Multi-stage hybrid YOLO-driven and MobileNetV2-CNN variants for robust fish freshness classification

Raseeda Hamzah[1], Rosniza Roslan[1], Amni Munira Khidir[1], Lala Septem Riza[2]
[1]Center of Computing Sciences, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Melaka, Malaysia
[2]Department of Computer Science Education, Faculty of Mathematics and Science Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

Article Info

Article history:

Received Sep 25, 2025
Revised Jun 9, 2026
Accepted Jul 9, 2026

Keywords:

Convolutional neural network variants
Deep learning
Fish freshness
Transfer learning-MobileNetV2
You only look once-driven

ABSTRACT

This study presents a multi-stage hybrid representation learning framework for robust fish freshness classification to address the critical challenge of reliable quality assessment in unpreserved food supply chains. The proposed pipeline operates in two stages: stage-1 employs you only look once (YOLO)v8n as feature-gated detector to validate inputs and eliminate non-fish images, while stage-2 leverages transfer-learned MobileNetV2 variants enhanced with convolutional neural network (CNN) layers, fine-tuning, adaptive learning rate schedulers, and expanded fully connected layers for hierarchical classification into three freshness classes i.e., highly fresh, fresh, and not fresh. The framework has been trained and evaluated on two curated datasets comprising 4,500 fish and non-fish images and 11,111 freshness-labeled including augmented dataset to increase diversity. The experimental results showed significant improvements. The best performance model transfer learning (TL)-MobileNetV2 + CNN + fine-tuning achieved 98.07% training accuracy and 67.72% validation accuracy on 80:20 split, and training accuracy of 97.57%, validation accuracy of 97.21% through 10-fold cross-validation. The comparative benchmarking confirmed that dual-stage design outperformed baseline MobileNetV2 and YOLOv5s models across precision, recall, and F1-score. The findings highlighted significant value of integrating detection-driven validation with transfer-learning classification, and propose new benchmark for intelligent freshness monitoring. For future work, this study aims to explore attention-based models, data integration, and species diversity.

Corresponding Author:

Rosniza Roslan
Center of Computing Sciences, Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
77300 Merlimau, Melaka, Malaysia
Email: rosniza@tmsk.uitm.edu.my

1. INTRODUCTION

The global fish industry is very important for food security and nutrition, especially in developing regions, as fish is the primary source of protein and essential micronutrients such as iodine, selenium, and vitamin B12 [1]. However, fish is one of the most perishable food products, with freshness rapidly deteriorating due to enzymatic, microbial, and physicochemical changes during storage and distribution. When fish is stored improperly after harvest, it can turn into histidine. This process can cause scombroid poisoning, gastrointestinal illness, respiratory complications, and potentially fatal outcomes [2]–[4]. Beyond health risks, spoilage creates ripple effects across supply chains, undermining consumer trust, contributes to food waste, and threaten to economic exporting nations [4].

The traditional methods such as the quality index method are widely used to score sensory attributes including color, odor, and eye or gill clarity, but remain operator-dependent and inconsistent across evaluators [5]. Human evaluators are likely to fatigue and bias, while related costs to maintain skill teams are expensive [6], [7]. The sensory evaluation based on phenotypic markers such as eye clarity, gill coloration, and body structure remain practiced but constrained with subjectivity, inconsistency, and dependence on human experts [5], [6], [8]. The instrumental methods such as torrymeter and spectroscopy-based provide more objectivity but require high acquisition costs, impractical design, and lack of scalability [4], [8]. The manual methods are time-consuming, and some existing digital systems, such as fisheye-based classifiers, have achieved modest accuracy with 77.3%, further limiting adoption [9], [10]. The emerging technologies such as biosensors and electronic noses can detect chemical markers of spoilage, but remain at experimental stage with high costs and technical barriers to commercialization.

The recent advancements in deep learning, convolutional neural networks (CNNs) and lightweight object detectors such as you only look once (YOLO)v8 have transformed automated visual inspection by enabling real-time feature extraction and object localization [9]–[16]. Furthermore, transfer learning with backbones like MobileNetV2 allows pre-trained models to adapt effectively to specialized agricultural domains with limited annotated data [17], [18]. Despite these advances, the existing vision-based approaches show some limitations of single-feature CNN classifiers lack reliability across diverse species [9]–[11], robotic visual systems demand strictly controlled studio environments [4], and standalone lightweight CNNs lack input validation mechanisms to filter out background noise or non-fish distractors [19]–[23]. Thus, these shortcomings highlight integrated dual-stage framework that unifies real-time detection-driven input gating with fine-grained, multi-region classification under a scalable pipeline [15]–[21], [24]–[27].

Therefore, this study introduces FusionNet-YOLO model, dual-stage hybrid deep learning framework for fish freshness classification. In the first stage, YOLOv8n as intelligent input validator, detection of fish versus non-fish images and to reduce false error. In the second stage, transfer learning (TL)-MobileNetV2 variants with CNN layers, fine-tuning strategies, adaptive learning rate schedulers, and additional fully connected layers to perform hierarchical classification of freshness levels. This dual-stage integration establishes a methodological between detection and classification as to improve robustness, enhance discriminative capability, and scalability across datasets.

The main contributions of this study are as follows:

- Dual-stage hybrid framework: to propose novel architecture that integrates YOLOv8n as feature-gated detection stage with TL-MobileNetV2 variants for fish freshness classification. This framework to ensure that only validated fish images are further processed as to reduce error propagation and improve overall application reliability.
- Multi-region feature integration: to enhance the various datasets robustness through the capture localized anatomical features i.e., body, eyes, and gills via data augmentation transformation. This approach allows the framework to generalize across various species, imaging conditions, and varying capture environments.
- Comprehensive benchmark: to validate layered evaluation to represent the significant proposed variants model. At stage-1, YOLOv8n consistently outperformed YOLOv5s in detection accuracy and false-positive reduction. At stage-2, the enhanced TL-MobileNetV2 models outperform baseline MobileNetV2 across every metrics.
- Advanced TL enhancements: to extend the MobileNetV2 backbone with additional CNN layers, fine-tuning, adaptive learning rate schedulers, and fully connected layers that increased validation accuracy from 63.67% (baseline MobileNetV2) to 67.72% (TL-MobileNetV2 + CNN + fine-tuning), and stable generalization at 67.01% with adaptive learning rate scheduling.
- Transferable methodological framework: to generate dual-stage methodology for hierarchical classification tasks. The scalability and adaptability as to extend unpreserved food quality monitoring domains and intelligent image classification.

The structure of this paper is organized as follows. Section 1 introduces research study background, motivation, literature review, proposed solution, and contributions. Section 2 describes the proposed methods i.e., image dataset preparation, data pre-processing, stage-1 feature-gated detection models i.e., YOLOv5s vs. YOLOv8n, and stage-2 MobileNetV2-based variants. Section 3 presents the experimental results and comparative benchmarking against existing methods. Finally, section 4 concludes the research study, contributions to the state-of-the-art, and recommendations for future research.

2. METHOD

The aim of this study is to improve multi-stage hybrid representation learning framework for robust fish freshness classification [19], [20]. This study integrates YOLOv5s vs YOLOv8n feature-gated detection at the first stage with MobileNetV2 variants at the second stage to achieve fine-grained freshness prediction [14], [18], [23]. This dual-stage representation aims to reduce error propagation, improve model stability, and enhance generalization across various datasets, and imaging conditions. The overall proposed model is shown in Figure 1 and outlines as follows: i) image dataset preparation, ii) data pre-processing, iii) stage-1 feature-gated detection via YOLOv5s versus YOLOv8n, and iv) stage-2 robust fish freshness classification with MobileNetV2 variants.

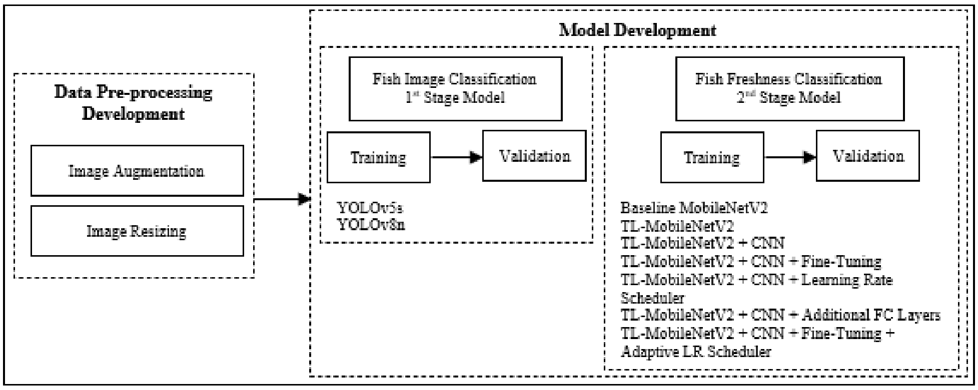

Figure 1. Proposed multi-stage hybrid pipeline

2.1. Image dataset preparation

The image dataset preparation for this study has been conducted to experiment the dual-stage framework that integrates stage-1 feature-gated detection via YOLOv8 for fish versus non-fish validation, and stage-2 freshness classification for fish categorization into freshness levels [4], [14]. This study to ensure that only valid fish images proceed to freshness classification. This approach reduces false positives, mitigate error propagation, and strengthening robustness [9]–[13].

2.1.1. Stage-1 feature-gated detection

For stage-1, the total image dataset of 4,692 images has been collected across two categories: fish and non-fish [14], [16]. The fish class contained 1,560 images, which also used as input for stage-2 freshness classification. These images have been obtained from three repositories with multiple anatomical ranges of body, eye, and gill regions [4], [13]. The non-fish class contained 3,132 images, including marine backgrounds, market environments, and irrelevant objects as to build invariance against environmental noise. The details of the image dataset of stage-1 are presented in Table 1.

Table 1. Stage-1 image dataset description for feature-gated detection

Dataset	Image dataset source	Total number of images
Fish dataset	Fish freshness classification dataset from Kaggle, fish eye freshness dataset from Mendeley, and primary dataset of fish images that captured and collected at local fish markets in Malaysia	1,560
Non-fish dataset	COVID-19 chest X-ray image dataset, jellyfish image dataset, animal image dataset-cats, dogs, and foxes, and garbage image dataset from Kaggle	3,132
Total		4,692

2.1.2. Stage-2 robust freshness classification

For stage-2, the total of 11,736 number of image dataset that has been collected to extend the experiment on robust fish freshness into three classes: highly fresh, fresh, and not fresh. This dataset integrated multiple sources to ensure diversity in anatomical cues and spoilage indicators, i.e. i) fish freshness classification dataset, ii) freshness of the fish eyes dataset, and iii) primary dataset of fish images that has been captured and collected at local fish markets in Malaysia. All datasets emphasized into anatomical regions (body, eye, and gill), to ensure that classification relied on multi-regional features rather than single attribute such as fish eye appearance. The details of image dataset description of stage-2 tabulated as in Table 2.

Table 2. Stage-2 image dataset description for fish freshness classification

Image dataset source	Fish freshness class	Total number of images
Fish freshness classification dataset from Kaggle	Fresh	2,581
	Not fresh	1,895
Fish eye freshness dataset from Mendeley	Highly fresh	1,764
	Fresh	1,320
	Not fresh	1,306
Primary dataset	Highly fresh	949
	Fresh	957
	Not fresh	964
Total		11,736

2.2. Data pre-processing

The primary objective of the data pre-processing phase is to normalize input representations and to quantify the progressive representation of TL strategies across 11,736 multi-regional images [17], [18]. The raw image has been experimented to standard image dimension and feature enhancement as to achieve fine-grained freshness levels across three-class freshness (highly fresh, fresh, and not fresh) [17], [18]. This phase further evaluates seven experimental model variants from the baseline MobileNetV2 backbone that incorporates custom CNN head, partial fine-tuning, adaptive learning rate schedulers, and expanded fully connected classifier layers [17], [27]. Furthermore, targeted data augmentation and image resizing has been applied to reduce overfitting, optimize gradient descent convergence, and improve feature invariance across body, eye, and gill anatomical regions [21], [25].

2.2.1. Data augmentation

The aim of this phase is to increase the diversity of the dataset, prevent overfitting, and increase the number of fish image dataset [12], [15], [21]. For stage-1, feature-gated detection, data augmentation on geometric and photometric transformations to strengthen invariance in detection. For stage-2, robust freshness classification, augmentation emphasized variability in illumination and contrast to simulate real-world capture conditions. The augmentation operations that have been employed for stage-1 and stage-2 are tabulated as Table 3.

Table 3. Data augmentation transformations for stage-1 and stage-2 model

Augmentation Technique	Stage-1	Stage-2
Flipping	Horizontal and vertical	Horizontal and vertical
Rotation	90° (clockwise, counter clockwise upside down); random −15° to +15°	Random −25° to +25°
Cropping/zoom	0–20% zoom applied	Not applied
Grayscale conversion	15% of images converted	Not applied
Brightness adjustment	Between −15% and +15% range	Randomized brightness shifts
Contrast/sharpness	Not applied	Randomized contrast and sharpness

2.2.2. Image resizing

All image datasets have been resized to standard 224×224 pixels to normalize the input image dimension and match the required input size of the MobileNetV2-based model [18], [22]. The standard image dimensional is to ensure batch-processing compatible and uniform matrix operations across heterogeneous images from multiple repository sources. Furthermore, this spatial resolution optimizes the hybrid model performance by balancing computational efficiency and feature preservation, thereby retain fine-grained visual details across the eye, gill and body during the model training [22], [25].

2.3. Stage-1 feature-gated detection

The aim of stage-1 model is to detect and classify the feature-gating of fish image representation [14], [16]. Two models from the YOLO family i.e., YOLOv5s and YOLOv8n have been experimented and evaluated as to benchmark the performance [23]. This study employed YOLOv5s, and YOLOv8n as to validate fish image. By filtering out non-fish inputs, this stage acted as feature-gating mechanism that reduced false positives and minimized error propagation into stage-2.

2.3.1. YOLOv5s

This study experimented YOLOv5s as the baseline object detection model, and a lightweight member of the YOLOv5 family [16]. YOLOv5s combines real-time detection with bounding box localization, enabling identification of fish objects in cluttered environments [23]. The stage-1 dataset contained 4,692 labeled images split between fish (1,560 images) and non-fish (3,132 images). For YOLOv5s training, bounding boxes have been manually annotated for the fish class, and non-fish images labelled as background. The experimented key features are to enhance multilevel feature fusion and decoupled attention mechanisms, reduce false positives and improve detection of small objects, simple, parameter-free attention module (SimAM) and custom channel attention modules strengthen focus on relevant features in complex scenes, to evaluate advanced intersection over union (IoU)-based losses, including focal minimum point distance IoU and complete intersection over union (CIoU), address class imbalance and improve convergence; and bidirectional feature pyramid network (BiFPN) with coordinate attention refines feature representation, boosting detection accuracy. The training experiment has been performed for 5 epochs with an input size of 640 × 640 pixels. The experimental results achieved an acceptable 76.745% precision and 62.766% recall at epoch four. However, this study showed two main issues: i) increased computational overhead due to bounding box regression as to reduce suitability for lightweight deployment; and ii) reduced precision in cluttered backgrounds, allowing some non-fish images to pass through as false positives.

2.3.2. YOLOv8n

The YOLOv8n has been experimented to compare with YOLOv5s for fish vs non-fish detection. The aim is to remove the bounding box annotations and simplify the pipeline while producing high discriminative performance [14], [23]. Key factors of YOLOv8n pipeline are to represent binary problem (fish and non-fish), then bounding box annotations is eliminated, hence simplify the detection pipeline while provides high discriminative performance, optimize for real-time detection, and suitable for fisheries applications. YOLOv8n also combine FasterNet blocks and efficient multi-scale attention (EMA) attention as to enhance feature extraction and detection accuracy, employs BiFPN-based multi-scale feature fusion to optimize feature pyramid sampling, reduces parameters for efficient real-time inference in resource-constrained environments, achieves lower complexity and faster inference, and offers strong performance for speed and accuracy, achieving higher overall performance metrics compared to YOLOv5s [14], [23].

2.4. Stage-2 robust freshness classification with TL-MobileNetV2 variants

Following stage-1 validation with YOLOv8n, only verified fish images have been processed into stage-2 for hierarchical freshness classification into highly fresh, fresh, and not fresh [15], [26]. This stage aims to experiment the lightweight MobileNetV2-based architecture as to improve efficiency, accuracy, and suitable for high-dimensional visual tasks under computational constraints [18], [25]. The baseline MobileNetV2 adapted by removing the original top layers and integrating customized classification head comprising average pooling, dense layers, dropout, and softmax output. The objective is to modify the extraction of domain-specific features, mainly from eye, gill, and body regions [15]–[27]. The total of seven refinements TL–MobileNetV2 variants have been developed that incorporate CNN extensions, fine-tuning, adaptive learning rate schedulers, and expanded fully connected layers as experimented in Table 4. The aims are to advance representation learning, mitigate overfitting, and improve generalization across heterogeneous datasets, thus establishing a robust freshness classification pipeline that outperforms conventional single-model approaches. The seven experimental refinement models are summarized as shown in Table 4.

Table 4. The seven proposed TL–MobileNetV2 variants model

Model	Description	Performance
Baseline MobileNetV2	Pretrained on ImageNet, used as lightweight feature extractor.	Achieved strong training accuracy (94.26%) but low validation accuracy (63.67%), indicating overfitting and weak generalization.
TL-MobileNetV2	Incorporated transfer learning with pretrained weights and custom classification head.	Improved validation accuracy to 67.11%, showing better adaptability to domain-specific features.
TL-MobileNetV2 + CNN	Combine convolutional layers on top of MobileNetV2 to capture finer local textures.	Boosted training accuracy (95.39%) and improved stability in body and gill datasets.
TL-MobileNetV2 + CNN + fine-tuning	Combined CNN extensions with selective fine-tuning of top MobileNetV2 layers.	Achieved the high training accuracy (98.07%) and highest validation accuracy (67.72%), demonstrating improved feature discrimination and robustness across augmented datasets.
TL-MobileNetV2 + CNN + learning rate scheduler	Introduced dynamic schedulers to stabilize convergence and reduce loss oscillations.	Reached validation accuracy of 66.78%, showing balanced generalization across datasets.
TL-MobileNetV2 + CNN + additional fully connected layers	Expanded fully connected layers to increase representational depth.	Improved performance moderately (65.10% validation accuracy), but with diminishing returns and higher computational cost.
TL-MobileNetV2 + CNN + fine-tuning + adaptive learning rate scheduler	Integrated all refinements into the most advanced configuration.	Achieved the highest training accuracy (98.63%) and strong validation accuracy (67.01%), with robustness.

3. RESULTS AND DISCUSSION

This section presents the experimental findings of the proposed multi-stage hybrid representation learning framework for robust fish freshness classification [4], [12], [19]. The detailed explanation of stage-1 feature-gated detection with YOLOv5s and YOLOv8n, stage-2 robust freshness classification using MobileNetV2 variants as follows. The analysis results are presented below with accuracy, precision, recall, and F1-score. The comparative evaluations with prior studies are also discussed [10], [13], [21], [26].

3.1. Stage-1 feature-gated detection via YOLOv5s–YOLOv8n: fish versus non-fish

The first stage of the pipeline evaluated YOLOv5s and YOLOv8n as competing architectures for feature-gated detection of fish versus non-fish classification prior to freshness analysis [14], [16]. Based on Table 5, it is shown that YOLOv8n consistently outperformed YOLOv5s across all key evaluation metrics. This feature-gated represents pre-filter that effectively suppresses background distractors, market environment noise, and non-fish objects, thereby to prevent false-positive error propagation into stage-2 [16]. Furthermore, YOLOv8n demonstrated faster inference speeds and reduced parameter complexity compared to YOLOv5s, thereby valid for practical suitability for real-time edge deployment in fisheries quality monitoring [14], [23].

Table 5. YOLOV5s and YOLOv8n

Epoch	YOLOv5s		YOLOv8n		
	Precision	Recall	Training accuracy	Training loss	Validation loss
0	0.18036	0.25897	0.9837	0.19066	0.07398
1	0.4865	0.47819	0.99457	0.04636	0.08152
2	0.49789	0.45945	0.98913	0.04373	0.06151
3	0.64529	0.55567	0.99457	0.03114	0.07025
4	0.76745	0.62766	0.99457	0.01522	0.07883

3.2. Stage-2 robust freshness classification on 80:20 split: baseline to hybrids

The aim of stage-2 is to benchmark the effectiveness of progressive MobileNetV2 enhancements under an 80:20 training–validation split, from the baseline backbone to hybrid variants [15], [26]. Table 6 shows average training and validation accuracy for MobileNetV2 variants model performance. Based on Table 6, baseline MobileNetV2 achieved 63.67% validation accuracy transfer learning and CNN integration provided incremental gains. Fine-tuning further improved stability, reaching 67.72%, while the hybrid with adaptive learning rate scheduling achieved the best trade-off between training accuracy (98.63%) and validation accuracy (67.01%). These results highlight the stage-2 model particularly CNN extensions and adaptive schedulers, mitigate overfitting and strengthen generalization compared to the baseline.

Table 6. Average training and validation accuracy for MobileNetV2 variants model performance

Model	Training accuracy	Validation accuracy
Baseline MobileNetV2	94.26	63.67
TL-MobileNetV2	93.03	67.11
TL-MobileNetV2 + CNN	95.39	64.72
TL-MobileNetV2 + CNN + fine-tuning	98.07	67.72
TL-MobileNetV2 + CNN + learning rate scheduler	96.81	66.78
TL-MobileNetV2 + CNN + Additional FC layers	95.25	65.10
TL-MobileNetV2 + CNN + fine-tuning + adaptive learning rate scheduler	98.63	67.01

3.3. Stage-2 evaluation metrics for MobileNetV2 variants across dataset

The aim of this section is to evaluate and compare the performance of TL-MobileNetV2 variants across various datasets. These evaluations have been conducted to determine the effect of architectural refinements, optimization strategies, and training duration on the robustness and generalization of the freshness classification system [15], [25], [26]. This section experimented for each variant for training and validation on heterogeneous datasets consisting of body, eye, and gill images across three freshness levels (highly fresh, fresh, and not fresh), and across 5, 10, and 15 epochs, focusing on training accuracy, validation accuracy, and generalization trends based on seven refinements variants [18], [27] as shown in Tables 7 to 14.

Based on the Table 14, the baseline MobileNetV2 achieved modest accuracy (92.23% training, 62.54% validation), while transfer learning improved validation to 65.90%. The CNN layers extension produced small significant, while incorporating fine-tuning significantly produced validation accuracy to 67.88% from TL-MobileNetV2. The learning rate scheduler maintained balanced performance (93.39% training, 65.52% validation), whereas additional fully connected layers led to overfitting with reduced validation (61.89%). The best performance model, TL-MobileNetV2 with CNN, and fine-tuning, resulted 94.97% ± 0.04 training and 67.88% ± 2.10 validation accuracy. Thus, fine-tuning integration enhance robustness more effectively than increased architectural depth. This model demonstrates both strong accuracy and stable convergence.

Table 7. Stage-2–training, and validation accuracy for baseline MobileNetV2 model performance

Dataset	5 epochs		10 epochs		15 epochs	
	Training	Validation	Training	Validation	Training	Validation
Fish freshness classification dataset	100	100	100	100	100	100
The freshness of the fish eyes dataset	65.02	41.51	70.88	41.16	72.93	41.85
Primary dataset–body	88.45	54.79	92.39	67.02	93.83	67.55
Primary dataset–eyes	92.49	52.88	96.89	49.21	98.83	53.40
Primary dataset–gills	90.74	56.84	97.78	60.00	98.17	57.37
Primary dataset with data augmentation–body	90.28	72.31	92.46	74.64	93.99	74.64
Primary dataset with data augmentation–eyes	94.70	50.93	96.68	56.35	98.29	56.72
Primary dataset with data augmentation–gills	93.66	56.31	96.90	57.73	98.06	57.80
Average	89.42	60.70	93.00	63.26	94.26	63.67

Table 8. Stage-2–training, and validation accuracy for TL–MobileNetV2 model performance

Dataset	5 epochs		10 epochs		15 epochs	
	Training	Validation	Training	Validation	Training	Validation
Fish freshness classification dataset	99.78	100	100	100	100	100
The freshness of the fish eyes dataset	59.78	41.73	66.30	43.79	70.20	43.33
Primary dataset–body	89.37	64.89	93.18	64.89	95.14	70.21
Primary dataset–eyes	92.36	54.45	98.06	54.97	99.22	58.18
Primary dataset–gills	93.61	61.05	98.83	64.74	99.48	67.37
Primary dataset with data augmentation–body	89.78	75.47	92.33	75.55	92.99	76.90
Primary dataset with data augmentation–eyes	93.18	56.94	96.57	59.91	97.26	55.75
Primary dataset with data augmentation–gills	93.19	61.09	96.87	65.20	97.11	65.12
Average	88.88	64.45	92.77	66.13	93.93	67.11

Table 9. Stage-2–training, and validation accuracy for TL–MobileNetV2 + CNN model performance

Dataset	5 epochs		10 epochs		15 epochs	
	Training	Validation	Training	Validation	Training	Validation
Fish freshness classification dataset	100	100	100	99.89	99.69	99.44
The freshness of the fish eyes dataset	51.64	42.42	61.20	41.28	79.05	41.96
Primary dataset–body	87.90	70.74	91.38	65.96	94.09	68.62
Primary dataset–eyes	95.08	60.21	99.61	52.88	99.87	53.93
Primary dataset–gills	94.52	57.89	98.17	65.26	98.70	69.47
Primary dataset with data augmentation–body	90.62	68.10	92.60	67.72	94.78	70.58
Primary dataset with data augmentation–eyes	95.24	56.57	97.26	54.79	98.39	53.45
Primary dataset with data augmentation–gills	93.00	58.03	97.89	56.83	98.58	60.27
Average	88.50	64.25	92.26	63.08	95.39	64.72

Table 10. Stage-2–training, and validation accuracy for TL–MobileNetV2 + CNN + fine-tuning model performance

Dataset	5 epochs		10 epochs		15 epochs	
	Training	Validation	Training	Validation	Training	Validation
Fish freshness classification dataset	99.92	99.89	100	100	99.97	100
The freshness of the fish eyes dataset	73.38	40.25	70.05	41.51	93.82	41.62
Primary dataset–body	94.62	67.55	94.36	65.96	70.74	95.14
Primary dataset–eyes	99.22	60.73	99.87	70.16	97.67	60.73
Primary dataset–gills	99.22	59.47	99.81	62.11	100	65.79
Primary dataset with data augmentation–body	95.49	72.46	97.52	67.12	98.53	74.87
Primary dataset with data augmentation–eyes	99.30	60.73	99.20	58.57	99.78	61.25
Primary dataset with data augmentation–gills	98.36	67.21	98.75	69.53	99.66	66.77
Average	94.94	66.00	94.95	66.87	95.02	70.77

Table 11. Stage-2–training, and validation accuracy for TL–MobileNetV2 + CNN + learning rate scheduler model performance

Dataset	5 epochs		10 epochs		15 epochs	
	Training	Validation	Training	Validation	Training	Validation
Fish freshness classification dataset	99.97	99.89	100	100	100	100
The freshness of the fish eyes dataset	51.38	40.94	75.18	42.42	83.63	41.51
Primary dataset–body	85.83	62.77	93.18	67.02	95.67	70.74
Primary dataset–eyes	94.82	61.26	99.22	54.45	100	60.21
Primary dataset–gills	95.57	57.89	97.39	61.05	99.61	63.16
Primary dataset with data augmentation–body	89.95	73.97	94.64	72.46	96.73	77.05
Primary dataset with data augmentation–eyes	96.46	57.54	99.81	61.40	99.50	55.83
Primary dataset with data augmentation–gills	94.37	60.64	99.10	64.53	99.33	65.80
Average	88.54	64.36	94.82	65.42	96.81	66.79

Table 12. Stage-2–training, and validation accuracy for TL-MobileNetV2 + CNN + additional fully connected layers model performance

Dataset	5 epochs		10 epochs		15 epochs	
	Training	Validation	Training	Validation	Training	Validation
Fish freshness classification dataset	99.89	99.66	99.86	99.66	100	99.89
The freshness of the fish eyes dataset	50.90	41.16	69.31	41.62	81.21	45.72
Primary dataset–body	56.96	56.38	85.17	57.98	90.94	61.70
Primary dataset–eyes	66.58	38.22	83.42	51.83	97.02	57.07
Primary dataset–gills	53.32	42.11	84.35	48.95	97.65	64.21
Primary dataset with data augmentation–body	87.78	74.27	94.72	70.35	96.81	74.57
Primary dataset with data augmentation–eyes	92.64	57.02	98.18	59.24	99.22	52.19
Primary dataset with data augmentation–gills	93.49	60.87	97.95	65.35	99.12	65.47
Average	75.20	58.71	89.12	61.87	95.25	65.10

Table 13. Stage-2–training, and validation accuracy for TL-MobileNetV2 + CNN + fine-tuning + adaptive learning rate scheduler model performance

Dataset	5 epochs		10 epochs		15 epochs	
	Training	Validation	Training	Validation	Training	Validation
Fish freshness classification dataset	99.94	100	100	99.89	100	100
The freshness of the fish eyes dataset	73.87	44.81	87.53	39.00	94.73	39.45
Primary dataset–body	95.67	67.02	94.88	65.43	95.14	76.06
Primary dataset–eyes	99.09	57.59	100	60.21	99.74	51.31
Primary dataset–gills	99.87	56.32	100	62.63	100	67.37
Primary dataset with data augmentation–body	96.84	68.62	97.58	66.37	99.68	69.22
Primary dataset with data augmentation–eyes	99.02	62.51	99.81	59.69	99.93	58.65
Primary dataset with data augmentation–gills	99.27	60.46	99.76	68.98	99.81	73.99
Average	95.45	64.67	97.44	65.21	98.63	67.01

Table 14. Stage-2 for TL-MobileNetV2 variants model–average ± standard deviation (SD)

Model	Training ± SD	Validation± SD	Epochs
Baseline MobileNetV2	92.23 ± 2.06	62.54 ± 1.32	5, 10, 15
TL-MobileNetV2	91.86 ± 2.15	65.90 ± 1.10	5, 10, 15
TL-MobileNetV2 + CNN	92.05 ± 2.82	64.02 ± 0.67	5, 10, 15
TL-MobileNetV2 + CNN + fine-tuning	94.97 ± 0.04	67.88 ± 2.10	5, 10, 15
TL-MobileNetV2 + CNN + learning rate scheduler	93.39 ± 3.48	65.52 ± 1.22	5, 10, 15
TL-MobileNetV2 + CNN + additional FC layers	86.52 ± 8.36	61.89 ± 2.62	5, 10, 15
TL-MobileNetV2 + CNN + fine-tuning + adaptive learning rate scheduler	97.17 ± 1.29	65.30 ± 1.00	5, 10, 15

Overall, the MobileNetV2 variants show strong generalization pattern across 5, 10, and 15 epochs. The baseline MobileNetV2 and TL-MobileNetV2 show small validation rates that indicate transfer learning alone improves adaptation but does not fully address overfitting. The additional CNN (TL-MobileNetV2 + CNN) raises training accuracy with relatively small fluctuations in validation accuracy, while fine-tuned variant (TL-MobileNetV2 + CNN + fine-tuning) achieves the highest average validation accuracy with slightly larger variability across epochs. The learning rate scheduler model maintains balanced trade-off between fit and generalization as configuration with additional fully connected layers clearly overfits, as reflected by high training but weaker validation accuracy. Therefore, the most advanced model (TL-MobileNetV2 + CNN + fine-tuning) delivers high average accuracy with comparatively small changes across epochs, indicating more stable optimization behavior and robust generalization on heterogeneous freshness datasets.

3.4. Stage-2 robustness fresh classification via 10-fold cross-validation

The aim of this experiment is to evaluate the robustness and stability of the proposed TL-MobileNetV2 + CNN + fine-tuning variant across combination datasets. The single 80:20 split experiments may produce bias results depending on the train–test structure dataset 10-fold cross-validation provides a more comprehensive measure of generalization by averaging outcomes across multiple folds [10], [15], [21]. Tables 15 and 16 show the 10-fold cross-validation results.

Table 15. Stage-2–training, and validation accuracy for 10-fold cross-validation

Fold	Training accuracy	Validation accuracy
1	0.9736	0.967626
2	0.9719	0.966697
3	0.9757	0.964897
4	0.9727	0.967597
5	0.975	0.955896
6	0.9719	0.972097
7	0.9724	0.957696
8	0.9732	0.948695
9	0.9755	0.951395
10	0.9749	0.957696
Average	0.97369	0.961029

Table 16. Stage-2–average precision, recall, and F1-score for 10-fold cross validation across class variants

Class	Precision	Recall	F1-score
Highly fresh	0.90	0.92	0.91
Fresh	0.97	0.95	0.96
Not fresh	0.99	0.99	0.99

Based on Table 15, the model maintained consistently high training and validation performance across all folds, with average training accuracy of 97.37% and validation accuracy of 96.10%. The small variance between folds demonstrates that framework generalizes well to unseen data, mitigate overfitting, and prove stability across various datasets. For class-wise evaluation further strengthens these findings. Table 16 reports the averaged precision, recall, and F1-score for each freshness category.

The not fresh class achieved near-perfect performance (precision = 0.99, recall = 0.99, F1-score = 0.99), outperformance proposed model performance with high reliability. The fresh class also demonstrated strong outcomes (precision = 0.97, recall = 0.95, F1-score = 0.96), proposed the mid-range freshness levels are consistently captured. For the highly fresh class, although precision (0.90) and recall (0.92) slightly lower than other classes, the average F1-score of 0.91 remains competitive, highlighted the challenges of distinguishing fine-grained freshness cues at the upper quality boundary. Overall, these results indicate that the dual-stage architecture achieves both accuracy and robustness, outperform the traditional baseline MobileNetV2 model reported in prior studies.

4. CONCLUSION

This paper presented multi-stage hybrid representation learning framework that integrates YOLOv8-driven feature-gated detection with MobileNetV2-CNN transfer learning for fish freshness classification. The two stages pipeline validates input images before predicts freshness across three categories: highly fresh, fresh, and not fresh. The enhanced model across MobileNetV2, CNN extensions, fine-tuning, and adaptive learning rate scheduling backbone achieved highest training accuracy of 98.07% and validation accuracy of 67.72% using 80:20 split, while 10-fold cross-validation presented peak training accuracy of 97.57%, and validation accuracy of 97.21%. The findings highlight the value of dual feature-gated detection with various TL-MobileNetV2 classification in the state of the art for fish freshness monitoring. However, there have some limitations such as misclassifications occasionally occurred between adjacent freshness categories when minimal visual differences and limited to fish species dataset. Beyond fish application, the proposed dual-stage framework offers transferable AI methodology for broader perishable food quality monitoring. The YOLO-based feature-gated detection stage can be repurposed to validate meat cuts, poultry surfaces, or fruit and vegetable deformities, while the MobileNetV2-CNN classifier can be fine-tuned to learn domain freshness cues such as bruising, discoloration, and dehydration. The architecture may strengthen the extraction of subtle spatial textures and progressive degradation indicators.

ACKNOWLEDGMENTS

The authors would like to thank the Universiti Teknologi MARA for research funding and support.

FUNDING INFORMATION

This work supported by the Universiti Teknologi MARA through the project "Lightweight deep neural network pruning modelling based on adaptive approximate general matrix multiplication and iterative structured filter fusion" under Grant 100-TNCPI/INT 16/6/2 (060/2024).

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Raseeda Hamzah		✓		✓	✓		✓		✓			✓	✓	✓
Rosniza Roslan				✓	✓		✓			✓	✓	✓		
Amni Munira Khidir	✓	✓	✓			✓	✓	✓	✓		✓		✓	
Lala Septem Riza	✓									✓		✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest

DATA AVAILABILITY

The data that support the findings of this study are as follows:

- Fish freshness classification in Kaggle at https://www.kaggle.com/datasets/muhammadaburayan/fish-freshness-classification.
- The freshness of the fish eyes dataset in Mendeley data at https://data.mendeley.com/datasets/xzyx7pbr3w/1.
- Primary dataset of fish images that captured and collected at local fish markets in Malaysia at https://drive.google.com/drive/folders/1-o6wqm02ekLmzDLEUAOzSFTykLSi6o25?usp=drive_link
- Chest X-ray (COVID-19 & Pneumonia) image dataset in Kaggle at https://www.kaggle.com/datasets/prashant268/chest-xray-covid19-pneumonia.
- Jellyfish image dataset in Kaggle at https://www.kaggle.com/datasets/anshtanwar/jellyfish-types
- Animal image dataset-cats, dogs, and foxes in Kaggle at https://www.kaggle.com/datasets/snmahsa/animal-image-dataset-cats-dogs-and-foxes.
- Garbage image dataset in Kaggle at https://www.kaggle.com/datasets/farzadnekouei/trash-type-image-dataset.

REFERENCES

[1] B. D. Abera and M. A. Adimas, "Health benefits and health risks of contaminated fish consumption: current research outputs, research approaches, and perspectives," *Heliyon*, 2024, doi: 10.1016/j.heliyon.2024.e33905.
[2] M. F. A. Razak *et al.*, "Scombroid poisoning among prisoners of a prison in Johor, Malaysia," *Malaysian Journal of Public Health Medicine*, vol. 24, no. 1, pp. 232–238, 2024.
[3] M. A. Ibrahim, S. M. Ibrahim, and F. H. Abozaid, "Guidelines of fish supply chain from post harvesting to marketing," *Egyptian Journal of Aquatic Biology & Fisheries*, vol. 28, no. 5, 2024, doi: 10.21608/ejabf.2024.378875.
[4] Y. E. T. Yasin, I. A. Ozkan, and M. Köklü, "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.
[5] E. Esteves and J. Aníbal, "Sensory evaluation of seafood freshness using the quality index method: a meta-analysis," *International Journal of Food Microbiology*, vol. 334, 2020, doi: 10.1016/j.ijfoodmicro.2020.108934.

[6] R. Saeed, H. Feng, X. Wang, X. Zhang, and Z. Fu, "Fish quality evaluation by sensor and machine learning: a mechanistic review," *Food Control*, vol. 139, 2022, doi: 10.1016/j.foodcont.2022.108902.
[7] S. Falahatnejad, Z. Arabi, S. Ghafari, and A. S.-Akbari, "Fish quality assessment using hyperspectral imaging and computer vision: a review," *IEEE Sensors Journal*, vol. 25, no. 14, pp. 26255–26268, 2025, doi: 10.1109/JSEN.2025.3573947.
[8] M. N. Madhubhashini, C. P. Liyanage, A. U. Alahakoon, and R. P. Liyanage, "Current applications and future trends of artificial senses in fish freshness determination: A review," *Journal of Food Science*, vol. 89, no. 1, pp. 33-50, 2023, doi: 10.1111/1750-3841.16865.
[9] M. Hasan, N. Vasker, M. M. Hossain, M. I. Bhuiyan, J. Biswas, and M. R. A. Rashid, "Framework for fish freshness detection and rotten fish removal in Bangladesh using mask R--CNN method with robotic arm and fisheye analysis," *Journal of Agriculture and Food Research*, vol. 16, 2024, doi: 10.1016/j.jafr.2024.101139.
[10] M. G. Lanjewar and K. G. Panchbhai, "Enhancing fish freshness prediction using NasNet-LSTM," *Journal of Food Composition and Analysis*, vol. 120, 2023, doi: 10.1016/j.jfca.2023.105945.
[11] A. Yudhana, R. Umar, and S. Saputra, "Fish freshness identification using machine learning: performance comparison of k-NN and naïve Bayes classifier," *Journal of Computing Science and Engineering*, vol. 16, no. 3, 2022, doi: 10.5626/JCSE.2022.16.3.153.
[12] R. Karmaker, M. A. Tinky, and S. Akter, "Deep learning and image processing approach for fish freshness detection," in *2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON)*, 2024, pp. 1-6, doi: 10.1109/PEEIACON63629.2024.10800710.
[13] İ. Akgül, V. Kaya, and Ö. Z. Tanir, "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, 2023, doi: 10.1371/journal.pone.0284804.
[14] O. Kılcı and M. Koklu, "Automated classification of biscuit quality using YOLOv8 models in food industry," *Food Analytical Methods*, vol. 18, no. 5, pp. 815–829, May 2025, doi: 10.1007/s12161-025-02755-5.
[15] R. F. S. Peries, S. Adeeba, M. F. S. Ahamed, and B. T. G. S. Kumara, "AI-driven solutions for automated fish freshness classification using CNN architectures," in *2025 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, 2025, pp. 1-6, doi: 10.1109/SCSE65633.2025.11031016.
[16] B. Xu, B. Gao, and Y. Li, "Improved small object detection algorithm based on YOLOv5," *IEEE Intelligent Systems*, vol. 39, no. 5, pp. 57–65, 2024, doi: 10.1109/MIS.2024.3399053.
[17] H. Elaggoune, M. Belahcene, and S. Bourennane, "Hybrid descriptor and optimized CNN with transfer learning for face recognition," *Multimedia Tools and Applications*, vol. 81, pp. 9403–9427, 2022, doi: 10.1007/s11042-021-11849-1.
[18] Q. Xiang, X. Wang, R. Li, G. Zhang, J. Lai, and Q. Hu, "Fruit image classification based on MobileNetV2 with transfer learning technique," in *3rd International Conference on Computer Science and Application Engineering (CSAE '19)*, 2019, pp. 1–7, doi: 10.1145/3331453.3361658.
[19] P. Patel and A. Thakkar, "The upsurge of deep learning for computer vision applications," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 1, pp. 538–548, 2020, doi: 10.11591/ijece.v10i1.pp538-548.
[20] I. H. Sarker, "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions," *SN Computer Science*, vol. 2, 2021, doi: 10.1007/s42979-021-00815-1.
[21] S. Biswas, C. Nahata, S. Ghosh, S. Sahoo, and D. Biswas, "Fish freshness detection via hybrid CNN-LSTM: an interpretable deep learning model," in *Biologically Inspired Techniques in Many Criteria Decision-Making (BITMDM 2024)*, 2025, pp. 365–373, doi: 10.1007/978-3-031-82706-8_37.
[22] K. Dey, K. Bajaj, K. S. Ramalakshmi, S. Thomas, and S. Radhakrishna, "FisHook -- an optimized approach to marine species classification using MobileNetV2," in *OCEANS 2023 - Limerick,* Limerick, Ireland, 2023, pp. 1-7, doi: 10.1109/OCEANSLimerick52467.2023.10244558.
[23] W. Ren, D. Shi, Y. Chen, L. Song, Q. Hu, and M. Wang, "A lightweight fine-grained pelagic fish recognition algorithm based on object detection," *Aquaculture International*, vol. 33, no. 2, 2025, doi: 10.1007/s10499-024-01737-4.
[24] J. H. Cheng, D. W. Sun, X. A. Zeng, and D. Liu, "Recent advances in methods and techniques for freshness quality determination and evaluation of fish and fish fillets: a review," *Critical Reviews in Food Science and Nutrition*, vol. 55, no. 7, pp. 1012–1225, 2015, doi: 10.1080/10408398.2013.769934.
[25] E. Prasetyo, R. Purbaningtyas, R. D. Adityo, N. Suciati, and C. Fatichah, "Combining MobileNetV1 and depthwise separable convolution bottleneck with expansion for classifying the freshness of fish eyes," *Information Processing in Agriculture*, vol. 9, no. 4, 2022, doi: 10.1016/j.inpa.2022.01.002.
[26] M. B. Yildiz, E. T. Yasin, and M. Koklu, "Fisheye freshness detection using common deep learning algorithms and machine learning methods with a developed mobile application," *European Food Research and Technology*, vol. 250, no. 7, pp. 1919–1932, 2024, doi: 10.1007/s00217-024-04493-0.
[27] Y. L. Khaleel, M. A. Habeeb, and G. G. Shayea, "Integrating image data fusion and ResNet method for accurate fish freshness classification," *Iraqi Journal for Computer Science and Mathematics*, vol. 5, no. 4, 2024, doi: 10.52866/2788-7421.1226.

BIOGRAPHIES OF AUTHORS

Raseeda Hamzah is currently a senior lecturer at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Melaka Branch, Malaysia. Before joining Universiti Teknologi MARA, she had 3 years of working experience in the telecommunications industry. She has a Ph.D. in Information Technology and Quantitative Sciences from the Universiti Teknologi MARA. Her research interests are in pattern recognition, artificial intelligence, machine learning, and the internet of things (IoT). She can be contacted at email: raseeda@uitm.edu.my.

Rosniza Roslan is a senior lecturer at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Melaka, Malaysia. She received her bachelor's degree of science in Computational Mathematics and master's degree in Information Technology and Quantitative Sciences from Universiti Teknologi MARA, Malaysia. Her research interests include image processing, medical imaging, pattern recognition, artificial intelligence, machine learning, and internet of things. She can be contacted at email: rosniza@tmsk.uitm.edu.my.

Amni Munira Khidir is a final-year computer science (Hons) student at Universiti Teknologi MARA (UiTM) Melaka Branch, specializing in soft computing, image processing, and mobile app development. She has earned multiple dean's list awards, recognition as an Endowment UiTM Scholar, and developed projects such as a fish freshness detection mobile application and a social good volunteer app. With strong skills in Python, Java, SQL, and Android Studio, she has also gained experience through hackathons, including the JOMHACK Varsity Challenge and Melaka Data Hackathon, community leadership with YouthCare Malaysia, and volunteering in national STEM programs. She can be contacted at email: 2022612638@student.uitm.edu.my.

Lala Septem Riza received Ph.D. in Computer Science from Universidad de Granada, Spain, in 2015. He works in the Department of Computer Science Education, Universitas Pendidikan Indonesia, Indonesia. He teaches machine learning, big data platforms, and statistical data science. His research interests are in machine learning, data science, and education. He can be contacted at email: lala.s.riza@upi.edu.