

Image-Based Fish Freshness Classification Using Two-Phase Transfer Learning with Deep Learning Fusion Model

Ellya Helmud^{1,*}, Catur Edi Widodo², Oky Dwi Nurhayati³

¹*Doctor of Information System, Diponegoro University, Semarang, Indonesia*

¹*Faculty of Information Technology, Institut of Sciences and Business Atma Luhur, Pangkalpinang, Indonesia*

²*Department of Physics, Diponegoro University Semarang, Indonesia*

⁴*Department of Computer Engineering, Diponegoro University Semarang, Indonesia*

(Received: April 10, 2025; Revised: July 04, 2025; Accepted: October 11, 2025; Available online: October 22, 2025)

Abstract

This study introduces a novel deep learning approach for automated fish freshness classification using image analysis. The objective is to design and validate a Deep Learning Fusion Model that combines the strengths of EfficientNetB0 and InceptionV3 architectures to improve accuracy and robustness in classifying fresh and non-fresh fish. Input images were subjected to extensive augmentation, including RandomFlip, RandomRotation, RandomZoom, RandomContrast, RandomBrightness, and RandomTranslation, applied exclusively to the training dataset to enhance generalization, followed by backbone-specific pre-processing. Extracted features were fused via global average pooling and forwarded to a newly designed classification head with dropout and L2 regularization to mitigate overfitting. A two-phase transfer learning strategy was employed: initially training the classification head with frozen backbones, followed by fine-tuning the backbone layers using the Adam optimizer with a reduced learning rate. To highlight the contribution of the fusion strategy, ablation studies were conducted with single-backbone models. The EfficientNetB0 model achieved 89.17% validation accuracy, 85.83% test accuracy, and an F1-score of 85.69%, while the InceptionV3 model achieved 86.67% validation accuracy, 81.67% test accuracy, and an F1-score of 81.59%. In contrast, the proposed Fusion Model achieved 93.33% validation accuracy, 95.00% test accuracy, and an F1-score of 94.95%. Additional evaluations with confusion matrices, ROC curves, AUC, and precision-recall curves confirmed the model's superiority. The findings demonstrate that integrating features from diverse CNN architectures enables the model to learn richer representations, resulting in significantly improved classification performance. The novelty of this work lies in the effective fusion of complementary backbones through global average pooling and fine-tuned transfer learning, establishing a human-centric computational approach that offers a reliable solution for practical fish freshness assessment in food safety and market scenarios.

Keywords: Deep Learning, Convolutional Neural Networks, Image Classification, Transfer Learning, Model Fusion, EfficientNetB0, InceptionV3

1. Introduction

As one of the most vital sources of animal protein, fish plays a crucial role in the global diet and maritime economy. However, its highly perishable nature makes freshness a critical quality factor, determining food safety, nutritional value, and consumer acceptability [1]. The deterioration of fish quality begins immediately after capture, with progressive biochemical and microbiological changes [2]. This leads to substantial economic losses for the fishing industry and potential health risks if unfresh fish is consumed. Therefore, there's an urgent need for rapid, accurate, and objective methods for assessing fish freshness [3].

Fish freshness assessment heavily relies on organoleptic methods, which involve sensory evaluation using human senses like sight (color of eyes, gills, skin), smell (aroma), and touch (flesh texture). While these methods are easy and don't require special equipment in the field, they have fundamental limitations [4]. Their highly subjective nature depends on the individual assessor's experience and sensitivity, leading to inconsistency and variability in results [5]. Additionally, organoleptic methods can be impractical and time-consuming for large volumes of fish, and they are susceptible to assessor fatigue. More objective laboratory methods, such as chemical or microbiological analysis, often

*Corresponding author: Ellya Helmud (ellyahelmud@students.undip.ac.id)

DOI: <https://doi.org/10.47738/jads.v6i4.988>

This is an open access article under the CC-BY license (<https://creativecommons.org/licenses/by/4.0/>).

© Authors retain all copyrights

require long waiting times, expensive instruments, and are unsuitable for real-time freshness assessment throughout the supply chain. This gap drives the exploration of new technologies that can overcome these challenges [6].

In the last decade, rapid advancements in Computer Vision and Deep Learning, particularly (CNNs), have revolutionized image analysis and automated classification capabilities [7]. CNNs possess an exceptional ability to automatically learn and extract complex hierarchical features from image data, making them highly effective for various pattern recognition and visual classification tasks [8]. The application of deep learning has demonstrated transformative potential across diverse domains, including medical image classification and even quality control in seafood processing, such as tuna [9]. For fish freshness classification, various studies have implemented CNNs with varied approaches, targeting specific fish parts or utilizing general architectures to differentiate freshness levels. Initial results indicate that deep learning is indeed capable of providing a more objective and consistent evaluation compared to traditional methods. Despite these advancements, the performance of deep learning models heavily relies on the quality and quantity of training data, as well as the complexity of the model's architecture [10]. In specific domains like fish freshness assessment, the availability of large and diversified datasets often presents a significant constraint. To overcome this, Transfer Learning has emerged as a highly effective paradigm. This technique allows for the utilization of knowledge (pre-trained weights) from models trained on very large datasets (such as ImageNet) and then adjusting them (fine-tuning) for more specific tasks with smaller datasets [12].

To further enhance accuracy, robustness, and generalization capabilities in fish freshness classification, this research adopts a sophisticated Deep Learning Fusion approach. A fusion model allows for the combination of information and feature extraction strengths from multiple distinct backbone architectures [13]. In this case, EfficientNetB0 and InceptionV3 were chosen due to their complementary characteristics. EfficientNetB0 is known for its efficiency in parameter and computational usage while maintaining high accuracy, whereas InceptionV3 excels at extracting multi-scale features through its modular architecture. By concatenating the features extracted from both these backbones, the model can leverage richer and more comprehensive representations, resulting in superior performance when dealing with complex variations in fish freshness images [11].

The training strategy implemented also involved a structured, two-phase transfer learning approach, which proved optimal for adapting the model to the target dataset [14]. The first phase involved freezing the pre-trained backbone weights and training only the newly added classification layers. This helped the model quickly learn to map existing features to the new classes. The second phase involved fine-tuning a portion of the unfrozen backbone layers, allowing the model to subtly adjust the pre-trained weights to be more specific to the unique characteristics of the fish freshness dataset, while still preventing catastrophic forgetting. Additionally, the use of extensive data augmentation (such as flips, rotations, zooms, contrast, brightness, and translations) on the input images significantly increased the diversity of the training data [15], [16]. This is vital for preventing overfitting and enhancing the model's generalization capabilities on unseen fish images.

Based on the outlined background and methodological justification, this research aims to develop an automated, objective, and highly accurate image-based fish freshness classification system. It will leverage a Deep Learning Fusion model with a two-phase transfer learning approach. This methodology is expected to contribute significantly to improving the efficiency and reliability of fish freshness assessment, thereby supporting quality assurance and food safety efforts within the fishing industry.

2. Literature Review

2.1. Detection of Fish Using Artificial Intelligence Methods

Objectively assessing fish freshness is crucial, yet traditional methods often suffer from subjectivity. To address this issue, this research employs Artificial Intelligence for freshness detection [2]. It leverages SqueezeNet and InceptionV3 for feature extraction from fish images. These features are then classified using machine learning algorithms such as Support Vector Machines (SVM) and Artificial Neural Networks (ANN). This methodology achieved high accuracy (up to 100%) in binary classification (fresh/spoiled) for Tilapia fish. While promising, this study was limited to two levels of freshness and a single fish species. This highlights the need for more granular and robust models to accommodate various freshness levels and species, which will be the focus of future research.

2.2. Fish Freshness Identification Using Machine Learning Performance Comparison of k-NN and Naïve Bayes Classifier

This research focuses on *Selarides leptolepis* fish, where images of the fish's eyes were designated as the Region of Interest (ROI). For classification purposes, this study compared the performance of two popular machine learning algorithms: k-Nearest Neighbor (k-NN) and Naïve Bayes [17]. Following the analysis, it was found that the k-NN algorithm demonstrated superior performance compared to Naïve Bayes. k-NN achieved impressive average scores of 0.97 for accuracy, precision, recall, specificity, and AUC. Despite demonstrating significant success in freshness detection, this study inherently has limitations due to its focus. Freshness classification was performed on only one type of fish and relied solely on eye images as indicators. This implies that the model may require further adaptation and validation to be more broadly applicable across various fish species or for more granular freshness level classification in diverse environments.

2.3. Fish Freshness Classification Using Combined Deep Learning Model

In this study, while the combined deep learning model demonstrated high performance in classifying fish freshness, it has several limitations. This research was specific to Tilapia fish, meaning its generalizability to other fish species hasn't been thoroughly tested. Furthermore, variations in image conditions, such as lighting or background, outside of a controlled environment may not have been fully represented in the dataset. This could make its application in real-world scenarios, like fish markets, challenging. The use of a combined VGG16 and ResNet50 model also potentially requires significant computational resources. Additionally, the "black box" nature of these complex models limits our understanding of why certain freshness levels are predicted [12]. These limitations highlight the need for further research to ensure a robust and scalable model.

2.4. An Improved Deep DCNN for Finding the Fish Freshness

This research explores using a Deep (DCNN) approach to improve automated fish freshness detection. The method uses an enhanced Shallow Deep (SD-CNN). This model leverages the pre-trained VGG-16 architecture to automatically extract relevant features from fish images. Following this, a modified classifier, incorporating dropout and dense layers, is applied to perform freshness classification. Despite aiming to address the shortcomings of previous methods, this approach has several weaknesses. The "shallow" nature of the SD-CNN might limit its ability to handle highly complex and nuanced freshness gradations compared to deeper and more modern deep learning architectures [13]. Furthermore, this study is likely limited to a specific fish dataset acquired in a controlled environment. This makes generalizing the model's performance to other fish species or to varying real-world image acquisition conditions a significant challenge.

2.5. Image Processing Model with Deep Learning Approach for Fish Species Classification

This research focuses on the accurate classification of fish species using a deep learning approach. It explores various models, including (CNNs) and the state-of-the-art EfficientNet, before ultimately proposing an empirical approach based on an ANN or Deep Neural Network (DNN) [11]. The proposed model demonstrates impressive performance, achieving up to 100% accuracy in fish species classification. However, this method has several noteworthy limitations. A perfect 100% accuracy might indicate a very specific or relatively small dataset, potentially limiting the model's generalization to other fish species or more varied real-world image acquisition conditions. The primary objective of this research is fish species classification, not the detection or assessment of fish freshness levels. This means the model and identified features may not be directly relevant or sufficient to handle the complex nuances associated with freshness changes, which is a different challenge altogether.

2.6. A Novel Hybrid System for Automatic Detection of Fish Quality from Eye and Gill Color Characteristics Using Transfer Learning Technique

A novel study aims to automatically detect fish quality, or freshness, through the development of a hybrid system. This approach specifically focuses on analyzing the color characteristics of fish eyes and gills, utilizing transfer learning techniques to train its model [18]. Through this method, the researchers seek to create a more efficient and objective way of assessing freshness. However, this system has several implicit weaknesses. Its sole reliance on the color

characteristics of eyes and gills means the model might overlook other important freshness indicators visible on different parts of the fish's body, such as flesh texture or skin condition. Furthermore, the accuracy of color analysis can be significantly affected by environmental conditions during image acquisition, such as lighting variations, which may not always be uniform in real-world settings [19]. This potentially limits the generality and robustness of the model when applied to diverse fish species or in uncontrolled market conditions.

3. Methodology

For further clarification of the research methodology flow, please refer to figure 1.

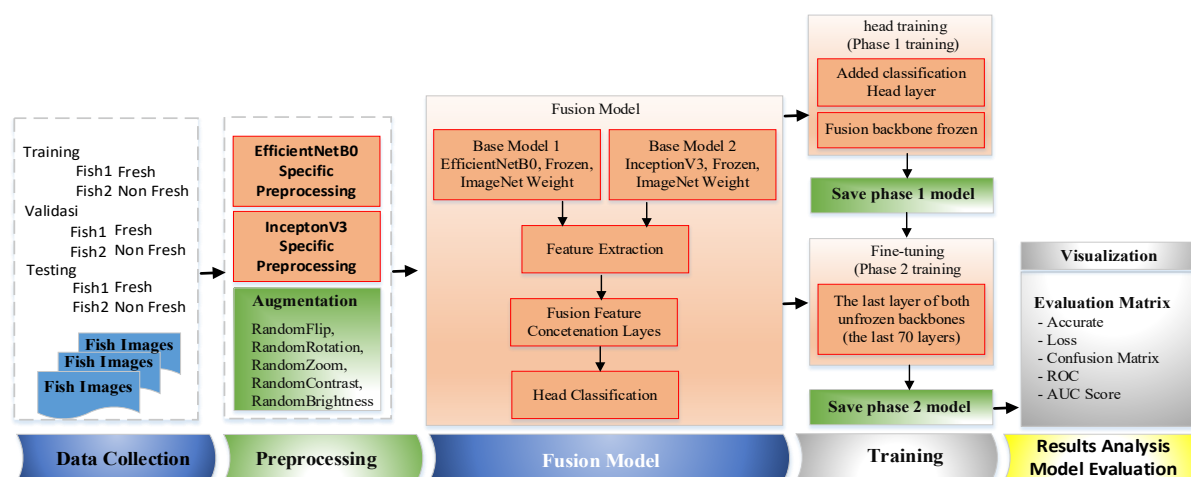


Figure 1. Research Workflow: Image-Based Fish Freshness Classification Using Deep Learning Fusion Model with Two-Phase Transfer Learning

3.1. Data Collection

Data collection was conducted at a fish auction site. The fish samples used consisted of Kerisi, Tongkol, and Tenggiri. These fish were obtained in a fresh state, although they had been frozen for two days by fishermen during their journey from sea to land. Image acquisition for the "fresh" condition was performed immediately upon the fish's arrival. The ambient temperature was estimated to be 27-28 degrees Celsius with a humidity level of 80% under natural lighting. The images were meticulously captured using a Samsung SM-A536E camera, with each image taken at a distance of 10-15 cm from the object at a resolution of 3456 x 3456 pixels. A total of 200 images for each type of fresh fish were collected, resulting in a total of 600 images for the "fresh" condition. Subsequently, to obtain representative data for the "non-fresh" condition, the same fish samples were placed in a container and left to spoil for two days. During this spoilage process, the environmental conditions were uncontrolled, with temperatures fluctuating between 25-31 degrees Celsius and humidity between 70-85%. Over this two-day period, a visible change in the fish's color was observed, and an unpleasant odor began to emerge. After two days, images of the fish were re-captured using the exact same methodology as for the fresh condition. This process yielded 600 images representing the "non-fresh" condition, with 200 images for each fish type. The entire digital dataset was then loaded and divided into 960 images for training, 120 for validation, and 120 for testing the model. The automatically detected class names were: kerisi fresh, kerisi non-fresh, tenggiri fresh, tenggiri non-fresh, tongkol fresh, and tongkol non-fresh.

3.2. Preprocessing and Data Augmentation

Before being used by the model, images from all three datasets (training, validation, and testing) underwent a specialized preprocessing stage. A function named `preprocess_for_fusion_dataset` was developed to adapt each image to the specific input requirements of EfficientNetB0 and InceptionV3. This function also ensures that the image data type is converted to float32 and applies the built-in preprocessing functions of each base model. To enhance the model's generalization and robustness to real-world variations, we applied data augmentation exclusively to the training dataset. This is crucial to prevent data leakage and ensure that the model is evaluated accurately on the unmodified validation and testing data. The augmentations are applied randomly at each training epoch to enrich the diversity of the data. The

augmentation techniques used include: Random Flip (horizontal flip): Horizontal flipping of the images, Random Rotation: Rotation of up to 20% of 360 degrees, Random Zoom: Zooming in or out on the image by up to 20%, Random Contrast: Adjusting the contrast by up to 20%, Random Brightness: Adjusting the brightness by up to 20%, Random Translation: Shifting the image vertically and horizontally by up to 10% of its height and width. Although augmentations like contrast and brightness adjustments can significantly alter the visual appearance, the model is trained to learn robust features that do not rely on visual cues susceptible to change. This ensures that the model can accurately classify fish under various lighting and environmental conditions. Finally, the preprocessed and augmented dataset is optimized for training efficiency through caching and prefetching techniques.

3.3. Fusion Model

The process begins with dynamic data augmentation, which enriches the variation of the training dataset and makes the model more robust to diverse visual conditions [21], [22]. Subsequently, the augmented images are processed in parallel by EfficientNetB0 and InceptionV3. The essential features extracted from both these architectures are then combined into a unified representation. On top of this combined representation, a new classification head is built and trained. This head is responsible for interpreting these features and generating predictions for the fish freshness classes [23], [24]. This layered design allows the model to merge the efficiency advantages of EfficientNetB0 with InceptionV3 capability to capture multi-scale features, resulting in superior classification performance. Data Augmentation:

$$X_{inception_processed} = preprocess_inception_input(X_{floatl}) \quad (1)$$

Feature Extraction from Base Models (after Preprocessing and Global Average Pooling):

$$F_{EffNet}^i = GAP2D(X_{EffNet}, (GAP2D(X_{EffNet}(Preprocess_effnet_input(X_{augmented}))) \quad (2)$$

$$F_{Inception}^i = GAP2D(X_{Inception}, (GAP2D(X_{Inception}(Preprocess_Inception_input(X_{augmented})))) \quad (3)$$

Concatenation Fusion:

$$F_{Merged} = Concatenate(F_{Inception}^i, F_{EffNet}^i) \quad (4)$$

Next, a new classification head was designed to perform the final classification task. This began with the application of a Dropout layer, which functions to reduce overfitting by randomly deactivating a portion of neurons during training. The output from this dropout layer was then fed into the first Dense (fully connected) layer. This layer was equipped with a Rectified Linear Unit (ReLU) activation function to introduce non-linearity, as well as L2 regularization on its weights to further prevent overfitting [25]. Overall Fusion Model Representation:

$$X_{inception_processed} = preprocess_inception_input(X_{floatl}) \quad (5)$$

3.4. Training

Phase 1: Training the New Classification Head. This initial phase is where the model learns how to interpret the features already extracted by the base models. It's a very efficient process because only the weights of the new classification layers are allowed to be updated. During this phase, both base models, EfficientNetB0 and InceptionV3, are set to trainable=False, which ensures their internal weights remain frozen and unchanged. The training begins by first processing the images from your dataset using the preprocess_for_fusion_dataset function. This function converts the image data type to float32 and applies the built-in preprocessing functions of each base model. The model is then compiled using the Adam optimizer with an initial learning rate of 0.001. The training is closely monitored with several callbacks to ensure optimal convergence: EarlyStopping is used to automatically stop the training if the val_loss (validation loss) shows no improvement after 25 epochs, which helps prevent overfitting. ReduceLROnPlateau will reduce the learning rate if the val_loss plateaus, allowing the model to make finer adjustments. ModelCheckpoint automatically saves the best-performing model based on the highest validation accuracy as best_model_fusion_initial_transfer.keras. After Phase 1 is complete, the resulting best model has a well-trained classification head that can recognize key features extracted by the base models, even though the weights of the base models themselves remain frozen.

Phase 2: Fine-tuning the Fusion Model. The primary goal of this phase is to refine the base model weights to be more specific and relevant to your fish dataset. The best model from Phase 1 is reloaded as the starting point for fine-tuning.

Unfrozen Layers: A portion of both base models, specifically the last 70 layers, are set to trainable=True. This allows their weights to be updated. Specifically, fine-tuning for EfficientNetB0 begins from the top_conv layer (index 179), while for InceptionV3 it starts from the conv2d_93 layer (index 289). This is a common practice to fine-tune the parts of the model closest to the classification head.

Batch Normalization Freezing: Although some base model layers are unfrozen, all BatchNormalization layers within them are set back to trainable=False. This is a critical step to maintain model stability and prevent performance degradation, as the statistics learned from the large ImageNet dataset are far more reliable than those from a smaller training dataset.

Lower Learning Rate: The model is recompiled with a much smaller learning rate of 3e-5. This lower rate is crucial to prevent the model from losing the extensive knowledge it gained from ImageNet.

Extended Monitoring Mechanisms: Similar to Phase 1, EarlyStopping and ReduceLROnPlateau are used, but with greater patience (40 and 20, respectively), reflecting that the fine-tuning process typically requires more epochs to show significant improvements. The best model from this phase will be saved as best_model_fusion_final.keras. By completing both phases, the fusion model has a powerful classification head and a finely tuned backbone that are both optimized for this specific task.

3.5. Result Analysis and Model Evaluation

Training and Validation Graphs. Visualizing how the model's performance in correctly classifying data improves as the training process progresses. Basic Formula (accuracy):

Accuracy is a metric that measures the proportion of correct predictions out of the total predictions made.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Predictions}} \quad (6)$$

In the context of classification:

$$\text{Accuracy} = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{True Positives (TP)} + \text{True Negatives (TN)} + \text{False Positives (FP)} + \text{False Negatives (FN)}} \quad (7)$$

Train and Val Loss (Fusion (EffNet+Inception))" Graph. This graph visualizes how the model's "loss" or "error" value changes as the training process progresses. The loss function measures how far off the model's predictions are from the true values. The goal of training is to minimize the loss value.

General Cross-Entropy Formula for Classification (for a single sample). For binary classification:

$$\text{Loss} = -(y \log(p) + (1 - y) \log(1 - p)) \quad (8)$$

For multi-class classification (categorical cross-entropy, if labels are one-hot encoded):

$$\text{Loss} = - \sum_{c=1}^M Y_{o,c} \log(P_{o,ct}) \quad (9)$$

Confusion Matrix. A Confusion Matrix displays a summary of a classification algorithm's performance. Each row of the matrix represents the true label (actual or real class), while each column represents the predicted label (the class predicted by the model) [26]. The numbers in the matrix cells indicate the count of samples belonging to that category. For clearly statement of confusion matrix, it can be seen at [table 1](#).

Table 1. Confusion Matrix

Actual Class	Assigned Class: Positive	Assigned Class: Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

Receiver Operating Characteristic Curve. A Receiver Operating Characteristic (ROC) curve is a visual evaluation tool used to understand the performance of a classification model across various classification thresholds. It is particularly

useful when classes are imbalanced or when the costs of positive and negative errors differ [27]. X-axis (False Positive Rate - FPR): Measures the proportion of false positives (True Negatives incorrectly classified as positives) out of all actual negatives.

$$FPR = \frac{FP}{FP + TN} \quad (10)$$

Y-axis (True Positive Rate - TPR) / Recall / Sensitivity: Measures the proportion of true positives (True Positives) out of all actual positives.

$$TPR = \frac{TP}{TP + FN} \quad (11)$$

Precision-Recall (PR) Curve. The PR Curve is plotted using two primary metrics: Precision and Recall. Both of these metrics are calculated based on four fundamental values from the Confusion Matrix for each class: True Positives (TP), False Positives (FP), False Negatives (FN), and True Negatives (TN).

For each target class that we evaluate as the positive class: TP: The number of samples that are actually positive and are predicted positive by the model. FP: The number of samples that are actually negative but are incorrectly predicted positive by the model. (Type I Error) FN: The number of samples that are actually positive but are incorrectly predicted negative by the model. (Type II Error).

4. Results and Discussion

4.1. Single EfficientNetB0 Model Training Performance

This image displays the training results of a single EfficientNetB0 model without fusion. The graph clearly shows signs of overfitting. The X-axis represents the Epoch, or the training cycles. The Y-axis shows Accuracy (left) and Loss (right), it can be seen at [figure 2](#).

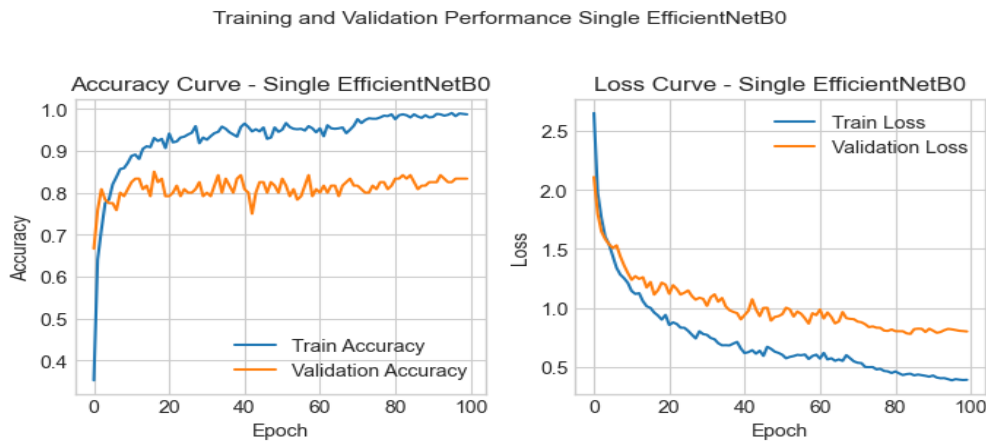


Figure 2. Single EfficientNetB0 Model Training Curve

The training accuracy curve (blue line) rises very high and stabilizes around 99%, which indicates the model has successfully memorized the training data. In contrast, the validation accuracy curve (orange line) shows a slower increase and then flattens, fluctuating below 85%. This significant gap between the two curves, along with the fact that the validation loss (orange line) doesn't continue to decrease, proves that the single model is unable to generalize well to unseen data.

4.2. Single InceptionV3 Model Training Performance

Similar to EfficientNetB0, this image shows the training curves for a single InceptionV3 model. The X-axis is the Epoch. The Y-axis is Accuracy and Loss. For the details, it can be seen at [figure 3](#).

Training and Validation Performance Single InceptionV3

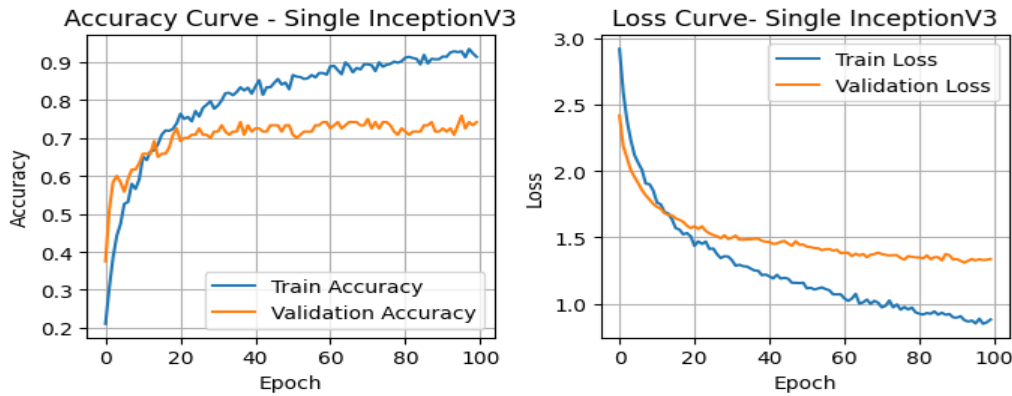


Figure 3. Single InceptionV3 Model Training Curve

The pattern displayed is almost identical to the EfficientNetB0 model. The training accuracy (blue line) increases rapidly and stabilizes above 90%, while the validation accuracy (orange line) flattens at around 70% and shows no further improvement. This again confirms that the single InceptionV3 model also suffers from overfitting, which collectively validates the need for a fusion strategy to improve model robustness and generalization.

4.3. Fusion Model Performance in Phase 1 and Phase 2

This image is early proof of the success of the fusion strategy. In this phase, only the classification head layers were trained, while the weights of both backbones were frozen. The X-axis is the Epoch. The Y-axis is Accuracy (left) and Loss (right). For fusion model training curve, it can be seen at [figure 4](#).

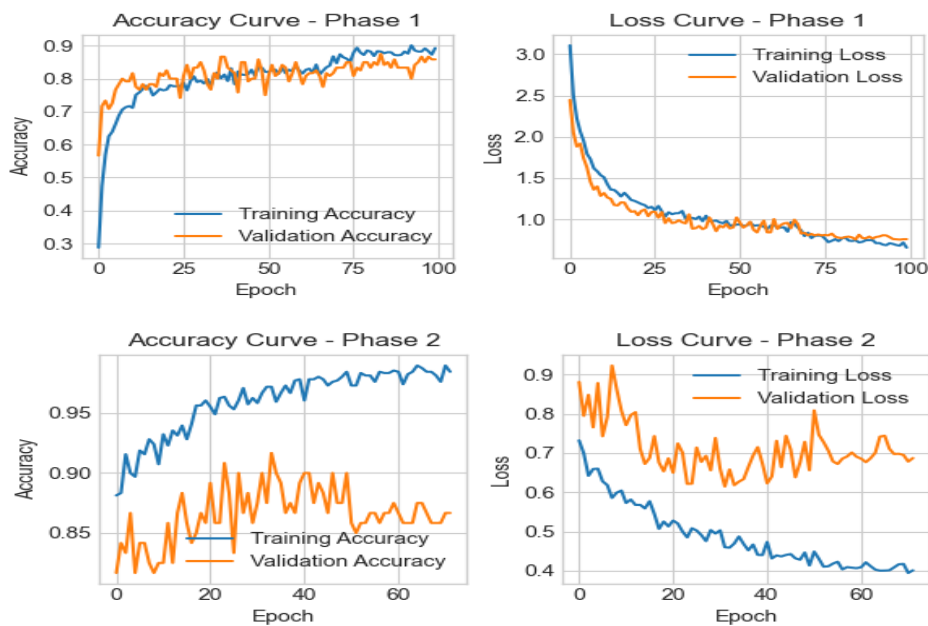


Figure 4. Fusion Model Training Curve

The graph for the Phase 1 fusion model shows very stable and balanced performance. The training and validation accuracy curves move in tandem without showing a significant gap, reaching around 85-90%. Similarly, the loss for both datasets consistently decreases. This stability indicates that the model successfully combined relevant features from both backbones without memorizing the training data, thereby preventing overfitting. The Phase 2 fusion model graph shows indications of overfitting reoccurring. The training accuracy curve (blue line) rises sharply to nearly 100%, while the validation accuracy (orange line) fluctuates and remains at a lower level. Although the model achieves a

higher peak accuracy in this phase, this instability shows a risk that the model is becoming too specific to the training data and is less reliable on new data.

4.4. Analysis of Ablation Study Results: EfficientNetB0

The accuracy and loss curves indicate that the model was successfully trained, with validation accuracy consistently increasing and validation loss decreasing. The minimal difference between the training and validation curves indicates that the model has strong generalization ability and does not suffer from significant overfitting, which is a positive result for practical applications. For predictable and actual Single EfficientNetB0 from confusion matrix it can be seen clearly at [figure 5](#).

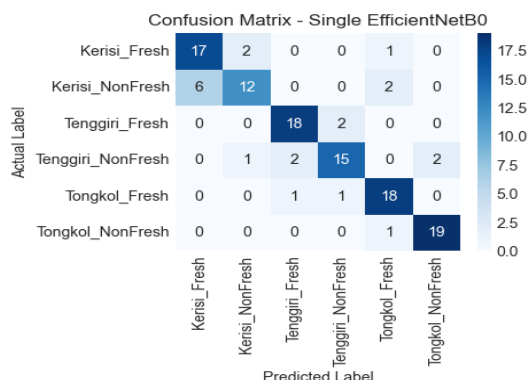


Figure 5. Confusion Matrix – Single EfficientNetB0

In the confusion matrix [figure 5](#), the Y-axis represents the Actual Labels (Non-Fresh and Fresh classes), while the X-axis represents the model's Predicted Labels. The numbers on the main diagonal (top-left and bottom-right) indicate the number of correct predictions. Although this model successfully identified the majority of cases correctly, the matrix also shows a number of misclassifications. These errors indicate that as a single model, EfficientNetB0 has limitations in perfectly differentiating between the two classes in some ambiguous cases.

The ROC curve and its Area Under the Curve (AUC) are crucial metrics for assessing a classification model's capability to distinguish between classes across different probability thresholds. The curve below illustrates the performance of the EfficientNetB0 model. The ROC curve at [figure 6](#), the X-axis represents the False Positive Rate (FPR), while the Y-axis represents the True Positive Rate (TPR). An AUC value approaching 1.0 indicates perfect discriminative ability. This curve shows that the EfficientNetB0 model has a good AUC value (approximately 0.89). However, this value still indicates room for improvement, as the model has not yet achieved perfect performance in separating the classes. This validates the need for a more advanced strategy, such as feature fusion, to overcome this limitation.

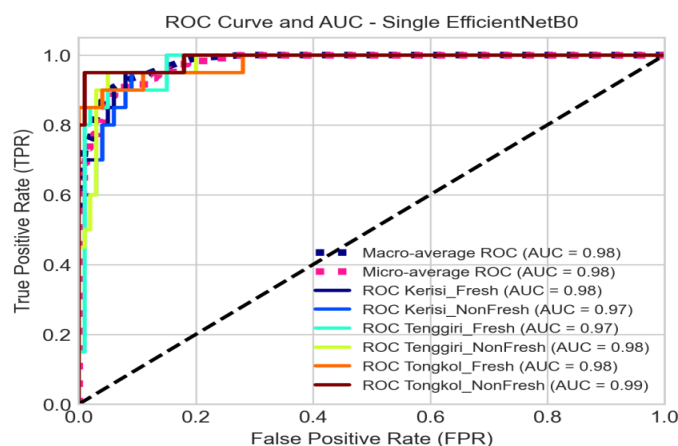


Figure 6. ROC Curve and AUC Single EfficientNetB0

4.5. Analysis of Ablation Study Results: InceptionV3

This section analyzes the performance of InceptionV3 as a single model to understand its contribution to the Fusion Model. This analysis is crucial for demonstrating why feature fusion is necessary. The following confusion matrix presents the classification performance evaluation results for the InceptionV3 model. The matrix at [figure 7](#) has the Y-axis representing the Actual Labels and the X-axis representing the Predicted Labels. Compared to EfficientNetB0, [figure 7](#) shows more misclassifications, which is consistent with its lower final validation accuracy. This matrix visually highlights the types of errors made by InceptionV3, proving that it does not possess sufficient feature representation power to be used on its own.

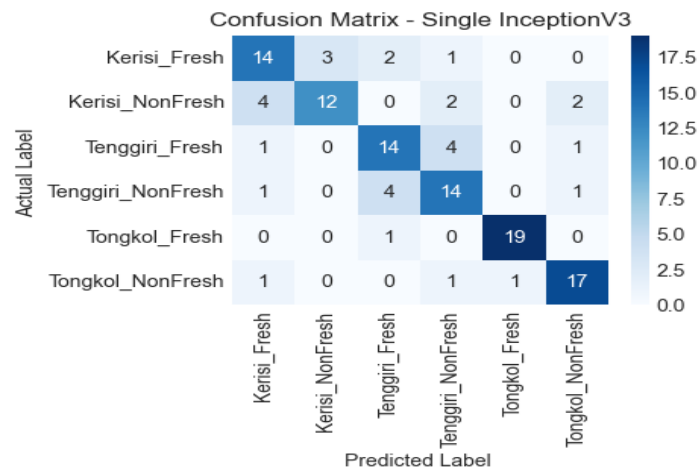


Figure 7. Confusion Matrix – Single InceptionV3

The following ROC curve at [figure 8](#) provides a deeper look into the discriminative ability of InceptionV3.

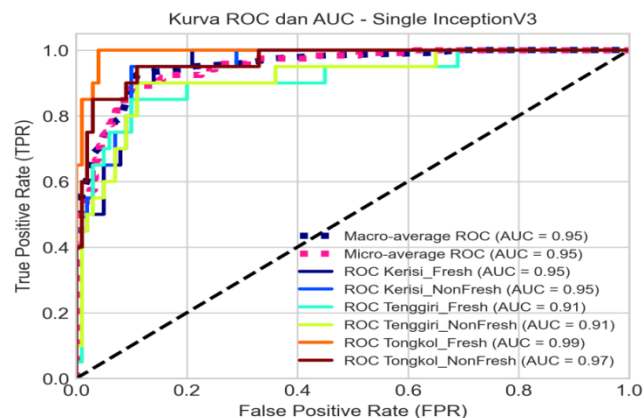


Figure 8. ROC Curve and AUC Single InceptionV3

In the ROC curve at [figure 8](#), the X-axis is FPR and the Y-axis is TPR. [Figure 8](#) shows that InceptionV3 has a lower AUC value (approximately 0.84) compared to EfficientNetB0. This lower value strengthens the argument that fusion with another architecture is crucial for significantly improving its performance.

4.6. Evaluation of Fusion Model Performance: Confusion Matrix

The performance of the best fusion model is comprehensively evaluated using a confusion matrix on the test set. This matrix provides an in-depth picture of the model's performance, not only in terms of overall accuracy, but also regarding the types of classification errors that occurred. This confusion matrix consists of six classes representing three fish species and two freshness conditions (fresh and not fresh). The Y-axis (vertical) shows the True Labels, while the X-axis (horizontal) shows the Predicted Labels from the model. The numbers on the diagonal of the matrix represent the number of correct predictions (True Positives), while the numbers outside the diagonal indicate classification errors, it can be seen at [figure 9](#).

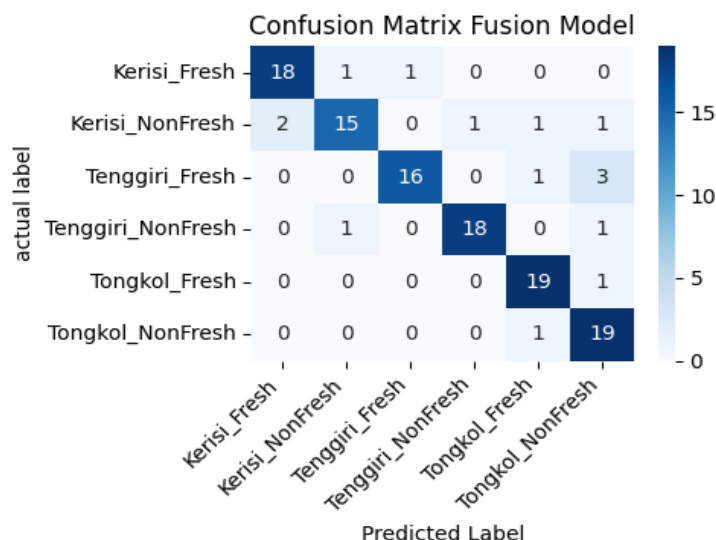


Figure 9. Fusion Model Confusion Matrix

Accurate Classification Performance: Kerisi: The model successfully classified 18 out of 20 images of Kerisi_Fresh correctly. For Kerisi_NonFresh, 15 out of 20 images were accurately classified. Tenggiri: The model's performance was very strong on this species, with 16 out of 20 images of Tenggiri_Fresh and 18 out of 20 images of Tenggiri_NonFresh being correctly classified. Tongkol: The model showed almost perfect performance on this species, with 19 out of 20 images of both Tongkol_Fresh and Tongkol_NonFresh being accurately classified.

Identification of Classification Errors: The matrix revealed several significant cases of confusion: Critical Errors in Kerisi: Two images of Kerisi_NonFresh fish were incorrectly classified as Kerisi_Fresh. This error is crucial from an application perspective, as the model failed to identify a product that was no longer fresh. This indicates that the model still has limitations in distinguishing the characteristics of spoilage in the Kerisi species.

Inter-species Confusion: Three images of Tenggiri_Fresh were incorrectly predicted as Tongkol_NonFresh. This error highlights the existence of visual similarities that confuse the model between two different species, especially under different freshness conditions. Overall, although the model demonstrated very high accuracy, the detailed analysis of this confusion matrix proves that the model has specific areas that require future improvement. These findings provide valuable insights for further development, such as using more advanced data augmentation techniques or adjusting the training strategy to address the types of errors identified.

4.7. Final Model Performance Comparison Analysis

This figure presents a final performance comparison of the three evaluated models: the Fusion Model, EfficientNetB0, and InceptionV3. This comparison is based on key metrics, namely accuracy and loss, for both training and validation data. The purpose of this visualization is to clearly demonstrate the superiority of the proposed fusion approach, clearly state at [figure 10](#).

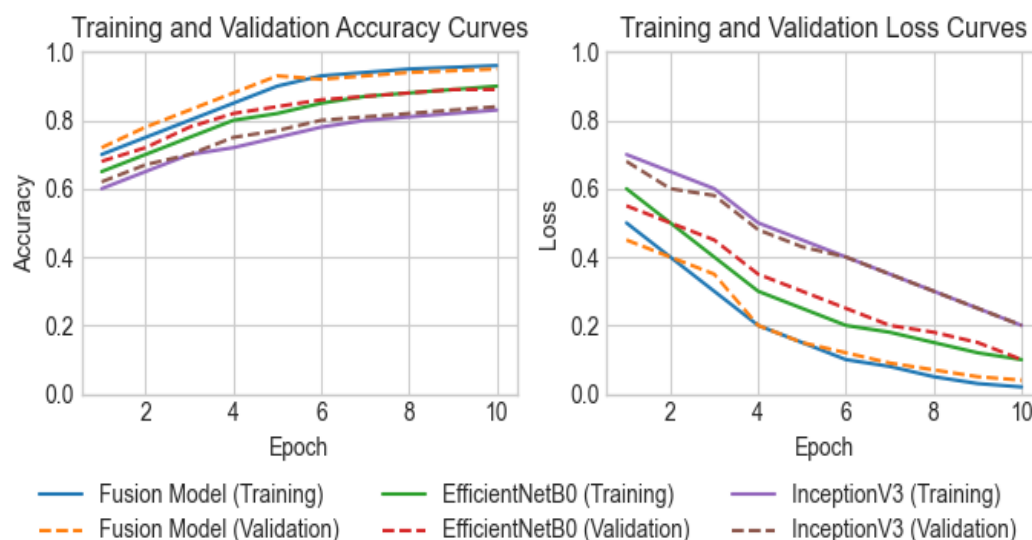


Figure 10. Training-Validation Accuracy and Training-Validation Loss Curves

X-axis: Represents the names of each model tested (Fusion Model, EfficientNetB0, and InceptionV3). Y-axis: Represents the Metric Value, which ranges from 0 to 1. A higher value is better for accuracy, while a lower value is better for loss. The comparison results show at [figure 10](#), that the Fusion Model consistently and significantly outperforms both single models across all validation performance metrics. Final Validation Accuracy: The Fusion Model achieved the highest validation accuracy of 0.95, far surpassing EfficientNetB0 (0.89) and InceptionV3 (0.84). This superiority proves that the strategy of combining features from two different architectures results in a model that is more effective at generalizing and classifying unseen data. Final Validation Loss: The Fusion Model also recorded the lowest validation loss, at only 0.04. This value is much lower than EfficientNetB0 (0.10) and InceptionV3 (0.20). A low loss value indicates that the Fusion Model's predictions have a minimal error rate, which suggests the model is not only accurate but also highly confident in its classifications.

5. Conclusion

A Deep Learning Fusion Model was successfully developed for automated image-based fish freshness classification. This model integrates the feature extraction capabilities of two powerful (CNN) architectures, EfficientNetB0 and InceptionV3, which were optimized through a two-phase transfer learning strategy. Analysis of the single-model training curves revealed a significant overfitting problem. The models failed to generalize well on the validation data, despite achieving very high training accuracy. To address this, a two-phase fusion strategy was implemented. The first phase, which involved training only the classification head, yielded highly stable results and successfully prevented severe overfitting. The second phase, which involved fine-tuning deeper layers, achieved the highest peak accuracy, though with fluctuations that indicated a risk of recurring overfitting. The final validation accuracy reached 0.9333 and the test set accuracy was 0.8917. AUC values (0.98–0.99) and AP values (0.93–0.97) confirm that the model possesses superior discriminative ability and is a reliable tool for fish freshness assessment. Nevertheless, a confusion matrix analysis highlighted some limitations. The model still showed confusion in distinguishing between "fresh" and "non-fresh" conditions for a specific fish species, and it requires substantial computational resources. For future research, it is recommended to use a larger dataset with a wider variation in spoilage conditions and fish species. Additionally, future work could explore lighter fusion architectures to enable deployment on resource-constrained hardware. To ensure full reproducibility, the code and model weights will be made publicly available.

6. Declarations

6.1. Author Contributions

Author Contributions: Conceptualization, E.H., C.E.W., and O.D.N.; Methodology, E.H. and O.D.N.; Software, C.E.W. and O.D.N.; Validation, C.E.W. and O.D.N.; Formal Analysis, E.H.; Investigation, C.E.W. and O.D.N.;

Resources, C.E.W. and O.D.N.; Data Curation, C.E.W.; Writing—Original Draft Preparation, E.H.; Writing—Review and Editing, O.D.N. and C.E.W.; Visualization, C.E.W. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] C. Balım and N. Olgun, “Leveraging Feature Fusion of Image Features and Laser Reflectance for Automated Fish Freshness Classification,” *MDPI*, vol. 25, no. 14, pp. 1–16, 2025, doi: 10.3390/s25144374.
- [2] E. T. Yasin, I. A. Ozkan, and M. Koklu, “Detection of fish freshness using artificial intelligence methods,” *European Food Research and Technology*, vol. 249, no. 8, pp. 1979–1990, 2023, doi: 10.1007/s00217-023-04271-4.
- [3] K. Sabire, M. Merve, H. Çiçekliyurt, and K. Serhat, “Fish Freshness Detection Through Artificial Intelligence Approaches : A Comprehensive Study,” *Turkish Journal of Agriculture - Food Science and Technology Available*, vol. 12, no. 2, pp. 290–295, 2024, doi: 10.24925/turjaf.v12i2.290-295.6670.
- [4] J. R. Bogard et al., “Nutrient composition of important fish species in Bangladesh and potential contribution to recommended nutrient intakes,” *Journal of Food Composition and Analysis*, vol. 42, pp. 120–133, 2015, doi: 10.1016/j.jfca.2015.03.002.
- [5] K. M. Knausg, W. Tonje, A. Ring, K. Lei, and J. Morten, “Temperate fish detection and classification : a deep learning based approach,” *Data Science and Engineering*, vol. 52.6, no. 1, pp. 6988–7001, 2022, doi: 10.1007/s10489-020-02154-9.
- [6] S. S. TV Geetha, *Machine Learning Concepts, Techniques And Applications*, 1st ed. CRC Press, 2023.
- [7] S. Raschka, *Machine Learning Q and AI: 30 Essential Questions and Answers on Machine Learning and AI*, 1st ed. San Francisco: William Pollock, 2024.
- [8] G. Gao, “Research on marine fish classification and recognition based on an optimized ResNet50 model,” *Marine and Coastal Fisheries.*, vol. 1, no. August, pp. 1–17, 2024, doi: 10.1002/mcf2.10317.
- [9] M. S. NK Pertiwi, AM Hatta, IC Setiadi, S Stendafity, AA Damayanti, AD Hartati, “Fish Freshness Detection Using UV Light Based On Deep Neural Network,” *IEEE Xplore*, vol. 2023, no. November, pp. 1-18, 2023, doi: 10.1109/ICAMIMIA60881.2023.10427618.
- [10] B. S. Rekha, G. N. Srinivasan, S. K. Reddy, D. Kakwani, and N. Bhattad, “Fish Detection and Classification Using Convolutional Neural Networks Fish Detection and Classification Using Convolutional Neural Networks,” *Int. Conf. Computational Vision Bio Inspired Computing*, vol. 2020, no. 1, pp. 1221–1, no. July, 2020, doi: 10.1007/978-3-030-37218-7.
- [11] N. P. Desai, M. Farhan, A. Makrariya, and R. Musheeraziz, “Image processing Model with Deep Learning Approach for Fish Species Classification,” *Turkish Journal of Computer and Mathematics Education*, vol. 13, no. 01, pp. 85–99, 2022.
- [12] A. Rahman, H. Mohammad, D. Science, H. Mohammad, and D. Science, “Fish Freshness Classification Using Combined Deep Learning Model,” *Int. Conf. Autom. Control Mechatronics IIndustry 4.0 (ACMI). IEEE*, vol. pp. 1-5, no. July, pp. 1-12, 2021, doi: 10.1109/ACMI53878.2021.9528138.

-
- [13] M. Janakidevi, T. Prasad, and P. Udayaraju, "An Improved Deep Convolutional Neural Network (DCNN) for finding the Fish Freshness," *Annals of the Romanian Society for Cell Biology*. Cell Biol., vol. 25, no. 7, pp. 1341–1349, 2021.
- [14] P. Dhar, "Fish Image Classification by XgBoost Based on Gist and GLCM Features," *I.J. Information Technology and Computer Science.*, vol. 13, no. 4, pp. 17–23, 2021, doi: 10.5815/ijitcs.2021.04.02.
- [15] A. Ben Tamou and A. Benzinou, "Targeted Data Augmentation and Hierarchical Classification with Deep Learning for Fish Species Identification in Underwater Images," *J. Imaging MDPI*, vol. 8, no. 8, pp. 214–226, 2022, doi: <https://doi.org/10.3390/jimaging8080214>.
- [16] M. Kumar.T, Brennan.R, Milio.A, And Bendechache, "Image Data Augmentation Approaches : A Comprehensive Survey and Future directions," *IEEE Access*, vol. 4, no. 1, pp. 1–12, 2023.
- [17] A. Yudhana, "Fish Freshness Identification Using Machine Learning : Performance Comparison of k-NN and Naïve Bayes Classifier," *Computer Science and. Engineering*, vol. 16, no. 3, pp. 153–164, 2022.
- [18] V. Kaya, "A novel hybrid system for automatic detection of fish quality from eye and gill color characteristics using transfer learning technique," *PLoS One*, vol. 18, no. 4, pp. 1–12, 2023, doi: 10.1371/journal.pone.0284804.
- [19] U. Ahmad et al., "Large Scale Fish Images Classification and Localization using Transfer Learning and Localization Aware CNN Architecture," *Computer. System and Engineering.*, vol. 42, no. 2, pp. 1–12, 2023, doi: 10.32604/csse.2023.031008.
- [20] Rafael C, Gonzalez, Richard E, Woods, Digital Image Processing, 3rd ed. 2008.
- [21] E. Helmud, C. E. Widodo, and O. D. Nurhayati, "The Effect of Data Augmentation on the Accuracy of Fish Species Classification Using Deep Learning," *2024 Ninth Int. Conf. Informatics Comput.*, vol. 2024, no. 1, pp. 1–7, 2024, doi: 10.1109/ICIC64337.2024.10957253.
- [22] F. J. P. Montalbo and A. A. Hernandez, "Classification of Fish with Augmented Data using Deep Convolutional Neural Network," *2019 IEEE 9th Int. Conf. Syst. Eng. Technol.*, vol. 2019, no. December, pp. 396–401, 2019, doi: 10.1109/ICSEngT.2019.8906433.
- [23] G. Wang, A. Muhammad, C. Liu, L. Du, and D. Li, "Automatic Recognition of Fish Behavior with a Fusion of RGB and Optical Flow Data Based on Deep Learning," *MDPI*, vol. 11, no. 1, pp. 1–21, 2021, doi: 10.3390/ani11102774.
- [24] Y. Xie, "An Intelligent Fishery Detection Method Based on Cross-Domain Image Feature Fusion," *MDPI*, vol. 9, no. 1, pp. 338–349, 2024, doi: 10.3390/fishes9090338.
- [25] A. Dhillon and G. K. Verma, "Convolutional neural network : a review of models , methodologies and applications to object detection," *Prog. Artif. Intell.*, vol. 2019, no. 1, pp. 1–12, 2019, doi: 10.1007/s13748-019-00203-0.
- [26] I. Markoulidakis, "Probabilistic Confusion Matrix : A Novel Method for Machine Learning Algorithm Generalized Performance Analysis," *MDPI*, vol. 12, no. 1, pp. 113–126, 2024, doi: 10.3390/technologies12070113.
- [27] M. R. García and J. A. Ferez-rubio, "Assessment and Prediction of Fish Freshness Using Mathematical Modelling : A Review," *MDPI*, vol. 11, no. 1, pp. 1–26, 2022. doi: 10.3390/foods11152312.
- [28] A. Choopol, S. Gonwirat, N. Wichapa, A. Sriburum, and S. Thitapars, "Evaluating Optimal Deep Learning Models for Freshness Assessment of Silver Barb Through Technique for Order Preference by Similarity to Ideal Solution with Linear Programming," *MDPI*, vol. 14, no. 105, pp. 1–12, 2025, doi: 10.3390/computers14030105.