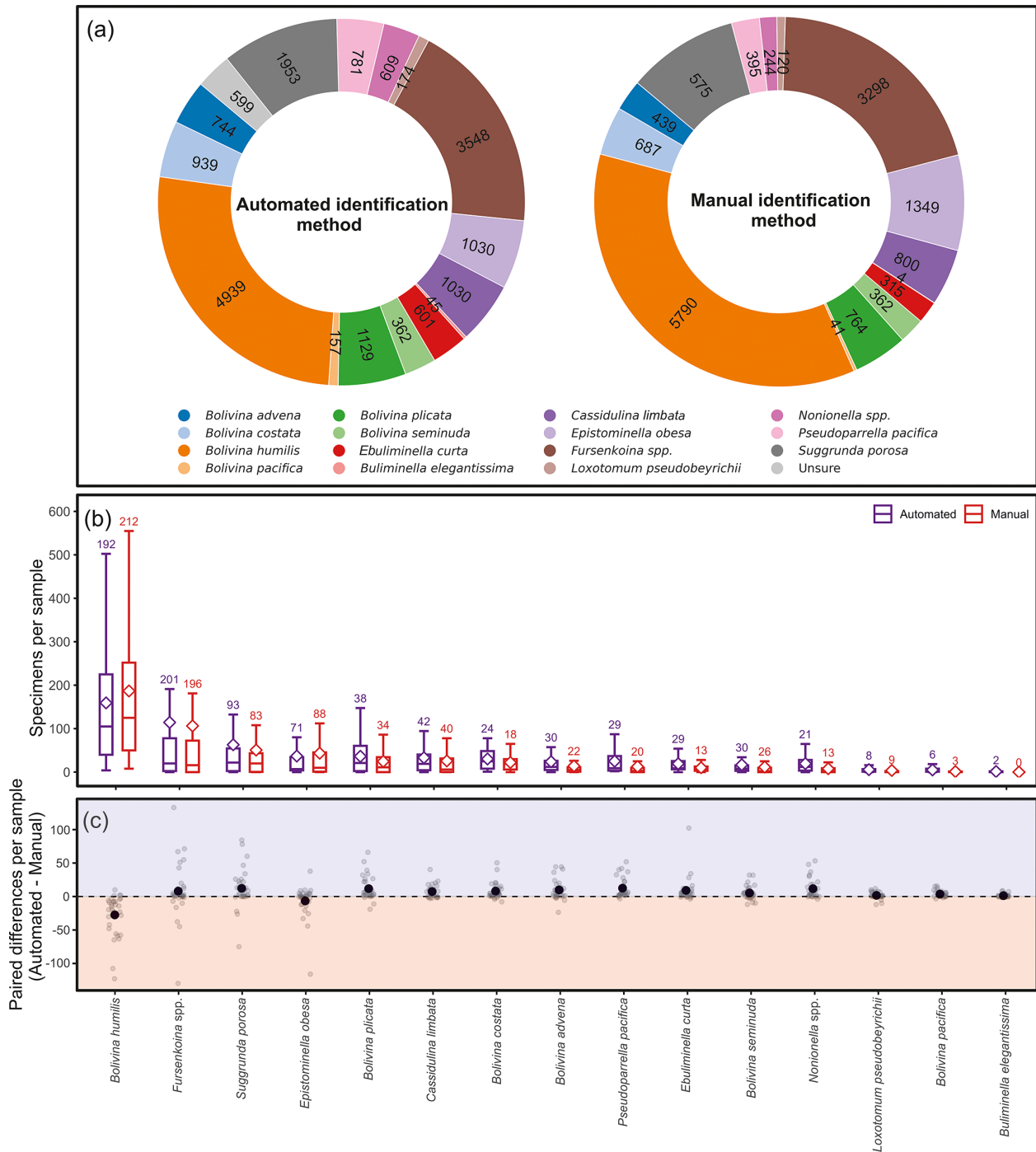


Figure 6. (a) Pie charts summarizing the relative proportions of species based on manual counts (left) and automated counts (right) from all 31 samples considered together. The numbers on the pie charts show the number of specimens of each species. (b) Box-and-whisker plots show the median and interquartile range of per-sample species' counts, comparing the automated vs. manual identification. Diamonds indicate means. (c) Paired per-sample differences in counts (automated – manual) for each species. Grey points represent individual samples, while black points indicate mean differences.

Table 1. Comparison of four classification models based on performance metrics and training-data distribution (80 % images were used for training and 20 % for validation).

Model		M-1	M-2	M-3	M-4
Recall		87.70 %	90.60 %	91.70 %	92 %
Precision		87.80 %	90.90 %	90.80 %	93.40 %
Overall accuracy		87.30 %	90 %	91 %	92 %
F1 score		87.8	90.74	91.74	92.9
Number of images per species	<i>Bolivina humilis</i>	105	310	415	535
	<i>Fursenkoina</i> spp.	105	205	400	555
	<i>Suggrunda porosa</i>	100	205	405	540
	<i>Bolivina costata</i>	105	210	405	405
	<i>Bolivina advena</i>	105	210	365	365
	<i>Bolivina plicata</i>	105	210	405	400
	<i>Ebuliminella curta</i>	105	205	395	395
	<i>Epistominella obesa</i>	100	205	395	395
	<i>Cassidulina limbata</i>	105	200	390	395
	<i>Bolivina seminuda</i>	100	215	400	405
	<i>Pseudoparrella pacifica</i>	100	205	395	395
	<i>Bolivina pacifica</i>	100	110	110	110
	<i>Nonionella</i> spp.	100	155	155	155
	<i>Loxotomum pseudobeyrichii</i>	100	110	160	160
	<i>Buliminella elegantissima</i>	50	55	55	55
	Planktonic foraminifera	100	140	200	195
	Other particles	105	205	405	400
Total images		1690	3155	5455	5860

We observed that *B. elegantissima* is overrepresented in the training dataset as compared to its actual abundance in test samples. As a result, the CNN identified *B. elegantissima* in samples where, according to human counts, it was absent (Fig. 6). This highlights the importance of a well-balanced training dataset that accurately reflects the real-world species distributions. Moreover, it highlights a fundamental limitation of CNN models trained on region-specific data. Such models may not generalize well to other areas, leading to erroneous classifications due to variations in species assemblage composition and distribution.

Species that were either absent from the training dataset or present but unrecognized due to variations in orientation or morphology were categorized as “unsure” by the model. One pitfall of automated identification can be the failure to recognize species absent from the training data. However, systematically reviewing the “unsure” image folder provides an opportunity to detect novel or misclassified species, thereby addressing this concern and ensuring continuous taxonomic discovery; e.g. we found some individuals of extremely rare species that were absent in training data, like *Virgulinella fragilis*, in the “unsure” folder. It should be noted that the most abundant species, i.e. *Bolivina humilis* and *Fursenkoina* spp., show the strongest correlation between manual and automated counts (Fig. 7) despite their morphological variability, suggesting that enlarging the sample size helps to smooth out discrepancies between the two methods.

3.3 Influence of imaging background

A sufficient number of high-quality images and consistency in imaging conditions (lighting, exposure, zoom) are the basis for developing an accurate CNN model. We observed that slight variations in the background do not significantly impact the performance of the CNN model. Two types of background materials were employed: Plummer cells with a black paper background for picked specimens and a darker 3D-printed non-reflective black plate for raw sediments. To assess whether background colour affected model performance, we compared manual and automated classifications using Welch’s *t* test (a variant of Student’s *t* test that does not assume equal variances between groups; Welch, 1947). Mean relative abundance errors were nearly identical between backgrounds (2.2 % vs. 2.5 %; $t = -0.84$, $p = 0.40$), and mean absolute count errors likewise showed no difference (15.1 vs. 17.6 specimens; $t = -1.00$, $p = 0.32$). These results demonstrate that background colour exerted no systematic influence on CNN performance.

4 Ecological implications

Paleo-ecology, paleo-oxygenation, and paleo-flux of organic matter are usually reconstructed using relative abundances of benthic foraminifera (e.g. Di Bella and Casieri, 2011; Goo-day, 2003; Sharon et al., 2021; Tavera Martínez et al., 2022),

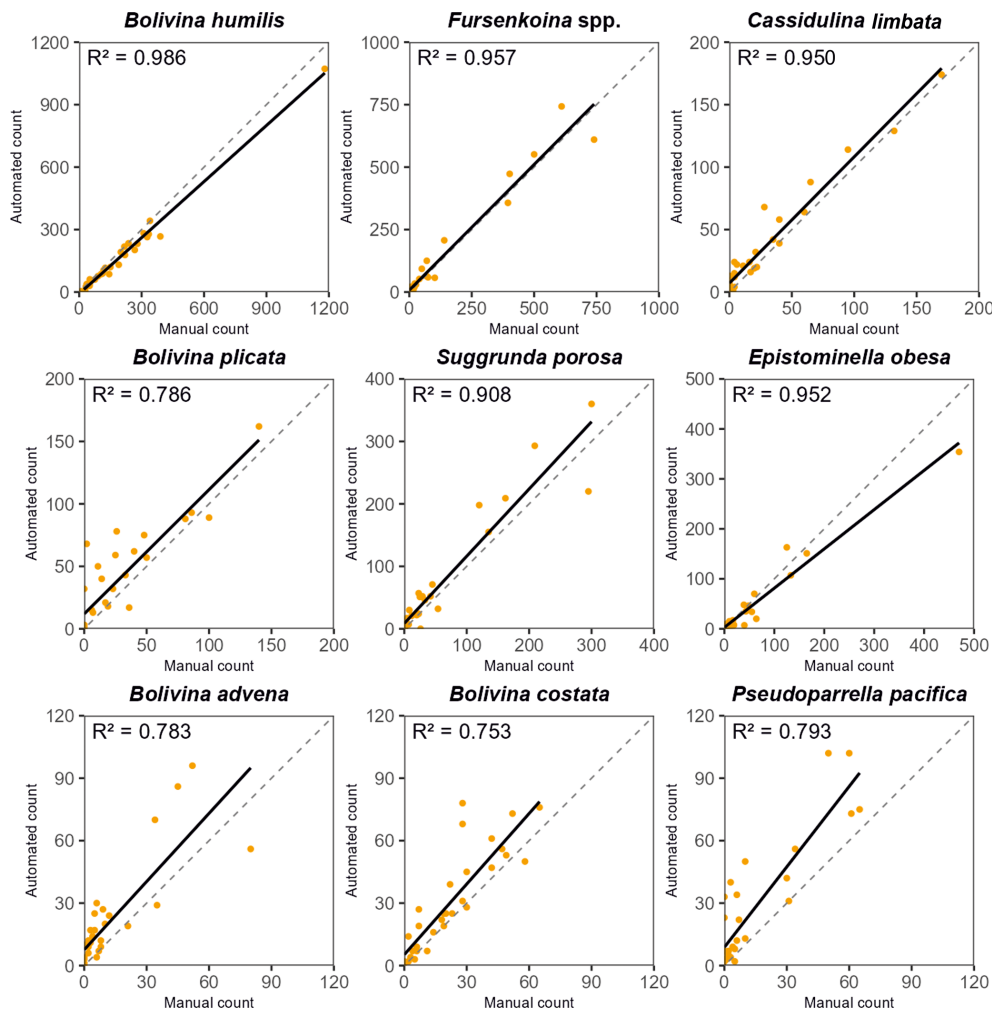


Figure 7. Regression plots comparing manual counts (x axis) and automated counts (y axis) for the most abundant species. Each plot shows samples as orange dots, a regression line (black) with the corresponding R^2 value, and a dashed perfect-agreement line.

and indices based on them, e.g. the infaunal / epifaunal (I/E) ratio (Rathburn and Corliss, 1994), Benthic Foraminiferal Oxygen Index (BFOI) (Kaiho, 1994, 1999), Extended Benthic Foraminiferal Assemblage (BFA_{ex}) index (Tetard et al., 2024), and diversity (e.g. Castillo et al., 2017; Erdem et al., 2020; Tetard et al., 2017), as the absolute abundances are influenced by sedimentation rate and sediment volume. The comparison of relative abundances (Fig. 8) obtained from manual and automated workflows demonstrated that, in the regression plots of major taxa, the data points cluster tightly along the 1 : 1 line, with high R^2 (> 0.8 in most cases). This indicates that the CNN reproduces species proportions almost exactly as a taxonomist would do (also see Fig. 9). Notably, there is no large bias; e.g. the dominant *Bolivina humilis* often thriving under low-oxygen conditions remains dominant in the automated counts, while other species showed comparable relative abundances in both methods. The low R^2 value in the case of *Bolivina costata* is mainly

due to a few outliers as the rest of the samples cluster around the regression line.

Figure 10a compares manual (red) versus automated (purple) counting based on three key indices, namely Evenness (Pielou, 1966), Shannon diversity (H) (Shannon, 1948), and Simpson diversity (1-D) (Simpson, 1949), of the same samples. The paired t test for automated and manual indices for Shannon H and Evenness (data following a normal distribution) gives p values of 3×10^{-11} and 8.9×10^{-4} , respectively, while the Wilcoxon signed-rank for Simpson 1-D (data not following a normal distribution) gives a p value of 1.8×10^{-6} . These p values show that there are differences in diversity index values between the two methods; however, the Bland–Altman plots (Fig. 10b) show that the differences are very minute, the trends are systematic, and there is no proportional bias (also see Fig. 11). For the Evenness index, both methods yield almost identical distributions (mean difference: 0.04), with automated counting-based values being slightly tighter around the median. The Shannon index based

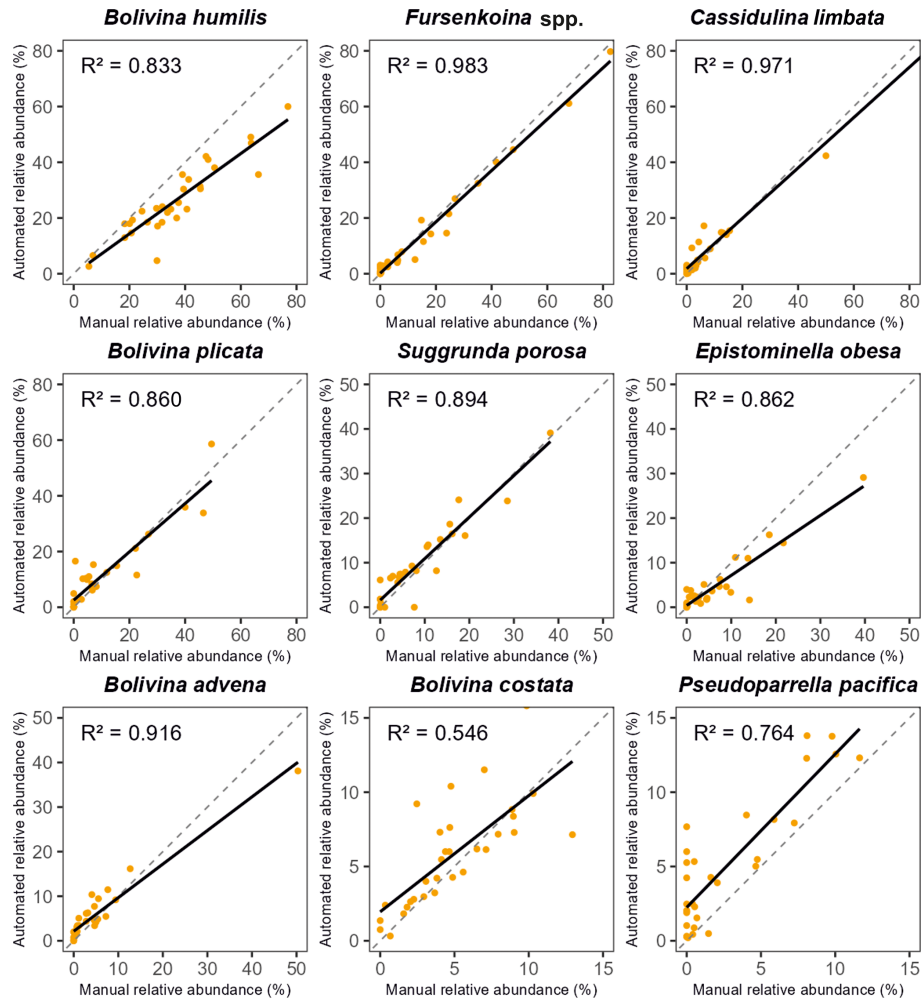


Figure 8. Regression plots comparing the relative abundance (RA) of major species based on manual counts (x axis) and automated counts (y axis). Each plot shows samples as orange dots, a regression line (solid black), a dashed perfect-agreement line (black), and the corresponding R^2 value.

on automated counts gives relatively higher diversity (mean difference: 0.3) and a broader range as compared to the manual workflow (Fig. 10a, b). However, the overall trend across the samples shows a similar pattern (see Fig. 11). The differences can be attributed to the sensitivity of the Shannon diversity index to rare taxa, which can be over-counted in automated identification (e.g. *B. elegantissima*, *B. pacifica*). The Simpson diversity index (1-D) shows tight clustering around the zero-difference line (Fig. 10b), indicating that both approaches capture dominance patterns equally well, with the manual approach yielding slightly higher values (mean difference: 0.08). This high degree of congruence is significant because benthic foraminiferal diversity has long been used as an environmental proxy, e.g. low-oxygen or high-organic-flux environments yield low diversity and strong dominance by a few tolerant species (Jorissen et al., 2007; Gooday, 2003).

For the visual comparison of differences in species composition between manual and automated methods, a Bray–Curtis-dissimilarity-based (Bray and Curtis, 1957) non-metric multidimensional scaling (nMDS) (Taguchi and Oono, 2005) was performed using the Vegan package in the R programming language (Oksanen et al., 2025; R Core Team, 2025). We used relative abundances of species, only including species with > 1 % abundances. Figure 12a presents a Bray–Curtis-dissimilarity-based nMDS ordination comparing foraminiferal assemblages derived from manual counts (red symbols) versus automated counts (purple symbols). Each point represents a single sample's species composition, positioned so that distances between points reflect their community dissimilarity. Each sample appears twice (once as a red dot for the manual count, once as a purple dot for the automated count), and these paired points mostly lie adjacent to each other, which shows that samples cluster by their inherent faunal similarity regardless of the identifi-

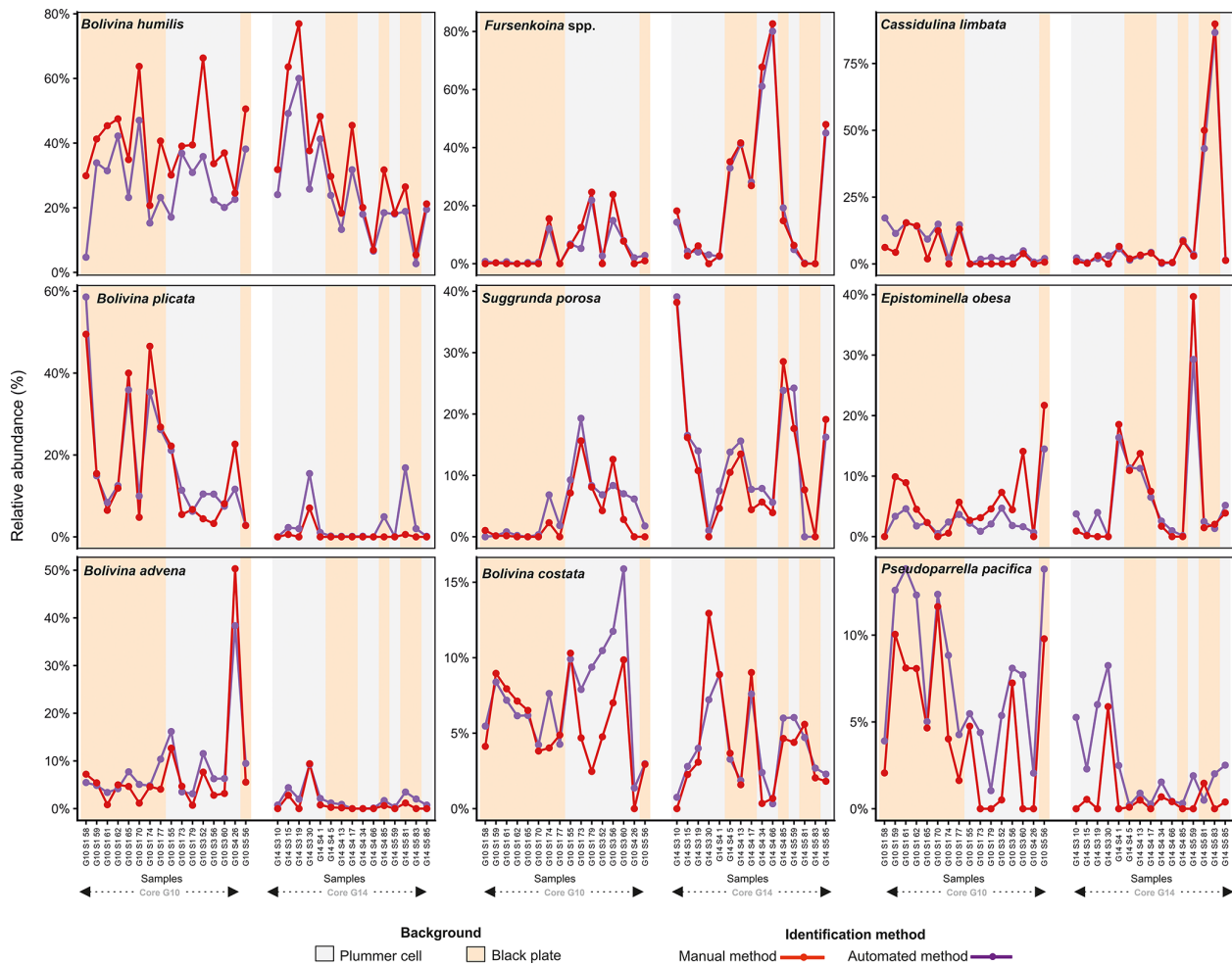


Figure 9. The relative abundance of major species in all of the studied samples in cores G10 and G14 obtained from manual and AI methods. The background of the samples is indicated as orange (black plate) and grey (Plummer cell) stripes.

cation method. The purple and red convex hulls, delineating the spread of samples in each method, overlap considerably. The stress value of 0.147 indicates a correct two-dimensional representation of community differences (Kruskal and Wish, 1978). Differences between different samples (but the same method) (median 0.35) are larger than the difference between the same sample in the two methods (median 0.17) (Fig. 12b), showing the effectiveness of automated identification in representing the species composition.

Finally, we estimated dissolved O_2 (mL L^{-1}) using BFA_{ex} (Eqs. 1 and 2) (Tetard et al., 2024), which provides oxygen class assignments for most benthic foraminiferal species in our cores. Unlike Kaiho's original BFOI (Kaiho, 1994, 1999), the BFA_{ex} framework does not treat biserial taxa as uniformly dysoxic: many biserial species occur outside dysoxic conditions, while some non-biserial species are characteristic of dysoxic to anoxic environments, highlighting the need for species-level identification rather than morphogroup-based classification. For species not covered

by BFA_{ex} , oxygen classes were assigned from published tolerances for the study area (Mallon et al., 2012; Cardich et al., 2015; Erdem et al., 2020). Only species with relative abundance $> 1\%$ were included. The oxygen classes were as follows: anoxic ($< 0.1 \text{ mL L}^{-1}$) – *Bolivina seminuda*, *B. humilis*, *Epistominella obesa*, *Suggrunda porosa*; dysoxic ($0.1\text{--}0.3 \text{ mL L}^{-1}$) – *B. advena*, *B. costata*; suboxic ($0.3\text{--}1.4 \text{ mL L}^{-1}$) – *B. plicata*, *Epistominella pacifica*, *Ebulimnella curta*; low-oxic ($1.4\text{--}3 \text{ mL L}^{-1}$) – *Fursenkoina* spp.; high-oxic ($> 3 \text{ mL L}^{-1}$) – *Cassidulina limbata*. The following formula was used to calculate BFA_{ex} :

$$BFA_{\text{ex}} = 2 \cdot (\% \text{anoxic}) + (\% \text{dysoxic}) \\ - \left(\% \text{low ox} + \frac{\% \text{other}}{2} \right) \\ - 2 \cdot \left(\% \text{high ox} + \frac{\% \text{other}}{2} \right). \quad (1)$$

BFA_{ex} was then converted into dissolved O_2 (mL L^{-1}) using:

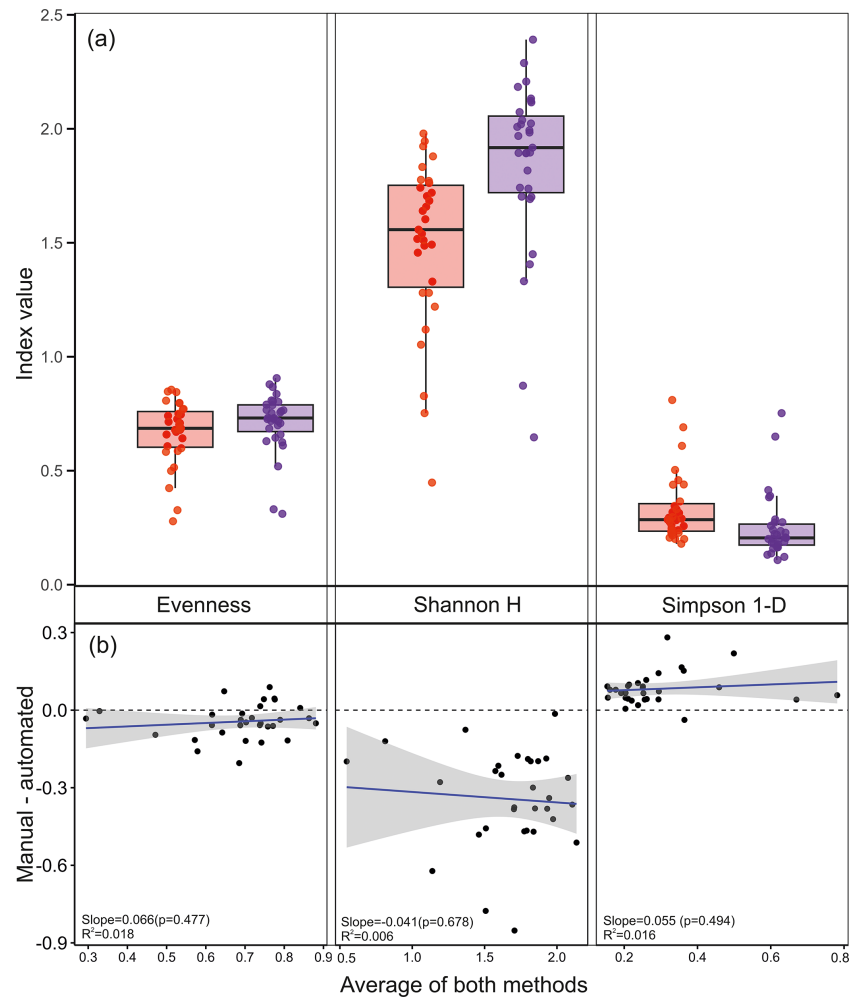


Figure 10. (a) Box-and-jitter plots for 31 samples showing the distribution of three indices (Evenness index and Shannon (H) and Simpson (1-D) diversity indices) measured using automated (purple) and manual (red) methods. (b) Bland–Altman regression: the difference between manual and automated values is plotted against their mean for all three indices. The dashed line marks zero difference, and the blue line is ordinary least-squares (OLS) fit of the difference of the mean with the 95 % confidence interval (grey). Slopes are small and non-significant, indicating no proportional bias. However, the vertical offset shows systematic differences.

$$O_2BFA_{ex} = 0.3441 \cdot \exp^{(-0.01479 \cdot BFA_{ex})}. \quad (2)$$

The dissolved- O_2 estimates based on automated and manual counting show agreement as shown in Fig. 13, indicating that automated identification can be used reliably to reconstruct the past oxygen using benthic-foraminifera-based indices.

5 Strengths and limitations of the automated identification

Our study shows that, once the CNNs are trained and calibrated to reach the required accuracy and recall, the time and effort required to process a sample is reduced significantly to 35 min per sample (Fig. 14). The most time-consuming

step is image acquisition, but it does not require continuous human supervision. Our setup used a plate with 10 compartments (Plummer cell size, each 24 cm^2), allowing for the programming of sequential imaging of 10 samples in a row. Most samples in this study are mud-dominated, which allows for an efficient concentration of foraminiferal tests by wet sieving through a $125 \mu\text{m}$ mesh. In sand-rich sediments, substantially larger bulk volumes are typically required to obtain comparable specimen yields that would lead to a higher imaging time. The Peruvian upwelling ecosystem represents an extreme environment (OMZ) where the species diversity of benthic foraminifera is low but abundance is very high; e.g. in our samples, sometimes after splitting the sample 10 times, we still end up with more than a thousand specimens in the final split. Automated identification can be a very valuable tool in high-temporal-resolution studies of such areas.

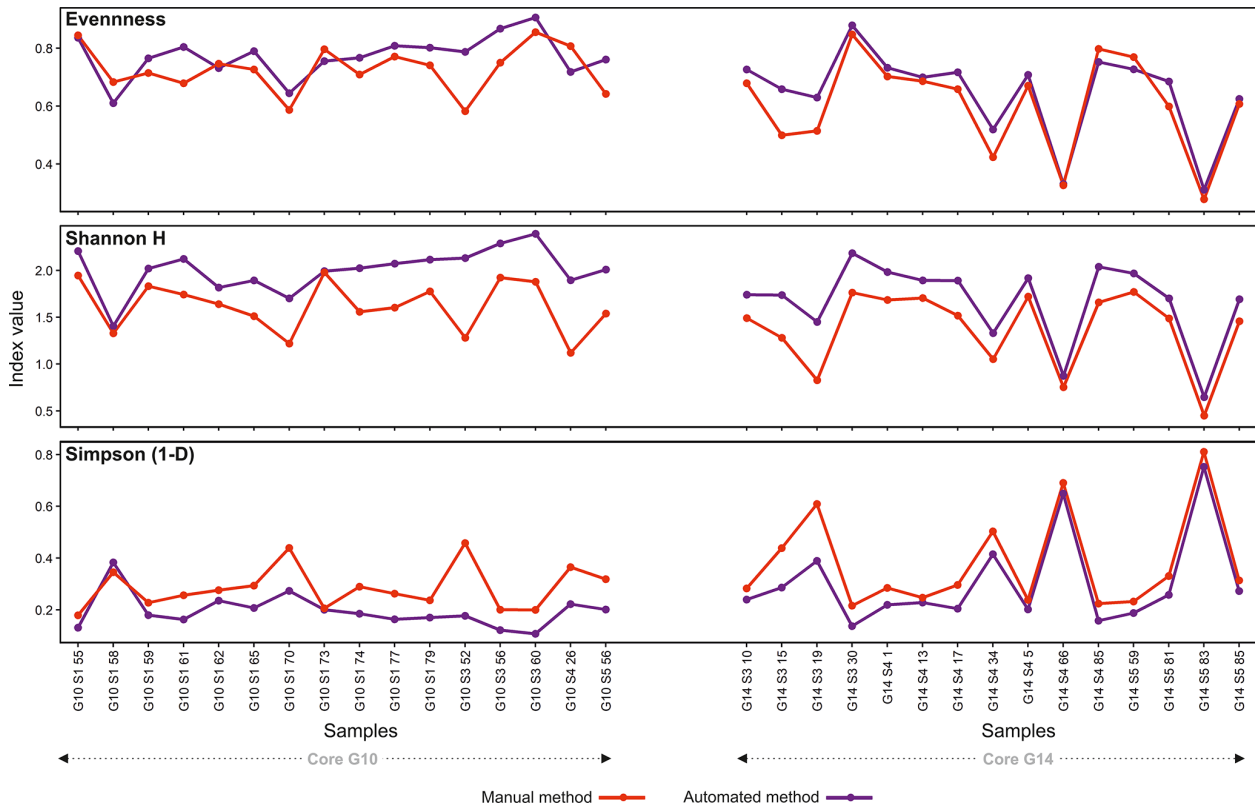


Figure 11. Diversity indices across the studied samples obtained using manual and AI methods.

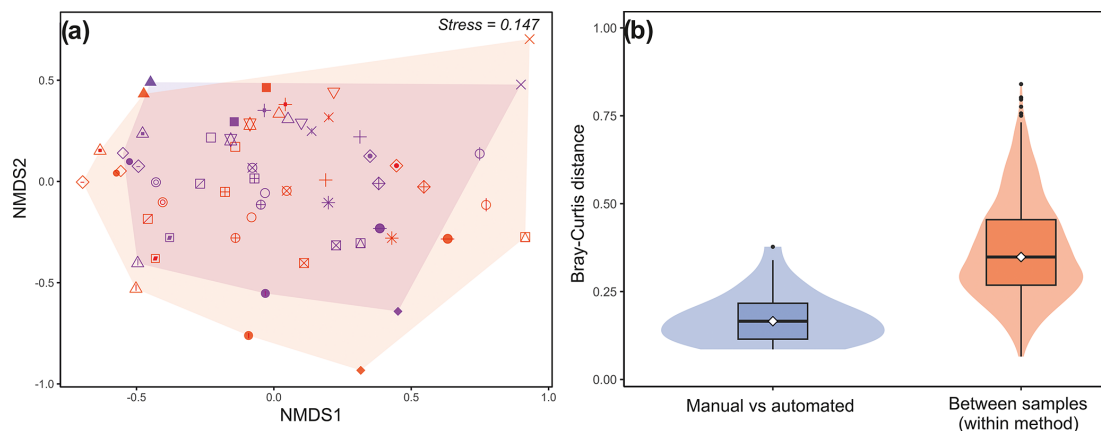


Figure 12. (a) Bray–Curtis dissimilarity-based non-metric multidimensional scaling (nMDS) of foraminiferal assemblages counted manually (red) and by CNN (purple). Convex hulls represent the spatial distribution of each group in ordination space. Different symbols indicate different samples. (b) Bray–Curtis distance distributions: left, within-sample dissimilarity between manual and automated methods; right, between-sample dissimilarities within method. Lower values indicate a more similar community composition.

The only major remaining manual component of the workflow is washing and sieving, which do not require advanced training and expertise. By eliminating the requirement of highly skilled labour, the automated system reduces not only resource allocation but also the chances of human error. If we compare this workflow with manual identification of benthic foraminifera, where we sometimes needed more than a day

for picking and identification from one sample, and if the accuracy is dependent on the level of tiredness, automated identification shows promise to become a reliable tool for paleo-ecologists. Moreover, species with fragile tests, e.g. *Fursenkoina* spp., often get fragmented during the picking and result in the loss of the specimen. Automated identification reduces such losses. Although the CNNs have the

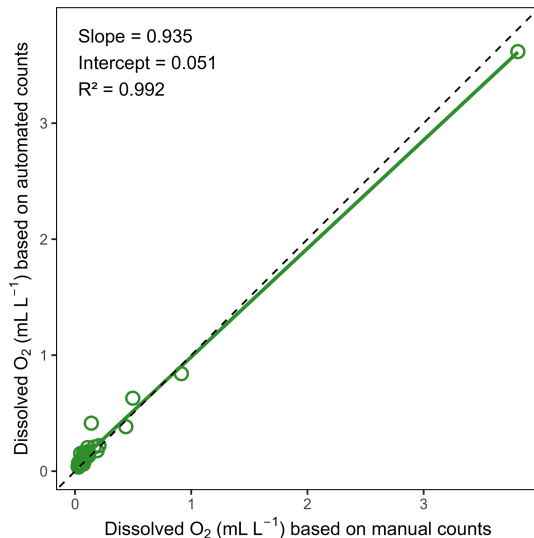


Figure 13. Comparison of dissolved O_2 (mL L^{-1}) estimated from automated and manual counts using BFA_{ex} (Tetard et al., 2024). Green circles are individual samples, the solid green line is the regression line, and the dashed line shows the 1 : 1 agreement.

potential to bring efficiency to the time-consuming process of foraminiferal identification, especially for well-preserved samples from the Late Quaternary to Holocene (e.g. Mitra et al., 2019; Marchant et al., 2020; Hsiang et al., 2019), the increase in speed and scale comes with limitations. For species whose key characteristics (e.g. aperture details or spiral–umbilical asymmetry) are only visible from a specific viewing angle, automated taxonomy based on a single resting view is inherently limited. This limitation particularly affects strongly trochospiral taxa, for which different sides can convey distinct taxonomic information. While manually reorienting each specimen or acquiring multiple views could improve classification, such approaches would substantially reduce throughput and undermine a core advantage of automated workflows. The best compromise is therefore to (i) accept the constraint of natural resting orientation and (ii) allow the model to learn from the range of orientations that occur naturally in the samples for each species. CNNs operate within predefined class boundaries and cannot accurately identify morphological intermediates, intraspecific variations, and partially preserved tests, cases where human expertise remains indispensable (Marchant et al., 2020; Harbowo and Muliawati, 2024). For instance, in the present study, the overcounting of *B. plicata* was due to its confusion with white carbonate particles or the confusion of *Fursenkoina* spp. with fragments. This limits the reuse of training data across different basins. Additionally, parameter differences during image capture can significantly affect the comparability of datasets between labs. Standardizing protocols for image capture, similar to established standards in foraminiferal studies such as those regarding mesh size, sam-

pling strategies, and the number of specimens per sample, could mitigate this issue.

Another major limitation of CNNs is their “black-box” nature, offering limited insight into the morphological criteria used for classification, unlike the interpretable logic employed by taxonomists (e.g. size, chamber and aperture shape, porosity, texture) (Pires De Lima et al., 2020). CNN-based methods are biased towards abundant taxa, reducing accuracy for rare morphotypes. Nonetheless, this limitation is unlikely to affect the main ecological and paleo-environmental conclusions that are mostly based on major species. A hybrid approach may address most of these limitations: CNNs can rapidly identify large numbers of specimens, while ambiguous cases are reviewed by a taxonomist. This method could achieve scalability without compromising accuracy.

6 Conclusions

In this study, we trained four CNN models to identify fossil benthic foraminifera from the Peruvian OMZ spanning the last 18 kyr to expedite the benthic foraminifera-based high-temporal-resolution studies of OMZ dynamics. The performance metrics of each model (trained on a different number of images) were compared, and the best-performing model – built on the most extensive, morphologically inclusive image set – was applied to 31 samples. For optimal model performance, we recommend (1) using a non-reflective black background, (2) imaging specimens in their natural resting orientation, (3) maintaining spacing between individual tests, and (4) including every morphological variant in the training set. The automated identification results closely matched manual identification for total counts, relative abundances, diversity metrics, oxygen estimation, and multivariate ordination. This agreement demonstrates the utility of automated identification and counting for paleo-ecological research and its capacity to accelerate high-temporal-resolution benthic foraminiferal studies; however, occasional overcounting of rare taxa still requires manual review.

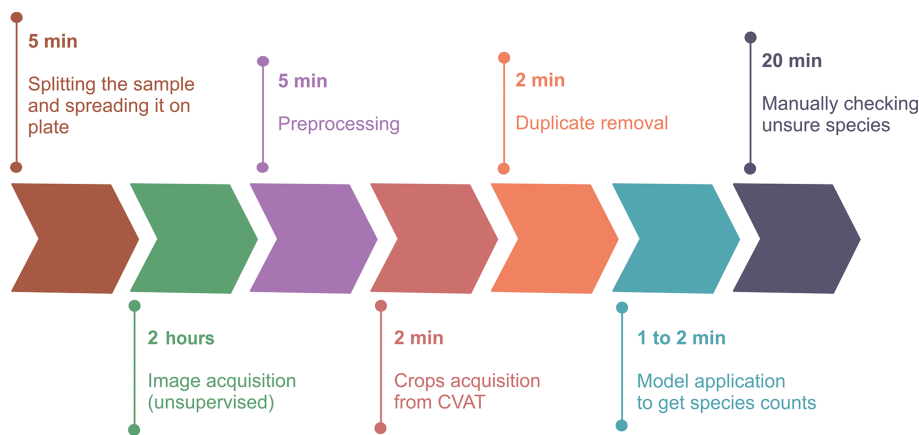


Figure 14. The time requirement of different steps in the automated identification workflow.

Appendix A

Table A1. The specification of the computer used for CVAT segmentation.

Component	Specification
GPU	NVIDIA RTX A5000 (24GB, 8192 CUDA Cores, 256 Tensor Cores)
CPU	Intel Core i7-13700 (16 Cores, 5.2 GHz Turbo)
RAM	64GB DDR5 (4400 MT s ⁻¹)
SSD (Primary Storage)	1TB M.2 PCIe NV Me (Class 40)
Hard Drive (Secondary Storage)	4TB 5400 RPM (3.5 in. (8.89 cm))
OS	Windows 11 Pro (64-bit)

Code and data availability. The raw data associated with this work are available online through the SEANO repository <https://doi.org/10.17882/108821> (Hayat et al., 2025a).

The full CNN training-image dataset is deposited on Zenodo at <https://doi.org/10.5281/zenodo.17965134> (Hayat et al., 2025b).

The code for stitching, recutting, and duplicate removal will be uploaded to <https://github.com/microfossil/> (last access: 8 August 2025; Marchant et al., 2020) upon publication of the article.

Author contributions. SH, MM, and CB designed the methodology. SH and CB carried it out. SH, MM, CB, JC, and MC analysed the results. SH prepared the paper with critical input and reviews from MM, CB, ME, JC, EG, TG, and MC.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Special issue statement. This article is part of the special issue “Advances and challenges in modern and benthic foraminifera research: a special issue dedicated to Professor John Murray”. It is not associated with a conference.

Acknowledgements. We warmly thank Geoffroy Couasnet and Louis Lanoy for their assistance in the development of the automation processes and Renato Salvatelli and Dimitri Gutierrez for generously providing us with the samples used in this study. We are grateful to the editors and reviewers for their valuable comments, which helped improve the paper. We thank Joao Barreira and Robin Fentimen for their assistance with the statistical analyses. We gratefully acknowledge funds from the Murray family that facilitated the publication of this work.

Financial support. This work was supported by the FOCUS (INSU–LEFE IMAGO CYBER programme) project. The doctoral fellowship to the first author is financed by the French Ministry of High Education and Research and by the University of Angers.

Review statement. This paper was edited by Francesca Sangiorgi and Babette Hoogakker and reviewed by Ellen Thomas and one anonymous referee.

References

- Al-Sabouni, N., Fenton, I. S., Telford, R. J., and Kučera, M.: Reproducibility of species recognition in modern planktonic foraminifera and its implications for analyses of community structure, *J. Micropalaeontol.*, 37, 519–534, <https://doi.org/10.5194/jm-37-519-2018>, 2018.
- Austen, G. E., Bindemann, M., Griffiths, R. A., and Roberts, D. L.: Species identification by experts and non-experts: comparing images from field guides, *Sci. Rep.*, 6, 33634, <https://doi.org/10.1038/srep33634>, 2016.
- Beaufort, L. and Dollfus, D.: Automatic recognition of coccoliths by dynamical neural networks, *Mar. Micropaleontol.*, 51, 57–73, <https://doi.org/10.1016/j.marmicro.2003.09.003>, 2004.
- Beaufort, L., Barbarin, N., and Gally, Y.: Optical measurements to determine the thickness of calcite crystals and the mass of thin carbonate particles such as coccoliths, *Nat. Protoc.*, 9, 633–642, <https://doi.org/10.1038/nprot.2014.028>, 2014.
- Bray, J. R. and Curtis, J. T.: An ordination of the upland forest communities of southern Wisconsin, *Ecol. Monogr.*, 27, 325–349, <https://doi.org/10.2307/1942268>, 1957.
- Brummer, G.-J. A. and Kučera, M.: Taxonomic review of living planktonic foraminifera, *J. Micropalaeontol.*, 41, 29–74, <https://doi.org/10.5194/jm-41-29-2022>, 2022.
- Cardich, J., Gutiérrez, D., Romero, D., Pérez, A., Quipúzcoa, L., Marquina, R., Yupanqui, W., Solís, J., Carhuapoma, W., Sifeddine, A., and Rathburn, A.: Calcareous benthic foraminifera from the upper central Peruvian margin: control of the assemblage by pore-water redox and sedimentary organic matter, *Mar. Ecol. Prog. Ser.*, 535, 63–87, <https://doi.org/10.3354/meps11409>, 2015.
- Carvalho, L. E., Fauth, G., Baecker Fauth, S., Krah, G., Moreira, A. C., Fernandes, C. P., and Von Wangenheim, A.: Automated microfossil identification and segmentation using a deep learning approach, *Mar. Micropaleontol.*, 158, 101890, <https://doi.org/10.1016/j.marmicro.2020.101890>, 2020.
- Castillo, A., Valdés, J., Sifeddine, A., Reyss, J.-L., Bouloubassi, I., and Ortlieb, L.: Changes in biological productivity and ocean-climatic fluctuations during the last ~1.5 kyr in the Humboldt ecosystem off northern Chile (27° S): a multiproxy approach, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 485, 798–815, <https://doi.org/10.1016/j.palaeo.2017.07.038>, 2017.
- Caulle, C., Koho, K. A., Mojtahid, M., Reichart, G. J., and Jorissen, F. J.: Live (Rose Bengal stained) foraminiferal faunas from the northern Arabian Sea: faunal succession within and below the OMZ, *Biogeosciences*, 11, 1155–1175, <https://doi.org/10.5194/bg-11-1155-2014>, 2014.
- Cushman, J. A. and McCulloch, I. A.: Some Virguliniinae in the collections of the Allan Hancock Foundation, University of Southern California Digital Library (USC.DL) [data set], <https://www.biodiversitylibrary.org/item/89276#page/204/mode/1up> (last access: 5 December 2024), 1942.
- Das, M., Singh, R. K., Gupta, A. K., and Bhaumik, A. K.: Holocene strengthening of the oxygen minimum zone in the northwestern Arabian Sea linked to changes in intermediate water circulation or Indian monsoon intensity?, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, 483, 125–135, <https://doi.org/10.1016/j.palaeo.2016.10.035>, 2017.
- Debenay, J.-P.: A Guide to 1,000 Foraminifera from Southwestern Pacific: New Caledonia, IRD Éditions, Marseille; Muséum national d'histoire naturelle, Paris, 378 pp., ISBN 978-2-7099-1729-2, 2012.
- Di Bella, L. and Casieri, S.: Paleoenvironmental reconstruction of Late Quaternary succession by foraminiferal assemblages of three cores from the San Benedetto del Tronto coast (central Adriatic Sea, Italy), *Quat. Int.*, 241, 169–183, <https://doi.org/10.1016/j.quaint.2011.03.010>, 2011.
- Dollfus, D. and Beaufort, L.: Fat neural network for recognition of position-normalised objects, *Neural Netw.*, 12, 553–560, [https://doi.org/10.1016/S0893-6080\(99\)00011-8](https://doi.org/10.1016/S0893-6080(99)00011-8), 1999.
- Erdem, Z. and Schönfeld, J.: Pleistocene to Holocene benthic foraminiferal assemblages from the Peruvian continental margin, *Palaeontol. Electron.*, 20.2.30A, <https://doi.org/10.26879/764>, 2017.
- Erdem, Z., Schönfeld, J., Rathburn, A. E., Pérez, M.-E., Cardich, J., and Glock, N.: Bottom-water deoxygenation at the Peruvian margin during the last deglaciation recorded by benthic foraminifera, *Biogeosciences*, 17, 3165–3182, <https://doi.org/10.5194/bg-17-3165-2020>, 2020.
- Fenton, I. S., Baranowski, U., Boscolo-Galazzo, F., Cheales, H., Fox, L., King, D. J., Larkin, C., Latas, M., Liebrand, D., Miller, C. G., Nilsson-Kerr, K., Piga, E., Pugh, H., Remmelzwaal, S., Roseby, Z. A., Smith, Y. M., Stukins, S., Taylor, B., Woodhouse, A., Worne, S., Pearson, P. N., Poole, C. R., Wade, B. S., and Purvis, A.: Factors affecting consistency and accuracy in identifying modern macroperforate planktonic foraminifera, *J. Micropalaeontol.*, 37, 431–443, <https://doi.org/10.5194/jm-37-431-2018>, 2018.
- Fuenzalida, R., Schneider, W., Garcés-Vargas, J., Bravo, L., and Lange, C.: Vertical and horizontal extension of the oxygen minimum zone in the eastern South Pacific Ocean, *Deep-Sea Res. II*, 56, 992–1003, <https://doi.org/10.1016/j.dsr2.2008.11.001>, 2009.
- García-Gallardo, Á., Machain-Castillo, M. L., and Almaraz-Ruiz, L.: Paleooceanographic evolution of the Gulf of Tehuantepec (Mexican Pacific) during the last ~6 millennia, Holocene, 31, 529–544, <https://doi.org/10.1177/0959683620981724>, 2021.
- Gimenez, B., Joannin, S., Pasquet, J., Beaufort, L., Gally, Y., De Garidel-Thoron, T., Combourieu-Nebout, N., Bouby, L., Canal, S., Ivorra, S., Limier, B., Terral, J., Devaux, C., and Peyron, O.: A user-friendly method to get automated pollen analysis from environmental samples, *New Phytol.*, 243, 797–810, <https://doi.org/10.1111/nph.19857>, 2024.
- Godbillot, C., Marchant, R., Beaufort, L., Leblanc, K., Gally, Y., Le, T. D. Q., Chevalier, C., and De Garidel-Thoron, T.: A new method for the detection of siliceous microfossils on sediment microscope slides using convolutional neural net-

- works, *J. Geophys. Res. Biogeosci.*, 129, e2024JG008047, <https://doi.org/10.1029/2024JG008047>, 2024.
- Gooday, A. J.: Benthic foraminifera (Protista) as tools in deep-water palaeoceanography: environmental influences on faunal characteristics, *Adv. Mar. Biol.*, 46, 1–90, [https://doi.org/10.1016/S0065-2881\(03\)46002-1](https://doi.org/10.1016/S0065-2881(03)46002-1), 2003.
- Gooday, A. J., Nomaki, H., and Kitazato, H.: Modern deep-sea benthic foraminifera: a brief review of their morphology-based biodiversity and trophic diversity, *Geol. Soc. London, Spec. Publ.*, 303, 97–119, <https://doi.org/10.1144/SP303.8>, 2008.
- Harbowo, D. G. and Muliawati, T.: Advancing the automated foraminifera fossil identification through scanning electron microscopy image classification: a convolutional neural network approach, *IOP Conf. Ser. Earth Environ. Sci.*, 1373, 012054, <https://doi.org/10.1088/1755-1315/1373/1/012054>, 2024.
- Hayat, S., Mojtahid, M., Elliot, M., Cardich, J., Geslin, E., de Garidel-Thoron, T., Carré, M., and Barras, C.: Automated and Manual counts of fossil benthic foraminifera for 31 sedimentary samples from Peruvian margin sedimentary core, SEANOE [data set], <https://doi.org/10.17882/108821>, 2025a.
- Hayat, S., Mojtahid, M., Elliot, M., Cardich, J., Geslin, E., de Garidel-Thoron, T., Carré, M., and Barras, C.: Benthic foraminifera images to train the CNN model (Automated identification of fossil benthic foraminifera from the Peruvian margin using convolutional neural networks), Zenodo [image dataset], <https://doi.org/10.5281/zenodo.17965134>, 2025b.
- Hayward, B. W., Grenfell, H. R., Carter, R., and Hayward, J. J.: World Foraminifera Database, World Register of Marine Species, <https://www.marinespecies.org/foraminifera/> (last access: 8 June 2025), 2025.
- He, K., Zhang, X., Ren, S., and Sun, J.: Deep residual learning for image recognition, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1512.03385>, 2015.
- Heinze, P.-M. and Wefer, G.: Coastal upwelling off Peru during the past 450,000 years: evidence from ODP Site 680B, *Geol. Soc. London, Spec. Publ.*, 64, 451–462, <https://doi.org/10.1144/GSL.SP.1992.064.01.30>, 1992.
- Hermelin, J. O. R.: Pliocene benthic foraminifera from the Blake Plateau: faunal assemblages and paleocirculation, *Mar. Micropaleontol.*, 10, 343–370, [https://doi.org/10.1016/0377-8398\(86\)90036-8](https://doi.org/10.1016/0377-8398(86)90036-8), 1986.
- Hsiang, A. Y. and Hull, P. M.: Automated community ecology using deep learning: a case study of planktonic foraminifera, *Ecology*, 104, e3985, <https://doi.org/10.1101/2022.10.31.514514>, 2022.
- Hsiang, A. Y., Brombacher, A., Rillo, M. C., Mleneck-Vautravers, M. J., Conn, S., Lordsmith, S., Jentzen, A., Henahan, M. J., Metcalfe, B., Fenton, I. S., Wade, B. S., Fox, L., Meilland, J., Davis, C. V., Baranowski, U., Groeneveld, J., Edgar, K. M., Movellan, A., Aze, T., and Hull, P. M.: Endless Forams: > 34 000 modern planktonic foraminiferal images for taxonomic training and automated species recognition using convolutional neural networks, *Paleoceanogr. Paleoclimatol.*, 34, 1157–1177, <https://doi.org/10.1029/2019PA003612>, 2019.
- Johansen, T. H., Sørensen, S. A., Møllersen, K., and Godtlielsen, F.: Instance segmentation of microscopic foraminifera, *Appl. Sci.*, 11, 6543, <https://doi.org/10.3390/app11146543>, 2021.
- Jorissen, F. J., Fontanier, C., and Thomas, E.: Paleoceanographical proxies based on deep-sea benthic foraminiferal assemblage characteristics, *Dev. Mar. Geol.*, 1, 263–325, [https://doi.org/10.1016/S1572-5480\(07\)01012-3](https://doi.org/10.1016/S1572-5480(07)01012-3), 2007.
- Kahanamoku-Meyer, S. S., Samuels-Fair, M., Kamel, S. M., Stewart, D., Wu, B., Kahn, L. X., Titcomb, M., Mei, Y. A., Bridge, R. C., Li, Y. S., Sinco, C., Moreno, J., Epino, J. T., Gonzalez-Marin, G., Latt, C., Fergus, H., Duijnste, I. A. P., and Finnegan, S.: An 800-year record of benthic foraminifer images and 2D morphometrics from the Santa Barbara Basin, *Sci. Data*, 11, 144, <https://doi.org/10.1038/s41597-024-02934-9>, 2024.
- Kaiho, K.: Benthic foraminiferal dissolved-oxygen index and dissolved-oxygen levels in the modern ocean, *Geology*, 22, 719–722, [https://doi.org/10.1130/0091-7613\(1994\)022<0719:BFDIOA>2.3.CO;2](https://doi.org/10.1130/0091-7613(1994)022<0719:BFDIOA>2.3.CO;2), 1994.
- Kaiho, K.: Effect of organic carbon flux and dissolved oxygen on the benthic foraminiferal oxygen index (BFOI), *Mar. Micropaleontol.*, 37, 67–76, [https://doi.org/10.1016/S0377-8398\(99\)00008-0](https://doi.org/10.1016/S0377-8398(99)00008-0), 1999.
- Kruskal, J. B. and Wish, M.: *Multidimensional Scaling*, Sage Publications, Beverly Hills, <https://doi.org/10.4135/9781412985130>, ISBN 9780803909403, 1978.
- Lo Giudice Cappelli, E. and Austin, W. E. N.: Size matters: analyses of benthic foraminiferal assemblages across differing size fractions, *Front. Mar. Sci.*, 6, 752, <https://doi.org/10.3389/fmars.2019.00752>, 2019.
- Mallon, J.: Benthic foraminifera of the Peruvian and Ecuadorian continental margin, Doctoral thesis, Christian-Albrechts-Universität zu Kiel, Kiel, Germany, urn:nbn:de:gbv:8-diss-75694, 2012.
- Mallon, J., Glock, N., and Schönfeld, J.: The response of benthic foraminifera to low-oxygen conditions of the Peruvian Oxygen Minimum Zone, in: *Anoxia*, edited by: Altenbach, A. V., Bernhard, J. M., and Seckbach, J., Springer, Dordrecht, 305–321, https://doi.org/10.1007/978-94-007-1896-8_16, 2012.
- Malmgren, K. A. and Funnell, B. M.: Benthic foraminifera from Middle to Late Pleistocene coastal upwelling sediments of ODP Hole 686B, Pacific Ocean, off Peru, *J. Micropaleontol.*, 9, 153–170, <https://doi.org/10.1144/jm.9.2.153>, 1990.
- Marchant, R., Tetard, M., Pratiwi, A., Adebayo, M., and de Garidel-Thoron, T.: Automated analysis of foraminifera fossil records by image classification using a convolutional neural network, *J. Micropaleontol.*, 39, 183–202, <https://doi.org/10.5194/jm-39-183-2020>, 2020.
- Marret, F.: The impact of artificial intelligence systems in micropaleontology, *Evolving Earth*, 1, 100022, <https://doi.org/10.1016/j.eve.2023.100022>, 2023.
- Mitra, R., Marchitto, T. M., Ge, Q., Zhong, B., Kanakiya, B., Cook, M. S., Fehrenbacher, J. S., Ortiz, J. D., Tripathi, A., and Lobaton, E.: Automated species-level identification of planktic foraminifera using convolutional neural networks, with comparison to human performance, *Mar. Micropaleontol.*, 147, 16–24, <https://doi.org/10.1016/j.marmicro.2019.01.005>, 2019.
- Murray, J. W.: *Ecology and Applications of Benthic Foraminifera*, Cambridge Univ. Press, Cambridge, xi + 426 pp., <https://doi.org/10.1007/s10933-007-9190-2>, ISBN 0 521 82839, 2008.
- Niu, Z.-B. and Xu, H.-H.: AI-based graptolite identification improves shale gas exploration, *Palaeontology [preprint]*, <https://doi.org/10.1101/2022.01.17.476477>, 2022.

- Oberhänsli, H., Heinze, P., Diester-Haass, L., and Wefer, G.: Upwelling off Peru during the last 430,000 yr and its relationship to the bottom-water environment, as deduced from coarse grain-size distributions and analyses of benthic foraminifera at Holes 679D, 680B, and 681B, Leg 112, *Proc. Ocean Drill. Prog., Sci. Results*, 112, 485–497, <https://doi.org/10.2973/odp.proc.sr.112.166.1990>, 1990.
- Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., and Wagner, H.: *Vegan: community ecology package*, CRAN package page, <https://cran.r-project.org/package=vegan> (last access: 5 June 2025), 2025.
- Palmer, H. M., Hill, T. M., Roopnarine, P. D., Myhre, S. E., Reyes, K. R., and Donnenfield, J. T.: Southern California margin benthic foraminiferal assemblages record recent centennial-scale changes in oxygen minimum zone, *Biogeosciences*, 17, 2923–2937, <https://doi.org/10.5194/bg-17-2923-2020>, 2020.
- Patterson, R. T. and Fishbein, E.: Re-examination of the statistical methods used to determine the number of point counts needed for micropaleontological quantitative research, *J. Paleontol.*, 63, 245–248, <https://doi.org/10.1017/S0022336000019272>, 1989.
- Pawlowski, J., Holzmann, M., Berney, C., Fahrni, J. F., Gooday, A. J., Cedhagen, T., and Bowser, S. S.: The evolution of early Foraminifera, *Proc. Natl. Acad. Sci. U.S.A.*, 100, 11494–11498, <https://doi.org/10.1073/pnas.2035132100>, 2003.
- Pielou, E. C.: The measurement of diversity in different types of biological collections, *J. Theor. Biol.*, 13, 131–144, [https://doi.org/10.1016/0022-5193\(66\)90013-0](https://doi.org/10.1016/0022-5193(66)90013-0), 1966.
- Pires De Lima, R., Welch, K. F., Barrick, J. E., Marfurt, K. J., Burkhalter, R., Cassel, M., and Soreghan, G. S.: Convolutional neural networks as an aid to biostratigraphy and micropaleontology: a test on Late Paleozoic microfossils, *PALAIOS*, 35, 391–402, <https://doi.org/10.2110/palo.2019.102>, 2020.
- Plavetić, M., Hsiang, A. Y., Josefson, M., Hulthe, G., and Polovodova Asteman, I.: Deep learning accurately identifies fjord benthic foraminifera, *J. Micropalaeontol.*, 44, 693–711, <https://doi.org/10.5194/jm-44-693-2025>, 2025.
- Rathburn, A. E. and Corliss, B. H.: The ecology of living (stained) deep-sea benthic foraminifera from the Sulu Sea, *Paleoceanography*, 9, 87–150, <https://doi.org/10.1029/93PA02327>, 1994.
- R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, <https://www.R-project.org/> (last access: 5 June 2025), 2025.
- Resig, J. M.: Biogeography of benthic foraminifera of the northern Nazca Plate and adjacent continental margin, *Geol. Soc. Am. Mem.*, 154, 619–666, <https://doi.org/10.1130/MEM154-p619>, 1981.
- Salvatteci, R., Field, D., Sifeddine, A., Ortlieb, L., Ferreira, V., Baumgartner, T., Caquineau, S., Velazco, F., Reyss, J.-L., Sanchez-Cabeza, J.-A., and Gutierrez, D.: Cross-stratigraphies from a seismically active mud lens off Peru indicate horizontal extensions of laminae, missing sequences, and a need for multiple cores for high-resolution records, *Mar. Geol.*, 357, 72–89, <https://doi.org/10.1016/j.margeo.2014.07.008>, 2014.
- Salvatteci, R., Gutierrez, D., Sifeddine, A., Ortlieb, L., Druffel, E., Boussafir, M., and Schneider, R.: Centennial- to millennial-scale changes in oxygenation and productivity in the Eastern Tropical South Pacific during the last 25 000 years, *Quat. Sci. Rev.*, 131, 102–117, <https://doi.org/10.1016/j.quascirev.2015.10.044>, 2016.
- Salvatteci, R., Gutierrez, D., Field, D., Sifeddine, A., Ortlieb, L., Caquineau, S., Baumgartner, T., Ferreira, V., and Bertrand, A.: Fish debris in sediments from the last 25 kyr in the Humboldt Current reveal the role of productivity and oxygen on small pelagic fishes, *Prog. Oceanogr.*, 176, 102114, <https://doi.org/10.1016/j.pocean.2019.05.006>, 2019.
- Schiebel, R. and Hemleben, C.: *Planktic Foraminifera in the Modern Ocean*, Springer, Berlin/Heidelberg, 358 pp., <https://doi.org/10.1007/978-3-662-50297-6>, 2017.
- Schiebel, R., Smart, S. M., Jentzen, A., Jonkers, L., Morard, R., Meilland, J., Michel, E., Coxall, H. K., Hull, P. M., de Garidel-Thoron, T., Aze, T., Quillévéré, F., Ren, H., Sigman, D. M., Vohhof, H. B., Martínez-García, A., Kučera, M., Bijma, J., Spero, H. J., and Haug, G. H.: Advances in planktonic foraminifer research: new perspectives for paleoceanography, *Earth-Sci. Rev.*, 185, 739–765, <https://doi.org/10.1016/j.revmic.2018.10.001>, 2018.
- Schönfeld, J., Alve, E., Geslin, E., Jorissen, F., Korsun, S., Spezzaferri, S., Abramovich, S., Almogi-Labin, A., Armynot du Chatelet, E., Barras, C., Bergamin, L., Bicchi, E., Bouchet, V., Cearreta, A., Di Bella, L., Dijkstra, N., Trevisan Disaro, S., Ferraro, L., Frontalini, F., Gennari, G., Golikova, E., Haynert, K., Hess, S., Husum, K., Martins, V., McGann, M., Oron, S., Romano, E., Mello Sousa, S., and Tsujimoto, A.: The FOBIMO (FORaminiferal Bio-MONitoring) initiative – Towards a standardised protocol for soft-bottom benthic foraminiferal monitoring studies, *Mar. Micropaleontol.*, 94–95, 1–13, <https://doi.org/10.1016/j.marmicro.2012.06.001>, 2012.
- Sen Gupta, B. K.: Foraminifera in marginal marine environments, in: *Modern Foraminifera*, Springer, Dordrecht, 141–159, https://doi.org/10.1007/0-306-48104-9_9, 1999.
- Shannon, C. E.: A mathematical theory of communication, *Bell System Technical Journal*, 27, 379–423, <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>, 1948.
- Sharon, S. N., Belanger, C. L., Du, J., and Mix, A. C.: Reconstructing paleo-oxygenation for the last 54,000 years in the Gulf of Alaska using cross-validated benthic foraminiferal and geochemical records, *Paleoceanogr. Paleoclimatol.*, 36, e2020PA003986, <https://doi.org/10.1029/2020PA003986>, 2021.
- Simpson, E. H.: Measurement of diversity, *Nature*, 163, 688, <https://doi.org/10.1038/163688a0>, 1949.
- Smith, P. B.: Quantitative and Qualitative Analysis of the Family Boliviniidae (Recent Foraminifera off Central America), *Geological Survey Professional Paper*, United States Government Printing Office, Washington, 1963.
- Taguchi, Y.-H. and Oono, Y.: Relational patterns of gene expression via non-metric multidimensional scaling analysis, *Bioinformatics*, 21, 730–740, <https://doi.org/10.1093/bioinformatics/bti067>, 2005.
- Tavera Martínez, L., Marchant, M., Muñoz, P., and Abdala Díaz, R. T.: Spatial and vertical benthic foraminifera diversity in the oxygen minimum zone of Mejillones Bay, northern Chile, *Front. Mar. Sci.*, 9, 821564, <https://doi.org/10.3389/fmars.2022.821564>, 2022.
- Tetard, M., Licari, L., and Beaufort, L.: Oxygen history off Baja California over the last 80 kyr: a new foraminiferal-based record, *Paleoceanography*, 32, 246–264, <https://doi.org/10.1002/2016PA003034>, 2017.

- Tetard, M., Marchant, R., Cortese, G., Gally, Y., de Garidel-Thoron, T., and Beaufort, L.: Technical note: A new automated radiolarian image acquisition, stacking, processing, segmentation and identification workflow, *Clim. Past*, 16, 2415–2429, <https://doi.org/10.5194/cp-16-2415-2020>, 2020.
- Tetard, M., Prebble, J. G., and Cortese, G.: Dissolved oxygen affinities of hundreds of benthic foraminiferal species, *Mar. Micropaleontol.*, 190, 102380, <https://doi.org/10.1016/j.marmicro.2024.102380>, 2024.
- Welch, B. L.: The generalization of “Student’s” problem when several different population variances are involved, *Biometrika*, 34, 28–35, <https://doi.org/10.2307/2332510>, 1947.
- Yayan, K., Bağlum, C., and Yayan, U.: Multi-label benthic foraminifera identification with convolutional neural networks, *IEEE Access*, 12, 196769–196785, <https://doi.org/10.1109/ACCESS.2024.3520633>, 2024.
- Zhong, B., Ge, Q., Kanakiya, B., Marchitto, T. M., and Lobaton, E.: A comparative study of image classification algorithms for Foraminifera identification, in: 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 1–8, <https://doi.org/10.1109/SSCI.2017.8285164>, 2017.