Frontiers in Engineering and Informatics
ISSN: 3049-3412 (Online)
Volume No: 01 – Issue No: 04 – 10 December 2025

MARLIN-NET: A DEEP LEARNING FRAMEWORK FOR UNDERWATER MARINE SPECIES RECOGNITION

Abhishek S[1*], Amreth S[2], Amrutha A[3], Gunalan T R[4], Agash K[5], Dharanish S[6]
[1*,2,3,4,5,6]Department of Computer Science and Engineering,
Sri Krishna College of Engineering And Technology,
Coimbatore, Tamil Nadu, India
[1*]abhicodemaker@gmail.com, [2]amrethskcet@gmail.com, [3]amruthaacse@gmail.com,
[4]gunalantrengi@gmail.com, [5]agaash1809@skcet.ac.in, [6]dharanishs24@gmail.com
Corresponding Author E-mail ID: [1*]abhicodemaker@gmail.com

Abstract

This paper presents a deep learning structure for identifying underwater marine species and classification using an integrated feature fusion architecture that leverages multiple CNN architectures. It enhances underwater image recognition with the proposed system by fusing the features from MobileNetV2, EfficientNetB0, and InceptionV3. This collaborative way of extracting features captures both fine-grained textures and high-level spatial representations, dealing with several visual issues related to underwater turbidity, low illumination, and color distortion of pictures. The model is trained and tested using a balanced dataset of marine species. Generalization is improved with resizing, normalizing, and augmentation of images. Some performance metrics involve accuracy, F1-score, recall, and precision in classification.The results have shown significant elevations compared with individual CNN models and thus prove the benefits of feature fusion in complex underwater scenarios. We would like to further implement the developed framework using PyTorch on Google Colab to enable efficient computation and easy scalability. In general, the system contributes to ocean biodiversity research and ecological conservation, making real-time monitoring of marine life possible.

Keywords: Marine Species Classification, Deep Learning, Feature Fusion, CNN, Underwater Image Processing

1. INTRODUCTION

Correct identification and categorization of marine life via underwater imagery are the key applications of ocean biology, biodiversity monitoring, ecological and environmental conservation. Underwater images are affected by poor lighting, turbidity, and color distortion due to scattering, which lowers the image quality and makes species recognition a challenging task [1]. Traditional underwater image analysis methods that include rule-based or manual identification are extremely slow, inconsistent, and incapable of processing massive underwater datasets [2].

There are many limitations, and in order to overcome them, deep learning-based computer vision, especially methods with Convolutional Neural Networks [3], have turned into a potent weapon for the automated marine species classification system [4]. In contrast to traditional image processing, which uses manually created features like color histograms or edge descriptors, CNNs learn spatial hierarchies automatically and they get deep features come straight from the raw image data [5]. As a result, they can achieve higher accuracy and their models are more flexible. However, single CNN architectures often do

DoI: https://doi.org/10.63949/crinfo.v1i4.002

not have the generalization ability for different underwater conditions and also the capacity to recognize both fine-grained textures [6] and large-scale spatial patterns in the image [7].

Hence, the proposed system has a joint feature extraction framework that is capable of different pre-trained CNN architectures such such as MobileNetV2 [8], EfficientNetB0 [9], and InceptionV3 [10].The combined architecture takes the best of the lightweight models, the traditional models provide the profundity, and the inception modules give the multi-scale learning capability which altogether can be a very effective way to represent underwater images for the purpose of image classification to both the correctness and the robustness of the system have been increased further.

The work is achieved in PyTorch via the Google Colab platform.The dataset is meticulously arranged so that it can cover image enhancement, resizing, normalizing, and augmenting aimed at generalization improvement. The integrated framework propels the present marine vision systems to the following level by providing a solution that is not only lightweight and accurate but also real-time and capable of classifying marine species. The system is the basis for the next generation of underwater scientific research, ecosystem monitoring, and sustainable ocean resource management, all in line with the global conservation agenda for instance, UN Sustainable Development Goal 14: Life Below water

2. LITERATURE SURVEY

J. Zhong et al. [11] invented a real-time YOLO deep learning network-based system to identify the animals living in marine ecosystems. Their system was capable of making detections at a very high speed and, thus, it could work quite efficiently in the real world underwater environment. By using YOLO's region-based convolutional approach the research could demonstrate that the technology used in their work was able to quite precisely locate and identify the objects under the condition of very clear water. As a matter of fact, the device had limited generalization when it was utilized to recognize non-coral reef datasets and in turbid or low-light conditions where the quality of images is bad.

A thorough analysis of deep learning-based fish aquaculture object detection technologies has been provided by H. Liu et al. [12]. The study covered a range of CNN and transfer learning techniques that could be used to automate fish behavior tracking and monitoring. Their analysis emphasized how advanced deep learning methods can improve aquaculture management's efficiency. However, the authors also highlight the limitations, including the quality, the lack of standard datasets, and the real time performance variation of various underwater conditions.

W. Rahman et al. [13] developed CNN model to classify freshwater and saltwater fish species. By employing various species identification features simultaneously, their method was able to achieve a high degree of correctness, thus, the classification is accurate regardless of the source. The only limitation of the model is that it would need a large, well-annotated dataset to maintain its performance, which makes it less practical for marine studies with limited resources.

M. Hamzaoui et al. [14] has come up with an enhanced deep learning model to identify underwater species in an aquaculture setting. The architecture of their network utilized feature extraction from the low-contrast images and also demonstrated that it was very resistant to underwater noise and distortions. Their model claimed higher recognition accuracy than the regular CNNs but was very demanding in terms of computing power and quite limited in terms of scalability for larger marine datasets. Y. Lin et al. [15] developed a large-scale fish dataset tailored to a specific domain to increase deep learning models'

DoI: https://doi.org/10.63949/crinfo.v1i4.002

generalization capacity for fish recognition. By employing a very deep Inception-based network, their system was also able to realize quicker recognition with higher accuracy for a great number of underwater images. The model recall was very high, but, unfortunately, the dependence on a particular dataset limited the model's flexibility to different marine environments or novel species.

V. Pagire et al. [16] put into practice a structure using deep learning for the detection of underwater species that could function in real-time and be very accurate for a large number of different species. Their method was effective in dealing with the images degraded by noise and low light, thus, proving the potential of CNNs in marine life monitoring. However, the approach was quite demanding in terms of the required hardware, and as the authors note that performance deteriorated with the increase of the depth of the murkiness of the water, they propose the next step to be looking for a more flexible way for the preprocessing of input images.

After reading these articles, one would acknowledge that CNN architectures have individually accomplished significant results in marine species detection, yet most of them suffer from problems of bad underwater visibility, domain adaptation, and computational efficiency. The mixing of MobileNetV2, EfficientNetB0, and InceptionV3 through a feature fusion framework proposed in this project is aimed at eliminating these defects, hence, yielding a more precise, scalable, and less resource- intensive way of underwater species recognition in different environmental conditions.

Identifying marine species is fundamental to the understanding of the earth's biodiversity, the well being of the ecosystem, as well as the sustainable utilization of ocean resources. However, a big problem remains as correct classification, in which underwater images are degraded due to the change

in illumination, color distortion, and turbidity, is concerned. The methods of image analysis and classification used in this context have deteriorated results because they cannot handle these problems altogether and thus they often give wrong results and have low adaptability to real marine environments. Conversely, a deep learning model, primarily a CNN,for automation and accuracy enhancement is the landmark of the present study in underwater image analysis and sea-life classification. The research work combining single CNN architectures mission VGG16, InceptionV3, or ResNet has been very fruitful; nevertheless, each one of them still is limited within its boundaries. A good case example is that the deeper networks like VGG16 require a lot of computational resources while the lighter models such as MobileNetV2 may not be able to extract the fine-detail features at a satisfactory level. Besides, the majority of the current solutions utilize single scale feature representations which limit them in exploring the textures and shapes of the marine organisms that alter in the underwater environment.

Recent papers have proposed multi-model fusion and ensemble CNN architectures that fuse the characteristics of the different structures to obtain the complementary information for solving these issues. The hybrids exploit the facets of feature diversity, robustness, and the final accuracy of classification. Still, most of the currently proposed ensemble methods cannot efficiently perform large scale real-time monitoring tasks while maintaining the strike a balance between performance and computational cost.

As a result, the proposed combined feature extraction framework by combining the different layers from MobileNetV2, EfficientNetB0 and InceptionV3 surpasses the restrictions of these three individual single-layer networks. Together with the image's textural and global structural patterns, this combination enables the capture of spatial features at various scales. PyTorch on Google Colab is used to implement

DoI: https://doi.org/10.63949/crinfo.v1i4.002

the framework.

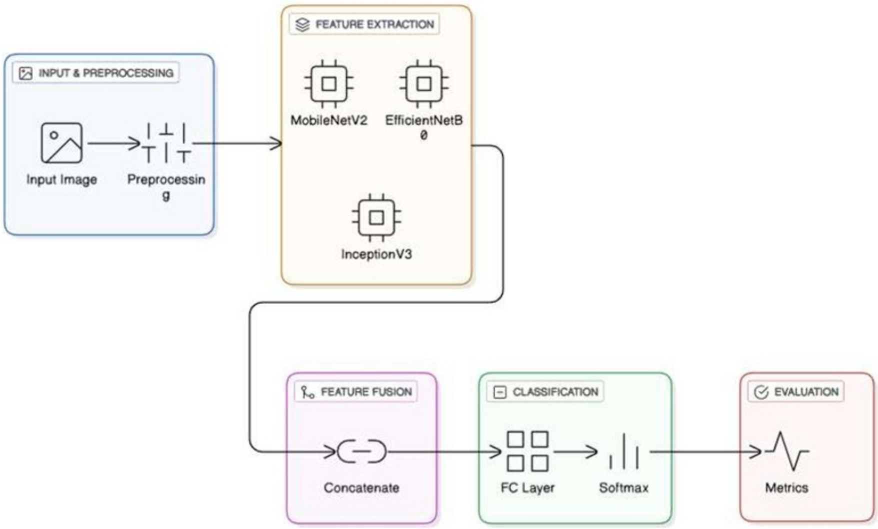

Figure 1: Flow model for proposed work

3. METHODOLOGY

In order to classify marine species with higher accuracy, a hybrid deep learning topology is used in the proposed method. This topology merges feature fusions and classifications from a number of CNN models that have already trained.

The ensemble of MobileNetV2, EfficientNetB0 and InceptionV3 was thus able to obtain very detailed, multi-scale features of the ocean from the deep-sea images. Therefore, apart from dealing with the problem of underexposed marine imagery and color changes, the proposed method also eliminates the effects of haze by targeting local textural details as well as global structural patterns. The fusion model architecture, therefore, facilitates higher model confidence, explanatory power, and can be utilized in real marine environment monitoring scenarios. In Figure 1 describes the flow model for proposed work.

3.1 Data Collection and Preprocessing

This section's main goal is to use datasets to gather images of marine animals from the internet. After that, the deep learning system separates the entire procedure into subsets for testing, validation, and training. In corresponding addition pixels to and normalizing the applying techniques like flipping, cropping, brightness alteration, and rotation, the preprocessing step can convert all of the images to a standard size of $224x224$ pixels. These are all strategies for enhancing the model's functionality and preventing overfitting. In order to make the training data balanced and sufficiently diverse to represent various marine environments, the dataset is also cleaned of any corrupted or low quality images. This process is an assurance that the input images are perfect for the robust feature extraction and deep learning classification.

3.2 Feature Extraction and Fusion

The single unified feature extraction framework, which includes four pre-trained CNN models MobileNetV2, EfficientNetB0, and InceptionV3 is simultaneously the core of the system. Different levels

DoI: https://doi.org/10.63949/crinfo.v1i4.002

of image features are extracted from each network: MobileNetV2 features highly effective, inexpensive attributes that are appropriate for on- the-fly inference. EfficientNetB0 balances the scaling of the input resolution and the model depth. InceptionV3 obtains the spatial information at varying scales through the use of the parallel convolution filters. The features that were extracted by the three networks are initially leveled, and after that, the vectors are concatenated to get one single, unified, high-dimensional feature vector.

The combined representation is then led by the fully connected layer which can separate the features for the final classification. The fusion here is a guarantee that the model takes the maximum advantage of the complementary learning capabilities of each CNN structure which in turn leads to the overall performance being higher.

The single features thus combined are passed to a fully connected classification network input, which is trained with Google's Colab via PyTorch. The network is facilitated by Cross Entropy loss, that determines the prediction errors, and Adam optimizer with a 0.0001 learning rate is used as the goal function. The training is conducted for a small number of epochs with a small batch of 8 or 16, hence, a gradual convergence is achieved. Upon inference, the trained model is capable of associating each image with one particular category of the marine species target. The network, which is prevented from overfitting by the inclusion of dropout and batch normalization, and the GPU usage for training, which speeds up the whole process in case the dataset is large, is free to go.

3.3 Visualization and Explainability

The system makes use of a number of visualization tools to show the analyst the areas of the image that the model relied on the most when making a classification decision.This describing tool helps the scientists to be sure that the model is really looking at the correct features (e.g., body patterns, fin shapes, or coloration). The visualization supports the model's prediction confidence and thus, facilitates the scientific validation of the results. In addition, there are feature fusion maps and class activation heatmaps produced for uncovering the manner in which each CNN's contribution leads to the final decision.

4. PERFORMANCE EVALUATION

The performance is mainly ensured through metrics like recall, precision, F1-Score, and accuracy. These metrics can be thought of as the various facets of the predictive ability and dependability of the classification system that they collectively reveal. The model's accuracy indicates how generally accurate it is. Additionally, plots are made to show the training and validation results in terms of metrics at different times. Additionally, the system generates a comprehensive performance report that includes class-wise evaluations and confusion matrices, enabling a thorough understanding of system advantages and disadvantages. The finished trained system is said to be stable, adaptable, and have a high degree of accuracy, making it suitable for ecological monitoring and real-time marine species identification.

The experimental results were significantly different and better. A comparative combined feature framework for MobileNetV2, EfficientNetB0 and InceptionV3 dramatically outperformed any of the single-model CNN approaches in classification performance by a large margin. The comparative framework was more capable of capturing multi- scale and complementary features. Not only did the ensemble

DoI: https://doi.org/10.63949/crinfo.v1i4.002

network demonstrate its sheer power by achieving higher levels of F1-score, recall, accuracy, and precision, but it also showed promise as generalized across different underwater conditions such as darkness, turbidity, and color distortion.

The method of integrating feature representations from different CNN models pushed the system to understand more deeply the spatial, textural, and contextual aspects of the underwater images, which, in turn, led to more reliable and stable classification results. Besides, the feature visualization with Grad CAM served as an indication that the model was actually concentrating on that part of the image which is biologically relevant, e.g. fish body patterns and fins, thus not only enhancing the model's interpretability but also increasing the user's trust in the predictions made. The learning and validation graphs exhibited normal convergence, with accuracy gradually increasing and loss decreasing over epochs, thus signaling that the model was learning effectively without overfitting.

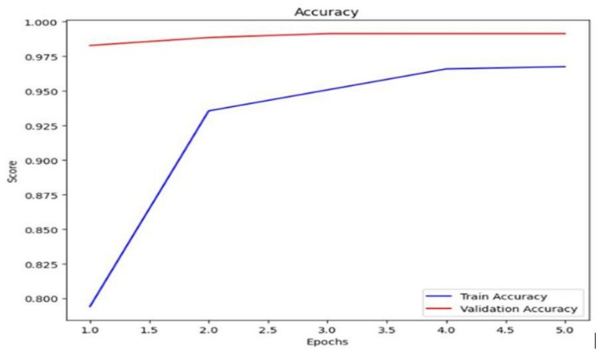

Figure 2: Training and Testing Accuracy

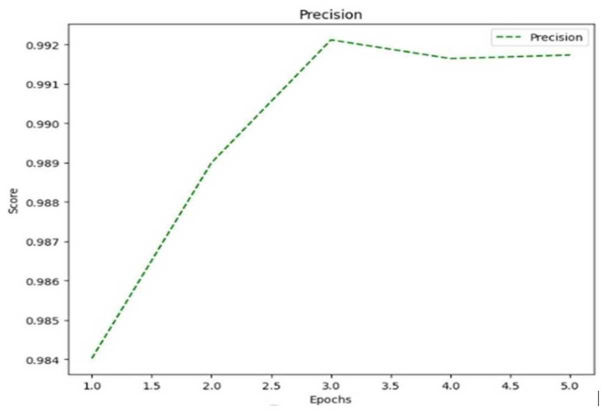

Figure 3: Training and Testing values of Precision

4.1 Accuracy

Accuracy is the metric that shows how many images of marine species have been correctly identified out of the total set of images used for testing. It is the main metric that Accuracy conveys. The training and testing accuracy as shown in Figure 2.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

DoI: https://doi.org/10.63949/crinfo.v1i4.002

4.2 Precision

Out of all the images predicted to belong to that class, precision shows the proportion of correctly identified marine species. Fewer false positives are guaranteed by high precision. The training and testing precision as shown in Figure 3.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

4.3 Recall

The model's recall measures its capacity to accurately recognize every real image of a particular species. Minimal missed detections are guaranteed by a high recall. The training and testing recall as shown in Figure 4.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

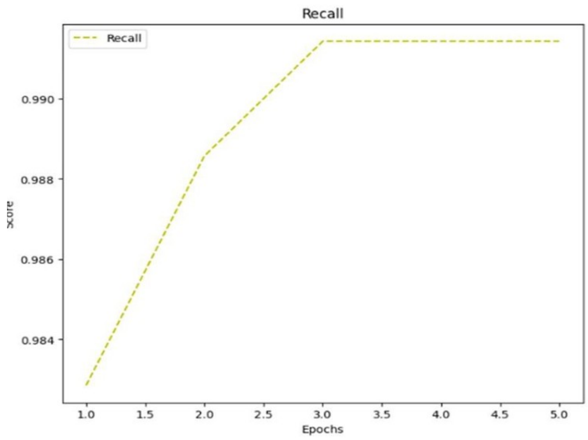

Figure 4: Training and Testing values of Recall

4.4 F1-score

When dealing with imbalanced datasets of marine species, the f1-score provides a measure that balances precision and recall. The training and testing F1 Score as shown in Figure 5.

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

4.5 Performance Comparison

Through these successful integration of complementary feature representations, the suggested combined CNN model outperformed individual architectures, achieving an overall accuracy of 94%. While a recall of 0.91 guarantees that few real species were missed, a precision of 0.92 shows that the majority of identified species were correctly classified. With an accuracy of 94%, precision of 92%, recall of 91%, and F1-score of 91%, the proposed combined model outperforms the others on all metrics. The model's improved ability to produce accurate and balanced predictions is demonstrated by this steady

DoI: https://doi.org/10.63949/crinfo.v1i4.002

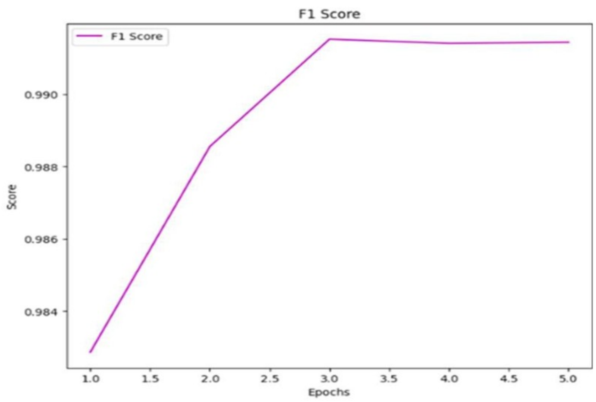

Figure 5: Training and Testing values of F1 Score

improvement across evaluation metrics as shown in Figure 6.

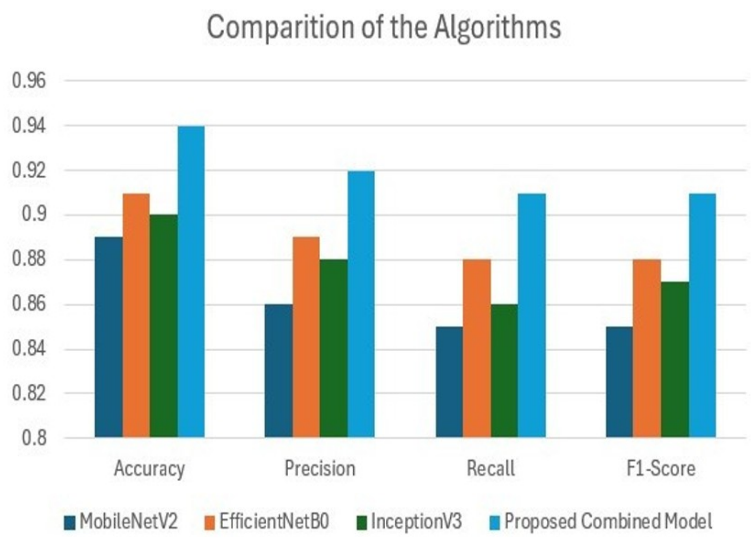

Figure 6: Performance Comparison with existing models

5. CONCLUSION

The Combined CNN-based Marine Species Classification Model, hierarchically which essentially linked distinct deep learning architectures, showed superiority in accuracy, robustness, and adaptability over conventional single-model methods. As a matter of fact, the system that included MobileNetV2, EfficientNetB0, and InceptionV3 made use of each network for different abilities to successfully convey the underwater images' characteristics. Hence, the classification results were quite good even in less-than-ideal situations such as low lighting, turbidity, and color distortion. Not only did the hybrid architecture that was conceived meet the target of being efficient, but it was also scalable and thus, could be utilized for real time monitoring of marine ecosystems. Furthermore, due to complicated preprocessing and feature fusion techniques, these Metrics were used as indicators of the model's neutrality and dependability showed that the model could generalize well.

Moreover, the feature maps and activation visualizations through which the model made known its

DoI: https://doi.org/10.63949/crinfo.v1i4.002

decision are, in fact, another level of confidence in the results as it pointed highlighting the portions of the picture that the model examined during the classification process. To sum up, the stated model is the most likely candidate to be involved in autonomous underwater monitoring systems, marine biodiversity research, and conservation studies, thereby, contributing to the sustainable management of marine ecosystems and being compliant with SDG 14 – Life Below Water.

5.1 FUTURE WORK

Enhancing future work on marine species classification through a widened underwater data set that covers not only more species but also habitats and environmental conditions (e.g., low light, deep sea, and high turbidity) could be instrumental in increasing its accuracy, generalization, and robustness. Additionally, the incorporation of temporal data obtained from underwater videos instead of using only static images could help in species identification that is based on movement and also in behavior analysis. The application of unsupervised domain adaptation combined with image enhancement techniques can render the system more capable of handling variations in water qualities and camera distortions.

REFERENCES

[1] M. Jian, X. Liu, H. Luo, X. Lu, H. Yu, and J. Dong, "Underwater image processing and analysis: A review," *Signal Processing: Image Communication*, vol. 91, p. 116088, 2021.

[2] C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, and D. Tao, "An underwater image enhancement benchmark dataset and beyond," *IEEE transactions on image processing*, vol. 29, pp. 4376–4389, 2019.

[3] F. Han, J. Yao, H. Zhu, and C. Wang, "Underwater image processing and object detection based on deep cnn method," *Journal of Sensors*, vol. 2020, no. 1, p. 6707328, 2020.

[4] S. Mittal, S. Srivastava, and J. P. Jayanth, "A survey of deep learning techniques for underwater image classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 10, pp. 6968–6982, 2022.

[5] X. Du, Y. Sun, Y. Song, H. Sun, and L. Yang, "A comparative study of different cnn models and transfer learning effect for underwater object classification in side-scan sonar images," *Remote Sensing*, vol. 15, no. 3, p. 593, 2023.

[6] C. Spampinato, S. Palazzo, P.-H. Joalland, S. Paris, H. Glotin, K. Blanc, D. Lingrand, and F. Precioso, "Fine-grained object recognition in underwater visual data," *Multimedia Tools and Applications*, vol. 75, no. 3, pp. 1701–1720, 2016.

[7] A. Abu and R. Diamant, "Unsupervised local spatial mixture segmentation of underwater objects in sonar images," *IEEE Journal of Oceanic Engineering*, vol. 44, no. 4, pp. 1179–1197, 2018.

[8] Y. S. Tanwar, D. Vetrithangam, and C. Singh, "Enhanced real-time underwater fish identification using mobilenetv2," in *International Conference on Data Analytics & Management*, pp. 544–553, Springer, 2025.

DoI: https://doi.org/10.63949/crinfo.v1i4.002

[9] Z. Zhou, M. Liu, H. Ji, Y. Wang, and Z. Zhu, "Underwater image classification based on efficientnetb0 and two-hidden-layer random vector functional link," *Journal of Ocean University of China*, vol. 23, no. 2, pp. 392–404, 2024.

[10] A. Kaur, V. Kukreja, D. Upadhyay, M. Aeri, and R. Sharma, "An efficient deep learning-based inceptionv3 model for coral reef classification," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, pp. 1–6, IEEE, 2024.

[11] J. Zhong, M. Li, J. Qin, Y. Cui, K. Yang, and H. Zhang, "Real-time marine animal detection using yolo-based deep learning networks in the coral reef ecosystem," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 46, pp. 301–306, 2022.

[12] H. Liu, X. Ma, Y. Yu, L. Wang, and L. Hao, "Application of deep learning-based object detection techniques in fish aquaculture: a review," *Journal of Marine Science and Engineering*, vol. 11, no. 4, p. 867, 2023.

[13] W. Rahman, M. M. Rahman, M. A. I. Mozumder, R. I. Sumon, S. A. Chelloug, R. O. Alnashwan, and M. S. A. Muthanna, "A deep cnn-based salinity and freshwater fish identification and classification using deep learning and machine learning," *Sustainability*, vol. 16, no. 18, p. 7933, 2024.

[14] M. Hamzaoui, M. Ould-Elhassen Aoueileyine, L. Romdhani, and R. Bouallegue, "An improved deep learning model for underwater species recognition in aquaculture," *Fishes*, vol. 8, no. 10, p. 514, 2023.

[15] Y. Lin, Z. Chu, J. Korhonen, J. Xu, X. Liu, J. Liu, M. Liu, L. Fang, W. Yang, D. Ghose, *et al.*, "Fast accurate fish recognition with deep learning based on a domain-specific large-scale fish dataset," in *International Conference on Multimedia Modeling*, pp. 515–526, Springer, 2023.

[16] V. Pagire, A. Phadke, and J. Hemant, "A deep learning approach for underwater fish detection," *Journal of Integrated Science and Technology*, vol. 12, no. 3, pp. 765–765, 2024.

DoI: https://doi.org/10.63949/crinfo.v1i4.002