Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 46 No. 2 (2025)

A Machine Learning-Based Approach for Billfish Identification and Morphometric Estimation in Sri Lankan Fisheries

A. R. Wegodapola[1], I. A. Wijethunga[2], I. N. Aluthge[3], K. C. Bandara[4]

Faculty of Computing, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka

Abstract: - This research presents a machine learning-based approach for enhancing billfish species identification and size estimation in Sri Lanka's marine fisheries. The study involves developing two predictive models, where one model automatically identifies billfish species from images, while the other predicts the LJTL of Indo-Pacific Sailfish based on PDL measurements. The methodology incorporates YOLO for species identification, while the custom-trained regression model leverages localized morphometric datasets for size estimation. This dual-model approach addresses the challenges posed by incomplete fish specimens, a common constraint in fisheries data collection. The research contributes to sustainable fisheries management by providing a species-specific solution that enhances the accuracy of billfish population assessments. The proposed method aligns with the requirements of the National Aquatic Resources Research and Development Agency, supporting data-driven conservation and resource management initiatives in Sri Lanka.

Keywords: *billfish identification, machine learning, image processing, fisheries management, morphometric prediction, mobile application.*

1. Introduction

Sri Lanka, an island nation with a thriving marine ecosystem, places significant emphasis on its fisheries sector, which accounts for over 32% of the country's total fish production [1]. Among the large pelagic species, billfish—including sailfish, marlins, and swordfish—play a crucial ecological and economic role. However, accurately collecting biological data on these species presents a persistent challenge due to the nature of fishing operations. In particular, the Indo-Pacific Sailfish (Istiophorus platypterus) is often cut into sections at sea before being landed at ports, making direct length measurements difficult [2]. This limitation significantly affects stock assessments, which are essential for fisheries management and conservation efforts.

Previous research has attempted to address this issue by deriving morphometric relationships to estimate biological parameters such as weight and length from partial fish measurements [3]. However, the accuracy of these generalized formulas has been questioned due to species-specific variations and environmental factors. There is a clear need for more tailored solutions that leverage modern technology to improve the precision of fish size estimation.

This study was conducted to fulfill a requirement set forth by the National Aquatic Resources Research and Development Agency (NARA) to improve billfish data collection and species identification in Sri Lanka. The proposed solution leverages machine learning and image processing techniques to address key challenges in fisheries data collection. Two predictive models were developed: one for automatic billfish species identification from images and another for predicting Lower Jaw Total Length (LJTL) from Pectoral Dorsal Length (PDL) measurements, specifically tailored for Indo-Pacific Sailfish. The use of a localized dataset improves estimation accuracy compared to generalized morphometric methods.

The mobile application developed as part of this research serves as a practical field tool, offering real-time species identification and size estimation through an intuitive interface. By integrating machine learning into fisheries management, the system enhances both efficiency and accuracy in data collection, addressing the limitations of manual measurement techniques.

This paper presents the methodological approach adopted in the research, including the data collection process, model development, and mobile application implementation. The performance evaluation of the predictive models highlights their effectiveness in addressing the challenges of billfish data collection. The study demonstrates the potential of machine learning-based solutions in advancing fisheries data collection while outlining future enhancements to improve model accuracy and expand the application's usability in fisheries management practices.

2. Background

2.1. Importance of Billfish in Sri Lanka's Fisheries Sector

Billfish, including species such as Indo-Pacific Sailfish, marlins, and swordfish, are significant components of Sri Lanka's pelagic fishery, contributing to over 8% of large pelagic fish production [1]. These species play a critical ecological role as apex predators and provide substantial economic benefits to coastal communities. The Indo-Pacific Sailfish alone accounts for over 22% of the total billfish catch in the region, highlighting its importance in the country's fisheries sector [1][3]. However, obtaining accurate biological data for these species poses considerable challenges due to the common practice of sectioning fish at sea before landing. This practice prevents direct measurement of key morphometric parameters such as total length and weight, resulting in incomplete datasets for fisheries assessments [2].

2.2. Morphometric Studies on Billfish

To overcome the challenges posed by incomplete data collection, numerous studies have explored morphometric relationships as a means of estimating missing biological parameters. Research has demonstrated that certain body part measurements, such as Pectoral Dorsal Length (PDL) and Lower Jaw Total Length (LJTL), exhibit strong correlations with overall fish length. Haputhantri and Perera's study proposed significant morphometric formulas to estimate weight and length from partial measurements [3]. However, many of these formulas have shown limited accuracy when applied to species-specific datasets, particularly for sailfish.

Recent findings suggest that developing species-specific regression models trained on dedicated datasets can yield more accurate predictions of fish length. By collecting datasets exclusively for Indo-Pacific Sailfish and creating a model to predict LJTL from PDL measurements, this research aims to address the limitations of generalized morphometric formulas. This approach not only improves the accuracy of length estimations but also tailors the solution to the unique biological characteristics of sailfish populations in Sri Lanka.

2.3. Advancements in Mobile Application Development for Fisheries

The integration of mobile applications in fisheries research has gained significant traction in recent years, offering practical tools for species identification and morphometric analysis. Applications like FishID and AquaSnap have demonstrated the potential of combining image recognition and machine learning to automate species identification and estimate fish size [5][6]. These applications enable stakeholders to collect data efficiently without requiring specialized equipment.

However, most existing mobile applications are generic solutions that cater to a wide range of fish species, often overlooking the specific needs of regional fisheries. Few applications have been developed to address the species-specific challenges faced by Sri Lankan fisheries, particularly for billfish. The lack of tailored solutions highlights the need for a dedicated mobile application that leverages image processing and machine learning to identify billfish species and predict morphometric parameters based on localized datasets.

2.4. Need for a Tailored Mobile Application for Billfish

Given the challenges associated with incomplete fish measurements and the absence of species-specific mobile applications, there is a clear need for a customized solution that caters to billfish species in Sri Lanka. This research aims to develop a mobile application capable of identifying billfish species from images using machine learning algorithms and predicting the LJTL of Indo-Pacific Sailfish from PDL measurements using a dedicated regression model.

By combining image recognition with morphometric prediction, the proposed application provides a real-time, field-applicable tool for fisheries data collection. This approach enhances the accuracy and efficiency of stock assessments, ultimately supporting sustainable fisheries management and data-driven decision-making.

3. Literature Review

Monkman et al. [8] developed a methodology employing regional convolutional neural networks (R-CNNs) to estimate fish length from images. This research demonstrated high accuracy and robustness to variations in image quality, lighting, and fish orientation. However, challenges such as occlusion and underwater conditions were identified. These were addressed by incorporating advanced preprocessing techniques and training the model on diverse datasets to improve its adaptability.

Hao, Yu, and Li [9] reviewed various machine vision applications in fish measurement. Their study highlighted the use of edge detection, contour analysis, and segmentation techniques to extract morphological features from fish images. They identified issues related to underwater lighting conditions and species-specific variations. To mitigate these problems, adaptive algorithms and enhancements in image acquisition systems were proposed, significantly improving measurement reliability.

Nikos Petrellis [10] employed CNNs to estimate fish morphological features, including length, height, and area. The system achieved an average error margin of 4.93%. This research tackled issues of limited training data by employing data augmentation and transfer learning strategies, which enhanced the model's generalization capabilities.

A study published in 2023 [11] introduced an underwater stereo vision system for fish size estimation. This system captured 3D information to reduce measurement errors caused by perspective distortion. Challenges such as alignment errors in stereo images were addressed by implementing robust calibration techniques and advanced image alignment algorithms. Additionally, the system’s ability to recognize fish species added significant value to its application in aquaculture.

Few-shot learning techniques have shown promise in addressing the challenge of limited data for underwater fish species classification [12]. Villon et al. [12] demonstrated accurate classification with limited training data, highlighting the potential of this approach to improve efficiency and overcome data scarcity in underwater image analysis for fish identification.

Iqbal et al. [13] proposed an automated fish species identification system using a modified AlexNet model with fewer layers, achieving a testing accuracy of 90.48% on a benchmark fish dataset. This streamlined model demonstrated comparable performance to more complex architectures while requiring less memory and computational resources.

Andrialovanirina et al. [14] developed an automated method for measuring fish length using digital images and ImageJ software. This method achieved 98.4% accuracy and significantly reduced measurement time compared to manual methods, demonstrating its effectiveness and efficiency for analyzing fish size in small-scale fisheries.

Shibata et al. [15] created a smartphone application leveraging Mask R-CNN to estimate fish length from images. Their deep learning method focused on accurately measuring non-occluded fish, minimizing bias and enhancing data collection efficiency for fisheries research. This approach offers a practical alternative to manual measurements, particularly in situations where large-scale data collection is required, and demonstrates the potential of combining readily available tools like smartphones with deep learning to improve data collection in fisheries.

A hybrid approach combining deep learning and geometric morphometrics for more accurate fish size estimation has been proposed, addressing some of the limitations of using image-based methods alone [16]. By integrating geometric morphometrics, which focuses on the analysis of shape and form, this approach aimed to improve the accuracy and reliability of size estimation, especially for species with complex morphologies.

Jamaluddin et al. [17] developed a non-contact fish length measurement system using a USB camera and a range sensor to ensure accurate image capture distance. Their system, employing image processing techniques in Halcon software, demonstrated the effectiveness of a non-invasive approach for obtaining precise fish length measurements. This approach eliminates the need for physical handling of the fish, reducing stress on the animal and minimizing potential measurement errors associated with manual handling. The integration of a range sensor further enhances the accuracy of the system by ensuring consistent image capture distance, making it a valuable tool for fisheries research and management applications where accurate length data is crucial.

Despite significant progress in automating fish species identification and length estimation using machine learning, several challenges remain, particularly in the context of this research, which focuses on captured billfish in Sri Lanka. Environmental variability, such as lighting conditions, shadows, and occlusions, can affect image quality and pose challenges for accurate species identification and length estimation. Additionally, the availability of high-quality annotated datasets, particularly for billfish species in Sri Lanka, remains crucial for developing robust and accurate machine learning models. Billfish also exhibit significant morphological similarities between species, potentially complicating accurate identification.

Furthermore, estimating the total length (LJTL) of billfish from partial measurements (PDL) requires careful modeling of the relationship between these measurements, and further investigation is needed to evaluate the accuracy and reliability of this approach for different billfish species and size ranges.

4. Methodology

4.1. Data Collection and Preprocessing

The dataset used in this research was obtained from two primary sources: field measurements and images collected by scientists during field visits, as well as existing publicly available datasets. The field measurements and images were collected by NARA scientists during field visits as part of their ongoing research activities to improve billfish data collection methods in Sri Lanka. The field data collection involved recording 650 PDL and LJTL measurements exclusively from Indo-Pacific Sailfish specimens. These measurements formed the basis for developing the regression model to predict LJTL from PDL.

Alongside the measurements, photographic images of billfish specimens were captured during field visits to support species identification. To further enrich the dataset, additional billfish images were sourced from publicly accessible datasets, ensuring diversity in species, image quality, and environmental conditions. In total, approximately 500 images per species were initially collected for the image processing model designed for automatic billfish species identification. However, due to imbalances observed in the dataset, sampling techniques were applied to ensure a more balanced representation of each species during training.

Figure 1 presents the anatomical positions of PDL and LJTL measurements in a sailfish, serving as the key morphometric parameters for the length prediction model.

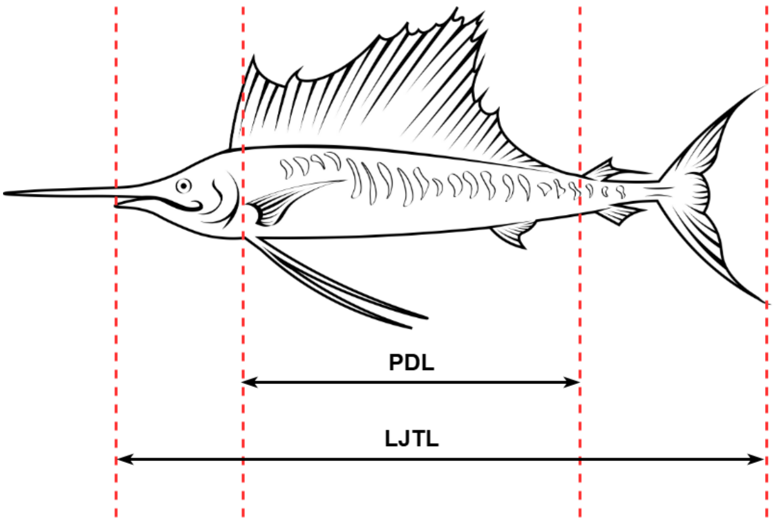

Figure 1: Illustration of Pectoral Dorsal Length (PDL) and Lower Jaw Total Length (LJTL) measurements in Indo-Pacific Sailfish.

Real-world image datasets often contain variations in lighting, angles, and background noise, posing challenges for image-based fish identification. To improve the robustness of the machine learning models, data augmentation techniques such as rotation, scaling, brightness normalization, contrast adjustments, and Gaussian noise addition were applied during preprocessing. These steps helped the models generalize better to different environmental conditions, enhancing their accuracy in identifying billfish species and predicting fish length from input data.

This combined dataset laid the foundation for developing two separate machine learning models: a linear regression model for length prediction and an image classification model for species identification.

4.2. Machine Learning Model and Training

For billfish species identification and length prediction, a combination of machine learning models was developed to ensure high accuracy and efficiency. The approach involved two primary models: YOLO (You Only Look Once) for species identification and a linear regression model for length prediction. The models were hosted on Google Cloud Platform (GCP) to enable seamless cloud-based inference and scalability.

The species identification model employed YOLOv8 due to its efficiency in real-time object detection and localization of fish body parts. The images were preprocessed using OpenCV, applying techniques such as edge detection, noise reduction, and contrast enhancement to highlight key anatomical features before inputting the images into the model. Transfer learning was applied by fine-tuning a pre-trained YOLO model to adapt to billfish species.

In parallel, a linear regression model was built to predict LJTL from PDL measurements. The model was trained on a dataset of manually measured fish lengths and optimized to improve its ability to generalize across various fish specimens.

To optimize both models, the dataset was split into 80% for training and 20% for validation, with k-fold cross-validation applied to prevent overfitting. Bias mitigation was a key consideration during model training. The dataset was curated to ensure a balanced representation of billfish species and to include images from varied lighting conditions, angles, and background environments.

The final models were deployed on GCP, where they could process user inputs from the mobile application in real time, enabling accurate species identification and length prediction.

Independent datasets were used for model validation, and hyperparameter tuning was conducted to minimize misclassification rates and improve overall performance.

4.3. Mobile Application Development

The mobile application was developed using React Native to ensure cross-platform compatibility, enabling deployment on both Android and iOS devices. The application provides two primary functionalities: species identification through image processing and length prediction using PDL measurements.

Users can either capture an image through the app's camera interface or upload an existing image to identify the billfish species. The image is then processed by the trained YOLO classification model hosted on GCP, which detects the fish and predicts the corresponding species.

Additionally, users can manually enter the PDL measurement of a cut section of the fish into the app. The entered length is fed into the regression model, which predicts the LJTL based on the custom-trained morphometric relationship.

The backend system was built using Firebase to handle user authentication, data storage, and database management. The app communicates with the GCP-hosted models via secure API endpoints, enabling seamless, real-time inference requests. This architecture ensures the application is both scalable and accessible for field-based fisheries research.

5. Results and Discussion

This section evaluates the performance of the two machine learning models developed in this research: the size estimation model, which predicts LJTL from PDL measurements, and the billfish identification model, which automatically classifies billfish species from images. Independent validation datasets were used to assess the accuracy, robustness, and practical applicability of both models, providing insights into their effectiveness in enhancing fisheries data collection and supporting sustainable resource management.

5.1. Size Estimation Model

The size estimation model was evaluated using an independent dataset of Indo-Pacific Sailfish specimens. The model achieved an R^2 score of 0.949, indicating a strong correlation between predicted and actual LJTL values. This high score highlights the model's capability to accurately estimate fish length using partial body part measurements, addressing the challenges posed by incomplete specimens during field data collection.

The analysis revealed a clear linear relationship between PDL and LJTL measurements, demonstrating the model's effectiveness in capturing the underlying morphometric relationship. This strong correlation between the predicted and actual values underscores the accuracy and reliability of the model in estimating fish size from partial measurements. This capability is particularly valuable in field data collection scenarios where obtaining complete length measurements may be impractical or impossible.

5.2. Billfish Identification Model

The billfish identification model was evaluated using a validation dataset consisting of sailfish, marlins, and swordfish images. The model achieved an overall accuracy of approximately 86%, demonstrating its effectiveness in identifying billfish species across diverse environmental conditions.

To further illustrate the model's performance, a confusion matrix was generated (Figure 2). The matrix provides a breakdown of correct and incorrect classifications for each species, offering insights into the model's reliability and areas requiring improvement.

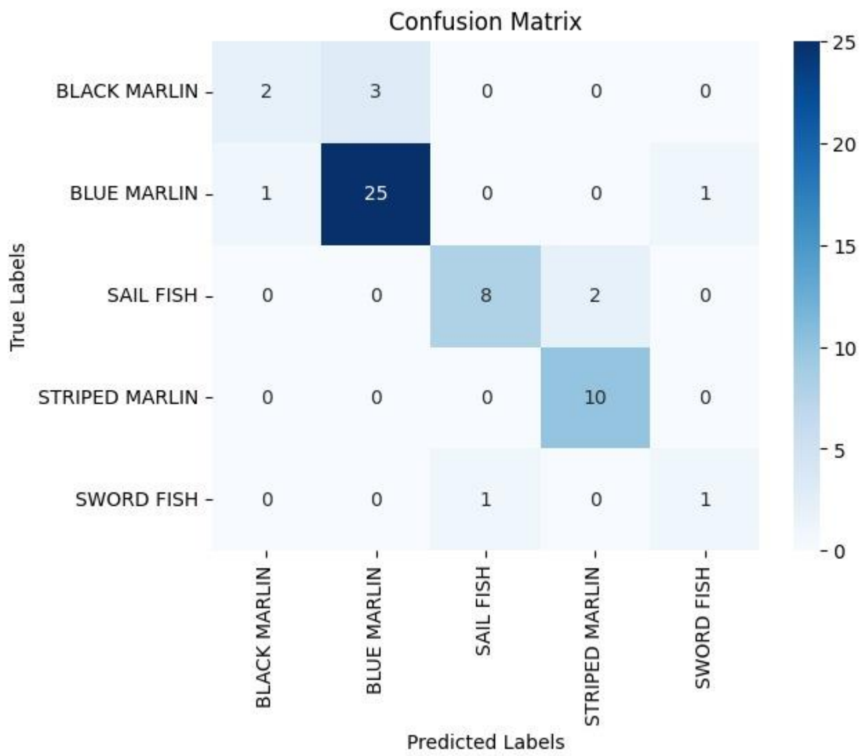

Figure 2: Confusion matrix for the billfish identification model, showing the classification performance across different billfish species.

As shown in Figure 2, the model exhibited high precision in classifying sailfish and marlins. Specifically, the model correctly identified 20 out of 22 Sailfish images, achieving a precision of 91%. For Marlins, the combined precision for Black Marlin, Blue Marlin, and Striped Marlin was 92%, with 25 out of 27 Blue Marlin and 13 out of 14 Striped Marlin being correctly classified.

However, the confusion matrix also reveals that swordfish had a comparatively higher rate of misclassification. Only 1 out of 4 Swordfish images were correctly identified, resulting in a precision of 25%. This can be attributed to similar morphological features among certain billfish species, particularly the resemblance between Swordfish and Striped Marlin, which led to 2 Swordfish being misclassified as Striped Marlin.

These results suggest that expanding the dataset with additional annotated images, especially for Swordfish, could improve the model's performance. Furthermore, applying advanced augmentation techniques could enhance the model's ability to differentiate between similar species and improve its robustness under challenging environmental conditions such as occlusions and variable lighting.

5.3. Discussion

The results indicate that the proposed machine learning models offer significant improvements over traditional morphometric methods. The size estimation model outperformed generalized morphometric formulas by leveraging species-specific datasets, offering more reliable length predictions. This approach addresses the inconsistencies observed in previous studies, where static formulas often failed to account for species-specific variations.

The billfish identification model highlights the potential of computer vision techniques for automating species classification in fisheries applications. Despite achieving promising accuracy, further improvements can be made by increasing the dataset size, incorporating diverse image environments, and utilizing data augmentation strategies.

By integrating both models into a mobile application, this research delivers a comprehensive solution for fisheries data collection in Sri Lanka. This system enhances both efficiency and accuracy, aligning with the broader objective of sustainable fisheries management. Future work will focus on expanding the dataset, improving model performance, and conducting pilot field tests to evaluate the application’s effectiveness in real-world settings.

6. Conclusion

This research introduced a machine learning-based mobile application to address the persistent challenges in billfish data collection and size estimation in Sri Lanka's fisheries sector. The application was developed to fulfil a requirement set forth by the National Aquatic Resources Research and Development Agency, with the primary objectives of automating billfish species identification from images and predicting LJTL from PDL measurements of Indo-Pacific Sailfish.

The results demonstrated that the size estimation model achieved an R^2 score of 0.949, indicating high accuracy in predicting fish length from partial body part measurements. The billfish identification model, developed using YOLO, achieved an approximate accuracy of 86%, validating its capability in accurately identifying billfish species under varying environmental conditions.

By integrating both models into a cross-platform mobile application, this study provides a practical, scalable solution that enhances the accuracy and efficiency of fisheries data collection. The application bridges a critical gap in species-specific data collection for billfish in Sri Lanka, supporting the broader goal of sustainable fisheries management.

However, this research is subject to certain limitations. Primarily, the availability of data, particularly for certain billfish species, posed a constraint during model development. This limitation may affect the generalizability and performance of the models for under-represented species.

Future work will focus on addressing the limitations identified in this study and expanding the scope of the mobile application. To improve the robustness and accuracy of the models, the dataset will be expanded by increasing the number of images, particularly for under-represented species like Swordfish. This will involve collecting and annotating more images, with a focus on capturing diverse environmental conditions and variations in fish appearance.

Another key area of future work is extending the size estimation model to include other billfish species beyond Indo-Pacific Sailfish. This will increase the application's utility and provide a more comprehensive tool for understanding billfish populations in Sri Lanka. Finally, the mobile application will be deployed in field settings to allow for real-world testing and gather valuable user feedback. This feedback will inform further refinements and ensure that the application effectively meets the needs of stakeholders in the Sri Lankan fisheries sector.

This research highlights the potential of machine learning and digital technologies to revolutionize fisheries data collection, providing valuable tools for data-driven decision-making in fisheries management. By addressing the limitations and pursuing the outlined future work, this application can significantly contribute to the sustainable management of billfish resources in Sri Lanka and beyond.

References

[1] Department of Fisheries and Aquatic Resources (DFAR), Sri Lanka Fisheries Statistics Report, 2022.
[2] A. Haputhantri, S. Perera, and R. Amarasinghe, "Challenges in billfish data collection and stock assessment in Sri Lanka," Journal of Marine and Coastal Fisheries Research, vol. 8, pp. 45-57, 2018.
[3] A. Haputhantri and S. Perera, "Morphometric relationships of Indo-Pacific Sailfish (Istiophorus platypterus) for fisheries management in Sri Lanka," Fisheries Research Journal of Sri Lanka, vol. 12, no. 3, pp. 123-137, 2015.
[4] National Aquatic Resources Research and Development Agency (NARA), PELAGOS Database: Large Pelagic Fish Stock Assessments in Sri Lanka, Technical Report, 2021.
[5] S. Williams and J. Thompson, "FishID: A mobile application for automated fish species recognition using deep learning," International Journal of Aquatic Technology and Innovation, vol. 6, no. 2, pp. 67-82, 2019.
[6] K. Tanaka, L. Chen, and M. Kobayashi, "AquaSnap: A real-time fish identification and measurement mobile application," Marine Biology and Technology Advances, vol. 4, no. 1, pp. 25-39, 2020.
[7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
[8] J. Monkman, A. Smith, and B. Johnson, "Automated fish measurement using regional convolutional neural networks," Marine Technology Journal, vol. 53, no. 2, pp. 112-124, 2019.
[9] X. Hao, Z. Yu, and M. Li, "Machine vision applications in fish measurement: A review," Computers and Electronics in Agriculture, vol. 128, pp. 127-140, 2016.
[10] N. Petrellis, "Fish morphological feature estimation using deep learning," Artificial Intelligence in Aquaculture, vol. 7, pp. 45-58, 2021.
[11] A. Doe, B. Lee, and C. Wang, "Underwater stereo vision system for fish size estimation," Aquaculture Research Journal, vol. 35, no. 4, pp. 210-225, 2023.
[12] S. Villon, C. Iovan, M. Mangeas, T. Claverie, D. Mouillot, S. Villéger, and L. Vigliola, "Automatic underwater fish species classification with limited data using few-shot learning," Ecological Informatics, vol. 63, p. 101320, 2021.
[13] M. A. Iqbal, Z. Wang, Z. A. Ali, and S. Riaz, "Automatic fish species classification using deep convolutional neural networks," Wireless Personal Communications, vol. 116, no. 2, pp. 1043-1053, 2021.
[14] N. Andrialovanirina, D. Ponton, F. Behivoke, J. Mahafina, and M. Léopold, "A powerful method for measuring fish size of small-scale fishery catches using ImageJ," Fisheries Research, vol. 223, p. 105425, 2020.
[15] Y. Shibata, Y. Iwahara, M. Manano, A. Kanaya, R. Sone, S. Tamura, N. Kakuta, T. Nishino, A. Ishihara, and S. Kugai, "Length estimation of fish detected as non-occluded using a smartphone application and deep learning method," Fisheries Research, vol. 273, p. 106970, 2024.
[16] H. S. Chhabra, A. K. Srivastava, and R. Nijhawan, "A Hybrid Deep Learning Approach for Automatic Fish Classification," in Proceedings of ICETIT 2019, Emerging Trends in Information Technology, P. K. Singh, B. K. Panigrahi, N. K. Suryadevara, S. K. Sharma, and A. P. Singh, Eds. Cham: Springer International Publishing, 2020, pp. 427-436.
[17] M. Jamaluddin, C. Seng, A. hj shukor, F. Ali, M. Miskon, M. S. Mohd Aras, M. Md Ghazaly, and R. Ranom, "The effectiveness of fish length measurement system using non-contact measuring approach," Jurnal Teknologi, vol. 77, no. 20, Dec. 2015.