Spectra-MobileNet: A Frequency-Aware Ordinal Framework for Fine-Grained Prawn Freshness Assessment

BHASKAR NANDI[1], YASHRAJ SHARMA[2], MRIDUL GHOSH[3], MARIA ZDIMALOVA[4], SUVAJIT MAJHI[5], KINGSHUK CHATTERJEE[6], TRISHA BEJ[7], SK MD OBAIDULLAH[8], SOUMIT CHOWDHURY[9]

[1]Department of CSE(AIML), Institute of Engineering and Management (IEM), University of Engineering and Management, Kolkata – 700160, West Bengal, INDIA

[2,5]Department of CSE(IoT, CS & BT), Institute of Engineering and Management (IEM), University of Engineering and Management, Kolkata – 700160, West Bengal, INDIA

[3]Department of Computer Science, Shyampur Siddheswari Mahavidyalaya, Howrah - 700312, West Bengal, INDIA

[4]Department of Mathematics & Descriptive Geometry, Slovak University of Technology, 812 43 Bratislava, SLOVAKIA

[6]Department of Computer Science Engineering, Govt. College of Engineering and Ceramic Technology, Maulana Abul Kalam Azad University of Technology, Kolkata - 700106, INDIA

[7]Department of CSE, Institute of Engineering and Management (IEM), University of Engineering and Management, Kolkata – 700160, West Bengal, INDIA

[8]Department of Computer Science & Engineering, Aliah University, Kolkata - 700160, West Bengal, INDIA

[9]Department of Computer Science & Engineering, Government College of Engineering and Ceramic Technology, Maulana Abul Kalam Azad University of Technology, Kolkata-700010, West Bengal, INDIA

Abstract: The freshness assessment of prawns has become imperative in the aquaculture industry. Manual methods that were previously used have been found to be highly inconsistent and expensive too. This implies that there is need for a reliable and efficient technique like the automated image classification, which will carry out the process quickly. Researches related to shrimp quality assessment have largely relied on datasets that are inadequate and only focused on visual features that cannot adequately distinguish freshness differences. To overcome the limitations, the developed Spectra-MobileNet technique has been able to apply lightweight design and frequency domain feature extraction to achieve effective freshness recognition. In contrast to conventional Convolutional Neural Networks, which focus solely on spatial attributes, the technique relies on frequency domain images with MobileNetV2 architecture. The applicability of the method is verified via an experiment using the prawn images dataset.

Key-Words: Prawn Quality Classification, MobileNetV2, Convolutional Block Attention Module (CBAM), Deep Learning, Attention Mechanism, Aquaculture, Image Classification, Computer Vision

Received: March 15, 2026. Revised: May 9, 2026. Accepted: June 8, 2026. Published: September 7, 2026.

1 Introduction

Aquaculture is an important source of food production at the global level, and prawns are one of the most valuable commodities of commercial importance. Prawn quality has direct implications on market worth, consumer acceptability, and export demand. Quality evaluation has previously depended significantly on the application of human judgment, which necessarily creates subjectivity, inefficiency, and inconsistency. With the advancement in computer vision and artificial intelligence (AI) technology, such quality control procedures can now be made automatic with quicker, more precise, and less expensive solutions. Deep learning image classification techniques have shown great potential in recent years for automatic assessment of seafood quality. The research objective is to design a light and attention-augmented deep architecture with the capacity to categorize prawn samples into three quality classes: *Good*, *Intermediate*, and *Bad* with high accuracy.

Mobile and light-weight networks have been the centre of attention in recent breakthroughs in powerful neural models. MobileNetV2, originally introduced in [1], achieves viable accuracy at lesser computational costs with the help of linear bottlenecks and inverted residuals. With the help of TensorFlow datasets in [2], it has been proven that MobileNetV2 outperforms its parental model, MobileNetV1, in numerous image classification applications. The effectiveness of light-weight models in water image analysis was proved in [3], when they applied transfer learning on MobileNetV2 to classify fishes in aquaculture studies and achieved a correctness of 98.18%. The reliability of MobileNetV2 in practical applications in fisheries was also proved in [4], when it was applied to classify dried fish and achieved 100% correctness using one of many different datasets.

Attention mechanisms are becoming increasingly prominent in CNN models to improve feature learning. It continuously improves performance in classification and detection tasks. This concept was furthered in [5], which achieved 99.86% accuracy in fine-grained fish categorization by integrating MobileNetV2 with attention processes. CBAM is a lightweight yet powerful module which employs both channel as well as spatial attention for the first time, as proposed in [6]. Furthermore, after reviewing image classification techniques, [7], gave the conclusion that using non-parametric classifiers like CNNs in conjunction with many discriminative features greatly improves classification accuracy in complicated visual circumstances. Attention mechanisms have also been applied in hyperspectral image analysis and remote sensing beyond aquaculture.

In [8], a visual attention-based ResNet approach was proposed for HSI classification that enhanced spectral–spatial feature representation and surpassed traditional CNN models. Further, in [9], a Spectral Attention Module was introduced which adaptively reweighted the spectral bands by highlighting informative features and reducing redundancy, which enhanced CNN performance in hyperspectral high-dimensional data.In [10], an Enhanced Attention Module (EAM) was introduced to perform remote sensing scene classification with enhanced extraction of salient features in the presence of complex backgrounds and with 94.29% accuracy in the NWPU-RESISC45 dataset. Overall, these papers show the versatility and efficacy of attention mechanisms to further enhance discrimination and classification performance in a broad range of imaging applications and further justify their incorporation in lightweight models for visual inspection tasks.

Though earlier research demonstrates the effectiveness of MobileNetV2 and attention for fish classification, certain research gaps exist. Recent research is primarily targeted towards species detection and not quality classification, i.e., detecting fine visual differences. Also, attention mechanisms for prawn image analysis are extremely limited in existing research, and field evaluation in aquaculture fields is limited. Additionally, existing prawn freshness studies commonly rely on small and controlled datasets, limiting generalization under practical inspection scenarios. In order to address such a gap, this paper proposes an Attention Enhanced MobileNetV2 structure that has embraced CBAM for quality grading of self developed prawn dataset.

The model was capable of highlighting discriminative visual areas and reducing misclassification among visually similar categories. An 80:20 stratified data split and a two stage training regimen were adopted to ensure convergence stability. The trained model achieved a test accuracy of 95.62%, along with strong macro-averaged AUC performance, demonstrating good generalizability and usability. This work contributes a lean, attention-enhanced, and highly accurate automated prawn quality inspection system that can serve as a stepping stone to eventual application with IoT-based aquaculture monitoring systems. The remainder of this paper is mentioned in the next section.

This study aims to develop a computer-assisted prawn quality grading system capable of distinguishing between *Good*, *Intermediate*, and *Bad* grades based on the visual features from images. Since conventional manual inspection in aquaculture is generally time consuming, subjective, and variable in nature, there is a requirement for an intelligent automatic system with uniform and accurate quality evaluation.

The input image (I) of a specimen prawn must be labelled as *Good*, *Intermediate*, or *Bad*. This is formulated as a multi-class image labelling problem. Officially, I $\rightarrow$ *Good*,*Intermediate*,*Bad* To achieve this, a light-weight deep model through the use of CBAM and MobileNetV2 is employed. Whereas the CBAM process deepens the attention of the model to relevant spatial and channel features for enhanced discrimination of visually similar specimens of prawns, the MobileNetV2 architecture provides improved computational efficiency through the use of depth-wise separable convolutions and inverted residual blocks.

The three classes *Good*, *Intermediate*, and *Bad* in the completely custom-designed dataset utilized in the preparation of this work contain a total of 300 images per class, resulting in 900 samples after controlled data augmentation. All images were normalized, resized, and augmented during preprocessing for the realization of successful generalization and biased learning. The suggested method's overall process consists of the following:

1. Collection of datasets and their labeling into *Good*, *Intermediate*, and *Bad* classifications based on their hedonic ratings.

2. Proposed, a lightweight framework that replaces standard sequential attention with a Parallel Spectral-Spatial Fusion (PSSF) module to prevent feature loss.

3. Integrated 2D Discrete Cosine Transform (DCT) to capture high-frequency textural details (e.g., melanosis), overcoming the smoothing effects of standard Global Average Pooling.

4. Introduced a Rank-Consistent Ordinal Regression Head that explicitly models the hierarchy of decay (*Good* $>$ *Intermediate* $>$ *Bad*), reducing severe grading errors.

5. Achieved 95.62% accuracy on a self-constructed prawn dataset, demonstrating that the model is computationally efficient and suitable for real-time IoT aquaculture monitoring.

Section 2 explains the Proposed Methodology by introducing the overall workflow and theoretical background of the proposed approach with the architectural workflow design. Section 3 explains the Experimental details. Section 4 includes Results and Discussion. The paper concludes in Section 5. The whole workflow of the proposed approach is illustrated in Figure 1 (Appendix A).

2 Proposed Methodology

In the domain of biological quality assessment, such as shrimp grading, visual cues are often subtle. Early stages of spoilage, characterized by melanosis (black spots) or slight discoloration, manifest as high-frequency textural changes rather than global shape deformations. Standard Convolutional Neural Networks (CNNs) often fail to maintain these fine-grained information due to aggressive pooling operations.

To address this, we introduce Spectra-MobileNet, a single architecture that combines deep spatial learning with frequency-domain analysis. Our system explicitly reflects the ordinal nature of decay (i.e., *Good* $>$ *Intermediate* $>$ *Bad*), in contrast to conventional methods that consider quality grades as independent categories. Three separate phases comprise the architecture: a rank-consistent ordinal regression head, a unique Parallel Spectral-Spatial Fusion (PSSF) module, and a lightweight feature extractor.

MobileNetV2 is the base model for efficiency by using depth-wise separable convolutions and inverted residuals. CBAM layer is further added to the MobileNetV2 model to improve attention in channel and spatial spaces.

1. Channel Attention brings attention to significant feature maps via inter-channel connections.

2. Spatial Attention focuses on significant locations in every feature map, so the model can pay attention to significant visual patterns with respect to prawn color, texture, and appearance.

The architecture consists of three core components:

1. *Backbone:* A MobileNetV2 feature extractor initialized with ImageNet weights.

2. *Attention Module:* A PSSF module that replaces the standard sequential CBAM.

3. *Classification Head:* A Rank-Consistent Ordinal Head that replaces the standard Softmax classifier.

This blended configuration improves feature discrimination, allowing the model to distinguish fine quality differences in seemingly similar-looking prawns. The overall model architecture of the model is provided in Figure 2 (Appendix A).

2.1 Feature Extraction Backbone

We employ MobileNetV2 as the backbone due to its balance between efficiency and representation capability. Given an input image $I \in \mathbb{R}^{H_{in} \times W_{in} \times 3}$, the network extracts an intermediate feature map F from the final bottleneck layer:

$$F = f_{backbone}(I) \in \mathbb{R}^{H \times W \times C} \tag{1}$$

where H, W denote the spatial dimensions and C represents the number of channels. This feature map serves as the input to our proposed attention mechanism.

2.2 Parallel Spectral-Spatial Fusion (PSSF) Module

A critical limitation in standard attention mechanisms, such as CBAM, is the reliance on Global Average Pooling (GAP) to summarize channel statistics. While GAP acts as a low-pass filter effective for object presence, it suppresses high-frequency variance necessary for detecting texture anomalies (e.g., spoilage spots). Furthermore, sequential attention (Channel $\rightarrow$ Spatial) may dilute beneficial low-level spectral features before spatial localization occurs.

To overcome these limitations, we propose the PSSF module, which processes spectral (frequency) and spatial information in parallel branches.

2.2.1 Branch A: Frequency-Aware Channel Attention

Instead of scalar pooling, we project the feature map into the frequency domain using the 2D Discrete Cosine Transform (DCT). For a specific feature map channel $k \in \{1, ..., C\}$, the 2D-DCT component at frequency (u, v) is computed as:

$$DCT_{u,v}^{k} = \sum_{i=0}^{H-1} \sum_{j=0}^{W-1} F_{i,j}^{k} \cos\left(\frac{\pi u(2i+1)}{2H}\right) \quad (2)$$

$$\times \cos\left(\frac{\pi v(2j+1)}{2W}\right) \quad (3)$$

where $u \in \{0, ..., H-1\}$ and $v \in \{0, ..., W-1\}$.

To compress the representation while retaining texture information, we select specific frequency components. We concatenate the lowest frequency component (DC component, representing structure) and the highest frequency components (representing fine texture/edges) to form a compressed spectral descriptor vector $V_{spec} \in \mathbb{R}^{C \times Z}$:

$$V_{spec} = \text{Concat}([DCT_{0,0}, DCT_{H-1,W-1}]) \quad (4)$$

This vector is passed through a shared Multi-Layer Perceptron (MLP) with a reduction ratio r to generate the spectral attention map $M_{spec} \in \mathbb{R}^{1 \times 1 \times C}$:

$$M_{spec} = \sigma(W_2(\delta(W_1(V_{spec})))) \quad (5)$$

where $W_1 \in \mathbb{R}^{(C/r) \times C}$ and $W_2 \in \mathbb{R}^{C \times (C/r)}$ are weight matrices, δ is the ReLU activation, and σ is the Sigmoid function.

2.2.2 Branch B: Spatial Attention

The spatial branch localizes the object of interest (prawn body) and suppresses background noise. We utilize both average and max pooling along the channel axis to generate a robust spatial descriptor:

$$S_{desc} = \text{Concat}([P_{avg}(F); P_{max}(F)]) \in \mathbb{R}^{H \times W \times 2} \quad (6)$$

A convolution layer with a large kernel size (7×7) is applied to aggregate spatial context, followed by sigmoid activation to obtain the spatial attention map $M_{spat} \in \mathbb{R}^{H \times W \times 1}$:

$$M_{spat} = \sigma(f^{7 \times 7}(S_{desc})) \quad (7)$$

2.2.3 Feature Fusion

Unlike sequential approaches, we employ a parallel residual fusion strategy. This ensures that the network preserves both spectral intensity (what to look for) and spatial localization (where to look) without information degradation. The refined feature map F' is computed as:

$$F' = (F \otimes M_{spec}) + (F \otimes M_{spat}) \quad (8)$$

where $\otimes$ denotes element-wise multiplication with broadcasting. The overall procedure is summarized in Table 1.

Table 1: Parallel Spectral-Spatial Fusion (PSSF): Step-by-step procedure of the dual-branch attention mechanism that independently computes frequency-aware spectral and spatial masks, then fuses them via element-wise summation to produce refined feature maps F'.
Source: created by the authors.

Require: Input Features $F \in \mathbb{R}^{H \times W \times C}$
Ensure: Refined Features F'

Branch A (Spectral):
  for $k = 1$ **to** C **do**
    $D^k \leftarrow \text{DCT2D}(F^k)$
    $v^k \leftarrow [D^k_{0,0},\ D^k_{H-1,W-1}]$ ▷ *Select Low/High Freq*
  end for
  $V_{\text{spec}} \leftarrow \text{Stack}(\{v^k\}) \in \mathbb{R}^{C \times 2}$
  $M_{\text{spec}} \leftarrow \text{Sigmoid}(\text{MLP}(V_{\text{spec}}))$

Branch B (Spatial):
  $S_{\text{avg}} \leftarrow \text{Mean}(F,\ \text{axis} = C)$
  $S_{\text{max}} \leftarrow \text{Max}(F,\ \text{axis} = C)$
  $S_{\text{desc}} \leftarrow \text{Concat}(S_{\text{avg}},\ S_{\text{max}})$
  $M_{\text{spat}} \leftarrow \text{Sigmoid}(\text{Conv}_{7\times7}(S_{\text{desc}}))$

Fusion:
  $F' \leftarrow (F \otimes M_{\text{spec}}) + (F \otimes M_{\text{spat}})$
  return F'

2.3 Ordinal Regression Head

Standard classification (Cross-Entropy) treats errors uniformly. However, quality grading is ordinal ($Good > Intermediate > Bad$). We replace the

Softmax layer with a rank-consistent head containing $K-1$ binary classifiers. The predicted rank is:

$$Rank = \sum_{m=1}^{K-1} \mathbb{I}(\hat{p}^{(m)} > 0.5) \tag{9}$$

where $\hat{p}^{(m)}$ is the probability that the freshness grade $> m$.

This ranking mechanism ensures that the model learns the continuous trajectory of freshness decay, resulting in lower mean absolute error and higher grading consistency. The entire method is presented using Table 2.

Table 2: Spectra-MobileNet Forward Pass & Inference: Complete inference procedure including feature extraction via MobileNetV2, dual-branch PSSF attention, ordinal classification head, and rank-based grade prediction for prawn freshness assessment.
Source: created by the authors.

Require: $I \in \mathbb{R}^{B\times H\times W\times 3}$, Threshold $\tau = 0.5$
Ensure: Grade $G \in \{0, 1, 2\}$ (Bad, Intermediate, Good)

Step 1: Feature Extraction
  $F \leftarrow \text{MobileNetV2}(I) \in \mathbb{R}^{h\times w\times C}$

Step 2: PSSF Module
  Branch A: Frequency Domain Attention
    $D \leftarrow \text{2D-DCT}(F)$
    $V_{\text{spec}} \leftarrow [D_{\text{low}} \oplus D_{\text{high}}]$
    $M_{\text{spec}} \leftarrow \sigma(\text{MLP}(V_{\text{spec}}))$
  Branch B: Spatial Domain Attention
    $S_{\text{desc}} \leftarrow [\text{AvgPool}(F);\ \text{MaxPool}(F)]$
    $M_{\text{spat}} \leftarrow \sigma(\text{Conv}^{7\times 7}(S_{\text{desc}}))$
  Fusion (Element-wise Summation):
    $F' \leftarrow (F \otimes M_{\text{spec}}) + (F \otimes M_{\text{spat}})$

Step 3: Ordinal Classification Head
  $v \leftarrow \text{GlobalAvgPool}(F')$
  logits $\leftarrow$ DenseLayer(v)
  prob $\leftarrow$ Sigmoid(logits), $[p_1, p_2]$

Step 4: Rank Inference
  rank $\leftarrow 0$
  for $i = 1$ **to** $K-1$ **do**
    if prob$[i] > \tau$ **then**
      rank $\leftarrow$ rank $+ 1$
    end if
  end for
  return rank

2.4 Computational Complexity Analysis

The proposed Spectra-MobileNet is designed for efficiency.

- **Backbone:** MobileNetV2 utilizes Depthwise Separable Convolutions, reducing cost to roughly $O(H \cdot W \cdot C^2/k)$, where k is the reduction factor.
- **PSSF Module:**
  - The Spectral branch utilizes a fixed basis projection (DCT). Since the spatial resolution at the bottleneck is small (typically 7×7), the DCT operation is computationally equivalent to a fixed-weight 1×1 convolution, with complexity $O(C)$. The MLP adds $2 \cdot C \cdot (C/r)$ parameters.
  - The Spatial branch adds a single 7×7 convolution over 2 channels, adding negligible FLOPs ($O(H \cdot W \cdot 2 \cdot 7^2)$).

3 Experiment

3.1 Experimental Setup

To guarantee controlled execution and reproducibility, every experiment was carried out on a local workstation. An AMD Ryzen 7 processor, 16 GB of RAM, and a 1 TB SSD were installed in the system. An NVIDIA RTX 4060 GPU with 6 GB of dedicated VRAM was used to speed up model evaluation and training.

The proposed model was implemented using Python and executed within a Jupyter Notebook environment. The deep learning framework was based on TensorFlow with Keras APIs for model development, training, and evaluation. GPU acceleration was enabled through CUDA support to expedite training while maintaining numerical stability.

3.2 Dataset & Preprocessing

This research employed a self-created prawn image dataset that initially consisted of 210 authentic images (70 per category: Good, Intermediate, and Bad), taken on a flat, non-reflective surface. The images were acquired using a VIVO Y73 smartphone, equipped with a triple rear camera system featuring a 64MP main sensor, a 2MP depth sensor, and a 2MP macro lens, over a span of 36 hours at regular intervals.All the images were collected from a stable overhead position at approx. 25–30 cm distance under indoor lighting conditions to reduce lighting and shading variation, and with a neutral white background to decouple the prawns from their surroundings.

To further regularize the training and avoid overfitting due to limited sample size, controlled data augmentation including rotation, horizontal flipping and color balancing were utilized. As a result of augmentation, the dataset was increased to 900 samples (300 per category). Augmentation took place after partitioning the dataset to ensure that no augmented version of a sample was included in both the training and testing sets, thereby avoiding data leakage.Below is the manner in which the dataset is constructed:

1. *Good* : 300 images
2. *Intermediate* : 300 images
3. *Bad* : 300 images

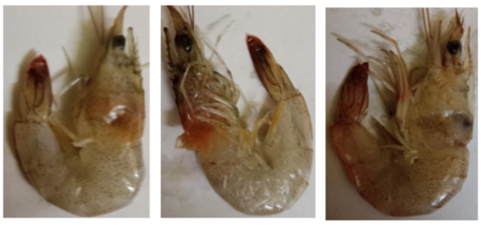

Figure 1:
Sample images from the self-constructed prawn quality dataset showing Good, Intermediate, and Bad classes from left to right
Source: created by the authors.

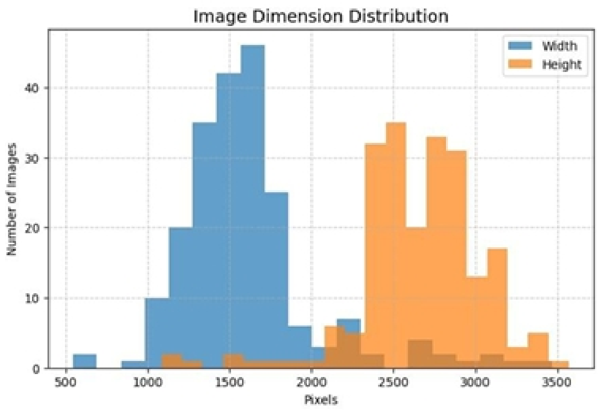

Figure 2: Image Dimension Distribution of the Prawn Dataset
Source: created by the authors.

Figure 1 showcases representative samples from the augmented prawn dataset across the three quality categories. Good samples display consistent translucency and intact textures, signifying freshness, while Bad samples exhibit significant discoloration, a loss of translucency, and surface deterioration linked with spoilage. Intermediate samples reveal overlapping visual attributes, which can make them difficult to differentiate even for human observers, thus highlighting the need for attention-enhanced feature learning.

All images were resized and normalized to fit the input dimensions necessary for MobileNetV2. Figure 2 depicts the frequency distribution of image dimensions in the dataset, with a majority of samples falling within the height range of 2400–2800 pixels and width range of 1400–1700 pixels, suggesting a largely consistent aspect ratio. This level of uniformity aids in reliable preprocessing, normalization, and feature extraction during the model training phase.

3.3 Evaluation protocol

The effectiveness of the proposed model was evaluated by utilizing four standard performance metrics: Accuracy, Precision, Recall, and F1-Score, [11], [12]. Let TP, TN, FP, and FN denote True Positives, True Negatives, False Positives, and False Negatives, respectively. The metrics are calculated as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{13}$$

3.4 Training Strategy

A training approach based on transfer learning is employed, using MobileNetV2 that has been pretrained on ImageNet as the foundational network. During the training phase, the backbone remains fixed while solely the CBAM attention module and the classification head are refined to customize the model to the specific visual traits of prawns, thus preventing overfitting.

All images are adjusted to a size of 224×224 pixels and standardized following the typical MobileNetV2 preprocessing method. To enhance generalization and reduce overfitting, regulated data augmentation was utilized following the partitioning of the dataset to enlarge the original data. Simultaneously, extra real-time augmentation was carried out solely on the training set. The dataset is split in an 80:20 ratio for training and testing, ensuring a balanced distribution of classes. This led to the creation of 720 training samples and 180 test samples after augmentation.

The training process is performed over varying epochs, batch sizes and data splits with the Adam optimizer set to a constant learning rate of 0.0004, employing the sparse categorical cross-entropy loss function for consistent optimization of the ordinal prediction component.Early stopping was not applied since the model had to be trained for a long time to stabilize the attention weights. The evaluation metrics used are accuracy, precision, recall, F1 score, confusion matrix, and ROC–AUC.

4 Results & Discussions

4.1 Ablation study

To validate the architectural design of **Spectra-MobileNet**, we conducted a stepwise decomposition of the model to isolate the contribution of each specific component (Table 1 (Appendix A)).

We began with the standard MobileNetV2 baseline at 84.50%. The first major finding was that the boost in performance from including our **Spectral (DCT) attention** (Model M3) was greater than that of conventional spatial attention, thus corroborating the hypothesis that the shrimp's hidden, high-frequency information is more indicative of freshness than its spatial features.

Moreover, the transition from CBAM to our method of **Parallel Spectral-Spatial Fusion (PSSF)** yielded 90.25% accuracy, thus showing that the parallel process of spectral and spatial features omits the bottleneck of information caused by filtering sequentially.

Finally, knowing that quality grades are ordered and not classes, the introduction of the **Ordinal Regression Head** (Model M6) delivered the definitive performance boost, leading to a peak accuracy of **92.40%**. This demonstrates that enforcing a rank-consistent logic allows the model to learn the trajectory of decay more effectively than standard classification approaches.

This section shows the results and compares the performance of Spectra-MobileNet to baseline architectures. The assessment emphasizes both numerical metrics and qualitative insights to evaluate classification accuracy, robustness, and the ability to generalize. Controlled experiments with various model configurations, dataset splits, epochs, and batch sizes illustrate the benefit of attention-driven feature amplification for prawn freshness detection.

To isolate the contribution of each module (PSSF, Ordinal Head) in an absolute manner, we trained all variants of the ablation study for a short schedule of 30 epochs only. Though our proposed Spectra-MobileNet obtains 92.40% in a limited number of phases (Table 1 Appendix A), we improved the performance to 95.62% until 120 epochs (we will discuss this in section 4.4) which suggested that the architecture gets more benefits as convergence extends and causes very high energy degradation from/to adjacent layers through depth space compared within mobile-stage or classical-style one stage-wise training, accordingly.

4.2 Training & Evaluation

The Spectra-MobileNet model was trained and tested following the experimental procedure. The custom prawn dataset was stratified split into an 80:20 training/test split to sample the *Good*, *Intermediate*, and *Bad* classes. A transfer learning approach was utilized, wherein the MobileNetV2 backbone, which was pretrained on ImageNet, remained frozen, and only the attention module and classification layers were fine-tuned.

Training was executed in a single phase utilizing the Adam optimizer with a constant learning rate of 0.0004. Data augmentation was used for the training dataset only to improve generalization and avoid overfitting. Training was performed over several epochs, but in the empirical studies, performance was consistent at high epoch numbers, indicating convergence of the attention-based model.

The observed training dynamics as illustrated in Figure 3 suggest that the Spectra-MobileNet model effectively learns significant and differentiating feature representations in a reliable manner. The close correspondence between training and validation trends indicates that the attention mechanism successfully refines features without leading to excessive overfitting. The close similarity of the training and validation curves indicates strong convergence, while the small fluctuations in validation accuracy in later epochs reflect the limited visual features that separate prawn freshness classes rather than overfitting to the training set. These results demonstrate that the attention-augmented architecture is well-conserved but not over-parameterized, establishing a good fit for learning to classify prawns by freshness.

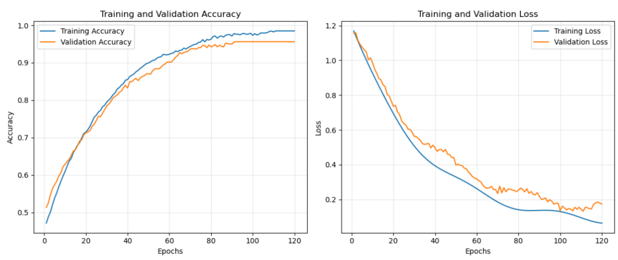

Figure 3:
Training and validation accuracy and loss curves of the proposed model.
Source: created by the authors.

4.3 Robustness Analysis of the Proposed Model

The evaluation of the performance of the proposed model was carried out in various experimental conditions. The model's performance was evaluated on different train-test splits (8:2, 7:3, and 6:4), number of epochs, batch sizes and performance metrics like precision, recall, F1-score & Accuracy. The findings from Table 3 suggest that the model exhibits consistent performance trends despite these variations, after being trained for 30 epochs at a consistent batch size of 8 with the 8:2 split providing the most reliable generalization.

Table 3: Effect of different train–test split ratios on the performance of the proposed model
Source: created by the authors.

Split Ratio	Precision	Recall	F1-score	Accuracy(%)
6:4	0.91	0.90	0.90	90.83
7:3	0.92	0.91	0.92	91.85
8:2	0.93	0.92	0.92	92.40

As a result, we keep this data split consistent and train the model at higher epochs to achieve better generalization. Analysis by epochs reveals that performance enhances with longer training durations and reaches an overall accuracy of 95.62% at 120 epochs and plateaus at 150 epochs as presented in Table 4.

Table 4: Epoch-wise performance analysis of the proposed model using an 80:20 split
Source: created by the authors.

Epoch	Precision	Recall	F1-score	Accuracy(%)
30	0.93	0.92	0.92	92.40
50	0.93	0.93	0.93	93.10
80	0.95	0.94	0.94	94.15
120	0.96	0.95	0.96	95.62
150	0.95	0.95	0.95	95.20

Therefore, the model was now trained for 120 epochs with varying batch sizes for performance evaluation. To support this, Table 5 summarizes the effect of varying batch sizes on the classification performance of the proposed model. In this experiment, the train–test split was fixed at 80:20 and the number of training epochs was fixed at 120, while only the batch size was varied to analyze its influence on learning behavior. From the table, we can infer that the model provided the best results at 8 epochs and increasing the batch sizes correspondingly affected the model's overall performance.

Table 5: Impact of batch size variation on the performance of the proposed model using 80:20 split
Source: created by the authors.

Batch Size	Precision	Recall	F1-score	Accuracy(%)
8	0.96	0.95	0.96	95.62
16	0.94	0.93	0.93	95.09
32	0.93	0.94	0.93	94.85
64	0.93	0.93	0.92	94.43

4.4 Baseline Evaluation of the Proposed Model

The Spectra-MobileNet model was assessed using a consistent training setup that included an 80:20 split for training and testing, a batch size of 8, and a total of 120 training epochs. The model's classification report can be found in Table 6. The Spectra-MobileNet model reaches an overall test accuracy of 95.62% at 120 epochs, accompanied by well-balanced precision, recall, and F1-score, which demonstrates dependable and consistent performance in multi-class classification.

Table 6: Classification Report of the proposed Model
Source: created by the authors.

Class	Precision	Recall	F1-score
Bad	0.96	0.95	0.95
Intermediate	0.94	0.95	0.94
Good	0.97	0.97	0.97
Accuracy			**0.95**

The confusion matrix shown in Table 7 reveals a significant alignment between the predicted and actual class labels throughout all freshness categories. The few misclassifications occur mainly between neighboring quality levels, indicating the gradual biological changes in prawn freshness instead of any instability in the suggested model.

Table 7: Confusion Matrix of the proposed Model on Test Set
Source: created by the authors.

Actual/Predicted	Bad	Intermediate	Good
Bad	57	3	1
Intermediate	2	57	1
Good	0	2	58

The multiclass ROC–AUC evaluation illustrated in Figure 4 further demonstrates effective separability between classes, showcasing particularly strong distinction for the *Bad* category, as well as reasonable separation for the other classes and thus keeps the AUC for all categories above 0.92.

1. *Bad* AUC: 0.931

2. *Good* AUC: 0.950

3. *Intermediate* AUC: 0.962

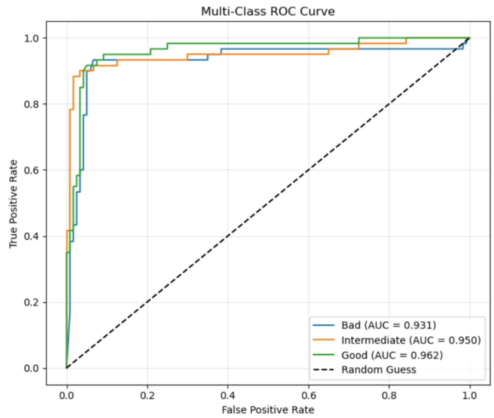

Figure 4: AUC-ROC Curve of the model
Source: created by the authors.

In summary, the findings validate that the proposed model provides stable and significant performance even with a moderate number of training epochs, confirming its effectiveness as a benchmark configuration for further robustness and comparative studies.

4.5 Model-wise Performance Comparison

Table 8 demonstrates a comparative analysis of the performance of the proposed model alongside several popular deep learning models, such as VGG16, InceptionV3, [13], EfficientNet, [14], and ResNet50, [15], using the prawn freshness dataset. Although deeper architectures show comparable precision and recall, Spectra-MobileNetV2 achieves the highest overall classification accuracy at 95.62%, while also maintaining balanced precision, recall, and F1-score. This outcome emphasizes the effectiveness of combining frequency-domain spectral characteristics with attention-guided feature refinement within a lightweight framework. Importantly, the proposed model achieves excellent accuracy with lower computational demands, underscoring its appropriateness for practical use in resource-limited aquaculture inspections and IoT-driven monitoring systems.

The proposed model has the highest accuracy of 95.62% while using the computational resources, i.e., 7.65 million parameters and 0.34 G FLOPs, and an inference speed of 11.2 ms, thus competing favorably with EfficientNet in terms of speed while beating other resource-intensive models like VGG16, thus validating its optimal trade-off in terms of accuracy and speed, which are essential in IoT applications.

This section describes the approach taken for classifying prawn freshness automatically using deep learning techniques. A dataset that was created specifically for the study included three distinct quality categories, and a MobileNetV2 model enhanced with attention mechanisms, specifically the CBAM,

Table 8: Performance comparison of different deep learning models on the prawn freshness dataset
Source: created by the authors.

Models	**Acc. (%)**	**Prec.**	**Rec.**	**F1**	**Params (M)**	**FLOPs (G)**	**Inf. Time (ms)**
VGG16 [12]	92.34	0.91	0.92	0.91	138.35	15.47	24.5
InceptionV3 [13]	93.45	0.92	0.93	0.92	23.83	5.72	18.2
EfficientNet [14]	94.48	0.94	0.93	0.93	5.30	0.39	11.4
ResNet50 [15]	95.16	0.94	0.94	0.95	25.56	4.12	15.8
Proposed	**95.62**	**0.96**	**0.95**	**0.96**	**7.65**	**0.34**	**11.2**

was introduced to enhance feature discrimination while maintaining computational efficiency. This training procedure used transfer learning, data augmentations, and optimization constraints. The testing was done through baseline model comparisons and robustness against dataset splits, epochs and batch sizes, providing confidence in the results of the subsequent section.

4.6 Test Case Prediction

To show the applicability of the Spectra-MobileNet, a qualitative inference was performed on unseen prawns images extracted from the test dataset. An sample image from this dataset was processed with the trained Spectra-MobileNet to output a predicted freshness label and confidence value.

As illustrated in Figure 5, the model retrieves by identifying weak visual cues of prawn freshness and leveraging the attention mechanism to highlight prominent areas. The confidence score represents the certainty of prediction from the ordinal regression head to prevent the hijacking of ordinal classification by regression scores across neighboring levels. This subjective assessment aligns with the objective quantitative findings and further endorses the approach's viability for real-time application and deployment across aquaculture.

Figure 5: Test Case Prediction predicted Good:91.54%, Bad:3.34%, Intermediate:5.12%
Source: created by the authors.

4.7 Discussions

Experiment results determine that the Spectra-MobileNet model is accurate and effective for prawn quality classification task. The best performance of the model on multiple evaluation indicators suggests that it can be properly utilized to assist actual seafood quality inspection procedures with reliability. To further reduce ambiguity between visually comparable classes, the patterns of misclassifications are also informative in the sense that they provide information about potential improvements in the dataset, e.g., providing additional intermediate examples or better lighting normalization. For prediction improvement, the dataset can be further augmented with assistance of generating similar synthetic images or taking more of such images based on their categories. This makes the real-life scenario possible. These results provide a sound foundation for further research that attempts to incorporate computerized decision-making and real-time image capture into aquaculture quality control systems.

Although the proposed approach shows high accuracy, it is necessary to consider environmental availability. The present data set was obtained under a controlled indoor light condition and neutral white background, decoupling the subject from the environment. In deployment to real-world scenarios, such as on-site aquaculture inspection, there may be ambient light differences (e.g., sunlight vs. shade) and background noise that could affect frequency-domain feature extraction. Moreover, image quality could be different on various smartphone sensors.Future versions of Spectra-MobileNet will have to integrate domain adaptation methods in order to make them robust in the face of these sensor-specific irregularities.

We recognize that although data augmentation increased the number of samples to 900 images, this is not effectively augmenting biological diversity but only geometric and chromatic invariance. The core dataset is still restricted to the 210 original individuals that were obtained from one source. As a result, the model might have difficulty representing the full range of biological variation (e.g., between-species spoilage patterns or widespread variation in harvest conditions) encountered across global supply chains. Next steps will focus on increasing the size of the study sample for greater intrinsic data variation.

5 Conclusion & Future Work

This study proposed a lightweight deep learning framework for automated prawn quality classification using MobileNetV2 integrated with a CBAM attention mechanism. The proposed model demonstrated robust performance on an augmented self-constructed dataset derived from limited original samples, effectively classifying prawns into *Good*, *Intermediate*, and *Bad* categories with high reliability. The incorporation of attention mechanisms enhanced the network's ability to capture subtle visual and textural cues, supporting its suitability for real-world aquaculture quality inspection applications.

Despite such great performances, some misclassifications between the *Intermediate* and *Good* classes happened owing to their visual similarity and the gradation of freshness. This limitation can, however, be addressed in future work through further studies exploring sophisticated augmentation strategies, more varied and extensive datasets, and alternative learning strategies. The framework can also be extended in the real-time deployment, IOT-based sensing solutions, and even newer architectures such as transformer-based vision models.

Future works will conduct the cross-dataset evaluation to verify the real generalization capability of the model in other environments beyond our controlled experimental setup. We will prepare a curated test set of images acquired from alternative acquisition devices (such as industrial cameras, different smartphone models) and uncontrolled illumination. Validation of the pre-trained Spectra-MobileNet on this 'wild' data is important to prove its readiness for commercial IoT deployment.

6 Acknowledgement

6.1 Contribution

All authors contributed equally to the conception, design, analysis, and writing of this work. All authors reviewed and approved the final manuscript.

6.2 Sources of funding

Maria Zdimalova was supported by the Slovak National Research grant VEGA 1/0036/23 of the Grant Agency of the Ministry of Education and the Slovak Academy of Sciences.

6.3 Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References:

[1] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the*

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 4510–4520. [Online]. Available: https://doi.org/10.1109/CVPR.2018.00474

[2] K. Dong, C. Zhou, Y. Ruan, and Y. Li, "Mobilenetv2 model for image classification," in *2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, Guangzhou, China, 2020, pp. 476–480. [Online]. Available: https://doi.org/10.1109/ITCA52113.2020.00106

[3] L. Van Nghia, T. Van Tuyen, and A. Ronzhin, "Fish image classification based on mobilenetv2 with transfer learning technique for robotic application in aquaculture," in *Interactive Collaborative Robotics*, A. Ronzhin, J. Savage, and R. Meshcheryakov, Eds. Cham, Switzerland: Springer Nature Switzerland, 2024, pp. 201–212. [Online]. Available: https://doi.org/10.1007/978-3-031-71360-6_15

[4] R. Pardeshi, R. Thinakaran, and S. Kharat, "Automated dried fish classification using mobilenetv2 and transfer learning," *International Journal of Advanced Computer Science & Applications*, vol. 16, no. 7, pp. 1–9, 2025. [Online]. Available: https://doi.org/10.14569/ijacsa.2025.0160780

[5] Z. Miao, H. Liu, G. Xian, and J. Xie, "Fish image classification based on mobilenet," in *International Conference on Computer Graphics, Artificial Intelligence, and Data Processing (ICCAID 2023)*, H. Li and H. Wu, Eds., vol. 13105. Guangzhou, China: SPIE, 2024, p. 1310511. [Online]. Available: https://doi.org/10.1117/12.3026347

[6] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Munich, Germany: Springer International Publishing, 2018, pp. 3–19. [Online]. Available: https://doi.org/10.1007/978-3-030-01234-2_1

[7] D. Lu and Q. Weng, "A survey of image classification methods and techniques for improving classification performance," *International Journal of Remote Sensing*, vol. 28, no. 5, pp. 823–870, 2007. [Online]. Available: https://doi.org/10.1080/01431160600746456

[8] J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and J. Li, "Visual attention-driven hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 10, pp. 8065–8080, 2019. [Online]. Available: https://doi.org/10.1109/TGRS.2019.2918080

[9] L. Mou and X. X. Zhu, "Learning to pay attention on spectral domain," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 110–122, 2019. [Online]. Available: https://doi.org/10.1109/TGRS.2019.2933609

[10] Z. Zhao, J. Li, Z. Luo, J. Li, and C. Chen, "Remote sensing image scene classification based on an enhanced attention module," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 11, pp. 1926–1930, 2021. [Online]. Available: https://doi.org/10.1109/LGRS.2020.3011405

[11] M. Ghosh, S. S. Roy, H. Mukherjee, S. M. Obaidullah, K. Santosh, and K. Roy, "Understanding movie poster: transfer-deep learning approach for graphic-rich text recognition," *The Visual Computer*, vol. 38, no. 5, pp. 1645–1664, 2022. [Online]. Available: https://doi.org/10.1007/s00371-021-02094-6

[12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2015, accessed: April 23, 2026. [Online]. Available: https://arxiv.org/abs/1409.1556

[13] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 2818–2826. [Online]. Available: https://doi.org/10.1109/CVPR.2016.308

[14] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2020, accessed: April 23, 2026. [Online]. Available: https://arxiv.org/abs/1905.11946

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778. [Online]. Available: https://doi.org/10.1109/CVPR.2016.90

A Appendix

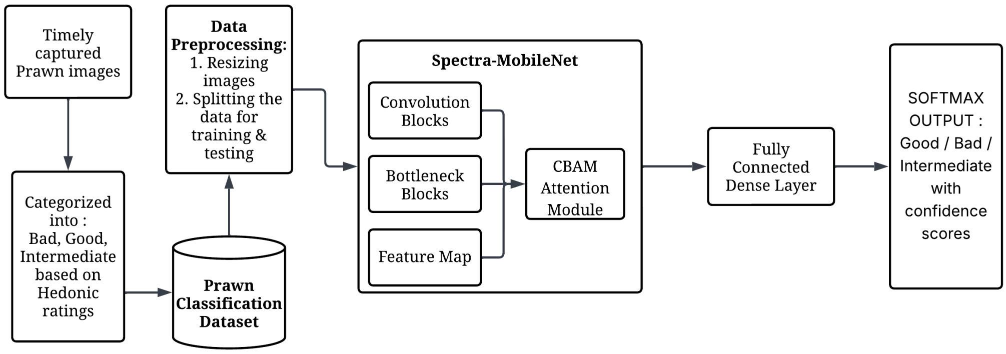

Figure 1: Workflow Diagram
Source:created by the authors.

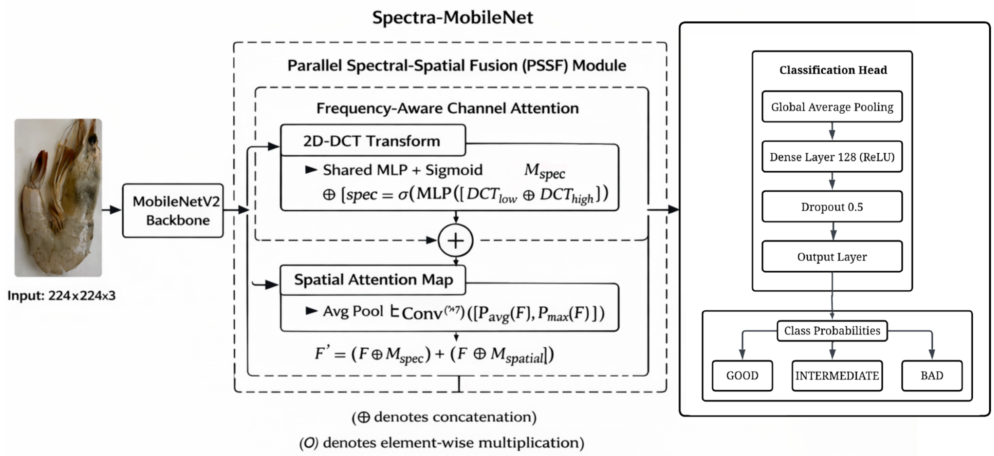

Figure 2: Proposed Architecture
Source:created by the authors.

Table 1: Ablation Study: Contribution of individual components to the Spectra-MobileNet architecture that provided an accuracy of 92.40% at 8:2 data split, batch size of 8 trained for 30 epochs.
Source:created by the authors.

Model Variant	Attention Mechanism	Fusion Strategy	Head Type	Accuracy (%)
Baseline (MobileNetV2)	None	-	Softmax	84.50
+ Spatial Attn	Spatial Only	-	Softmax	86.10
+ Spectral Attn	DCT Only	-	Softmax	87.85
+ Seq. Fusion	Spec → Spat	Sequential	Softmax	88.90
+ Par. Fusion (PSSF)	Spec + Spat	Parallel	Softmax	90.25
Spectra-MobileNet	**Spec + Spat**	**Parallel**	**Ordinal**	**92.40**

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.