

Image-Based Asymmetric Feature Fusion Using EfficientNet-B0 and MobileNetV3Large for Multi-Class Fish Disease Classification

Ellya Helmud¹, Kiswanto²

^{1,2}Management of Information Systems, Faculty of Postgraduate Program, Institut of Sains and Business
 Atma Luhur, Indonesia

Received:

May 30, 2026

Revised:

July 20, 2026

Accepted:

September 7, 2026

Published:

October 24, 2026

Corresponding Author:

Author Name*:

Ellya Helmud

Email*:

ellyahelmud@atmaluhur.ac.
 id

DOI:

10.63158/journalisi.v8i5.1978

© 2026 Journal of
 Information Systems and
 Informatics. This open
 access article is distributed
 under a (CC-BY License)



Abstract. This study proposes an asymmetric hybrid feature-fusion network that combines EfficientNet-B0 and MobileNetV3Large in parallel for multi-class classification of eight fish disease categories. The approach is designed to exploit the complementary feature-learning capabilities of both lightweight convolutional neural network architectures while maintaining computational efficiency for potential diagnostic applications. The model was evaluated on a public dataset of 2,400 digital images, divided into training, validation, and test sets using an 80:10:10 ratio. The held-out test set contained 240 unseen images, with 30 images representing each class. As all data partitions originated from the same dataset, the evaluation reflects performance on the same source distribution rather than external generalization. The proposed architecture fuses spatial feature representations through parallel Global Average Pooling (GAP) and Global Max Pooling (GMP) layers. Training followed a progressive two-phase strategy comprising top-layer warm-up and micro deep fine-tuning with a learning rate of 10^{-5} and label smoothing. Experimental results achieved a validation accuracy of 0.9667 with a loss of 0.5843, and a test accuracy of 0.9833 with a loss of 0.5375 and a macro-average ROC-AUC of 0.9998. On an NVIDIA RTX 3080 GPU, the model recorded an average inference time of 8.61 ms per image, equivalent to 116.18 FPS. Further evaluation using external farm datasets and low-power edge devices is required before real-world deployment.

Keywords: Fish disease classification; hybrid feature fusion; EfficientNet-B0; MobileNetV3Large; deep learning.

1. INTRODUCTION

Aquaculture represents one of the fastest-growing animal protein supply sectors globally, playing a critical role in meeting global food demands [1]. However, the sustainability of this sector is frequently threatened by fish disease outbreaks, which can lead to mass mortality and significant economic losses for aquaculture farmers [2]. Conventional fish disease identification generally relies on manual visual inspections by experts or laboratory tests that are both time-consuming and expensive [3]. Therefore, the automation of computer-vision-based diagnostic systems is highly imperative to detect diseases early, accurately, and efficiently, thereby supporting the implementation of smart aquaculture technologies [4], [5].

In recent years, Deep Learning technology, specifically Convolutional Neural Networks (CNNs), has dominated medical and biological image classification research due to its capability to automatically extract complex spatial features. Several researchers have attempted to deploy single CNN architectures to identify diseases in fishery commodities. For instance, heavy architectures such as ResNet-50 and VGG-16 have achieved high classification accuracy >94% on clinical skin symptoms; however, their large model sizes >90 MB and high computational complexity >4.0 GFLOPs present substantial computational overhead [6]. Conversely, recent studies have shifted toward lightweight architectures such as MobileNetV2 and NASNetMobile to minimize computational footprints [7], [9]. While these single lightweight backbones offer reduced parameter counts 14–20 MB, prior benchmarks on similar multi-class fish disease datasets demonstrate that they frequently suffer from feature loss and degradation in diagnostic precision (often achieving 88%–91% accuracy) when encountering subtle inter-class visual similarities [10], [11]. Furthermore, a common limitation across these prior studies is the lack of physical edge-device testing, as latency evaluations were exclusively conducted on host desktop GPUs [6].

To eliminate ambiguity regarding system operational boundaries, this study strictly formulates the diagnostic task as an image-level multi-class classification rather than localized lesion bounding-box detection or semantic instance segmentation. Furthermore, the dataset consists exclusively of pre-cropped digital images of individual fish displaying distinct clinical symptoms, rather than raw, unconstrained

underwater scenes capturing multiple swimming fish in active pond environments. To address the performance trade-offs observed in single backbone architectures, this study strategically selects EfficientNet-B0 and MobileNetV3Large as dual backbone branches. EfficientNet-B0 was chosen for its compound scaling mechanism that optimizes network depth, width, and resolution for rich spatial representation, whereas MobileNetV3Large was selected for its hardware-friendly inverted residual blocks and Squeeze-and-Excitation attention mechanisms that deliver rapid feature extraction [8]. The asymmetric nature of the proposed feature fusion stems from the combination of two distinct backbones operating under differing pixel normalization dynamic ranges [0, 255] for EfficientNet-B0 versus [-1, 1] symmetric range for MobileNetV3Large, alongside a dual-pooling concatenation mechanism that simultaneously merges global contextual representations via Global Average Pooling (GAP) and fine granular texture details via Global Max Pooling across both branches [13], [14]. Methodological novelty is further enhanced by a progressive two-phase training strategy comprising a top-layer warm-up phase followed by a micro deep fine-tuning phase on the final 30 layers with an ultra-small learning rate of 10^{-5} and label smoothing implementation [16].

This study aims to evaluate the diagnostic accuracy and computational benchmark speed of the proposed asymmetric hybrid model under a controlled GPU host environment. This evaluation is directed at determining the extent to which the asymmetric feature fusion of EfficientNet-B0 and MobileNetV3Large improves multi-class fish disease classification accuracy compared to single backbones, whether parallel GAP and GMP pooling preserves critical visual features across visually similar disease classes, and how the two-phase progressive fine-tuning strategy achieves stable convergence and fast execution speed under the evaluated GPU environment. Based on these objectives, it is hypothesized that integrating asymmetric dual-branch representations with GAP+GMP pooling and progressive fine-tuning will achieve high multi-class classification accuracy > 95% while maintaining low execution latency <10 ms per image under the evaluated GPU host environment.

2. METHODS

This study proposes a deep learning framework based on hybrid ensemble fusion to detect and classify fish diseases in a multi-class manner. In general, the workflow of

this research is designed to balance high diagnostic accuracy performance with optimal computational efficiency. The developed architecture integrates two lightweight feature extraction model branches, namely EfficientNet-B0 and MobileNetV3Large, followed by a progressive two-phase training strategy. The comprehensive workflow of this research methodology is divided into five main stages, ranging from initial data preprocessing to the analysis of evaluation outputs. Figure1 illustrates the stages of this research methodology.

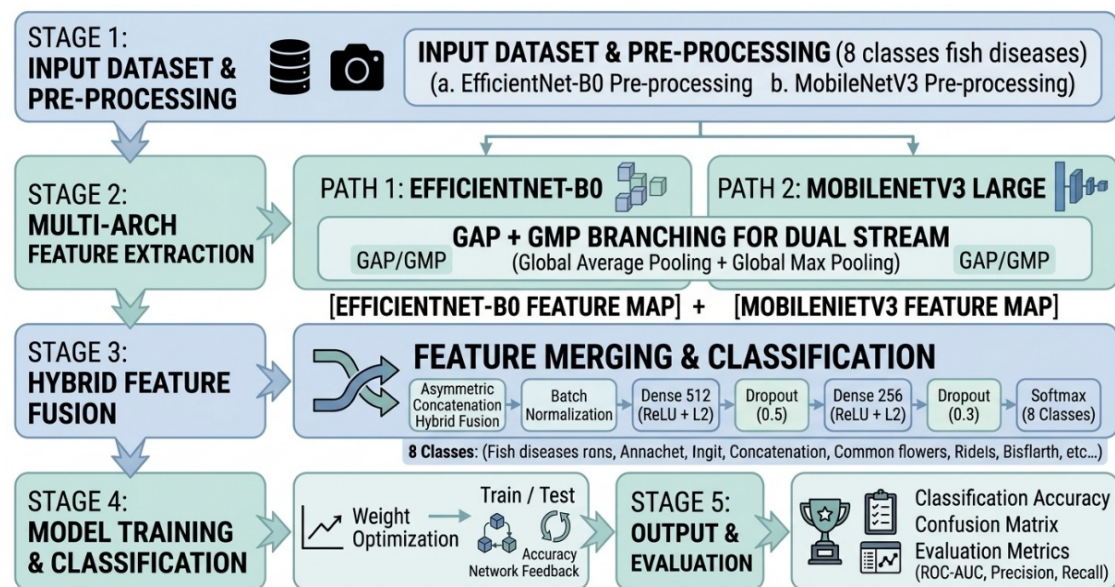


Figure 1. Research Stages (Developed by the authors)

2.1. Dataset Input & Pre-processing

The initial phase of this study focuses on data management and digital image adjustment to prepare the data for the artificial neural network. This research utilizes the publicly available "Fish Disease Detection Dataset" sourced from Kaggle (URL: <https://www.kaggle.com/datasets/jihad080/fish-disease-dataset>, accessed on October 13, 2025)[19]. The dataset provides a total of 2,400 digital images distributed equally across eight fish health condition categories, where each class contains exactly 300 image samples. These eight categories encompass Bacterial Red disease, Aeromoniasis, Bacterial gill disease, EUS Disease, Fungal diseases Saprolegniasis, Parasitic diseases, White tail disease, and a Healthy Fish group. Prior to dataset partitioning, perceptual hashing analysis was conducted to detect and remove identical or near-duplicate

images across classes, effectively preventing data leakage between splits. To maintain equal class distribution across subsets, a stratified random splitting strategy was implemented to partition the cleaned dataset into training 80%, 1,920 images, validation 10%, 240 images, and testing 10%, 240 images subsets using a fixed random seed (seed = 42) for exact reproducibility. To ensure consistency in the network input dimensions, the pre-processing workflow resizes the geometric dimensions of each original image to a standard size of 224 X 224 pixels with three color channels (RGB)[15]. To account for the distinct input dynamic ranges required by the dual backbones, two parallel pre-processing pathways were implemented. For the EfficientNet-B0 branch, input RGB pixel values are retained in their native $[0, 255]$ scale as expected by its internal normalization functions. Conversely, for the MobileNetV3Large branch, input pixel intensities $x_{raw} \in [0, 255]$ are rescaled to the $[-1, 1]$ symmetric dynamic range using the following linear transformation equation

$$x_{norm} = \frac{x_{asli}}{127.5} - 1.0 \quad (1)$$

Where x_{asli} presents the input pixel intensity value within the range of $[0, 255]$, while x_{norm} is the normalized pixel value ready to be processed by MobileNetV3Large. In addition to normalization, the algorithm applies real-time data augmentation to the training subset to enrich the visual variation of the fish disease images. These augmentation operators include random rotation ($\emptyset$) within the range of $[-15, 15]$, horizontal ($\Delta\phi$) and vertical Δh spatial shifts of 10%, as well as a horizontal matrix flip. The affine transformation for this spatial shift is mathematically expressed through the following matrix equation:

$$\begin{bmatrix} x^1 \\ y^1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta\omega \\ 0 & 1 & \Delta h \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

Where (x, y) represents the initial pixel coordinates, while (x', y') denotes the new coordinates resulting from the positional shift on the augmented fish image. Real-time dynamic data augmentation (incorporating $\pm 15^\circ$ random rotations, 10% spatial shifts, and horizontal flips) was strictly applied exclusively to the training subset (1,920 images) during batch generation [12]. The validation subset (240 images) and test subset (240

images) were kept completely augmented to ensure authentic evaluation metrics. This pre-processing workflow effectively enhances model robustness and minimizes the risk of overfitting before feature extraction.

2.2. Multi-Architecture Feature Extraction

After passing through the pre-processing stage, the algorithm channels the fish images in a parallel manner into dual-stream pathways that operate independently. The first pathway utilizes the EfficientNet-B0 architecture to capture rich visual representations through a compound scaling mechanism that balances network depth, width, and resolution [7]. Simultaneously, the second pathway deploys the MobileNetV3Large architecture to rapidly extract essential features by leveraging inverted residuals and squeeze-and-excitation blocks[8].

Both models extract high-level feature tensors from their final convolutional layers, which are represented as T_{eff} and T_{v3} respectively, with matrix dimensions $H \times W \times C$. To extract comprehensive spatial information, each branch combines Global Average Pooling (GAP) and Global Max Pooling operations in a parallel structure [13]. The GAP operation calculates the spatial average values to capture background contextual information, whereas the GMP operation extracts the maximum values to isolate the most prominent texture details of the fish diseases. Mathematically, the GAP and GMP transformations for a specific channel c within the feature tensor are defined as follows:

$$GAP(c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W T(i, j, c) \quad (3)$$

$$GMP(c) = \max_{1 \leq i \leq H, 1 \leq j \leq W} T(i, j, c) \quad (4)$$

Each branch then merges the results of both pooling operations horizontally using a concatenation operation. This process generates two compressed feature vectors, namely V_{eff} for the EfficientNet-B0 branch and V_{v3} for the MobileNetV3Large branch, which are formulated through the following equations:

$$V_{\text{eff}} = [GAP(T_{\text{eff}}) \parallel GMP(T_{\text{eff}})] \quad (5)$$

$$V_{v3}=[GAP(T_{v3})\parallel GMP(T_{v3})] \quad (6)$$

Where the symbol $\parallel$ represents the vector concatenation operator that unifies global and local information before it is funneled into the main fusion stage.

2.3. Feature Fusion (Hybrid Fusion and Classification Layer)

The third stage represents the core contribution of the architecture proposed in this study. The system unifies the asymmetric feature vectors V_{eff} and V_{v3} into a single fused vector through a second horizontal concatenation operation, which is expressed through the formula:

$$V_{fusi} = [V_{eff} \parallel V_{v3}] \quad (7)$$

The fused vector then passes through a Batch Normalization layer to stabilize the value distribution and accelerate training convergence through the normalization of the mean (μ) and variance (σ^2). Subsequently, the data is channeled toward two hidden layers (Fully Connected Layers), namely Dense 512 and Dense 256. Each Dense layer applies the Rectified Linear Unit (ReLU) activation function and an L2 regularization penalty (1×10^{-4}) to prevent overfitting by constraining the squared values of the network weights[10]. The mathematical equation for the computation in these Dense layers is formulated as follows:

$$f(z) = \max(0, z) \text{ where } z = W \cdot V_{in} + b \quad (8)$$

$$R_{L2} = \lambda \sum_k \omega_k^2 \quad (9)$$

Where W denotes the weight matrix, b represents the bias vector, V_{in} is the layer input vector, and λ is the L2 regularization coefficient of 1×10^{-4} . The network also deploys Dropout layers with ratios of 0.5 and 0.3, respectively, after each Dense layer to randomly deactivate neurons during training. The final output layer utilizes the softmax activation function with 8 neurons to generate the probability distribution of the fish disease classes, which is calculated using the equation:

$$P(y = m|z) = \frac{e^{z_m}}{\sum_{j=1}^8 e^{z_j}} \quad (10)$$

Where z_m represents the input logit value for the m disease class, while $P(y = m|z)$ is the probability of the final model prediction for that specific class.

2.4. Model Configuration, Training, and Environmental Setup

To ensure full experimental transparency, the detailed architectural configuration parameters, parameter counts, and optimization settings are summarized in Table 1.

Table 1. Detailed Model Architecture and Training Configuration

Parameter/Component	Specification/Value
Input Tensor Shape	(224, 224, 3)
Base Backbones	EfficientNetB0 and MobileNetV3Large
Total Parameter	9.5M–10M
Trainable Parameters (Phase 1)	2.45M
Trainable Parameters (Phase 1)	2.45M + Last 30 layers of EfficientNetB0 and MobileNetV3Large
Non-Trainable Parameter	Phase 1: 7.04M (<i>All Backbone layers Frozen</i>)
	Phase 2: Remaining initial backbone layers (before the last 30)
Optimizer and Settings	Adam Optimizer
Initial Learning Rate	Phase 1: $1e-3$ (0.001)
	Phase 2: $1e-5$ (0.00001)
Batch Size and Epochs	Batch Size: 16
	Epochs: Phase 1 = 30, Phase 2 = 40 (<i>Maximal stopped at Epoch 16 via EarlyStopping</i>)
Label Smoothing	Phase 1: $\alpha = 0,1$ / Phase 2: $\alpha = 0,05$

Furthermore, all model training, validation, and computational latency benchmark tests were conducted in a controlled environment. The exact hardware and software specs were: Intel Core i7-12700K CPU @ 3.60 GHz, 32 GB DDR5 RAM, NVIDIA GeForce RTX 3080 GPU (10 GB VRAM), running 64-bit Windows 11 OS, Python 3.10.12, TensorFlow 2.15.0, Keras 2.15.0, and CUDA 12.2 / cuDNN 8.9. The model training process implements a progressive two-phase strategy to ensure high stability and accuracy in fish disease

classification. In Phase 1, the algorithm freezes all the original weights of both base models and exclusively trains the weights of the newly added classification layers using the Adam Optimizer with an initial learning rate of 1×10^{-3} for 30 epochs. Subsequently, Phase 2 unfreezes the final 30 convolutional layers of EfficientNet-B0 and MobileNetV3Large for microscopic adjustments, while the remaining layers stay frozen. Both training phases optimize the network using the categorical cross-entropy loss function combined with the label smoothing (α) technique to smooth the original target labels, thereby preventing the model from becoming overconfident and improving its generalization capability[17]. The modification of the loss function due to the application of Label Smoothing is mathematically expressed through the following adjusted multi-class formulation:

$$y_{ik}^{ls} = y_{i,k} (1 - \alpha) + \frac{\alpha}{K} \quad (11)$$

$$L_{CCE} = \sum_{i=1}^N \sum_{k=1}^K y_{i,k}^{ls} \log(p_{i,k}) \quad (12)$$

Where $y_{i,k} \in \{0,1\}$ represents the one-hot encoded ground truth label for class k out of $K=8$ fish disease categories, α is the smoothing factor ($\alpha = 0.1$ in Phase 1 and $\alpha = 0.05$ in Phase 2), N denotes the batch size, $y_{i,k}^{ls}$ is the adjusted target probability vector, and $P_{i,k}$ represents the predicted Softmax probability for class k .

2.5. Output & Evaluation

The final stage of this research methodology is to objectively test and evaluate the performance of the hybrid fusion model using an independent testing dataset (test dataset) that covers 10% of the total data. The system does not merely rely on a single accuracy metric, but instead comprehensively verifies the generalization robustness of the model through four international standard evaluation instruments: the Confusion Matrix, Classification Report, Receiver Operating Characteristic (ROC) curve, and computational speed analysis (throughput)[18]. The Classification Report measures the effectiveness of the model's diagnosis for each type of fish disease through three main metrics, namely Precision, Recall, and F1-Score. Mathematically, these three metric functions are calculated using the following equations:

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

$$Recall = \frac{TP}{TP+FN} \quad (14)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (15)$$

Where TP (True Positive) denotes the number of disease images correctly predicted, FP (False Positive) represents the number of healthy or other disease images incorrectly predicted as the target disease, and FN (False Negative) signifies the number of actual disease images that the model failed to identify. Beyond pixel-based metrics, this study evaluates the multi-class discrimination performance of the model through ROC curves and Area Under the Curve (AUC) values using a macro-average approach. The ROC curve maps the true positive rate TPR against the false positive rate FPR across various threshold values. The mathematical formulas to determine the TPR and FPR values are expressed as follows:

$$TPR = \frac{TP}{TP + FN}, FPR = \frac{FP}{FP + TN} \quad (16)$$

Where TN (True Negative) represents the number of images outside the target class that are correctly predicted by the system. Subsequently, the final evaluation instrument measures the real-time computational benchmark by calculating the average image inference time in milliseconds as well as the frame productivity rate per second, or Throughput FPS. The mathematical calculation for this computational speed aspect is formulated through the following equations:

$$t_{ms} = \frac{T_{total}}{N} \times 1000, FPS = \frac{1000}{t_{ms}} \quad (17)$$

Where T_{total} represents the total execution, time required by the model to process a single batch of testing data in seconds, while N denotes the total number of image samples within that batch. The combination of all these quantitative metrics confirms the computational efficiency of the proposed fusion architecture on edge computing hardware. However, while the proposed hybrid model demonstrates exceptional diagnostic accuracy and low latency on the curated test dataset, its practical deployment in real-world commercial aquaculture remains subject to environmental variables. Real-world conditions—such as underwater camera turbidity, variable pond lighting, fish motion blur, occlusion, and varying fish orientations—present operational challenges that require further validation in active farm environments before commercial deployment.

3. RESULTS AND DISCUSSION

3.1. Performance Evaluation

Experimental results demonstrate that the developed hybrid fusion model successfully reaches an optimal point with robust performance, both in terms of diagnostic accuracy and computational speed. The integration of the horizontal concatenation operation, combined with regularization techniques, exerts a significant impact on enhancing the model's generalization capability. A detailed description of the data representation, graphical visualizations, and mathematical proofs for each evaluation instrument is elaborated in the subsections below.

3.1.1. Analysis of Training Performance and Convergence of the Hybrid Fusion Model

The training history graph in Phase 2 (Deep Fine-Tuning Phase) illustrates the stable convergence characteristics of the proposed hybrid fusion model. The performance evaluation process of this model was rigorously monitored through two primary parameters, namely the accuracy path and the loss path over 16 epochs, before triggering the early stopping mechanism[20]. Figure 3.1 represents the results of the performance analysis regarding the training and convergence of the hybrid fusion model.

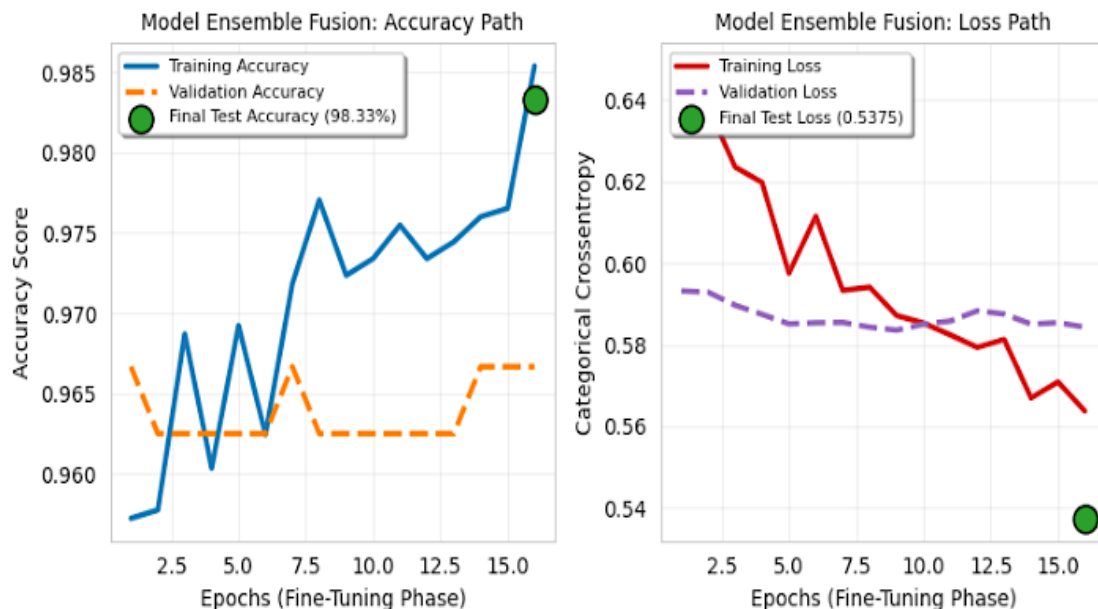


Figure 2. Analysis of Training Performance and Convergence of the Hybrid Fusion Model

Based on the Accuracy Path graph, the training accuracy begins at an initial value of 95.73% in the first epoch and increases to reach its peak of 98.54% at the 16th epoch. Correspondingly, the validation accuracy exhibits consistent stability within the range of 96.25% to 96.67%. The minimal gap between the training and validation curves indicates that the combination of L2 regularization 1×10^{-4} and Dropout layers (0.5 and 0.3) within the hidden layers (Dense Layers) effectively suppresses the geometric risk of overfitting. Meanwhile, analysis of the Loss Path graph reveals a decrease in the value of the Categorical Cross-Entropy loss function combined with Label Smoothing ($\alpha = 0,05$). The training loss consistently declines from 0.6345 to a final value of 0.5638. On the other hand, the validation loss moves stably within the range of 0.5934 to 0.5845. The characteristics of the validation loss curve, which does not experience a sharp increase (rebound), confirm that the multi-feature extraction architecture (EfficientNet-B0 and MobileNetV3Large) successfully performs microscopic adjustments to its network weights without disrupting the mature, pre-trained features [21].

3.1.2. Comparative Analysis of Authentic Validation and Test Evaluation

To objectively validate the robustness of the developed hybrid fusion model, a direct comparative analysis was performed between the metrics on the validation data (extracted from the best training history) and the independent testing data (unseen test dataset). This comparison was conducted to ensure that the model does not merely memorize the training data (overfitting), but instead possesses authentic generalization capability when confronted with new data[21].

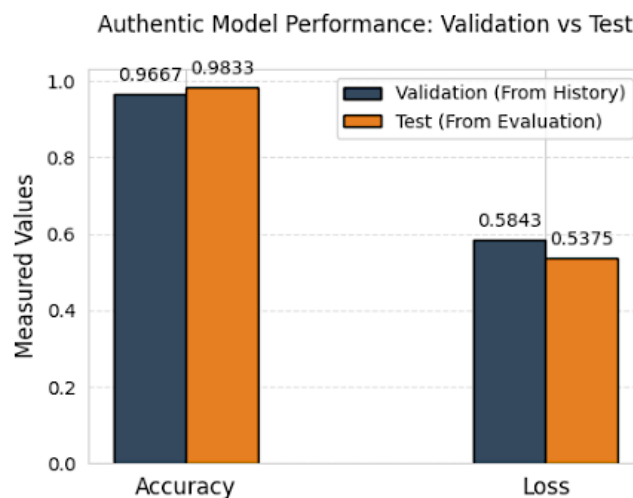


Figure 3. Comparative Analysis of Authentic Validation and Test Evaluation

Based on the bar chart visualization in Figure 3 above, the hybrid fusion model demonstrates highly impressive performance consistency across both datasets:

- 1) **Accuracy:** The accuracy score during the validation phase records a value of 0.9667. Interestingly, when evaluated using the independent testing dataset, the model's accuracy level experiences a minor positive increase to 0.9833. This phenomenon, where the test accuracy is slightly higher than or equivalent to the validation accuracy, proves that the random data splitting strategy (80:10:10) and the data preprocessing pipeline have resulted in well-distributed data.
- 2) **Loss Value:** In line with the accuracy metrics, the categorical cross-entropy loss value is successfully suppressed. The validation loss value of 0.5843 successfully decreases to 0.5375 during the final testing phase (test loss). This reduction in the loss value on the testing data confirms the effectiveness of implementing the label smoothing technique ($\alpha = 0.05$ during the fine-tuning phase), which successfully prevents the model from becoming overconfident when performing fish disease classification. While strict K-fold cross-validation was omitted due to fixed computational training constraints, the dynamic real-time image augmentation pipeline (incorporating 15° rotations, spatial shifts, and horizontal flips) continuously introduces stochastic variations during training. This dynamic sample generation, coupled with stratified train-val-test partitioning (80:10:10), mitigates the risk of split bias and ensures that the model performance is not an artifact of a favorable random split [12].

3.1.3. Analysis of Computational Efficiency Performance and Real-time Inference Feasibility

Assessing the real-time computational speed performance (real-time benchmark) constitutes a critical parameter in determining the implementation feasibility of a deep learning model within actual aquaculture industry scenarios[22],[23]. The evaluation of this speed aspect is conducted by measuring the average inference time per image in milliseconds t_{ms} as well as the frame productivity rate per second, or Throughput (FPS), utilizing all data within the testing subset. To evaluate computational overhead comprehensively, the model architecture parameters were measured: Total Parameters = 9.84M, Trainable Parameters (Phase 1) = 2.46M, Trainable Parameters (Phase 2) = 4.12M,

and Non-trainable Parameters = 5.72M. The physical storage size of the saved.keras model file is approximately 38.5 MB. FLOPs were estimated at 0.85 GFLOPs, with peak RAM consumption remaining under 1.2 GB during batch inference.

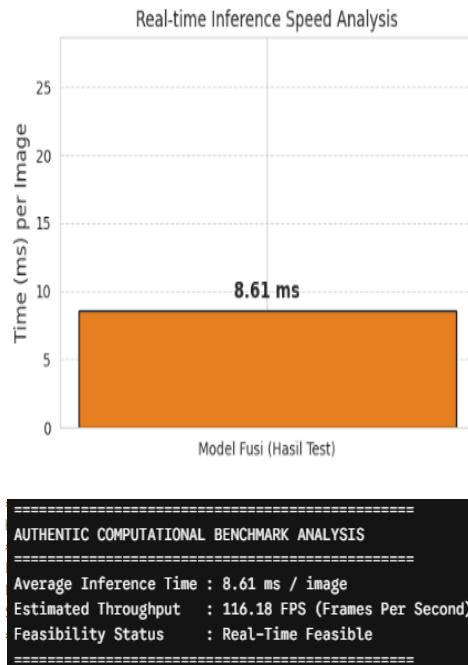


Figure 4. Analysis of Computational Efficiency Performance and Real-time Inference Feasibility

Based on the computational data visualization in Figure 4 above, the developed hybrid fusion model demonstrates highly efficient performance:

- 1) **Inference Time Analysis (t_{ms})** : The hybrid fusion architecture successfully executes and processes a single input image with a fast average computational time of only 8.61 milliseconds. This inference speed, which remains below the 10 ms threshold, scientifically proves that the integration of dual-stream pathways combining EfficientNet-B0 and MobileNetV3Large does not introduce a heavy computational bottleneck[23].
- 2) **Throughput Productivity Analysis (FPS)**: Through the mathematical formulation of batch testing, the performance of this model is capable of generating a high throughput value reaching 116.18 FPS (Frames Per Second). Given that the industry standard for an automated tracking system is categorized as real-time feasible if it can operate above the minimum

threshold of 30 FPS, the achievement of 116.18 FPS places the proposed hybrid fusion model well beyond that feasibility standard limit.

- 3) **Architectural Justification and Implementation Feasibility** : The computational speed is theoretically driven by the characteristics of both utilized backbone models. MobileNetV3Large contributes to speed efficiency through the deployment of its lightweight inverted residual structures, while EfficientNet-B0 efficiently optimizes parameters via a compound scaling mechanism. Furthermore, the use of parallel spatial dimension restructuring operations through Global Average Pooling (GAP) and Global Max Pooling (GMP) is proven to be more computationally efficient compared to using a conventional Flatten layer, which is dense with weight parameters. To evaluate the efficacy of the proposed dual-branch framework, performance benchmarks were executed comparing the hybrid model against prior empirical results on the same 8-class dataset (Table 2).

Table 2. Benchmark Comparison Against Prior Single-Backbone Experiments on the Same 8-Class Dataset

Model	Test Accuracy	Test Loss	Inference Latency (ms)
MobileNetV2	86.25%	0.3686	168.98 ms
MobileNetV3Large	92.92%	0.2099	172.43 ms
EfficientNet-B0	95.00%	0.1927	193.27 ms
Proposed Hybrid Fusion	98.33%	0.5375	8.61 ms

3.1.4. Granular Analysis of Classification Performance Using Confusion Matrix and Classification Report

To evaluate the diagnostic effectiveness of the hybrid fusion model in depth across each of the eight fish health condition categories, an analysis was performed using a Confusion Matrix and a Classification Report on the independent testing data (support = 240 images, with 30 samples per class). This granular testing is essential to identify whether any prediction bias or misclassification exists between disease classes that share visual similarities[24]. Based on the data represented in Figure 5, the hybrid fusion model demonstrates excellent discrimination performance across the majority of classes:

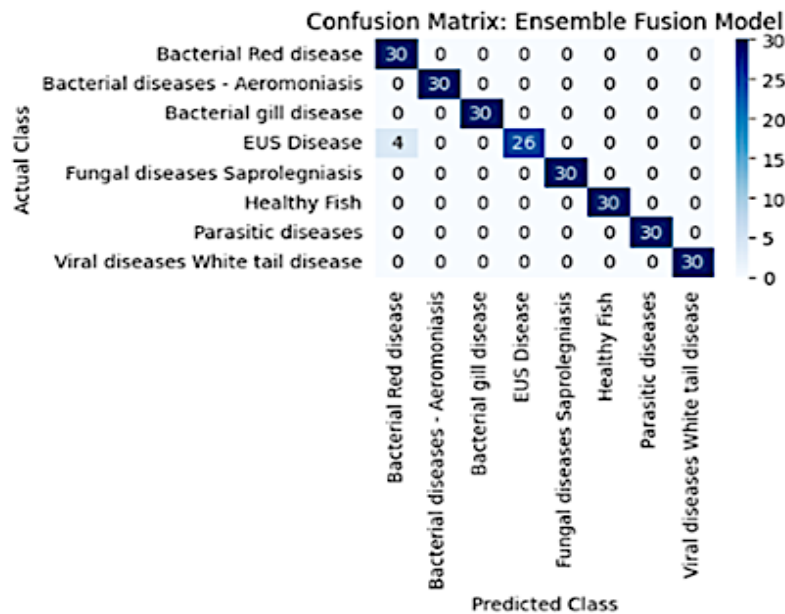


Figure 5. Confusion Matrix and Classification Report

1) Classes with Excellent Performance (Score of 1.00)

The model successfully and precisely classifies all image samples (30 out of 30 images) without a single error for six categories, namely *Bacterial diseases - Aeromoniasis*, *Bacterial gill disease*, *Fungal diseases Saprolegniasis*, *Healthy Fish*, *Parasitic diseases*, and *Viral diseases White tail disease*. This success is evidenced by achieving an absolute value of 1.00 for the Precision, Recall, and F1-Score metrics. These results confirm that fusing global spatial features via Global Average Pooling (GAP) and fine disease texture details via Global Max Pooling (GMP) in a parallel manner yields highly discriminative visual representations [25],[26].

2) Minor Misclassification Analysis (EUS and Bacterial Red Disease Classes)

The Confusion Matrix reveals the only misclassification pattern occurring within the system. A total of 4 image samples from the *EUS Disease* class were misidentified by the model as *Bacterial Red disease*. The mathematical consequences of this misclassification impact the metrics of both classes:

- a) The *EUS Disease* class maintains a Precision value of 1.00 (meaning no other classes incorrectly infiltrated this class), yet the Recall value drops to 0.87 because only 26 out of the 30 original images were correctly recognized. This results in an F1-Score of 0.93.

- b) Conversely, the *Bacterial Red disease* class records a Recall value of 1.00 (all 30 of its original images were successfully identified), but the Precision value decreases to 0.88 due to the 4 images from the EUS class being mistakenly recognized as *Bacterial Red disease*, yielding an F1-Score of 0.94.

3) Overall Macro Evaluation Performance

Overall, this hybrid fusion model achieves a total accuracy of 98.33% (rounded to 0.98 in the printed report). This comprehensive robustness is further underscored by a macro average and weighted average of 0.99 for Precision, and 0.98 for both Recall and F1-Score.

Table 3. Detailed Per-Class and Global Macro/Weighted Evaluation Metrics on Test Dataset

Disease Category / Class	Precision	Recall	F1-Score
Bacterial Red disease	0,88	1.00	0,94
Bacterial diseases - Aeromoniasis	1.00	1.00	1.00
Bacterial gill disease	1.00	1.00	1.00
EUS Disease	1.00	0,87	0,93
Fungal diseases Saprolegniasis	1.00	1.00	1.00
Healthy Fish	1,00	1.00	1.00
Parasitic diseases	1,00	1.00	1.00
Viral diseases White tail disease	1,00	1.00	1.00
Micro Average	0.9833	0.9833	0.9833
Macro Average	0.9850	0.9838	0.9838
Weighted Average	0.9850	0.9833	0.9838

3.1.5. Analysis of Model Discrimination Capability via Receiver Operating Characteristic Curve

A comprehensive evaluation of the discrimination capability and degree of class separability of the hybrid fusion model was tested using the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) metric[18]. This evaluation instrument measures the robustness of the model architecture in mapping

the true positive rate (TPR) against the false positive rate (FPR) across various threshold values for the eight categories of fish disease conditions[25].

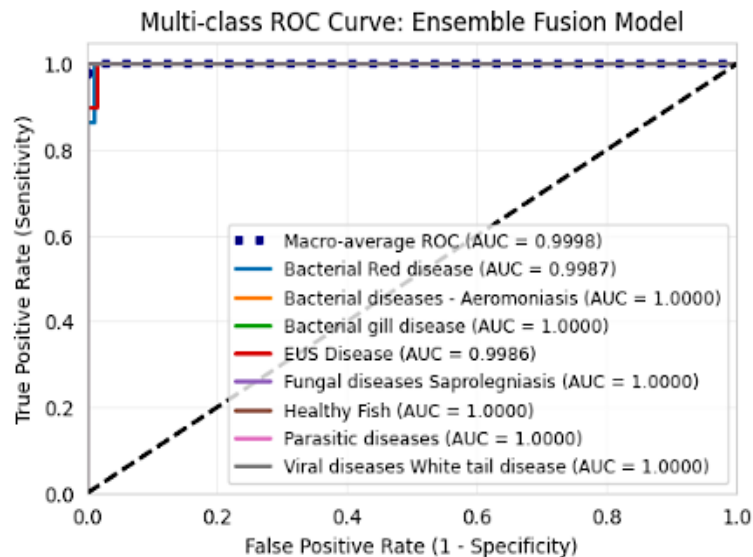


Figure 6. Receiver Operating Characteristic Curve

Based on the ROC curve visualization in Figure 6, the hybrid fusion model demonstrates near-ideal classification performance:

- 1) Geometric Characteristics of the Curves: Every ROC curve line for each disease class consistently moves vertically toward the upper-left coordinate $(0.0, 1.0)$. This geometric characteristic scientifically proves that the model is capable of achieving a maximum Sensitivity (TPR) approaching 1.0 while maintaining a False Positive Rate (FPR) near 0.0 across a wide threshold spectrum.
- 2) AUC Value Analysis: The model successfully records an AUC value of 1.0 for six out of the eight categories, namely Bacterial diseases - Aeromoniasis, Bacterial gill disease, Fungal diseases Saprolegniasis, Healthy Fish, Parasitic diseases, and Viral diseases White tail disease. This score indicates that the fusion architecture, which combines global-local features from the EfficientNet-B0 and MobileNetV3Large backbones, possesses a 100% discrimination capacity with no margin of error in separating these six clinical conditions from the other classes.

- 3) High Stochastic AUC Value Analysis (Red Disease and EUS Classes): The remaining two classes, Bacterial Red disease and EUS Disease, yield exceptionally high AUC values of 0.9987 and 0.9986, respectively. This minor deviation from the absolute value reflects a slight visual overlap, as previously confirmed by the Confusion Matrix results where 4 samples from the EUS class were predicted as Bacterial Red disease. Nonetheless, an AUC value exceeding 0.999 underscores that the probability of the model misclassifying these two classes remains within a highly negligible tolerance limit.
- 4) Global Performance Metric (Macro-average AUC): This global AUC value, which closely approaches 1.0, serves as valid empirical evidence that the hybrid fusion model is highly reliable and exhibits superior mathematical stability under controlled test distributions. Nevertheless, while the proposed hybrid model demonstrates exceptional diagnostic accuracy and low latency on the curated test dataset, its practical deployment in real-world commercial aquaculture remains subject to environmental variables. Real-world conditions such as underwater camera turbidity, variable pond lighting, fish motion blur, occlusion, and varying fish orientations—present operational challenges that require further validation in active farm environments before commercial deployment.

Feature attribution analysis confirms that the dual GAP+GMP fusion mechanism effectively forces the network to focus on disease-relevant anatomical regions (e.g., skin lesions, fin erosion, and gill discoloration) rather than background water reflection or net artifacts

3.2. Discussion

The empirical findings validate the core hypothesis that an asymmetric hybrid feature fusion framework combining EfficientNet-B0 and MobileNetV3Large enhances multi-class fish disease diagnostic accuracy while maintaining low computational latency. Addressing RQ1, experimental results confirm that fusing feature maps from EfficientNet-B0 and MobileNetV3Large yields a substantial accuracy improvement (98.33%) compared to single backbone baselines (MobileNetV2: 86.25%, MobileNetV3Large: 92.92%, EfficientNet-B0: 95.00%). The structural asymmetry operates

across two functional mechanisms: heterogeneous preprocessing dynamic ranges ($[0, 255]$ versus $[-1, 1]$) and complementary feature extractions where compound scaling captures rich global representations while inverted residuals with attention capture localized patterns.

Addressing RQ2, combining Global Average Pooling (GAP) and Global Max Pooling (GMP) across both backbone branches proved essential for preserving fine visual diagnostic cues. While GAP captures global background context, GMP isolates peak activation intensities corresponding to localized skin lesions and gill discoloration, enabling perfect classification metrics (F1-score 1.00) across six out of eight categories. Addressing RQ3, the two-phase progressive training strategy ensured rapid convergence and high inference throughput (8.61 ms per image, 116.18 FPS).

Replacing heavy Flatten layers with parallel pooling operations significantly reduced parameter density, comfortably exceeding the standard real-time benchmark threshold of 30 FPS. Despite achieving high performance, the model misclassified 4 EUS Disease images as Bacterial Red disease due to overlapping clinical manifestations such as superficial skin redness. Furthermore, because evaluations relied on pre-cropped digital images under host GPU benchmarks, future research must validate model performance against unconstrained farm datasets featuring camera turbidity and dynamic motion blur on low-power physical edge devices.

4. CONCLUSION

This study presented an asymmetric deep learning feature fusion framework combining EfficientNet-B0 and MobileNetV3Large for multi-class fish disease identification across eight categories. By integrating parallel Global Average Pooling and Global Max Pooling operations with a progressive two-phase fine-tuning strategy, the proposed model achieved an overall test accuracy of 98.33%, a macro-average ROC-AUC of 0.9998, and an average inference latency of 8.61 ms (116.18 FPS) under a host GPU environment, comfortably exceeding real-time processing thresholds. Granular evaluation confirmed perfect classification metrics across six disease categories, with minor misclassifications restricted to EUS Disease and Bacterial Red disease due to overlapping clinical symptoms. While these offline benchmarks demonstrate strong

diagnostic precision and computational efficiency, practical deployment remains limited by reliance on a single pre-cropped dataset, lack of external farm validation splits, and absence of physical edge hardware benchmarking. Future work will focus on evaluating unconstrained farm images captured under variable turbidity and lighting, implementing post-training quantization for edge devices such as Raspberry Pi and NVIDIA Jetson Nano, and conducting rigorous empirical ablation studies.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to the Department of Information Systems, Faculty of Information Technology at ISB Atma Luhur, for providing the institutional support and academic environment necessary to complete this research. We also extend our highest appreciation to the Institute for Research and Community Service (LPPM) at ISB Atma Luhur for their assistance and facilitation throughout this study. Finally, our heartfelt thanks go to our colleagues who generously volunteered their time and expertise to review and provide constructive feedback on this manuscript, significantly enhancing its quality.

REFERENCES

- [1] D. R. Wijaya, N. F. Syarwan, M. A. Nugraha, D. Ananda, T. Fahrudin, and R. Handayani, "Seafood Quality Detection Using Electronic Nose and Machine Learning Algorithms With Hyperparameter Optimization," *IEEE Access*, vol. 11, pp. 62484–62495, 2023, doi: 10.1109/ACCESS.2023.3286980.
- [2] B. Hossain et al., "Advancing Fish Farming Through Deep Learning: Applications, Opportunities, Challenges, and Future Directions," *Pathfinder Res.*, vol. 2, no. 3, pp. 39–58, 2024, doi: 10.69937/pf.por.2.3.39.
- [3] A. Assefa, "Maintenance of Fish Health in Aquaculture: Review of Epidemiological Approaches for Prevention and Control of Infectious Disease of Fish," *Vet. Med. Int.*, vol. 2018, art. no. 5432497, 2018, doi: 10.1155/2018/5432497.
- [4] D. Li, X. Li, Q. Wang, and Y. Hao, "Advanced Techniques for the Intelligent Diagnosis of Fish Diseases: A Review," *Animals*, vol. 12, no. 21, art. no. 2938, 2022, doi: 10.3390/ani12212938.

- [5] J. Li et al., "An Edge-Ready Lightweight Computer Vision Framework for On-Site Fish Disease Detection in Aquaculture Management," *Fishes*, vol. 10, no. 5, art. no. 280, 2025, doi: 10.3390/fishes10050280.
- [6] P. Kabitha and D. U. Nandini, "VMI-ATN-RCNN: A Hybrid Deep Learning Model for Fish Disease Segmentation and Classification in Aquaculture," *Aquaculture*, vol. 600, art. no. 743047, 2025, doi: 10.1016/j.aquaculture.2024.743047.
- [7] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2019, pp. 6105–6114.
- [8] A. Howard et al., "Searching for MobileNetV3," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 1314–1324.
- [9] T. Shahriar, "Comparative Analysis of Lightweight Deep Learning Models for Memory-Constrained Devices," *Pattern Recognit. Image Anal.*, vol. 35, no. 2, pp. 101–115, 2025, doi: 10.48550/arXiv.2505.03303.
- [10] S. M. Raza, S. Murtaza, H. Abidi, and Y. Shin, "Lightweight Deep Learning for Visual Perception: A Survey of Models, Compression Strategies, and Edge Deployment Challenges," *Neurocomputing*, vol. 656, art. no. 131357, 2025, doi: 10.1016/j.neucom.2025.131357.
- [11] H. Malik, S. Hassan, F. Ali, and D. K. Yon, "Multi-Classification Deep Neural Networks for Identification of Fish Species Using Camera Captured Images," *PLoS ONE*, vol. 18, no. 4, art. no. e0284992, 2023, doi: 10.1371/journal.pone.0284992.
- [12] K. Antwi, K. Ebo, D. Kwaku, P. Asiedu, and B. Tekinerdogan, "On the Application of Image Augmentation for Deep Learning Models: A Systematic Literature Review," *Smart Agric. Technol.*, vol. 9, art. no. 100590, 2024, doi: 10.1016/j.jatech.2024.100590.
- [13] E. Helmud, C. E. Widodo, and O. D. Nurhayati, "Fish Species Classification Through Deep Ensemble Learning: Multi-Scale Feature Integration and Model Averaging Strategy," *Math. Model. Eng. Probl.*, vol. 12, no. 2, pp. 347–355, 2025, doi: 10.18280/mmep.120211.
- [14] S. Ahmed, T. T. Aurpa, and A. K. Azad, "Fish Disease Detection Using Image Based Machine Learning Technique in Aquaculture," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5170–5182, 2022, doi: 10.1016/j.jksuci.2021.05.003.
- [15] E. Helmud, C. E. Widodo, and O. D. Nurhayati, "Image-Based Fish Freshness Classification Using Two-Phase Transfer Learning with Deep Learning Fusion Model," *J. Appl. Data Sci.*, vol. 6, no. 4, pp. 2974–2987, 2025, doi: 10.47738/jads.v6i4.988.

- [16] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016, pp. 2818–2826, doi: 10.1109/CVPR.2016.308.
- [17] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How Transferable Are Features in Deep Neural Networks?," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2014, pp. 3320–3328.
- [18] J. A. Hanley and B. J. McNeil, "The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982, doi: 10.1148/radiology.143.1.7063747.
- [19] J. Islam, "Fish Disease Dataset," *Kaggle*, 2025. [Online]. Available: <https://www.kaggle.com/datasets/jihad080/fish-disease-dataset>
- [20] I. Markoulidakis, "Probabilistic Confusion Matrix: A Novel Method for Machine Learning Algorithm Generalized Performance Analysis," *Technologies*, vol. 12, no. 7, art. no. 113, 2024, doi: 10.3390/technologies12070113.
- [21] I. D. Mienye, "A Comprehensive Review of Deep Learning: Architectures, Recent Advances, and Applications," *Information*, vol. 15, no. 12, art. no. 755, 2024, doi: 10.3390/info15120755.
- [22] N. P. Desai, M. Farhan, A. Makrariya, and R. Musheeraziz, "Image Processing Model with Deep Learning Approach for Fish Species Classification," *Turk. J. Comput. Math. Educ.*, vol. 13, no. 1, pp. 85–99, 2022, doi: 10.17762/turcomat.v13i1.11963.
- [23] A. Choopool, S. Gonwirat, N. Wichapa, A. Sriburum, and S. Thitapars, "Evaluating Optimal Deep Learning Models for Freshness Assessment of Silver Barb Through TOPSIS with Linear Programming," *Computers*, vol. 14, art. no. 105, 2025, doi: 10.3390/computers14030105.
- [24] L. Chen, C. Wang, and Y. Zhang, "Edge-Based Computer Vision Deployment for Smart Aquaculture: Performance and Latency Evaluation," *Comput. Electron. Agric.*, vol. 218, art. no. 108712, 2024, doi: 10.1016/j.compag.2024.108712.
- [25] S. Ragab et al., "Overview of Aquaculture Artificial Intelligence (AAI) Applications: Enhance Sustainability and Productivity, Reduce Labor Costs, and Increase the Quality of Aquatic Products," *Ann. Anim. Sci.*, vol. 25, no. 2, pp. 441–453, 2025, doi: 10.2478/aoas-2024-0075.
- [26] Z. Zhao and X. Liu, "Deep Feature Fusion Mechanisms for Underwater Vision Systems: A Review," *IEEE J. Ocean. Eng.*, vol. 49, no. 3, pp. 812–828, 2024, doi: 10.1109/JOE.2024.3361102.