Attention-Enhanced Yolov8 for Real-Time Milkfish Disease Detection with Synthetic Data Augmentation

***Nahum M. Quiros[1], Arnel C. Fajardo[2]**

[1]Information Technology Department, College of Computing, Pangasinan State University, Urdaneta City, Philippines

[2] College of CSICT, Isabela State University – Cauayan Campus, Cauayan City, Philippines.

DOI: https://dx.doi.org/10.51244/IJRSI.2026.1305000225

Received: 13 May 2026; Accepted: 18 May 2026; Published: 10 June 2026

ABSTRACT

Milkfish (Chanos chanos) aquaculture faces persistent challenges due to disease outbreaks that are difficult to detect at early stages using traditional manual monitoring methods. Varied underwater lighting conditions, turbidity, and fish movement further complicate visual assessment and may reduce detection reliability. This study proposes an automated disease detection framework that combines YOLOv8m object detection and the Convolutional Block Attention Module (CBAM) with CycleGAN-based synthetic data augmentation. CycleGAN is employed to generate additional diseased samples from unpaired image domains, addressing dataset imbalance and improving training diversity. CBAM is integrated into the YOLOv8m feature extraction pipeline to enhance spatial and channel attention while preserving real-time inference capability. Multiple experimental training configurations were evaluated, including raw-image training, traditional augmentation techniques, CycleGAN-augmented datasets, and attention-enhanced detection models. Performance was assessed using precision, recall, mAP@0.50, and mAP@0.50–0.95 metrics. The proposed YOLOv8m with CBAM and CycleGAN framework achieved the best overall performance, with a precision of 0.940, a recall of 0.910, mAP@0.50 of 0.945, and mAP@0.50–0.95 of 0.725. These results indicate that the combined use of GAN-based augmentation and attention mechanisms significantly improves detection performance under a realistic aquatic environment.

Keywords: Milkfish disease detection, YOLOv8, CycleGAN, CBAM, Synthetic data augmentation

INTRODUCTION

Milkfish (Chanos chanos) is a key aquaculture species in Southeast Asia, contributing substantially to regional food security and the livelihoods of coastal populations (Martínez, Tseng, & Yeh, 2006). Despite its importance, disease outbreaks remain a persistent challenge, spreading rapidly and causing considerable economic losses. Early detection of diseases in aquaculture is difficult due to subtle initial symptoms that often become noticeable only as the disease advances. In most fish pens, particularly in the Philippines, health monitoring relies on manual observation, which is labor-intensive and dependent on the experience of fish farmers.

Several technological advances, particularly in computer vision, have been adopted to perform this task. Non-intrusive technologies such as sensors, cameras, and IoT have been used in some regions to assist fish farmers in managing their pens (Chang, et al., 2022). However, despite of these advances, disease detection still poses an ever-existing challenge especially when working in the underwater environments, such as unfavourable water turbidity, varying light conditions, the movement of the fish, and the angles of view from the camera, which usually tend to lower the quality of the images taken and make it hard even for the trained observer to make accurate judgments about the image data being observed. Manual inspection tends to be inconsistent, and the observer may often miss subtle abnormalities, especially when monitoring a large number of fish.

Deep learning models have recently been used to automate underwater image analysis. Convolutional Neural Networks (CNNs) have been effectively applied to tasks such as fish counting, species identification, and behavior monitoring. The You Only Look Once (YOLO) family of detectors, based on CNNs, has gained

popularity due to its architecture, which enables fast and accurate detection. Unlike traditional two-stage models, YOLO treats detection as a single regression problem, allowing simultaneous prediction of bounding boxes and class probabilities at over 45 frames per second (FPS) (Deepika, Raghuveer shetty, & G N, 2020). This reason makes this model a powerful tool for early signs of fish disease detection underwater in real time.

However, the performance of deep learning models depends strongly on the availability and diversity of training data (Alfred, Leo, & Kaijage, 2025). In the context of milkfish disease detection, collecting large numbers of diseased samples is often impractical, resulting in significant class imbalance between healthy and diseased fish. Models that are trained solely on raw images may struggle to generalize to new environments or novel disease manifestations (Komori & Eguchi , 2019). While traditional data augmentation techniques can partially mitigate this issue, they typically introduce only limited variation and may not capture the complexity of real-world visual conditions (Prova, 2024). An alternative way to increase the number of diseased samples is to use a Generative Adversarial Network (GAN). Among these, CycleGAN has gained popularity for its unpaired image-to-image translation framework. This is particularly helpful when dealing with a limited class in the dataset, where controlled acquisition of a paired dataset is rarely feasible (Chen, Pan, & Wu, 2023). For instance, if the diseased fish sample is relatively small in number compared to the healthy fish sample, generating synthetic data would significantly help balance the dataset.

Model architectural refinements have been explored in several studies demonstrating that they further improve performance. An example is integrating attention mechanisms, particularly the Convolutional Block Attention Model (CBAM). Doing so allows neural networks to highlight the informative spatial and channel-wise information and reduce the influence of background noise. Integrating an attention mechanism can prove useful, especially in underwater imaging, where disease indicators can cover only small parts of the fish body (Khabusi, Huang, & Lee, 2023) (Biswal, Kumar, & Punuri, 2024). The YOLOv8m model typically treats all parts of the image equally during detection; however, an underwater image may contain a lot of unimportant information, such as the water background, reflections, light noise, floating particles, shadows, and fish parts that are not affected by disease. CBAM will focus on important regions with physical abnormalities and pay less attention to irrelevant regions.

This paper will provide a systematic evaluation of YOLOv8-based milkfish disease detection models across various architecture configurations and dataset configurations. Different experimental setups were performed using raw data, traditional augmentations, GAN-based dataset augmentation, and finally, the integration of the Convolutional Block Attention Model to the YOLOv8 medium variant. The performance is evaluated using standard detection metrics, including precision, recall, mAP (0.50), mAP (0.50-0.95), and processing speed.

The novelty of this study lies in the integration of CycleGAN-based augmentation and YOLOv8 model enhancement with CBAM design for real-time disease detection. It mainly addresses the challenges of the underwater detection task by improving the focusing mechanism of YOLOv8 and addresses the limitation of disease samples through synthetic data augmentation.

RELATED WORKS

This section reviews relevant studies across the two categories: (1) Enhancement and modification of YOLO-Based Object Detection in Aquatic and Real-Time Applications, and (2) GAN-Based Data Augmentation for Limited and Imbalanced Datasets, to establish the technical foundation of the proposed approach and to identify existing research gaps that motivate the design of YOLOv8m with CycleGAN and the CBAM framework.

Enhancement and modification of YOLO-Based Object Detection in Aquatic and Real-Time Applications

YOLO-based models have become increasingly prominent in object detection tasks, particularly in a real-time monitoring setup. Unlike traditional two-stage detectors, such as R-CNN, Faster R-CNN, Feature Pyramids, etc., which are characterized by modular architectures that separate the detection processes and classification phase, YOLO processes the entire image in a single forward pass, enabling high detection accuracy

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

while maintaining high performance (Bharti, 2024). This architecture makes YOLO practically suitable for real-time, underwater object detection tasks.

The DyAqua-YOLO achieved a mAP@50 of 0.918 on the Dense Underwater Objects (DUO) public dataset at 21 FPS on embedded hardware (Li & Peng, 2025). Another enhanced YOLOv8, the CEH-YOL, which incorporated a high-order deformable attention (HDA) to enhance spatial feature extraction, improved spatial pyramid pooling-fast (ESPPF) for extraction of object attributes, and customized composite detection (CD) to improve accuracy and inclusivity, reached 88.4% mAP at 156 FPS (Feng & Jin, 2024). On the other hand, the UOD-YOLO, a framework based on YOLO11n, integrates three components: (1) the RepVit for feature representation, (2) a scale-sequence feature fusion (SSFF), and a triple feature encoding (TFE) to enhance multi-scale feature extraction and representation, and (3) an attention mechanism (CPAM) in the head layer to strengthen small-object detection. The integration improved the mAP@50 of its baseline model by 3.5% while reducing parameters by 36.3% and achieving 279.8 FPS (Xi & Yin, 2025). The RTL-YOLOv8n, with a modified head (LCD-Head), reduced the computational requirements by 31.6%, and an enhanced Focaler–MPDIoU loss function, showed a 1.5% increase in mAP@0.5 and a 5.2% improvement in precision compared to other models (Guanbo, et al., 2024). The AOD-YOLO, another improved model based on YOLOv11s, with four enhancements: first, incorporating Sobel operators and pooling operations (SPFE) to effectively extract features, particularly edges and global structural features. Second, the integration of RGL (RepConv and Ghost Lightweight) for feature mapping, parameter reduction, and computational cost reduction. Third, the integration of MDCS (Multiple Dilated Convolution Sharing Module) for improved multi-scale target recognition. Lastly, the C2PSA module is extended to C2PSA-M (Cascade Pyramid Spatial Attention—Mona) by integrating the Mona mechanism to enhance visual signal processing and information distribution. The model achieved a 2.6% improvement in accuracy over the YOLOv11s URPC2020 dataset and a 2.4% improvement on the RUOD dataset.

However, despite these significant improvements to the YOLO architecture, most of these focus on the architectural components, such as feature fusion, attention mechanisms, and multi-scale representation, to improve accuracy and speed. DyAqua-YOLO, CEH-YOLO, and UOD-YOLO exhibited strong performance on public datasets; however, these generally assume the availability of large and balanced datasets. One of the major limitations of these studies is the lack of emphasis on class imbalance and the scarcity of data, especially in mapping subtle information relevant to fish diseases. Additionally, many modified YOLO architectures incur higher computational costs due to their complexity, which can limit real-time deployment.

This study addresses these gaps by integrating a more lightweight CBAM attention mechanism into YOLOv8 to enhance feature representation. The dataset limitations and class imbalance were addressed by introducing GAN-based data augmentation.

GAN-Based Data Augmentation for Limited and Imbalanced Datasets

As previously mentioned, data scarcity has been a major challenge in fish disease detection, as it can lead to class imbalance. Data scarcity and class imbalance are two distinct challenges that often co-occur (Minouei, Soheili, & Stricker, 2024). Traditional techniques such as flipping, rotation, and color adjustments may be used to augment limited samples; however, they may not capture essential visual information in real-time settings (Gong, Wang, Li, Chandra, & Liu, 2021).

This section explores studies demonstrating how GAN-based augmentation effectively improves object detection performance while addressing class imbalance and dataset scarcity. In a study of Deng et al., 2025 (Deng, Ryu, & Lee, 2025), an image generation method was developed to create synthetic fish disease images from healthy fish, improving the performance of ResNet50 by 6.0% in precision, 4.4% in recall, 3.8% in F1-score, and 4.4% inaccuracy (Deng, Ryu, & Lee, 2025). The IDA-GAN, a GAN-based augmentation method, introduced a variational autoencoder coupled with a GAN to learn minority and majority classes. The generative model later learned useful features that enabled it to generate samples for the minority class. Compared with traditional GANs, the model can generate more diverse minority samples, thereby benefiting the imbalanced classification task (Yang & Zhou, 2021). Another variant of GAN, the StyleGAN2-ADA, has also demonstrated a 17.9% improvement in detection performance across all algal genera and a 32.1% improvement on rare genera

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

by enhancing algal image instance segmentation and generating artificial single-algal images to address data scarcity and imbalance (Fung, Chan, Lo, & Tsang, 2023). R. Yuwana et al. 2020 used GAN and DCGAN for data augmentation to detect tea leaf diseases, demonstrating DenseNet accuracy improvement by 88.84% compared to the baseline model (Yuwana, et al., 2020). Similar studies like the balancing GAN (BAGAN) which pragmatically applied autoencoders to initialize GAN, and collaborative discrimination-enabled GAN (CoD-GAN) which employed collaborative discrimination framework to improve discrimination robustness, both studies aim to address class imbalance through GAN-base data augmentation (G. Mariani, F. Scheidegger, R. Istrate,, C. Bekas, & A. Malossi, 2018) (Zhang, Li, & Liu, 2022).

Despite the effectiveness of these methods, IDA-GAN, StyleGAN2-ADA, BAGAN, and CoD-GAN in improving model performance and addressing class imbalance by generating synthetic images, these approaches are often evaluated in controlled environments and do not fully consider real-world challenges such as noise, blur, motion blur, and lighting variability. Additionally, integration of GAN-based data augmentation with real-time detection remains limited. On the other hand, this study proposes using CycleGAN for unpaired data augmentation. It employs the generated data in an attention-based YOLOv8m model to improve the detection and localization of milkfish disease features in complex underwater environments.

PROPOSED WORK

This study proposes a hybrid deep learning framework designed for real-time, early detection of milkfish disease in an underwater environment through visual analysis. This method utilizes the object detection capabilities of YOLOv8, medium variant, enhanced with the Convolutional Block Attention Module (CBAM) to improve feature representation by emphasizing relevant spatial and channel information in the input image. The objective of this modification is to enable the model to focus on disease-relevant regions, such as subtle lesions and abnormal texture patterns in milkfish.

System Architecture

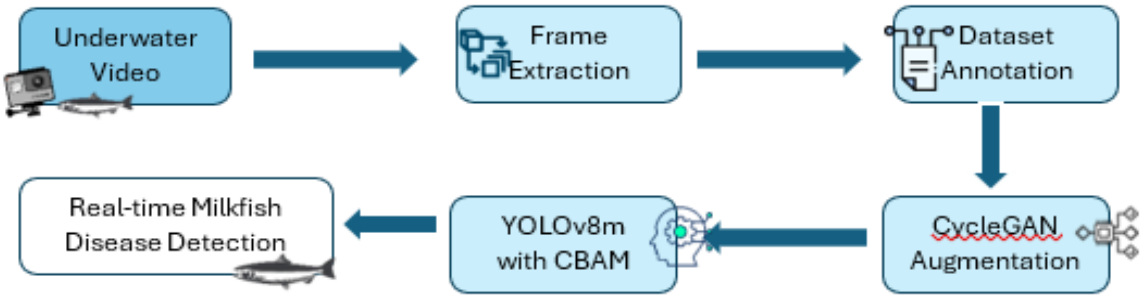

Figure 1. Proposed system workflow for real-time milkfish disease detection using YOLOv8m with CBAM and CycleGAN-based data augmentation.

The system architecture of the proposed model follows a sequential pipeline that transforms underwater video input data into detection output using a modified YOLOv8m model. As shown in Figure 1, the process begins with gathering video footage of milkfish in an actual underwater environment using an underwater camera. The video recordings are then processed to extract individual frames. After extraction, the images were annotated in YOLO format, with each frame labeled with bounding boxes. To generate more data and to balance the two classes, CycleGAN augmentation is applied, improving model generalization. The augmented dataset is used to train the YOLOv8m with CBAM enahncement. The attention module will help the model focus on regions of images relevant to milkfish diseases, such as lesions and abnormal texture patterns on fish scales. Finally, the trained model performs real-time early detection of milkfish disease, producing bounding-box predictions and classification outputs indicating whether the milkfish in the frame is diseased or healthy.

Dataset Preparation

The dataset used in this study was collected in a real aquaculture environment and is not publicly available. The use of real-world underwater milkfish data ensures that the model's performance will be evaluated

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

under realistic environmental conditions, including water turbidity, lighting conditions, and background noise. The milkfish video footage was collected from a fish pen in Dagupan City, Pangasinan, Philippines. A single fish pen was used to collect video data for training. The GoPro Hero 11 Black camera was used, positioned approximately 2 feet below the water surface. The footage was recorded within a 30-minute window at 30 FPS in a typical underwater environment under natural daylight. To use the videos as training data, an OpenCV Python script was used to extract individual frames at 2 fps, resulting in 17,211 images.

Water turbidity is a major issue in acquiring quality images underwater. Aquatic environments, especially fish pens, often exhibit high turbidity due to several factors. Fish pens contain many suspended particles from unconsumed fish food, fish waste, soil erosion, algae, plankton, and decaying organic matter. These cause images to be cloudy and noisy because the light penetrating the water is being scattered (Gaude & Borkar, 2020). This also lessens the contrast of the images, resulting in low transparency and reduced visibility.

The continuous fish movements likewise produce blurring in most images, especially when fish move quickly within the camera's field of view. Blurriness can cause the loss of crucial details, creating confusion in a model's algorithm and affecting its classification performance (Garcia, Alvarez, & Marcia, 2023). Blurred images obstruct definitive characteristics, making it difficult for a model to classify, identify, and localize important information that we are trying to detect, that is, subtle physical abnormalities.

Figure 2. Sample Repetitive images extracted from the underwater footage. Fish motion captured and extracted at 2 fps resulted in repetitive images that needed to be manually discarded during the selection process.

Another problem encountered when extracting video data is the generation of too many repetitive frames, as shown in Figure 2, due to fish movement, which may cause model overfitting. Because the images were captured at 2 fps, many consecutive frames had nearly identical positions, angles, lighting conditions, and backgrounds. Model overfitting may lower its ability to generalize to new data. The model treats these duplicate images as highly important patterns, causing it to memorize and specialize in them rather than learn the general pattern, leading to performance bias, overconfidence, and inaccurate performance on unseen details.

To ensure that the dataset used to train the model is optimal, the researchers manually filtered the extracted images, carefully removing unwanted blurry, repeated, and noisy images; the final dataset consisted of 1,232 healthy images and 172 diseased images. The filtering process considered image quality, visual usefulness, and image uniqueness. The resulting 1,404 dataset poses another challenge: class imbalance, which may cause the model to be biased towards the majority class.

The dataset was organized into detection classes: healthy milkfish and diseased milkfish. The healthy class refers to samples with normal physical appearance, no visible lesions, and clear scale patterns. The diseased

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

class, on the other hand, refers to milkfish having visible physical abnormalities such as lesions, discoloration, wounds, pale or dark patches, and abnormal scale patterns. Since the model is intended for deployment in a real-time setting, the disease category was based on observable external manifestations rather than laboratory-confirmed disease types. Therefore, the model was trained to detect whether the fish appeared healthy or diseased based on visible physical indicators. The dataset was annotated and labeled in Roboflow.

CycleGAN Dataset Augmentation

A variant of the Generative Adversarial Network, CycleGAN, has been employed to augment the disease samples for training. CycleGAN, as previously mentioned, is a deep learning framework designed for image-to-image translation that does not require paired training data, making it an ideal model for generating and adding more realistic diseased fish samples to train the modified YOLOv8m model (Zhu, Park, Isola, & Efros, 2017).

The CycleGAN training process starts with the careful selection of 359 healthy and 323 diseased milkfish images. The initial dataset, containing only 172 diseased samples, was augmented using traditional techniques such as saturation adjustments, background removal, increased sharpness, cropping, flipping, and additional public images from the web to improve diversity, thereby normalizing the sample size to around 300 images across the two domains. This pre-processing step ensures that the visual features relevant to milkfish disease detection, e.g., physical abnormalities such as lesions and discolorations, are clear and consistent, enabling CycleGAN to accurately map variations between healthy and diseased samples.

The transformation of healthy samples into visually realistic diseased variants starts by learning a mapping between two domains – healthy (X) and diseased (Y) – without requiring pixel-level correspondences (Zhu, Park, Isola, & Efros, 2017).

The adversarial loss for the mapping $G: X \longrightarrow Y$ with discriminator D_Y is defined as (Zhu, Park, Isola, & Efros, 2017):

$$L_{GAN}(G, D_Y, X, Y) = E_{y_{Pdata_{(y)}}}[log\, D_y(y)] + E_x{\sim}_{Pdata_{(x)}}[log(1 - D_Y(G(x)))] \quad (1)$$

On the other hand, drawing the adversarial loss for $F: Y \longrightarrow X$ with discriminator D_X, we get the equation (Zhu, Park, Isola, & Efros, 2017):

$$L_{GAN}(F, D_X, X, Y) = E_x{\sim}_{Pdata_{(x)}}[log\, D_x(x)] + E_y{\sim}_{Pdata_{(y)}}[log(1 - D_Y(G(y)))] \quad (2)$$

To ensure that the translated image from one domain and back remains close to the original, CycleGAN uses cycle consistency loss (Zhu, Park, Isola, & Efros, 2017):

$$L_{cyc}(G, F) = E_x{\sim}_{Pdata_x}[\|F(G(x) - x\|1] + E_y{\sim}_{Pdata_y}[\|G(F(y)) - y\|1] \quad (3)$$

The objective is to combine all the loss functions, weighted by a factor λ for cycle consistency (Zhu, Park, Isola, & Efros, 2017):

$$L(G, F, D_X, D_Y) = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, Y, X) + \lambda L_{CYC}(G, F) \quad (4)$$

Where:

- G represents the generator that maps images from the healthy domain (X) to the diseased domain(Y),
- F represents the inverse generator that maps images from the diseased domain (Y) back to the healthy domain (X),
- D_X and D_Y represent the discriminators that distinguish between real and generated images in the two domains,

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

- $E_{x \sim Pdata_{(x)}}$ represents an image sampled from the healthy domain,
- $E_{y_{Pdata_{(y)}}}$ represents an image sampled from the diseased domain,
- λ is the weighting parameter that controls the importance of the cycle consistency loss.

The cycle consistency loss ensures that the generated image, when translated back to its original domain, remains structurally similar to the input image. This is particularly important in this study, as it preserves key visual features such as fish shape and texture while introducing realistic disease characteristics.

The CycleGAN training process was conducted for 100 epochs to give the model time to move beyond initial random guessing and settle into a decent level of accuracy. A learning rate of 0.0002 was used to balance convergence speed and training stability. This small learning rate allows the Generator to gradually learn realistic disease patterns, such as lesions and discolorations, with less distortion and noise in the generated images. Additionally, a learning rate decay was applied after 50 epochs, so for the first 50 epochs, the learning rate was held constant at 0.0002, allowing a strong mapping between the two domains. After that, a gradual linear reduction in the learning rate is employed to enable fine-tuning in the later stages of training. Gradually reducing the learning rate allows the model to make subtler, more precise updates, thereby preventing overfitting. Moreover, the Adaptive Moment Estimation (Adam) Optimizer was used during training, which is widely adopted in machine learning due to its efficiency and stability. The Adam optimizer provides adaptive learning rates and helps stabilize training. It helps manage noisy gradients from small batches by combining momentum and variance-based updates to improve convergence. The Adam optimizer was used to adjust the learning rate during training.

Augmented Dataset

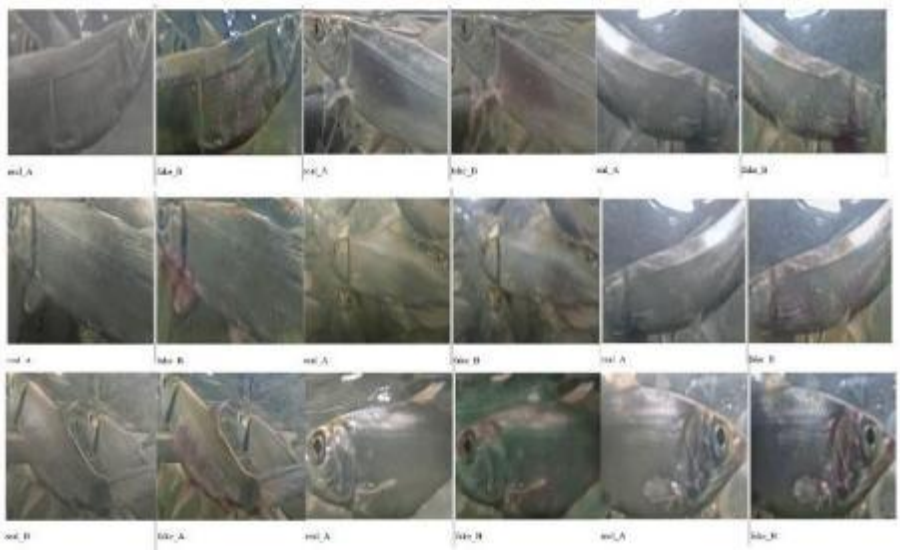

Figure 3. Sample output of CycleGAN milkfish dataset augmentation

The CycleGAN implementation produced an additional 700 synthetic diseased images. These images were then combined with the original raw images extracted from the video data. This created a more balanced dataset to be used in the YOLOv8m-CBAM training. This phase significantly improved the representation of the diseased class due to the aforementioned limitation. To further diversify the dataset, the additional CycleGAN training samples were combined with the original dataset, bringing the total to 2,647 images. The dataset was divided into training, validation, and test sets in a 70:20:10 ratio. The model learning was based on the training set, the validation set was used for hyperparameter tuning, and the final evaluation was applied on the test set.

Figure 3 shows original (real) and generated (fake) images generated by CycleGAN, where healthy fish variants were translated into synthetic diseased images. The generated images show visible lesion-like patterns and color variations, demonstrating the capability of the model to generate synthetic disease features that will be used in training the proposed model.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

Dataset Versions	Healthy	Diseased	Total
Raw Dataset	1,232	172	1,404
Raw + Basic Augmentation	1,710	658	2,368
After CycleGAN Augmentation	1,232	700	1,932
With the CycleGAN training dataset	1,232	1,415	2,647
Final Dataset (with CycleGAN augmentation & selection)	1,232	1,415	2,647

Table 1: Dataset Versions. Number of datasets used and their distribution. Source: Dataset collected and processed by the researchers.

To evaluate the quality of the CycleGAN-generated images, the Fréchet Inception Distance (FID) was computed between the real diseased images and the generated synthetic images. The computed FID score was 89.93, indicating a noticeable distribution difference between the real diseased images and the CycleGAN-generated diseased images. The result may be attributed to the limited number of real samples, along with environmental factors such as noise, blur, and lighting variations, as well as the difficulty of generating consistent disease-related patterns in underwater fish images. Despite the distributional differences, the generated dataset was still useful for the model, as it was used for data augmentation to improve diversity and address class imbalance.

The Proposed Attention-Enhanced YOLOv8m Architecture

The proposed model is based on the YOLOv8 medium variant and is enhanced with the Convolutional Block Attention Module (CBAM) to improve feature representation. The YOLOv8 model has been widely adopted as an efficient object detection model, a single-stage detector that is ideal for real-time detection, offering a well-balanced balance between detection accuracy and computational efficiency. However, in an underwater setting where the environment can be noisy, turbid, and subject to varying lighting, a standard convolutional layer may not capture subtle physical patterns associated with a fish disease. To address this limitation, the CBAM is integrated into the original YOLOv8m architecture to enable the model to specifically detect relevant spatial regions and informative feature channels to improve disease detection, especially in an underwater environment.

The Baseline YOLOv8m Architecture

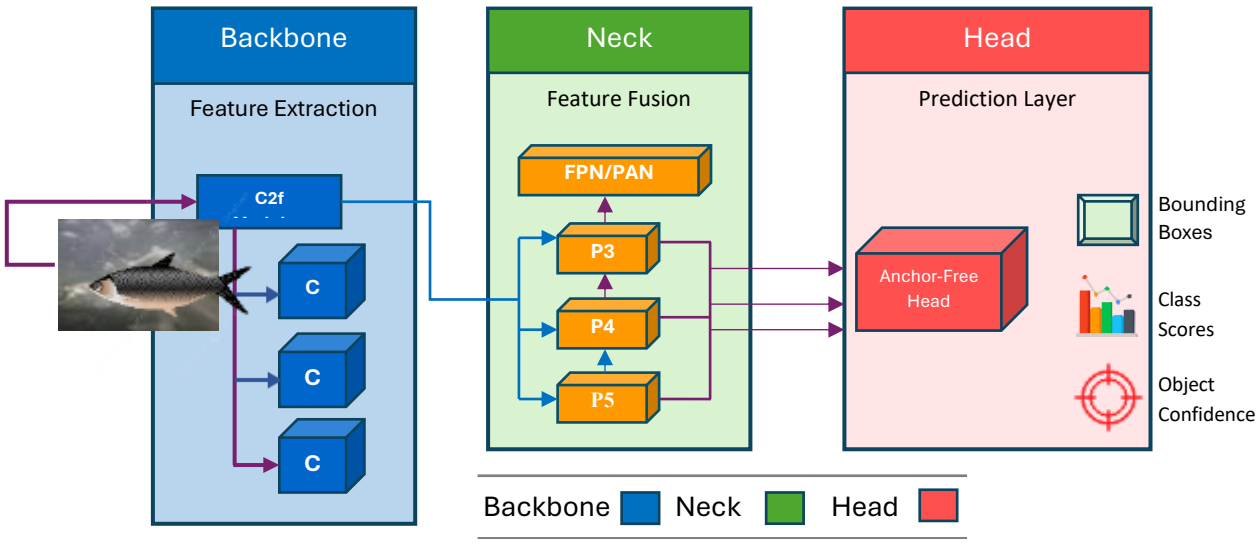

Figure 4. The YOLOv8 Architecture

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

The proposed model builds on the YOLOv8m architecture, a state-of-the-art single-stage object detection model that offers a superior balance between computational efficiency and precision, making it ideal for real-time detection tasks.

The YOLO architecture consists of three primary components: the backbone, the neck, and the head. Each of these components plays an important role in the detection tasks. The backbone is responsible for feature extraction. It consists of a convolutional layer combined with advanced modules such as the C2f (Cross-Stage Partial Bottleneck with 2 convolutions) blocks to improve gradient flow and reduce computational redundancy. These mechanisms progressively transform the input image into feature maps, capturing important information such as edges, textures, object shapes, and structures. These modules are crucial for identifying fish in the frame and for detecting subtle physical abnormalities, such as discoloration and texture irregularities. The neck outputs a set of feature maps, commonly denoted C3, C4, and C5, which represent different levels of abstraction. The neck, on the other hand, fuses the different feature maps from the backbone using a Feature Pyramid Network (FPN) and a Path Aggregation Network (PAN), allowing the model to detect objects of varying sizes by combining high-resolution spatial information with deep semantic features. The aggregated feature maps are denoted as P3, P4, and P5, which correspond to small, medium, and large object scales. This is important in addressing the distance, orientation, and size of an object in a given frame. Finally, the detection head of the YOLOv8 is responsible for the prediction tasks. For each of the feature maps, the head outputs bounding box coordinates, class probabilities, and object coordinate scores. The detection head produces a real-time prediction based on the refined feature maps.

The Integration of CBAM

To enhance the feature representation of the baseline model YOLOv8, this study proposes the integration of the Convolutional Block Attention Module (CBAM) block, a lightweight attention mechanism that applies channel attention and spatial attention to refine the feature map extracted from the backbone module. This will enable the model to focus on, or "pay more attention" to the regions and features relevant to milkfish disease detection, particularly in the challenging underwater environment.

The proposed model particularly integrates CBAM into the neck layer (feature fusion), specifically in the P3, P4, and P5, after they were generated by the neck layer. These layers correspond to layer 15 (small-scale layer), layer 18 (medium-scale layer), and layer 21 (large-scale layer). This placement is strategically selected because in this layer, the neck aggregates features from different scales and sizes, and disease patterns may appear at varying sizes and different locations, and refining these aggregated features will improve detection performance.

Layer	Feature Level	Purpose
15	P3	Detect small objects (tiny lesions)
18	P4	Detect medium features
21	P5	Detect large objects (whole fish context)

Table 2. P3, P4, and P5 layer representation and their purpose.

The CBAM uses two linear sub-modules, the channel-attention and the spatial attention. The channel-attention sub-module assigns weights to each of the channels depending on what features are important. The feature map is first processed using the Global Average Pooling (GAP)AvgPool(F), and Global Max Pooling (GMP)MaxPool(F). The output of this process will then be passed through a shared multi-layer perception (MLP). The results will then be combined with a sigmoid function σ. The generated channel attention map will then be multiplied by the input feature map (Woo, Park, Lee, & Kweon, 2018). The equation below demonstrates this process:

$$M_c(F) = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F)) + (\mathrm{MLP}(\mathrm{MaxPool}(F)) \quad (5)$$

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

The spatial attention module, on the other hand, focuses on where the important features are located within the input image. The process starts with the compression of the feature map along with the channel axis using average pooling AvgPool(F) and max pooling MaxPool(F). The compressed feature map will then be passed through convolutional layers to generate a spatial feature map using a sigmoid function. The result will then be multiplied by the input feature map. This process will highlight the regions where disease is present and improve the localization of the infected areas (Woo, Park, Lee, & Kweon, 2018). The equation below demonstrates the whole process of generating the spatial feature map:

$$M_s(F) = \sigma(f^{7x7}([AvgPool(F); MaxPool(F)])) \quad (6)$$

The final output of the combination of these two sub-modules will be (Woo, Park, Lee, & Kweon, 2018):

$$F' = M_s(M_c(F) \cdot f) \quad (7)$$

Where:

- F ∈ ℝ^{C × H × W} represents the input feature map, where C is the number of channels, H is height, and W is width;
- AvgPool(·) and MaxPool(·) denote global average pooling and global max pooling operations, respectively;
- MLP denotes a shared multi-layer perceptron used to compute channel attention;
- σ denotes the sigmoid activation function;
- f^{7×7} represents a convolution operation with a 7×7 kernel;
- $M_c(F)$ is the channel attention map;
- $M_s(F)$ is the spatial attention map;
- F' is the intermediate feature map after channel attention;

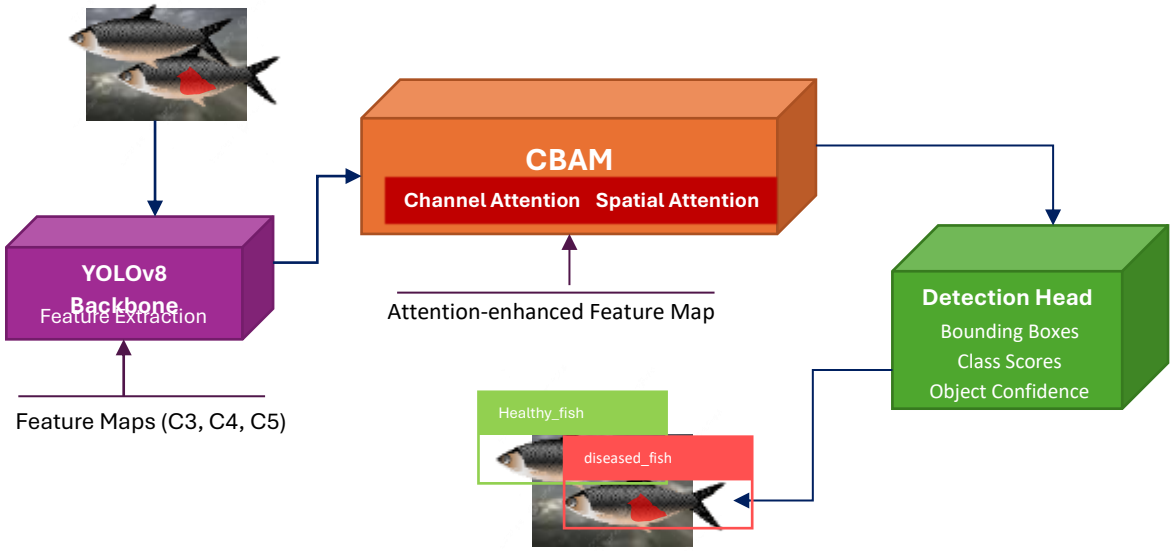

Figure 5. The proposed YOLOv8-CBAM architecture

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

In the underwater environment, fish disease can be small, subtle, and can blend into the background, along with other factors that may impair the quality of the image. The integration of the CBAM to the baseline model will enable it to focus on fine-grained information and ignore irrelevant visual information, thereby improving the detection accuracy in the real world.

The figure shows the process of detecting milkfish disease using the proposed model. A feature map is created from the YOLOv8 backbone, denoted as C3, C4, and C5. This feature map is then refined in the neck layer, where CBAM is integrated through channel attention and spatial attention. The attention-enhanced feature map is then forwarded to the detection head to predict bounding boxes, class labels, and confidence scores for the healthy and diseased milkfish.

Experimental Backbone Modification using ResNet50

To further evaluate the impact of the previously discussed architectural enhancement, this study likewise explored an alternative modification by replacing the CSPDarknet53 backbone of the YOLOv8m with a ResNet-based feature extractor. ResNet, or the Residual Network, is a deep convolutional neural network model that is known for its ability to solve the vanishing gradient problem using "skip-connections" to bypass layers, allowing training of deeper networks (He, Zhang, Ren, & Sun, 2016; Liu & Goh, 2025). This modification was performed to enable deeper feature extraction capability of the YOLOv8m model and potentially improve its ability to extract complex visual patterns and capture subtle disease-related features such as lesions, discolorations, and abnormal texture patterns. However, the integration of ResNet with the YOLOv8 architecture introduced several challenges, particularly the compatibility of the modified backbone's resulting feature maps and channel sizes with the neck and the detection head layers. This difference requires additional adjustments, which adds to the model's computational costs and complexity.

This modification is included as part of the experimental analysis for comparison with the performance of the proposed model. While the ResNet-based modification demonstrates improvement in feature extractions, the attention-focused CBAM-based enhancement offers a more efficient and stable solution to the real-time detection without significantly increasing computation cost and model complexity.

Hardware and Software details

All experiments conducted in this study were executed on an Acer Predator Helios 16 laptop equipped with an Intel Core i9 processor, NVIDIA RTX 4070 GPU with 8GB VRAM, and 16GB of RAM. This is for the computational requirements of the training runs. The YOLOv8 model, especially, requires significant processing power for object detection, and CBAM, although considered lightweight, adds to the computational load with its convolutional operations. The models were implemented using Python and trained using the Ultralytics YOLOv8 framework.

The Experimental Setup

In the experimental setup, all models were trained with the same baseline parameters to ensure a fair comparison. The YOLOv8m model was used as the base architecture, with an input image size of 640×640 pixels. An Adam optimizer with a learning rate of 0.001 and a batch size of 16 was applied all throughout the 100-epoch training process. The evaluation metrics include precision, recall, mean Average Precision at IoU threshold 0.50 (mAP@0.50), and mean Average Precision across multiple IoU thresholds from 0.50 to 0.95 (mAP@0.50–0.95).

Different training configurations (T1–T8) were designed to analyze the impact of dataset augmentation and model enhancement. These include training on raw data, traditional augmentation, CycleGAN-based augmentation, and integrating CBAM into the YOLOv8m architecture.

The performance of the attention-enhanced YOLOv8m was evaluated using four key metrics commonly used in object detection: precision, recall, mAP50, and mAP50-95. The experiments were conducted to evaluate the impact of data augmentation techniques and model enhancement and modification, particularly GAN-based synthetic augmentation and attention-based model enhancement.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

OVERVIEW OF THE RESULTS

As shown in Table 3, the baseline YOLOv8m model in T1, trained on raw images, achieved a precision of 0.817, a recall of 0.906, and a mAP@50 of 0.925. This indicates that the baseline model has strong detection capability; however, it also exhibits low precision due to false positives in a complex underwater environment. The introduction of ResNet into the backbone of the baseline model, as shown in T2 to T3, improved the precision score to 0.882, demonstrating the better feature-extraction capability of ResNet; however, recall decreased slightly, indicating reduced detection coverage, which means that there are missed subtle cases of disease detection. This also did not consistently improve the object detection performance because deeper feature extraction alone does not always guarantee improved performance, especially in underwater environments and

Training	Dataset Description	Model	Images	P	R	mAP50	mAP50-95
T1	Raw Images only	YOLOv8m	1,404	0.817	0.906	0.925	0.682
T2	Raw Images	YOLOv8mRN50	1,404	0.882	0.865	0.922	0.684
T3	Raw image with Basic Augmentation	YOLOv8mRN50	2,368	0.882	0.885	0.930	0.686
T4	Raw images with CycleGAN	YOLOv8mRN50	1,932	0.885	0.863	0.920	0.685
T5	Raw images with Basic Augmentation and CycleGAN Augmentation	YOLOv8mRN50	3,364	0.902	0.898	0.928	0.693
T6	Raw images with Extended augmentation (3×), and CycleGAN Augmentation	YOLOv8mRN50	4,749	0.882	0.862	0.921	0.679
T7	CycleGAN and CycleGAN Training images	YOLOv8mRN50	2,647	0.910	0.865	0.936	0.710
T8	CycleGAN and CycleGAN Training images	YOLOv8m-CBAM	2,647	0.940	0.910	0.945	0.725

Table 3. Comparative Performance Table.

Table 3 shows the comparative performance of YOLOv8-based models under different training configurations (T1-T8). with limited and imbalanced datasets. This was evident in T2 to T8, where the model became more selective but missed more disease samples. The result also suggests that the ResNet-based modification improved feature discrimination but reduced detection coverage. The addition of basic dataset augmentation, as shown in T3, improved recall to 0.885 and mAP@50 to 0.930. This demonstrates the benefit of increased data variability. However, improvements in mAP@50–95 remained minimal, indicating limited gains in terms of localization accuracy.

The use of CycleGAN-based augmentation indicated in T4 to T7 showed a more remarkable impact. While T4 produced modest improvements, combining CycleGAN with raw images and traditional augmentation, as shown in T5, resulted in a more balanced performance with a precision score of 0.902 and a recall of 0.898, demonstrating the importance of dataset variability and class balance. On the other hand, excessive augmentation resulted to performance degradation, as seen in T6, particularly in recall of 0.862 and mAP@50–95 of 0.679, indicating that excessive traditional augmentation introduces noise and redundancy. This meant that instead of effectively adding relevant data, excessive augmentation may introduce redundant, noisy, or less realistic visual patterns, which reduced the model's ability to generalize. The larger dataset also has likely increased deviation that was not fully representative of real underwater disease features, causing lower recall and localization accuracy. T7 demonstrated the best performance among the ResNet-based configurations, achieving a precision of 0.910, a recall of 0.865, an mAP@0.50 of 0.936, and an mAP@0.50–0.95 of 0.710, suggesting that a well-

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

balanced augmented dataset improves both detection and localization. The proposed model in T8, which integrates CBAM into YOLOv8m neck, demonstrated the highest overall performance with a precision score of 0.940, a recall of 0.910, a mAP@50 score of 0.945, and a mAP@50–95 of 0.725. Compared to T7, the improvements in precision, recall, and mAP@50 confirm that CBAM enhances feature discrimination by focusing on disease-relevant regions.

These results are considered to be satisfactory because the proposed model achieved remarkable improvement compared to the baseline YOLOv8m's performance. The baseline in T1 obtained a precision of 0.817, a recall of 0.906, an mAP@0.50 of 0.925, and an mAP@0.50–0.95 of 0.682. In comparison, the proposed T8 model achieved an improvement of 0.123 in precision, 0.02 in mAP@0.50, and 0.043 in mAP@0.50–0.95 over the baseline.

Overall, the results demonstrate that while backbone modification provides moderate improvements, the combination of CycleGAN-based augmentation and CBAM attention yields the most consistent and significant performance gains. The proposed YOLOv8m-CBAM model effectively balances detection accuracy and localization performance, making it suitable for real-time milkfish disease detection in underwater environments. The results also show that the performance improvement was not caused by augmentation alone but by the controlled combination of CycleGAN-based augmentation and CBAM attention. While T5 achieved balanced results, T6 showed performance degradation after excessive augmentation, indicating that increasing the dataset size without careful selection can reduce generalization. Therefore, the final T8 result supports the effectiveness of the proposed approach in improving disease detection under challenging underwater conditions.

The next table, Table 4, reflects the speed-related metrics of the three models, such as the processing time of each model during preprocessing, inference, and post-processing. Based on the results, the proposed model YOLOv8m-CBAM maintained real-time inference capability while improving detection accuracy. Compared with YOLOv8mRN50, the YOLOv8m-CBAM model was faster and more accurate. Therefore, the inclusion latency confirms that the proposed model is suitable for real-time milkfish disease detection.

Version	P	R	mAP50	mAP50-95	Preprocess (ms)	Inference (ms)	Postprocess (ms)
YOLOv8m	0.90	0.88	0.93	0.69	0.2	6.8	0.9
YOLOv8mRN50	0.91	0.87	0.94	0.71	0.3	12.32	0.5
YOLOv8m-CBAM	0.94	0.91	0.95	0.73	0.3	6.3	0.6

Table 4. Accuracy and speed performance comparison of the evaluated YOLO-based models using the T8 dataset version.

Detailed Performance Analysis

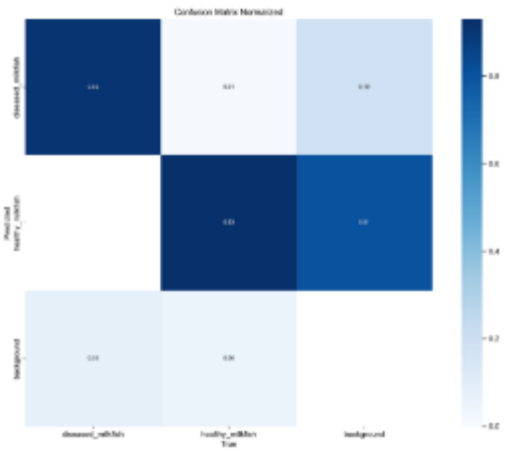

Figure 6. Confusion matrix of the proposed YOLOv8m-CBAM model showing classification performance for healthy and diseased milkfish.

ISSN No. 2321-2705 | DOI: 10.51244/IJRSI |Volume XIII Issue V May 2026

A detailed performance analysis, as shown in the following figures, was conducted to further evaluate the effectiveness of the proposed model YOLOv8-CBAM using confusion matrix, precision-recall, and qualitative detection analysis.

As shown in Figure 6, the normalized confusion matrix indicates a strong classification performance with high true positive rates for both healthy and diseased milkfish. It was included to evaluate how well the proposed YOLOv8m-CBAM model distinguished between healthy and diseased milkfish. This matrix shows minimal misclassification between classes, confirming that the model effectively distinguishes subtle visual differences. This improvement can be attributed to the integration of CBAM, which enhances feature discrimination, and CycleGAN augmentation, which improves class balance. Compared to the baseline YOLOv8m model, the proposed YOLOv8m with CBAM demonstrates improved classification consistency and reduced false positives.

The precision-recall (PR) curve further validates the robustness of the proposed model. The curve maintains high precision across a wide range of recall values, indicating that the model successfully detects diseased fish while minimizing false positives. The F1-score curve shows a strong peak, demonstrating an optimal balance between precision and recall, which is critical for real-time detection systems.

Qualitative evaluation of the detection results reveals that the model accurately localizes milkfish and identifies disease-related features such as lesions and discolorations under varying underwater conditions. The model performs well even in the presence of turbidity, lighting variation, and fish movement. However, minor limitations were observed in highly blurred frames and low-contrast conditions, where disease features are less distinguishable.

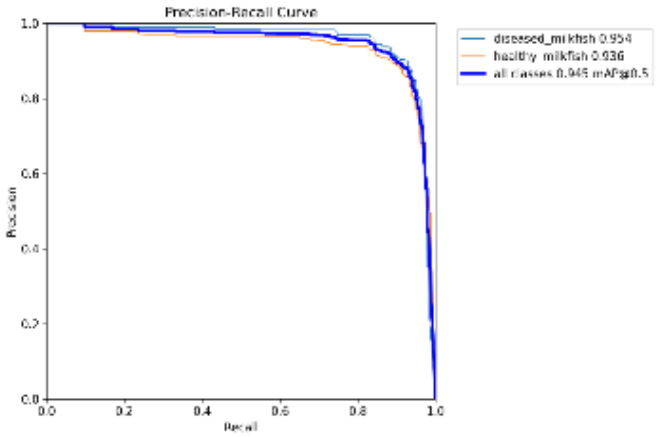

Figure 7. The Precision-recall curve of the proposed model demonstrating high detection performance across varying thresholds.

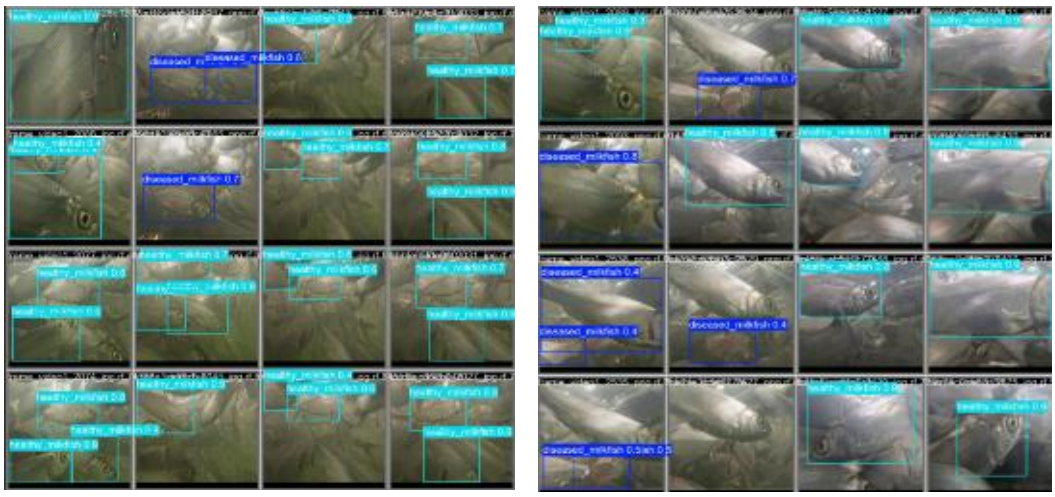

Figure 8. Sample detection outputs of the YOLOv8m-CBAM model showing accurate detection of healthy and diseased milkfish under varying underwater conditions.

Overall, the results confirm that the proposed YOLOv8m-CBAM model, combined with CycleGAN-based augmentation, achieves robust and reliable detection performance in real-world underwater environments, outperforming baseline configurations.

CONCLUSION

This study explores a YOLOv8m-based framework integrated with a Convolutional Block Attention Mechanism (CBAM) for real-time detection of milkfish diseases. The hybrid model was trained with an augmented milkfish dataset comprising traditional and GAN-augmented datasets. This approach addresses key challenges in aquaculture environments, including limited diseased milkfish samples, class imbalance, and other complexities of underwater imaging, such as water turbidity, motion, and varied lighting conditions.

The results directly support the effectiveness of combining GAN-based synthetic data augmentation and CBAM-based feature refinement on the milkfish disease detection task. The baseline YOLOv8m already shows a promising performance, achieving a precision of 0.817, a recall of 0.906, a mAP@0.50 of 0.925, and an mAP@0.50–0.95 of 0.682. However, its low precision indicates a high rate of false positives. On the other hand, the ResNet modification improved precision in some runs, but did not consistently improve the overall model performance, which suggests that deeper feature extraction alone was not sufficient for detecting subtle indicators of milkfish diseases, especially in the underwater environment.

CycleGAN-based data augmentation improved the diversity and reduced class imbalance, which was reflected in T7 having an improved precision of 0.910, recall of 0.865, mAP@0.50 of 0.936, and mAP@0.50–0.95 of 0.710. The performance drop in T6 indicates that excessive augmentation may introduce redundant or unrealistic patterns that reduce model generalization.

The integration of the attention mechanism, CBAM, into the YOLOv8m architecture achieved the best overall performance, with a precision of 0.940, a recall of 0.910, mAP@50 of 0.945, and mAP@50–95 of 0.725. These results confirm that enhancing the YOLOv8m model with an attention-based mechanism effectively improves feature discrimination by enabling the model to focus on relevant diseased regions, which is particularly challenging in underwater environments.

The findings of this study demonstrate that combining model enhancement and balancing data through synthetic augmentation provides a robust and efficient solution for milkfish disease detection, especially in an underwater environment. The proposed YOLOv8m-CBAM model improves detection performance while maintaining real-time capability, making it suitable for practical deployment in underwater monitoring systems, such as those in fish pens. However, future work may include the exploration of segmentation-based approaches, multi-species disease detection, and deployment on edge devices to further enhance real-time monitoring capabilities.

Ethical Considerations

This study used video recordings of milkfish from a selected fish pen in a non-intrusive manner. No human participants were involved. The dataset was used only for research purposes, and the model is intended only as an early detection support tool rather than a conclusive diagnostic system.

Conflict of Interest

The authors declare no conflict of interest.

Dataset Availability

The dataset used in this study was collected from a selected fish pen and is not publicly available due to ownership and privacy considerations. The dataset contains research-specific underwater milkfish image data. However, selected sample images, processed outputs, or related information may be made available from the corresponding author upon reasonable request and subject to research-use limitations.

REFERENCES

1. Alfred, R., Leo, J., & Kaijage, S. (2025). Optimizing dataset diversity for a robust deep-learning model in rice blast disease identification to enhance crop health assessment across diverse conditions. Smart Agricultural Technology, 10. doi:https://doi.org/10.1016/j.atech.2024.100726.
2. Bharti, A. (2024). Object Detection Using YOLO Algorithm. International Research Journal of Modernization in Engineering Technology and Science, 06(08). doi:DOI: https://www.doi.org/10.56726/IRJMETS60812
3. Biswal, S., Kumar, A., & Punuri, S. (2024). 2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS). 2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS). Kalaburagi, India: IEEE. doi:10.1109/ICIICS63763.2024.10860112
4. Chang, C.-C., Ubina, N. A., Cheng, S.-C., Lan, H.-Y., Chen, K.-C., & Huang, C.-C. (2022, October 7). A Two-Mode Underwater Smart Sensor Object for Precision Aquaculture Based on AIoT Technology. Development and Application of Methods and Algorithms to Smart Objects and Smart Sensors, 22(19)(7603). doi:https://doi.org/10.3390/s22197603
5. Chanos chanos, Milkfish : fisheries, aquaculture, gamefish, bait. (n.d.). Retrieved from https://www.fishbase.se/search.php: https://www.fishbase.se/summary/chanos-chanos.html
6. Chen, Y., Pan, J., & Wu, Q. (2023). Apple leaf disease identification via improved CycleGAN and Convolutional Neural Network. Soft Comput, 27, 9773–9786. doi:https://doi.org/10.1007/s00500-023-07811-y
7. Deepika, H., Raghuveer shetty, V., & G N, S. (2020, June). An Overview of You Only Look Once: Unified,Real-Time Object Detection . International Journal for Research in Applied Science & Engineering Technology (IJRASET), 8(VI). doi:https://doi.org/10.22214/ijraset.2020.6098
8. Deng, Q., Ryu, J., & Lee, D. (2025). A Novel Rare Image Data Augmentation Method for Fish Diseases Classification. Journal of the Korean Society of Multimedia, 28(2), 137 - 148 (12 pages). doi:10.9717/kmms.2025.28.2.137
9. Feng, J., & Jin, T. (2024). CEH-YOLO: A composite enhanced YOLO-based model for underwater object detection. Ecological Informatics, 82. doi:https://doi.org/10.1016/j.ecoinf.2024.102758
10. Fung, B. S., Chan, W., Lo, I. M., & Tsang, D. H. (2023). Freshwater Microscopic Algae Detection Based on Deep Neural Network with GAN-Based Augmentation for Imbalanced Algal Data. ACS EST Water 2024 (pp. 982–990). American Chemical Society. doi:https://doi.org/10.1021/acsestwater.3c00150
11. G. Mariani, F. Scheidegger, R. Istrate,, C. Bekas, & A. Malossi. (2018). BAGAN: Data Augmentation with Balancing GAN. ArXiv. doi:abs/1803.09655
12. Garcia, R., Alvarez, J., & Marcia, R. (2023). Machine Learning for Classifying Images with Motion Blur. 2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE. doi:10.1109/ICMLA55696.2022.00079
13. Gaude, G. S., & Borkar, S. (2020). Fish Detection And Tracking For Turbid Underwater Video. Madurai, India: IEEE. doi:10.1109/ICCS45141.2019.9065425
14. Gong, C., Wang, D., Li, M., Chandra, V., & Liu, Q. (2021). KeepAugment: A Simple Information-Preserving Data Augmentation Approach. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. doi:10.1109/CVPR46437.2021.00111
15. Guanbo, F., Xiong, Z., Pang, H., Gao, Y., Zhang, Z., Yang, J., & Ma, Z. (2024). RTL-YOLOv8n: A Lightweight Model for Efficient and Accurate Underwater Target Detection. Fishes. 2024, 8. doi: https://doi.org/10.3390/fishes908029
16. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778. doi:DOI:10.1109/cvpr.2016.90
17. Khabusi, S., Huang, Y.-P., & Lee, M.-F. (2023). Attention-Based Mechanism for Fish Disease Classification in Aquaculture. 2023 International Conference on System Science and Engineering (ICSSE). Ho Chi Minh, Vietnam: IEEE. doi:10.1109/ICSSE58758.2023.10227224
18. Komori , O., & Eguchi , S. (2019, July 3). Introduction to Imbalanced Data. In: Statistical Methods for Imbalanced Data in Ecological and Biological Studies. Statistical Methods for Imbalanced Data in Ecological and Biological Studies . doi:https://doi.org/10.1007/978-4-431-55570-4_1

19. Li, S., & Peng, X. (2025). DyAqua-YOLO: a high-precision real-time underwater object detection model based on dynamic adaptive architecture. (Y. Ju, Ed.) Front. Mar. Sci., 12:1678417. doi:doi: 10.3389/fmars.2025.1678417
20. Liu, X., & Goh, K. (2025). ResNet: Enabling Deep Convolutional Neural Networks through Residual Learning. arXiv.org. doi:DOI:10.48550/arXiv.2510.24036
21. Martínez, F. S., Tseng, M., & Yeh, S.-P. (2006, September 1). Milkfish (Chanos chanos) Culture: Situations and Trends. Environmental Science, Agricultural and Food Sciences, Biology. doi:DOI:10.29822/JFST.200609.0004
22. Minouei, M., Soheili, M., & Stricker, D. (2024). Multimodal approach for imbalanced document classification. Proceedings Volume 13517, Seventeenth International Conference on Machine Vision (ICMV 2024). Edinburg, United Kingdom: Proc. SPIE 13517. doi:https://doi.org/10.1117/12.3055119
23. Prova, N. (2024, December 3). Enhancing Fish Disease Classification in Bangladeshi Aquaculture through Transfer Learning, and LIME Interpretability Techniques. 2024 4th International Conference on Sustainable Expert Systems (ICSES). doi:10.1109/ICSES63445.2024.10763350
24. Woo, S., Park, J., Lee, J.-Y., & Kweon, I. (2018). CBAM: Convolutional Block Attention Module. Computer Vision – ECCV 2018, 11211. doi:https://doi.org/10.1007/978-3-030-01234-2_1
25. Xi, Y., & Yin, J. (2025). UOD-YOLO: a lightweight real-time model for detecting marine organisms. Front. Mar. Sci., 12. doi:https://doi.org/10.3389/fmars.2025.1728563
26. Yang, H., & Zhou, Y. (2021). IDA-GAN: A Novel Imbalanced Data Augmentation GAN. 2020 25th International Conference on Pattern Recognition (ICPR). Milan, Italy: IEEE. doi:10.1109/ICPR48806.2021.9411996
27. Yuwana, R., Fauziah, F., Heryana, A., Krisnandi, D., Kusumo, R., & Pardede, H. (2020, August 31). Data Augmentation using Adversarial Networks for Tea Disesse Detection. Jurnal Elektronika dan Telekomunikasi, 20. doi:http://dx.doi.org/10.14203/jet.v20.29-35
28. Zhang, Z., Li, Y., & Liu, C. (2022). Collaborative Discrimination-Enabled Generative Adversarial Network (CoD-GAN) for the Data Augmentation in Imbalanced Classification. 2022 IEEE 18th International Conference on Automation Science and Engineering (CASE). Mexico City, Mexico: IEEE. doi:10.1109/CASE49997.2022.9926707
29. Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE. doi:10.1109/ICCV.2017.244