JIPK
(JURNAL ILMIAH PERIKANAN DAN KELAUTAN)

Scientific Journal of Fisheries and Marine

Research Article

Deep Learning Models Performance on Marine Fish Species Classification

Nur Muhammad Afiq Anang [iD] , **and Ezmahamrul Afreen Awalludin*** [iD]

Faculty of Fisheries and Aquaculture Sciences, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia.

Abstract

Identifying marine fish species accurately can be difficult due to their subtle anatomical and color pattern similarities, which often result in misclassification during ecological assessments and fisheries operations. Manual identification methods are time-consuming and prone to errors especially in high throughput environments such as fish markets. In this study, transfer learning is used to evaluate three deep learning models ResNet-50, AlexNet and GoogLeNet on a total of 20,325 images from twenty marine fish species acquired from Kuantan (Pahang) and Mengabang Telipot (Kuala Nerus), Malaysia. All images were morphologically classified as complete fish, head, body and tail. The dataset was subjected to preprocessing procedures which encompassed image resizing, pixel normalization and data augmentation techniques that consists of random rotation (±15°), horizontal flipping, adjustments to brightness and contrast (±20%) and cropping. Subsequently, the dataset was partitioned into 80% training set (16,260 images), 10% validation set (2,032 images) and 10% testing set (2,033 images). The classification patterns were analyzed using confusion matrices and standard metrics such as accuracy, precision and recall. ResNet-50 outperformed other models achieving ideal results with 100% accuracy, precision and recall in every category. With 99.5% and 99.4% accuracy, GoogleNet and AlexNet came in second and third, respectively. This study shows that deep learning models especially ResNet-50 achieved an accurate and efficient way to classify fish species automatically. With multi-view images, data augmentation and transfer learning, the model performs well even in difficult visual conditions. These results support its use in real-time fisheries monitoring, biodiversity studies, and environmental impact assessments.

1. Introduction

Fish classification plays a vital role in ecological monitoring, biodiversity conservation, and sustainable fisheries management (Yang *et al.*, 2024). Traditionally, this process has relied on morphological characteristics and DNA barcoding techniques. While these methods provide results, they demand significant resources and require highly skilled personnel, limiting their practicality for large-scale applications (Iqbal *et al.*, 2021; Hilal *et al.*, 2023). Moreover, the hands-on aspect of these methods makes them susceptible to human mistakes and personal bias, particularly when addressing uncommon or morphologically akin species. Consequently, the scalability and practicality in contemporary fisheries science continue to pose significant challenges.

Recently, deep learning an advanced branch of artificial intelligence has emerged as a powerful tool for automating fish classification, offering high accuracy with minimal human intervention (Saleh *et al.*, 2022; Hasan *et al.*, 2022; Han *et al.*, 2022). Deep learning models, particularly Convolutional Neural Networks (CNNs) have significantly transformed the field of image recognition by effectively extracting complex hierarchical features directly from raw image data (Peddina and Mandava, 2025; Prasetyo *et al.*, 2022; Zhang *et al.*, 2022; Wang *et al.*, 2023, Iqtait *et al.*, 2024). In contrast to traditional computer vision methods that rely on manual feature extraction, CNNs autonomously learn spatial and temporal patterns, thereby improving accuracy and reducing the necessity for specialized domain knowledge. Nevertheless, the capacity of CNNs to classify various components of fish anatomy, including the head, body, and tail, remains challenging due to high visual similarity in texture and shapes across species, which is less distinct than in typically general objects classification (Lan *et al.*, 2020; Ahmad *et al.*, 2023).

A typical CNN model is made up of several essential elements: the input image, convolutional layer, pooling layer, activation function, fully connected layer, and output layer (Rawat and Wang, 2017; Song *et al.*, 2019). The process starts with the input image, which represents visual data from the pixels of a digital image. The convolutional layer employs a kernel or filter to automatically identify specific features, producing a comprehensive channel representation of the input images. Research indicates that CNN models like ResNet, AlexNet, and GoogLeNet effectively classify marine organisms, including various fish species, plankton, and coral (Allken *et al.*, 2019; Wang *et al.*, 2021; Veluswami and Panneerselvam, 2021; Zhou *et al.*, 2022; Alinsug *et al.*, 2024; Zhai *et al.*, 2023).

However, the growing potential of research reveals significant deficiencies in model evaluation, especially concerning classification accuracy across different fish body areas. This area remains poorly understood, even though comprehensive categorization is crucial for effective real-time underwater monitoring systems. Furthermore, many current models rely on broad benchmark datasets, which limit their capacity to adapt specific local marine environments.

Limited research has methodically evaluated various CNN architectures on a consistent dataset, employing uniform performance metrics specifically for fish classification (Dong *et al.*, 2023; Rawat and Wang, 2017; Sun *et al.*, 2020). Current studies infrequently investigate the performance of various CNNs across diverse data conditions or with images taken in authentic aquatic settings (French *et al.*, 2020; Ismail *et al.*, 2021; Kaya *et al.*, 2023). Additionally, the visual background and camera settings used during image acquisition are often overlooked, even though they significantly impact feature extraction and model learning. Therefore, it's crucial to conduct a more in-depth investigation into fish image classification that considers both full-body images and region-specific segments across different CNN models.

This study examines how effectively three prominent CNN architectures ResNet50, AlexNet, and GoogLeNet classify 20 local marine fish species using a dataset collected in controlled laboratory settings. The research aims to address this question on How accurate are CNN models in classifying fish species when employing augmented and region-specific image data? The experimental design includes both full-body images and cropped images that concentrate on the head, body, and tail, aimed at evaluating classification performance in specific regions. Model performance is evaluated through established metrics like accuracy, precision, and recall, complemented by confusion matrices and training loss curves (Ahmed *et al.*, 2023)

This research propels the field of ecological AI by addressing several research gaps through innovative dataset design, data augmentation techniques, and a comparative analysis of CNN architectures (Ben Tamou *et al.*, 2022; Deka *et al.*, 2023). The study concentrates on developing a carefully annotated fish image dataset, segmented by regions and representative of authentic marine environments (Zheng *et al.*, 2024). The study presents a comparison of three notable CNN models, highlighting their real-world uses and limitations. The results aim to improve scalable and accurate fish classification by offering an advanced AI-based approach to enhance fisheries management, species

monitoring, and marine conservation efforts. To tailor the models for fish species identification, transfer learning was applied (Deka *et al*., 2023; Zheng *et al*., 2024). Each model underwent training across ten separate sessions, using the same hyperparameters to ensure consistency and evaluate performance through various iterations. This standardized training approach guaranteed stable results despite the typical random variations associated with deep learning optimization methods.

2. Materials and Methods

2.1 Material

The tools used in this research were selected for their ability to obtain high-resolution aquatic images suitable for classification purposes and to recreate a controlled aquatic environment.

2.1.1 The equipments

Fish images were acquired using a Huawei P30 Lite smartphone equipped with a 24 MP wide lens and AI scene recognition, ensuring consistent high-resolution output suitable for training deep learning models. This device features a triple-lens setup comprising a 24 MP wide lens, an 8 MP ultra-wide lens, and a 2 MP depth sensor, all precisely designed to deliver high-resolution images of 1080 × 2312 pixels. The integration of AI scene recognition and image stabilization features guarantees consistent and sharp image quality, which is crucial for training deep learning models. Furthermore, the study used a standard glass aquarium measuring 30 cm in height, 30 cm in width, and 40 cm in length as the observation chamber. The transparent glass walls minimized distortion and maintained uniform lighting throughout the image acquisition process. The aquarium was designed with standardized dimensions and substrate depth to replicate natural marine conditions while optimizing image clarity and fish mobility for classification purposes (Zheng *et al*., 2024).

2.1.2 The materials

The research initiative concentrated on the classification of fish species utilizing Convolutional Neural Networks (CNN) and depended on two essential materials: sand and saltwater. These materials were chosen due to their significant role in the marine ecosystem, which is imperative for the study of various fish species. By using sand and saltwater, researchers were able to create environments that closely mimic the natural habitats of the fish being studied. Furthermore, the artificial saline mixture of saltwater was meticulously prepared and introduced into

the aquarium to a depth of 20 centimeters, yielding an estimated volume of 24 liters, determined by the tank's dimensions (40 cm length × 30 cm width × 30 cm height). The designated water depth was carefully selected to mimic the natural environment of the fish species and to enhance optimal camera focus from different angles. In addition, a consistent layer of 5.0 cm thick natural beach sand was evenly distributed along the tank's bottom. Adding beach sand achieved two main goals: (1) it replicated the fish species' natural habitat, and (2) it enhanced the visual background's complexity, which supports the development of effective computer vision models (Zheng *et al*., 2024).

2.1.3 Ethical approval

This research did not include the use of live experimental animals, humans, or any protected species. Consequently, there was no requirement for ethical approval from an institutional animal care and use committee. Nevertheless, the research adhered to established ethical principles for gathering image-based data in environmental contexts. As precaution, any future studies involving biological samples or live subjects will secure ethical clearance in accordance with institutional policies and national regulations.

2.2 Method

2.2.1 Experimental design

The experimental framework used in this study and provides a systematic overview of the research workflow. This framework encompasses data collection, data augmentation, dataset training, model evaluation, and classification test. These components are essential for comprehending the contributions of each phase to the overall success of the research objectives (Figure 1). By carefully following this framework, the study aims to ensure strong results and improve the reliability of the findings.

2.2.2 Fish specimen collection

Twenty marine fish species were collected for classification, as detailed in Table 1. Data collection took place over a 12-month period, from January to December 2024, enabling analysis of seasonal variations in species availability. Specimens were obtained from various local markets and vendors in Kuantan (Pahang) and Mengabang Telipot, Kuala Nerus (Terengganu), Malaysia. These locations were thoughtfully chosen to ensure a thorough representation of the marine biodiversity along the East Coast of Peninsular Malaysia.

2.2.3 Data augmentation

This study employed a range of image augmentation techniques on the unprocessed dataset to enhance the robustness and generalizability of the fish classification model. The methodologies utilized were meticulously developed to improve the dataset and incorporate variability that accurately reflects real-world scenarios in a synthetic environment. After acquiring images under controlled conditions, each original image was enhanced through various geometric and photometric transformations. Drawing from the research conducted by Yasin *et al.* (2023), this study examined various image manipulation techniques, such as rotations within a range of -30° to +30°, horizontal and vertical flips, scaling, cropping, and modifications to brightness and contrast. The selected methodologies endeavour to replicate the natural orientation of fish as well as the variations in lighting. This approach facilitates the model's ability to acquire invariant features across a range of conditions (Li *et al.*, 2024). The execution of these transformations was particularly imperative for species with limited initial sample sizes, thereby ensuring sufficient representation throughout the dataset.

environments. Images were captured using a stationary camera setup that ensured consistent lighting, maintained a 90-degree angle from above, and was positioned 15 cm away, which promoted uniformity among samples. This method was designed to minimize environmental noise and facilitate precise feature extraction during model training and validation.

2.2.4 Preparation of the training dataset

After augmentation, the dataset was divided into training and testing subsets with an 80:20 stratified split to maintain the proportional representation of all species in both sets. Before training began, each image was resized to 224 × 224 pixels to comply with the input specifications mandated by the convolutional neural network (CNN). Taxonomists performed extensive validation and label annotation on each image, ensuring ground truth accuracy. The research employed three advanced pre-trained CNN models ResNet50, AlexNet, and GoogLeNet chosen for their established image classification effectiveness and capacity for fine-tuning on specific datasets. Transfer learning was

Figure 1. The process of experimental design.

The operations resulted in an expansion of the dataset to 20,325 images, which showcased 20 distinct fish species. This enhanced dataset was specifically designed to address class imbalance and to enrich the variety of included visual features. Considerable efforts were undertaken to ensure that augmentation was implemented consistently across all classes, thereby preserving the balance of the dataset and mitigating the emergence of any potential bias. This approach aligns with the best practices noted in recent research (Okafor *et al.*, 2018), indicating that augmentation improves classification performance and acts as a regularization technique during model training.

To develop a varied image dataset for model training, images of each species were captured in two controlled settings: (i) on a flat surface with laminated white A3 paper as the background and (ii) in an artificial underwater environment, as shown in Figures 2 and Figure 3. The underwater scene included a transparent aquarium filled with saltwater and natural beach sand from local coastal areas to replicate genuine marine conditions. This two pronged background strategy sought to improve the model's versatility by integrating both controlled and naturalistic imaging

employed to tailor the models for the specific task of recognizing fish species. Each model underwent ten training sessions, utilizing the same hyperparameters to ensure consistency and evaluate performance across various iterations. This consistent training methodology guaranteed that the outcomes were not affected by the random fluctuations typically associated with deep learning optimization techniques. Furthermore, to assess the distinguishing capability of various anatomical areas, classification was conducted utilizing both full-body images and segmented images that concentrated on the fish head, mid-body, and tail. This comprehensive analysis across multiple regions sought to pinpoint the body part that holds the most significant visual cues for precise species classification, providing essential insights for the advancement of more effective and localized recognition systems.

2.2.5 Classification and assessment

After finishing the model training, evaluations for classification were carried out on all three CNN architectures. Performance evaluation was conducted through three main tools: the confusion matrix, training progress graphs, and classification test reports. The
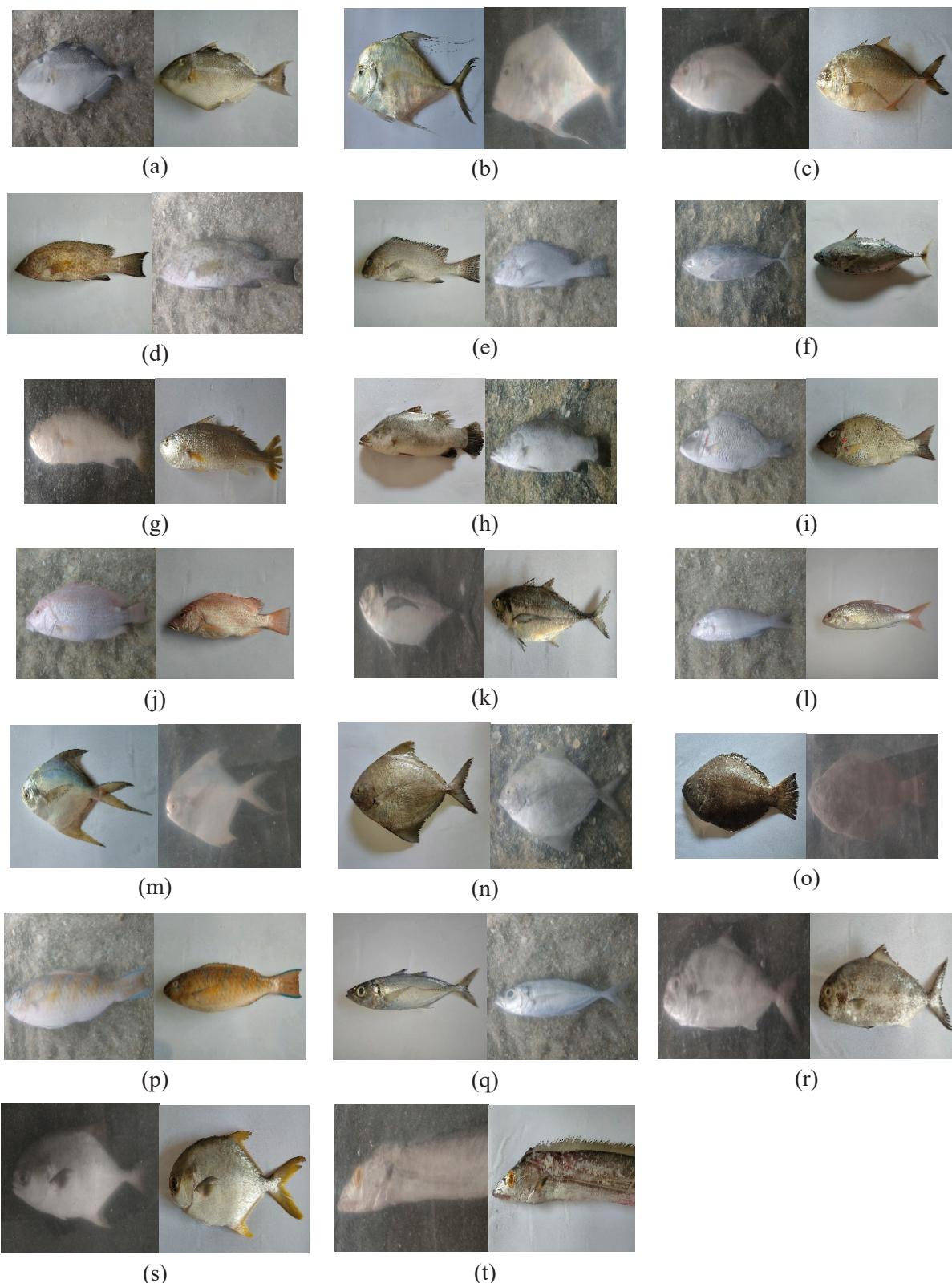
Figure 2. Images of fish species presented on a laminated A3 background as well as under underwater conditions: (a) *Abalistes stellatus*, (b) *Alectis indica/indicus*, (c) *Carangoides gymnosthetus*, (d) *Pampus argenteus*, (e) *Epinephelus bleekeri*, (f) *Euthynnus affinis*, (g) *Johnius dussumieri*, (h) *Lates calcarifer*, (i) *Lethrinus lentjan*, (j) *Lutjanus gibbus*, (k) *Megalaspis cordyla*, (l) *Nemipterus furcosus*, (m) *Pampus argenteus*, (n) *Parastromateus niger*, (o) *Psettodes erumei*, (p) *Scarus ghobban*, (q) *Selar crumenophthalmus*, (r) *Seriolina nigrofasciata*, (s) *Trachinotus blochii*, (t) *Trichiurus lepturus*.

confusion matrix facilitated the calculation of fundamental performance metrics, encompassing overall accuracy, precision, and recall. Accuracy is defined as the ratio of correctly classified samples to the total number of samples, whereas precision assesses the proportion of true positives among all positive predictions. Recall, also known as sensitivity, indicated the percentage of true positives accurately recognized by the model. The metrics were derived using the established formula.

Accuracy is calculated by taking the sum of true positives and true negatives, divided by the total of true positives, false positives, true negatives, and false negatives.

Accuracy = (Tp+ Tn)/(Tp+ Tn+ Fp+ Fn)×100%.(Eq 1)

Accuracy is determined by the ratio of true positives to the sum of true positives and false positives.

Precision = TP/(Tp + FP) x100%.....................(Eq 2)

Recall is calculated as the ratio of true positives to the sum of true positives and false negatives.

Recall = Tp/(Tp+ Fn) ×100%...........................(Eq 3)

Where true positives (*Tp*), true negatives (*Tn*), false positives (*Fp*), and false negatives (*Fn*) of the model's predictions.

These metrics provided a comprehensive view of model performance, enabling a detailed analysis that goes beyond mere accuracy numbers. Graphs illustrating training progress were created to monitor the changes in loss and accuracy throughout the training epochs. The visualizations played a crucial role in evaluating the learning behaviour of each CNN and pinpointing possible challenges like under fitting or overfitting. Given the constraints of the hardware, we presented representative graphs for each model to illustrate the trends in convergence. The classification tests involved comparing results across various anatomical regions to determine which segment yielded the highest accuracy. This segmentation method supports the overarching goal of creating optimised, resource-efficient models for immediate application in aquaculture systems.

2.3 Analysis Data

The evaluation of model performance was conducted comprehensively through both statistical and visual techniques utilizing MATLAB Software 2022b, which is licensed to Universiti Malaysia Terengganu. Confusion matrices were created for each model to as-

sess the accuracy of predictions for each fish species. The matrices offered valuable insights into particular inter-class misclassifications, especially among visually similar species, aiding in pinpointing the areas where the model faced challenges in distinguishing between different fish types. Class-wise analysis of precision and recall scores was conducted to evaluate the performance of each model in managing both prevalent and infrequent species. Classification reports compiled all performance metrics and were utilized to evaluate models comprehensively, allowing for the identification of the most effective architecture. The examination further encompassed reviewing training progress charts for each CNN and graphing training accuracy and loss over the epochs. The curves facilitated the evaluation of how the model converges and helped pinpoint any irregularities in the training process.

The study aimed to assess not only model accuracy but also the practicality of implementing AI-driven classification within actual aquaculture and marine monitoring settings. The research offered important insights into practical deployment considerations by analysing how effectively the models distinguish species based on various anatomical features. This analysis highlighted the importance of image quality control, camera positioning, and habitat variability. The data indicated that convolutional neural network-based models can attain impressive accuracy in species classification if trained with well-augmented and balanced datasets in controlled settings. These findings contribute to the advancement of automated fish identification systems designed to support sustainable fisheries and ecological monitoring efforts.

3. Results and Discussion

3.1 Result

The confusion matrix serves as a tabular representation employed to assess the performance of a classification model by juxtaposing its predicted labels with the actual labels. In the context of this study, it elucidates the precision with which the deep learning model differentiates each species of marine fish. The matrix exhibits precise predictions along the diagonal; however, the off-diagonal entries signify instances of misclassification. Through a comprehensive analysis of this matrix, it can differentiate specific species that the model consistently misclassifies, thereby enabling subsequent improvements.

The graph depicting training progress effectively visualizes the model's learning trajectory gradually, typically illustrating metrics such as accuracy and

loss throughout the training process epochs. In this research, the graphical representation demonstrates the enhancement of the model's performance across each epoch. A consistent rise in accuracy accompanied by a corresponding decline in loss indicates effective learning. Conversely, any irregularities may signal complications such as overfitting or under fitting. It is imperative to monitor this graph to ensure that the model is training as anticipated and to facilitate the implementation of necessary adjustments.

Figure 4 displays the confusion matrix results from the initial training sessions of each Convolutional Neural Network (CNN) model. In Figure 4(a), the ResNet50 model's confusion matrix shows outstanding performance, achieving an accuracy, precision, and recall of 100%. This result demonstrates that the ResNet50 model effectively identified all fish species without any false positives or negatives. Figure 4(b) presents the confusion matrix for the AlexNet model, which recorded an accuracy of 99.9%, alongside

Table 1. Taxonomy of twenty selected fish species in this study
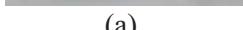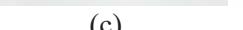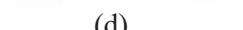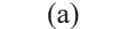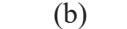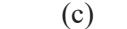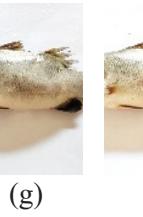
No.	Family	Genus	Species
1	Balistidae	*Abalistes*	*Abalistes stellatus*
2	Carangidae	*Scyris*	*Alectis indica/indicus*
3	Carangidae	*Carangoides*	*Carangoides gymnosthetus*
4	Serranidae	*Epinephelus*	*Epinephelus areolatus*
5	Serranidae	*Epinephelus*	*Epinephelus bleekeri*
6	Scombridae	*Euthynnus*	*Euthynnus affinis*
7	Sciaenidae	*Johnius*	*Johnius dussumieri*
8	Latidae	*Lates*	*Lates calcarifer*
9	Lethrinidae	*Lethrinus*	*Lethrinus lentjan*
10	Lutjanidae	*Lutjanus*	*Lutjanus gibbus*
11	Carangidae	*Megalaspis*	*Megalaspis cordyla*
12	Nemipteridae	*Nemipterus*	*Nemipterus furcosus*
13	Stromateidae	*Pampus*	*Pampus argenteus*
14	Carangidae	*Parastromateus*	*Parastromateus niger*
15	Psettodidae	*Psettodes*	*Psettodes erumei*
16	Scaridae	*Scarus*	*Scarus ghobban*
17	Carangidae	*Selar*	*Selar crumenophthalmus*
18	Carangidae	*Seriolina*	*Seriolina nigrofasciata*
19	Carangidae	*Trachinotus*	*Trachinotus blochii*
20	Trichiuridae	*Trichiurus*	*Trichiurus lepturus*

3.1.1 Confusion table

perfect precision and recall at 100%. While its accuracy

Table 2. The process of data augmentation applied to enhance the training dataset for deep learning models

| Pre-Processing	Images
Whole Picture Brightness/Contrast (sharpen 25%): (a) 0%/0% (b) 10%/10% (c) 20%/20% (d) 30%/30% Same Brightness/Contrast as above images (Sharpen -25%): (e), (f), (g), (h)	(a) (b) (c) (d) (e) (f) (g) (h)
Head (Cropped) Brightness/Contrast: (a) 0%/0% (b) 10%/10% (c) 20%/20% (d) 30%/30% Same Brightness/Contrast as above images (Sharpen -25%): (e), (f), (g), (h)	(a) (b) (c) (d) (e) (f) (g) (h)
Body (Cropped) Brightness/Contrast: (a) 0%/0% (b) 10%/10% (c) 20%/20% (d) 30%/30% Same Brightness/Contrast as above images (Sharpen -25%): (e), (f), (g), (h)	(a) (b) (c) (d) (e) (f) (g) (h)

Tail (Cropped)
Brightness/Contrast:
(a) 0%/0% (b) 10%/10%
(c) 20%/20% (d) 30%/30%

(a)　　(b)　　(c)　　(d)

Same Brightness/Contrast
as above images (Sharpen
-25%): (e), (f), (g), (h)
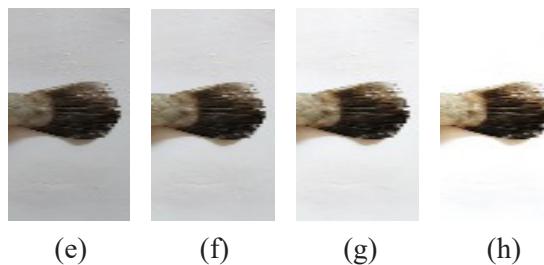
(e)　　(f)　　(g)　　(h)

Rotation
(a) 0˚ (b) 90˚
(c) 180˚ (d) 270˚

(a)　　(b)　　(c)　　(d)
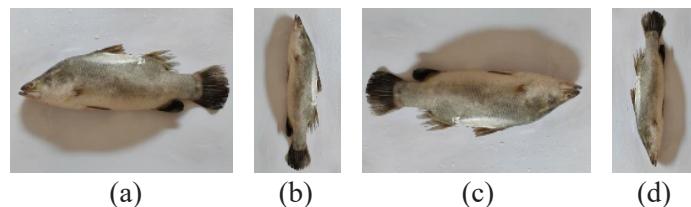
Same rotation as above
images (Sharpen -25%): (e),
(f), (g), (h)

(e)　　(f)　　(g)　　(h)
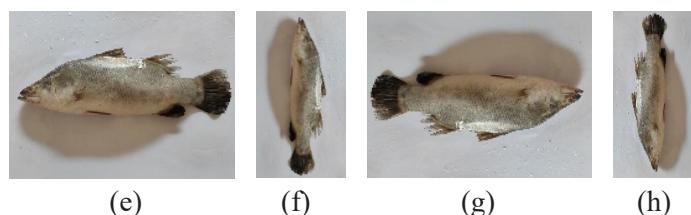
*The underwater images also through a similar data augmentation process as above.

Aquarium Water

Salt Water
= 20.0 cm
height

Beach Sand
= 5.0 cm
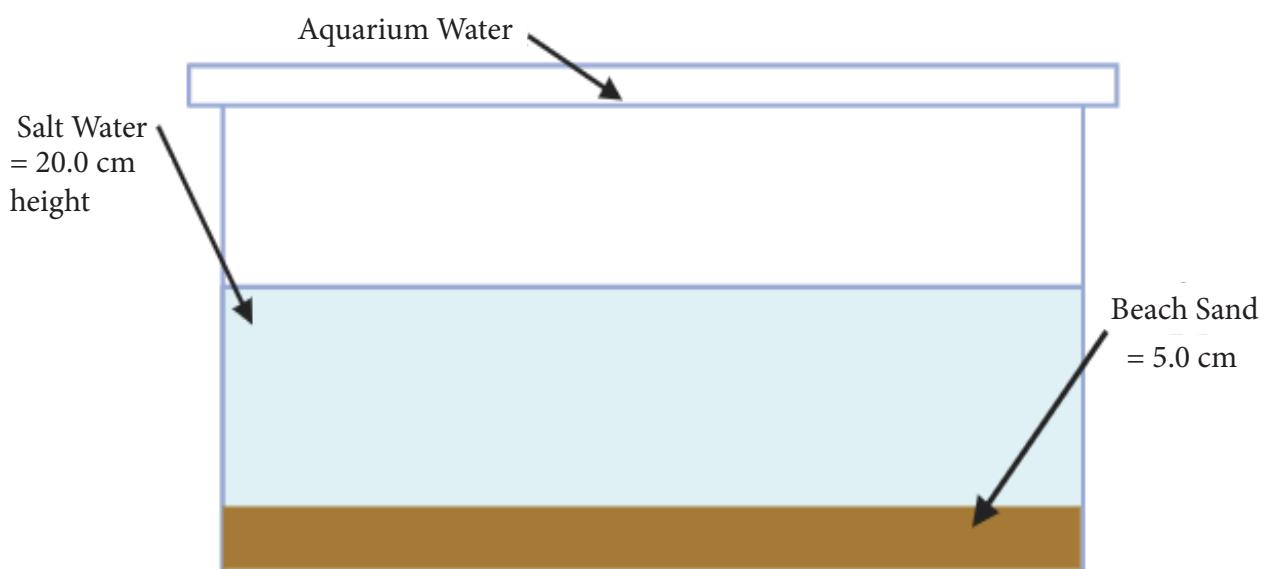
Figure 3. The aquarium has been thoughtfully organized to replicate an underwater ecosystem, thereby enabling the acquisition of authentic images of fish.

is slightly lower than that of ResNet50, its flawless precision and recall illustrate its capability in classifying fish species in this training period. Figure 4(c) shows the confusion matrix for the GoogLeNet model, which reached an accuracy of 99.6%, with perfect precision and recall (100%). Although its accuracy is a bit lower than that of AlexNet and ResNet50, GoogLeNet still exhibited strong performance in the classification task, resulting in no false positives or negatives. Table 3 summarizes the confusion matrix results for each model and training session, enabling a thorough comparison of performance across different models. The results highlight the effectiveness of the Convolutional Neural Network (CNN) models in classifying fish species, with ResNet50 noted as the model demonstrating the best performance efficiency. The discrepancies in performance levels among these models may be ascribed to their respective architectural intricacies, particularly the utilization of deeper layers and skip connections in ResNet50, which contribute to its enhanced accuracy. Although AlexNet and GoogLeNet display marginally lower accuracy rates, both models demonstrate commendable performance, achieving flawless precision and recall metrics, thereby establishing them as dependable alternatives for this classification task.

for the various CNN models: (a) ResNet50, (b) AlexNet, and (c) GoogLeNet. Due to hardware limitations, specifically the constraints imposed by a single-core CPU, training was executed solely once for each model. The dataset, comprising 20,325 images, posed challenges for the system in terms of repeated processing. Each graph illustrates the validation value, which serves as an indicator of the model's performance concerning accuracy following each training session. The loss value denotes the number of incorrect predictions made during the training process, functioning as a critical metric for assessing the model's learning progression over time. A reduced loss value signifies superior model performance, whereas an elevated value indicates opportunities for enhancement. The recorded training time encompasses the total duration taken for the model to complete the training, including the time required for data loading. Progress in training was monitored subsequent to each designated epoch, with live updates reflecting the evolution of various parameters throughout the session. This approach enabled real-time tracking of model performance. Nonetheless, hardware limitations restricted us to a single training session for each model. The training was limited to two epochs, using sixty-four mini batches and a learning rate of 0.001 to enhance the model's

Table 3. Comparison of classification performance through confusion matrices for ResNet-50, AlexNet, and GoogleNet

Training	ResNet50			AlexNet			GoogLeNet		
	A	P	R	A	P	R	A	P	R
1	100%	100%	100%	99.9%	100%	100%	99.6%	100%	100%
2	100%	100%	100%	100%	100%	100%	99.6%	100%	100%
3	100%	100%	100%	99.9%	100%	100%	99.7%	100%	100%
4	100%	100%	100%	99.9%	100%	100%	99.4%	100%	100%
5	100%	100%	100%	100%	100%	100%	99.4%	100%	100%
6	100%	100%	100%	100%	100%	100%	99.7%	100%	100%
7	100%	100%	100%	99.9%	100%	100%	99.5%	100%	100%
8	100%	100%	100%	100%	100%	100%	99.7%	100%	100%
9	100%	100%	100%	99.9%	100%	100%	99.4%	100%	100%
10	100%	100%	100%	99.9%	100%	100%	99.5%	99.5%	99.5%
Average	100%	100%	100%	99.4%	100%	100%	99.5%	99.9%	99.5%

3.1.2 Training progress graph

Figure 5 presents the training progress graphs

accuracy, considering the limited computational resources. In future projects, we might leverage more advanced hardware, allowing for additional training

sessions and more significant potential improvements to these models.

Table 4. The classification test on 10 fish species

Model	Training	Classification									
		1	2	3	4	5	6	7	8	9	10
ResNet50	1	✓	✓	x	✓	✓	✓	x	✓	✓	x
	2	✓	✓	x	✓	✓	✓	x	✓	✓	x
	3	✓	✓	x	✓	✓	✓	x	✓	✓	x
	4	✓	✓	x	✓	✓	✓	x	✓	✓	x
	5	✓	✓	x	✓	✓	✓	x	✓	✓	x
	6	✓	✓	x	✓	✓	✓	x	✓	✓	x
	7	✓	✓	x	✓	✓	✓	x	✓	✓	x
	8	✓	✓	x	✓	✓	✓	x	✓	✓	x
	9	✓	✓	x	✓	✓	✓	x	✓	✓	x
	10	✓	✓	x	✓	✓	✓	x	✓	✓	x
AlexNet	1	✓	✓	x	✓	x	✓	✓	x	x	✓
	2	✓	✓	x	✓	x	✓	✓	x	x	✓
	3	✓	✓	x	✓	x	✓	x	x	x	✓
	4	✓	✓	x	✓	x	✓	✓	x	x	✓
	5	✓	✓	x	✓	x	✓	✓	x	x	✓
	6	✓	✓	x	✓	x	✓	x	x	x	✓
	7	✓	✓	x	✓	x	✓	✓	x	x	✓
	8	✓	✓	x	✓	x	✓	✓	x	x	✓
	9	✓	✓	x	✓	x	✓	x	x	x	✓
	10	✓	✓	x	✓	x	✓	✓	x	x	✓
GoogLeNet	1	✓	✓	x	✓	x	✓	✓	✓	✓	x
	2	✓	✓	x	✓	x	✓	✓	✓	✓	x
	3	✓	✓	x	✓	x	✓	✓	✓	✓	x
	4	✓	✓	x	✓	x	✓	✓	✓	✓	x
	5	✓	✓	x	✓	x	✓	✓	✓	✓	✓
	6	✓	✓	x	✓	x	✓	✓	✓	✓	x
	7	✓	✓	x	✓	x	✓	✓	✓	✓	x
	8	✓	✓	x	✓	x	✓	✓	✓	✓	x
	9	✓	✓	x	✓	x	✓	✓	✓	✓	x
	10	✓	✓	x	✓	x	✓	✓	✓	✓	x

*This compiles results including for whole part of the fish, head, body, and tail (similar results)

3.1.3 Classification test

The classification test was conducted after each training session to validate the results shown in the confusion matrix, as presented in Table 4. For this study, we selected 10 fish species from a pool of 20 distinct species, ensuring that new images were used for testing, which were excluded from the training dataset to reduce bias. The tested species included *Az Pampus argenteus, Psettodes erumei, Scarus ghobban,* and *Trachinotus blochii.*

Figure 4. The confusion matrix table in training 1: (A) ResNet50, (B) AlexNet, (C) GoogLeNet
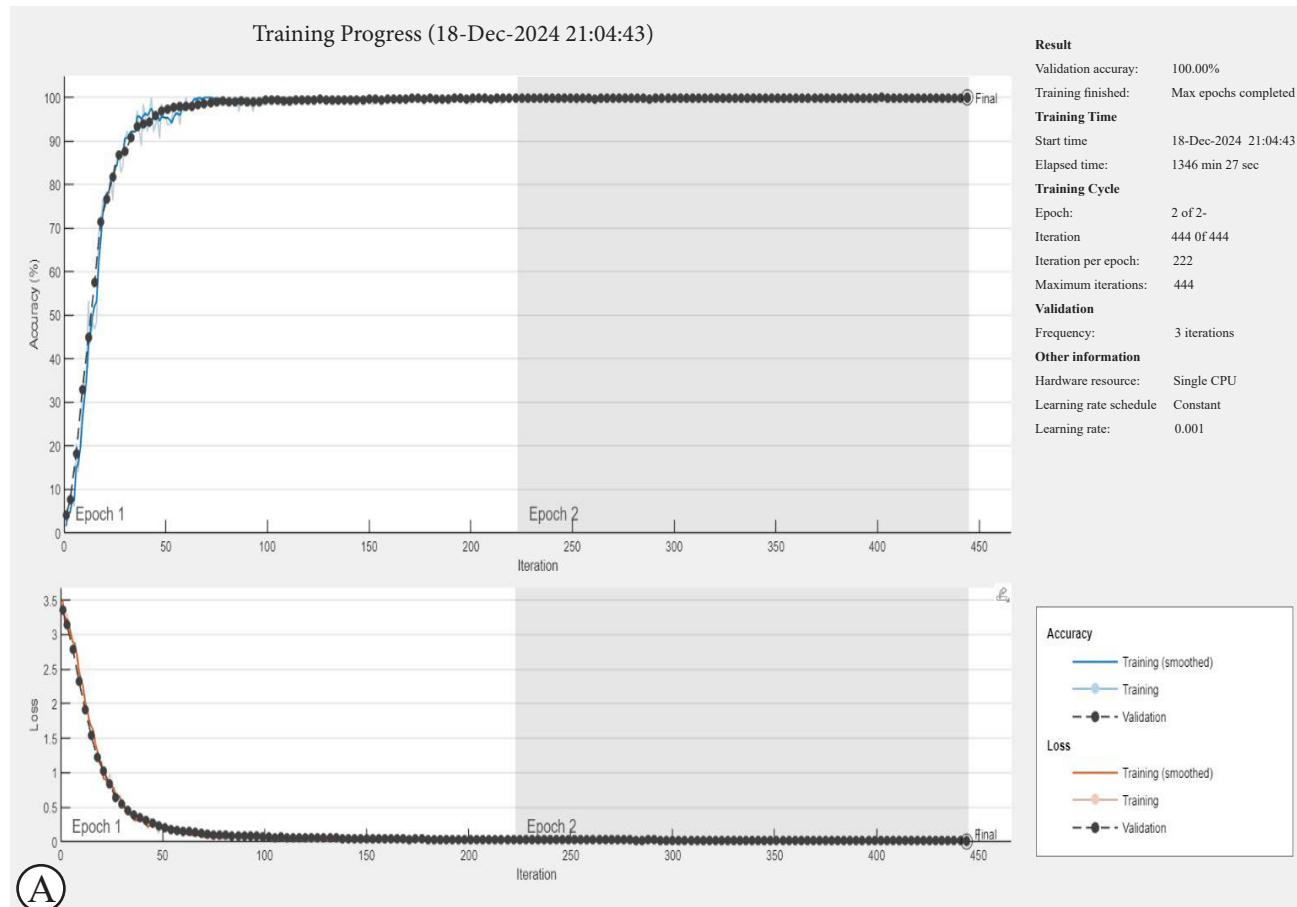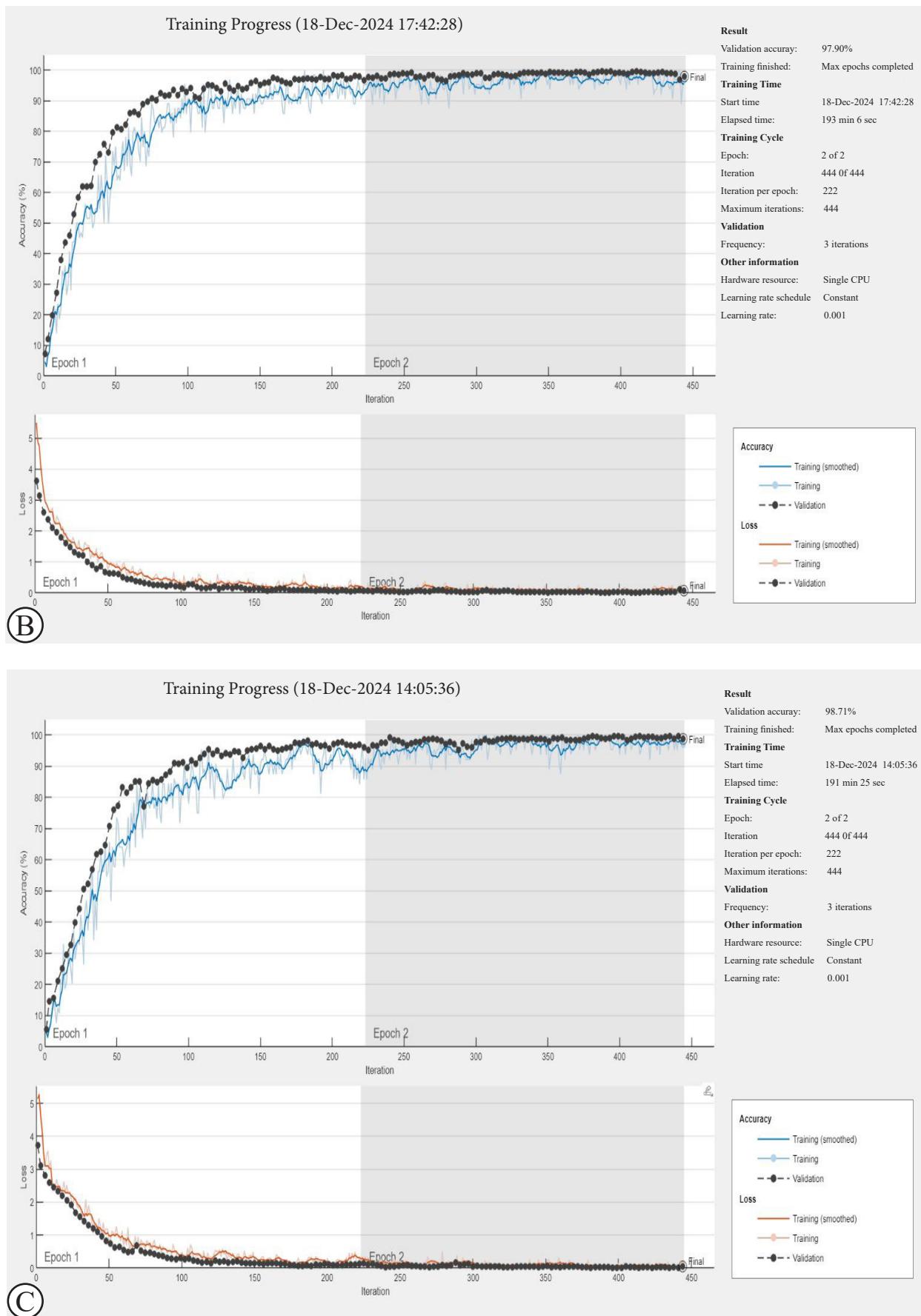
Figure 5. Training progress graphs for deep learning models: (A) ResNet-50, (B) AlexNet, and (C) GoogLeNet.

Table 5. Accuracy comparison from confusion matrix table and training progress graph on similar models in other studies

| Studies | CNN Models | | | | | |
| | ResNet50 | | AlexNet | | GoogLeNet | |
	CMT	TPG	CMT	TPG	CMT	TPG
(Deka *et al.*, 2023)	99.76%	100.0%	97.84%	97.84%	-	-
(Zhou *et al.*, 2022)	91.38%	-	90.16%	-	92.34%	-
(Zhou *et al.*, 2023)	92.22%	-	89.84%	-	90.28%	-
(Rauf *et al.*, 2019)	91.05%	-	85.54%	-	81.77%	-
Proposed Study	100.0%	100.0%	99.40%	97.90%	99.50%	98.71%

*CMT= Confusion Matrix Table, TPG= Training Progress Graph

The accuracy results derived from the confusion matrix indicate that ResNet 50 attained a perfect accuracy of 100%, whereas AlexNet achieved an accuracy of 99.4%, and GoogLeNet reached an accuracy of 99.5%. ResNet 50 maintained consistent performance, while AlexNet reached a peak training accuracy of 100% in sessions 2, 5, 6, and 8, despite fluctuations during these sessions. Although AlexNet did not match ResNet 50's reliability, its simpler architecture with fewer layers produced noteworthy results for a Convolutional Neural Network (CNN), mainly because it was one of the pioneering architectures for deep learning tasks. GoogLeNet also produced strong results, achieving a maximum accuracy of 99.7%, which occurred during training sessions 3, 6, and 8.

As illustrated in Table 5, the Training Progress Graph ResNet 50 consistently exhibited the best performance with 100% validation accuracy. In contrast, AlexNet and GoogLeNet attained validation accuracies of 97.9% and 98.71%, respectively. ResNet 50 reached 90% accuracy early in the training process (around iteration 50), while the other models required more time to reach that threshold. The deeper layers in ResNet 50 allowed it to capture more intricate patterns in the images, thus achieving superior accuracy. However, adding these deeper layers led to a longer training time for the model. Specifically, ResNet 50 needed 1,346 minutes and 27 seconds to finish its training considerably longer than AlexNet, which took 193 minutes and 6 seconds, and GoogLeNet at 191 minutes and 25 seconds. While the deeper layers of ResNet 50 enhance its accuracy, longer training times can pose difficulties on limited hardware, as high temperatures during prolonged training may negatively affect outcomes. Nonetheless, all three models exhibited strong performance, with ResNet 50 especially recognized for its consistent and high accuracy.

Examining the confusion matrix in conjunction with the training progress graph indicates that all three Convolutional Neural Network (CNN) models ResNet 50, AlexNet, and GoogLeNet performed effectively in classifying fish species, albeit with differing degrees of accuracy and reliability. Table 4 presents the classification outcomes for ten chosen fish species, evaluated across four anatomical regions, including whole body, head, body, and tail. This multi-segment classification method was employed to thoroughly evaluate each model's robustness in recognizing both partial and obstructed views of the fish. ResNet 50 showed a strong alignment with the dataset, accurately classifying seven out of ten fish species across all recognition segments. Misclassifications occurred for *Epinephelus areolatus, Pampus argenteus*, and *Trachinotus blochii*. ResNet 50's strong performance across several sessions highlights its advanced architecture, enabling effective learning and generalization of intricate features in the dataset. On the other hand, AlexNet managed to classify just five to six species in similar conditions accurately. Additional misclassifications were noted with *Epinephelus areolatus, Lethrinus lentjan, Pampus argenteus, Psettodes erumei,* and *Scarus ghobban*. This outcome is anticipated due to AlexNet's relatively shallow architecture and restricted feature extraction capacities, which hinder

its ability to discern subtle distinctions among closely related species.

GoogLeNet produced slightly better classification outcomes, accurately identifying seven to eight species. This improvement is linked to the model's unique Inception modules, which facilitate multi-scale feature extraction. However, while GoogLeNet occasionally outperformed ResNet 50 in terms of raw classification numbers, its results were less stable across repeated sessions. Misclassifications noted included *Epinephelus areolatus, Lethrinus lentjan*, and *Trachinotus blochii*, with only one instance occurring during the fifth training session, as indicated in Table 4. All models exhibited substantial validation accuracy during training: ResNet 50 (100%), GoogLeNet (71%), and AlexNet (97%), as illustrated by the training progress graphs. These results highlight the capability of the models to generalize to new datasets, while ResNet 50 demonstrating the highest reliability and consistency as a classifier across various input perspectives and sessions. It is crucial to understand the connection between architectural depth, training time, and classification accuracy when choosing a suitable CNN model for large-scale fish species classification endeavors.

3.2 Discussion

3.2.1 Discussion on the confusion table

A thorough examination of the confusion matrix and training progress graphs indicates that the ResNet50 model consistently outperformed both AlexNet and GoogLeNet regarding classification accuracy, robustness, and reliability. As highlighted in Table 5, ResNet50 achieved nearly perfect classification across various training sessions, showing minimal misclassifications. The model's capacity to accurately identify species from diverse anatomical perspectives, including the whole body, head, body, and tail, highlights its outstanding generalization capabilities, thereby making it particularly suitable for high-resolution fish datasets. Nevertheless, this remarkable performance is accompanied by considerable limitations; ResNet50 required a significantly longer training duration of 1,346 minutes and 27 seconds, as opposed to AlexNet's 193 minutes and GoogLeNet's 191 minutes. The extended duration of training can be attributed to the intricate architecture of the model and the heightened computational requirements. An enhancement of CPU hardware may reduce this duration (Dong *et al.*, 2023); however, the employment of GPU or TPU acceleration, coupled with architectural advancements such as model pruning, dropout techniques, and batch normalization, may present more efficient and scalable solutions (Xu *et al.*, 2022, Knausgård *et al.*, 2022).

3.2.2 Discussion on the training progress graph

The training progress graph reinforces these findings. Figure 5(a) shows that ResNet50 achieved over 90% validation accuracy within the first 50 iterations and reached perfect accuracy by epoch 1. This quick convergence suggests that the model effectively captured intricate patterns and features early in training. In contrast, AlexNet and GoogLeNet required nearly 100 iterations to surpass the 90% validation threshold, as illustrated in Figures 5(b) and 5(c). Although these models also attained high accuracy (AlexNet: 97.9%, GoogLeNet: 98.71%), the disparity in their convergence rates emphasizes the differing capacities of each model in processing complex image data.

In spite of their inherent strengths, all three models of Convolutional Neural Networks (CNNs) experienced difficulties with certain species, particularly *Pampus argenteus*, which was predominantly misclassified across all models. This persistent misclassification suggests that the current dataset may be deficient in high-quality and diverse images pertaining to this specific species. Factors contributing to these inaccuracies may include reflections of light, insufficient contrast, and morphological similarities to other species (Sun *et al.*, 2020; Hou *et al.*, 2021).

In order to address these challenges, the following enhancements are recommended: firstly, the implementation of advanced augmentation techniques, such as varying rotation angles, random cropping, contrast adjustments, and the incorporation of species-specific features, for example, body stripe coloration and tail morphology. Moreover, establishing controlled lighting during image capture and using post-processing tools to standardize brightness and reduce glare is essential. Additionally, increasing the number of images for commonly misclassified species will promote a fairer class distribution and boost visual diversity. ResNet50 exhibits superior classification performance in this study, corroborated by both quantitative metrics and visual convergence assessments. Future research should concentrate on optimizing training durations and enhancing dataset quality to improve the reliability and scalability of deep learning models employed for automated fish species classification.

4. Conclusion

In this study, three Convolutional Neural Network architectures, ResNet–50, AlexNet and GoogLeNet were compared in classifying 20 local fish species. Using a well-annotated dataset of about 20,000 underwater images captured from different an-

atomical perspectives, all proposed models achieved high classification accuracy. Among the three proposed models, ResNet–50 had the highest classification accuracy. This is likely due to the residual skip connections that help mitigate degradation in deeper networks. The minor misclassifications that did occur like in *Pampus argenteus* were likely a result of overlapping visual features and/or inconsistent image quality. Most importantly, current work shows that a well-labeled dataset is essential for a model to identify more subtle features and symmetries, rather than depending solely on its architecture. Without enough labelled data, even state-of-the-art architectures may struggle to achieve good classification performance in heterogeneous underwater environments. Results demonstrated that deep learning is a viable and robust method to classify underwater species if a sufficient labelled dataset is available. Going forward, results identified several major research gaps: (i) broader taxonomic coverage, i.e., more species and environments, (ii) optimization for real-time deployment (e.g., transfer learning, lightweight architecture), and (iii) seamless integration of classification models into fisheries monitoring systems. Resolutions to these gaps can help improve adaptability to varying ecological contexts. Looking ahead, our results highlight a few exciting areas for further research: (1) expanding our taxonomic coverage by including more species and environments; (2) improving models for real-time application using techniques like transfer learning and lightweight architectures; (3) integrating classification models into fisheries monitoring systems to make them even more effective. Collectively, these directions can improve adaptability across various ecological settings and resource-limited environments. By connecting deep learning with fisheries science, this study lays the foundation for better ecosystem monitoring, sustainable fisheries management and marine biodiversity conservation. Future efforts should aim to improve the models' capacity to generalize, transfer knowledge and perform inference directly on devices in order to better address the complex requirements of marine resource management.

Acknowledgement

Authors'Contributions

All authors have played crucial roles in the creation of the final manuscript. Their specific contributions are detailed as follows: Nur Muhammad Afiq Anang gathered the data, executed data augmentation, carried out data processing, managed the training process, and composed the manuscript. Ezmahamrul Afreen Awalludin refined the main conceptual ideas of the article, organized its structure, and offered comments and suggestions for improvement. In addition, all authors collaborated to discuss the results and enhance the final manuscript together.

Conflict of Interest

The authors hereby declare that no conflicts of interest are associated with this publication's writing.

Declaration of Artificial Intelligence (AI)

The author confirms that this manuscript is entirely composed and developed independently. However, the author has used Grammarly software to correct the grammar, sentence structure, and use of appropriate words in the manuscript. All content presented here is assumed to be their own creativity and dedication, demonstrating its originality and integrity.

Funding Information

References

Ahmad, U., Ali, M. J., Khan, F. A., Khan, A. A., Ur Rehman, A., Shahid, M. M. A., Haq, M. A., Khan, I., Alzamil, Z. S., & Alhussen, A. (2023). Large scale fish images classification and localization using transfer learning and localization aware CNN architecture. *Computer Systems Science and Engineering*, 45(2):2125-2140.

Ahmed, F., Basak, B., Chakraborty, S., Karmokar, T., Reza, A. W., Imam, O. T., & Arefin, M. S. (2023). Developing a classification CNN model to classify different types of fish. In J. Kacprzyk (Ed.), Lecture notes in networks and systems. (pp. 529-539). Switzerland: Springer International Publishing AG.

Alinsug, M. V., Reyes I. V. D., Maquiling, A. C. C., Bercades, A. J., & Deocaris, C. (2024). Skipjack tuna (*Katsuwonus pelamis*) otolith developmental stage classification using deep learning. *Philippine Journal of Fisheries*, 31(2):291-298.

Allken, V., Handegard, N.O., Rosen, S., Schreyeck, T., Mahiout, T., & Malde, K. (2019). Fish species identification using a convolutional neural net-

work trained on synthetic data. *Journal of Marine Science*, 76(1):342-349.

Ben Tamou, A., Benzinou, A., & Nasreddine, K. (2022). Live fish species classification in underwater images by using convolutional neural networks based on incremental learning with knowledge distillation loss. *Journal of Machine Learning and Knowledge Extraction*, 4(3):753-767.

Deka, J., Laskar, S., & Baklial, B. (2023). Automated freshwater fish species classification using deep CNN. *Journal of the Institution of Engineers (India): Series B*, 104(3):603-621.

Dong, G., Wang, N., Xu, T., Liang, J., Qiao, R., Yin, D., & Lin, S. (2023). Deep learning-enabled morphometric analysis for toxicity screening using zebrafish larvae. *Journal of Environmental Science and Technology*, 57(46): 18127-18138.

French, G., Mackiewicz, M., Fisher, M., Holah, H., Kilburn, R., Campbell, N., & Needle, C. (2020). Deep neural networks for analysis of fisheries surveillance video and automated monitoring of fish discards. *Journal of Marine Science*, 77(4):1340-1353.

Han, Y., Chang, Q., Ding, S., Gao, M., Zhang, B., & Li, S. (2022). Research on multiple jellyfish classification and detection based on deep learning. *Journal of Multimedia Tools and Applications*, 81(14):19429-19444.

Hasan, N., Ibrahim, S., & Azlan, A. A. (2022). Fish diseases detection using convolutional neural network (CNN). *International Journal of Nonlinear Analysis and Applications*, 13(1):1977-1984.

Hilal, A. M., Hashim, A. H. A., Mohamed, H. G., Nour, M. K., Asiri, M. M., Al-Sharafi, A. M., Othman, M., & Motwakel, A. (2023). Malicious URL classification using artificial fish swarm optimization and deep learning. *Computers, Materials and Continua*, 74(1):607-621.

Hou, Y., Canul-Ku, M., Cui, X., Hasimoto-Beltran, R., & Zhu, M. (2021). Semantic segmentation of vertebrate microfossils from computed tomography data using a deep learning approach. *Journal of Micropalaeontology*, 40(2):163-173.

Iqbal, M. A., Wang, Z., Ali, Z. A., & Riaz, S. (2021). Automatic fish species classification using deep convolutional neural networks. *Wireless Personal Communications*, 116(2):1043-1053.

Iqtait, M., Alqaryouti, M. H., Sadeq, A. E., Aburomman, A., Baniata, M., Mustafa, Z., & Chan, H. Y. (2024). Enhanced fish species detection and classification using a novel deep learning approach. *International Journal of Advanced Computer Science and Applications*, 15(10):1063-1067.

Ismail, H., Ayob, A. F. M., Muslim, A. M. S. M., & Zulkifli, M. F. R. (2021). Convolutional neural network architectures performance evaluation for fish species classification. *Journal of Sustainability Science and Management*, 16(5):124-139.

Kaya, V., Akgül, İ., & Tanir, Ö. Z. (2023). IsVoNet8: A proposed deep learning model for classification of some fish species. *Journal of Agricultural Sciences*, 29(1):298-307.

Knausgård, K. M., Wiklund, A., Sørdalen, T. K., Halvorsen, K. T., Kleiven, A. R., Jiao, L., & Goodwin, M. (2022). Temperate fish detection and classification: a deep learning-based approach. *Applied Intelligence*, 52(6):6988-7001.

Li, B., Liu, Y., & Duan, Q. (2024). T-KD: two-tier knowledge distillation for a lightweight underwater fish species classification model. *Aquaculture International Journal*, 3(2):3107-3128.

Lan, X., Bai, J., Li, M., & Li, J. (2020). Fish image classification using deep convolutional neural network. *ACM International Conference Proceeding Series*, 18-22.

Okafor, E., Schomaker, L., & Wiering, M.A. (2018). An analysis of rotation matrix and color constancy data augmentation in classifying images of animals. *Journal of Information and Telecommunication*, 2(4):465-491.

Peddina, K., & Mandava, A. K. (2025). The intelligent object detection framework for detecting fish from underwater image. *International Journal of Communication Networks and Distributed Systems*, 31(1):63-88.

Prasetyo, E., Suciati, N., & Fatichah, C. (2022). Multi-level residual network VGGNet for fish species classification. *Journal of King Saud University - Computer and Information Sciences*, 34(8):5286-5295.

Rauf, H. T., Lali, M. I. U., Zahoor, S., Shah, S. Z. H., Rehman, A. U., & Bukhari, S. A. C. (2019). Visual features based automated identification of fish species using deep convolutional neural networks. *Computers and Electronics in Agri-*

culture, 167(12):1-10.

Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural Computation*, 29(9):2352-2449.

Saleh, A., Sheaves, M., & Azghadi, M. R. (2022). Computer vision and deep learning for fish classification in underwater habitats: A survey. *Fish and Fisheries*, 23(4):977-999.

Song, J., Gao, S., Zhu, Y., & Ma, C. (2019). A survey of remote sensing image classification based on CNNs. *Big Earth Data*, 3(3):232-254.

Sun, M., Yang, X., & Xie, Y. (2020). Deep learning in aquaculture: A review. *Journal of Computer*, 31(1):294-319.

Veluswami, J. R. S., & Panneerselvam, N. (2022). Multi-species fish identification using hybrid deep CNN with refined squeeze and excitation architecture. *Journal of Aquatic Sciences and Engineering*, 37(4):220-228.

Wang, P., Fan, E., & Wang, P. (2021). Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognition Letters*, 141(1):61-67.

Wang, F., Zheng, J., Zeng, J., Zhong, X., & Li, Z. (2023). S2F-YOLO: An optimized object detection technique for improving fish classification. *Journal of Internet Technology*, 24(6):1211-1220.

Xu, M., Yang, H., Liu, G., Tang, Y., & Li, W. (2022). In silico prediction of chemical aquatic toxicity by multiple machine learning and deep learning approaches. *Journal of Applied Toxicology*, 42(11):1766-1776.

Yang, C., Zhou, P., Wang, C. S., Fu, G. Y., Xu, X. W., Niu, Z., Zhu, L., Yuan, Y., Shen, H. Bin, & Pan, X. (2024). FishAI: Automated hierarchical marine fish image classification with vision transformer. *Engineering Reports*, 6(12):1-16.

Yasin, E. T., Ozkan, I. A., & Koklu, M. (2023). Detection of fish freshness using artificial intelligence methods. *Journal of European Food Research and Technology*, 249(8):1979-1990.

Zhai, J., Han, L., Xiao, Y., Yan, M., Wang, Y., & Wang, X. (2023). Few-shot fine-grained fish species classification via sandwich attention CovaMNet. *Frontiers in Marine Science*, 10:1149186.

Zhang, S., Liu, W., Zhu, Y., Han, W., Huang, Y., & Li, J. (2022). Research on fish identification in tropical waters under unconstrained environment based on transfer learning. *Earth Science Informatics*, 15(2):1155-1166.

Zheng, S., Wang, R., Zheng, S., Wang, L., & Jiang, H. (2024). Adaptive density guided network with CNN and transformer for underwater fish counting. *Journal of King Saud University - Computer and Information Sciences*, 36(6):1-13.

Zhou, W., Wang, H., & Wan, Z. (2022). Ore image classification based on improved CNN. *Computers and Electrical Engineering*, 99(4):1-9.

Zhou, Z., Yang, X., Ji, H., & Zhu, Z. (2023). Improving the classification accuracy of fishes and invertebrates using residual convolutional neural networks. *ICES Journal of Marine Science*, 80(5):1256-1266.