MEKATRONIKA: JOURNAL OF INTELLIGENT MANUFACTURING AND MECHATRONICS
E-ISSN: 2637-0883
VOLUME 8, ISSUE 1, 2026, 12 - 27
DOI: https://doi.org/10.15282/mekatronika.v8i1.13187

RESEARCH ARTICLE

Analysis of MobileNetV2 for fish species classification in underwater images

Kiran Kumar Vijayakumar[1], Ahmad Shahrizan Abdul Ghani[1]*, Nur Najmiyah Jaafar[1], Mohd Yazid Abu[1], Yulvia Sabri[2], Mohd Fauzi Hassan[3], Fusaomi Nagata[4], Xin Jin[5], Mohd Azri Hizami Abd Rasid[1]

[1]Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, Pahang, Malaysia
[2]BMM Nagano Sdn Bhd, Pasir Gudang, Johor, Malaysia
[3]Universiti Kuala Lumpur, Malaysian Spanish Institute, Kedah, Malaysia
[4]Sanyo-Onada City University, Daigaku-dori, Sanyo-Onada, Japan
[5]School of Software and AI, Yunnan University, China

Abstract - This study investigates deep learning-based fish species classification in underwater images using the MobileNetV2 architecture combined with sophisticated data augmentation techniques, such as rotation and scaling. Robust preprocessing methods, including noise reduction, contrast enhancement, and normalisation, were employed as preprocessing steps. The system facilitates seamless dataset loading and classification procedures via a user-friendly GUI, enabling effective underwater image enhancement and accurate fish classification. Evaluation metrics such as accuracy, precision, recall, F1-score, and k-fold cross-validation were used to assess the project's performance, offering insights into the model's capabilities and generalisability across various underwater conditions. In conclusion, MobileNetV2, a novel underwater image enhancement and fish classification model, has the potential to revolutionise aquaculture. However, it faces limitations, such as a limited preprocessing pipeline owing to underwater conditions and the need for adapted data augmentation techniques. Solutions include developing specific techniques for different underwater environments, optimising annotation processes, and using transfer learning methods. Challenges in dataset availability and complexity include inadequate datasets, computational complexity, and complex annotation and classification difficulties. Alternatives include developing preprocessing methods specifically for underwater environments, using model compression techniques, improving the annotations, customising larger datasets, and leveraging transfer learning.

Article History
Received : 25 September 2025
Revised : 10 January 2026
Accepted : 9 June 2026
Published : 30 June 2026

Keywords
Fish Classification
MobileNetV2
Preprocessing methods
Image enhancement

1. Introduction

Underwater imaging poses unique challenges owing to colour distortion, poor visibility, and noise, which complicate fish species identification. Traditional methods often fail to address these issues. Advances in deep learning, particularly convolutional neural networks (CNNs), have provided powerful tools for image classification tasks. This project aims to leverage deep learning and tailored preprocessing techniques to enhance underwater images and improve the accuracy of fish species classification. Challenges in underwater imaging are exacerbated by the lack of a comprehensive framework that integrates MobileNetV2 with sophisticated data augmentation and domain-specific preprocessing. This shortfall hampers the image enhancement and species identification processes. Accurate species identification is crucial for effective fish stock management, disease control, and resource allocation in aquaculture industry. The current technological gaps, particularly the absence of comprehensive underwater image enhancement frameworks, significantly impact ecological research, marine biodiversity monitoring, and fish identification in aquaculture. Integrating advanced techniques to address underwater challenges, such as colour distortion, noise, and low visibility, is essential for more effective and sustainable fish farming and ecological monitoring. This study focused on the classification of 19 fish species in freshwater environments in Malaysia. Images were sourced from online databases, such as Kaggle, Fish4Base, and other relevant repositories, such as Google Images, ensuring a diverse representation of underwater conditions and species diversity. The objectives of this study are as follows: Development and Integration of Deep Learning with Tailored Techniques: to develop a framework that blends deep learning with customised data augmentation and preprocessing techniques for underwater image enhancement; improvement of Fish Species Identification Accuracy: Enhance the accuracy of fish species identification in underwater settings by addressing colour distortion, poor visibility, and noise; and Analysis and Evaluation of the Developed Framework/Techniques: Conduct comprehensive analysis and evaluation of the framework's performance in fish classification tasks.

2. Related Works

This review explores the intersection of computer vision, marine biology, and imaging technology, emphasising the optimisation of MobileNetV2 for underwater image enhancement and fish classification. MobileNetV2, with its inverted residuals and linear bottlenecks [1], offers computational efficiency, making it suitable for aquatic monitoring. By integrating preprocessing (e.g. photometric correction and noise filtering) and data augmentation techniques, recent studies have improved the robustness against underwater challenges, such as low visibility, spectral distortion, and species variability [2][3][4][5]. The reviewed works highlight the progress in enhancing fish species recognition accuracy while also identifying gaps in imaging hardware, acoustic modalities, and automated aquatic life assessment. These findings

*CORRESPONDING AUTHOR | Ahmad Shahrizan Abdul Ghani | ✉ shahrizan@umpsa.edu.my

underscore the importance of cross-domain transfer learning and multidisciplinary collaborations. Overall, combining MobileNetV2 with augmentation and preprocessing frameworks shows strong potential for modernising underwater image analysis and enabling reliable marine biodiversity monitoring. Deep learning has significantly advanced underwater imaging, particularly through architectures such as MobileNetV2. Convolutional Neural Networks (CNNs) address fundamental challenges, including colour distortion, resolution loss, and low contrast in aquatic environments, enabling enhanced image quality and accurate fish species classification [6]. Introduced by Sandler et al. [1], MobileNetV2 integrates inverted residuals and linear bottlenecks, offering computational efficiency and strong performance in mobile and embedded vision applications. Its ability to capture local and global contextual features makes it particularly effective for interpreting fine morphological details under variable underwater conditions [7][8]. Despite these strengths, limitations such as overfitting, dataset scarcity, and domain variability persist in these models. Recent studies [8][9][10][11][12] suggest that integrating MobileNetV2 with hybrid architectures, including U-Net, CNNs, and Pyramid Vision Transformers (PVT), can further improve the robustness of segmentation and classification tasks. In addition, UGIF-Net was proposed for underwater image enhancement using a fully guided information flow network [13].

The integration of MobileNetV2 with data augmentation and preprocessing pipelines has proven pivotal in enhancing the robustness of models across diverse imaging domains [14]. Zheng et al. [5] demonstrated improved colorectal tumour segmentation on MRI through resizing, normalisation, random flipping, zooming, and CLAHE-based contrast enhancement, thereby refining tumour morphology and boundary detection. Similarly, Sfakianakis et al. [11] optimised cardiac ultrasound segmentation using U-Net combined with ROI centralisation, downscaling, and specialised augmentation strategies to enable precise cardiac structure extraction. These studies underscore the synergy between U-Net and augmentation techniques in object extraction, intelligent inversion, and medical image segmentation. Building on this line of work, Xu et al. [2] highlighted the joint role of U-Net with preprocessing and augmentation, whereas Wang et al. [15] emphasised RGB equalisation and CIE-Lab colour space modification, thereby significantly improving image fidelity and colour accuracy. Image enhancement through an intensity-randomised approach with swarm intelligence [7] and median-based dual-image fusion [16] has been shown to produce higher-quality images, increasing visibility and improving detection and classification capabilities. Automatic background removal may be useful for increasing the degree of detection by eliminating background areas [12]. Other CNN-based methods have been proposed for image enhancement, including UW-GAN [17] combined with MobileNetV2 and VGG16 for fish classification [18] and a combination of MobileNet and Transfer Learning [19]. In summary, MobileNetV2 demonstrated high adaptability for underwater image enhancement and fish taxonomy, supporting advancements in marine ecological monitoring and biodiversity conservation. Its scalability and continuous refinement highlight its transformative role in next-generation aquatic imaging systems.

3. Materials and Methods

This project presents a technically enhanced framework for underwater fish species classification by integrating the lightweight yet high-performance MobileNetV2 architecture with advanced image preprocessing and data augmentation pipelines. The proposed methodology is specifically designed to mitigate common underwater imaging challenges, including colour attenuation, backscatter noise, low illumination, motion blur, and reduced visibility, all of which significantly impact feature extraction and classification accuracy in marine ecological monitoring and conservation studies. The system was developed using a deep learning environment in PyCharm and Google Colab, leveraging hardware acceleration via Intel Iris Xe Graphics and NVIDIA GeForce MX450 to optimise computational efficiency during model training and inference. The preprocessing framework incorporates domain-specific enhancement techniques, including denoising filters, adaptive contrast enhancement, colour correction, histogram equalisation, and image normalisation, to improve underwater image clarity and preserve discriminative visual features.

To strengthen model generalisation and reduce overfitting, the methodology applies extensive data augmentation strategies, including image rotation, scaling, flipping, translation, brightness adjustment and random cropping. The enhanced dataset was subsequently used to train and validate the MobileNetV2-based classification model using rigorous evaluation procedures, including cross-validation, accuracy assessment, confusion matrix analysis, and performance metrics such as precision, recall, and F1-score. In addition, the integration of an intuitive graphical user interface (GUI) improves system accessibility and operational efficiency by enabling streamlined workflows for image uploading, preprocessing, prediction, and visualisation. Collectively, the proposed framework establishes a scalable, computationally efficient, and highly accurate solution for underwater fish species identification, positioning it as a strong reference model and potential benchmark for future research in underwater computer vision, marine biodiversity monitoring and intelligent aquatic ecosystem management.

3.1 Dataset Acquisition

A systematic and domain-focused data acquisition strategy was employed to collect underwater imagery representing 19 freshwater fish species native to aquatic ecosystems in Pahang, Malaysia (Figure 1). The dataset was compiled from reputable open-access repositories, including Kaggle and Fish4Knowledge, ensuring a broad species representation and variability in underwater imaging conditions. The collected images encompass diverse environmental factors, such as fluctuating illumination, turbidity levels, occlusion, motion blur, background complexity, and varying camera perspectives, thereby enhancing the robustness and generalisation capability of the classification model. This heterogeneous dataset construction supports the development of a more resilient deep learning framework capable of accurately identifying fish species under real-world underwater conditions that are commonly encountered in freshwater habitats.

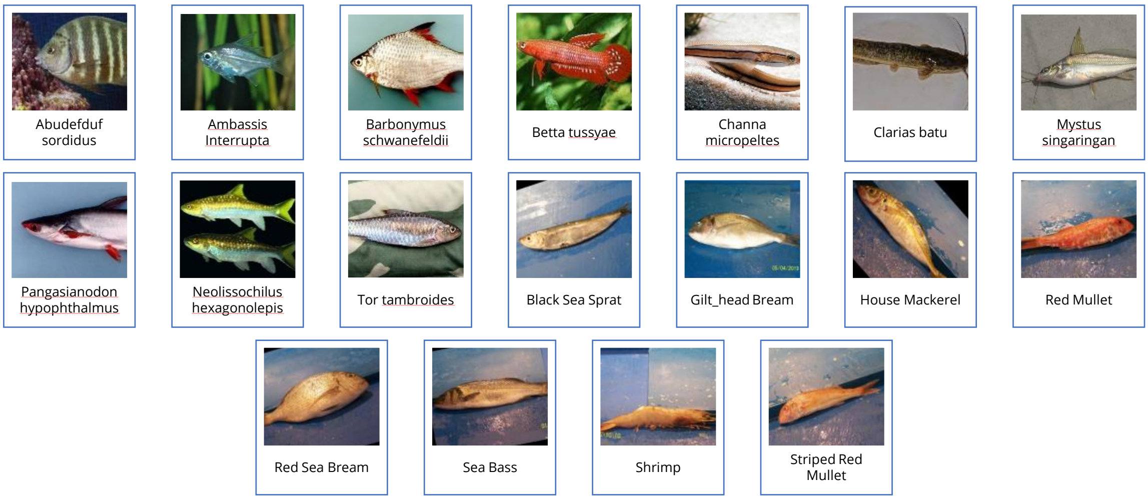

Figure 1. The dataset of 19 fish species images from the *FishBase* online database, showcasing the diverse aquatic life in Pahang, Malaysia, illustrates the local ecosystem

3.2 Preprocessing and Data Augmentation

As shown in Figure 2, a comprehensive image preprocessing pipeline was implemented to improve the underwater image quality and optimise feature extraction for deep learning-based fish species classification. Noise reduction was performed using variance-based filtering techniques to suppress underwater artefacts, sensor noise, and backscatter interference, while preserving critical edge and texture information. Contrast enhancement was achieved through integrated image smoothing approaches, as proposed by Kamarul Baharin et al. [20], combined with intensity-randomised optimisation techniques utilising swarm intelligence algorithms introduced by Mohd Azmi et al. [7]. These enhancement methods significantly improve image sharpness, local contrast, colour consistency, and object visibility in low-illumination underwater environments. In addition, image normalisation based on the dataset mean and standard deviation was applied to standardise the pixel intensity distributions, reduce the illumination variance, and stabilise the convergence behaviour of the deep neural network during training. Collectively, these preprocessing operations enhanced the discriminative visual characteristics of fish morphology, thereby improving the classification accuracy and model robustness. To further strengthen the model generalisation and mitigate overfitting, extensive data augmentation techniques were incorporated into the training pipeline. These included geometric transformations, such as rotation, scaling, flipping, translation, and random cropping, which artificially increased dataset diversity and enabled the model to learn invariant spatial and orientation features across varying underwater imaging conditions.

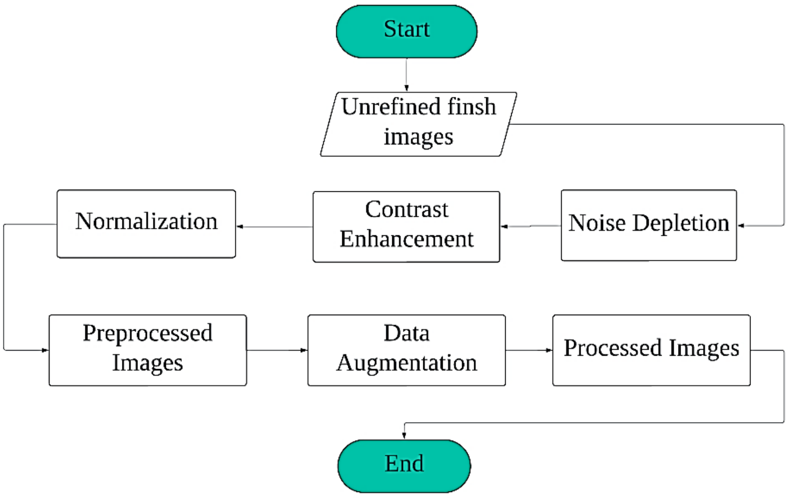

Figure 2. The flowchart shows a structured Python-based preprocessing pipeline that includes noise reduction, contrast enhancement, colour correction, normalisation, and data augmentation techniques

3.3 MobileNetV2 Architecture and Configuration

MobileNetV2 was selected as the primary deep learning architecture because of its lightweight design, computational efficiency, and strong performance in image classification tasks in resource-constrained environments. The architecture incorporates depth-wise separable convolutions and inverted residual bottleneck layers, significantly reducing the number of trainable parameters and computational complexity while maintaining a high feature extraction capability and classification accuracy. This efficient architectural design enables faster inference, lower memory consumption, and improved deployment feasibility in systems with limited hardware resources. The implemented MobileNetV2 framework served as the core backbone of the underwater fish species-classification system. Its configuration was carefully optimised to enhance feature representation from challenging underwater imagery while minimising the computational overhead. The training pipeline used a comprehensive and heterogeneous dataset processed through advanced preprocessing techniques, including variance-based noise reduction, adaptive contrast enhancement, colour normalisation, and illumination correction, to improve image clarity and preserve biologically relevant morphological features. In addition, extensive data augmentation operations, such as rotation, scaling, translation, flipping, and random cropping, were applied to improve dataset diversity, strengthen model generalisation, and reduce overfitting during training.

To maximise the model performance, multiple training configurations and hyperparameters were systematically optimised, including learning rate scheduling, batch size selection, optimiser tuning, dropout regularisation, epoch configuration, and activation function adjustments. These optimisations contributed to improved convergence stability, reduced training loss, and enhanced predictive accuracy under varying underwater environmental conditions. To rigorously assess the effectiveness of MobileNetV2, a comparative performance evaluation was conducted against several state-of-the-art baseline convolutional neural network architectures, including ResNet50, InceptionV3, VGG16, and EfficientNetB0. The comparative analysis examined key performance metrics, such as classification accuracy, precision, recall, F1-score, inference speed, model complexity, and computational efficiency. The experimental results demonstrated that MobileNetV2 achieved superior performance in balancing accuracy and computational cost, particularly in handling the visual complexities of underwater imagery characterised by noise, low contrast, colour distortion, and environmental variability. Its ability to deliver high classification performance with reduced computational demand reinforces its suitability as the optimal architecture for real-time underwater fish species identification and intelligent aquatic ecosystem monitoring applications. The proposed integrated image preprocessing with the MobileNetV2 architecture is shown in Figures 3 and 4.

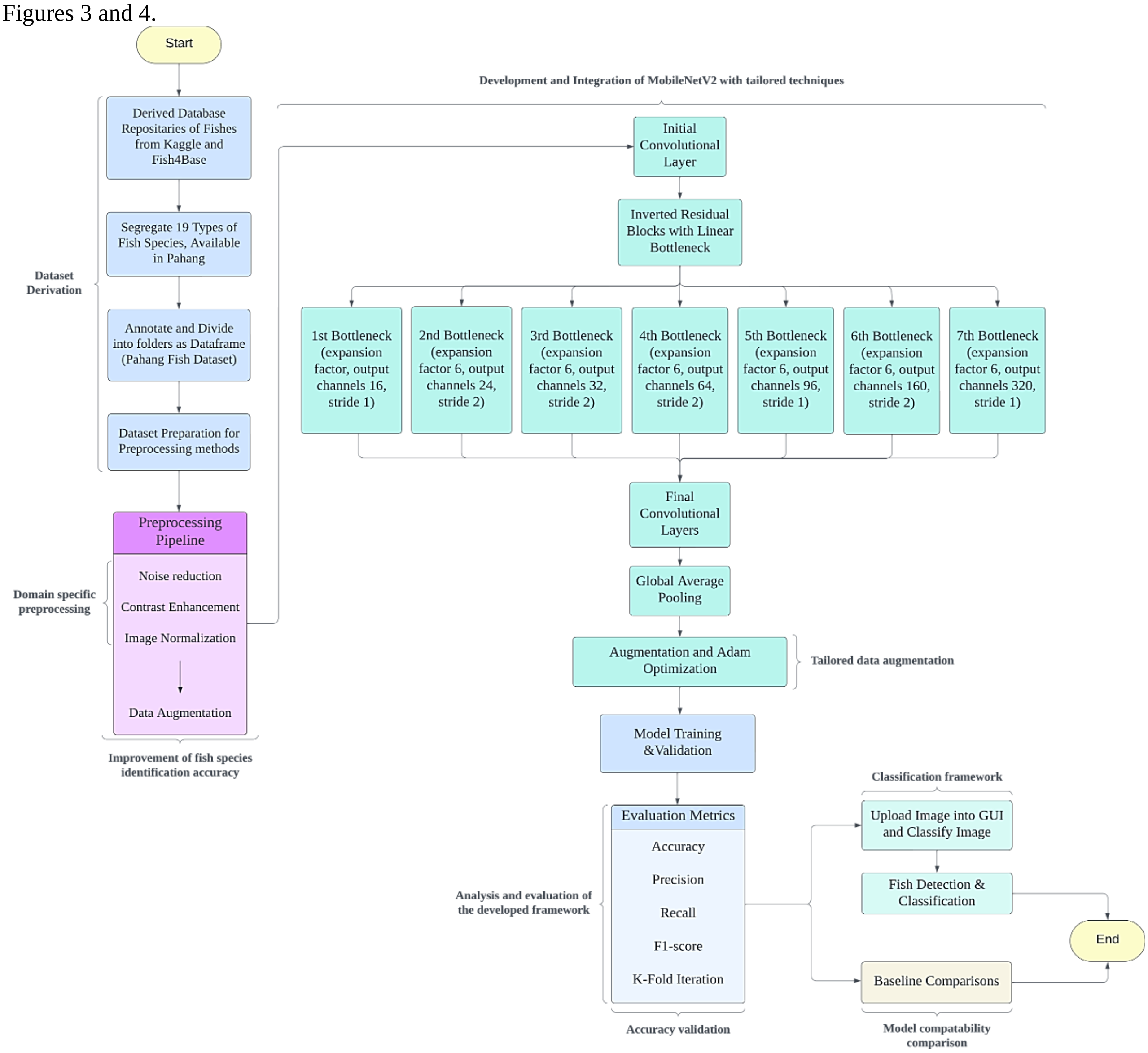

Figure 3. The MobileNetV2 architecture, combined with specialised database preprocessing techniques, data augmentation, and comprehensive validation metrics, enhances underwater images for accurate fish classification

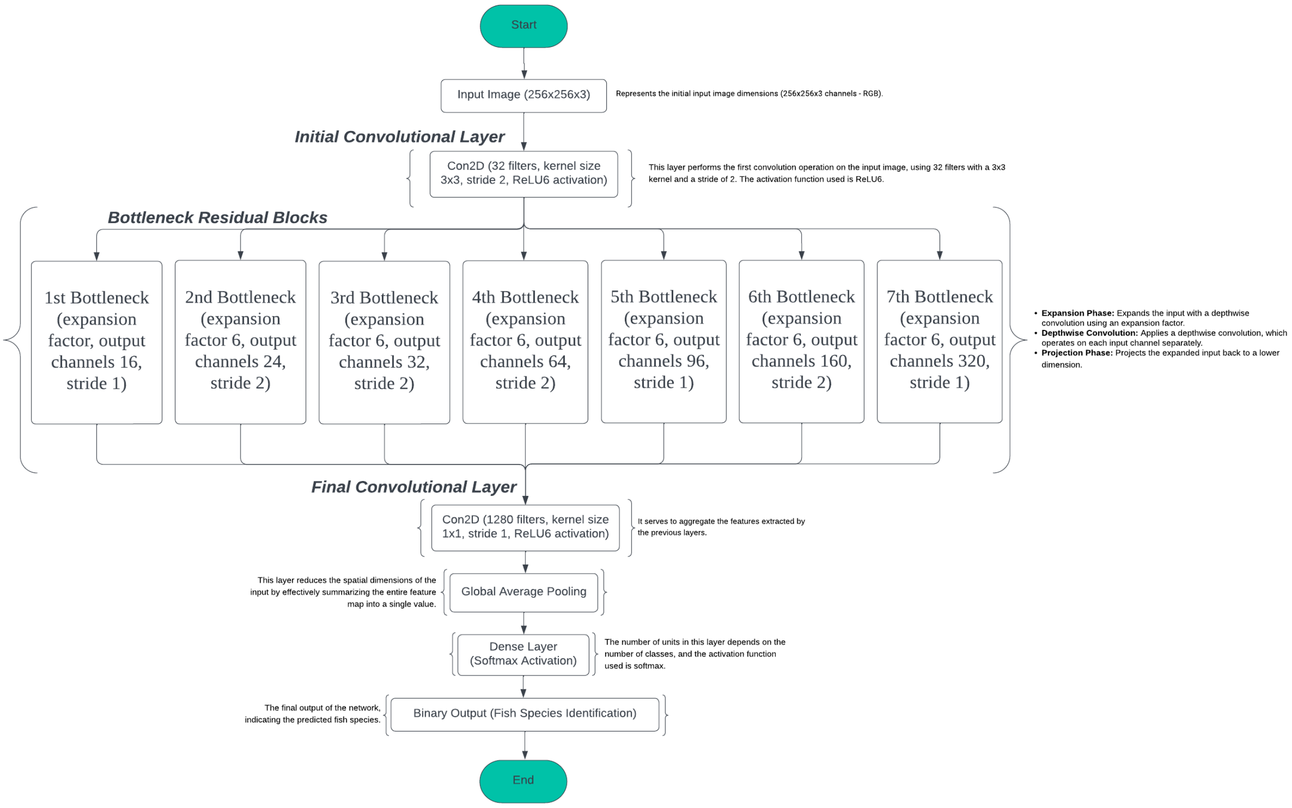

Figure 4. The architecture and proposed modifications to the MobileNetV2 structure are detailed in a flowchart with clearly defined specifications and explanations

4. Results and Discussion

4.1 Evaluation Metrics and Graphical User Interface

The project's performance was thoroughly evaluated using various metrics, including accuracy, precision, recall, F1-score, and k-fold cross-validation. These metrics offered a detailed assessment of the model's capabilities and its ability to generalise across various underwater conditions. K-fold cross-validation was employed to ensure the robustness and generalisability of the model. This technique involves partitioning the dataset into k subsets, training the model on k-5 subsets, and validating it on the remaining subset of data. This process was repeated five times, providing a comprehensive evaluation of the model's performance and ensuring that it performed well across different data subsets. In addition to the quantitative metrics, a qualitative evaluation was conducted through visual inspection of the preprocessing and augmentation results. This step ensured that the applied techniques effectively enhanced the underwater images, thereby improving feature extraction for classification. Visual inspection confirmed that preprocessing methods, such as noise reduction, contrast enhancement, and normalisation, along with augmentation techniques, such as rotation and scaling, significantly improved the clarity and detail of the images. This comprehensive approach, which combines both quantitative and qualitative evaluations, ensured a robust assessment of the model's performance in underwater fish species classification.

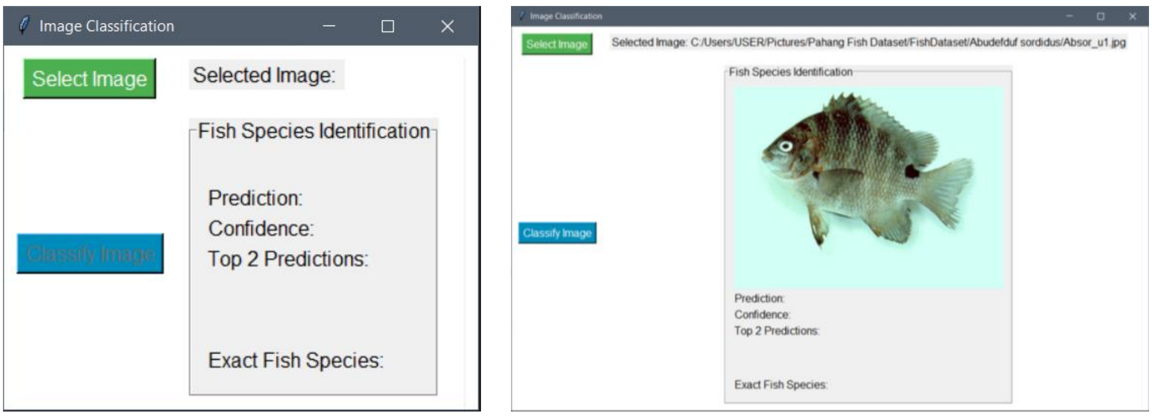

Figure 5. The initial design and first/second phase of GUI interpretation for fish classification

The GUI for this project, which was developed using Python and the Tkinter library, provides a user-friendly interface for image classification tasks. Users can Select Image from their file system via the "Select Image" button and classify it using the "Classify Image" button. The selected image is displayed within the GUI and resized to fit the designated area. The script uses a deep learning model loaded in.h5 format with TensorFlow to classify the image after preprocessing steps such as noise reduction and normalisation. The GUI displays the predicted fish species and their confidence levels, showcasing the top two predictions and their respective probabilities. In addition, a pie chart generated using Matplotlib visually represents the classification results, providing clear insights into the model's predictions. Figure 5 illustrates the initial display of the GUI and the subsequent interaction stages, which enhance the user experience through intuitive design and functionality. The system underwent several refinements based on the preliminary results and feedback. Adjustments to the preprocessing methods, augmentation techniques, and model parameters were made to further enhance the classification accuracy and robustness.

4.2 Utilisation of k-fold Iteration

In evaluating the MobileNetV2 model for underwater fish species classification, this study demonstrated a strong predictive capability based on a single train–test split evaluation strategy. The model achieved a test accuracy of 88.89%, precision of 91.96%, recall of 88.03%, and F1-score of 88.39%, indicating effective feature extraction and discriminative learning from raw underwater image datasets. These performance metrics suggest that the lightweight deep convolutional architecture can capture complex visual patterns, including variations in fish morphology, illumination conditions, water turbidity, and background noise, which are commonly encountered in underwater environments. Nevertheless, the evaluation methodology lacks k-fold cross-validation, which limits the statistical reliability and generalisability of the results. Relying solely on a single data partition increases the susceptibility to sampling bias, dataset imbalance, and variance caused by random train–test selection. Incorporating k-fold cross-validation would provide a more comprehensive assessment of the model robustness, stability, and reproducibility across multiple data distributions, thereby offering stronger evidence of the model's real-world deployment capability in underwater aquatic monitoring and intelligent marine biodiversity applications.

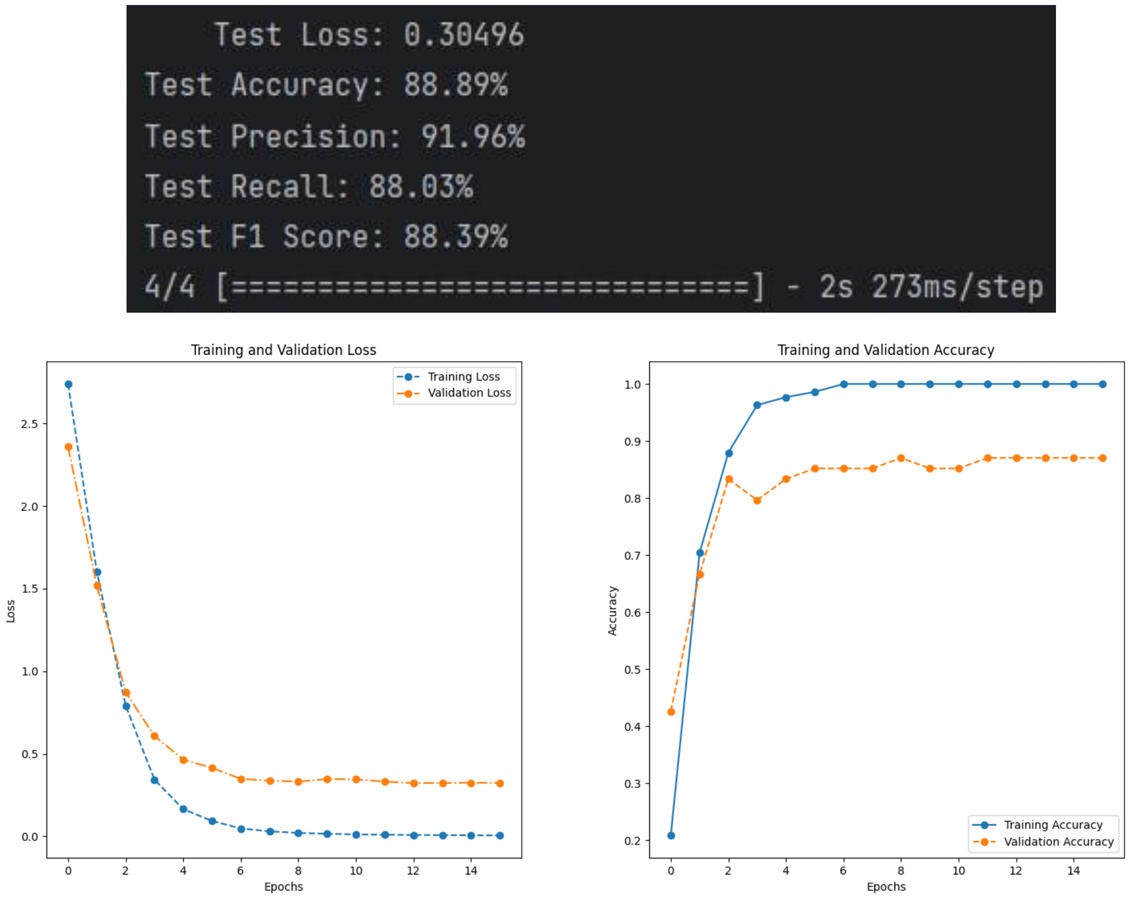

Figure 6. The derived results of evaluation metrics and loss/accuracy graph on non-utilisation of K-fold iteration

The reported performance metrics indicate high classification accuracy with a well-balanced precision–recall trade-off, demonstrating the effectiveness of the model in distinguishing underwater fish species across multiple classes. Furthermore, the training and validation curves presented in Figure 6 illustrate a stable learning behaviour and satisfactory model convergence throughout the training process. However, intermittent fluctuations in the validation loss suggest mild overfitting, potentially caused by limited dataset diversity, class imbalance, or sensitivity to underwater image variations, such as illumination inconsistency, motion blur, and water turbidity. To improve the robustness and statistical reliability of the evaluation, future studies should incorporate k-fold cross-validation, enabling the model to be assessed across multiple data partitions and reducing the risk of sampling bias and over-optimistic performance estimation. In evaluating the MobileNetV2 model for underwater fish species classification, this study demonstrated strong predictive performance based on a single train–test split evaluation framework. The model achieved a test accuracy of 88.89%, precision of 91.96%, recall of 88.03%, and F1-score of 88.39%, reflecting its capability to perform reliable feature extraction and discriminative learning directly from raw underwater image data. These results indicate that the lightweight deep convolutional neural network architecture effectively captures complex visual characteristics, including fish morphology, texture patterns, scale variations, and environmental distortions commonly encountered in underwater imaging systems. Nevertheless, the absence of k-fold cross-validation limits the comprehensiveness of the evaluation because the reported metrics may be influenced by dataset partitioning and sampling variance. A more rigorous cross-validation strategy would provide deeper insights into the model's consistency, robustness, and generalisation capability across diverse data distributions and unseen underwater scenarios.

In this study, the MobileNetV2 model was further evaluated using a rigorous 5-fold cross-validation framework to enhance the reliability and reproducibility of underwater fish species classification results. By systematically partitioning the dataset into multiple training and validation subsets, the approach minimised evaluation bias and enabled a comprehensive assessment of the model's generalisation performance. During the initial training epochs, the model exhibited relatively high loss values and low classification accuracy, indicating early stage adaptation of the network weights and feature learning processes. However, as training progressed, the model demonstrated substantial performance improvements, characterised by consistent reductions in loss and significant increases in accuracy, precision, recall, and F1-score. The progressive convergence behaviour confirmed the effectiveness of the optimisation process and the network's ability to learn discriminative underwater visual features. Overall, the findings demonstrate that the model achieved a balanced classification performance by maximising true-positive detections while minimising false-positive and false-negative detections, thereby supporting its suitability for intelligent marine monitoring, automated fisheries management, and underwater biodiversity assessment applications. To optimise performance, the training configuration used a learning rate of 0.0001 and the Adam optimiser with categorical cross-entropy loss. The dataset was diversified using augmentation techniques, such as rotation, shift, shearing, zooming, and horizontal flipping. The model was run for 100 epochs on a dataset of 224 × 224 resized images. Key findings include the model achieving impressive results on the test dataset with a test loss of 0.04118, test accuracy of 98.71%, precision of 98.96%, recall of 98.45%, and an F1 score of 98.73%. These metrics underscore the robustness and effectiveness of the proposed method in handling the complexities of underwater image segmentation tasks. The use of 5-fold cross-validation ensured consistent performance across different data subsets, validating the model's reliability and enhancing confidence in its generalisation capabilities.

4.3 Non-utilisation of Preprocessing Methods

When evaluating the performance of the MobileNetV2 model specifically for the classification of Abudefduf sordidus under non-pre-processed image conditions, several significant observations and technical challenges were identified.

- **Prediction Accuracy**: The classification accuracy for Abudefduf sordidus was moderate, reflecting the model's ability to extract discriminative visual features directly from raw underwater image inputs. However, the absence of preprocessing operations, such as image normalisation, contrast enhancement, denoising, and colour correction, reduced the consistency of feature representation across samples. Consequently, the predictive accuracy of the model became highly dependent on the image quality, underwater visibility conditions, and object prominence within the frame. As illustrated in Figure 7, similar fish species achieved classification accuracies of approximately 79.9%, indicating that the model was only moderately successful in differentiating species with closely related morphological characteristics.

- **Prediction Confidence Levels**: The confidence scores generated by the model provide important insights into the certainty and reliability of its classification decisions. For Abudefduf sordidus, comparatively lower confidence levels suggest increased uncertainty during inference, likely caused by overlapping visual features, inconsistent illumination, motion blur and underwater scattering effects. These factors reduce the separability of feature embeddings within the latent feature space, making it more difficult for the neural network to establish high-confidence decision boundaries between visually similar species.

- **Recognition and feature extraction challenges**: Recognition of Abudefduf sordidus presents several challenges when preprocessing techniques are omitted. Underwater environments inherently introduce visual distortions, such as low contrast, colour attenuation, suspended particle noise, non-uniform lighting, and complex background textures. In addition, subtle morphological similarities among reef fish species complicate feature extraction. These environmental and biological factors reduce the ability of convolutional layers to learn robust, invariant feature representations, increasing the likelihood of inter-class misclassification.

- **Potential for Performance Enhancement**: The experimental findings highlight substantial opportunities for improving model robustness and classification reliability by integrating advanced image preprocessing techniques. Implementing methods such as adaptive histogram equalisation, Gaussian filtering, underwater image enhancement, colour normalisation, and noise suppression can significantly improve feature visibility and reduce input variability. Enhanced preprocessing would allow the model to learn more stable spatial and texture-based features, ultimately improving the classification accuracy, prediction confidence, and generalisation performance for Abudefduf sordidus across diverse underwater imaging conditions.

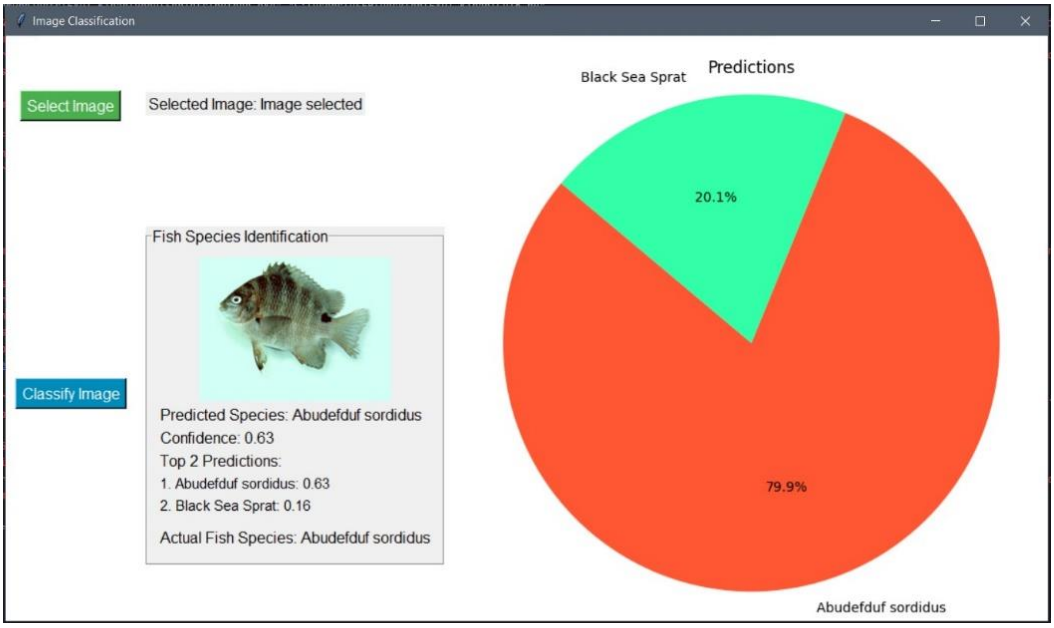

Figure 7. Classification of *Abudefduf sordidus* shows a higher percentage of prediction, at 79.9% but is still correlated with 20.1% of *Black Sea Sprat* species

4.4 Utilisation of Preprocessing Methods

When focusing specifically on Abudefduf sordidus and the integration of preprocessing techniques such as noise reduction and image normalisation, the MobileNetV2 model demonstrates highly effective and robust classification performance in underwater fish species recognition tasks.

- **Classification Accuracy and Prediction Confidence**: As illustrated in Figure 8, Abudefduf sordidus achieved a classification accuracy of 100% with a prediction confidence score of 1.00, indicating absolute certainty in the model's inference process. This result demonstrates the effectiveness of the deep convolutional architecture in learning highly discriminative spatial and texture-based features when high-quality preprocessed input data are provided. The exceptionally high confidence value further suggests that the extracted feature representations exhibit strong separability within the learned feature space, thereby enabling precise inter-class discrimination.

- **Impact of Preprocessing Techniques**: Incorporating preprocessing operations significantly enhanced the quality and consistency of the input images prior to feature extraction. Techniques such as image normalisation, denoising, contrast enhancement, and illumination correction reduce unwanted variability caused by underwater environmental distortions. These preprocessing methods improved the statistical distribution of pixel intensities, minimised background interference, and enhanced edge and texture visibility, thereby allowing the model to more effectively capture the subtle morphological characteristics unique to Abudefduf sordidus. Consequently, the preprocessing pipeline contributed directly to improved classification accuracy and inference stability.

- **Model Reliability and Classification Consistency**: The consistently high accuracy and confidence levels observed in Figure 9 demonstrate the reliability, repeatability, and robustness of the model in identifying Abudefduf sordidus across multiple test instances. This consistency indicates that the learned feature representations remain stable despite minor variations in the pose, orientation, scale, and underwater environmental conditions. Such reliability is essential for automated marine monitoring systems, intelligent fishery management, and biodiversity assessment applications, where accurate species-level identification is critical.

- **Real-world Deployment and Generalisation Capability**: In real-world underwater environments, image quality is often degraded by factors such as low illumination, water turbidity, motion blur, suspended particles, and colour attenuation. The integration of preprocessing techniques enables the MobileNetV2 model to maintain a robust classification performance under these challenging conditions by reducing noise sensitivity and improving feature consistency. As a result, the model exhibited a stronger generalisation capability across diverse underwater datasets and operational environments, supporting its suitability for practical deployment in autonomous underwater vision systems and real-time aquatic species recognition applications.

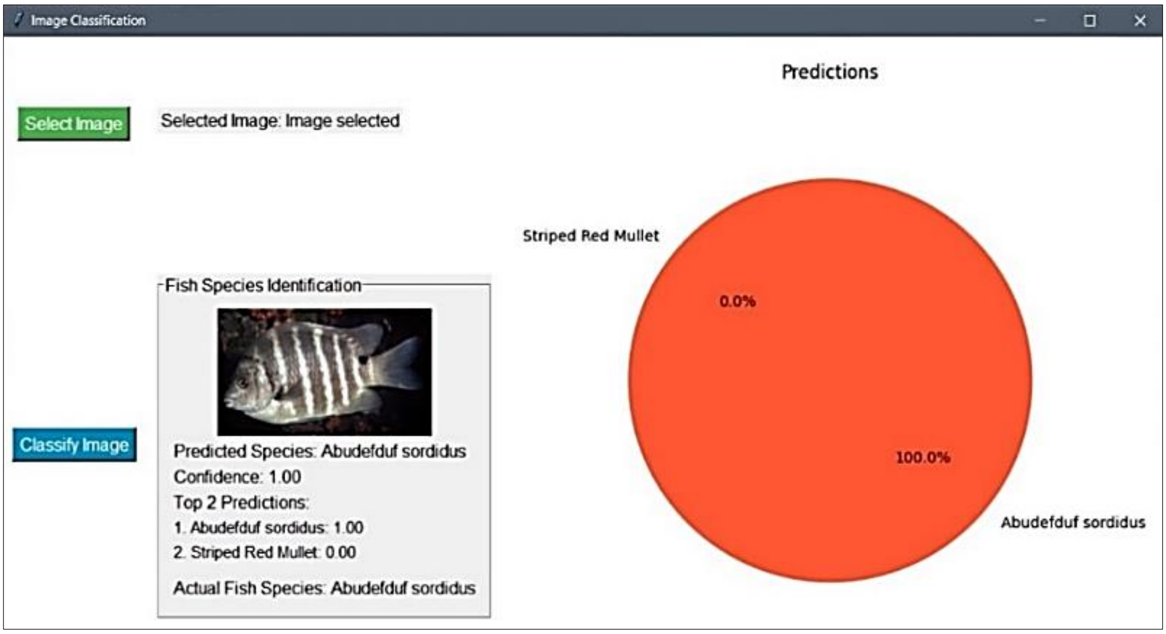

Figure 8. Classification of *Abudefduf sordidus* shows a higher prediction accuracy, with 100.0% success

Table. 1 Comprehensive comparison of key performance metrics for MobileNetV2, ResNet50, InceptionV3, VGG16, and EfficientNetB0 on the task of underwater fish classification

Models/Metrics	Accuracy	Precision	Recall	F1-score	Loss
MobileNetV2	97.93	98.19	97.93	97.97	0.05354
ResNet50	57.11	92.66	26.10	56.37	1.43117
InceptionV3	91.99	95.53	88.37	92.02	0.26820
VGG16	97.42	97.65	96.64	97.48	0.09074
EfficientNetB0	15.25	88.89	2.07	9.04	2.76139

4.5 Comparative Analysis of Fish Classification Models

A comprehensive comparative analysis was conducted to evaluate the performance of multiple deep learning architectures for underwater fish species classification and image segmentation. The experimental results demonstrated that the MobileNetV2 model consistently outperformed the baseline models across key evaluation metrics, including accuracy, precision, recall, F1-score, and classification loss. This superior performance highlights MobileNetV2's effectiveness in extracting robust and discriminative visual features from complex underwater image datasets while maintaining computational efficiency. Among the evaluated architectures, MobileNetV2 achieved the best balance between classification accuracy and model efficiency, making it highly suitable for underwater-image analysis applications. Its depth-wise separable convolution mechanism reduces the computational complexity and parameter count while preserving the strong feature learning capability. This lightweight architecture is particularly advantageous for real-time underwater monitoring systems, embedded devices, and edge-based marine vision applications, where processing power and memory resources are limited. Other evaluated deep convolutional neural network models, including ResNet50, InceptionV3, VGG16, and EfficientNetB0, also demonstrated competitive performance in underwater fish species classification. These models exhibited strong feature extraction capabilities and achieved relatively high classification metrics across several species. However, ResNet50 exhibited limitations in consistently identifying true positive samples, resulting in lower recall, moderate F1-scores, and higher classification loss values. This behaviour suggests potential difficulties in handling class imbalances, subtle interspecies visual similarities, and underwater image distortions. Overall, as shown in Table 1 and Figure 9, the findings indicate that MobileNetV2 provides the most optimal trade-off between the predictive performance, computational efficiency, and generalisation capability for underwater fish species classification. Its ability to maintain high precision and recall while minimising loss demonstrates its suitability for deployment in intelligent marine biodiversity monitoring, automated fishery management systems, and real-time underwater computer vision applications.

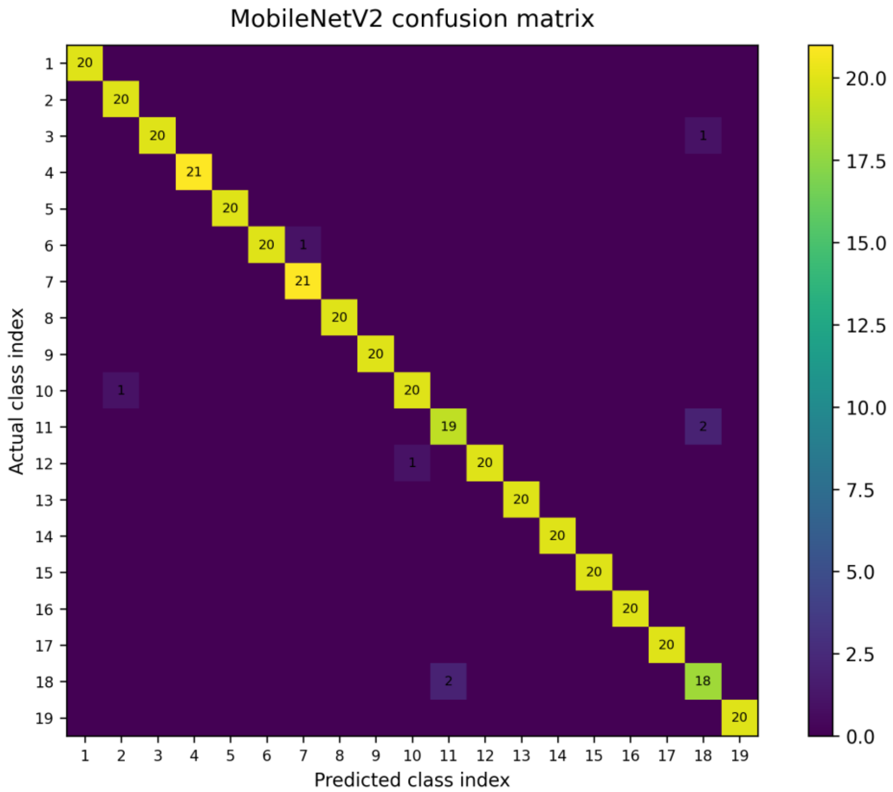

Figure 9(a). MobileNetV2's confusion matrix. The class indices correspond to the dataset order listed in Table 2

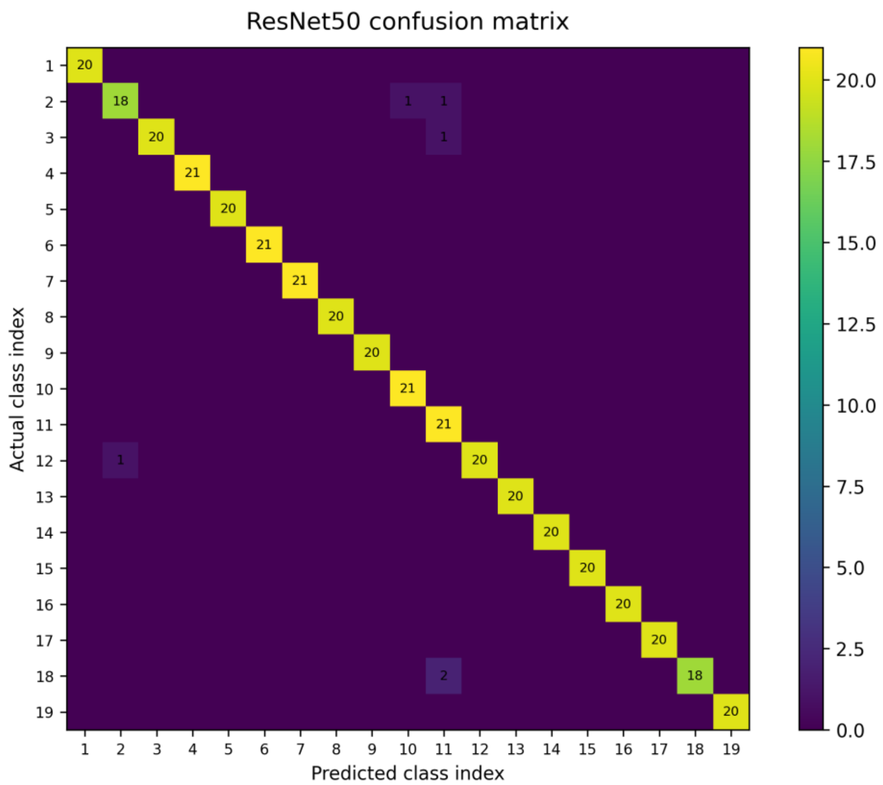

Figure 9(b). ResNet50 confusion matrix. The class indices correspond to the dataset order listed in Table 2

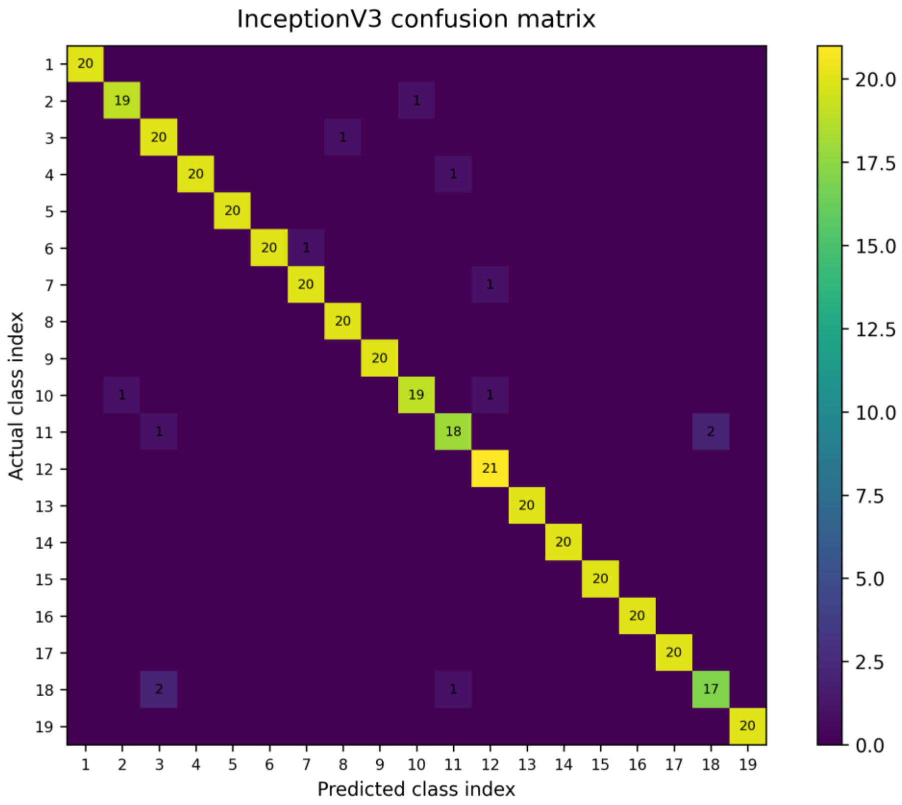

Figure 9(c). InceptionV3 confusion matrix. The class indices correspond to the dataset order listed in Table 2

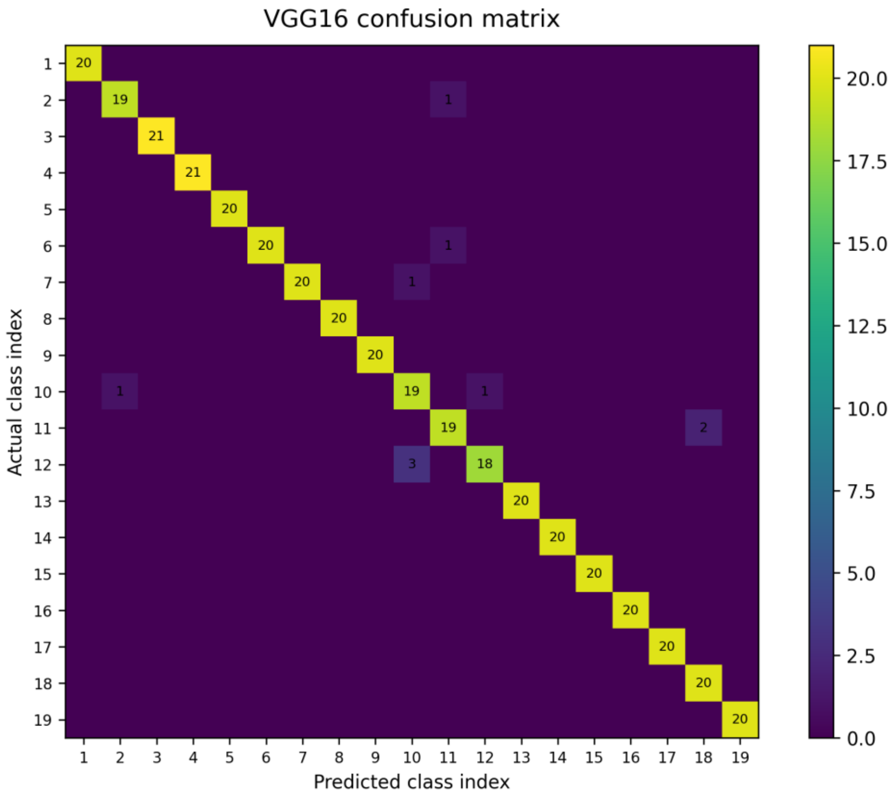

Figure 9(d). VGG16 confusion matrix. The class indices correspond to the dataset order listed in Table 2

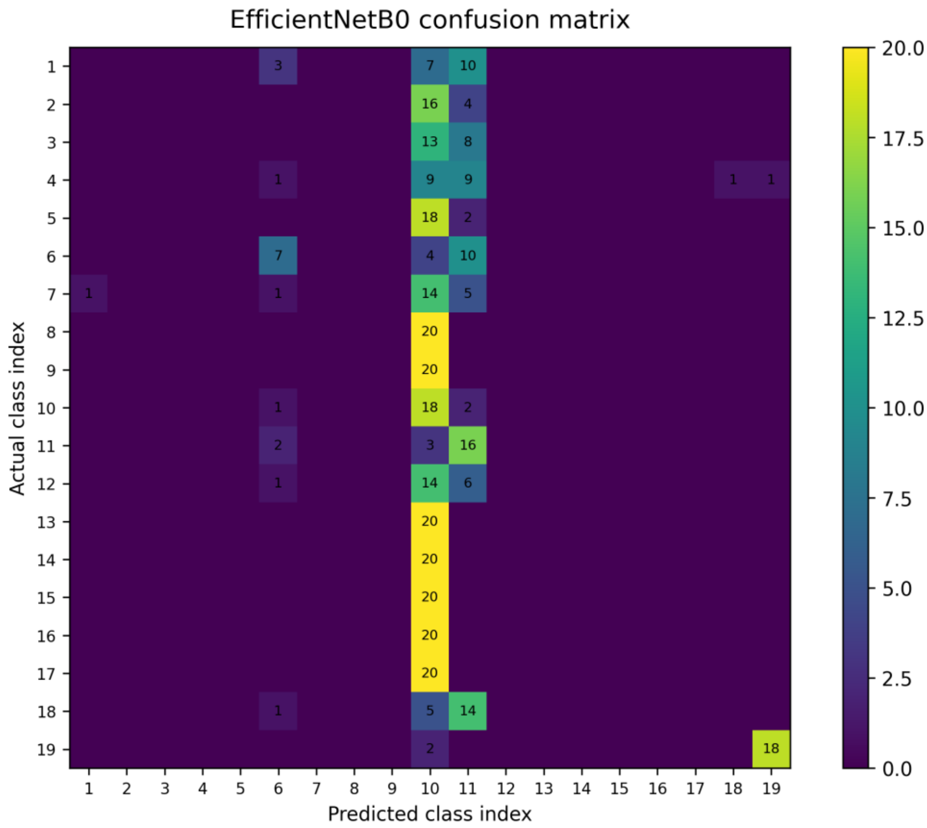

Figure 9(e). EfficientNetB0 confusion matrix. The class indices correspond to the dataset order listed in Table 2

Table 2. Class indices used in the confusion matrices

Class index	Species/class name
1	Abudefduf sordidus
2	Ambassis interrupta
3	Barbonymus schwanefeldii
4	Betta tussyae
5	Black Sea Sprat
6	Channa micropeltes
7	Clarias batu
8	Gilt-Head Bream
9	Horse Mackerel
10	Mystus singaringan
11	Neolissochilus hexagonolepis
12	Pangasianodon hypophthalmus
13	Red Mullet
14	Red Sea Bream
15	Sea Bass
16	Shrimp
17	Striped Red Mullet
18	Tor tambroides
19	Trout

As shown in Figure 10, integrating MobileNetV2 with advanced preprocessing and data augmentation techniques resulted in a robust and accurate fish classification system. The user-friendly GUI facilitated seamless dataset management and classification, thereby contributing to the success of the project. A comprehensive performance evaluation was conducted to compare five state-of-the-art deep learning architectures—MobileNetV2, ResNet50, InceptionV3, VGG16, and EfficientNetB0—for underwater fish species classification using image datasets. The models were assessed using multiple evaluation metrics, including classification accuracy, precision, recall, F1-score, and loss, to determine their effectiveness, robustness, and generalisation capability in complex underwater imaging environments.

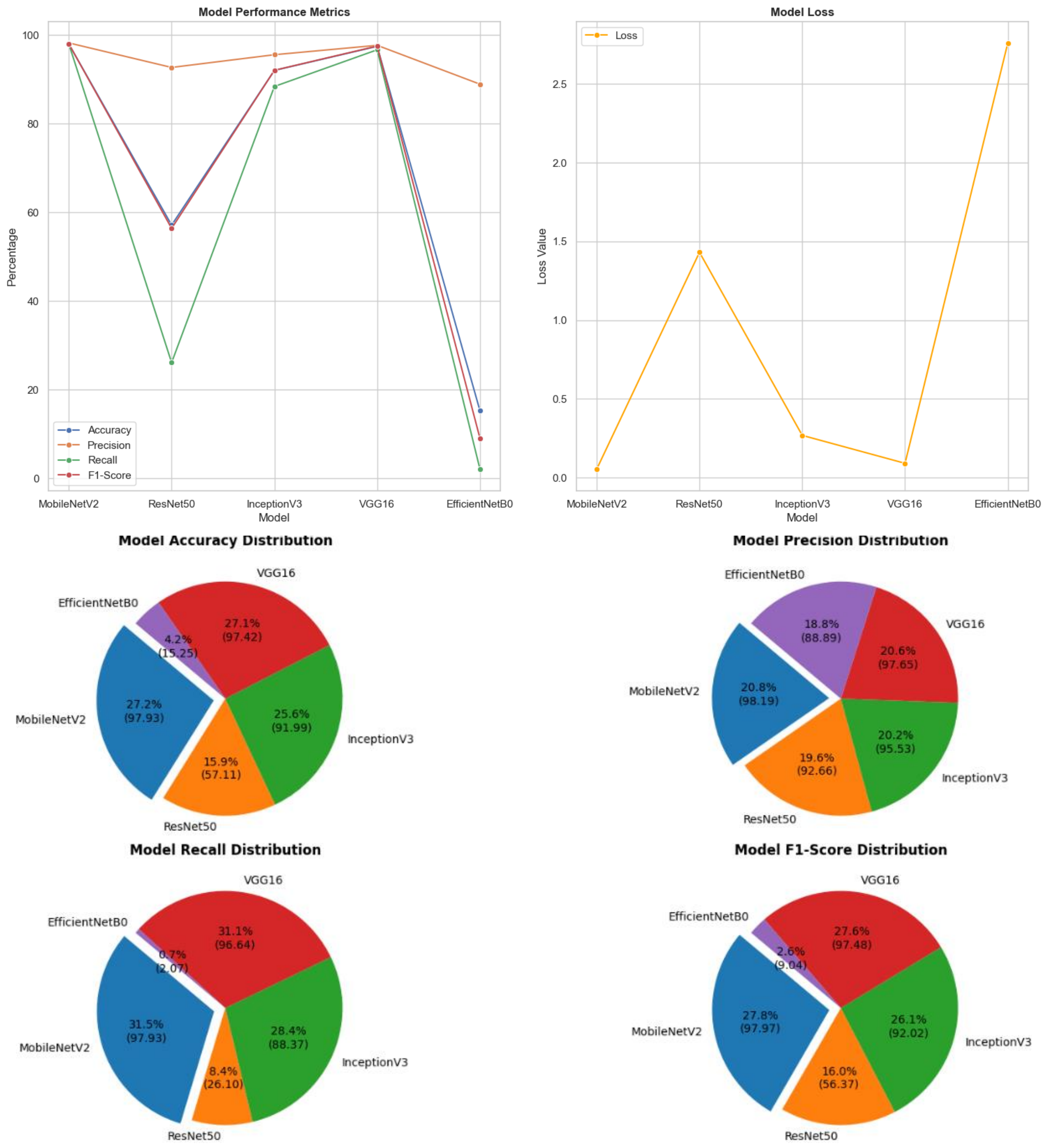

Figure 10: The evaluation metrics of all involved models for accuracy, precision, recall, F1-score and K-fold iteration with utilization of preprocessing methods

i. **MobileNetV2 Performance Analysis**
MobileNetV2 emerged as the best-performing model, achieving the highest classification accuracy and precision while maintaining the lowest loss of 0.05354. The model consistently demonstrated superior performance across all evaluation metrics, indicating excellent feature extraction and strong discriminative learning performances. Its depth-wise separable convolution architecture significantly reduces computational complexity and parameter size without compromising classification accuracy, making it highly suitable for real-time underwater fish recognition systems, embedded vision platforms, and resource-constrained edge computing applications. The balanced precision–recall performance further indicates a strong capability to minimise both false-positive and false-negative classifications.

ii. ResNet50 Performance Analysis
ResNet50 achieved relatively high precision, suggesting a strong confidence in correctly predicting the positive classes. However, the model exhibited comparatively lower recall and F1-score values, indicating reduced sensitivity in detecting true positive samples. This suggests that although the model performs well in confirming predictions, it struggles to consistently identify all relevant fish species. The higher loss values and moderate overall performance may be attributed to challenges such as underwater image distortion, class imbalance, and interspecies morphological similarity, which can negatively affect residual feature learning and classification stability.

iii. InceptionV3 Performance Analysis
InceptionV3 demonstrated robust and stable performance across all evaluation metrics, achieving high accuracy, precision, recall, and F1-score values. Its multiscale convolutional architecture effectively captures both local and global visual features, enabling improved classification of underwater fish images with varying scales and orientations. Although its overall performance was slightly lower than that of MobileNetV2, the model proved highly effective for underwater species recognition tasks requiring complex feature representation.

iv. VGG16 Performance Analysis
VGG16 achieved a competitive classification performance with high accuracy, precision, recall, and F1-score values. The model's deep-layered structure enabled effective hierarchical feature extraction, contributing to reliable fish species recognition performance. However, compared with MobileNetV2, VGG16 requires significantly higher computational resources, memory allocation, and training time because of its large number of parameters. This increased architectural complexity may limit the suitability of these systems for real-time underwater monitoring or low-power embedded applications.

v. EfficientNetB0 Performance Analysis
EfficientNetB0 exhibited substantial underperformance across all evaluation metrics, recording very low accuracy, recall, and F1-score values, along with a notably high loss value of 2.76139. These results indicate significant learning instability, insufficient feature generalisation, or potential overfitting during training. The poor performance may also suggest sensitivity to underwater image variability, inadequate optimisation convergence, or limitations in adapting to the dataset characteristics used in this study.

The comparative findings strongly establish MobileNetV2 as the most effective and reliable architecture for underwater fish species classification. Its superior balance between accuracy, precision, recall, F1-score, and minimal loss demonstrates exceptional classification capability and strong generalisation performance across diverse underwater-imaging conditions. Furthermore, the implementation of k-fold cross-validation enhanced the statistical reliability and robustness of the evaluation by validating the model across multiple dataset partitions and reducing the sampling bias. This project successfully developed an intelligent deep-learning-based fish species recognition system capable of accurately classifying underwater fish images. The optimised MobileNetV2 model achieved outstanding performance metrics, including an accuracy of 97.93%, precision of 98.19%, recall of 97.93%, and F1-score of 97.97%, confirming its effectiveness for automated marine species identification. During development, several challenges related to model convergence, training instability, and validation performance were addressed through improvements in optimisation strategies, hyperparameter tuning, architectural refinements, and enhanced data preprocessing techniques. Additionally, k-fold cross-validation was employed to ensure reliable model evaluation and improved generalisation across multiple data subsets. To further improve usability and operational accessibility, the system was integrated with a graphical user interface (GUI), enabling efficient interaction, visualisation, and deployment for end users. Overall, this study contributes a robust and practical solution for automated fish species classification, with significant potential applications in marine biodiversity conservation, environmental monitoring, intelligent aquaculture systems, fishery resource management, and real-time underwater computer vision analytics.

4. Conclusions

MobileNetV2 demonstrates strong potential for advancing underwater fish species classification and intelligent aquaculture systems through efficient deep learning-based image analysis. Despite its high classification performance, several technical limitations remain, particularly in handling underwater image distortions, such as low illumination, colour attenuation, turbidity, suspended particles, and background noise. Conventional preprocessing methods are often insufficient for adapting to varying underwater environments, highlighting the need for adaptive enhancement techniques, including denoising, contrast enhancement, colour correction, and dehazing algorithms. In addition, limited underwater datasets, class imbalance, and labour-intensive annotation processes reduce the model generalisation capability and increase classification difficulty, especially for visually similar species. Although MobileNetV2 offers lower computational complexity than larger convolutional architectures, real-time deployment in embedded or edge-based underwater systems still requires optimisation through model compression, pruning, quantisation, and lightweight inference strategies. Future improvements should incorporate underwater-specific data augmentation techniques, transfer learning, semi-supervised learning, and attention-based architectures to improve feature extraction robustness and classification accuracy. Expanding datasets with diverse environmental conditions and integrating multimodal sensing data, such as sonar, depth, and temporal information, could further enhance model reliability. Moreover, the implementation of real-time processing pipelines, continual learning mechanisms, and user feedback systems would support adaptive performance improvement in practical marine applications. Strengthening collaboration between computer vision researchers, marine biologists, and the aquaculture industry is essential for developing standardised datasets, improving annotation quality, and ensuring reliable deployment in real-world environments. Overall, the integration of advanced machine learning techniques with optimised preprocessing and robust validation strategies positions MobileNetV2 as a highly promising solution for automated marine biodiversity monitoring, sustainable fisheries management, and intelligent underwater vision systems.

Acknowledgements
We would like to thank Universiti Malaysia Pahang Al-Sultan Abdullah and BMM Nagano Sdn Bhd (Malaysia) for their support through a collaborative industrial matching research grant.

Funding
This study was supported by a collaborative industrial matching research grant between Universiti Malaysia Pahang Al-Sultan Abdullah (UMPSA) and BMM Nagano Sdn Bhd (Malaysia) (RDU232401 and UIC230810) entitled "Integrated Machine Performance Monitoring System for Improvement of Printed Circuit Board (PCB) Machine Efficiency".

Declaration of Competing Interest
The authors declare no conflicts of interest.

Credit Authorship Contribution Statement
Kiran Kumar Vijayakumar: Conceptualisation, Methodology, Software, Data curation, Investigation, Formal analysis, Visualisation, Writing – original draft.
Ahmad Shahrizan Abdul Ghani: Conceptualization; Supervision; Project administration; Funding acquisition; Validation; Writing – review and editing.
Nur Najmiyah Jaafar: Methodology; Data curation, Investigation; Validation; Writing – review, and editing.
Mohd Yazid Abu: Formal analysis; Validation; Resources; Writing – review and editing.
Yulvia Sabri: Investigation; Resources; Validation; Writing – review and editing.
Mohd Fauzi Hassan: Resources; Project administration; Validation; Writing – review and editing.
Fusaomi Nagata: Supervision, Methodology, Validation, Writing – review and editing.
Xin Jin: Methodology; Software; Formal analysis; Validation; Writing – review and editing.
Mohd Azri Hizami Abd Rasid: Methodology; Data curation, visualisation, validation, writing, review, and editing.

Availability of Data and Materials
The datasets generated and/or analysed during the current study are available from the corresponding author upon reasonable request.

Ethics Declarations
This study did not involve human participants or animal subjects. Therefore, ethical approval was not required.

Generative Artificial Intelligence Declarations
Generative artificial intelligence tools were used solely to improve the readability and language quality. All scientific content, analyses, interpretations, and conclusions were reviewed and approved by the authors, who took full responsibility for the manuscript.

References
[1] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. MobileNetV2: Inverted residuals and linear bottlenecks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018: 4510-4520. https://doi.org/10.1109/cvpr.2018.00474
[2] Xu S, Zhang M, Song W, Mei H, He Q, Liotta A. A systematic review and analysis of deep learning-based underwater object detection. Neurocomputing. 2023;527;204–232. https://doi.org/10.1016/j.neucom.2023.01.056
[3] Tang Y, Iwaguchi T, Kawasaki H, Sagawa R, Furukawa R. AutoEnhancer: Transformer on U-net architecture search for underwater image enhancement. In: Wang L, Gall J, Chin TJ, Sato I, Chellappa R, editors. Computer Vision – ACCV 2022. Lecture Notes in Computer Science. Vol. 13846. Cham: Springer; 2023: 120–137. https://doi.org/10.1007/978-3-031-263132_8
[4] Haider A, Arsalan M, Nam SH, Sultan H, Park KR. Computer-aided fish assessment in an underwater marine environment using parallel and Progressive Spatial Information Fusion. Journal of King Saud University - Computer and Information Sciences, 2023;35(3):211–226. https://doi.org/10.1016/j.jksuci.2023.02.016
[5] Zheng S, Lin X, Zhang W, He B, Jia S, Wang P, Jia F. MDCC-Net: multiscale double-channel convolution U-Net framework for colorectal tumor segmentation. Computers in Biology and Medicine, 2021;130:104183. https://doi.org/10.1016/j.compbiomed.2020.104183
[6] Teli SG, Shelke RJ. Review of Underwater Image Enhancement using CNN and U-net. Journal of Artificial Intelligence and Imaging. 2023:1(1);5–10. https://doi.org/10.48001/joaii.2023.115-10
[7] Mohd Azmi KZ, Abdul Ghani AS, Md Yusof Z, Mohammad-Noor N, Ismail HN, Abu MY. 2024. Underwater image contrast enhancement through an intensity-randomised approach incorporating a swarm intelligence technique with unsupervised dual-step fusion. International Journal of Image and Data Fusion. 2024;15(3):310-344. https://doi.org/10.1080/19479832.2023.2289490
[8] Liu Y, An D, Ren Y, Zhao J, Zhang C, Cheng J, Liu J, Wei Y. DP-FishNet:Dual-path pyramid vision transformer-based Underwater Fish Detection Network. Expert Systems with Applications. 2024;238:122018. https://doi.org/10.1016/j.eswa.2023.122018

[9] Huang H, Lin L, Tong R, Hu H, Zhang Q, Iwamoto Y, et al. UNet 3+: A full-scale connected unet for medical image segmentation. In: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing; 2020: 1055-1059. https://doi.org/10.1109/icassp40776.2020.9053405

[10] Chuang YC., Hu Z Han B Wang P, Zhao Y, Wu H. . Intelligent Measurement of Morphological Characteristics of Fish Using Improved UNet. Electronics. 2021;10(12):1426. https://doi.org/10.3390/electronics10121426

[11] Sfakianakis C, Simantiris G, Tziritas G. Gudu: Geometrically constrained ultrasound data augmentation in U-Net for echocardiography semantic segmentation. Biomedical Signal Processing and Control 2023; 82:104557. https://doi.org/10.1016/j.bspc.2022.104557

[12] Kamarul Baharin MAS, Abdul Ghani AS, Syamsul Amri SQ, Mohammad-Noor N, Ismail HN. Enhancement of Low-Quality Diatom Images using Integrated Automatic Background Removal (IABR) ethod from Digital Microscopic Image. In: 2023 IEEE 13th Symposium on Computer Applications & Industrial Electronics; 2023: 256-261 https://doi.org/10.1109/ISCAIE57739.2023.10165175

[13] Zhou J, et al. UGIF-Net: An Efficient Fully Guided Information Flow Network for Underwater Image Enhancement. IEEE Transactions on Geoscience and Remote Sensing. 2023;61:1-17. Article 4206117. https://doi.org/10.1109/TGRS.2023.3293912

[14] Ben Tamou A, Benzinou A, Nasreddine K 2022. Targeted Data Augmentation and Hierarchical Classification with Deep Learning for Fish Species Identification in Underwater Images. Journal of Imaging. 2022;8(8):214. https://doi.org/10.3390/jimaging8080214

[15] Wang X, Mao L, Wang J. Underwater Fish Image Enhancement method based on color correction. Highlights in Science, Engineering and Technology. 2022;1:415–425. https://doi.org/10.54097/hset.v1i.498

[16] Mohd Azmi KZ, Abdul Ghani AS, Md Yusof Z. Red and blue channels correction based on green channel and median-based dual-intensity images fusion for turbid underwater image quality enhancement. In: Md Zain Z, et al., editors. Proceedings of the 11th National Technical Seminar on Unmanned System Technology 2019. Lecture Notes in Electrical Engineering. Vol. 666. Singapore: Springer; 2021: 73–86. https://doi.org/10.1007/978-981-15-5281-6_6

[17] Hambarde P, Murala S, Dhall A, UW-GAN: Single-Image Depth Estimation and Image Enhancement for Underwater Images. IEEE Transactions on Instrumentation and Measurement. 2021;70:1–12. Article 5018412. https://doi.org/10.1109/TIM.2021.3120130

[18] Fitrianah D, Suryaningrum K M, Sagala NT, Ayumi V, Lim SM. Fine-tuned mobilenetv2 and VGG16 algorithm for Fish Image Classification. In: 2022 International Conference on Informatics, Multimedia, Cyber and Information System; 2022: 384–389. https://doi.org/10.1109/ICIMCIS56303.2022.10017901

[19] Liu X, Jia Z, Hou X, Fu M, Ma L, Sun Q. Real-time marine animal images classification by embedded system based on Mobilenet and Transfer Learning. In: OCEANS 2019 - Marseille; 2019: 1–5. https://doi.org/10.1109/OCEANSE.2019.8867190

[20] Kamarul Baharin MAS, Abdul Ghani AS, Mohammad-Noor N, Ismail HN, Syamsul Amri SQ. Automatic phytoplankton image smoothing through integrated dual image histogram specification and enhanced background removal method. The Imaging Science Journal. 2021;69(1-4), 65-91. https://doi.org/10.1080/13682199.2022.2149067