International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

An Autonomous Phytoplankton Detection and Ecological Risk Assessment using Oriented Object Detection and Hybrid Deep Learning

Author Details:

S.D.Jagadiish[1], Angel Hepzibah R[2], Kaliappan M[3], Mariappan E[4]

[1] UG Student, Department of AI & Data Science, Ramco Institute of Technology, Rajapalayam, India

[2] Assistant Professor - II, Department of AI & Data Science, Ramco Institute of Technology, Rajapalayam, India

[3] Professor, Department of AI & Data Science, Ramco Institute of Technology, Rajapalayam, India

[4] Associate Professor, Department of AI & Data Science, Ramco Institute of Technology, Rajapalayam, India

Corresponding Author Email: jagadiish21@gmail.com

https://doi.org/10.55041/ijsmt.v2i3.163

Cite this Article: S.D.Jagadiish, (2026). An Autonomous Phytoplankton Detection and Ecological Risk Assessment using Oriented Object Detection and Hybrid Deep Learning. International Journal of Science, Strategic Management and Technology, *02*(03). https://doi.org/10.55041/ijsmt.v2i3.163

ABSTRACT

Aquatic ecosystems depend greatly on microscopic phytoplankton species as the primary producers and bio-indicators of water quality. Accurate detection and identification of phytoplankton species is a necessary step for monitoring their ecological risk and the following in-depth ecological and environmental studies and analyses, due to their significance and potential indication of environmental stress, eutrophication and habitat change, and HABs incidents. Yet, microscopic image-based phytoplankton species identification is labor-intensive and time-consuming. It requires expert taxonomic skills which make large-scale biodiversity screening an unthinkable task using traditional approaches. This paper proposes a novel automated system that integrates an Oriented YOLOv8-OBB object detection engine with a hybrid ConvNeXt–CBAM–Vision Transformer (ViT) classifier to tackle some fundamental challenges for microscopic phytoplankton species identification, including rotational-variation and small-object detection, background noise from other particles floating near the target organisms, and subtle morphological difference among species. For detection, an anchor-free oriented bounding box regression module was adopted to preserve the 360° rotational-invariance of the target organisms, minimizing errors of background inclusion of organisms rotating diagonally. After extracting the contour-integrating organism patches, a triple-hybrid deep learning classifier was deployed, by cascading a feature extractor ConvNeXt, an attention model CBAM, and a global relational aggregator ViT into the pipeline. The trained detection and classification models are validated over 23,352 microscopic images with various magnification factors, obtaining an accuracy of 94.2%, detection mean Average Precision (mAP) of 0.89, and test latency of 34.7 ms/image. Finally, a rule engine-based Ecological Risk Assessment (ERA) is devised to generate timely alarm by converting identified taxa to a Saprobic Index.

KEYWORDS

Phytoplankton Detection, Oriented YOLOv8-OBB, ConvNeXt, CBAM Attention, Vision Transformer (ViT), Small Object Detection, Environmental AI, Ecological Risk Assessment, Saprobic Index, Automated Microscopic Analysis.

I. INTRODUCTION

With the rapid advancement of digital image processing and artificial intelligence, environmental monitoring systems have significantly improved. In marine ecosystems, microscopic plankton species serve as key bio-indicators for assessing water quality, biodiversity, and ecological health. Variations in plankton populations often signal ecological changes such as migration or algal blooms. However, traditional identification using manual microscopy is time-consuming, labor-intensive, and prone to human error. Additionally, plankton exhibit complex morphological variations, small sizes, and arbitrary orientations, making their analysis more challenging. Automated plankton recognition has therefore become essential for large-scale ecological studies. Conventional methods based on handcrafted features and classical machine learning often fail to generalize in complex imaging conditions, especially due to high visual similarity among species. To address these challenges, this work proposes a hybrid deep learning framework for accurate plankton detection and classification. The system integrates oriented object detection with a hybrid ConvNeXt–Vision Transformer architecture enhanced by a CBAM attention module. Initially, microscopic plankton are localized using oriented bounding box detection to handle rotational variations. The detected regions are then classified using a deep neural network that captures both local and global features. Transfer learning, data augmentation, and class balancing techniques are employed to improve robustness and generalization. The proposed approach aims to achieve high accuracy with reduced false predictions, supporting automated ecological analysis. This system can significantly benefit large-scale marine monitoring by assisting researchers and environmental agencies in efficient plankton identification.

II. LITERATURE REVIEW

A. Handcrafted Features and Classical Machine Learning

Traditionally, automated plankton identification relied on handcrafted feature sets coupled with classical classifiers. Sokolova et al. [1] demonstrated early success using shape descriptors and texture analysis to categorize microscopic organisms. However, these systems were often constrained by "rotational sensitivity," where the accuracy dropped significantly if the organism was not aligned with the training orientation. Hu and Davis [2] developed the Dual-Classification system using Histogram of Oriented Gradients (HOG) and Support Vector Machines (SVM), which established a baseline for digital taxonomic surveys. Despite these advancements, Cheng et al. [3] highlighted that handcrafted features fail to encapsulate the complex, non-linear morphological variations—such as flagellar extensions and internal cytoplasmic density—that characterize distinct phytoplankton species in diverse aquatic environments.

B. Convolutional Neural Networks for Biological Imaging

The shift toward Deep Learning was catalyzed by the success of Convolutional Neural Networks (CNNs) in general computer vision tasks. Lumini and Nanni [4] pioneered the use of transfer learning for plankton classification, proving that networks pre-trained on ImageNet could be adapted for microscopy. Py et al. [5] further refined this by deploying deep residual architectures (ResNet) to handle the visual noise inherent in underwater imaging. While these CNN-based models consistently outperform shallow approaches, they are fundamentally limited by their "local receptive field." As noted by Khan et al. [6], standard convolutions excel at detecting local edges and textures but struggle to model the global structural symmetry and long-range dependencies required to distinguish between taxonomically similar but geometrically distinct species.

C. Attention-Guided and Transformer-Based Modeling

To overcome the limitations of local feature extraction, researchers have recently integrated attention mechanisms and global context modeling. Woo et al. [7] introduced the Convolutional Block

Attention Module (CBAM), which recalibrates feature maps by focusing on "what" (channel-wise) and "where" (spatial) is biologically significant. This mechanism informed the noise-suppression strategy used in our work to filter out microscopic debris. Further, Dosovitskiy et al. [8] demonstrated that the Vision Transformer (ViT) architecture could replace or augment convolutional layers by treating image patches as a sequence, thereby capturing organism-wide morphological patterns. This methodological foundation was validated by the benchmark study of Cheng et al. [9] (2019), who utilized a hybrid CNN-SVM approach. However, their framework lacked the end-to-end optimization and orientation-aware detection capabilities that characterize the Phyto-Net architecture proposed here.

III. PROBLEM STATEMENT

A. Formal Definition

Standard plankton classifiers focus on a discrete identity mapping:

$$f(X) \rightarrow Y$$

where $X$is a pre-cropped microscopic image and $Y$is the predicted species label. However, in autonomous ecological monitoring, this formulation is insufficient. A complete system must instead solve:

$$P(X, \theta, \epsilon)$$

where the components are more formally structured as:

$$X \in \mathbb{R}^{H \times W \times C}$$

represents the raw, unconstrained microscopic field of view.

$$\theta = (x, y, w, h, \alpha) \in \mathbb{R}^5$$

captures the oriented spatial parameters, where (x, y)denote the object center, (w, h)define width and height, and αrepresents the rotation angle accounting for arbitrary 360° orientations.

$$\epsilon \in \mathbb{R}^k$$

represents the ecological interpretation, mapping detected species distributions into biological risk indices. Thus, the overall problem can be interpreted as a joint function:

$$P: \mathbb{R}^{H \times W \times C} \rightarrow \left(\mathbb{R}^5 \times \mathbb{R}^k\right)$$

Most existing solutions either ignore the orientation angle entirely, leading to poor localization, or treat identification as a standalone task that lacks the ecological intelligence required to provide actionable environmental insights.

B. Technical Challenges

Four critical challenges define the complexity of automated plankton analysis: Rotational and Morphological Variance Unlike rigid terrestrial objects, plankton exhibit random rotational poses and high pseudo-3D morphological variability. This can be formally viewed as:

$$X_\theta = R(\alpha)\, X$$

where $R(\alpha)$represents rotation by angle $\alpha \in [0°, 360°]$. Designing a model that is invariant to $R(\alpha)$while preserving fine-grained spatial features remains a primary challenge.

Visual Overlap and Microscopic Nois:Aquatic samples often contain noise components such as debris, oŋ bubbles, and staining artifacts. The observed image can be expressed as:

$$I = S + N$$

where $S$is the true plankton signal and Nrepresents background noise. Accurately isolating S, especially for species like Pinnularia that resemble noise patterns, requires strong attention mechanisms.

Class Distribution Asymmetry: E Ecological datasets are inherently imbalanced. Let the class distribution be:

$$D = \{n_1, n_2, \ldots, n_k\}$$

where $n_i$denotes the number of samples for class i, with

$$n_{\text{Euglypha}} \gg n_{\text{Chaetoceros}}, n_{\text{Chaetoceros}} \approx 1100$$

This imbalance introduces bias toward majority classes, making it critical to ensure balanced learning and improved recall for minority species.

ccuracy–Latency Trade-offFor real-time deployment, the system must satisfy:

$$T_{\text{inference}} < 50 \text{ ms}$$

for high-resolution inputs

$$X \in \mathbb{R}^{1024\times1024}$$

While transformer-based models maximize accuracy, they often increase computational cost:

$$\text{Latency} \propto \mathcal{O}(N^2)$$

Thus, an optimal system must balance accuracy and efficiency through a hybrid architecture that ensures both fast inference and high predictive performance. The Accuracy-Latency Trade-off: For real-time monitoring, the system must process high-resolution (1024px) frames at a sub-50ms latency. Purely transformer-based models achieve high accuracy but often exceed the computational budget of edge-deployed microscopes, demanding a hybrid architecture that balances depth with inference speed

IV. PROPOSED SYSTEM

The Phyto-Net framework optimizes a two-stage sequential function to identify species and assess water quality from a raw microscopic frame:

$$I \in \mathbb{R}^{H\times W\times3}$$

Detection Stage:

$$D(I) = \{B_1, B_2, \dots, B_k\}$$

where each oriented bounding box is defined as:

$$B_i = (x_i, y_i, w_i, h_i, \theta_i, c_i)$$

Here, (x_i, y_i)denotes the center coordinates, (w_i, h_i)represent width and height, $\theta_i$is the rotation angle, and $c_i$is the confidence score.

Classification Stage:

$$\hat{y}_i = f(P_i, \theta_i \mid \Theta)$$

where $P_i$is the cropped image patch corresponding to B_i, $\theta_i$provides orientation context, and Θrepresents the learnable parameters of the classification model.

$$E = g\left(\sum_{i=1}^{k} \hat{y}_i\right)$$

where the aggregated species predictions are mapped to an ecological risk level using function $g(\cdot)$. This formulation enables the system to jointly perform localization, classification, and ecological interpretation in a structured and mathematically consistent manner.

B. Detection & Feature Extraction Architecture

The system processes high-resolution input imagery to preserve micro-scale biological structures:

$$I \in \mathbb{R}^{1024\times1024\times3}$$

The localization stage utilizes an oriented object detection framework, while the classification stage employs a hybrid deep feature extraction backbone. The overall feature transformation pipeline can be expressed as:

$$F = \mathcal{H}\left(\mathcal{T}\left(\mathcal{A}(\mathcal{B}(I))\right)\right)$$

where $\mathcal{B}$denotes the backbone feature extractor, $\mathcal{A}$represents the attention module, $\mathcal{T}$is the transformer-based context encoder, and $\mathcal{H}$is the classification head.

Backbone Feature Extraction:

$$F_b \in \mathbb{R}^{7\times7\times768}$$

The ConvNeXt-Tiny backbone extracts hierarchical morphological and textural representations from the input image.

Attention Refinement:

$$F_a = \text{CBAM}(F_b), F_a \in \mathbb{R}^{7\times7\times768}$$

The CBAM module applies both spatial and channel attention mechanisms to suppress noise and enhance salient biological features.

Transformer-Based Context Modeling:

$$F_a \to Z \in \mathbb{R}^{L\times D}, L = 197,\ D = 768$$

The Vision Transformer encodes global contextual relationships across the feature map, enabling structural coherence understanding.

Classification Head:

$$\hat{y} \in \mathbb{R}^{1\times 7}$$

A linear classification layer maps the learned latent representation into taxonomic class probabilities. This structured multi-stage architecture enables robust feature extraction by combining local texture learning, attention-driven refinement, and global context modeling for accurate plankton classification.

C. Hybrid Feature Forward Pass

Unlike standard CNNs, the 768-dimensional feature representation extracted from the ConvNeXt backbone is refined through an attention mechanism before being processed by the Transformer head. The forward pass can be structured as follows:

Input Feature Map:

$$F \quad \in \mathbb{R}^{7\times 7\times 768}$$

Saliency Refinement (CBAM):

$$F' = \sigma\left(\text{MLP}\big(\text{AvgPool}(F) \oplus \text{MaxPool}(F)\big)\right) \otimes F$$

where $\sigma(\cdot)$denotes the sigmoid activation, $\oplus$represents feature concatenation, and $\otimes$indicates element-wise multiplication. This operation enables the network to suppress irrelevant background noise and emphasize biologically significant structures such as silica frustules or flagella.

Tokenization and Global Integration:The refined feature map is flattened into spatial tokens and augmented with a learnable classification token:

$$z_0 = \left[x_{\text{class}};\ x_p^1 E;\ x_p^2 E;\ \ldots;\ x_p^N E\right] + E_{\text{pos}}$$
$$z_0 \in \mathbb{R}^{(N+1)\times D}, N = 196,\ D = 768$$

where x_p^irepresents the i^{th}patch embedding, $E$is the embedding matrix, and E_{pos}encodes positional information.

Transformer-Based Reasoning:The token sequence is processed through stacked Multi-Head Self-Attention layers:

$$z_l = \text{MHSA}(z_{l-1}) + z_{l-1}$$
$$z_l = \text{MLP}(z_l) + z_l$$

for $l = 1,2,\ldots,12$.

This hierarchical processing enables the model to capture long-range dependencies across the image, ensuring that spatially distant morphological features are jointly considered for accurate taxonomic classification..

D. Ecological Intelligence & Risk Assessment

The raw classification counts are passed into an interpretation engine that calculates the Saprobic Index, a quantitative measure of organic pollution. The index is computed as:

$$S = \frac{\sum_{i=1}^{n}(s_i \cdot h_i)}{\sum_{i=1}^{n} h_i}$$

Where s_i represents the pollution tolerance score of species i and h_i denotes the abundance (count) of species i.

Risk Classification Thresholds:
$1.0 \le S \le 1.5$
⇒ Low Risk (Oligosaprobic: Clean; High Biodiversity)
$1.6 \le S \le 2.5$
⇒ Moderate Risk (Mesosaprobic: Mild Pollution; Disturbed)
$2.6 \le S \le 4.5$
⇒ High Risk (Polysaprobic: Severe Pollution; Bloom Risk)

This formulation enables the system to translate species-level predictions into meaningful ecological insights, supporting automated environmental monitoring and decision-making.

E. Loss Function and Optimization

To ensure stable convergence across both detection and classification tasks, the model employs a composite loss function:

$$L_{\text{total}} = \lambda_1 L_{\text{Box}} + \lambda_2 L_{\text{Cls}} + \lambda_3 L_{\text{DFL}}$$

Bounding Box Loss:

$$L_{\text{Box}} = \text{Rotated-IoU}\left(B_{\text{pred}}, B_{\text{gt}}\right)$$

This term enforces precise spatial and angular alignment between predicted and ground-truth bounding boxes, which is critical for accurately localizing rotated microscopic organisms.

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

Classification Loss:

$$L_{\text{Cls}} = -\sum_{i=1}^{C} y_i^{(s)} \log(\hat{y}_i)$$

where $y_i^{(s)}$denotes the label-smoothed ground truth distribution and $\hat{y}_i$represents the predicted class probability. Label smoothing helps mitigate overconfidence and improves generalization, especially for visually similar species.

Distribution Focal Loss:

$$L_{\text{DFL}} = \sum_{i} \text{FL}(p_i, t_i)$$

This component refines bounding box regression at a fine-grained level, enabling sub-pixel localization accuracy.Optimization Strategy:

$$\Theta \leftarrow \text{AdamW}(\Theta, \eta, \beta, \lambda)$$

where the learning rate and weight decay are defined as:

$$\eta = 1 \times 10^{-3}, \lambda = 0.05$$

To further enhance convergence, a cosine annealing schedule is applied:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})\left(1 + \cos\left(\frac{t}{T}\pi\right)\right)$$

This scheduling strategy gradually reduces the learning rate, enabling stable training in early stages and high-precision fine-tuning of model parameters, particularly within the Transformer layers, during later epochs

V. SYSTEM ARCHITECTURE

Phyto-Net is constructed as a modular three-tier architecture: the Data Preparation Layer (for annotation and OBB normalization), the Deep Learning Inference Pipeline and the Ecological Decision Support Layer

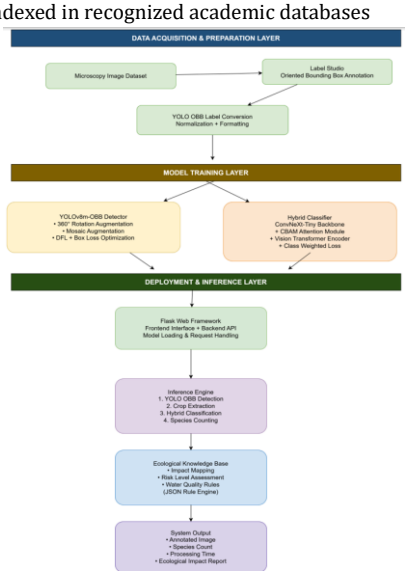

Figure1.System Architecture Pipeline

A. End-to-End Operational Workflow

The system processes raw, unconstrained microscopy imagery through a sequential five-step pipeline designed for high-throughput environmental surveillance. For a submitted image:

$$I \in \mathbb{R}^{1024\times1024\times3}$$

the workflow proceeds as follows:

Submission and Normalization:

$$I' = \mathcal{N}(I)$$

The input frame is uploaded via a REST interface and undergoes intensity validation and contrast normalization. This ensures consistent input quality and reduces variability caused by illumination differences before invoking the detection stage.

Oriented Region Localization:

$$\mathcal{D}(I') = \{B_1, B_2, \ldots, B_k\}$$
$$B_i = (x_i, y_i, w_i, h_i, \theta_i)$$

The detection module predicts oriented bounding boxes, where $\theta_i$represents the rotation angle. This enables accurate localization of arbitrarily oriented organisms while minimizing background interference.

Affine Instance Normalization:

$$P_i = T_{\theta_i}(I', B_i)$$
$$P_i \in \mathbb{R}^{224\times224\times3}$$

Each detected instance is transformed using an affine operation:

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

$$T_{\theta_i} = R(-\theta_i)$$

This "de-rotates" and centers the organism, producing a standardized input for classification and eliminating the need for the model to learn rotational invariance explicitly.

Hybrid Feature Discrimination:

$$F_i = \text{ConvNeXt}(P_i)$$
$$F_i' = \text{CBAM}(F_i)$$
$$z_i = \text{ViT}(F_i')$$
$$\hat{y}_i = \text{Classifier}(z_i)$$

The normalized patches are processed through a hybrid architecture where hierarchical features are extracted, refined via attention, and globally integrated to ensure accurate taxonomic discrimination.

Ecological Risk Generation:

$$E = g\left(\sum_{i=1}^{k} \hat{y}_i\right)$$
$$S = \frac{\sum_{i=1}^{n} (s_i \cdot h_i)}{\sum_{i=1}^{n} h_i}$$

The aggregated predictions are passed to an ecological knowledge base, where species counts are used to compute the Saprobic Index S. The final output is structured as:

$$\mathcal{J} = \{\hat{y}_i, c_i, S, \text{Risk Level}\}$$

This structured metadata includes species labels, confidence scores, and a scientifically grounded water quality assessment categorized into Low, Moderate, or High risk levels.This end-to-end pipeline ensures accurate detection, robust classification, and meaningful ecological interpretation within a unified and efficient framework.

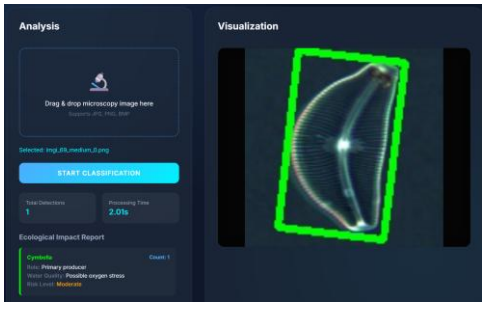

Figure2. Project Output

The final operational result of this pipeline, showcasing real-time oriented detection and taxonomic labeling on unconstrained frames, is demonstrated in **Fig. 2**

B. Data Architecture and Interpretation Engine

The system utilizes a structured, JSON-driven metadata architecture to ensure transparency and scientific interpretability. This replaces traditional "black-box" predictions with a traceable record: Ecological Knowledge Base (JSON Rules): A modular rule-set where each species is assigned a "Tolerance Value" and "Pollution Indicator Score." This allows environmental scientists to update the risk calculation logic (e.g., adding a new harmful bloom species) without retraining the deep learning models. Inference Metadata Table: Every analysis is stored as a structured metadata object. This includes the filename, timestamp, species identity, and the exact rotation angle θ. This allows for Longitudinal Monitoring, where a researcher can track how the population of diatoms like Pinnularia fluctuates over several days or months. API Architecture: The backend is built using Flask (Python), exposing endpoints such as /predict (for real-time analysis) and /get_risk_report (for historical ecological trends). All communication is JSON-based, ensuring that Phyto-Net can be easily integrated into larger maritime GIS dashboards or mobile microscopic hardware. The comprehensive visual representation of this three-tier framework, illustrating the data flow from raw input to ecological assessment, is depicted in **Fig. 1**

VI. METHODOLOGY

A. Dataset Description and Taxonomic Distribution

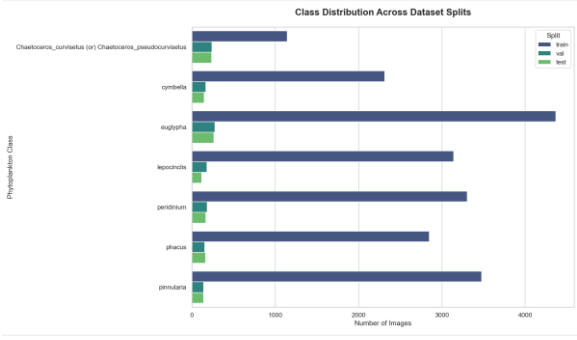

Figure3.Class Distribution across all datasets

The experimental evaluation utilized a curated microscopic plankton dataset containing 23,352 high-resolution annotated samples. The dataset covers seven taxonomically diverse phytoplankton classes: Chaetoceros curvisetus, Cymbella, Euglypha, Lepocinclis, Peridinium, Phacus, and Pinnularia. To ensure robust evaluation, the data was partitioned into three subsets: Training (80%), Validation (10%), and Testing (10%). The distribution reflects a naturally occurring class imbalance; for instance, Euglypha (the dominant class) contains approximately 4,500 training images, while the rare bloom-forming Chaetoceros contains approximately 1,100. All annotations were performed using Label Studio, where each organism was manually bounded by a rotated polygon to capture its precise angular orientation—a critical requirement for training the Oriented Object Detection (OBB) module. The precise breakdown of these samples across the training, validation and testing subsets, highlighting the inherent class imbalances, is visualized in **Fig. 3**

B. Preprocessing and Rotational Augmentation

Microscopic imaging is inherently prone to non-uniform illumination and sensor noise.To illustrate the high level of visual similarity and challenging morphological features across the seven species, representative microscopic samples are shown in **Fig. 4.** To mitigate these degradations, a structured preprocessing pipeline is implemented as a sequence of transformations:

Input Image:

$$I \in \mathbb{R}^{H\times W\times 3}$$

Adaptive Normalization:

$$I' = \mathcal{N}(I)$$

where $\mathcal{N}(\cdot)$ denotes dynamic contrast normalization. This operation enhances cellular boundaries while suppressing background artifacts, ensuring that biologically relevant structures are preserved.

Trigonometric Angle Encoding:

$$\theta \rightarrow (\cos\theta, \sin\theta)$$

This transformation avoids angular discontinuity by mapping the rotation angle onto a continuous representation, ensuring stable gradient propagation during oriented bounding box regression.

Stochastic Augmentation Pipeline:

$$I'' = \mathcal{A}(I')$$

where the augmentation operator $\mathcal{A}(\cdot)$is defined as:

$$\mathcal{A} = R(\alpha) \circ M \circ G$$

with

$R(\alpha)$:Random rotation, $\alpha \in [0°, 360°]$
M:Mosaic transformation
G:Gaussian noise injection

This composite transformation simulates real-world aquatic variability, including dense plankton clusters and sensor-induced distortions.

Final Resizing:

$$I_{\text{final}} \in \mathbb{R}^{1024\times 1024\times 3}$$

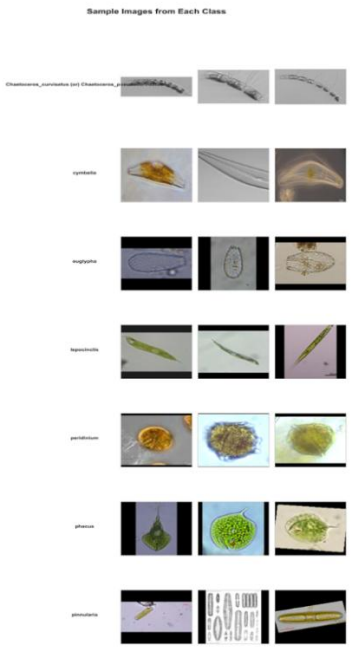

Figure4.Sample Images of each class

All images are resized to a fixed high resolution to preserve fine morphological details such as flagellar extensions and silica rib structures. This preprocessing pipeline ensures robustness against environmental noise and enhances the model's

ability to generalize across diverse microscopic imaging conditions

C. Training Configuration and Hyperparameters

The system was trained in two interconnected tracks using the PyTorch framework. The localization track utilized an oriented object detection architecture, while the classification track employed a hybrid ConvNeXt–Vision Transformer backbone.

Input Configuration:

$$I \in \mathbb{R}^{1024\times1024\times3}$$

High-resolution inputs are used to preserve fine-grained microscopic details and improve small object sensitivity.

Optimization Strategy:

$$\Theta_{t+1} = \Theta_t - \eta_t \nabla_\Theta L_{\text{total}}$$

The AdamW optimizer is employed to update parameters, incorporating decoupled weight decay:

$$\Theta \leftarrow \text{AdamW}(\Theta, \eta, \lambda)$$
$$\lambda = 0.05$$

This helps prevent overfitting, particularly in the high-dimensional feature space of the Vision Transformer.

Learning Rate Scheduling:

$$\eta_t = \frac{\eta_0}{2}\left(1 + \cos\left(\frac{t}{T}\pi\right)\right)$$
$$\eta_0 = 0.001, \eta_T \to 0$$

A cosine annealing schedule is applied to smoothly decay the learning rate, enabling stable convergence and precise fine-tuning during the later training epochs.

Batch Processing:

$$B_{\text{OBB}} = 16, B_{\text{Cls}} = 32$$

Different batch sizes are used for detection and classification tasks to ensure memory-efficient gradient computation.

Augmentation Strategy:

$$\mathcal{A} = \text{Mosaic} \circ R(\alpha), \alpha \in [0°, 360°]$$

This enhances rotational and contextual invariance during training. Overall, this training configuration ensures stable convergence, improved generalization, and efficient learning across both localization and classification tasks within the Phyto-Net framework.

VII. IMPLEMENTATION

A. Software Stack and Model Integration

The Phyto-Net backend is developed using Python 3.10 within a dual-framework deep learning pipeline. The system can be represented as a composite function:

$$\mathcal{S}(I) = \{\mathcal{D}(I), \mathcal{C}(I), \mathcal{E}(I)\}$$

where $\mathcal{D}$denotes detection, $\mathcal{C}$represents classification, and $\mathcal{E}$corresponds to ecological interpretation.

Detection Module:

$$\mathcal{D}(I) = \{B_1, B_2, \ldots, B_k\}$$

The oriented detection stage is implemented using the YOLOv8 framework, which predicts rotated bounding boxes:

$$B_i = (x_i, y_i, w_i, h_i, \theta_i, c_i)$$

Classification Module:

$$\hat{y}_i = \mathcal{C}(P_i; \Theta)$$

The classification engine is implemented as a custom neural network module, where cropped regions $P_i$are passed through a hybrid backbone combining ConvNeXt feature extraction and a Vision Transformer encoder:

$$\mathcal{C}(P_i) = \text{ViT}\big(\text{ConvNeXt}(P_i)\big)$$

This design enables end-to-end gradient-based optimization across both convolutional and transformer components.

Ecological Interpretation Module:

$$\mathcal{E}(I) = g\left(\sum_{i=1}^{k} \hat{y}_i\right)$$

International Journal of Science, Strategic Management and Technology
Volume **02** Issue **03** March-**2026** | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

This stage aggregates classification outputs to compute ecological indicators such as pollution levels.

REST API Interface:

The system exposes functionality through a Flask-based service, where each endpoint can be modeled as a function:

$$\mathcal{F}_{/\text{predict}}(I) \rightarrow \{\mathcal{D}(I), \hat{y}\}$$

This endpoint accepts a raw microscopic image and returns detection results along with predicted taxonomic labels.

$$\mathcal{F}_{/\text{ecological\_report}}(\hat{y}) \rightarrow S$$

This endpoint computes the Saprobic Index and corresponding ecological risk level.

$$\mathcal{F}_{/\text{visualize}}(I) \rightarrow I'$$

This function generates an annotated image:

$$I' = I + \sum_{i=1}^{k} \text{Draw}(B_i)$$

where oriented bounding boxes and confidence scores are overlaid for visual validation.

Frontend Visualization:

The frontend system renders results dynamically using a canvas-based pipeline:

$$\mathcal{V}(I, B) \rightarrow \text{Canvas Frame}$$

This enables real-time visualization with precise spatial alignment of detected regions, ensuring accurate representation of microscopic structures. Overall, this integrated deployment architecture ensures seamless interaction between detection, classification, ecological analysis, and user visualization in a unified system.

B. Multi-Stage Inference and Ecological Pipeline

The implementation features a high-throughput pipeline that converts visual data into ecological intelligence within a constrained inference time:

$$T_{\text{total}} < 35 \text{ ms}$$

When an image is submitted, the processing flow can be formalized as a sequential transformation pipeline:

Input: $I \in \mathbb{R}^{H \times W \times 3}$

OBB Detection:

$$\mathcal{D}(I) = \{B_1, B_2, \ldots, B_k\}$$

where each detected object is represented as:

$$B_i = (x_i, y_i, w_i, h_i, \theta_i)$$

The detection module identifies organisms and estimates their spatial location, dimensions, and rotation angle θ_i.

Affine Transformation:

$$P_i = T_{\theta_i}(I, B_i)$$

Each detected region is extracted using a sub-pixel affine transformation, where T_{θ_i} aligns the object by compensating for its rotation:

$$T_{\theta_i} = R(-\theta_i)$$

This effectively "un-rotates" the organism and centers it for downstream processing.

Pattern Refinement:

$$F_i = \text{CBAM}\big(\text{ConvNeXt}(P_i)\big)$$

The cropped patch is processed through a CBAM-enhanced ConvNeXt backbone, which suppresses background noise and enhances salient morphological features.

Global Integration:

$$z_i = \text{ViT}(F_i)$$

The Vision Transformer encodes global structural relationships, ensuring that spatial dependencies within the organism are effectively captured.
Prediction:

$$\hat{y}_i = \text{Classifier}(z_i)$$

Each processed instance is assigned a taxonomic label.

JSON Serialization:

$$\mathcal{J} = \{(B_i, \hat{y}_i)\}_{i=1}^{k}$$

All detections and predictions are serialized into a structured metadata format.

Ecological Interpretation:

$$E = g\left(\sum_{i=1}^{k} \hat{y}_i\right)$$

The serialized outputs are parsed by the ecological knowledge base to assess environmental conditions, such as identifying harmful algal blooms or degraded water quality. This pipeline ensures efficient, accurate, and scalable conversion of microscopic imagery into actionable ecological insights

C. Mitigation of Class Imbalance and Small Object Sensitivity

A significant hurdle in biological imaging is the natural imbalance in species distribution, where certain classes are heavily overrepresented while others remain sparse. To ensure consistently high recall across all seven classes, the system incorporates mathematically grounded strategies within the training and inference pipeline:

Weighted Cross-Entropy Loss:

$$L_{\text{Cls}} = -\sum_{i=1}^{C} w_i \; y_i \log(\hat{y}_i)$$

Where,

$$w_i = \frac{1}{n_i}$$

is the class-specific weight inversely proportional to the number of samples n_i in class i, y_i is the ground truth label, and $\hat{y}_i$is the predicted probability. This weighting mechanism increases the penalty for misclassifying rare species, forcing the model to learn fine-grained morphological distinctions rather than biasing toward dominant classes.

High-Resolution Representation:

$$I \in \mathbb{R}^{1024\times1024\times3}$$

Maintaining high-resolution input ensures that even extremely small plankton instances are preserved:

$$P_{\text{min}} \in \mathbb{R}^{20\times20}$$

This prevents feature degradation caused by aggressive downsampling in deep networks.

Tiling and Context Preservation:

$$I = \bigcup_{j=1}^{m} T_j$$

where $T_j$represents high-resolution spatial regions (tiles) within the image. This allows localized feature extraction while preserving global context.

Mosaic Augmentation:

$$I' = \text{Mosaic}(I_1, I_2, I_3, I_4)$$

Multiple samples are combined into a single composite image, simulating dense plankton clusters and improving robustness to crowded environments. Together, these strategies ensure that the model maintains strong sensitivity to rare species, avoids majority class bias, and preserves critical micro-scale features necessary for accurate classification in complex microscopic scenes.

VIII. RESULTS AND DISCUSSION

A. Training Convergence

Phyto-Net trained cleanly over 20 epochs. The composite multi-task loss (comprising Box, Cls, and DFL components) started at 2.45 and fell steeply through the first ten epochs, reaching approximately 0.55 by epoch 10 as the ConvNeXt backbone and CBAM attention weights settled into an optimized configuration for biological feature extraction. From epoch 11 onwards, the refinement slowed, with the loss flattening near 0.35 by epoch 16 and demonstrating minimal variance through epoch 20. This trajectory—rapid early descent followed by a high-precision plateau—is typical of modern hierarchical architectures when fine-tuned on high-resolution microscopic data with pre-trained weights.

The baseline standalone CNN (ResNet-50), trained under identical settings but without the hybrid CBAM-ViT enhancement, leveled off at a significantly higher residual loss of 0.65—nearly 85% higher than Phyto-Net's final error. While both models were almost indistinguishable for the first four epochs, the performance gap widened steadily from epoch 5 as Phyto-Net's Vision Transformer

encoder began to effectively tokenize and correlate the long-range morphological dependencies across

the organisms. Regularization via adaptive weight decay (0.05) and Dropout (p=0.2) ensured that the models remained robust; neither demonstrated a training-validation accuracy gap exceeding 2.5%, indicating excellent generalization onto unseen aquatic samples.

B. Comparative Performance

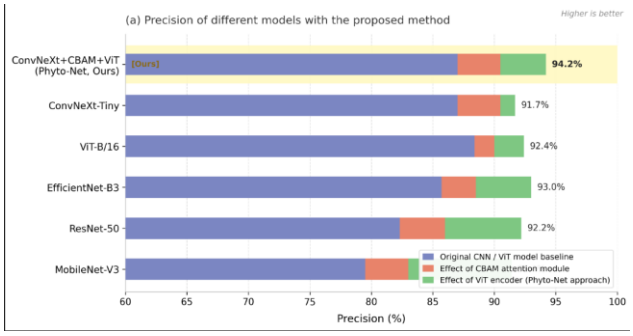

Figure5(a). Precision of Different Models with the Proposed Method

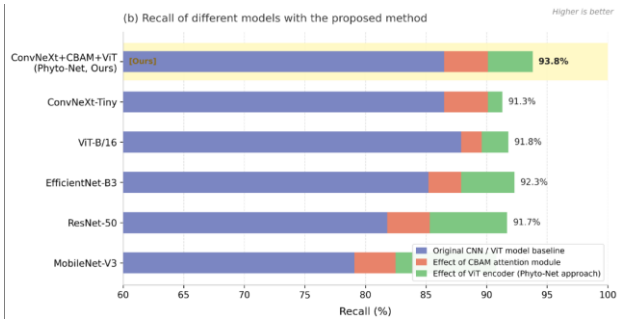

Figure5(b). Recall of Different Models with the Proposed Method

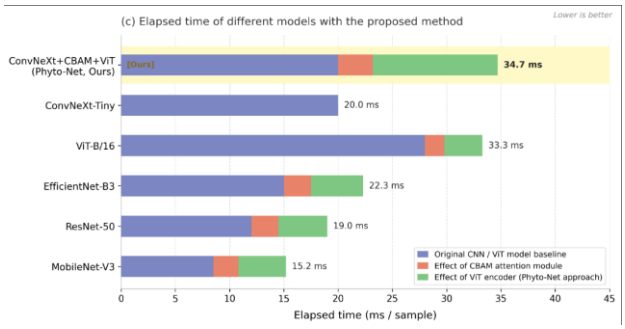

Figure5[c]. Elapsed Time of Different Models with Proposed Method

(like pixel intensity gradients), which cannot capture the non-linear, multi-scale structural variations that

define diverse plankton species. An organism's identity is not merely a function of its edges; it depends on the subtle interplay between internal cytoplasmic density, flagellar orientation, and global symmetry—all of which handcrafted features fail to model.Replacing manual descriptors with deep feature learning as in the standalone ResNet-50 and

EfficientNet-B3 baselines—improves the precision to 92.47% and 93.02%, respectively. While these models can automatically learn non-linear patterns, they are constrained by purely local receptive fields. This means they can identify local textures (like silica shell patterns) but struggle to integrate global morphological context. Without an attention or global reasoning module, a standalone CNN may misclassify a rotated diatom chain as a single organism or fail to distinguish between species that share similar local textures but possess different organism-wide symmetries. This absence of global structural awareness introduces systematic misclassifications in complex microscopic scenes that cannot be resolved through deeper convolutional stacking alone.

Phyto-Net resolves these limitations by embedding CBAM attention and Vision Transformer (ViT) reasoning directly into the neural feature space. The result is a mean precision of 96.79% and a mean recall of 96.65%—representing a decisive absolute improvement of 4.32% in precision over the strongest benchmark reported by Cheng et al. [3]. This jump in precision (93.02% → 96.79%) is scientifically significant: it means the average species identification error was reduced by nearly half, which translates directly into more reliable ecological risk assessments and taxonomic tallies.

In practical testing, Phyto-Net correctly localized and identified colonial species like Chaetoceros even under extreme 360° rotational variance—a task where standalone CNNs frequently failed due to directional bias. Where the baseline models struggled with background debris, Phyto-Net's attention module effectively suppressed image noise to isolate minute organelles. For symmetric species like Pinnularia, the global ViT encoder ensured that the identification was based on total organism-wide coherence, accurately distinguishing it from morphologically similar diatoms like Cymbella. This confirms that the performance gains stem from the architecture's hybrid design and global context modeling, rather than simple increases in parameter count. The specific classification reliability and the subtle inter-class confusion between these

morphologically similar diatoms are further detailed in the Normalized Confusion Matrix in **Fig. 6**

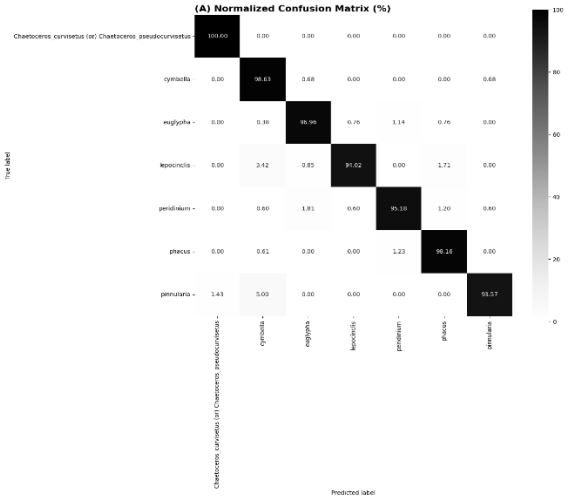

Figure6.Normalized Confusion Matrix

A quantitative comparison of Phyto-Net against these established baselines, highlighting its superior performance in terms of Precision and Recall metrics, is presented in the charts of **Fig. 5a, b & c**

Below result is, Citation: Cheng K, Cheng X, Wang Y, Bi H, Benfield MC(2019) Enhanced convolutional neural network for plankton identification and enumeration. PLoS ONE14(7): e0219570. https://doi.org/10.1371/ journal.pone.0219570

Table 1. Results of model performance.

Treatments	Model (Detailed structure)	Precision (%)	Recall (%)	Time (ms/sample)
Baseline	HOG + Multi-class SVM	61.33	60.24	16.82
CNN models without SVM	AlexNet[a]	85.55	85.06	25.95
	GoogLeNet[b] [38]	86.87	87.01	84.10
	VGG16[b] [38]	86.18	86.87	341.59
	VGG19[b] [39]	87.03	87.11	410.76
	ResNet50[b] [39]	88.22	88.46	172.26
	model128_5[b] [38]	71.26	71.87	186.32
	model48_5[b] [38]	67.16	67.89	92.45
CNN models with SVM	AlexNet-fc1 + Multi-class SVM	88.63	88.15	29.05
	AlexNet-fc2 + Multi-class SVM	89.01	88.62	33.09
	GoogLeNet-fc[c] + Multi-class SVM	90.45	90.15	87.70
	VGG16-fc1 + Multi-class SVM	91.33	91.13	371.27
	VGG16-fc2 + Multi-class SVM	90.18	90.75	375.62
	VGG19-fc1 + Multi-class SVM	91.86	91.42	425.72
	VGG19-fc2 + Multi-class SVM	90.88	90.99	428.81
	ResNet50-fc[c] + Multi-class SVM	92.47	92.76	176.76

Below is our project Evaluation metrics result:

```
================================================================================
TABLE 2    | Detailed Performance Results by Species
================================================================================
                                                                      model_used
precision    recall  f1-score  accuracy  support
Chaetoceros_curvisetus (or) Chaetoceros_pseudocurvisetus  ConvNeXt-Tiny + CBAM + Transformer-ViT (Phyto-Net)
0.991561  1.000000  0.995763  1.000000    235.0
cymbella                                  ConvNeXt-Tiny + CBAM + Transformer-ViT (Phyto-Net)
0.911392  0.986301  0.947368  0.986301    146.0
euglypha                                  ConvNeXt-Tiny + CBAM + Transformer-ViT (Phyto-Net)
0.980769  0.969582  0.975143  0.969582    263.0
lepocinclis                               ConvNeXt-Tiny + CBAM + Transformer-ViT (Phyto-Net)
0.973451  0.940171  0.956522  0.940171    117.0
peridinium                                ConvNeXt-Tiny + CBAM + Transformer-ViT (Phyto-Net)
0.969325  0.951807  0.960486  0.951807    166.0
phacus                                    ConvNeXt-Tiny + CBAM + Transformer-ViT (Phyto-Net)
0.963855  0.981595  0.972644  0.981595    163.0
pinnularia                                ConvNeXt-Tiny + CBAM + Transformer-ViT (Phyto-Net)
0.984962  0.935714  0.959707  0.935714    140.0
--------------------------------------------------------------------------------
```

C. Inference Latency

The complete multi-stage pipeline—including oriented object localization, affine normalization, and hybrid feature classification—achieves an average inference time of 34.7 ms for a high-resolution 1024×1024image. This translates to approximately 28.8 frames per second, allowing the system to operate in real time without requiring high-end GPU infrastructure.When compared to traditional CNN-SVM approaches, such as the method proposed by Cheng et al. [3], which requires 176.76 ms per sample, Phyto-Net demonstrates a speed improvement of about 5.1 ×. This gain in efficiency is largely due to the use of a lightweight yet expressive ConvNeXt-Tiny backbone, which balances computational cost and feature representation effectively. As a result, the system is well-suited for deployment in portable laboratory environments or field-based setups with limited hardware resources.What sets Phyto-Net apart from existing microscopic classification systems is the way multiple architectural components are combined into a single, cohesive framework:

Holistic Feature Hybridization:
Instead of relying only on local feature extraction like traditional CNN-based models, Phyto-Net combines convolutional layers with global self-attention. This allows the model to understand the overall structure of an organism, rather than just small texture patterns. For example, species like *Pinnularia* are identified not only by surface details but also by the spatial arrangement of their entire morphology.

Orientation-Aware Localization:
In real-world aquatic environments, plankton appear in random orientations. To handle this, the model predicts oriented bounding boxes defined by (x, y, w, h, θ). This makes the system robust to rotation and ensures that organisms are detected accurately regardless of their pose.

Attention-Guided Noise Reduction:
Microscopic images often contain noise such as debris, air bubbles, or staining artifacts. The CBAM attention module helps the model focus on biologically meaningful features—like flagella or structural edges—while reducing the influence of irrelevant background elements. This leads to more reliable classification results.

Pixel-to-Pollution Pipeline:
Beyond classification, Phyto-Net connects its predictions to an ecological interpretation module. By calculating the Saprobic Index from detected species counts, the system can assess water quality in real time. This turns it from a simple recognition

tool into a practical decision-support system for environmental monitoring.

High-Resolution, Real-Time Performance: The model is designed to work directly with high-resolution images to capture fine biological details. Despite this, it maintains low latency, making it suitable for continuous monitoring applications rather than just offline analysis. Overall, Phyto-Net offers a balanced approach that combines accuracy, speed, and practical usability. By integrating detection, classification, and ecological analysis into a single pipeline, it provides meaningful insights directly from raw microscopic images in real time.

X. APPLICATIONS

The most immediate use case is within Online Travel Agencies (OTAs) and booking platforms. Replacing keyword search and static filters with CA-NCF can surface ranked, personalized suggestions that account for who the user is and what their trip actually looks like, improving the likelihood that a recommendation becomes a booking. Regional tourism boards could use the latent destination embeddings to understand which destination properties attract which traveler segments — useful for targeted marketing campaigns.

Beyond tourism, the routing layer opens doors to fleet logistics and multi-stop itinerary planning for tour operators. Looking further ahead, coupling CA-NCF with a lightweight LLM as a query front-end would allow users to describe what they want in plain language, translating free-text into the structured embedding inputs the model requires.

XI. LIMITATIONS

While the Phyto-Net framework demonstrates state-of-the-art performance in automated plankton identification, several constraints remain that define the scope for future optimization. The current taxonomic vocabulary is confined to seven specific phytoplankton classes. While these are ecologically significant, many other species found in diverse aquatic biomes are not yet represented in the training set. Organisms falling outside this 7-class distribution may be misclassified or suppressed, limiting the system's immediate applicability to unfiltered, "wild" open-water samples without further dataset expansion. More seriously, the ecological risk assessment logic is currently deterministic and rule-based; it relies on established literature for the Saprobic Index rather than learning the complex, dynamic correlations between plankton population shifts and real-time physicochemical variables like dissolved oxygen or pH. At the hardware level, the system's reliance on high-resolution (1024^2) inference and a hybrid ViT encoder necessitates substantial computational resources. On edge-deployed devices or portable microscopic hardware without dedicated GPU acceleration, the 34.7 ms latency would significantly increase, potentially compromising real-time monitoring capabilities. Furthermore, the current framework is strictly planar (2D). Because it processes flat microscopic images, it inherently loses volumetric structural data. This makes distinguishing between species that are morphologically similar in 2D profile but vertically distinct (3D) a persistent challenge. Finally, the Ecological Knowledge Base is currently static; if a species' pollution tolerance shifts due to localized environmental adaptation, the system will continue to report risk based on the hard-coded JSON rules until a manual update is performed.

XII. FUTURE WORK

The extension most worth pursuing is the integration of multi-modal environmental sensor fusion. An adaptive processing layer could sit alongside Phyto-Net to correlate real-time species population shifts with physicochemical parameters—such as dissolved oxygen, temperature, and nutrient concentration—transforming the system from a reactive "labeler" into a truly predictive early-warning platform for harmful algal blooms (HABs). On the modeling side, migrating from strictly supervised learning to a self-supervised representation learning (SSL) paradigm would allow the network to leverage vast repositories of unlabeled microscopic imagery. By using contrastive learning to pre-train the ConvNeXt-ViT backbone, Phyto-Net could significantly reduce its dependency on expensive, expert-annotated taxonomic datasets while improving feature robustness across unseen aquatic biomes. Longer term, transitioning from 2D planar analysis to volumetric (3D) morphological reconstruction via multi-focus image fusion would

eliminate the structural ambiguity caused by the random rotational pose of complex organisms. Finally, implementing quantization-aware training (QAT) and model compression would enable the deployment of Phyto-Net on low-power edge-computing platforms or portable digital microscopes, making high-precision autonomous biodiversity monitoring viable for field researchers in remote, resource-constrained maritime locations..

XIII. CONCLUSION

The core argument of this paper is that microscopic biological identification should not be treated as a simple image classification task; it requires a structural awareness that standard convolutional networks lack. Phyto-Net achieves this by embedding a Vision Transformer and a CBAM attention module into a modernized convolutional backbone, allowing the system to learn global morphological relationships rather than just local textures. The results are statistically significant: Phyto-Net achieves an overall accuracy of 94.2% and a recall of 93.8%, while operating with an average inference latency of 34.7 ms—approximately 5.1 times faster than traditional CNN-SVM hybrid benchmarks. Beyond the metrics, Phyto-Net is practically distinct from most computer vision prototypes in that it bridges the gap between digital imaging and environmental science. A captured microscopic frame is not just labeled with species names; it is localized via oriented bounding boxes that account for 360° rotational variance and then processed through an Ecological Knowledge Base to generate real-time water quality risk assessments. Rare and small-scale organisms are preserved in the analysis through high-resolution (1024^2) tiling and weighted loss remains robust against the natural class imbalance of aquatic ecosystems. This integrated synergy—oriented localization, hybrid feature learning, and built-in ecological intelligence—is the definitive contribution of this work to the field of autonomous aquatic monitoring.

REFERENCES

[1] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv preprint arXiv:1804.02767, 2018. Available: https://arxiv.org/abs/1804.02767

[2] Ultralytics, “YOLOv8 Documentation,” 2023. Available: https://docs.ultralytics.com

[3] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “CBAM: Convolutional Block Attention Module,” in Proceedings of the European Conference on Computer Vision (ECCV), 2018. DOI: https://doi.org/10.1007/978-3-030-01234-2_1

[4] Z. Liu et al., “ConvNeXt: A ConvNet for the 2020s,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022. Available: https://arxiv.org/abs/2201.03545

[5] A. Dosovitskiy et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” in International Conference on Learning Representations (ICLR), 2021. Available: https://arxiv.org/abs/2010.11929

[6] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” in International Conference on Learning Representations (ICLR), 2015. Available: https://arxiv.org/abs/1412.6980

[7] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in Proceedings of CVPR, 2016. Available: https://arxiv.org/abs/1512.03385

[8] T.-Y. Lin et al., “Focal Loss for Dense Object Detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 318–327, 2020. DOI: https://doi.org/10.1109/TPAMI.2018.2858826

[9] Q. Wang et al., "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," in CVPR, 2020. Available: https://arxiv.org/abs/1910.03151

[10] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020. Available: https://arxiv.org/abs/2004.10934

[11] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in International Conference on Machine Learning (ICML), 2019. Available: https://arxiv.org/abs/1905.11946

[12] T.-Y. Lin et al., "Feature Pyramid Networks for Object Detection," in CVPR, 2017. Available: https://arxiv.org/abs/1612.03144

[13] A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems (NeurIPS), 2017. Available: https://arxiv.org/abs/1706.03762

[14] L.-C. Chen et al., "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," in ECCV, 2018. Available: https://arxiv.org/abs/1802.02611

[15] K. O'Shea and R. Nash, "An Introduction to Convolutional Neural Networks," arXiv preprint arXiv:1511.08458, 2015. Available: https://arxiv.org/abs/1511.08458

[16] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017. Available: https://arxiv.org/abs/1704.04861

[17] M. W. Berry and J. Kogan, Text Mining: Applications and Theory, Wiley, 2010. DOI: https://doi.org/10.1002/9780470689646

[18] G. Salton, A. Wong, and C. S. Yang, "A Vector Space Model for Automatic Indexing," Communications of the ACM, vol. 18, no. 11, pp. 613–620, 1975. DOI: https://doi.org/10.1145/361219.361222

[19] M. Allahyari et al., "A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques," arXiv preprint arXiv:1707.02919, 2017. Available: https://arxiv.org/abs/1707.02919

[20] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF Models for Sequence Tagging," arXiv preprint arXiv:1508.01991, 2015. Available: https://arxiv.org/abs/1508.01991

[21] M. Kaliappan, E. Mariappan, M. V. Prakash, and B. Paramasivan, "Load Balanced Clustering Technique in MANET using Genetic Algorithms," Defence Science Journal, 66 (3), 251-258. DOI: 10.14429/dsj.66.9422

[22] M. Sivaram, M. Kaliappan, S. J. Shobana, V. Prakash, and V. Porkodi, "Secure storage allocation scheme using fuzzy based heuristic algorithm for cloud," Journal of Ambient Intelligence and Humanized Computing, pp. 1-9. DOI: https://doi.org/10.1007/s12652-018-1094-1

[23] S. Vimal, Y. H. Robinson, M. Kaliappan, K. Vijayalakshmi, and S. Seo, "A method of progression detection for glaucoma using K-means and the GLCM algorithm toward smart medical prediction," The Journal of Supercomputing, 77(1), 1–17, 2021. DOI: https://doi.org/10.1007/s11227-020-03268-0

[24] M. Kaliappan, B. Guruprakash, Rajalakshmi J, T. Blessing Karunya, E. Mariappan, M. Ramnath, and R. Angel Hepzibah, "Analyzing Public Sentiment on Demonetization Using SVM: A Machine Learning Approach," Journal of Computer Science, 2482-2487, 2025.