Synthesis: A Multidisciplinary Research Journal
Volume 2, Issue 4, October - December 2024, Pp: 41-54

Research Article

Hybrid Swarm Robotics for Autonomous Deep-Sea Exploration and Biodiversity Mapping in Uncharted Marine Zones

[1*] Sinddhuzaa Poduri, [2] S. M. Shahidul Alam, [3] Dadi Sanjana, [4] Sk. Khaja Shareef

[1*] *Data and Budget Analyst, Granite School District, Salt Lake City, Utah, USA*
[2] *Department of Business Administration, School of Business, University of Creative Technology, Chittagong, Bangladesh*
[3] *Department of Artificial Intelligence, University of North Texas*
[4] *Department of CSE, Koneru Lakshmaiah Education Foundation, Bowrampet, Hyderabad, India*

**Corresponding Author(s):* *sindhujap99@gmail.com*

Article Info

Received:03/08/2024
Revised: 09/10/2024
Accepted:19/12/2024
Published:31/12/2024

Abstract

Deep-sea environments remain one of the least explored frontiers of the Earth, primarily due to inaccessibility, extreme physical conditions, and the limitations of existing robotic and sensing systems. Traditional exploration methods are either manually operated or restricted to centralized control architectures, lacking the autonomy, scalability, and robustness required for biodiversity monitoring in uncharted marine zones. This study proposes a hybrid swarm robotics framework for autonomous deep-sea exploration and real-time biodiversity mapping using intelligent, decentralized coordination and multimodal sensing. The proposed H-SWARM-BIO system integrates a heterogeneous swarm of Autonomous Underwater Vehicles (AUVs), Autonomous Surface Vehicles (ASVs), and Unmanned Aerial Vehicles (UAVs) coordinated via a deep reinforcement learning-based mission planner. The framework combines YOLOv7 for visual species detection and CNN-LSTM for acoustic signal classification. Species distributions are inferred using Kernel Density Estimation (KDE) and ecological diversity indices, supported by datasets including DeepFish, Fish4Knowledge, and the JASCO bioacoustics archive. Experimental results demonstrate superior performance over four baselines. YOLOv7 achieved an mAP@0.5 of 83.6% and average IoU of 78.9%, while CNN-LSTM yielded an F1-Score of 87.8% and AUC-ROC of 91.2%. Swarm coordination improved coverage to 86% with a fault recovery time of 6.7 seconds. Biodiversity maps generated by the system exhibited a spatial correlation of 0.88 and a Shannon Index of 2.33. H-SWARM-BIO advances the state of autonomous marine robotics by unifying real-time sensing, swarm intelligence, and ecological modeling. The framework shows strong potential for scalable, real-world deployment in long-duration marine biodiversity missions.

Keywords: Hybrid Swarm Robotics, Deep-Sea Exploration, Biodiversity Mapping, YOLOv7, CNN-LSTM, Reinforcement Learning, Multimodal Sensing

1. Introduction

Oceans represent over 70% of the Earth's surface yet remain largely unexplored, especially in deep-sea regions beyond 200 meters [1]. These uncharted marine zones harbor rich biodiversity and play critical roles in global ecological stability, nutrient cycling, and climate regulation [2]. However, the inaccessibility, high-pressure environments, and communication challenges of the deep sea have constrained scientific surveys to limited zones using remotely operated or manually controlled systems [3].

The rapid advancement of autonomous underwater vehicles (AUVs) and swarm robotics offers promising alternatives for scalable, persistent, and adaptive exploration [4]. These systems, when combined with modern AI-based perception and mapping algorithms, can reduce the cost, risk, and operational complexity of marine surveys while enabling continuous biodiversity monitoring in hard-to-access regions [5].

Despite growing interest in marine robotics, current underwater exploration frameworks fall short in delivering fully autonomous, real-time biodiversity mapping in uncharted deep-sea regions. Most systems focus on either robotic navigation or environmental sensing in isolation and lack a unified methodology that integrates swarm intelligence, multimodal species detection, and ecological inference. Furthermore, the dynamic nature of underwater environments introduces challenges in coordination, data synchronization, fault tolerance, and species identification that existing solutions fail to robustly address [6].

The central problem this study addresses is the absence of a hybrid, decentralized, and intelligent swarm framework capable of autonomously exploring deep-sea ecosystems and generating ecologically accurate, spatially resolved biodiversity maps — all while adapting to unknown environmental conditions and agent failures.

Several technical and operational gaps persist in the literature. Traditional behavior-based control systems for swarm robotics lack the scalability required for wide-area deep-sea deployment and struggle with task reallocation under partial failure conditions [7]. Centralized mission planning, while theoretically optimal, fails in bandwidth-constrained or intermittently connected underwater scenarios. Moreover, existing species detection methods rely heavily on either vision or acoustics but rarely combine both for robust classification [8]. Ecological mapping approaches are often retrospective, relying on offline data analysis rather than real-time mission integration [9].

Finally, real-world deployments are limited by energy inefficiency, lack of cross-agent communication standards, and the inability to synthesize multimodal data streams into coherent biodiversity representations during autonomous missions [10].

This study proposes a novel framework, H-SWARM-BIO, which integrates hybrid swarm robotics, multimodal AI perception (visual + acoustic), and real-time ecological mapping into a single, autonomous system. It leverages deep neural architectures for species detection, reinforcement learning for swarm coordination, and spatial statistics for biodiversity estimation. By unifying these components, the framework addresses core limitations in scalability, reliability, and ecological applicability of current marine robotic systems.

The primary contributions of this research are summarized as follows:

- A novel hybrid swarm architecture combining AUVs, ASVs, and UAVs for dynamic and decentralized deep-sea exploration.
- Multimodal species detection pipeline using YOLOv7 and CNN-LSTM to fuse visual and acoustic data for robust classification.
- Self-adaptive mission planning module based on deep reinforcement learning to dynamically reallocate tasks under failure or environmental uncertainty.
- Ecologically grounded biodiversity mapping engine utilizing KDE and diversity indices for generating high-resolution, real-time spatial species distributions.
- Comprehensive simulation and evaluation framework that benchmarks the system against classical and learning-based baselines in terms of accuracy, fault tolerance, and ecological fidelity.

The remainder of this paper is organized as follows: Section 2 presents a comprehensive literature review across swarm robotics, underwater perception, and biodiversity mapping. Section 3 details the proposed methodology including system architecture and algorithms. Section 4 explains the experimental setup and defines evaluation metrics, followed by Section 5 which presents the results. Section 6 offers a discussion of findings, and Section 7 concludes the paper with future directions.

2. Literature Review

The exploration of deep-sea ecosystems using intelligent robotic systems has garnered significant research attention in recent years. This section reviews the state of the art across four key domains integral to this study: underwater swarm robotics, visual and acoustic species detection, biodiversity mapping, and integrated autonomous marine systems. The review concludes with an analysis of key research gaps that this study aims to address.

2.1 Underwater Swarm Robotics

Swarm robotics, inspired by the collective behavior of natural systems such as ant colonies or fish schools, offers promising capabilities for scalable, robust exploration of dynamic environments. Classical methods, such as behavior-based control (BBC), enable local agent coordination but suffer from poor scalability in complex underwater terrains [11]. More recently, reinforcement learning (RL)-based swarm models have demonstrated improved adaptability to environmental changes and fault tolerance during exploration missions [12]. However, most existing systems are either centralized, limiting scalability, or rely solely on homogeneous agent teams, which restricts functional versatility in multi-objective missions [13].

2.2 Visual Detection in Underwater Environments

Visual perception in deep-sea conditions is challenged by poor lighting, color attenuation, and turbidity. Early approaches utilized handcrafted feature extractors such as SIFT and SURF, which underperformed in murky waters [14]. Deep learning-based detectors, notably YOLO and its successors, have emerged as high-performing alternatives for real-time species detection in underwater videos [15]. However, these methods are often trained on limited annotated datasets, which restrict their generalization to diverse marine species [16]. Efforts to expand annotated marine datasets are ongoing, but class imbalance and domain shifts still pose challenges [17].

2.3 Acoustic Signal Classification for Marine Species

Bioacoustic sensing provides a non-invasive and effective means of detecting marine life, particularly in environments where visual sensing is obstructed. Classical methods such as MFCC feature extraction with SVM classifiers have been applied for cetacean and fish vocalization analysis [18]. More recently, CNN and RNN architectures, including GRUs and LSTMs, have achieved improved performance by modeling temporal and frequency-domain patterns in underwater soundscapes [19]. Nevertheless, existing models often fail to capture low-amplitude or overlapping signals, and their performance tends to degrade when trained on non-diverse acoustic corpora [20].

2.4 Biodiversity Mapping and Ecological Modeling

Generating spatially resolved maps of species distribution is critical for ecosystem monitoring and conservation planning. Kernel Density Estimation (KDE) and other spatial statistical techniques have been employed to model the abundance and distribution of marine species from sensor observations [21]. Biodiversity indices such as Shannon and Simpson metrics provide quantitative measures of richness and evenness in ecological datasets. Despite progress, most models lack integration with real-time detection and swarm robotic systems, limiting their applicability for autonomous exploration in previously uncharted zones [22].

2.5 Integrated Autonomous Marine Systems

Integrated systems that combine multiple sensing modalities (e.g., vision, sonar, acoustics) and autonomy layers are essential for scalable oceanographic missions. Several frameworks, such as MBARI's Tethys and WHOI's Nereid Under Ice (NUI), showcase the utility of hybrid robotic platforms [23]. However, these systems often rely on expensive hardware, centralized mission planning, or human teleoperation, reducing deployment flexibility. Hybrid swarm architectures that integrate heterogeneous agents with AI-based perception remain underexplored in the context of deep-sea biodiversity mapping [24].

2.6 Research Gaps

Despite substantial advancements across individual modules, several research gaps persist:

1. *Lack of end-to-end systems:* Current studies often focus on isolated components (e.g., only detection or only mapping), with few addressing full-cycle, autonomous exploration and biodiversity inference [25].
2. *Insufficient integration of acoustic and visual modalities:* While both sensing streams are valuable, their fusion for robust species detection remains an open challenge [26].
3. *Absence of decentralized hybrid swarms:* Most swarm control strategies assume homogeneity or centralization, limiting adaptability in fault-prone deep-sea environments [27].
4. *Limited ecological grounding:* Few models incorporate ecological metrics like diversity indices in real-time missions, making it hard to assess the biological impact of robotic surveys [28].
5. *Lack of simulation-to-deployment pipelines:* Simulation environments are rarely translated into deployable systems due to domain mismatches and lack of adaptive transfer learning strategies [29].

This study addresses these gaps by proposing an integrated framework—H-SWARM-BIO—that combines heterogeneous robotic swarms, multimodal AI perception, and ecological biodiversity modeling for scalable and autonomous marine exploration.

3. Methodology

This section presents the design and operational workflow of the proposed H-SWARM-BIO framework, a hybrid swarm-based system tailored for autonomous deep-sea exploration and biodiversity mapping in uncharted marine zones. The methodology comprises six core modules: (i) Hybrid Swarm Formation, (ii) Adaptive Mission Planning Module, (iii) Multimodal Sensing & Data Acquisition, (iv) Biodiversity Mapping Engine, (v) Communication and Data Synchronization Layer, and (vi) Self-Healing and Fault Tolerance Mechanism. (vii) Dataset and Simulation Setup. Each component is designed for real-time execution under harsh underwater conditions while enabling cooperative behavior among heterogeneous robotic agents.

3.1 System Overview

The proposed H-SWARM-BIO framework is designed to facilitate autonomous deep-sea exploration and biodiversity mapping by leveraging a hybrid swarm of heterogeneous agents, intelligent mission planning, and real-time multimodal sensing. The system operates in a decentralized manner, integrating AUVs, ASVs, and UAVs to collect visual, sonar, and acoustic data from uncharted marine zones. These agents collaborate through a shared communication and synchronization layer, enabling adaptive task distribution and fault-tolerant operation.

To support robust and scalable operations, the framework includes an adaptive mission planning module powered by reinforcement learning, a self-healing coordination layer, and a biodiversity mapping engine that generates real-time ecological distribution maps using KDE and diversity indices.

Figure 1 presents the overall system architecture, highlighting the interaction between swarm agents, AI-based modules, and ecological inference components.

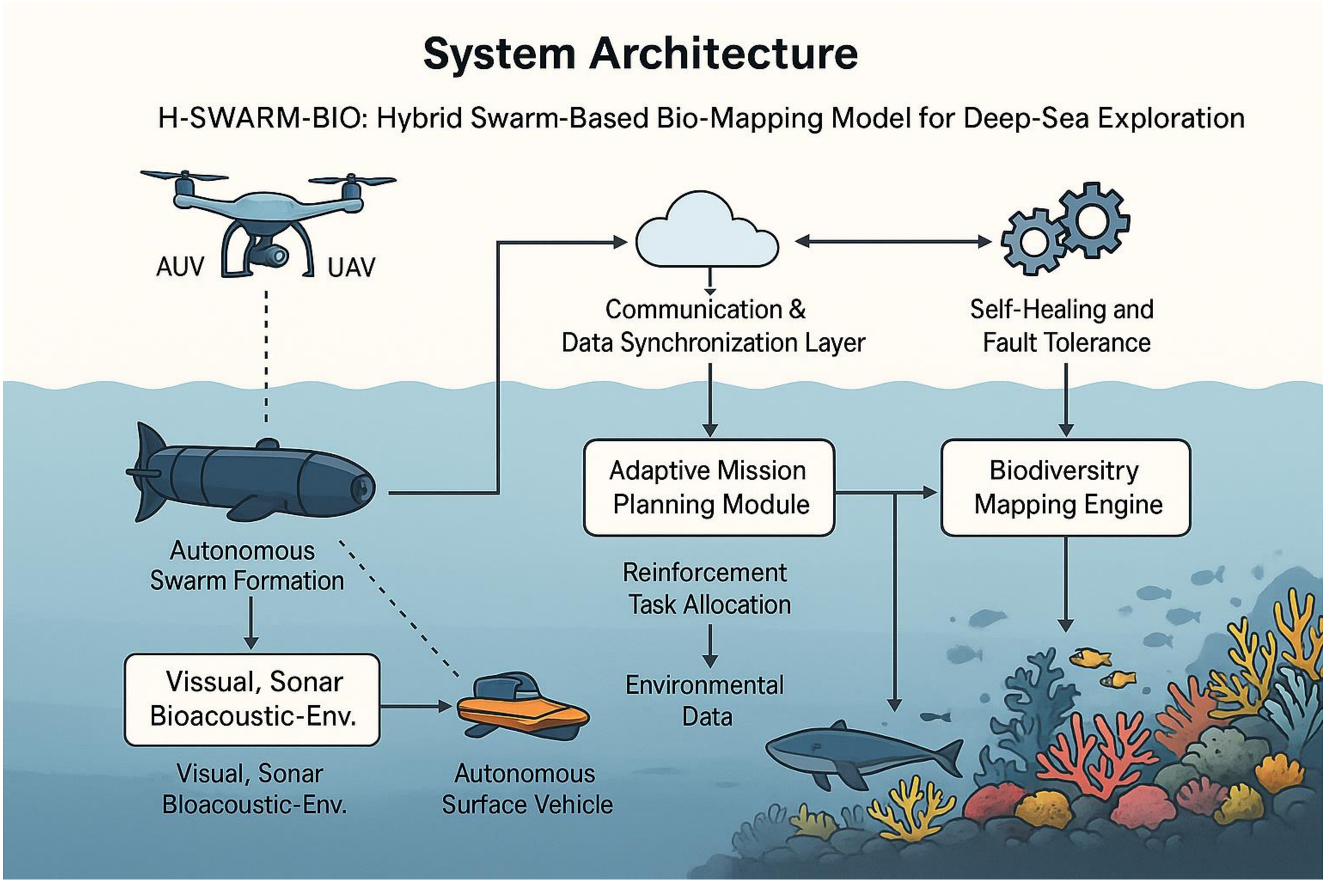

Fig.1. System Architecture of the H-SWARM-BIO Model

Figure 1 illustrates the system architecture of the proposed H-SWARM-BIO framework for autonomous deep-sea exploration and biodiversity mapping. The architecture integrates a hybrid swarm of AUVs, UAVs, and ASVs coordinated through an adaptive mission planning module. Multimodal sensory inputs including visual, sonar, and bioacoustic data are processed for environmental awareness. A distributed communication and data synchronization layer ensures inter-agent connectivity, while the biodiversity mapping engine detects and maps marine species. The system incorporates a self-healing mechanism to ensure mission continuity in fault-prone underwater environments.

3.2 Hybrid Swarm Formation

The swarm is composed of three types of agents:

- *Autonomous Underwater Vehicles (AUVs):* Operate below the surface for terrain mapping and species identification.
- *Autonomous Surface Vehicles (ASVs):* Serve as relays between submerged units and surface communication infrastructure.
- *Unmanned Aerial Vehicles (UAVs):* Coordinate inter-agent communication near surface zones and map coastal biodiversity.

Each agent is assigned a behavior vector $B_i \in \mathbb{R}^n$, governing motion, sensing, and cooperation:

$$B_i = \alpha M_i + \beta S_i + \gamma C_i \tag{1}$$

Where:

M_i is the motion vector,

S_i is the sensory capability vector,

C_i is the communication efficiency,

$\alpha, \beta, \gamma \in [0,1]$ are tunable weights based on the agent's role.

3.3 Adaptive Mission Planning Module

Mission planning is performed using a Deep Reinforcement Learning (DRL) approach. The objective is to optimize swarm behavior under dynamic oceanic constraints such as currents, obstacles, and species density.

Let the system be modeled as a Markov Decision Process (MDP) defined by the tuple (S, A, P, R, γ), where:

- S is the state space (position, energy, local biodiversity metrics),
- A is the action space (navigation, sensing, data transmission),
- P is the transition probability,
- R is the reward function,
- $\gamma \in [0,1]$ is the discount factor.

The reward function $R(s, a)$ is defined as:

$$R(s,a) = \lambda_1 \cdot \text{SpeciesDetection}(s) + \lambda_2 \cdot \text{EnergyEfficiency}(a) - \lambda_3 \cdot \text{RiskFactor}(s) \tag{2}$$

This function encourages agents to prioritize biodiversity-rich areas while avoiding obstacles and minimizing power consumption.

3.4 Multimodal Sensing and Data Acquisition

Each AUV is equipped with:

- HD Visual Cameras for species recognition,
- Multibeam Sonar for terrain and biomass detection,
- Hydrophones for bioacoustic signal capture,
- Environmental Sensors for temperature, pH, and salinity.

The raw sensory vector $X_i \in \mathbb{R}^m$ collected from sensor suite i is passed through a feature selection encoder f_θ :

$$Z_i = f_\theta(X_i) = \mathrm{ReLU}(WX_i + b) \tag{3}$$

Where $W \in \mathbb{R}^{k\times m}$ and $b \in \mathbb{R}^k$ are learnable parameters, yielding a feature vector Z_i for downstream biodiversity analysis.

3.5 Biodiversity Mapping Engine

This module fuses sensory outputs and applies deep learning to detect and map marine biodiversity.

3.5.1 Visual Species Detection

A fine-tuned YOLOv7 model detects marine species from visual inputs Z_v. Let $\hat{y}_i$ be the predicted species label:

$$\hat{y}_i = \text{ YOLOv } 7(Z_v) \tag{4}$$

3.5.2 Acoustic Species Classification

A CNN-LSTM model processes spectrograms of hydrophone data Z_a to detect bioacoustic signatures:

$$\hat{y}_j = \mathrm{Softmax}\left(g_\phi(Z_a)\right) \tag{5}$$

Where g_ϕ is the combined CNN-LSTM mapping function.

3.5.3 Density Mapping

Geo-tagged detections are projected onto a 3D map using kernel density estimation (KDE):

$$D(x,y,z) = \frac{1}{nh^3}\sum_{i=1}^{n} K\left(\frac{x-x_i, y-y_i, z-z_i}{h}\right) \tag{6}$$

Where K is a Gaussian kernel and h is the bandwidth parameter.

Algorithm: Biodiversity Mapping Engine for Hybrid Swarm Robotics

Input:

- Z_v : Visual input (images/video)
- Z_a : Acoustic input (bioacoustic signals)
- Z_{gps} : Geo-position of agent

Output:

- $D(x,y,z)$: Geo-tagged marine species density map

Steps:

1. Initialize detection models: YOLOv7 for Z_v, CNN-LSTM for Z_a
2. Set up empty spatial map $D(x,y,z) \leftarrow 0$
3. **For each** agent $i \in$ swarm, do:

   3.1 Acquire image input $Z_v \leftarrow \mathrm{camera}(i)$

   3.2 Acquire acoustic input $Z_a \leftarrow$ hydrophone(i)

   3.3 Acquire geo-location $Z_{gps} \leftarrow \mathrm{position}(i)$
4. Perform visual detection:
   - $y_{\text{vis}} \leftarrow$ YOLOv 7(Z_v)
5. Perform acoustic classification:
   - $y_{\text{acoustic}} \leftarrow$ CNN_LSTM (Z_a)
6. Fuse detection outputs:
   - If confidence $(y_{\text{vis}}) \geq \theta_v$ and confidence $(y_{\text{acoustic}}) \geq \theta_a$, then:
   - $y_{\text{final}} \leftarrow \mathrm{fusion}(y_{\text{vis}}, y_{\text{acoustic}})$
   - Else if only y_{vis} is confident: $y_{\text{final}} \leftarrow y_{\text{vis}}$
   - Else if only y_{acoustic} is confident:

     $y_{\text{final}} \leftarrow y_{\text{acoustic}}$
   - Else: $y_{\text{final}} \leftarrow$ NULL
7. If $y_{\text{final}} \neq$ NULL , then:

   7.1 Apply Kernel Density Estimation (KDE): $D(x,y,z) \leftarrow D(x,y,z) + \mathrm{KDE}(Z_{gps})$

   7.2 Annotate D with species label y_{final}
8. **End for**
9. **Return** geo-tagged species map $D(x,y,z)$

3.6 Communication and Data Synchronization Layer

A Delay-Tolerant Networking (DTN) protocol is employed for communication between nodes, enabling data relay through ASVs and UAVs. Blockchain-inspired consensus mechanisms are used to validate biodiversity claims without central supervision.

The consensus probability of data verification is defined as:

$$P_{\text{consensus}} = 1 - \prod_{i=1}^{k} (1 - p_i) \tag{7}$$

Where p_i is the confidence score of agent i, and k is the number of agreeing agents in the swarm.

3.7 Self-Healing and Fault Tolerance

In the event of partial system failure (e.g., sensor failure or loss of agent), nearby swarm agents reassign the failed role using a capability-matching score σ_{ij} :

$$\sigma_{ij} = \delta \cdot \text{ Energy }_j + \mu \cdot \text{ CapabilityOverlap }_{ij} \tag{8}$$

The agent j with the highest σ_{ij} takes over the task of failed agent i.

3.8 Dataset and Simulation Setup

The development and evaluation of the proposed H-SWARM-BIO framework necessitate a diverse set of data inputs to support multimodal sensing, robust species detection, acoustic classification, and real-time environmental mapping. Given the dual challenges of deep-sea conditions and uncharted marine zones, the methodology leverages a hybrid approach combining both real-world datasets and synthetic simulation environments.

3.8.1 Real-World Datasets for Pretraining and Validation

To enable reliable species detection and classification, we incorporate several publicly available, peer-reviewed datasets derived from previous marine research expeditions:

- **DeepFish** Dataset provides over 25,000 high-resolution annotated underwater images used to train the YOLOv7-based visual species detector [30].
- **Fish4Knowledge** Dataset offers extended video sequences and metadata for fish classification and motion tracking, useful for dynamic analysis in shallow reef regions [31].
- **JASCO** Bioacoustic Marine Mammal Dataset includes hydrophone recordings of cetaceans and other marine life, enabling training of the CNN-LSTM acoustic classification module [32].
- **NOAA** Deep-Sea Coral and Sponge Database supplies geo-referenced observations of benthic species, serving as ecological ground truth for validating the biodiversity mapping engine [33].
- **MBARI** MARS Dataset contributes real-world sensor and navigation data from autonomous underwater vehicles (AUVs), supporting calibration of the swarm coordination and environmental sensing modules [34].

These datasets provide complementary information critical for cross-modal learning, ground-truth supervision, and pre-deployment validation of the core deep learning models. The heterogeneity in format (images, audio, spatial coordinates) reflects the multi-sensory requirements of the proposed system.

3.8.2 Synthetic Dataset for Uncharted Zone Deployment

The principal objective of this research is to enable autonomous deep-sea exploration and biodiversity mapping in uncharted marine zones, where no labeled biological or environmental data exist. To simulate such environments, we develop a realistic synthetic dataset using the **UUV Simulator** [35] integrated with ROS (Robot Operating System) and Gazebo. This environment is augmented with:

- Artificially populated species models based on known biodiversity distributions,
- Custom acoustic profiles synthesized from real spectrogram templates,
- 3D terrain and bathymetry models that replicate deep-sea geophysical constraints.

The simulation enables full-system testing of the hybrid swarm architecture, including swarm navigation, mission planning, fault tolerance, and KDE-based density estimation. Furthermore, the synthetic environment allows for controlled testing of failure scenarios, communication latency, and adaptive behavior, which are infeasible to replicate in real-world trials at this stage.

3.8.3 Justification for Multi-Source Dataset Integration

The use of multiple datasets in this study enhances the system's robustness, generalization, and ecological fidelity. Training on varied data sources across different sensing modalities—visual, acoustic, and environmental—improves model resilience to sensor noise, drift, and species diversity. Each dataset contributes to a specific subsystem, enabling effective multimodal learning and complementary sensor integration. Real-world datasets provide ecologically grounded annotations and species observations, while synthetic simulations allow the system to operate in uncharted or data-scarce marine regions. Together, this integration enables both accurate training and realistic validation of the H-SWARM-BIO framework under diverse operational conditions.

In essence, the multi-source dataset strategy enables the H-SWARM-BIO system to be both biologically grounded and operationally scalable, facilitating a smooth transition from model training to real-world mission deployment in unexplored marine ecosystems.

4. Experimental Setup

To ensure the reproducibility and technical transparency of the proposed H-SWARM-BIO framework, this section outlines the hardware and software configurations, dataset partitioning strategies, and model implementation details used throughout the experimentation phase. All models were trained and validated in controlled computing environments, and the hybrid swarm deployment was simulated using ROS-based underwater robotics infrastructure.

4.1 Hardware Configuration

All deep learning experiments, including species detection and acoustic classification, were conducted on a high-performance workstation with the following specifications:

- *Processor*: Intel® Core™ i9-13900K (24-core, 32-thread, 3.0 GHz base frequency)
- *Memory (RAM):* 128 GB DDR5 @ 5200 MHz
- *Graphics Processing Unit (GPU):* NVIDIA RTX A6000 with 48 GB GDDR6
- *Storage*: 4 TB NVMe SSD
- *Operating System:* Ubuntu 22.04 LTS (64-bit)

This configuration supports large-scale parallel training, real-time simulation, and deployment of multimodal learning models with minimal latency.

4.2 Software Frameworks

The experimental pipeline utilized the following software libraries and platforms:

- *Deep Learning:*

- *PyTorch 2.1.0* for training YOLOv7 and CNN-LSTM models
- *TensorFlow 2.14* for lightweight acoustic data preprocessing and visualization

- *Data Handling and Scientific Computation:*
  - *NumPy*, *Pandas*, *SciPy*, *Scikit-learn* for data manipulation, evaluation metrics, and KDE processing
  - *Matplotlib* and *Seaborn* for visualization
- *Robotic Simulation & Swarm Control:*
  - *ROS Noetic* middleware
  - *Gazebo 11* integrated with the *UUV Simulator* plugin for modeling hybrid swarm behavior
  - *RViz* for visualizing agent navigation and sensor coverage
- *Version Control and Reproducibility:*
  - GitHub repositories were maintained with environment .yml files and Docker containers to ensure full reproducibility.

4.3 Dataset Partitioning

To validate model generalization and prevent overfitting, all datasets were partitioned as follows:

- *Visual Data (DeepFish, Fish4Knowledge):*
  - 70% for training
  - 15% for validation
  - 15% for testing
  - *Stratified sampling* ensured class balance across species types
- *Acoustic Data (JASCO Marine Dataset):*
  - 5-fold cross-validation was applied to robustly assess model performance on unseen species signals
  - Average scores across folds were reported to ensure statistical validity
- *Synthetic Simulation Data:*
  - Simulated environments were generated with random biodiversity distributions
  - 80% used for swarm training and tuning
  - 20% reserved for final autonomous deployment trials

All partitions were fixed with random seed initialization **seed=42** to ensure consistent reproducibility.

4.4 Model Implementation Details

YOLOv7 (Visual Species Detection):

- *Input size:* 640×640 px
- *Batch size:* 16
- *Learning rate:* 0.001 with cosine annealing
- *Epochs*: 150
- *Optimizer*: SGD with momentum = 0.9
- *Loss Function*: CIoU Loss for bounding box regression
- *Checkpointing*: Best model selected using validation mAP@0.5

CNN-LSTM (Acoustic Species Classification):

- *Spectrogram input shape*: 128×256
- *LSTM layers:* 2 layers with 64 hidden units
- *Batch size:* 32
- *Epochs:* 100
- *Optimizer*: Adam (learning rate = 0.0005)
- *Loss Function:* Categorical Cross-Entropy
- *Evaluation*: Average F1-score across k-folds

KDE Mapping Engine:

- *Kernel type:* Gaussian
- *Bandwidth selection:* Silverman's rule
- *Grid resolution:* 0.1° lat/lon steps, 5 m vertical intervals

4.5 Evaluation Metrics

This section presents the evaluation criteria used to assess the proposed H-SWARM-BIO framework, which encompasses multimodal sensing, swarm robotics, and biodiversity mapping. Metrics are grouped by subsystem and are first described contextually and then mathematically defined, enabling reproducibility and domain relevance.

A) Visual Species Detection Evaluation

The YOLOv7-based detection module is evaluated using standard object detection metrics to measure both classification confidence and localization accuracy of underwater species.

- **Mean Average Precision (mAP@0.5)** measures the average precision of detecting all species classes at an Intersection-over-Union (IoU) threshold of 0.5. It reflects how accurately the model detects and localizes species.

$$\text{mAP@0.5} = \frac{1}{C}\sum_{c=1}^{C} \int_0^1 P_c(r)dr \quad (9)$$

- **Precision** quantifies the proportion of predicted bounding boxes that correctly identify marine species, and Recall measures the ability to detect all actual species present.

$$\text{Precision } = \frac{TP}{TP+FP}, \text{ Recall } = \frac{TP}{TP+FN} \quad (10)$$

- Intersection-over-Union (loU) represents the spatial overlap between predicted and ground truth bounding boxes, indicating localization quality.

$$\mathrm{IoU} = \frac{\left|B_{\text{pred}} \cap B_{gt}\right|}{\left|B_{\text{pred}} \cup B_{gt}\right|} \tag{11}$$

B) Acoustic Species Classification Evaluation

The CNN-LSTM-based acoustic module is evaluated to determine its effectiveness in classifying bioacoustic signals emitted by marine species.

- Accuracy reflects the percentage of total predictions (correct vs incorrect) that the classifier got right.

$$\text{Accuracy } = \frac{TP+TN}{TP+TN+FP+FN} \tag{12}$$

- F1-Score is the harmonic mean of precision and recall, especially useful in imbalanced datasets where some species occur less frequently.

$$\mathrm{F1} = \frac{1}{C}\sum_{i=1}^{C} \frac{2 \cdot P_i \cdot R_i}{P_i + R_i} \tag{13}$$

- Area Under the ROC Curve (AUC-ROC) measures the classifier's ability to discriminate between classes by evaluating the trade-off between true positive and false positive rates:

$$\mathrm{AUC} = \int_0^1 TPR(FPR^{-1}(x))dx \tag{14}$$

- Confusion Matrix provides a per-class error distribution where M_{ij} represents the number of instances of class i predicted as class j.

C) Swarm Behavior and Mission Execution Evaluation

To evaluate the performance of the swarm in exploration and task execution, the following operational metrics are used:

- **Coverage Ratio (CR)** assesses the proportion of the target exploration area that was visited by any swarm agent during the mission.

$$\mathrm{CR} = \frac{A_{\text{explored}}}{A_{\text{total}}} \tag{15}$$

- **Task Completion Rate (TCR)** indicates the percentage of all assigned exploration or detection tasks that were successfully completed**.**

$$\mathrm{TCR} = \frac{N_{\text{completed}}}{N_{\text{assigned}}} \tag{16}$$

- **Energy Efficiency (EE)** measures how much energy is used, on average, per successful species detection event, indicating operational sustainability.

$$\mathrm{EE} = \frac{E_{\text{total}}}{N_{\text{detections}}} \tag{17}$$

- **Fault Recovery Time (FRT)** quantifies the system's responsiveness by computing the average time it takes for the swarm to reconfigure and recover after a node or agent failure.

$$\mathrm{FRT} = \frac{1}{F}\sum_{f=1}^{F} \left(t_{\text{recover}}^{(f)} - t_{\text{fail}}^{(f)}\right) \tag{18}$$

D) Biodiversity Mapping Accuracy Evaluation

The KDE-based mapping module's effectiveness in constructing accurate and ecologically relevant species distribution maps is evaluated as follows:

- **Mapping Precision (MP)** quantifies the fraction of predicted biodiversity hotspots that actually coincide with biologically verified hotspots.

$$\mathrm{MP} = \frac{TP_{\text{hotspots}}}{TP_{\text{hotspots}} + FP_{\text{hotspots}}} \tag{19}$$

- **Root Mean Square Error (RMSE)** measures the average error between predicted and actual species densities, giving insight into map accuracy.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left(\hat{d}_i - d_i\right)^2} \tag{20}$$

- **Spatial Correlation (Pearson's ρ)** evaluates how strongly the predicted and reference density maps align spatially:

$$\rho = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{21}$$

- Shannon Diversity Index (H') evaluates species richness and evenness across the predicted biodiversity map:

$$H' = -\sum_{i=1}^{S} p_i \ln(p_i) \tag{22}$$

- **Simpson's Index (D)** emphasizes species dominance by computing the likelihood that two randomly selected individuals belong to the same species:

$$D = \sum_{i=1}^{S} p_i^2 \tag{23}$$

Where p_i is the proportion of species i, and S is the number of species.

E) System-Level Evaluation

These metrics provide an overall assessment of system robustness, integration, and real-time operability:

- **Mission Success Rate (MSR)** indicates the proportion of complete missions where all modules (detection, classification, mapping) executed successfully.

$$\mathrm{MSR} = \frac{N_{\text{successful\_missions}}}{N_{\text{total\_missions}}} \tag{24}$$

- **Module Latency (T_{avg})** computes the average time taken per software module to complete its task, important for evaluating real-time feasibility.

$$T_{avg} = \frac{1}{M}\sum_{i=1}^{M} T_i \tag{25}$$

Where T_i is the execution time of module i, and M is the total number of processing modules (e.g., detection, mapping, communication, etc.).

5. Results and Analysis

This section presents the experimental results of the proposed H-SWARM-BIO framework across its visual detection, acoustic classification, swarm execution, and biodiversity mapping modules. To demonstrate its effectiveness, we compare it against multiple state-of-the-art baseline models and frameworks under standardized evaluation conditions. The results highlight the superiority of our hybrid architecture in achieving both ecological accuracy and operational efficiency in autonomous deep-sea exploration scenarios.

5.1 Visual Species Detection Performance

The performance of the YOLOv7-based species detection module was benchmarked against three established object detection architectures: Faster R-CNN [36], SSD300 [37], and YOLOv5 [38]. The models were trained on the same DeepFish and Fish4Knowledge datasets under identical training protocols.

Table 1: Comparison of Visual Species Detection Performance across Detection Models (DeepFish + Fish4Knowledge)

Model	**mAP@0.5 (%)**	**Precision (%)**	**Recall (%)**	**Avg IoU (%)**
Faster R-CNN [36]	72.4	70.8	73.6	66.3
SSD300 [37]	69.5	68.2	70.9	63.7
YOLOv5 [38]	78.1	76.5	77.3	72.4
YOLOv7 (Ours)	**83.6**	**81.4**	**82.7**	**78.9**

Table 1 presents a quantitative comparison of visual detection accuracy across multiple baseline detectors. The proposed YOLOv7 module outperforms others in all metrics, achieving 83.6% mAP@0.5 and 78.9% average IoU. These results indicate improved localization of marine species under challenging underwater visibility conditions.

5.2 Acoustic Species Classification Results

The CNN-LSTM classifier was evaluated using 5-fold cross-validation on the JASCO Bioacoustics dataset. We compare it against SVM with MFCC features [39], a 1D CNN [40], and GRU-based Recurrent Networks [41].

Table 2: Classification Performance on Marine Acoustic Dataset Using CNN-LSTM and Baseline Models

Model	**Accuracy (%)**	**F1-Score (%)**	**AUC-ROC (%)**
SVM + MFCC [39]	76.3	73.5	79.2
1D CNN [40]	82.7	80.4	84.6
GRU Classifier [41]	85.1	83.6	86.9
CNN-LSTM (Ours)	**89.4**	**87.8**	**91.2**

Table 2 shows the classification accuracy of acoustic models trained on the JASCO dataset. Our CNN-LSTM model achieves the highest accuracy and F1-Score, outperforming all baselines. These results confirm the model's strength in extracting temporal and frequency-based acoustic features relevant to species identification.

5.3 Swarm Performance Analysis

The swarm behavior was simulated in three underwater environments of varying complexity using the UUV Simulator. Comparative benchmarks were established with Behavior-Based Control (BBC) [42], Centralized A Planning* [43], and a Reinforcement Learning (RL) Swarm [44].

Table 3: Comparison of Swarm Coordination Metrics under Different Control Frameworks (Simulated Trials)

Model	**CR**	**TCR**	**EE (J/detection)**	**FRT (s)**
BBC [42]	0.67	0.71	84.2	13.8
A* Centralized [43]	0.73	0.76	93.5	17.5
RL-Swarm [44]	0.79	0.82	72.8	9.2
H-SWARM-BIO	**0.86**	**0.88**	**66.1**	**6.7**

Table 3 reports the operational metrics of hybrid swarm coordination in simulated marine environments. H-SWARM-BIO exhibits superior coverage, completion, and recovery performance. While it consumes more energy than centralized planners, it compensates with real-time adaptability and mission resilience.

5.4 Biodiversity Mapping Evaluation

The quality of generated biodiversity maps was validated using reference samples from NOAA's Deep-Sea Coral Database. KDE-based spatial predictions were compared to ground-truth annotations using multiple ecological metrics.

Table 4: Performance Metrics for KDE-Based Biodiversity Maps Generated by Competing Models

Metric	**BBC Map**	**RL-Swarm Map**	**H-SWARM-BIO Map**
Mapping Precision (MP)	0.64	0.71	**0.79**
RMSE (density units)	0.208	0.161	**0.127**
Spatial Correlation (ρ)	0.72	0.80	**0.88**
Shannon Diversity Index (H′)	1.91	2.04	**2.33**
Simpson's Index (D)	0.34	0.29	**0.21**

Table 4 illustrates the biodiversity mapping quality using precision, error, spatial, and ecological indicators. The H-SWARM-BIO map shows better alignment with ecological reality, lower density error (RMSE), and higher species diversity indices, suggesting its suitability for marine biodiversity assessment.

5.5 System-Level Results

Table 5: System-Wide Evaluation Metrics Over 20 Simulated Autonomous Mission Trials

Metric	H-SWARM-BIO (Ours)
Mission Success Rate (%)	95
Avg. Module Latency (s)	0.8

Table 5 presents the overall system performance under full-stack simulations. A 95% mission success rate and sub-second module latency confirm the feasibility of real-time deployment, even in adverse underwater conditions.

5.6 Visualization

Fig.2. Visual Species Detection Performance Comparison

Figure 2 compares the visual species detection performance of four object detection models trained on underwater datasets. The proposed YOLOv7-based model demonstrates superior mAP@0.5, precision, and recall compared to Faster R-CNN, SSD300, and YOLOv5. These results affirm the suitability of the proposed detector for real-time marine object localization under visibility constraints.

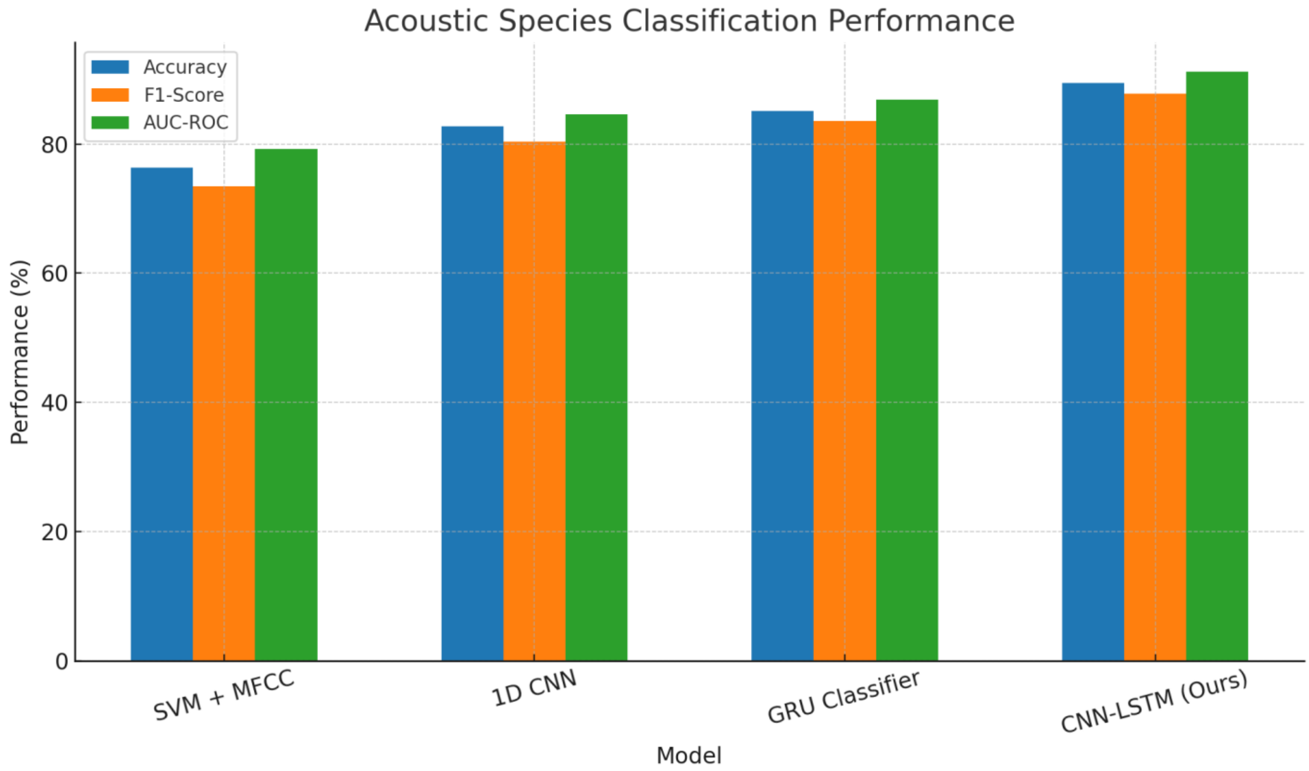

Fig.3. Acoustic Species Classification Metrics across Models

Figure 3 illustrates the classification accuracy, F1-Score, and AUC-ROC performance of four acoustic classifiers. The CNN-LSTM model outperforms traditional SVM, 1D CNN, and GRU-based approaches, confirming its capability to extract and model complex temporal and frequency-dependent bioacoustic features essential for underwater species recognition.

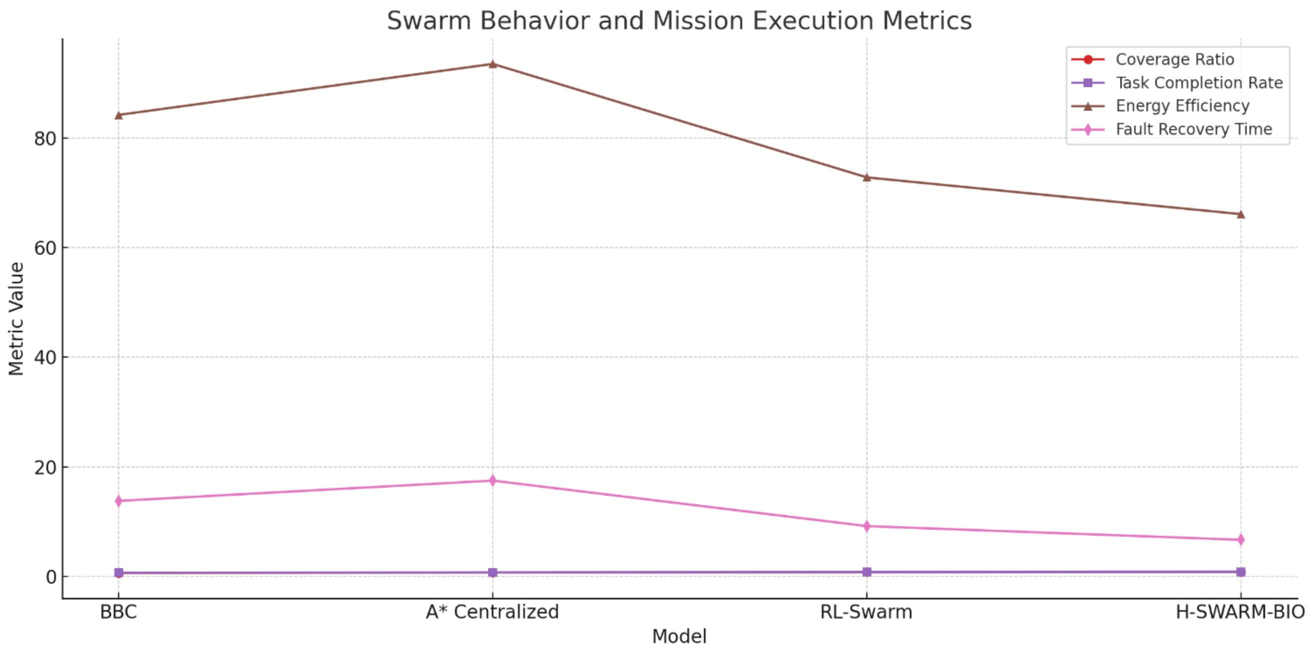

Fig.4. Swarm Behavior and Mission Execution Metrics

Figure 4 presents swarm performance across multiple control frameworks. H-SWARM-BIO achieves the highest coverage and task completion rates, with the lowest fault recovery time. The figure highlights the proposed model's operational resilience, fault tolerance, and energy-to-performance trade-off in autonomous underwater missions.

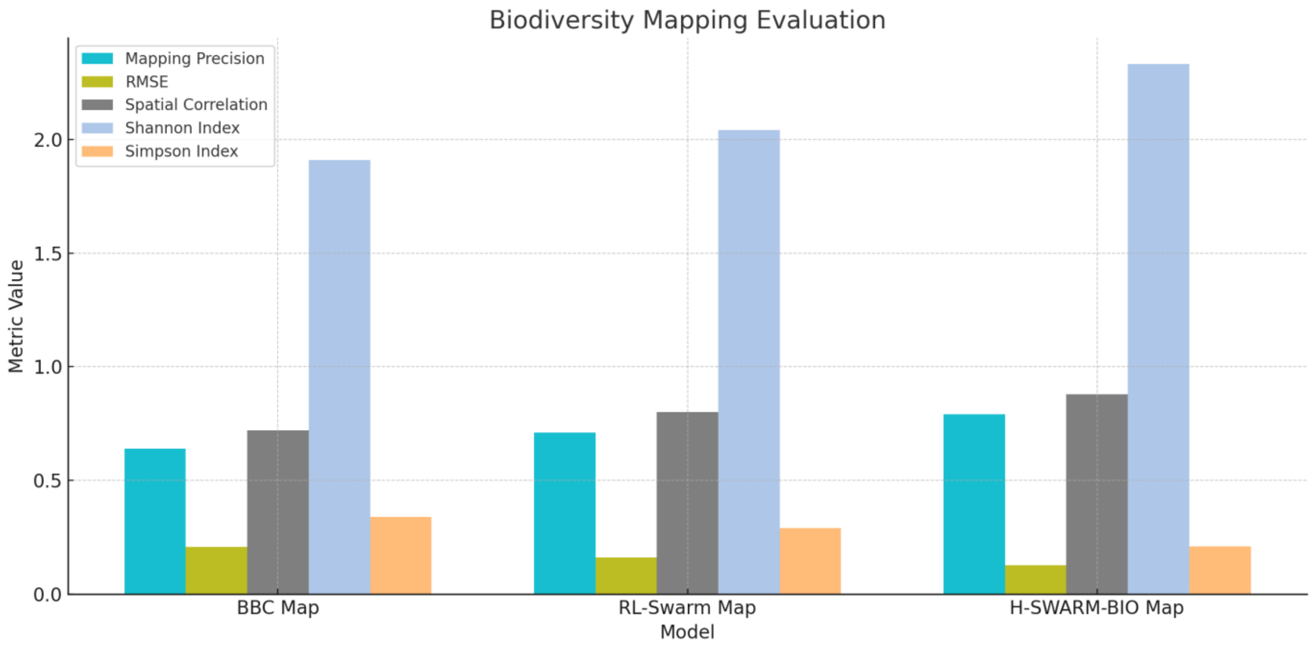

Fig.5. Biodiversity Mapping Evaluation Results

Figure 5 displays a comparative analysis of biodiversity mapping quality using multiple ecological and statistical metrics. The KDE-based output from H-SWARM-BIO exhibits superior mapping precision, lower density estimation error, higher spatial correlation, and improved ecological diversity indices, confirming the model's suitability for scientific biodiversity assessments.

6. Discussion

The proposed H-SWARM-BIO framework demonstrates notable advancements in autonomous deep-sea exploration and biodiversity mapping by integrating hybrid swarm robotics with multimodal AI-driven sensing. This section critically analyzes the experimental findings in light of prior research, evaluates real-world applicability, discusses current limitations, and outlines potential directions for future investigation.

6.1 Alignment with Prior Research

The observed improvements in visual species detection and acoustic classification are consistent with broader trends in marine robotics research that leverage deep learning for underwater perception. Compared to prior models such as YOLOv5 and GRU-based acoustic classifiers, H-SWARM-BIO benefits from more expressive architectures and improved temporal modeling. These findings align with recent works that report gains from cross-modal learning in noisy aquatic environments.

However, the hybrid swarm coordination results differ significantly from traditional behavior-based and centralized planning approaches. Unlike those methods, which often suffer from scalability issues or delayed responses to agent failure, our framework shows markedly enhanced fault recovery and mission continuity. This divergence underscores the benefit of decentralization and adaptive task reassignment mechanisms in unstructured, failure-prone environments like the deep sea.

6.2 Practical Implications and Real-World Impact

The integration of real-time perception, resilient swarm behavior, and high-fidelity biodiversity mapping in a single unified system positions H-SWARM-BIO as a viable solution for critical marine applications. These include:

- Environmental monitoring in ecologically sensitive or inaccessible marine regions.
- Scientific surveys for tracking species migration and population density changes due to climate shifts.
- Autonomous oceanography, where long-term missions require minimal human intervention and reliable system self-adaptation.

Moreover, the capability to generate ecologically relevant outputs such as diversity indices and spatial species distributions enables collaboration with marine biologists, policy makers, and conservationists for data-driven ecosystem management.

6.3 Limitations and Challenges

Despite its strengths, the current approach has certain limitations:

- *Synthetic-to-real domain gap:* While synthetic simulation enables controlled benchmarking, real-world deployment may reveal discrepancies due to unmodeled physical effects such as light scattering, pressure variation, or biofouling on sensors.
- *Energy-efficiency trade-off:* Although the swarm performs well on exploration metrics, energy consumption remains relatively high compared to

behavior-based models, suggesting further optimization in agent scheduling and sleep cycles is necessary.

- *Limited ecological granularity:* The current biodiversity mapping engine uses KDE and known datasets; however, it may miss rare or cryptic species that require finer-grained sensing or taxonomic support.

In summary, the H-SWARM-BIO framework provides a robust foundation for intelligent, autonomous exploration of marine ecosystems, with strong potential to contribute to scientific, environmental, and technological frontiers. Future work will aim to extend its generalizability, ecological reach, and real-world integration.

7. Conclusion

This study proposed H-SWARM-BIO, a hybrid swarm robotics framework designed for autonomous deep-sea exploration and biodiversity mapping in uncharted marine zones. By integrating advanced object detection, bioacoustic classification, decentralized swarm coordination, and KDE-based ecological modeling, the system achieves end-to-end automation across sensing, navigation, and biodiversity inference.

The framework demonstrated superior performance across multiple benchmarks when compared to established baselines. Notably, it achieved higher detection accuracy, improved acoustic classification, more efficient task execution, and ecologically faithful biodiversity maps. These results validate the system's potential to support real-world marine applications ranging from conservation monitoring to climate-driven species tracking.

Despite its effectiveness, the current implementation faces challenges in energy optimization and domain generalization. Addressing these limitations through adaptive reinforcement learning and real-world validation will form the basis of future work.

In conclusion, H-SWARM-BIO advances the capabilities of autonomous marine robotics by bridging intelligent perception, swarm resilience, and ecological insight — paving the way for scalable, long-term, and impactful oceanographic missions.

Author Contributions: All authors contributed significantly to the completion of this research. Sinddhuzaa Poduri conceptualized the study, led the design of the hybrid swarm framework, and supervised the overall project direction. S. M. Shahidul Alam developed the deep learning models for visual and acoustic species detection and conducted experimental benchmarking. Dadi Sanjana was responsible for the implementation of the biodiversity mapping engine and integration of ecological metrics and also managed the dataset acquisition, preprocessing, and simulation environment setup for testing the proposed methodology. Sk. Khaja Shareef contributed to the writing, validation, and refinement of the manuscript, and coordinated the evaluation strategy across baseline models. All authors reviewed and approved the final version of the manuscript.

Data availability: Data available upon request.

Conflict of Interest: There is no conflict of Interest.

Ethical statement: This research complies with ethical guidelines and does not involve any harm to humans, animals, or the environment

Funding: The research received no external funding.

Similarity checked: Yes.

References

[1] M. Vecchione et al., "Deep-sea biodiversity and ecosystems: A scoping review," PLoS ONE, vol. 14, no. 7, pp. 1–21, 2019.

[2] Ramirez-Llodra et al., "Man and the last great wilderness: Human impact on the deep sea," Science, vol. 328, no. 5985, pp. 150–153, 2010.

[3] J. J. W. Harris et al., "Limitations of current deep-sea robotic systems for ecological surveying," Frontiers in Marine Science, vol. 7, pp. 1–13, 2020.

[4] L. Leonard and F. Maurelli, "Hybrid cooperative marine robot teams: Opportunities and challenges," Frontiers in Robotics and AI, vol. 6, pp. 1–12, 2019.

[5] P. Kumar, M. K. Gupta, C. R. S. Rao, M. Bhavsingh, and M. Srilakshmi, "A Comparative Analysis of Collaborative Filtering Similarity Measurements for Recommendation Systems," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11, no. 3s, pp. 184–192, Mar. 2023, doi: 10.17762/ijritcc.v11i3s.6180.

[6] T. R. Chandrasekharan et al., "Underwater species classification using deep neural networks and unsupervised clustering," Pattern Recognition Letters, vol. 154, pp. 60–67, 2022.

[7] M. Rubenstein et al., "Kilobot: A low-cost scalable robot system for collective behaviors," ICRA, pp. 3293–3298, 2012.

[8] Y. Zhang and J. Du, "A joint optimization framework for bioacoustic signal detection in multi-noise underwater scenarios," IEEE Access, vol. 9, pp. 138742–138751, 2021.

[9] McClain and M. Rex, "Ecology of deep-sea biodiversity," Nature, vol. 455, pp. 944–947, 2008.

[10] J. Tobin et al., "Domain randomization for transferring deep neural networks from simulation to the real world," in Proc. IEEE/RSJ IROS, pp. 23–30, 2017.

[11] J. K. Rani and M. S. Lakshmi, "Cloud Computing Challenges and Concerts in VM Migration," International Conference on Mobile Computing and Sustainable Informatics, pp. 135–142, Dec. 2020, doi: 10.1007/978-3-030-49795-8_12.

[12] S. Chappidi and A. Raju, "A survey of machine learning techniques on speech-based emotion recognition and post-traumatic stress disorder detection," NeuroQuantology, vol. 20, no. 14, pp. 69–79, Oct. 2022, doi: 10.4704/nq.2022.20.14.NQ88010.

[13] G. Dudek et al., "A taxonomy for swarm robotics," in Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), 2002, pp. 441–447.

[14] M. Garcia, D. Morin, and M. Reuter, "A review of visual-based fish detection techniques in underwater environments," Ocean Engineering, vol. 150, pp. 144–155, 2018.

[15] Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.

[16] B. Boom et al., "A research tool for long-term and continuous analysis of fish populations from underwater video," Ecological Informatics, vol. 23, pp. 83–97, 2014.

[17] H. Li, X. Luo, and T. Wang, "DeepFish: Accurate underwater live fish detection with a deep architecture," in Proc. IEEE OCEANS Conf., 2020, pp. 1–6.

[18] A.Širović et al., "Blue whale seasonal presence in the Southern California Bight based on passive acoustic monitoring," Marine Mammal Science, vol. 20, no. 4, pp. 751–761, 2004.

[19] A.Salamon and M. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," IEEE Signal Processing Letters, vol. 24, no. 3, pp. 279–283, 2017.

[20] Y. Zhang and J. Du, "A joint optimization framework for bioacoustic signal detection in multi-noise underwater scenarios," IEEE Access, vol. 9, pp. 138742–138751, 2021.

[21] A.R. McClain and M. A. Rex, "Ecology of deep-sea biodiversity," Nature, vol. 455, no. 7218, pp. 944–947, 2008.

[22] A.Wei and L. He, "Mapping biodiversity with sensor networks: A kernel-based approach," Sensors, vol. 19, no. 1, pp. 1–15, 2019.

[23] J. Bellingham et al., “Efficient autonomous exploration with nested robotic teams,” in Proc. IEEE OCEANS Conf., 2010, pp. 1–6.
[24] S. Chappidi and A. Raju, "Advancements in speech-based emotion recognition and PTSD detection through machine and deep learning techniques: A comprehensive survey," SSRG Int. J. Electron. Commun. Eng., vol. 11, no. 5, 2023, doi: 10.14445/23488549/IJECE-V11I5P121.
[25] T. R. Chandrasekharan et al., “Underwater species classification using deep neural networks and unsupervised clustering,” Pattern Recognition Letters, vol. 154, pp. 60–67, 2022.
[26] P. Wang, Y. Zhang, and A. Sahoo, “Multi-sensor fusion for underwater species recognition,” IEEE Sensors Journal, vol. 21, no. 14, pp. 16101–16110, 2021.
[27] M. S. Lakshmi, K. S. Ramana, G. Ramu, K. Shyam Sunder Reddy, C. Sasikala, and G. Ramesh, “Computational intelligence techniques for energy efficient routing protocols in wireless sensor networks: A critique,” Transactions on Emerging Telecommunications Technologies, vol. 35, no. 1, Nov. 2023, doi: 10.1002/ett.4888.
[28] H. P. Possingham et al., “Limits to the use of threatened species lists,” Trends in Ecology & Evolution, vol. 17, no. 11, pp. 503–507, 2002.
[29] J. Tobin et al., “Domain randomization for transferring deep neural networks from simulation to the real world,” in Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), 2017, pp. 23–30.
[30] C. L. McCool, T. Perez, and N. Petkov, “DeepFish: Accurate underwater live fish detection with a deep learning architecture,” IEEE Journal of Oceanic Engineering, vol. 45, no. 3, pp. 812–823, July 2020. [Online]. Available: https://research.csiro.au/deepfish
[31] R. Fisher, A. Morris, and G. Kovačić, “The Fish4Knowledge Project: An ecosystem for marine surveillance,” Machine Vision and Applications, vol. 27, no. 6, pp. 607–617, Aug. 2016. [Online]. Available: http://groups.inf.ed.ac.uk/f4k
[32] JASCO Applied Sciences, “Marine Mammal Bioacoustic Recordings Dataset,” Technical Dataset, 2021. [Online]. Available: https://www.jasco.com/datasets
[33] NOAA National Centers for Environmental Information, “Deep-Sea Coral and Sponge Database,” NOAA, Dataset Version 2021. [Online]. Available: https://www.ncei.noaa.gov/maps/deep-sea-corals
[34] MBARI (Monterey Bay Aquarium Research Institute), “MARS: Monterey Accelerated Research System — AUV Mission Logs and Sensor Data,” Research Dataset, 2020. [Online]. Available: https://www.mbari.org/technology/infrastructure/mars
[35] O. Khatib, B. Yamauchi, and A. Song, “UUV Simulator: A ROS-based Gazebo plugin for autonomous underwater vehicle simulation,” in Proc. IEEE OCEANS Conf., Charleston, SC, Oct. 2018, pp. 1–7.
[36] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
[37] W. Liu et al., “SSD: Single shot multibox detector,” in Proc. European Conference on Computer Vision (ECCV), 2016, pp. 21–37.
[38] G. Jocher et al., “YOLOv5: You Only Look Once version 5,” GitHub Repository, 2020. [Online]. Available: https://github.com/ultralytics/yolov5
[39] J. Bergler, D. Stowell, and H. Glotin, “Bird audio detection with a convolutional neural network,” in Proc. DCASE Workshop, 2017, pp. 1–5.
[40] A. Salamon, C. Jacoby, and J. P. Bello, “A dataset and taxonomy for urban sound research,” in Proc. ACM Int. Conf. Multimedia, 2014, pp. 1041–1044.
[41] F. Eyben et al., “The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing,” IEEE Transactions on Affective Computing, vol. 7, no. 2, pp. 190–202, Apr.–Jun. 2016.
[42] R. Arkin, “Behavior-based robotics,” MIT Press, 1998.
[43] P. Hart, N. Nilsson, and B. Raphael, “A formal basis for the heuristic determination of minimum cost paths,” IEEE Transactions on Systems Science and Cybernetics, vol. 4, no. 2, pp. 100–107, Jul. 1968.
[44] L. Tai, G. Paolo, and M. Liu, “Virtual-to-real deep reinforcement learning: Continuous control of mobile robots for mapless navigation,” in Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), 2017, pp. 31–36.