



Deep learning accurately identifies fjord benthic foraminifera

Marko Plavetić^{1,★}, Allison Yi Hsiang^{2,3,★}, Mats Josefson⁴, Gustaf Hulthe⁴, and Irina Polovodova Asteman⁵

¹Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, 413 90, Sweden

²Department of Geological Sciences, Stockholm University, Stockholm, 114 18, Sweden

³Department of Ecology, Environment and Plant Sciences, Stockholm University, Stockholm, 114 18, Sweden

⁴Global Product Development, Pharmaceutical Technology and Development, Operations, AstraZeneca, Mölndal, 431 83, Sweden

⁵Department of Marine Sciences, University of Gothenburg, Gothenburg, 413 90, Sweden

★These authors contributed equally to this work.

Correspondence: Irina Polovodova Asteman (irina.polovodova@marine.gu.se)

Received: 1 November 2024 – Revised: 17 November 2025 – Accepted: 28 November 2025 – Published: 17 December 2025

Abstract. Over the past several decades, there has been increasing interest in using foraminifera as environmental indicators for coastal marine environments. Foraminifera provide equally good environmental quality status assessment as compared to large invertebrates (macrofauna), which are currently used as biological quality elements. However, foraminifera offer several distinct advantages as bioindicators, including short response and generation times, a high number of individuals per small sample volume, and hard and fossilizing shells with a potential of paleoecological record. One of the major challenges in foraminifera identification is the reliance on manual morphological methods, which are not only time-consuming and error-prone but also highly dependent on the expertise of taxonomic specialists. Deep learning, a subfield of machine learning (ML), has emerged as a promising solution to this challenge, since a neural network can learn to recognize subtle differences in shell morphology that may be difficult for the human eye to distinguish. In addition, the speed and ease afforded by deep learning methods would allow experts and non-experts alike to use foraminifera more extensively in their work, thus helping to integrate the use of foraminifera in biomonitoring programs by agencies and industry. This study focuses on benthic foraminifera from several Skagerrak fjords, including Gullmar Fjord, Hakefjord, Sannäs Fjord, and Idefjord (Fig. 1a). Sediment archives from these fjords provide extensive records of past and ongoing climate and environmental changes. Fjord foraminifera mounted on microslides were imaged using a stereomicroscope (3003 images), and individual foraminifera were labeled using the Roboflow online platform (22 138 individuals). Using the labeled images, we trained a You Only Look Once (YOLO) v7 deep learning model, which demonstrates state-of-the-art speed and performance for object detection as of the time of writing. The models can distinguish among 29 species with 90.3 % and 78.8 % mean average precision in the best- and the worst-performing models, respectively. Even though the imaging and labeling was done in a short amount of time (~300 h over a course of 2 months), the results show that even a relatively small dataset can be used for training a reliable deep learning species identification model.

1 Introduction

Benthic foraminifera are single-celled eukaryotes that are widely used in environmental monitoring due to their sensitivity to environmental changes (e.g., Bouchet et al., 2012; Dolven et al., 2013; Polovodova Asteman et al., 2015; Alve et al., 2019; O'Brien et al., 2021). The majority of benthic foraminifera construct calcareous or agglutinated shells; the high preservation potential of these shells, in combination with their high abundance in small sample volumes and the rapid response of benthic foraminifer populations to environmental changes, makes benthic foraminifera ideal for tracking both short- and long-term ecological shifts in modern and ancient marine environments (Alve et al., 2009). Foraminiferal distribution, diversity, and morphology can provide key insights into oceanographic conditions, including temperature, salinity, oxygen levels, and pollution (Alve, 1995; Murray, 2006). The above-mentioned qualities make these microeukaryotes useful proxies in paleoceanography and paleoclimate research and valuable bioindicators in environmental research. Over the past several decades, benthic foraminifera have been increasingly used in environmental monitoring in coastal waters and were shown to characterize environmental quality status equally well as traditional macrofaunal bioindicators (Dolven et al., 2013; Alve et al., 2019). Foraminifera have been applied to monitor pollutants such as heavy metals (e.g., Alve, 1991; Frontalini and Cocchioni, 2008; Dolven et al., 2013; Polovodova Asteman et al., 2015; Boehnert et al., 2020; Lintner et al., 2021) and hydrocarbons (e.g., Morvan et al., 2004; Ernst et al., 2006; Brunner et al., 2013; Schwing et al., 2015; Frontalini et al., 2020), with their population dynamics offering insights into ecosystem health, particularly in areas experiencing stress from hypoxia or eutrophication (Bernhard and Alve, 1996; Filipsson and Nordberg, 2004).

Traditionally, the identification of foraminifera has relied on manual identification methods based on shell morphology (e.g., pore patterns, wall texture, sutures, chamber arrangement), requiring expert knowledge and significant time investment. This process involves the examination of samples under a stereomicroscope to manually pick specimens (generally < 1 mm in size, using a fine brush) and to count and classify the organisms to the lowest possible taxonomic level. Manual identification is not only slow but also prone to human error (e.g., Austen et al., 2016; Austen et al., 2018; Fenton et al., 2018), leading to potential biases and misinterpretations of results (see Hsiang et al., 2019, for discussion of rarity biases in human classifiers of planktonic foraminifera). Additionally, manual identification limits the scale of analysis, as it is difficult to process large datasets manually in a reasonable time frame, resulting in a lengthy process of data collection taking months to years depending on the size of the team involved (Hsiang et al., 2019). Given the increasing demand for high-throughput environmental monitoring data, particularly in response to pressing issues such as climate

change and pollution, there is a growing need to automate the identification process of foraminifera to facilitate their use as bioindicators, as advocated by the FOBIMO protocol (Schönfeld et al., 2012).

Recent advances in machine learning (ML), particularly deep learning techniques, offer promising solutions to these challenges. Convolutional neural networks (CNNs), a class of deep learning algorithm designed to process, classify, and segment images, have revolutionized the field of image recognition (Krizhevsky et al., 2012; LeCun et al., 2015). CNNs are highly effective at learning complex patterns and features in images, making them ideal for tasks such as species identification from microscopy images. CNNs achieve this by automatically learning to detect features such as edges, textures, and shapes in images by using convolutional layers, which apply filters to the input images, creating feature maps that highlight specific visual characteristics of the objects in the image (Dumoulin and Visin, 2016). By stacking multiple convolutional layers, CNNs can learn to recognize complex patterns such as the intricate morphological details of foraminiferal shells. Unlike traditional image recognition methods, CNNs can perform feature extraction without manual pre-processing and hand-crafted features by learning features directly from the data. This allows CNNs to generalize better to new images (Castro et al., 2017), allowing more widespread deployment and application of trained models.

While CNNs are effective for image classification, many ecological studies require not only classification but also object detection. Object detection involves identifying and localizing specific objects within an image. This is particularly important in cases where multiple objects of interest (such as different species of foraminifera) may appear in the same image, and the model needs to detect and classify each object individually. The You Only Look Once (YOLO) family of models represents a state-of-the-art approach to object detection in deep learning (Redmon et al., 2015). YOLO models are known for their speed and accuracy, making them particularly well suited for real-time applications where quick detection is crucial, such as for face recognition algorithms in mobile devices and smart cameras (Chen et al., 2021; Yu et al., 2024). In the context of foraminiferal research, machine learning models like YOLO can automate the detection and classification of species, dramatically reducing the time required for analysis. Several studies have successfully applied deep learning to planktonic foraminifera (Mittra et al., 2019; Hsiang et al., 2019; Marchant et al., 2020; Karaderi et al., 2022; Ferreira-Chacua and Koeshidayatullah, 2023), but there has been comparatively limited research focused on the application of these techniques to benthic foraminifera, particularly those from fjord environments (but see Govindakutty Menon et al., 2023, deep learning for pore detection).

The Skagerrak (North Sea) and its fjords, located along the Swedish and Norwegian coastlines, are home to a rich diversity of benthic foraminifera with communities ranging

from estuarine to oceanic species (Fig. 1b). The fjords represent contrasting environmental conditions, ranging from well oxygenated (Hakefjord: Watts et al., 2024; O'Brien et al., 2025) to severely hypoxic (Gullmar Fjord and Sannäs Fjord: Nordberg et al., 2000, 2017; Choquel et al., 2021) and from relatively deep (Gullmar Fjord: > 110 m) to shallow water conditions (Sannäs Fjord: 7–25 m; Nordberg et al., 2017). Despite serving as important carbon sinks (Watts et al., 2024), some of the Skagerrak fjords today experience high natural and anthropogenic pressures, including eutrophication, various industrial discharges, introduction of alien and invasive species, and climate-related changes in bottom water temperature, salinity, and oxygenation (e.g., Filipsson and Nordberg, 2004; Dolven et al., 2013; Polovodova Asteman et al., 2015; Karlsson et al., 2018; Mattsson et al., 2022; Choquel et al. 2023; O'Brien et al., 2025; McGann et al., 2025). For instance, species typical of the so-called Skagerrak–Kattegat (S–K) fauna (*Cassidulina laevigata*, *Textularia earlandi*, *Bulimina marginata*, *Liebusella goesi*, *Hyalinea balthica*, and *Nonionellina labradorica*), the opportunistic species *Stainforthia fusiformis*, and the non-indigenous and putatively invasive *Nonionella* sp. T1 are commonly used as bioindicators for monitoring environmental changes in these regions (Nordberg et al., 2000; Dolven et al., 2013; Polovodova Asteman and Schönfeld, 2015). The S–K fauna and *S. fusiformis* show clear responses to bottom water hypoxia, making them valuable for reconstructing past and present ecological conditions (Alve, 2003; Polovodova Asteman and Nordberg, 2013). Furthermore, the recent discovery of putatively invasive *Nonionella* sp. T1 in the Skagerrak has raised concerns about the shifting dynamics within the foraminiferal communities there (e.g., Polovodova Asteman and Schönfeld, 2015; Deldicq et al., 2019; Brinkmann et al., 2023; Morin et al., 2023). This species now thrives in nitrate-rich fjord sediments and may outcompete native species under severely hypoxic conditions due to its ability to respire nitrate through denitrification (Choquel et al., 2021).

With the ongoing climate change shifting climatic zones and intense marine traffic, more and more coastal settings will become a new home for alien and invasive foraminifera species, further emphasizing the need for new and more efficient monitoring methods offered by deep learning and image recognition. This study explores the application of convolutional neural networks (CNNs) and the You Only Look Once (YOLO) architecture to identify benthic foraminifera species from the Skagerrak and its fjords. Our aim is to demonstrate the feasibility of these models for performing complex morphological identification and detection of benthic foraminifera, offering a potential solution to the limitations of manual identification of benthic foraminifera and thus opening up the path towards their widespread usage as bioindicators for rapid, automated monitoring of marine coastal environments.

2 Materials and methods

2.1 Image acquisition

The dataset for this study was generated from > 70 sediment samples collected from the Skagerrak fjords, including the Gullmar Fjord, the Hakefjord, the Sannäs Fjord, and the Idefjord over the time period 2009–2023 (Figs. 1 and 2a). These fjords are located along the western coastline of Sweden and are known for their ecological diversity and dynamic marine environments.

The foraminiferal specimens were picked and mounted on micropaleontological slides (Fig. 2c). Prior to this, all sediment samples were washed over 63 and 1000 µm sieves and dried at 50 °C, and at least 300 individuals were picked from non-stained samples. No sample treatment was applied to concentrate foraminifera because of fjord sediments being composed of fine-grained organic-rich clay. The foraminiferal specimens were imaged using a Nikon SMZ-10 stereomicroscope equipped with a DeltaPix DP450 microscope camera (Fig. 2d). The camera provided a resolution of 1600 × 1200 pixels (1.92 MP), and images were captured at 30× magnification, resulting in an optical resolution of approximately 4.16 µm per pixel. To maintain consistency across images, the exposure time, lighting angle, light intensity, and magnification were kept as constant as possible throughout the imaging process. Light intensity was kept at the maximum of what the ring lamp could supply, and we opted for the ring lamp because the illumination with such a setup is more constant than with arm lights which can be bumped or moved during different imaging sessions. A total of 3095 sample images (each containing multiple individual specimens) were obtained over a 2-month period (Fig. 2e). The foraminifera in these images were oriented in various ways due to the use of both mounted (i.e., glued-on slides with fixed orientations) and non-mounted (i.e., not glued, with free movement allowed during manipulation) specimens, which proved beneficial for training the machine learning model. By capturing different views of each specimen, the model was better able to learn species-specific morphological features that might not be apparent from a single orientation.

Additionally, during the SEEPS II research cruise conducted in April 2023, a second dataset of benthic foraminifera was prepared using surface sediment samples (> 63 mm fraction sediment residue) collected in the western Skagerrak and the Kattegat (Fig. 2b; 2011 images). This dataset was processed differently from the initial Skagerrak fjord dataset: the foraminifera were not picked; thus the samples containing both foraminifera and detrital material left after sieving were imaged from a micropaleontological tray (Fig. 2f). This provided an additional opportunity to evaluate the model's performance on real-world samples, as opposed to only on picked and cleaned specimens mounted on or stored in slides.

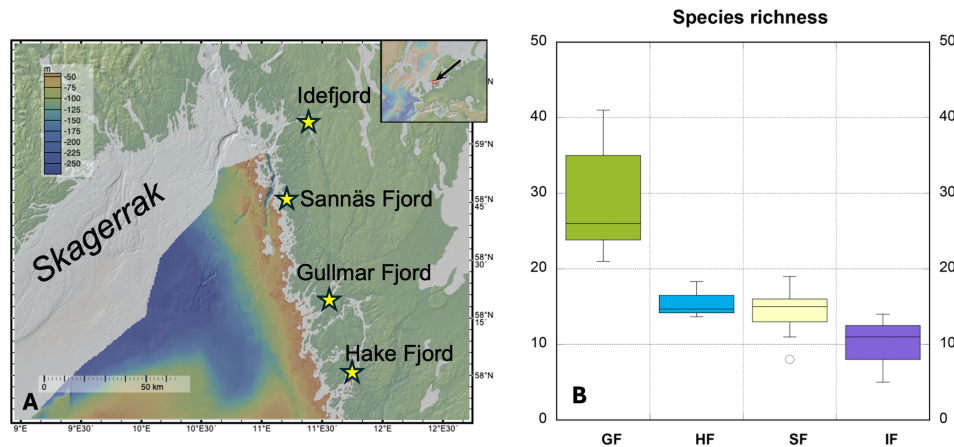


Figure 1. (a) Map over the Skagerrak fjords on the Swedish west coast, where the foraminiferal samples used in this study originate. The figure was created with GeoMapApp (<http://www.geomapp.org>, last access: 10 October 2025)/CC BY. (b) Range of foraminifera species richness observed in the studied fjords showing clear diversity changes moving from the oceanic and deep-sea-like conditions (Gullmar Fjord) to progressively more estuarine conditions with increased riverine input (Idefjord). Abbreviations GF, HF, SF, and IF stand for Gullmar Fjord, Hakefjord, Sannäs Fjord, and Idefjord, respectively. Horizontal lines show median richness values.

2.2 Image processing and labeling

The collected images were saved in TIFF format for high-quality preservation, as this format does not compromise image quality due to compression artifacts. These images were then processed using Roboflow, a web-based application programming interface (API) designed for machine learning tasks. Roboflow allows users to easily label images and create datasets suitable for training machine learning models (Fig. 3). One of the most valuable features of Roboflow is its content-aware selection tool, which automatically detects the boundaries of objects, significantly reducing the time required for manual labeling. A total of 22 138 individual foraminifera were manually labeled across the dataset by the first author over a 2-month period (Fig. 2g). On average the first author spent 1 h labeling per working day, equaling ~ 80 h total. The labeling process involved manually drawing polygons around each foraminifera specimen in the images and assigning a species label to each specimen. This labeled dataset was then exported for use in model training.

2.3 Dataset creation

Once the images were labeled, the dataset was split into three subsets (70/20/10 %) for the training, validation, and testing sets, respectively. In total, 3003 images were used to create the dataset, containing 22 138 labeled individuals. However, not all species identified in the original sediment samples were included in the training dataset. Out of 59 identified species, 29 species were chosen for training (Fig. 4). The remaining 30 species were excluded due to having too few labeled individuals (fewer than 72 individuals), which could have reduced the accuracy of the model during training. The cutoff point for species inclusion was set at 72 la-

beled individuals based on initial model testing. Model performance for those less abundant species with an individual count lower than 72 resulted in precision less than 0.5; therefore only species with precision higher than 0.5 were included.

To increase the effective size of the dataset, data augmentation techniques were applied (Fig. 2h). Data augmentation involves making systematic changes to the images, such as adjusting brightness, contrast, and vibrance, as well as horizontally and vertically flipping the images. The augmentations used in this work included flipping the images horizontally and vertically and rotating them 90° clockwise, counter-clockwise and upside down. Hue, vibrance, brightness, and exposure were adjusted randomly between -25% and $+25\%$. Artificial noise was added up to 5 % of total pixels. These augmentations create new versions of the original images that are slightly different from each other, allowing the model to learn to recognize species under varying conditions. After augmentation, the dataset contained 7089 images and 49 459 labeled foraminifera. As the augmentation process was done after annotation, it did not require additional manual annotation, as the annotations from the original images were retained.

2.4 Object detection with You Only Look Once (YOLO)

Unlike object detection methods which rely on sliding windows or region proposals, such as Mask-RCNN (He et al., 2017), YOLO models divide the input image into a grid and predict bounding boxes and class probabilities for each grid cell simultaneously. This allows YOLO to detect objects in a single forward pass through the network, significantly reducing computation time (Wang et al., 2023). YOLOv7, the version used in this study, builds on previous iterations by

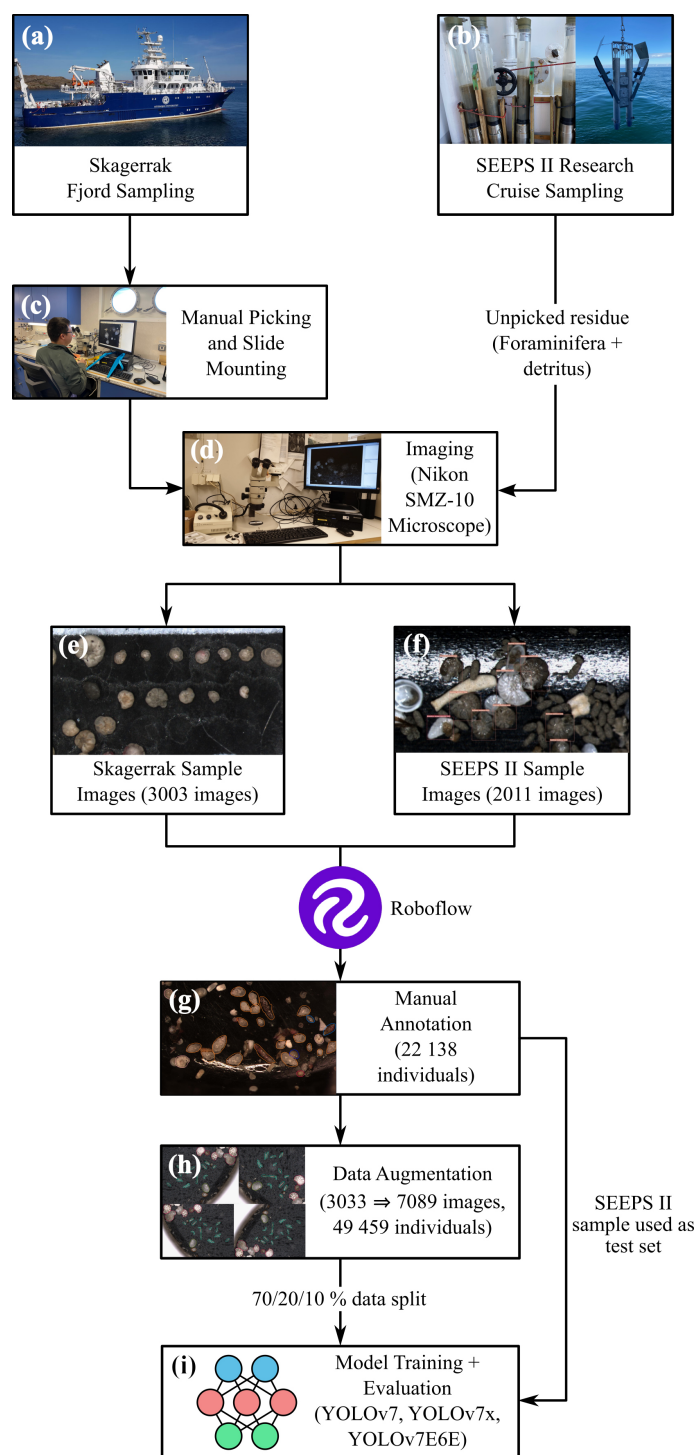


Figure 2. Analysis pipeline for training deep learning models using benthic foraminifera. The samples are sourced from two sampling initiatives: the Skagerrak fjord sampling (a) and the SEEPS II research cruise sampling (b). For the Skagerrak samples, benthic foraminifera were manually picked and mounted on slides (c). The SEEPS II samples were not manually picked and were instead imaged with foraminifera and detritus spread together in a picking tray. The samples were imaged using a Nikon SMZ-10 microscope (d), resulting in 3003 images from the Skagerrak sample (e) and 2011 images from the SEEPS II sample (f). Roboflow was then used to manually annotate 22 138 individuals from the combined samples (g). Data augmentation increased the number of images to 7089, resulting in a total of 49 459 individuals (h). This aggregated dataset was then split into training, validation, and test datasets (70/20/10 % split) and used to train the YOLOv7, YOLOv7x, and YOLOv7E6E models (i).

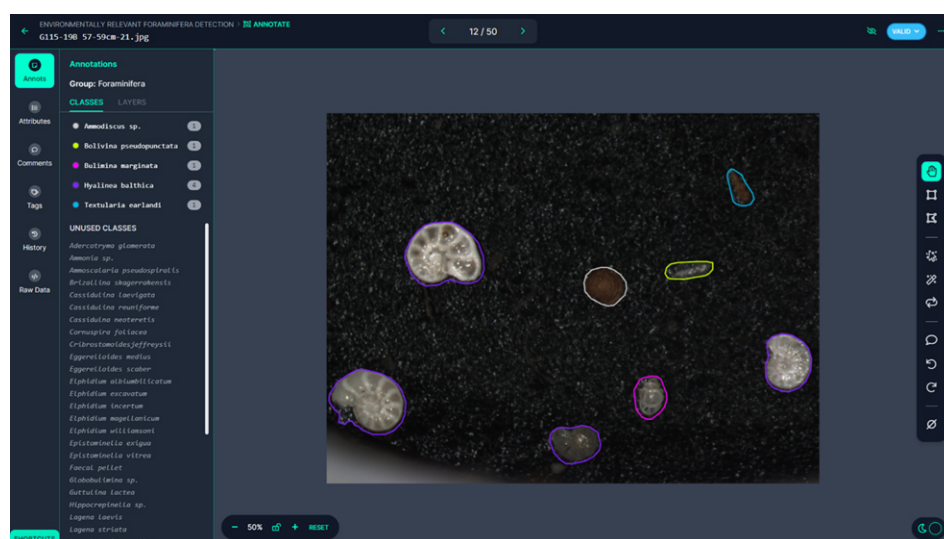


Figure 3. Roboflow interface for classifying individual foraminifera in an image. On the left side of the window, there is a color-coded list of labeled foraminifera in an image on the right. On the rightmost side, there is a tool selection panel, from which a user can label foraminifera either by using an automatic feature extraction tool or by manually drawing a polygon around the foraminifera. In this study, the specimens were labeled using manually drawn polygons.

improving detection accuracy through reparametrized convolutions and model scaling (Wang et al., 2023). These innovations enable YOLOv7 to perform well on complex datasets while maintaining fast processing times. The model is particularly efficient in detecting small objects, which is a critical feature for identifying foraminifera in images taken from micropaleontological slides, where the minute organisms are non-mounted and can be closely packed together.

In this study, YOLOv7 was chosen for its balance between speed and accuracy based on its performance on the Common Objects in Context (COCO) dataset, a standard dataset for computer vision model evaluation and benchmarking (Lin et al., 2014). To find an optimal model it is important to look at their average precision (AP), which measures accuracy of the model, and to look at batch 1 frame per second (FPS) and batch 32 average time (Table 1). Batch 1 FPS measures frames per second when processing one image at a time. Higher FPS means faster processing, indicating that the model can process images more quickly. Such performance is valuable for real-time applications. Batch 32 average time measures the average time needed to process a batch of 32 images in milliseconds. Lower times are better as they indicate faster processing, valuable for processing images and videos. The model's ability to detect multiple objects in a single image makes it ideal for processing large datasets of benthic foraminifera, where multiple individuals may be present in each frame. The use of YOLOv7 also allows real-time detection applications, which could potentially be extended to mobile or desktop-based tools for species identification in the future.

2.5 Transfer learning for small datasets

One of the challenges in applying deep learning models to ecological datasets is the relatively small size of the datasets compared to those typically used in fields like computer vision. In this study, the dataset consisted of 3003 images with 22 138 labeled individuals, which is relatively small compared to large-scale datasets like ImageNet (Deng et al., 2009) or the COCO database (Lin et al., 2014), which contain millions of labeled images. Training deep learning models from scratch on such small datasets can lead to overfitting, where the model memorizes the training data but performs poorly on new, unseen data (i.e., poor generalizability). To address this issue, transfer learning was used in this study. Transfer learning involves using a model that has already been trained on a large dataset for a different task and fine-tuning it for a new task with a smaller dataset (He et al., 2015). In the case of YOLOv7, the model was pre-trained on the COCO database (Table 1), which contains a diverse set of objects from everyday scenes. By starting with a model that already “knows” how to detect general objects, we can fine-tune it to recognize specific objects like foraminifera with far fewer training images. This approach allows the model to leverage the knowledge it gained from the large dataset and apply it to the new task, thereby improving its performance even with a relatively small training set. In this study, transfer learning was used to initialize the weights of the YOLOv7 model before training it on the benthic foraminifera dataset. This enabled the model to achieve high accuracy despite the limited number of labeled data available.















































Species	Count	Image	Species	Count	Image	Species	Count	Image	Species	Count	Image	Species	Count	Image
<i>Bulinina marginata</i>	2325		<i>Nonionella labradorica</i>	664		Inner Organic Lining (IOL)	169		<i>Elphidium albumbilicatum</i>	21		<i>Milliotinella subrotunda</i>	8	
<i>Textularia earlandi</i>	2105		<i>Cibicides lobatulus</i>	549		<i>Elphidium magellanicum</i>	134		<i>Cassidulina neoteretis</i>	15		<i>Trifarina angulosa</i>	7	
<i>Stainforthia fusiformis</i>	1737		<i>Reophax</i> sp.	396		<i>Eggerelloides medius</i>	97		<i>Cribrostomoides jeffreysii</i>	15		<i>Trochammina rotuliformis</i>	7	
<i>Hyalinea balthica</i>	1719		<i>Spiroplectamina biformis</i>	341		<i>Bathysiphon acuta</i>	87		<i>Textularia bocki</i>	15		<i>Milliammina fusca</i>	6	
<i>Eggerelloides scaber</i>	1590		<i>Globobulimina</i> sp.	335		<i>Pyrgo williamsoni</i>	73		<i>Elphidium williamsoni</i>	12		<i>Psammospaera bowmanni</i>	6	
<i>Nonionella</i> sp. T1	1507		<i>Brizallina skagerrakensis</i>	307		<i>Quinqueloculina seminula</i>	72		<i>Elphidium macellum</i>	11		<i>Hormosinella gracilis</i>	4	
<i>Elphidium excavatum</i>	1343		<i>Nonionella turgida</i>	299		<i>Epistominella vitrea</i>	68		<i>Pullenia osloensis</i>	10		<i>Lagena laevis</i>	4	
<i>Ammonia</i> sp.	1279		<i>Uvigerina peregrina</i>	215		<i>Cornuspira foliacea</i>	37		<i>Elphidium incertum</i>	9		<i>Recurvoides trochammini-forme</i>	4	
<i>Bolivina pseudopunctata</i>	1221		<i>Liebusella goesi</i>	211		<i>Mellonis barleeianum</i>	37		<i>Glandulina laevigata</i>	9		<i>Lagena mollis</i>	2	
<i>Cassidulina laevigata</i>	804		<i>Leptohalysis catella</i>	203		<i>Haplophragmoides bradyi</i>	31		<i>Nonionella iridea</i>	9		<i>Epistominella exigua</i>	1	
<i>Ammodiscus</i> sp.	774		<i>Quinqueloculina stalkerii</i>	187		<i>Cribrostomoides</i> sp.	26		<i>Oolina hexagona</i>	9				
<i>Adercotryma glomerata</i>	725		<i>Ammoscalaria pseudospiralis</i>	169		<i>Guttulina lactea</i>	23		<i>Lagena striata</i>	8				

Figure 4. List of all benthic foraminifer species imaged in this study, with total counts and representative images. Of the 50 total species imaged, 29 were selected for training the model (those with ≥ 72 images; *Quinqueloculina seminula* and above).

2.6 Model training

For this study, three different models from the YOLOv7 family were trained: YOLOv7, YOLOv7x, and YOLOv7E6E (Fig. 2i). These models differ in their architecture, the num-

ber of layers, and the size of the input images they can process. YOLOv7 and YOLOv7x use 640×640 pixel images as input, while YOLOv7E6E uses 1280×1280 pixel images, which allows it to detect smaller objects in greater detail. All models were trained on a workstation equipped with an Intel

Table 1. A comparison of YOLOv7 models' performance on the COCO dataset (Lin et al., 2014) prior to the training with foraminifera images.

Model	Type	Test Size	APtest ¹	AP50test	AP75test	Batch 1 FPS ²	Batch 32 average time (milliseconds)
YOLOv7 P5	P5	640	51.40 %	69.70 %	55.90 %	161	2.8
YOLOv7-X	P5	640	53.10 %	71.20 %	57.80 %	114	4.3
YOLOv7-W6	P6	1280	54.90 %	72.60 %	60.10 %	84	7.6
YOLOv7-E6	P6	1280	56.00 %	73.50 %	61.20 %	56	12.3
YOLOv7-D6	P6	1280	56.60 %	74.00 %	61.80 %	44	15.0
YOLOv7-E6E	P6	1280	56.80 %	74.40 %	62.10 %	36	18.7

¹ AP: Average precision. ² FPS: Frames per second, a measure of the frequency (rate) at which consecutive images are captured or displayed.

i7 9700K processor (8 cores, 8 threads at 3.60 GHz), 32 GB of RAM, and a Nvidia RTX A4000 graphics card running Ubuntu 22.04. The models were trained using the PyTorch deep learning framework (version 2.0.0) with CUDA version 11.7 for GPU acceleration.

The training process consisted of 350 epochs for each model (one epoch = one complete pass of the entire training dataset through the CNN). The smallest model (YOLOv7) took approximately 18 h to train, the medium-sized model (YOLOv7x) took 25 h, and the largest model (YOLOv7E6E) required 120 h due to its increased complexity and higher input resolution. During training, the model weights were initialized using pre-trained weights from the COCO dataset and fine-tuned through subsequent training sessions.

2.7 Object detection metrics

In object detection tasks, model performance is typically evaluated using several key metrics, including precision (P), recall (R), and Intersection over Union (IoU). Precision measures the proportion of correctly identified objects (true positives) out of all positively identified objects, while recall measures the proportion of true positives out of all actual objects in the image, i.e.

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (1)$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (2)$$

where TP = true positives, FP = false positives, and FN = false negatives. IoU is used to quantify the overlap between the predicted bounding box and the ground truth bounding box, with higher IoU values indicating better accuracy at a given threshold value (Fig. S1 in the Supplement).

These metrics can be combined to compute more comprehensive performance indicators such as average precision (AP) and mean average precision (mAP). AP is the area under the precision–recall (PR) curve, which plots precision on the y axis and recall on the x axis for different confidence thresholds. In turn, mAP is calculated as the average of the

AP values over all classes (i.e., species) and is commonly used to evaluate the overall performance of object detection models (Padilla et al., 2021).

For this study, the performance of the trained YOLOv7 models was evaluated using mAP at two different IoU thresholds: 0.5 (mAP@0.5) and 0.5 to 0.95 in steps of 0.05 (mAP@0.5 : 0.95). The threshold of 0.5 is commonly used as a standard measure of detection accuracy, while the 0.5 : 0.95 range provides a more stringent evaluation by considering a range of IoU values.

The F_1 score was also used to evaluate model performance. The F_1 score combines precision and recall into a single metric and is particularly useful for evaluating models with unbalanced datasets. It is the harmonic mean of precision and recall and is defined as

$$F_1 = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}. \quad (3)$$

In this study, the confidence threshold was adjusted to maximize the F_1 score, which maximizes precision and recall simultaneously.

3 Results

3.1 Model training and performance overview

Over the course of the study, 58 training experiments were conducted using different configurations of the YOLOv7 architecture (Table S1 in the Supplement). Out of these, 23 models were successfully trained, while the remaining 35 failed due to memory limitations. We used the default YOLOv7 experiment suffix, “exp”, for the name of each experiment. Models exp 10 and exp 50 used a 16-image batch size, while exp 56 and 58 used an 8-image batch size. The E6E models exp 51, 52, and 53 used only a 2-image batch size. These batch size differences result from the amount of space taken up in GPU memory by the different model types, with the E6E model being the largest. The best-performing models are summarized in Table 2. Out of these, the best-performing model was exp 50, which achieved a precision of

0.877, a recall of 0.882, and a mean average precision (mAP) of 90.3 % at an IoU threshold of 0.5 and 69.9 % at a threshold range of 0.5 to 0.95. The model was based on the YOLOv7x architecture using 640×640 pixel images as input.

3.2 Training curves

The performance of the models was monitored throughout the training process using loss curves, which provide insight into model performance across training epochs. Figure 5 shows examples for the training curves for exp 50, which include three types of loss: box loss, objectness loss, and classification loss.

Following the definitions of Alexe et al. (2012):

- Box loss measures how accurately the model can predict the location and size of bounding boxes around detected objects. The lower the box loss, the higher the model accuracy generally, but the model must also correctly classify objects and avoid false positives and negatives.
- Objectness loss represents the probability that an object exists within a given region of interest. If objectness is high, it indicates that the model exists within a given region, while a low objectness score indicates that the region is background or that it does not contain any object of interest.
- Classification loss measures how well the model can classify the detected objects into the correct species category. The lower the classification loss, the higher the model accuracy generally, but the model must also optimize objectness and box loss to achieve high accuracy in all respective categories.

During training, the loss values decreased steadily over time, indicating that the model was learning to better detect and classify foraminifera species (Fig. 5). For classification, the validation loss closely followed the training loss, suggesting that the model did not suffer from significant overfitting.

3.3 Confusion matrix

The confusion matrix (Fig. 6) provides a detailed view of the model's classification performance by showing the number of correct and incorrect predictions for each species on the test set of images. Figure 6 displays the confusion matrix for the best-performing model, exp 50.

The confusion matrix reveals that the model performed well across most species, with most predictions falling along the diagonal (correct classifications). However, there were some instances where the model confused species with similar morphology. For example, species in the genus *Nonionella* and *Eggerelloides* were occasionally misclassified, reflecting the challenge of distinguishing between species within the same genus and with subtle morphological differences.

The matrix also shows that the model had a small number of false positives and false negatives for the background (1 %–13 %), indicating that it sometimes incorrectly identified sediment particles or other debris as foraminifera (Fig. 6).

3.4 Object detection metrics

Figure S2 shows the precision and recall results for the best-performing model, exp50. At a confidence threshold of 0.540, the model achieved its highest F_1 score (Fig. S2a), indicating a balanced trade-off between precision and recall. This threshold was used in subsequent evaluations of the model's performance. The precision–recall (PR) curve (Fig. S2b) provides a visualization of the model's precision and recall across different confidence thresholds. A high area under the PR curve indicates strong performance, with both high precision (low false positive rate) and high recall (low false negative rate).

The precision and recall curves (Fig. S2c, d) provide additional insights into the model's behavior during training. Typically, as the confidence threshold increases, precision increases while recall decreases. The precision curve shows that precision increases as the confidence threshold increases, indicating fewer false positives at higher confidence levels.

The recall curve shows that recall decreases as the confidence threshold increases, indicating fewer false negatives at lower confidence levels. The model provided output images (Fig. 7), where the input image of benthic foraminifera is processed by the model, resulting in an output image with bounding boxes and species labels. The confidence level of the model predictions is displayed next to each species name (Fig. 7).

4 Discussion

4.1 Performance of the YOLOv7 and comparison with existing models

While YOLO family models have been used on other microfossils including ichthyoliths (Mimura et al., 2024), fish scales (Hanson et al., 2024), and pollen (Kubera et al., 2022; Endo et al., 2024; Jofre et al., 2025), to the best of our knowledge this study is one of the first to apply the YOLO architecture to the identification of foraminifera as of time of writing and is one of the very few models used on benthic foraminifera specifically to date. The results from this study demonstrate that deep learning, specifically the YOLOv7 family of models, can be effectively applied to the task of benthic foraminifera identification from the diverse species communities such as in the Skagerrak fjords. Among the models tested, YOLOv7x (exp 50) performed the best, achieving a mean average precision (mAP) of 90.3 % at an IoU threshold of 0.5 % and 69.9 % at the stricter 0.5 to 0.95 range. These results are competitive with, and in some cases

Table 2. Performance comparison of trained models.

Model	Precision	Recall	mAP ¹ @0.5	mAP@0.5–0.95	F ₁ score	Training time	Model architecture	Batch size
exp 10	0.868	0.855	0.889	0.696	0.861	17 h 57 m 41 s	YOLOv7@640	16
exp 50 ²	0.877	0.882	0.903	0.699	0.879	24 h 49 m 37 s	YOLOv7x@640	16
exp 51	0.797	0.849	0.867	0.661	0.822	90 h 11 m 40 s	YOLOv7E6E@640	2
exp 52	0.862	0.863	0.895	0.690	0.862	118 h 44 m 27 s	YOLOv7E6E@1280	2
exp 53	0.833	0.853	0.870	0.680	0.843	89 h 52 m 30 s	YOLOv7E6E@640	2
exp 56	0.845	0.874	0.887	0.695	0.859	25 h 53 m 23 s	YOLOv7x@640	8
exp 58	0.827	0.883	0.888	0.691	0.854	18 h 43 m 29 s	YOLOv7@640	8

¹ mAP: mean average precision. ² Best-performing model

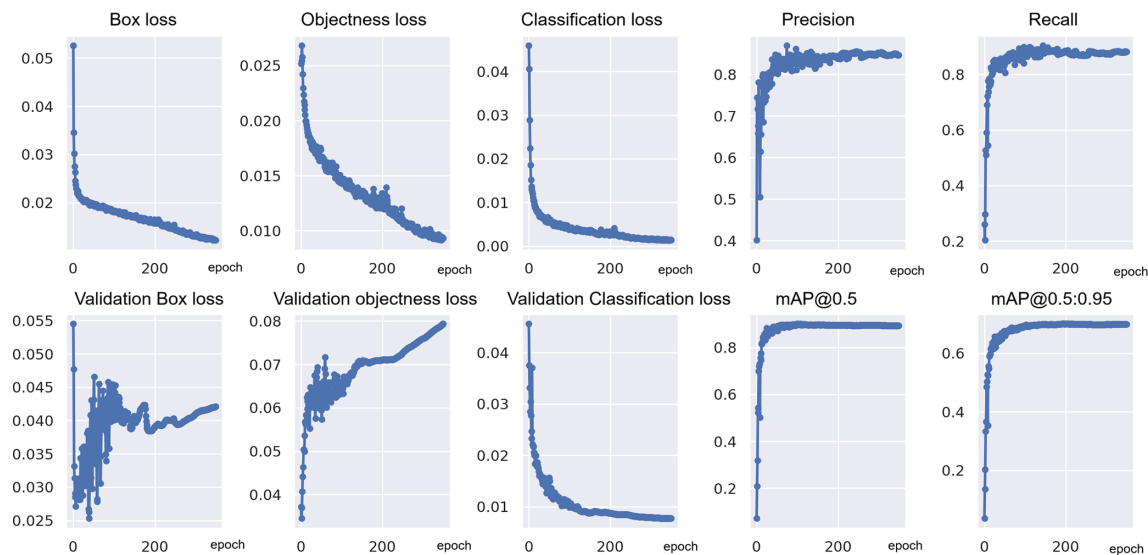


Figure 5. Training and validation curves for the best-performing model, exp 50, showing box loss, objectness loss, classification loss, precision, recall, and mean average precision (mAP) (corresponding y axes) across training epochs (x axes).

exceed, the performance of other machine learning models that have been used for foraminifera identification (Table 3). For instance, Hsiang et al. (2019) employed a VGG16 CNN to classify 35 planktonic foraminiferal species, achieving a precision of 87.4 %. Their study led to the creation of the *Endless Forams* dataset, which included over 34 000 individual foraminifera. This dataset was then used by Karaderi et al. (2022) to train a deep metric learning model which achieved an accuracy of 92 % at an F_1 score of 0.84. Although these studies focused on planktonic species, the performance of the VGG16 and deep metric learning models is comparable to the YOLOv7x model used in this work, which both achieved a precision of 87.7 % and an mAP@0.5 of 90.3 %, respectively.

Similarly, Marchant et al. (2020) used a customized ResNet50 model on the *Endless Forams* dataset, achieving a precision of 90.3 %. This model was also trained on both planktonic and benthic foraminifera, with a slightly lower precision of 89 % for the benthic species. While the precision achieved by Marchant and colleagues is higher than

the YOLOv7x model on some species, the key advantage of YOLO lies in its ability to perform object detection in addition to classification. This makes YOLO suitable for scenarios where foraminifera need to be identified within unpicked sediment samples that contain non-foraminifera objects, which would typically require manual sorting before classification to reduce task complexity. In another study, Mitra et al. (2019) combined ResNet50 and VGG16 models to classify planktonic foraminifera, achieving a precision of 80 %. The lower precision in their study can be attributed to the smaller dataset of 1437 individuals, highlighting the importance of dataset size for model performance. In contrast, the dataset used for the YOLOv7x model in this study was augmented to include over 7000 images, leading to improved performance. Johansen and Sørensen (2020) implemented a VGG16 model for the classification of both sediment particles and benthic foraminifera, achieving a 98.5 % precision for this simpler task of detecting and differentiating between broad categories. However, the focus of their model was not species-level classification, and its application is lim-

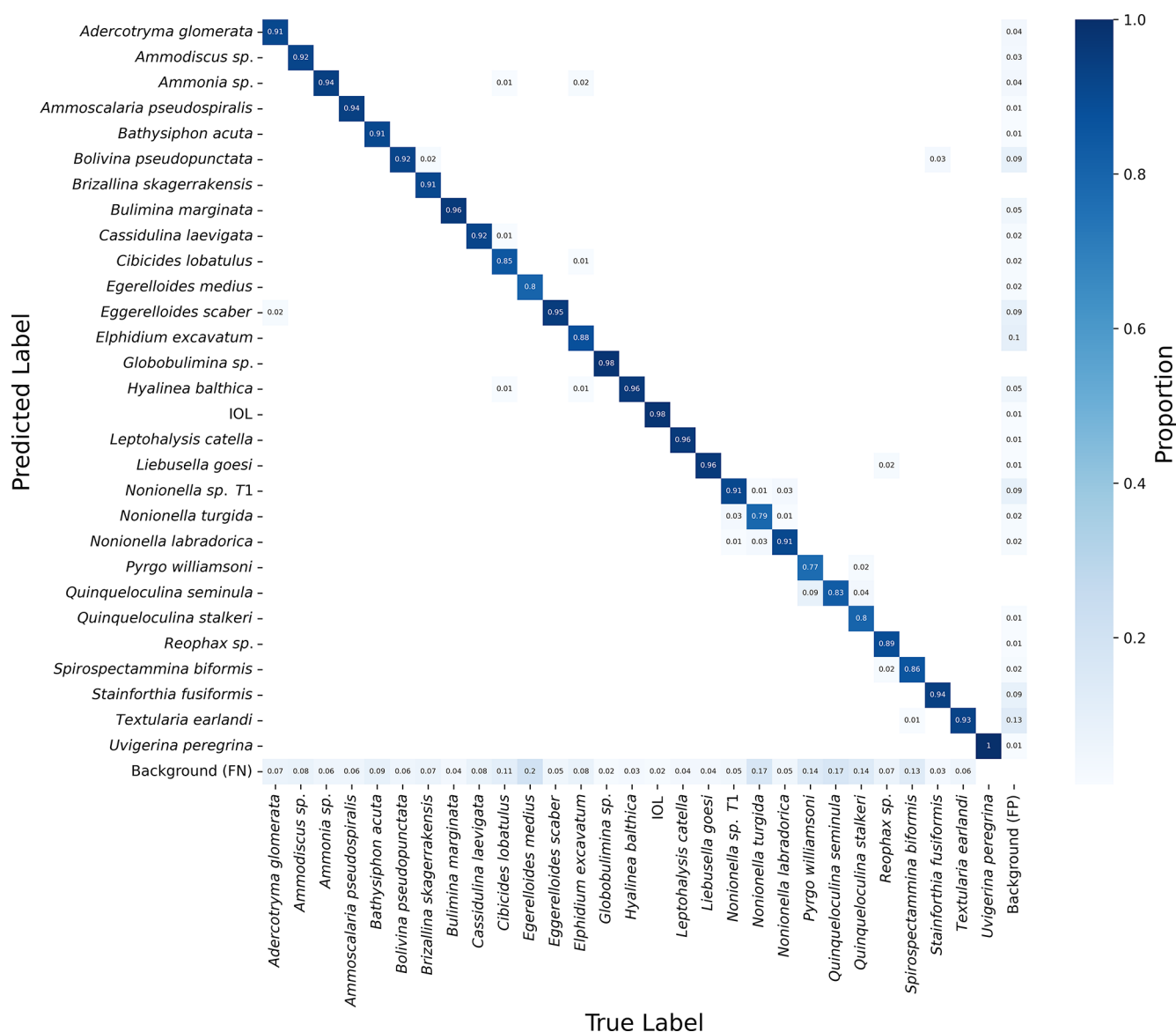


Figure 6. Confusion matrix of the best-performing model, exp 50. On the left-hand side are the predicted labels made by the model, and on the bottom of the matrix are the ground truth labels. The shade of blue indicates the proportion of images in the test set that the model correctly identified as the given species (only values > 0 are shown). The model has false negatives of background in all species, while mistaking only a small percentage of one foraminifera species with other foraminifera species. IOL: inner organic lining; FN: false negative; FP: false positive.

ited to basic detection tasks, whereas the YOLOv7x model presented in this study demonstrates significant capabilities for species-level identification.

One key factor contributing to the success of the YOLOv7x model is its ability to balance detection accuracy with speed. YOLO models, in general, are designed to process images in a single forward pass, making them exceptionally fast compared to other object detection algorithms. For example, Wang et al. (2023) show that the base YOLOv7 model is more than 6 times faster than the fastest non-YOLO model tested (Deformable DETR), with an FPS

of 118 vs. 19 on a V100 GPU. This makes them ideal for real-time applications and for processing large datasets efficiently. The inclusion of YOLOv7E6E, which accepts larger input images (1280 × 1280 pixels), showed that increasing image resolution improved detection in some cases but at a significant computational cost. As a result, the YOLOv7x model has shown the most promise for use in real time applications, since its high inference speed allows it to be used directly during foraminifera picking (for instance, allowing non-experts to pick specimens when training students or future applications using robotics) owing to its fast

Table 3. Overview of deep learning classification models applied to foraminifera discussed.

Study	Year	Group	Task	Model type	# Species/classes	Dataset size	Performance
Mitra et al.	2019	Planktonic	Classification (species-level)	ResNet50 + VGG16 CNN	6	1437	80 % (precision), 0.81 (F_1 score)
Hsiang et al.	2019	Planktonic	Classification (species-level)	VGG16 CNN	36	34 640	87.4 % (accuracy)
Johansen and Sørensen	2020	Planktonic + benthic + sediment	Classification (group-level)	VGG16 CNN	4	2673	98.5 % (accuracy)
Marchant et al.	2020	Planktonic + benthic	Classification (species-level)	Base-Cyclic CNN	36 (planktonic), 9 (benthic)	27 729 (planktonic), 15 274 (benthic)	90.3 % (accuracy, planktonic), 89 % (accuracy, benthic)
Karaderi et al.	2022	Planktonic	Classification (species-level)	Metric learning, ResNet50 backbone	36	34 640	92 % (accuracy), 0.84 (F_1 score)
This study	2025	Benthic	Localization + classification (species-level)	YOLOv7	29	49 459 (augmented from 22 138)	87.7 % (precision), 90.3 % (mAP@0.5)

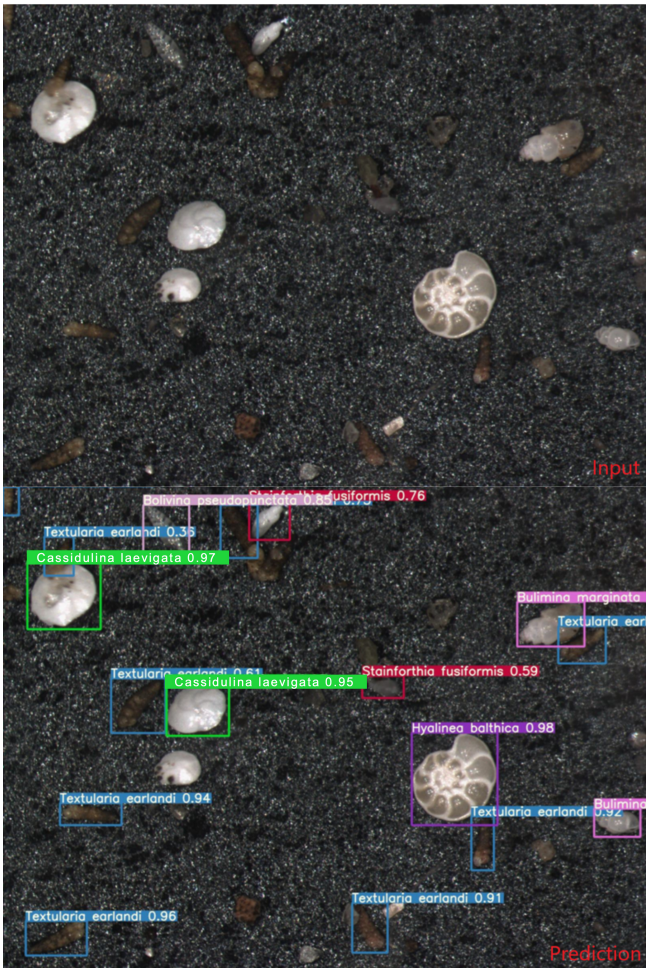


Figure 7. Input image of foraminifera (top figure) is processed by a model and creates a new image (bottom figure) with bounding boxes and species name. The number next to the species is the confidence level of the model between 0–1.

frame-per-second performance (≈ 65 FPS). However, the integrated nature of the YOLO workflow (i.e., the identification and segmentation steps are combined into a single, one-step model) does present some disadvantages in comparison to two-step workflows (i.e., a segmentation model is trained first to localize objects, followed by a separate classification model to perform identification) (not to be confused with one- vs. multi-stage segmentation models, e.g., YOLO vs. Mask RCNN). The latter has been shown to perform advantageously in problems varying from skin lesion identification (Shakya et al., 2025) to the identification of diatoms in sediment traps (Godbillot et al., 2025) and allows a separate evaluation of problems in segmentation and classification. The trade-off between speed/computational cost (in which one-step models are advantaged) and accuracy (in which two-step models are advantaged) must be balanced when determining the most suitable model type for a particular task. For instance, it would likely be preferable to use YOLO-type

models for rapid, in situ, real-time biomonitoring applications. On the other hand, tasks that involve samples that must be substantially pre-processed (e.g., core-top sediment samples) – resulting in the analysis bottleneck being centered on sample preparation and imaging – might be better served by higher-accuracy two-step models. We note that our results show that one of the classical drawbacks of YOLO models – poor performance with small objects (Tariq and Javed, 2025) – does not appear to significantly impact accuracy in this case.

4.2 Species-level performance

While the overall performance of the YOLOv7x model was strong, there were notable variations in its species-level performance. For common species such as *Nonionella labradorica* and *Bulimina marginata*, precision and recall values exceeded 90 %, indicating the model's robustness in detecting and classifying species with abundant training data. However, for less abundant species like *Eggerelloides medius*, performance was notably lower, with a precision of 55.3 %. This discrepancy can largely be attributed to the limited number of training examples for these less abundant species, which is a common challenge in supervised learning models (Table S2).

In the case of *Eggerelloides medius*, the confusion with morphologically similar species, such as *Eggerelloides scaber*, highlights one of the primary limitations of the current model: its limited ability to differentiate between closely related species. This issue is exacerbated by the presence of subtle morphological differences that may not be sufficiently captured in lower-resolution images or that require more nuanced feature extraction techniques. Future work could address this limitation by incorporating higher-resolution images and additional morphometric data or by using hybrid models that integrate both image and textural data. Our data additionally show that the model struggled with certain species, particularly agglutinated foraminifera like *Eggerelloides medius* and *Textularia earlandi*. These species are characterized by their relatively simple yet irregular shell structures, which may have been more difficult for the model to distinguish against the dark background of microslides.

Generally, it appears that common denominators of poor performance of the model are:

- a. agglutinated species, which the model has difficulty distinguishing against the black background;
- b. large and voluminous (e.g., globular to spherically shaped) species, which require multiple focal points to distinguish details needed for accurate identification; and
- c. species with a low number of individuals in the training set (fewer than 100 individuals).

Another challenge observed in the confusion matrix of this study was the tendency for background elements to be misclassified as foraminifera in some instances, contributing to a false positive rate of 1 % to 13 % across different species. While this issue was relatively minor in comparison to the overall performance, it underscores the importance of careful image pre-processing and the potential value of incorporating background filtering algorithms or the use of different background colors to further enhance the model's accuracy.

4.3 Applications and implications

One of the key advantages of the YOLOv7 architecture is its ability to process images in real time, which opens new possibilities for automated species identification in the field. For example, researchers could deploy YOLO-based models on applications for mobile devices or desktop computers to identify foraminifera in real time during fieldwork or laboratory analysis. This would significantly reduce the time and expertise required for species identification, making it accessible to a broader range of users.

Additionally, the ability to identify foraminifera directly from unpicked sediment samples represents a major advancement in foraminiferal studies. Current methods for analyzing foraminifera require time-consuming manual sorting and picking of individual specimens, which limits the scalability of environmental monitoring programs. By applying deep learning models to sieved sediment samples that have yet to be picked for foraminifera, researchers can streamline the analysis process and obtain results more quickly, allowing more frequent and widespread monitoring of marine ecosystems.

4.4 Limitations and future directions

While this study demonstrates the effectiveness of deep learning for foraminifera identification, there are several limitations that should be addressed in future work. One limitation is the relatively small size of the dataset, particularly for certain species. As noted earlier, species with fewer than 100 labeled individuals tended to perform worse in the model, likely due to the lack of sufficient training data. Future studies should aim to increase the size and diversity of the dataset, either by collecting more images through taking multiple images of the same individual from multiple angles or by augmenting the existing data through additional augmentation techniques. The model results presented here also point the way for future targeted sampling initiatives, as species with lower identification accuracies are clearly demarcated.

Another limitation is the use of a single imaging setup, which could introduce bias into the model. All images in this study were captured under the same lighting and magnification conditions, which may limit the model's ability to generalize to images taken under different conditions. In future studies, it would be beneficial to test the model on images

collected using different microscopes, cameras, or lighting setups to evaluate the model's robustness to varying conditions. Additionally, the use of different background colors for microscopic slides could prove beneficial for distinguishing boundaries between foraminifera specimens and the background. One potential background color could be chroma key green, which has shown considerable success in the movie industry for decades in the visual effects field. Further, the use of *z*-stacking, which would provide higher-resolution images by incorporating multiple focal planes, would allow a deep learning model to learn to recognize species-specific features that are not visible or resolved in a single 2D image. This would be particularly useful for large individuals that exceed a given camera's depth of focus. While *z*-stacking is a standard component of many microfossil imaging pipelines (e.g., Elder et al., 2018; Hsiang et al., 2019; Marchant et al., 2020; Adebayo et al., 2023), incorporating *z*-stacking necessitates a corresponding increase in imaging time and file sizes (proportional to the number of slices taken). Furthermore, additional hardware and software are required, particularly for large imaging initiatives (e.g., automated microscope stages), which may be a limiting factor depending on funding and accessibility. The fact that we achieve high accuracies in this study without using *z*-stacking is notable, and it suggests that useful training data can be collected for species identification even when the imaging setup is not optimized for image resolution and quality.

While this study represents an important step forward in the use of automated segmentation models for rapid identification of benthic foraminifera, there remain many avenues for future work as we endeavor to develop such models for application in biomonitoring. These directions for future work include improving the model architecture, improving the training dataset (including incorporating metadata to enhance model accuracy), and expanding model deployment. The YOLO model used in this study, YOLOv7, is only one in a family of models that enjoys good support from developers and continued research interest; since the experimental part of this study was completed, as of the time of writing, there have been four new updates to the architecture, with the current version being YOLOv11 (Jocher and Qiu, 2024). This new model shows further improvements in object detection and inference time compared to its predecessors (Khanam and Hussain, 2024). In addition to using newer model architectures, future work could explore the use of ensemble learning techniques, where multiple models are combined to improve overall classification accuracy. By integrating the strengths of different architectures – such as YOLO models for object detection and ResNet or VGG models for fine-grained classification – it may be possible to achieve even higher precision and recall, particularly for challenging cases involving similar species.

While using alternate model architectures would likely improve accuracies, in certain cases one might be limited in one's model choice by hardware availability; for example,

in this study, many of our experiment setups were not feasible due to the lack of sufficient GPU memory to implement larger models (e.g., YOLOv7E6E; Table S1). As such, other ways of improving model performance should also be considered. One of the most important of these which is often overlooked is the quality of the actual training data. In this case we are referring not to the image quality (as in the discussion on *z*-stacking above) but to the quality of the annotations and labeling. To our knowledge, a systematic review of annotation practices for building species identification training sets does not yet exist (and is outside the scope of this study). However, it is our general observation that most studies training models for species identification, including this one, use training datasets that have been labeled/annotated by a single worker or equivalent (i.e., multiple workers that label non-overlapping image sets). This is often done to save time on annotation, which is a time-consuming manual process. However, having only a single labeler means that annotation mistakes are difficult to catch and are likely to be propagated through to training, leading to systematic biases. Hsiang et al. (2019) prioritized training set quality for the *Endless Forams* dataset by having all images labeled by four independent experts and only retaining images for which there was $\geq 75\%$ agreement on the species identity for the final training dataset. While such annotation procedures would be preferable, there is a significant trade-off in time spent annotating unless a large team can be assembled to label (as in the case of the *Endless Forams* study). This type of annotation strategy also requires agreement on a community taxonomy to which all labelers agree to adhere. While the current study serves as a promising proof of concept, initiatives moving forward must carefully consider their labeling procedures in order to minimize systematic biases in their training sets.

Another data-based approach to improving model performance is the incorporation of metadata together with image data in fused networks. Previous work has shown that data fusion of this kind, where camera trap images are combined with metadata such as location, improves model accuracy (Tøn et al., 2024). The integration of molecular data, such as species barcoding and metabarcoding (eDNA) data, could also enhance the accuracy of species identification models. Recent (non-deep-learning) studies have shown that combining morphological and genetic data can significantly improve the resolution of species identification in foraminifera, particularly for cryptic or closely related species (e.g., Darling et al., 2016; Bird et al., 2020; Richirt et al., 2021). Future studies should explore the potential of combining deep learning models with molecular techniques to develop more comprehensive tools for environmental monitoring. Additionally, since the detection of dead vs. living material is important for bioindicator applications, an important avenue for future work is the incorporation of stained (e.g., rose bengal) and/or labeled (e.g., CellTracker Green) individuals in the training dataset.

5 Conclusions

This study demonstrates the effectiveness of using deep learning models, specifically YOLOv7x, for the automated identification and classification of benthic foraminifera from the Skagerrak Fjords. The YOLOv7x model achieved high performance metrics, with a precision of 87.7 %, a recall of 88.2 %, and a mean average precision (mAP@0.5) of 90.3 %. These results suggest that deep learning models can significantly reduce the time and effort required for foraminiferal identification, offering a faster and more scalable solution compared to manual methods.

The two most important contributions of this study are as follows:

1. *Automation of species-level classification.* This study is one of the first to apply YOLOv7 to benthic foraminiferal classification, successfully automating the identification process while maintaining accuracy comparable to expert manual identification. By leveraging deep learning, it is possible to process thousands of images in minutes, a task that would typically take days or weeks with manual analysis.
2. *Comparison to existing models.* When compared to other models, such as VGG16 and ResNet50, which have been used in earlier studies, the YOLOv7x model presented here achieves competitive precision while offering the additional benefit of object detection, a critical feature for handling complex, unpicked samples.

In conclusion, the application of deep learning, specifically the YOLOv7x model, represents a significant advancement in the field of foraminiferal research. This study illustrates the potential of deep learning to automate and accelerate the process of species identification, making it more accessible for large-scale environmental monitoring. While challenges remain, particularly regarding underrepresented species, the findings presented here offer a strong foundation for future research and model improvement.

Lastly, this study does not aim to eliminate human expertise from the taxonomic identification process. Instead, an ML-based approach could be used as a labor-saving device to go through the bulk of the dataset, and later a person could validate and, if needed, correct the identifications. By reducing the time needed to identify foraminifera, one could focus more on the analysis of ecological interactions between the species.

Code availability. The dataset and code are available at Zenodo: <https://doi.org/10.5281/zenodo.17608881> (Plavetić et al., 2025)

Data availability. We have uploaded all the data to Zenodo, in both YOLOv7-PyTorch and COCO format. The associated DOI is <https://doi.org/10.5281/zenodo.17608881> (Plavetić et al., 2025).

Supplement. The supplement related to this article is available online at <https://doi.org/10.5194/jm-44-693-2025-supplement>.

Author contributions. MP developed an idea and designed the project with the help of MJ, AYH, and IPA. MP carried out model development, analyzed the data, and had a leading role in article writing. MJ, GH, and IPA contributed with resources and material. MJ, AYH, and IPA supervised the work. MP and AYH invested an equal amount of work in article writing and editing. All co-authors contributed to the article writing.

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. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. 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. The authors sincerely thank everybody who helped to facilitate this study. Crew and captains of the old and new R/V *Skagerak*, Kjell Nordberg and Lennart Bornmalm, assisted with sediment sampling. Katarina Abrahamsson organized the SEEPS II expedition and facilitated the participation of MP, MJ, GH, and IPA in the expedition. Students of the Departments of Earth and Marine Sciences (University of Gothenburg) prepared and picked foraminiferal samples.

Financial support. The publication of this article was funded by the Swedish Research Council, Forte, Formas, and Vinnova.

Review statement. This paper was edited by Christopher Smart and reviewed by Constance Choquel and two anonymous referees.

References

- Adebayo, M. B., Bolton, C. T., Marchant, R., Bassinot, F., Conrod, S., and de Garidel-Thoron, T.: Environmental controls of size distribution of modern planktonic foraminifera in the tropical Indian Ocean, *Geochem Geophys Geosyst*, 24, e2022GC010586, <https://doi.org/10.1029/2022GC010586>, 2023.
- Alexe, B., Deselaers, T., and Ferrari, V.: Measuring the Objectness of Image Windows, *IEEE T. Pattern Anal.*, 34, 2189–2202, <https://doi.org/10.1109/TPAMI.2012.28>, 2012.
- Alve, E.: Benthic foraminifera in sediment cores reflecting heavy metal pollution in Sør fjord, western Norway, *J. Foramin. Res.*, 21, 1–19, <https://doi.org/10.2113/gsjfr.21.1.1>, 1991.
- Alve, E.: Benthic foraminiferal responses to estuarine pollution, *J. Foramin. Res.*, 25, 190–203, <https://doi.org/10.2113/gsjfr.25.3.190>, 1995.
- Alve, E.: A common opportunistic foraminiferal species as an indicator of rapidly changing conditions in a range of environments, *Estuar. Coast. Shelf S.*, 57, 501–514, [https://doi.org/10.1016/S0272-7714\(02\)00383-9](https://doi.org/10.1016/S0272-7714(02)00383-9), 2003.
- Alve, E., Lepland, A., Magnusson, J., and Backer-Owe, K.: Monitoring strategies for re-establishment of ecological reference conditions: Possibilities and limitations, *Mar. Pollut. Bull.*, 59, 297–310, <https://doi.org/10.1016/j.marpolbul.2009.08.011>, 2009.
- Alve, E., Hess, S., Bouchet, V. M., Dolven, J. K., and Rygg, B.: Intercalibration of benthic foraminiferal and macrofaunal biotic indices: An example from the Norwegian Skagerrak coast (NE North Sea), *Ecol. Indic.*, 96, 107–115, <https://doi.org/10.1016/j.ecolind.2018.08.037>, 2019.
- 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.
- Austen, G. E., Bindemann, M., Griffiths, R. A., and Roberts, D. L.: Species identification by conservation practitioners using online images: accuracy and agreement between experts, *PeerJ*, 6, e4157, <https://doi.org/10.7717/peerj.4157>, 2018.
- Bernhard, J. M. and Alve, E.: Survival, ATP pool, and ultrastructural characterization of benthic foraminifera from Drømmensfjord (Norway): response to anoxia, *Mar. Micropaleontol.*, 28, 5–17, [https://doi.org/10.1016/0377-8398\(95\)00036-4](https://doi.org/10.1016/0377-8398(95)00036-4), 1996.
- Bird, C., Schweizer, M., Roberts, A., Austin, W. E., Knudsen, K. L., Evans, K. M., and Darling, K. F.: The genetic diversity, morphology, biogeography, and taxonomic designations of *Ammonia* (Foraminifera) in the Northeast Atlantic, *Mar. Micropaleontol.*, 155, 101726, <https://doi.org/10.1016/j.marmicro.2019.02.001>, 2020.
- Boehnert, S., Birkelund, A. R., Schmiedl, G., Kuhnert, H., Kuhn, G., Hass, H. C., and Hebbeln, D.: Test deformation and chemistry of foraminifera as response to anthropogenic heavy metal input, *Mar. Pollut. Bull.*, 155, 111112, <https://doi.org/10.1016/j.marpolbul.2020.111112>, 2020.
- Bouchet, V. M., Alve, E., Rygg, B., and Telford, R. J.: Benthic foraminifera provide a promising tool for ecological quality assessment of marine waters, *Ecol. Indic.*, 23, 66–75, <https://doi.org/10.1016/j.ecolind.2012.03.011>, 2012.
- Brinkmann, I., Schweizer, M., Singer, D., Quinchar, S., Barras, C., Bernhard, J. M., and Filipsson, H. L.: Through the eDNA looking glass: Responses of fjord benthic foraminiferal communities to contrasting environmental conditions, *J. Eukaryot. Microbiol.*, 70, e12975, <https://doi.org/10.1111/jeu.12975>, 2023.
- Brunner, C. A., Yeager, K. M., Hatch, R., Simpson, S., Keim, J., Briggs, K. B., and Louchouart, P.: Effects of oil from the 2010 Macondo well blowout on marsh foraminifera of Mississippi and Louisiana, USA, *Environ. Sci. Technol.*, 47, 9115–9123, <https://doi.org/10.1021/es401943y>, 2013.
- Castro, W., Oblitas, J., Santa-Cruz, R., and Avila-George, H.: Multilayer perceptron architecture optimization using parallel computing techniques, *PLoS ONE*, <https://doi.org/10.1371/journal.pone.0189369>, 2017.
- Chen, W., Huang, H., Peng, S., Zhou, C., and Zhang, C.: YOLO-face: a real-time face detector, *Visual. Comput.*, 37, 805–813, <https://doi.org/10.1007/s00371-020-01831-7>, 2021.
- Choquel, C., Geslin, E., Metzger, E., Filipsson, H. L., Risgaard-Petersen, N., Launeau, P., Giraud, M., Jauffrais, T., Jesus, B., and Mouret, A.: Denitrification by benthic foraminifera and their contribution to N-loss from a fjord environment, *Biogeosciences*, 18, 327–341, <https://doi.org/10.5194/bg-18-327-2021>, 2021.
- Choquel, C., Mütter, D., Ni, S., Pirzamanbein, B., Charrieau, L. M., Hirose, K., Seto, Y., Schmiedl, G., and Filipsson, H. L.: 3D morphological variability in foraminifera unravel environmental changes in the Baltic Sea entrance over the last 200 years, *Front. Earth Sci.*, 11, 1120170, <https://doi.org/10.3389/feart.2023.1120170>, 2023.
- Darling, K. F., Schweizer, M., Knudsen, K. L., Evans, K. M., Bird, C., Roberts, A., and Austin, W. E.: The genetic diversity, phylogeography and morphology of *Elphidiidae* (Foraminifera) in the Northeast Atlantic, *Mar. Micropaleontol.*, 129, 1–23, <https://doi.org/10.1016/j.marmicro.2016.09.001>, 2016.
- Deldicq, N., Alve, E., Schweizer, M., Asteman, I. P., Hess, S., Darling, K., and Bouchet, V. M.: History of the introduction of a species resembling the benthic foraminifera *Nonionella stella* in the Oslofjord (Norway): morphological, molecular and paleo-ecological evidences, *Aquat. Invasions*, 14, 182–205, <https://doi.org/10.3391/ai.2019.14.2.03>, 2019.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L.: ImageNet: A large-scale hierarchical image database, 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 248–255, <https://doi.org/10.1109/cvpr.2009.5206848>, 2009.
- Dolven, J. K., Alve, E., Rygg, B., and Magnusson, J.: Defining past ecological status and in situ reference conditions using benthic foraminifera: a case study from the Oslofjord, Norway, *Ecol. Indic.*, 29, 219–233, <https://doi.org/10.1016/j.ecolind.2012.12.031>, 2013.
- Dumoulin, V. and Visin, F.: A guide to convolution arithmetic for deep learning, *arXiv [preprint]*, <https://doi.org/10.48550/ARXIV.1603.07285>, 2016.
- Elder, L. E., Hsiang, A. Y., Nelson, K., Strotz, L. C., Kahanamoku, S. S., and Hull, P. M.: Sixty-one thousand recent planktonic foraminifera from the Atlantic Ocean, *Sci. Data*, 5, 180109, <https://doi.org/10.1038/sdata.2018.109>, 2018.
- Endo, K., Hiraguri, T., Kimura, T., Shimizu, H., Shimada, T., Shibasaki, A., Suzuki, C., Fujinuma, R., and Takemura, Y.: Estimation of the amount of pear pollen based on flowering stage detection using deep learning, *Sci. Rep.*, 14, 13163, <https://doi.org/10.1038/s41598-024-63611-w>, 2024.

- Ernst, S. R., Morvan, J., Geslin, E., Le Bihan, A., and Jorissen, F. J.: Benthic foraminiferal response to experimentally induced Erika oil pollution, *Mar. Micropaleontol.*, 61, 76–93, <https://doi.org/10.1016/j.marmicro.2006.05.005>, 2006.
- 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., Rimmelzwaal, 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. Micropaleontol.*, 37, 431–443, <https://doi.org/10.5194/jm-37-431-2018>, 2018.
- Ferreira-Chacua, I. and Koeshidayatullah, A. I.: ForamiViT-GAN: Exploring new paradigms in deep learning for micropaleontological image analysis, *IEEE Access*, 11, 67298–67307, <https://doi.org/10.1109/ACCESS.2023.3291620>, 2023.
- Filipsson, H. L. and Nordberg, K.: Climate variations, an overlooked factor influencing the recent marine environment. An example from Gullmar Fjord, Sweden, illustrated by benthic foraminifera and hydrographic data, *Estuaries*, 27, 867–881, <https://doi.org/10.1007/BF02912048>, 2004.
- Frontalini, F. and Coccioni, R.: Benthic foraminifera for heavy metal pollution monitoring: a case study from the central Adriatic Sea coast of Italy, *Estuar. Coast. Shelf S.*, 76, 404–417, <https://doi.org/10.1016/j.ecss.2007.07.024>, 2008.
- Frontalini, F., Nardelli, M. P., Curzi, D., Martín-González, A., Sabbatini, A., Negri, A., and Bernhard, J. M.: Benthic foraminiferal ultrastructural alteration induced by heavy metals, *Mar. Micropaleontol.*, 138, 83–89, <https://doi.org/10.1016/j.marmicro.2017.10.009>, 2018.
- Frontalini, F., Cordier, T., Balassi, E., Du Châtelet, E. A., Cermakova, K., Apothéloz-Perret-Gentil, L., and Pawlowski, J.: Benthic foraminiferal metabarcoding and morphology-based assessment around three offshore gas platforms: Congruence and complementarity, *Environ. Int.*, 144, 106049, <https://doi.org/10.1016/j.envint.2020.106049>, 2020.
- Godbillot, C., Pesenti, B., Leblanc, K., Beaufort, L., Chevalier, C., Di Pane, J., de Madron, D., and de Garidel-Thoron, T.: Contrasting trends in phytoplankton diversity, size structure, and carbon burial efficiency in the Mediterranean Sea under shifting environmental conditions, *J. Geophys. Res.-Oceans*, 130, e2025JC022486, <https://doi.org/10.1029/2025JC022486>, 2025.
- Govindankutty Menon, A., Davis, C. V., Nürnberg, D., Nomaki, H., Salonen, I., Schmiedl, G., and Glock, N.: A deep-learning automated image recognition method for measuring pore patterns in closely related bolivinids and calibration for quantitative nitrate paleo-reconstructions, *Sci. Rep.*, 13, 19628, <https://doi.org/10.1038/s41598-023-46605-y>, 2023.
- Hanson, N. N., Ounsley, J. P., Henry, J., Terzić, K., and Caneco, B.: Automatic detection of fish scale circuli using deep learning, *Biology Methods and Protocols*, 9, bpae056, <https://doi.org/10.1093/biomethods/bpae056>, 2024.
- 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.
- He, K., Gkoxari, G., Dollár, P., and Girshick, R.: Mask R-CNN, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.1703.06870>, 2017.
- Hsiang, A. Y., Brombacher, A., Rillo, M. C., Mleneck-Vautravers, M. J., Conn, S., Lordsmith, S., Jentzen, A., Henehan, 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., Dowsett, H. J., Miller, C. G., Rios, N., 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.
- Jocher, G. and Qiu, J.: Ultralytics YOLO (Version 11.0.0) Github [code], <https://github.com/ultralytics/ultralytics> (last access: 12 October 2025), 2024.
- Jofre, R., Tapia, J., Troncoso, J., Staforelli, J., Sanhueza, I., Jara, A., Muchuca, G., Rondanelli-Reyes, M., Lamas, I., Godoy, S. E., and Coelho, P.: YOLOv8-based on-the-fly classifier system for pollen analysis of Guindo Santo (*Eucryphia glutinosa*) honey and assessment of its monoflorality, *J. Agr. Food Res.*, 19, 101665, <https://doi.org/10.1016/j.jafr.2025.101665>, 2025.
- Johansen, T. H. and Sørensen, S. A.: Towards detection and classification of microscopic foraminifera using transfer learning, *arXiv [preprint]*, <https://doi.org/10.48550/ARXIV.2001.04782>, 2020.
- Karaderi, T., Burghardt, T., Hsiang, A. Y., Ramaer, J., and Schmidt, D. N.: Visual microfossil identification via deep metric learning, in: *Pattern Recognition and Artificial Intelligence*, edited by: El Yacoubi, M., Granger, E., Yuen, P. C., Pal, U., and Vincent, N., ICPRAI 2022, Lecture Notes in Computer Science, Springer, Cham, 13363, https://doi.org/10.1007/978-3-031-09037-0_4, 2022.
- Karlsson, T. M., Arneborg, L., Broström, G., Almroth, B. C., Gipperth, L., and Hassellöv, M.: The unaccountability case of plastic pellet pollution, *Mar Pollut Bull.*, 129, 52–60, <https://doi.org/10.1016/j.marpolbul.2018.01.041>, 2018.
- Khanam, R. and Hussain, M.: YOLOv11: An overview of the key architectural enhancements, *arXiv [preprint]*, <https://doi.org/10.48550/arXiv.2410.17725>, 2024.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.*, Curran Associates, Inc., 25, <https://doi.org/10.1145/3065386>, 2012.
- Kubera, E., Kubik-Komar, A., Kurasiński, P., Piotrowska-Weryszko, K., and Skrzypiec, M.: Detection and recognition of pollen grains in multilabel microscopic images, *Sensors*, 22, 2690, <https://doi.org/10.3390/s22072690>, 2022.
- LeCun, Y., Bengio, Y., and Hinton, G.: Deep learning, *Nature*, 521, 436–444, <https://doi.org/10.1038/nature14539>, 2015.
- Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., and Dollár, P.: Microsoft COCO: Common objects in context, *arXiv [preprint]*, <https://doi.org/10.48550/ARXIV.1405.0312>, 2014.
- Lintner, M., Lintner, B., Wanek, W., Keul, N., von der Kammer, F., Hofmann, T., and Heinz, P.: Effects of heavy elements (Pb, Cu, Zn) on algal food uptake by *Elphidium excavatum* (Foraminifera), *Heliyon*, <https://doi.org/10.1016/j.heliyon.2021.e08427>, 2021.
- Mattsson, K., Ekstrand, E., Granberg, M., Hassellöv, M., and Magnusson, K.: Comparison of pre-treatment methods and heavy density liquids to optimize microplastic extrac-

- tion from natural marine sediments, *Sci. Rep.*, 12, 15459, <https://doi.org/10.1038/s41598-022-19623-5>, 2022.
- 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. Micropalaeontol.*, 39, 183–202, <https://doi.org/10.5194/jm-39-183-2020>, 2020.
- McGann, M., Holzmann, M., Bouchet, V. M. P., Disaró, S. T., Eichler, P. P. B., Haig, D. W., Himson, S. J., Kitazato, H., Pavard, J.-C., Polovodova Asteman, I., Rodrigues, A. R., Tremblin, C. M., Tsuchiya, M., Williams, M., O'Brien, P., Asplund, J., Axelsson, M., and Lorenson, T. D.: Analysis of a human-mediated microbioinvasion: the global spread of the benthic foraminifer *Trochammina hadai* Uchio, 1962, *J. Micropalaeontol.*, 44, 275–317, <https://doi.org/10.5194/jm-44-275-2025>, 2025.
- Mimura, K., Nakamura, K., Yasukawa, K., Sibert, E. C., Ohta, J., Kitazawa, T., and Kato, Y.: Applicability of object detection to microfossil research: Implications from deep learning models to detect microfossil fish teeth and denticles using YOLO-v7, *Earth Space Sci.*, 11, e2023EA003122, <https://doi.org/10.1029/2023EA003122>, 2024.
- 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.
- Morin, F., Panova, M. A. Z., Schweizer, M., Wiechmann, M., Eliassen, N., Sundberg, P., Cluzel-Burgalat, L., and Polovodova Asteman, I.: Hidden aliens: Application of digital PCR to track an exotic foraminifer across the Skagerrak (North Sea) correlates well with traditional morphospecies analysis, *Environ. Microbiol.*, <https://doi.org/10.1111/1462-2920.16458>, 2023.
- Morvan, J., Le Cadre, V., Jorissen, F., and Debenay, J. P.: Foraminifera as potential bio-indicators of the “Erika” oil spill in the Bay of Bourgneuf: field and experimental studies, *Aquat. Living Resour.*, 17, 317–322, <https://doi.org/10.1051/alr:2004034>, 2004.
- Murray, J. W.: Ecology and applications of benthic foraminifera. Cambridge University Press, Cambridge, UK, <https://doi.org/10.1017/CBO9780511535529>, 2006.
- Nordberg, K., Gustafsson, M., and Krantz, A.-L.: Decreasing oxygen concentrations in the Gullmar Fjord, Sweden, as confirmed by benthic foraminifera, and the possible association with NAO, *J. Mar. Syst.*, 23, 303–316, [https://doi.org/10.1016/S0924-7963\(99\)00067-6](https://doi.org/10.1016/S0924-7963(99)00067-6), 2000.
- Nordberg, K., Asteman, I. P., Gallagher, T. M., and Robijn, A.: Recent oxygen depletion and benthic faunal change in shallow areas of Sannäs Fjord, Swedish west coast, *J. Sea Res.*, 127, 46–62, <https://doi.org/10.1016/j.seares.2017.02.006>, 2017.
- O'Brien, P. A. J., Polovodova Asteman, I., and Bouchet, V. M. P.: Benthic foraminiferal indices and environmental quality assessment of transitional waters: A review of current challenges and future research perspectives, *Water*, <https://doi.org/10.3390/w13141898>, 2021.
- O'Brien, P. A. J., Schweizer, M., Morin, F., Hylén, A., Robertson, E. K., Hall, P. O., Quinchar, S., and Polovodova Asteman, I.: Meiofaunal Diversity Across a Sharp Gradient of Fjord Oxygenation: Insights from Metabarcoding and Morphology Approaches in Foraminifera, *Mar. Pollut. Bull.*, in press, 2025.
- Padilla, R., Passos, W. L., Dias, T. L. B., Netto, S. L., and Da Silva, E. A. B.: A comparative analysis of object detection metrics with a companion open-source toolkit, *Electronics (Switzerland)*, 10, 1–28, <https://doi.org/10.3390/electronics10030279>, 2021.
- Plavetić, M., Hsiang, A., Josefson, M., Hulthe, G., and Polovodova Asteman, I.: Deep learning accurately identifies fjord benthic foraminifera, Zenodo [data set], <https://doi.org/10.5281/zenodo.17608881>, 2025.
- Polovodova Asteman, I., and Nordberg, K.: Foraminiferal fauna from a deep basin in Gullmar Fjord: The influence of seasonal hypoxia and North Atlantic Oscillation, *J. Sea Res.*, 79, 40–49, <https://doi.org/10.1016/j.seares.2013.02.001>, 2013.
- Polovodova Asteman, I., Hanslik, D., and Nordberg, K.: An almost completed pollution-recovery cycle reflected by sediment geochemistry and benthic foraminiferal assemblages in a Swedish–Norwegian Skagerrak fjord, *Mar. Pollut. Bull.*, 95, 126–140, <https://doi.org/10.1016/j.marpolbul.2015.04.031>, 2015.
- Polovodova Asteman, I. and Schönfeld, J.: Recent invasion of the foraminifer *Nonionella stella* Cushman & Moyer, 1930 in northern European waters: evidence from the Skagerrak and its fjords, *J. Micropalaeontol.*, 35, 20–25, <https://doi.org/10.1144/jmpaleo2015-007>, 2016.
- Redmon, J., Divvala, S., Girshick, R., and Farhadi, A.: You only look once: Unified, real-time object detection, *arXiv [preprint]*, <https://doi.org/10.48550/ARXIV.1506.02640>, 2015.
- Richirt, J., Schweizer, M., Mouret, A., Quinchar, S., Saad, S. A., Bouchet, V. M. P., Wade, C. M., and Jorissen, F. J.: Biogeographic distribution of three phylotypes (T1, T2 and T6) of *Ammonia* (foraminifera, Rhizaria) around Great Britain: new insights from combined molecular and morphological recognition, *J. Micropalaeontol.*, 40, 61–74, <https://doi.org/10.5194/jm-40-61-2021>, 2021.
- Schönfeld, J., Alve, E., Geslin, E., Jorissen, F., Korsun, S., and Spezzaferri, S.: The FOBIMO (FOraminiferal BIO-MOnitoring) initiative – Towards a standardised protocol for soft-bottom benthic foraminiferal monitoring studies, *Mar. Micropaleontol.*, 94, 1–13, <https://doi.org/10.1016/j.marmicro.2012.06.001>, 2012.
- Schwing, P. T., Romero, I. C., Brooks, G. R., Hastings, D. W., Larson, R. A., and Hollander, D. J.: A decline in benthic foraminifera following the Deepwater Horizon event in the northeastern Gulf of Mexico, *PLoS ONE*, 10, e0120565, <https://doi.org/10.1371/journal.pone.0120565>, 2015.
- Shakya, M., Patel, R., and Joshi, S.: A comprehensive analysis of deep learning and transfer learning techniques for skin cancer classification, *Sci. Rep.*, 15, 4633, <https://doi.org/10.1038/s41598-024-82241-w>, 2025.
- Tariq, M. F. and Javed, M. A.: Small object detection with YOLO: A performance analysis across model versions and hardware, *arXiv [preprint]*, 2504.09900v1, <https://doi.org/10.48550/arXiv.2504.09900>, 2025.
- Tøn, A., Ahmed, A., Imran, A. S., Ullah, M., and Azad, R. M. A.: Metadata augmented deep neural networks for wild animal classification, *Ecol. Inform.*, 83, 102805, <https://doi.org/10.1016/j.ecoinf.2024.102805>, 2024.
- Yu, Z., Huang, H., Chen, W., Su, Y., Liu, Y., and Wang, X.: Yolo-facev2: A scale and occlusion

- aware face detector, *Pattern Recognit.*, 155, 110714, <https://doi.org/10.1016/j.patcog.2024.110714>, 2024.
- Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M.: YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, <https://doi.org/10.1109/CVPR52729.2023.00721>, 2023.
- Watts, E. G., Hylén, A., Hall, P. O., Eriksson, M., Robertson, E. K., Kenney, W. F., and Bianchi, T. S.: Burial of organic carbon in Swedish fjord sediments: Highlighting the importance of sediment accumulation rate in relation to fjord redox conditions, *J. Geophys. Res.-Biogeo.*, 129, e2023JG007978, <https://doi.org/10.1029/2023JG007978>, 2024.