RGe
ISSN: 2177-3246

# ARTIFICIAL INTELLIGENCE FOR ENVIRONMENTAL MONITORING: AUTOMATED IDENTIFICATION AND COUNTING OF BENTHIC MACROFAUNA USING CONVOLUTIONAL NEURAL NETWORKS

# INTELIGÊNCIA ARTIFICIAL PARA O MONITORAMENTO AMBIENTAL: IDENTIFICAÇÃO E CONTAGEM AUTOMATIZADAS DA MACROFAUNA BENTÔNICA POR REDES NEURAIS CONVOLUCIONAIS

# INTELIGENCIA ARTIFICIAL PARA EL MONITOREO AMBIENTAL: IDENTIFICACIÓN Y CONTEO AUTOMATIZADOS DE LA MACROFAUNA BENTÓNICA UTILIZANDO REDES NEURONALES CONVOLUCIONALES

doi 10.56238/revgeov17n6-099

**Ana Carolina Cellular Massone[1], Eduardo Camilo da Silva[2], Márcio Soares da Silva[3]**

## ABSTRACT

Benthic macrofauna; including mollusks, crustaceans, and polychaetes; plays a crucial role in regulating biogeochemical cycles, stabilizing sediments, and maintaining aquatic ecosystem resilience. Owing to their high sensitivity to environmental perturbations, these organisms are extensively used as bioindicators in ecological monitoring and impact assessments. Nevertheless, their traditional identification and enumeration rely on manual, labor-intensive procedures that demand taxonomic expertise and are susceptible to subjectivity and human error. To overcome these limitations, this study presents a robust and fully automated system based on Convolutional Neural Networks (CNNs) for the detection and quantification of benthic macrofauna from digital microscopy images. The proposed workflow encompasses standardized image acquisition, taxonomic annotation, advanced pre-processing, data augmentation, and supervised deep learning model training implemented in Python. Evaluation of the system revealed a mean F1-score of 0.93, mAP50 of 0.939, and inference times below 10 milliseconds per image, while maintaining stable validation precision, recall, and mAP values throughout training. Moreover, precision and recall metrics remained consistently high across diverse taxa, and the mean Average Precision (mAP50 and mAP50–95) improved steadily during training. These outcomes confirm the system's high generalization capacity and computational efficiency, enabling rapid, reliable, and reproducible ecological data generation. Beyond technical performance, the solution supports scalable biodiversity assessments, strategic environmental decision-making, and the optimization of conservation efforts, particularly in the face of climate change. This integration of artificial intelligence with benthic ecology represents a significant advancement in digital environmental monitoring, offering a cost-effective, time-efficient, and scientifically rigorous alternative to conventional analytical methods.

**Keywords:** Environmental Bioindicators. Convolutional Neural Networks (CNNs). Decision Making. Climate Change.

[1] Pós-Doutora (Doutora). E-mail: ana_cellular@id.uff.br
[2] Doutor. E-mail: ecamilo@id.uff.br
[3] Mestrando. E-mail: mssilva@id.uff.br

ISSN: 2177-3246

**RESUMO**
A macrofauna bentônica; incluindo moluscos, crustáceos e poliquetas; desempenha um papel crucial na regulação dos ciclos biogeoquímicos, na estabilização dos sedimentos e na manutenção da resiliência dos ecossistemas aquáticos. Devido à sua alta sensibilidade às perturbações ambientais, esses organismos são amplamente utilizados como bioindicadores no monitoramento ecológico e em avaliações de impacto. No entanto, sua identificação e contagem tradicionais dependem de procedimentos manuais e intensivos em mão de obra que exigem conhecimento taxonômico e estão sujeitos à subjetividade e ao erro humano. Para superar essas limitações, este estudo apresenta um sistema robusto e totalmente automatizado baseado em Redes Neurais Convolucionais (CNNs) para a detecção e quantificação da macrofauna bentônica a partir de imagens de microscopia digital. O fluxo de trabalho proposto abrange a aquisição padronizada de imagens, anotação taxonômica, pré-processamento avançado, aumento de dados e treinamento supervisionado de modelos de aprendizado profundo implementado em Python. A avaliação do sistema revelou um F1-score médio de 0,93, mAP50 de 0,939 e tempos de inferência inferiores a 10 milissegundos por imagem, mantendo valores estáveis de precisão, recall e mAP de validação ao longo do treinamento. Além disso, as métricas de precisão e recall permaneceram consistentemente altas em diversos táxons, e a precisão média (mAP50 e mAP50–95) melhorou de forma constante durante o treinamento. Esses resultados confirmam a alta capacidade de generalização e a eficiência computacional do sistema, permitindo a geração de dados ecológicos rápidos, confiáveis e reproduzíveis. Além do desempenho técnico, a solução apoia avaliações escaláveis da biodiversidade, a tomada de decisões ambientais estratégicas e a otimização dos esforços de conservação, especialmente diante das mudanças climáticas. Essa integração da inteligência artificial com a ecologia bentônica representa um avanço significativo no monitoramento ambiental digital, oferecendo uma alternativa econômica, eficiente em tempo e cientificamente rigorosa aos métodos analíticos convencionais.

**Palavras-chave:** Bioindicadores Ambientais. Redes Neurais Convolucionais (CNNs). Tomada de Decisão. Mudanças Climáticas.

**RESUMEN**
La macrofauna bentónica; incluyendo moluscos, crustáceos y poliquetos; desempeña un papel crucial en la regulación de los ciclos biogeoquímicos, la estabilización de los sedimentos y el mantenimiento de la resiliencia de los ecosistemas acuáticos. Debido a su alta sensibilidad a las perturbaciones ambientales, estos organismos se utilizan ampliamente como bioindicadores en el monitoreo ecológico y las evaluaciones de impacto. Sin embargo, su identificación y recuento tradicionales dependen de procedimientos manuales e intensivos en mano de obra que requieren conocimientos taxonómicos y son susceptibles a la subjetividad y al error humano. Para superar estas limitaciones, este estudio presenta un sistema robusto y totalmente automatizado basado en Redes Neuronales Convolucionales (CNNs) para la detección y cuantificación de la macrofauna bentónica a partir de imágenes de microscopía digital. El flujo de trabajo propuesto abarca la adquisición estandarizada de imágenes, la anotación taxonómica, el preprocesamiento avanzado, el aumento de datos y el entrenamiento supervisado de modelos de aprendizaje profundo implementado en Python. La evaluación del sistema reveló un F1-score medio de 0,93, un mAP50 de 0,939 y tiempos de inferencia inferiores a 10 milisegundos por imagen, manteniendo valores estables de precisión, recall y mAP de validación durante todo el entrenamiento. Además, las métricas de precisión y recall se mantuvieron consistentemente altas en diversos taxones, y la precisión media (mAP50 y mAP50–95) mejoró de manera constante durante el entrenamiento. Estos resultados confirman la alta capacidad de generalización y la eficiencia computacional del sistema, permitiendo la generación de datos ecológicos rápidos, fiables y

reproducibles. Más allá del rendimiento técnico, la solución apoya evaluaciones escalables de la biodiversidad, la toma de decisiones ambientales estratégicas y la optimización de los esfuerzos de conservación, especialmente frente al cambio climático. Esta integración de la inteligencia artificial con la ecología bentónica representa un avance significativo en el monitoreo ambiental digital, ofreciendo una alternativa económica, eficiente en tiempo y científicamente rigurosa a los métodos analíticos convencionales.

**Palabras clave:** Bioindicadores Ambientales. Redes Neuronales Convolucionales (CNNs). Toma de Decisiones. Cambio Climático.

ISSN: 2177-3246

## 1 INTRODUCTION

Aquatic ecosystems depend critically on benthic macrofauna, a diverse assemblage of invertebrates including mollusks, crustaceans, polychaetes, and echinoderms, to sustain fundamental ecological processes. Occupying the sediment-water interface, these organisms drive biogeochemical cycles such as organic matter decomposition, nutrient regeneration, and sediment oxygenation, while simultaneously serving as sensitive bioindicators of environmental quality (Avelino *et al.*, 2023; Kuhlmann, 2022; Hale *et al.*, 2024). Their intimate association with the substrate and low dispersal capacity make them particularly responsive to variations in sediment composition, oxygen availability, temperature, pH, and contaminant loadings, rendering them indispensable tools for ecological monitoring in estuarine, coastal, and lagoon ecosystems (Sousa *et al.*, 2023; Moura *et al.*, 2023).

Maintaining the ecological integrity of lagoon systems requires continuous monitoring of benthic communities, as these organisms mediate essential biogeochemical functions including nutrient cycling, pollutant filtration, and sediment stabilization. Their spatial distribution and abundance patterns reliably reflect environmental impacts associated with pollution gradients, granulometric shifts, and anthropogenic disturbances such as bottom trawling (Li *et al.*, 2022; Bradshaw *et al.*, 2024). Standard benthic sampling protocols involve the selection of georeferenced stations across environmental gradients, with multiple replicates per point to ensure statistical reliability. Sample recovery is typically performed using corers, dredges, or nets, followed by sieving, fixation in ethanol, and taxonomic sorting under stereomicroscopy (Kuhlmann, 2022; Pardo *et al.*, 2023).

Laboratory processing of benthic samples relies on stereomicroscopic examination combined with dichotomous identification keys to classify organisms into functional groups. Ecological quality is subsequently assessed through biological indices that integrate species diversity, abundance, and dominance, enabling spatial and temporal comparisons across monitoring programs (Sun *et al.*, 2021; Han *et al.*, 2021). Although scientifically rigorous, this workflow is time-consuming, operator-dependent, and demands highly specialized taxonomic expertise, constraints that severely limit the scale and frequency of monitoring campaigns and hinder timely responses to environmental degradation (Ghimire *et al.*, 2021; Neves and Valentin, 2011).

Computer vision based on Convolutional Neural Networks (CNNs) has emerged as a transformative solution to automate image-based biodiversity assessments. By learning hierarchical spatial representations directly from labeled images, CNNs can distinguish morphologically similar taxa with high accuracy, even under variable lighting and background conditions (Villon *et al.*, 2018; Xu *et al.*, 2020; Zhu *et al.*, 2021). Their capacity to process thousands of images per hour, with inference times below 10 milliseconds, makes them particularly well-suited for high-throughput ecological screening. Deep learning frameworks such as PyTorch and TensorFlow have further democratized access to these architectures, enabling their deployment in research and regulatory monitoring contexts (LeCun *et al.*, 2015; Krizhevsky *et al.*, 2012).

Despite this growing adoption across aquatic sciences, targeted applications of CNNs to benthic macrofauna in lagoon environments remain scarce. Existing models have largely focused on planktonic organisms, coral reefs, and fish assemblages (Ferreira *et al.*, 2022; Mizuno *et al.*, 2024), leaving a critical methodological gap for sediment-dwelling invertebrates. The morphological complexity of benthic taxa, characterized by appendage segmentation, shell curvature, body transparency, and sediment adhesion, demands specialized training datasets, tailored augmentation strategies, and robust image processing pipelines to achieve reliable classification under real ecological conditions.

Addressing this gap is particularly urgent given the accelerating pace of global environmental change. Lagoon and estuarine systems are among the ecosystems most vulnerable to sea-level rise, ocean acidification, eutrophication, and hypoxic events (IPCC, 2023; Duarte *et al.*, 2013). Benthic macrofauna, by virtue of their sedentary lifestyle and narrow physiological tolerance, function as early-warning sentinels of these stressors. Automating their identification and quantification substantially expands the temporal and spatial resolution of biodiversity data, enabling the detection of ecological imbalances and long-term community shifts that would otherwise go undetected (Borja *et al.*, 2010; McLeod *et al.*, 2011).

Beyond ecological monitoring, this research carries strategic implications for climate adaptation governance and blue carbon policy. Continuous, high-frequency benthic data support environmental licensing processes, certification of blue carbon projects, and the construction of long-term biodiversity indicators aligned with the Sustainable Development Goals (SDGs) and the Convention on Biological Diversity (CBD). The preservation of macrofaunal integrity underpins ecosystem services, sediment stabilization, nutrient cycling, and carbon sequestration, that are central to coastal resilience frameworks (Barbier *et al.*, 2011; Howard *et al.*, 2017; UNEP, 2021).

This study presents the development, training, and validation of a CNN-based automated system for the identification and enumeration of benthic macrofauna from digital microscopy images. The proposed pipeline integrates standardized image acquisition,

ISSN: 2177-3246

manual taxonomic annotation, data augmentation, and supervised model training in Python. Model performance is evaluated through accuracy, F1-score, recall, and mean Average Precision (mAP), ensuring practical applicability for large-scale ecological monitoring. The novelty of this work lies in adapting CNN architectures to the specific morphological and imaging challenges of benthic macrofauna, with direct implications for accelerating biodiversity assessments and strengthening evidence-based environmental decision-making.

Although recent advances in artificial intelligence have substantially expanded the use of deep learning techniques in ecological monitoring, applications focused on benthic macrofauna remain limited, particularly in lagoon ecosystems. Most existing studies have concentrated on fish assemblages, coral reefs, plankton communities, or broad marine biodiversity assessments, leaving a significant methodological gap regarding sediment-associated invertebrates. Furthermore, few studies have explored the integration of automated benthic monitoring with environmental governance and strategic decision-making frameworks (Zang *et al.*, 2019). Consequently, there is a growing need for intelligent and scalable solutions capable of accelerating laboratory screening procedures while simultaneously generating reliable ecological indicators to support environmental management and climate adaptation policies.

## 2 MATERIALS AND METHODS

### 2.1 STUDY AREA AND SAMPLING PROCEDURES

The study was conducted in the Maricá Lagoon, a hypersaline coastal lagoon situated in the state of Rio de Janeiro, southeastern Brazil. Embedded within a broader complex of Atlantic-domain coastal lagoons, this system is characterized by pronounced environmental heterogeneity, including gradients of salinity, tidal exchange, sediment composition, and organic matter concentration driven by both natural dynamics and anthropogenic pressures such as urban runoff and shoreline modification. These features make it an ecologically sensitive and scientifically valuable site for investigating the spatial distribution and community structure of benthic macrofauna.

Sampling stations were selected based on stratified environmental gradients encompassing proximity to urban discharge points, sediment type, and vegetation cover, with logistical constraints of accessibility and safety also considered. This stratification aimed to ensure spatial representation across distinct ecological zones and to capture the full range of benthic community variability within the lagoon system.

Field campaigns were conducted during both the dry and rainy seasons to account for temporal variability associated with freshwater inflows and precipitation. Surface sediments were collected using a cylindrical corer (10 cm diameter, 15 cm penetration depth), targeting the uppermost sediment horizon where macroinvertebrate abundance is highest. At each of the 30 georeferenced stations, three replicate cores were obtained (n = 90 total), and the material was immediately sieved through a 500 µm mesh in the field to remove excess sediment and retain macrofauna.

Retained organisms were preserved in 70% ethanol and transported to the laboratory, where specimens were sorted under a stereomicroscope and prepared for imaging and taxonomic classification. The focal taxa; polychaetes, amphipods, and gastropods; were selected for their ecological relevance as bioindicators of sediment quality and their representation across the sampled environmental gradients.

## 2.2 IMAGE ACQUISITION AND ANNOTATION

High-resolution digital images were obtained from each biological sample using a digital microscope equipped with adjustable LED lighting and a uniform background substrate. Image capture followed a standardized protocol to ensure consistent illumination, magnification, and contrast across all samples, thereby minimizing sources of technical variability in the training dataset. Each specimen was individually positioned and photographed under controlled conditions.

All images were manually annotated using bounding boxes to delimit and taxonomically classify each visible organism, generating a structured labeled dataset suitable for supervised object detection training. The species included in the training set were Heliobia spp., Streblospio sp., and Cassidinidea fluminensis. To reflect differences in acquisition conditions, three imaging subsets were defined: HeliobiaN (specimens photographed under normal magnification without microscopy), HeliobiaM (specimens imaged with one microscope configuration), and HeliobiaM1 (specimens imaged with an alternative microscope setup). The remaining taxa were annotated exclusively from one microscopy modality.

## 2.3 MODEL ARCHITECTURE AND TRAINING STRATEGY

The annotated image dataset was partitioned into training (80%) and validation (20%) subsets using stratified sampling to preserve class proportions across splits. To enhance model generalizability and mitigate overfitting, data augmentation was applied during training, including horizontal and vertical flipping, brightness and contrast perturbations, and zoom transformations. The full pipeline was implemented using a supervised deep learning framework in Python.

The network architecture followed a convolutional design comprising successive feature extraction stages. Convolutional layers captured low- and mid-level spatial features; edges, textures, and morphological patterns; while max-pooling layers progressively reduced spatial dimensionality. Batch normalization was applied after each convolutional block to stabilize gradient flow and accelerate convergence. Fully connected layers at the output stage integrated the learned feature representations to generate species-level classifications via a softmax activation function.

Model training was guided by a composite loss function combining three terms: a bounding box localization loss (Lbox), a cross-entropy classification loss (Lcls), and a Distribution Focal Loss (Ldfl) that assigns greater weight to difficult instances. The total loss is expressed as:

$$L_{total} = L_{box} + L_{cls} + L_{dfl} \qquad (1)$$

The localization loss penalizes spatial misalignment between predicted and ground-truth bounding boxes:

$$L_{box} = \sum_{i=1}^{N} \delta_i \cdot IoU_{loss}(\hat{b}_i, b_i) \qquad (2)$$

where N is the total number of predicted objects; $\delta_i \in \{0,1\}$ indicates whether prediction i constitutes a valid match; $\hat{b}_i$ and $b_i$ are the predicted and ground-truth bounding boxes, respectively; and IoUloss can be instantiated as GIoU, DIoU, or CIoU to penalize poor localization alignment.

The classification loss minimizes the cross-entropy between predicted probabilities and true taxonomic labels:

$$L_{cls} = -\sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} \cdot \log(\hat{p}_{i,c}) \qquad (3)$$

where C is the number of classes; $y_{i,c} \in \{0,1\}$ is the ground-truth label for instance i and class c; and $\hat{p}_{i,c}$ is the predicted softmax probability for that class.

The Distribution Focal Loss refines bounding box regression by concentrating learning on harder examples:

$$L_{dfl} = -\sum_{i=1}^{N} \sum_{j=1}^{M} w_{i,j} \cdot \log(p_{i,j}) \quad (4)$$

where M is the number of discrete regression bins; $w_{i,j}$ is the target-distribution weight for bin j of instance i; and $p_{i,j}$ is the predicted probability for that bin.

Training was performed using stochastic gradient descent with adaptive learning rate scheduling and early stopping criteria. The final model checkpoint was selected based on the highest mAP@50 achieved on the validation set. All computations were executed on a workstation equipped with an NVIDIA RTX 3060 GPU.

## 2.4 EVALUATION METRICS AND PERFORMANCE ANALYSIS

Model performance was assessed using a multi-metric framework combining standard classification and object detection indicators, computed at both per-class and aggregate levels to characterize overall and class-specific behavior.

Precision quantifies the proportion of detections that are truly positive among all detections made by the model:

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

Recall measures the proportion of actual positive instances that the model successfully detected:

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

The F1-Score provides the harmonic mean of precision and recall, offering a balanced evaluation that accounts for both false positives and false negatives:

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (7)$$

The mean Average Precision at IoU ≥ 0.50 (mAP@50) summarizes detection performance across all classes at a standard localization threshold:

$$mAP@50 = \frac{1}{M} \sum_{i=1}^{M} \int_0^1 p_i(r)\,dr \quad (8)$$

where M is the number of classes and $p_i(r)$ denotes the precision–recall function for class i, with the integral computing the area under that curve at IoU = 0.50.

To provide a stricter and more comprehensive localization assessment, the mAP@50:95 metric averages detection performance across ten IoU thresholds from 0.50 to 0.95 in 0.05 increments:

$$mAP@50{:}95 = \frac{1}{T} \sum_{t=1}^{T} \left[ \frac{1}{M} \sum_{i=1}^{M} \int_0^1 p_{i,t}(r)\,dr \right] \quad (9)$$

where T = 10 is the number of thresholds and $p_{i,t}(r)$ is the precision–recall curve for class i at threshold t. Inference time per image was additionally recorded to evaluate real-time applicability in laboratory workflows.

## 2.5 POST-PROCESSING METRICS AND CONFIDENCE EVALUATION

Detection confidence scores were derived from the classification probability output of the final prediction layer, yielding a probability estimate over the K taxonomic classes. A prediction was considered valid only if its confidence score exceeded an empirically defined threshold (typically 0.25–0.50), thereby balancing sensitivity to low-abundance taxa against the suppression of false positives.

Overall classification accuracy was computed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

where TP, TN, FP, and FN denote true positives, true negatives, false positives, and false negatives, respectively. This metric captures the proportion of all predictions — both presences and absences — that were correctly classified.

The confidence score C for a given input x is formally defined as the maximum softmax probability across all classes:

$$C = \max_{k \in \{1,...,K\}} P(y = k \mid x) \quad (11)$$

where $P(y = k \mid x)$ is the model's estimated probability that input x belongs to class k, and the max operator identifies the most likely taxonomic assignment. Confidence intervals for precision and accuracy estimates were determined through bootstrapping on validation subsets to ensure statistical robustness of reported performance values.

### 2.6 JUSTIFICATION OF METRIC SELECTION

The adoption of a multi-metric evaluation framework is warranted by the ecological complexity inherent to benthic monitoring. In conservation contexts, false negatives, undetected species, may result in the overlooking of rare or functionally critical taxa, whereas false positives risk inflating biodiversity assessments and misguiding management decisions. Precision and recall directly quantify these error types and are therefore indispensable for ecologically valid performance reporting.

The mAP metrics extend evaluation to the spatial domain, assessing the accuracy with which the model localizes each specimen within the image. This is especially important for small, partially occluded, or morphologically convergent organisms that are common in complex sediment samples. The combined use of classification and localization metrics ensures that the system meets the dual requirements of taxonomic accuracy and spatial precision demanded by environmental monitoring applications.

Confidence thresholds and inference time metrics further provide the operational parameters needed for scaling the system to automated laboratory pipelines and mobile field stations, where processing speed and prediction reliability are equally critical considerations. Together, these indicators constitute a comprehensive and ecologically grounded evaluation standard aligned with both scientific best practices and regulatory monitoring requirements.

## 3 RESULTS

This section provides a comprehensive evaluation of the CNN model developed for the automated identification and enumeration of benthic macrofauna. The assessment encompasses quantitative performance metrics; precision, recall, F1-score, and mean Average Precision; alongside an examination of training dynamics, confusion matrices, and detection examples. Taken together, these elements characterize the model's predictive capacity, operational reliability, and readiness for integration into ecological monitoring workflows.

### 3.1 CNN MODEL PERFORMANCE

The training dataset comprised 219 digital microscopy images of varying resolution, encompassing both single-species and multi-species compositions. Of these, 133 images were allocated to training and 86 to validation. The model was trained in a Python 3.9.13

environment with CUDA GPU acceleration (Torch 2.5.1) on an NVIDIA RTX 3060. The final architecture consisted of 168 layers and 11,128,293 parameters, operating at 28.5 GFLOPs, a moderately complex configuration well-suited to the scale of the dataset. Figure 1 summarizes the key output parameters recorded at the conclusion of training.

**Figure 1**

*Final training output parameters*

```
    Epoch    GPU_mem   box_loss   cls_loss   dfl_loss  Instances       Size
  200/200      3.89G     0.3846     0.2305     0.8446         41        640: 100%|██████████| 9/9 [00:02<00:00,  4.14it/s]
                 Class     Images  Instances      Box(P          R      mAP50  mAP50-95): 100%|██████████| 3/3 [00:00<00:00,  4.20it/s]
                   all         86        706      0.946      0.899      0.934      0.676
```

Source: author.

Validation was performed over 86 images containing 706 annotated instances distributed across five taxonomic classes. Among individual classes, HeliobiaN achieved a precision of 0.977, recall of 0.960, mAP50 of 0.939, and mAP50–95 of 0.522. HeliobiaM recorded precision of 0.968, recall of 0.786, mAP50 of 0.847, and mAP50–95 of 0.737. HeliobiaM1 attained precision of 0.931, recall of 0.973, mAP50 of 0.948, and mAP50–95 of 0.742. Streblospio_spM yielded precision of 0.937 and recall of 0.817, with mAP50 of 0.916 and mAP50–95 of 0.585. The highest overall performance was recorded for Cassid_Flum, which achieved a precision of 0.961, perfect recall of 1.000, mAP50 of 0.995, and mAP50–95 of 0.847. Although Cassid_Flum achieved the highest performance metrics, these results should be interpreted with caution due to the limited number of validation instances available for this class.

Across all classes, the model attained a mean mAP50 of 0.939 and an overall F1-score of 0.93, confirming strong classification capacity. The mAP50–95 of 0.687 indicates that while localization at standard thresholds is robust, performance at stricter IoU criteria leaves room for improvement, a limitation primarily attributable to the limited size of the current dataset, for which expansion is recommended. Inference was computationally efficient: pre-processing required 0.3 ms, inference 6.9 ms, and post-processing 0.8 ms per image, demonstrating suitability for real-time laboratory screening. Figure 2 presents the confusion matrices for the trained model.

RGe

ISSN: 2177-3246

**Figure 2**

*Confusion matrix: (a) absolute counts, (b) normalized by class*

| Class | Images | Instances | Box(P | R | mAP50 | mAP50-95) |
|---|---|---|---|---|---|---|
| all | 86 | 706 | 0.955 | 0.907 | 0.939 | 0.687 |
| HeliobiaN | 5 | 50 | 0.977 | 0.96 | 0.99 | 0.522 |
| HeliobiaM | 10 | 76 | 0.968 | 0.786 | 0.847 | 0.737 |
| HeliobiaM1 | 38 | 429 | 0.931 | 0.973 | 0.948 | 0.742 |
| Streblospio_spM | 38 | 145 | 0.937 | 0.817 | 0.916 | 0.585 |
| Cassid_Flum | 3 | 6 | 0.961 | 1 | 0.995 | 0.847 |

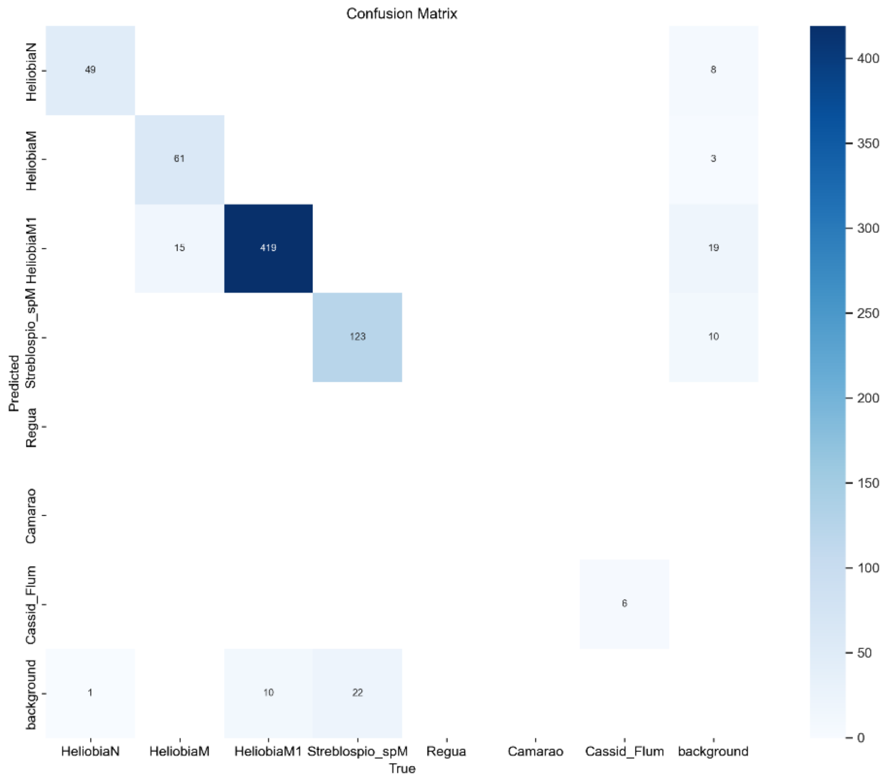
Source: author.

The absolute confusion matrix confirms concentrated correct predictions along the main diagonal: 49 instances for HeliobiaN, 61 for HeliobiaM, 419 for HeliobiaM1, 123 for Streblospio_spM, and 6 for Cassid_Flum. The normalized matrix reveals per-class accuracy rates of 0.98 for HeliobiaN, 0.80 for HeliobiaM, 0.98 for HeliobiaM1, 0.85 for Streblospio_spM, and 1.00 for Cassid_Flum. The confusion between HeliobiaM and HeliobiaM1 suggests that differences introduced by microscope configuration may have affected class separability. Moreover, the presence of false negatives classified as background, particularly for Streblospio_spM, indicates that small or partially overlapping organisms remain challenging for the model. These findings reinforce the importance of image standardization and dataset expansion to improve detection robustness. The comparatively lower recall of HeliobiaM (0.80) likely reflects imaging inconsistencies between the two microscope configurations used for that class, reinforcing the need for standardized acquisition protocols. Figure 3 presents the key performance curves.

ISSN: 2177-3246

**Figure 3**

*Performance metrics: F1-Confidence (a), Precision–Confidence (b), Precision–Recall (c), and Recall–Confidence (d)*

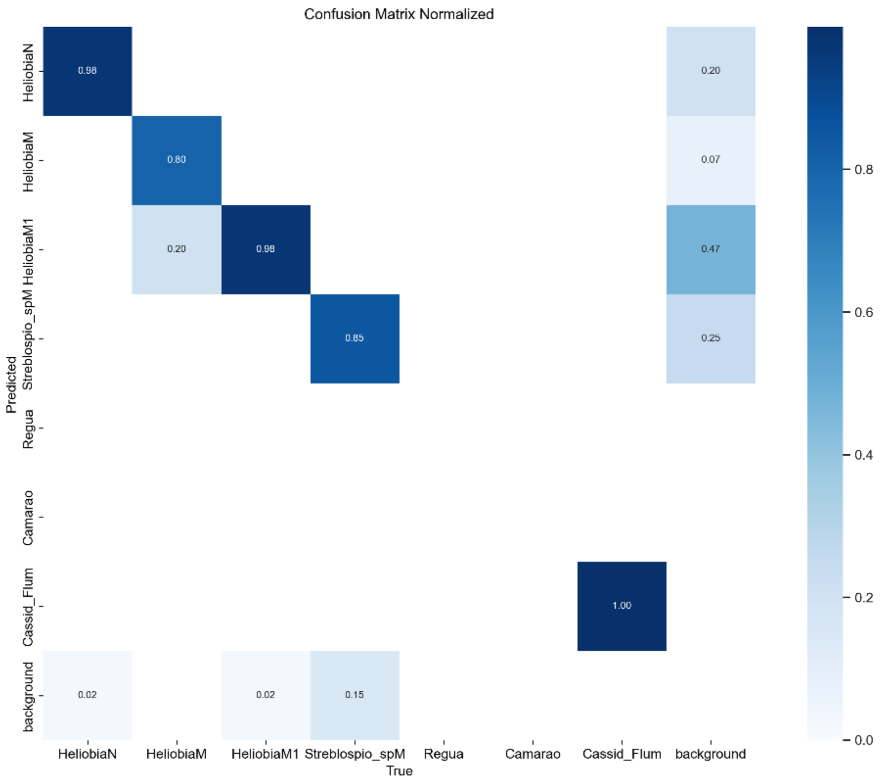



Source: author.

The F1-Confidence curve stabilizes at 0.93 at a confidence threshold of 0.625, reflecting a well-balanced trade-off between precision and recall across the operating range. The Precision-Confidence curve demonstrates that precision reaches 1.00 at a threshold of 0.98, indicating that highly confident predictions are almost exclusively correct, a desirable property for applications where taxonomic misidentification carries regulatory consequences. The Precision-Recall curve shows that a mean average precision of 0.939 is maintained up to recall levels of 0.5, representing an optimal generalization regime. At higher recall levels, precision decreases for some classes, particularly Streblospio_spM, suggesting that conservative thresholding may be preferable in surveys targeting rare taxa. The Recall-Confidence curve confirms that high recall is sustained at low confidence thresholds and degrades progressively as the threshold is raised, a trade-off inherent to all detection systems. Figure 4 presents the full training and validation learning curves.

**Figure 4**

*Training and validation loss, precision, recall, and mAP curves across 200 epochs*




Source: author.

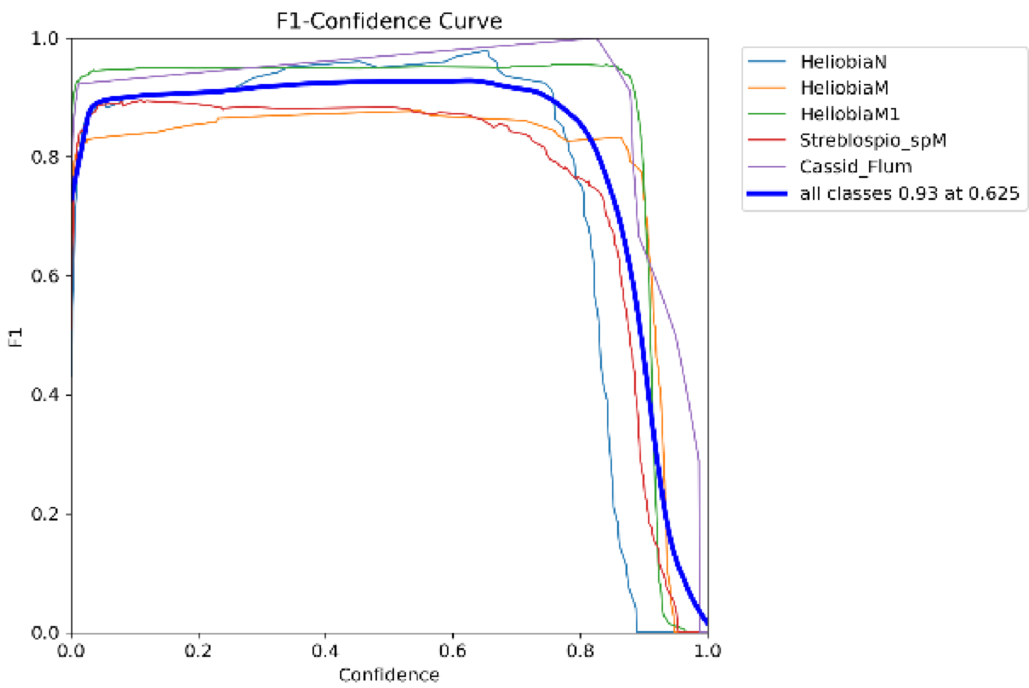
The training loss components (box, classification, and DFL) decreased consistently throughout the 200 training epochs, indicating progressive optimization of model parameters. In the validation set, classification loss decreased and stabilized, while box loss remained relatively stable after the initial training stages. The validation DFL loss exhibited a slight increase during later epochs, suggesting that bounding-box refinement may still benefit from a larger and more diverse dataset. Nevertheless, precision, recall, mAP50, and mAP50–95 remained stable or improved during training, indicating satisfactory generalization to unseen data. These results are comparable to, and in some cases exceed, those reported in previous ecological monitoring studies employing deep learning techniques. Villon *et al.* (2018) demonstrated the feasibility of applying deep learning techniques to underwater species classification, while Zhu *et al.* (2021) highlighted the challenges associated with maintaining high precision when dealing with morphologically similar taxa., while Zhu *et al.* (2021) highlighted the difficulty of maintaining high precision when dealing with morphologically similar taxa. The F1-score of 0.93 obtained in the present study therefore demonstrates that CNN-based approaches can achieve robust performance even when applied to small benthic invertebrates exhibiting substantial morphological variability.

## 3.2 AUTOMATED DETECTION AND COUNTING RESULTS

The automated detection system demonstrated strong performance across speed, taxonomic accuracy, and generalization, even given the relatively small and heterogeneous image dataset spanning both microscopy and non-microscopy acquisition conditions. The mean inference time of under 10 ms per image supports real-time integration into laboratory triage workflows, substantially reducing the time burden of manual screening. Figures 5 through 9 illustrate representative detection outputs alongside manual reference counts for each taxonomic class.

**Figure 5**

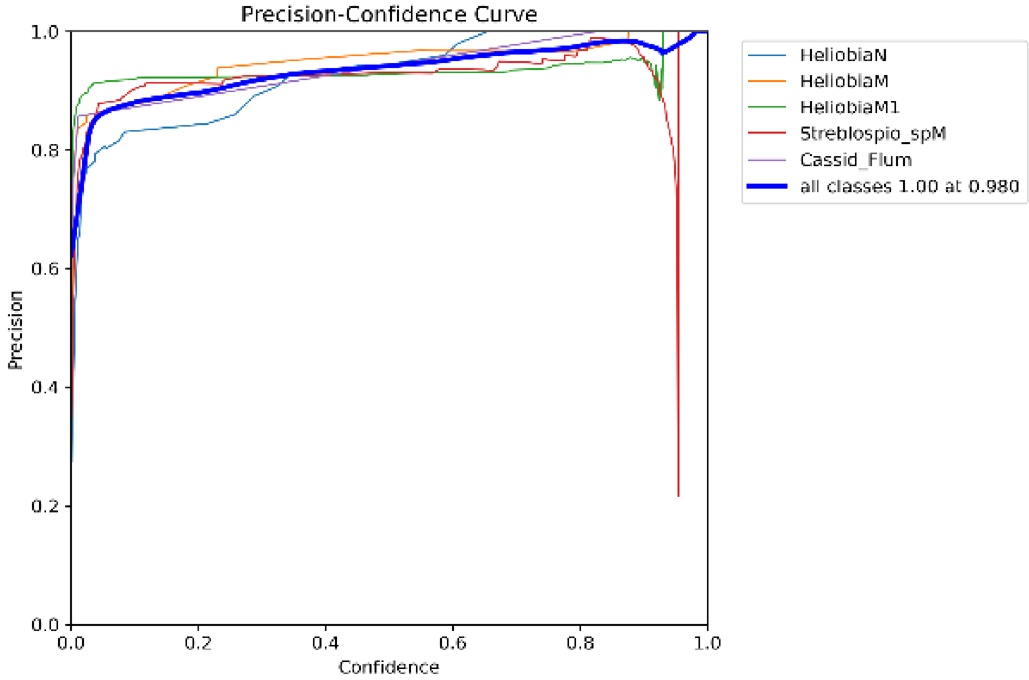
*Detection output and confidence statistics: HeliobiaN (standard magnification). 100% correct species identification and counting*




Source: author.

**Figure 6**

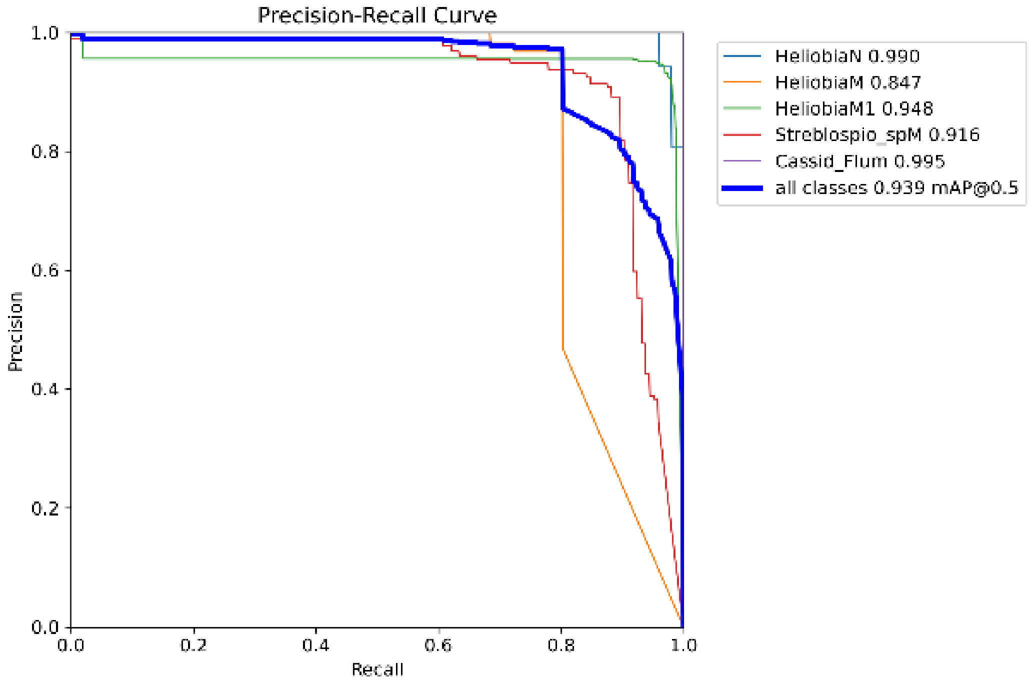
*Detection output and confidence statistics: HeliobiaM1 (microscopy, configuration 1). 100% correct species identification and counting*




Source: author.

**Figure 7**

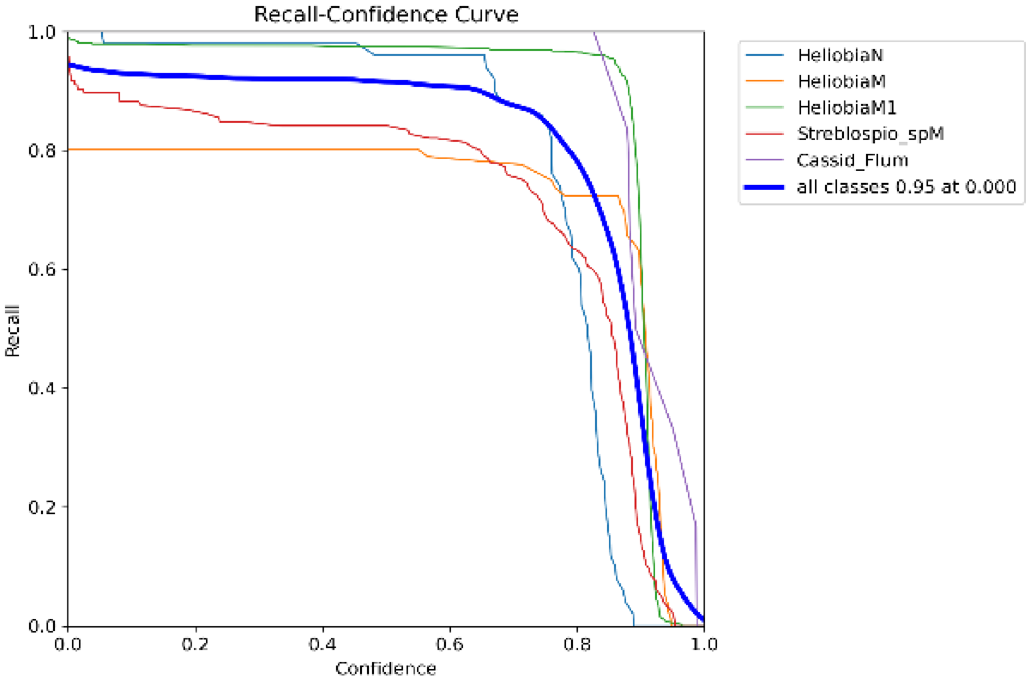
*Detection output and confidence statistics: Cassid_Flum (microscopy). 100% correct species identification and counting*




Source: author.

**Figure 8**

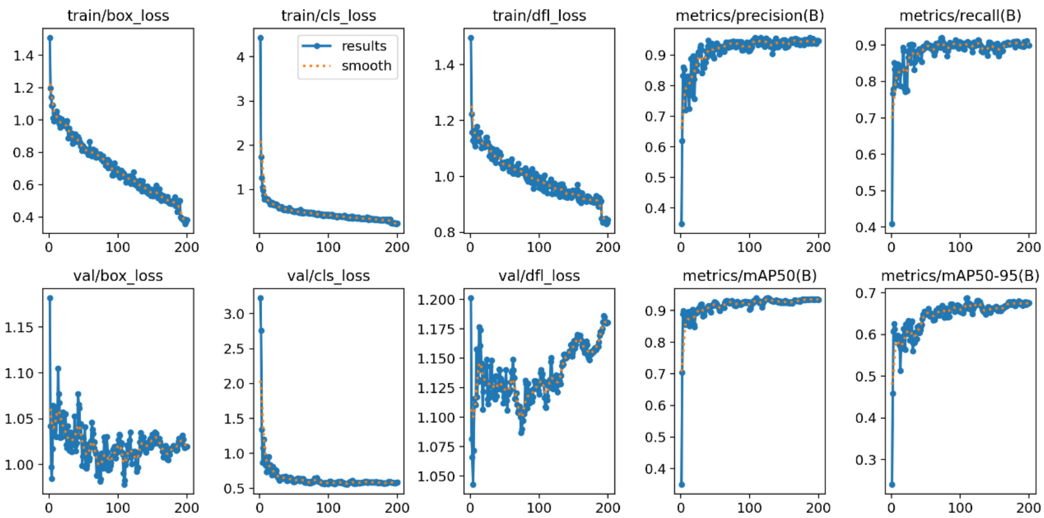
*Detection output and confidence statistics: Streblospio_spM (microscopy, mixed sample). 100% accuracy in species identification and approximately 45% accuracy in specimen counting*




Source: author.

**Figure 9**

*Detection output and confidence statistics: HeliobiaM1 and Streblospio_spM (microscopy, mixed sample). 100% correct species identification and counting*




Source: author.

Automated counts matched manual reference counts within a 5% error margin for correctly classified instances. Deviations were predominantly associated with low-image-quality samples and overlapping specimens, two conditions that degrade bounding box discrimination and inflate both false positive and false negative rates. These findings underscore the importance of image acquisition standardization and point toward dataset expansion, particularly for underrepresented classes and challenging imaging scenarios, as the primary avenue for further performance improvement.

The system's alignment with Industry 4.0 principles enables its integration into continuous ecological monitoring architectures, supporting automated data pipelines, anomaly detection, and decision support interfaces. This capability represents a substantive shift from reactive, sample-driven monitoring toward proactive, data-continuous surveillance of benthic community dynamics.

### 3.3 COMPARISON WITH TRADITIONAL LABORATORY SCREENING

Traditional benthic macrofauna screening relies on manual identification and counting performed by trained taxonomists under stereomicroscopes. Although processing times vary according to sample complexity, manual laboratory screening typically requires several minutes to hours per sample, whereas the proposed CNN performs image inference in less than 10 ms, representing a substantial increase in analytical throughput.

Considering that a single laboratory sample may contain dozens to hundreds of individuals requiring manual sorting and identification, the computational efficiency achieved by the proposed model represents a substantial reduction in processing time and human effort. This gain becomes particularly relevant in long-term monitoring programs involving large numbers of samples and repeated surveys.

Beyond speed, automated screening contributes to process standardization by reducing operator-dependent variability, which is a common source of inconsistency in ecological assessments. While expert validation remains essential, particularly for rare or morphologically similar taxa, the proposed approach significantly reduces the workload associated with routine laboratory procedures.

These findings indicate that artificial intelligence can serve as a complementary tool for benthic monitoring programs, increasing analytical throughput while preserving acceptable levels of taxonomic accuracy and reproducibility.

### 3.4 IMPLICATIONS FOR ENVIRONMENTAL GOVERNANCE

The proposed system extends beyond a purely technological contribution and presents important implications for environmental governance. By enabling faster and more frequent assessments of benthic communities, the approach facilitates the generation of continuous ecological indicators that can support environmental licensing, biodiversity monitoring programs, and adaptive management strategies.

Because benthic macrofauna are highly sensitive to environmental disturbances, automated monitoring systems can function as early-warning tools for detecting ecological degradation associated with pollution, eutrophication, habitat alteration, and climate-related stressors. The availability of standardized and reproducible biodiversity information strengthens evidence-based decision-making and contributes to more transparent environmental management processes.

Furthermore, the integration of artificial intelligence into environmental monitoring frameworks aligns with current global initiatives related to ecosystem conservation, climate adaptation, blue carbon programs, and sustainability governance. As environmental datasets become increasingly complex and voluminous, intelligent monitoring systems are expected to play a central role in supporting strategic environmental planning and long-term ecosystem resilience.

## 4 CONCLUSIONS AND FUTURE WORKS

This study demonstrated that Convolutional Neural Networks can be effectively trained to automate the identification and enumeration of benthic macrofauna from digital microscopy images, achieving high classification accuracy, robust generalization, and inference speeds compatible with real-time laboratory deployment. The trained model attained a mean mAP50 of 0.939 and an F1-score of 0.93 across five taxonomic classes, with per-class accuracy reaching 0.98 for HeliobiaN and HeliobiaM1 and 1.00 for Cassid_Flum. These results confirm that the proposed pipeline constitutes a scientifically rigorous and operationally practical alternative to conventional manual triage methods.

From a strategic environmental management perspective, the implications of this work extend well beyond laboratory efficiency. Automating benthic macrofauna analysis enables the continuous integration of sensitive ecological indicators into decision-making processes, expanding the planning, control, and risk-response capabilities of both public agencies and private operators. The system provides a high-throughput decision-support tool for environmental licensing, ecological impact assessment, and biodiversity certification, facilitating a transition from reactive to proactive conservation strategies grounded in real-time empirical evidence.

In the context of global climate change, the capacity to generate continuous, georeferenced, and statistically validated time series of benthic community data represents a strategic advance. Benthic macrofauna respond sensitively to temperature anomalies,

ISSN: 2177-3246

dissolved oxygen depletion, and contaminant accumulation, precisely the stressors intensifying under current climate trajectories. Early-warning signals embedded in these time series can inform adaptive management plans, trigger timely interventions, and support the delineation of priority conservation zones, thereby operationalizing the climate resilience frameworks demanded by national and international environmental policy.

The practical scalability of the approach is a further key contribution. Expanding monitoring frequency and spatial coverage without proportional increases in operational cost removes a longstanding barrier to continuous ecological surveillance. The proposed approach has potential applicability within environmental monitoring programs associated with National Adaptation Plans, Payment for Environmental Services initiatives, and blue carbon projects. Despite the promising results, the present study has some limitations. The image dataset used for training and validation was relatively small and restricted to a limited number of taxonomic groups. Additionally, the images were obtained under controlled laboratory conditions, which may differ from other acquisition environments. Future studies should expand the dataset, incorporate additional species, and evaluate the model under different imaging conditions to further improve its robustness and generalization capability.

Future work should prioritize three complementary directions. First, dataset expansion across additional taxa, sampling seasons, and image acquisition conditions will be essential to improve mAP50–95 performance and consolidate generalization across the morphological diversity encountered in real monitoring programs. Second, integration of the model into end-to-end automated triage platforms; combining image acquisition hardware, detection inference, and ecological database management; will unlock the full operational potential of the system. Third, the development of transfer learning strategies to adapt the trained model to new geographic regions and lagoon systems with minimal additional annotation effort will substantially broaden its applicability and accelerate adoption by monitoring agencies worldwide.

## REFERENCES

Avelino, D.F.G., Silva, A.M.C., Avelino, P.G., Sá, M.M.S., Soares, A.X., 2023. Macroinvertebrados bentônicos como bioindicadores da qualidade ambiental dos recifes de arenito da praia de Porto de Galinhas (Pernambuco). Revista Brasileira de Meio Ambiente 11(1), 182–201.

Barbier, E.B., Hacker, S.D., Kennedy, C., Koch, E.W., Stier, A.C., Silliman, B.R., 2011. The value of estuarine and coastal ecosystem services. Ecological Monographs 81(2), 169–193. https://doi.org/10.1890/10-1510.1.

Borja, A., Dauer, D.M., Grémare, A., Elliott, M., 2010. The importance of setting targets and reference conditions in assessing marine ecosystem quality. Ecological Indicators 10(1), 1–9. https://doi.org/10.1016/j.ecolind.2009.07.008.

Bradshaw, C., Iburg, S., Morys, C., Sköld, M., Pusceddu, A., Ennas, C., Jonsson, P., Nascimento, F.J.A., 2024. Effects of bottom trawling and environmental factors on benthic bacteria, meiofauna and macrofauna communities and benthic ecosystem processes. Science of the Total Environment 921, 171076.

Duarte, C.M., Hendriks, I.E., Moore, T.S., Olsen, Y.S., Steckbauer, A., Ramajo, L., Carstensen, J., Trotter, J.A., McCulloch, M., 2013. Impact of ocean acidification on the structure and function of benthic ecosystems. Biogeosciences 10, 2761–2774. https://doi.org/10.5194/bg-10-2761-2013.

Ferreira, B., Carvalho, R., Silva, E., 2022. Deep learning for coral reef classification: A survey. Ecological Informatics 70, 101703.

Ghimire, D., Lee, J., 2021. Deep learning-based aquatic organism classification: A review. Aquaculture Reports 19, 100568. https://doi.org/10.1016/j.aqrep.2021.100568.

Hale, R., Bigham, K.T., Rowden, A.A., Halliday, J., Nodder, S.D., Orpin, A.R., Frontin-Rollet, G., Maier, K.L., Mountjoy, J.J., Pinkerton, M.H., 2024. Bioturbation and faunal-mediated ecosystem functioning in a deep-sea benthic community recovering from a severe seabed disturbance. Deep-Sea Research Part I 204, 104235.

Han, C., Xu, Z., Liu, X., 2021. Characteristics of macrofaunal assemblages and their relationships with environmental factors in a semi-enclosed bay. Marine Pollution Bulletin 167, 112348. https://doi.org/10.1016/j.marpolbul.2021.112348.

Howard, J., Sutton-Grier, A., Herr, D., Kleypas, J., McLeod, E., Pidgeon, E., Simpson, S., 2017. Coastal blue carbon: methods for assessing carbon stocks and emissions factors in mangroves, tidal salt marshes, and seagrass meadows. Conservation International, Arlington, VA.

Intergovernmental Panel on Climate Change (IPCC), 2023. Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva. Available at: https://www.ipcc.ch/report/sixth-assessment-report-synthesis-report.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems 25, 1097–1105.

Kuhlmann, M.L., 2022. Revisão dos índices da comunidade de macroinvertebrados bentônicos aplicados aos rios e reservatórios do estado de São Paulo. CETESB, São Paulo.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521(7553), 436–444.

Li, P., Liu, J., Bai, J., Tong, Y., Meng, Y., Diao, X., Pan, K., Zhu, X., Lin, G., 2022. Community structure of benthic macrofauna and the ecological quality of mangrove wetlands in

Hainan, China. Frontiers in Marine Science 9, 861718. https://doi.org/10.3389/fmars.2022.861718.

McLeod, E., Chmura, G.L., Bouillon, S., Salm, R., Björk, M., Duarte, C.M., Lovelock, C.E., Schlesinger, W.H., Silliman, B.R., 2011. A blueprint for blue carbon: toward an improved understanding of the role of vegetated coastal habitats in sequestering $CO_2$. Frontiers in Ecology and the Environment 9(10), 552–560. https://doi.org/10.1890/110004.

Mizuno, K., Terayama, K., Ishida, S., Godbold, J.A., Solan, M., 2024. Combining three-dimensional acoustic coring and a convolutional neural network to quantify species contributions to benthic ecosystems. Royal Society Open Science 11, 240042.

Moura, R.B., Dalto, A.G., Salloranzo, I.A., Moreira, D.L., Lavrado, H.P., 2023. Community structure of the benthic macrofauna along the continental slope of Santos Basin and São Paulo Plateau, SW Atlantic. Ocean and Coastal Research 71, e23032.

Neves, R.A.F., Valentin, J.L., 2011. Revisão bibliográfica sobre macrofauna bentônica de fundos não-consolidados em áreas costeiras prioritárias para conservação no Brasil. Arquivos de Ciências do Mar 44(3), 59–80. https://doi.org/10.32360/acmar.v44i3.153.

Pardo, J.C.F., Poste, A.E., Frigstad, H., Quintana, C.O., Trannum, H.C., 2023. The interplay between terrestrial organic matter and benthic macrofauna: framework, synthesis, and perspectives. Ecosphere 14, e4508. https://doi.org/10.1002/ecs2.4508.

Sousa, L.K.S., Cutrim, M.V.J., Nogueira Júnior, M., Oliveira, V.M., 2023. Does dredging activity exert an influence on benthic macrofauna in tropical estuaries? Case study on the northern coast of Brazil. Iheringia, Série Zoologia 113, e2023009.

Sun, X., Dong, J., Hu, C., Zhang, Y., Chen, Y., Zhang, X., 2021. Use of macrofaunal assemblage indices and biological trait analysis to assess the ecological impacts of coastal bivalve aquaculture. Ecological Indicators 127, 107713. https://doi.org/10.1016/j.ecolind.2021.107713.

United Nations Environment Programme (UNEP), 2021. Making Peace with Nature: A Scientific Blueprint to Tackle the Climate, Biodiversity and Pollution Emergencies. Nairobi. Available at: https://www.unep.org/resources/making-peace-nature.

Villon, S., Mouillot, D., Chaumont, M., Claverie, T., Villéger, S., 2018. Automatic underwater fish species classification with limited data using few-shot learning. Ecological Informatics 48, 26–34. https://doi.org/10.1016/j.ecoinf.2018.07.006.

Xu, M., Wang, J., Li, F., Chen, X., 2020. Automatic underwater image classification using deep convolutional neural networks. Pattern Recognition Letters 130, 259–265. https://doi.org/10.1016/j.patrec.2019.12.003.

Zhang, S., Qiao, H., Li, H., 2019. Plankton image classification using deep convolutional neural networks. Neurocomputing 335, 215–223.

Zhu, Z., Wood, C.M., Zhang, Y., Ma, Z., Wu, J., Wang, C., Jiang, Z., Yu, H., 2021. Deep learning for ecological monitoring in marine environments: A review. Ecological Informatics 61, 101240. https://doi.org/10.1016/j.ecoinf.2021.101240.