ISSN: 2458-8989

Natural and Engineering Sciences

NESciences, 2024, 9 (3): 144-153 doi: 10.28978/nesciences.1606623

# A Deep Learning-based Intelligent Automatic Detection and Classification of Fish Species in Marine Environment

Ashu Nayak [1*], Dr.F. Rahman [2]

[1*] Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India.
E-mail: ku.ashunayak@kalingauniversity.ac.in

[2] Assistant Professor, Department of CS & IT, Kalinga University, Raipur, India.
E-mail: ku.frahman@kalingauniversity.ac.in

## Abstract

As marine environments encounter escalating threats and obstacles, accurate and effective Fish Species Classification (FSC) has become crucial for managing fisheries, preserving biodiversity, and ecological surveillance. Considering the substantial volume of georeferenced fish photographs gathered daily by fishermen, artificial intelligence (AI) and computer vision (CV) technologies provide significant potential to automate their analysis via species recognition and classification. This study investigates utilizing Deep Learning (DL) techniques alongside appearance-based feature selection to automatically and precisely determine fish species from images. The research utilizes many aquatic fish images, including diverse species, sizes, and ecological settings. Conventional DL models struggle to capture long-term dependencies and necessitate fixed input sizes, rendering them less adaptable when processing images of varying dimensions. The Vision Transformer (VT) mitigates these limitations using the transformer model's Self-Attention Mechanisms (SAM). This paper employs a VT to address the FSC problem and provides Intelligent Automatic Detection and FSC in Marine Environment (IAD-FSC-ME). VT's efficacy is evaluated compared to pre-trained Convolutional Neural Network (CNN) models: VGG19, DenseNet121, ResNet50v2, InceptionV3, and Xception. The investigations utilize an open data set (Fish4Knowledge), wherein both the object detection and classification systems are enhanced with subtropical fish species of interest. It has been observed that VT surpassed the prevailing literature by attaining 99.14% accuracy in efficient FSC.

## Keywords:

*Deep learning, vision transformer, fish species classification, convolutional neural network, intelligent automatic detection.*

## Article history:

Received: 04/08/2024, Revised: 22/09/2024, Accepted: 21/10/2024, Available online: 31/12/2024

**Corresponding Author: Ashu Nayak, E-mail: ku.ashunayak@kalingauniversity.ac.in*

## Introduction

Over 85% of worldwide fish captures occur in fisheries that lack critical data, assets, and infrastructure necessary for conducting inventory evaluations (Scherrer & Galbraith, 2020). This particularly applies to commercial fisheries, which are growing in prominence and provide significant health and economic advantages but are challenging to monitor, regulate, and evaluate (Freire et al., 2020). Considering the substantial participation in recreational fisheries and the prevalent use of advanced technologies, there exists a significant opportunity for extensive data gathering that might enhance our understanding of catches and the state of the fish population (Agnes Pravina et al., 2024). Fishermen are generally inclined to preserve fish populations, and experiences with other organizations indicate that participation in citizen science initiatives not only facilitates the compilation of extensive datasets but also enhances understanding and fosters a feeling of responsibility. Citizen science would significantly enhance recreational fisheries management by facilitating cooperative study and management efforts (Harris et al., 2021).

AI has transformed weather prediction, wildfire disaster management, medical care, and transportation (Trivedi et al., 2023). AI and CV technologies provide an unexploited potential to revolutionize fishing administration. AI facilitates the rapid analysis of photos sent by anglers, enabling automated FSC and perhaps supplying data regarding fish body size. Despite the growing use of AI in fisheries, shown by around 40 scholarly papers annually (Honarmand Ebrahimi et al., 2021), this remains significantly constrained compared to other disciplines. AI-driven models are mostly used in commercial fisheries, often focusing on a limited number of species within a particular context (Uyan, 2022; Xue, 2024).

Furthermore, the methodologies, instruments, and scripts used in this research are often inaccessible to the public, limiting their broader adoption, utilization, and community-led enhancement. Implementing AI applications and creating novel models need a robust foundation in computing and sophisticated technology. As new tools and databases are produced, programming may become more simplified, allowing smaller study teams globally to use these resources for their objectives (Kim et al., 2019). The most labor-intensive and time-consuming aspect of constructing species recognition models is acquiring expert-identified photos rather than computing resources. Public participation would be particularly impactful, enabling research organizations globally to include local scientists in collecting pictures and creating FSC models pertinent to regional fishing operations and ecological inquiries. These methods might facilitate expediting the processing of freshly collected local science information, advancing marine fishing evaluations into a resource-rich epoch (Larkin et al., 2022).

Current DL models, such as CNNs, may encounter difficulties in fine-grained categorization tasks where minute distinctions between species are essential. These models often emphasize general characteristics instead of the intricate, nuanced differences essential for precise FSC (Wei et al., 2021). Numerous DL models have been developed on static databases and may struggle to adapt to new or dynamic inputs. As fresh marine species are identified or environmental circumstances evolve, current models may need retraining or refinement to include these modifications, which may be resource-demanding and time-consuming.

The shortcomings of conventional methods and current DL technologies underscore the need for more sophisticated strategies to tackle these difficulties. In recent years, VT has gained prominence in picture categorization challenges (Han et al., 2022). VTs, due to their capacity to collect worldwide context and manage fluctuation in picture quality, provide a viable approach to enhance the precision and effectiveness

of species categorization. VTs have emerged as a transformative architecture in the area of CV, signifying a paradigm shift from traditional DL models.

## Literature Survey

The past few years have seen substantial advancements in categorizing marine species, propelled by the development and use of advanced Machine Learning (ML) algorithms. These methods have shown that they can achieve FSC with enhanced accuracy, especially in difficult underwater conditions marked by fluctuating illumination, background interference, and turbulent water.

DL has been used in several industries, from gaming to healthcare, although its potential for fish classification remains only partially investigated. A particular CNN known as Fast R-CNN has been used for object recognition to isolate fish from photos captured in real settings while effectively disregarding noise from the surroundings (Li et al., 2024). This method involves pre-training an AlexNet on the ImageNet repository and then adapting it to learn on a portion of the Fish4Knowledge database (Iqbal et al., 2021). In the concluding phase, Fast R-CNN utilizes the previously trained weights and the area suggestions generated by AlexNet as inputs, attaining a mean average accuracy of 81.4%. Another method involves applying pre-training to a CNN akin to AlexNet, with three fully connected (FC) layers and five layers of convolution (CL). Pre-training is conducted using 1500 images over 1500 classes from the ImageNet file, and a CNN uses the acquired scores after its adaptation to the Fish4Knowledge database. Post-training is conducted using 60 images for each of the 12 categories from Fish4Knowledge. The Fish4Knowledge photos undergo pre-processing using image de-noising, achieving an accuracy of 87.12% on 1,550 test photographs.

The maximum documented accuracy for Fish4Knowledge in the literature preceding our research is 97.28%, attained by initially applying filters to the initial images to delineate the fish's form and eliminate the background, followed by the use of a CNN in conjunction with a Support Vector Machine (SVM) for categorization (Zhang et al., 2023). The method is called DeepFish, which includes three conventional CL and three FC layers. A prevalent characteristic of prior methods is their typical use of a pre-processing approach for photos to minimize noise in the desired image and specifically to delineate the outline of the fish (Qu et al., 2024). While this strategy may enhance system efficiency, the pre-processing operation must be thoroughly calibrated, as it may exclude valuable data and lead to detrimental performance effects. Individual species inherently possess unique living environments, as seen by their backgrounds. Deliberately omitting the species' context during pre-processing may eradicate valuable information. It is essential to implement a robust methodology capable of effectively managing noise and accommodating categorization variance to maximize the utilization of historical data while minimizing the impact of background noise on outcomes (Hyder et al., 2020).

Liu et al., (2018) have developed a virtual fish monitoring system that employs YOLO and concurrent correlation filters, including detection and classification in a comprehensive end-to-end framework. Jahanbakht et al., (2022) conducted similar research by training a YOLO architecture to recognize several fish species using three databases, achieving an estimated average accuracy score of 0.5392. Authors in (Pedersen et al., 2019) expanded their research to include marine animals and fish, using the same YOLO methodologies. All of these techniques have the characteristic of training their networks in an end-to-end manner. The literature study indicates that fish classification systems mostly use conventional machine learning and deep learning techniques. The literature has often used transfer-learning approaches to conserve time and computing resources. The use of VT in FC is restricted, with just a few research using VT techniques on reference fish species databases (Teng & Zhao, 2020).

## Proposed Method

### *Database Information*

The Fish4Knowledge database comprises pictures derived from marine videos of fish. A collection of 25,552 images has been cataloged across 20 distinct species. The leading 12 species constitute 96.5% of the photos, with the foremost species representing around 44% of the total images. The quantity of photos per species varies from 20 to 10,998. This results in a highly skewed sample. The image sizes vary from roughly 32 × 32 pixels to 256 × 256 pixels. Another discovery in the collection is that most photos are captured from a perspective along the anterior-posterior axis or are somewhat inclined off that axis. In that collection of photos, the majority are captured from the left or right medial perspective, revealing the entire ventral skeletal plan. There are many photographs from the front perspective but few from the back end. Few photos were captured from the authentic frontal perspective. Many of the chosen species have a compact morphology, such as dorsoventrally elongated forms. This produces a distinctive form when the photos are captured from a lateral perspective. Consequently, photographs captured from the dorsal perspective produce a slender, abbreviated form. The images have a comparatively light backdrop, accentuating the fish's outline (Silva et al., 2022).

### *Transformer Encoder (TE)*

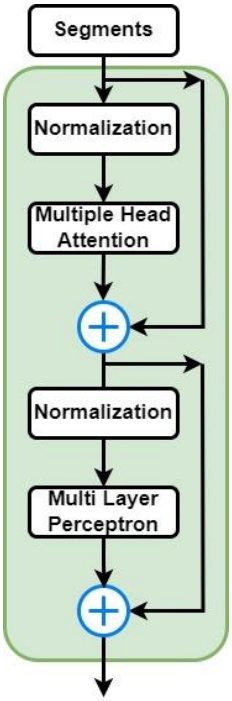


Figure 1. Transformer encoder (TE)

The domain of CV has traditionally depended on CNN until the introduction of transformers, which demonstrate exceptional performance. VT employs the conventional TE architecture for image-related tasks, as seen in Fig. 1. The graphic is first divided into many regions and then subjected to element encoding and position encoding before being entered into the TE with class encoding. Ultimately, it is sent into the Multi-Layer Perceptron (MLP) for classification prediction. Multi-head attention (MHA) constitutes the fundamental architecture of the TE. The input consists of three vectors: request, key, and significance, while the output is the weighted sum of all values. The SAM matrix is denoted by Equation (1) using matrices $R, K, and\ S$.

$$SAM(R,K,S) = softmax[\frac{R^t K}{\sqrt{d_K}}S \quad (1)$$

The matrices $R, K, and\ S$ in Equation (1) pertain to matrix computation. The concatenation of several attention mechanisms achieves MHA. The MHA shown in Equation (2) has a diverse subdomain representation across several positions, which may be articulated as:

$$MHA(R,K,S) = \sum[H1, H2, \ldots . Hn)W0 \quad (2)$$

$$Hi = SAM(RW_i^R, KW_i^K, SW_i^S) \quad (3)$$

***VT for FSC***

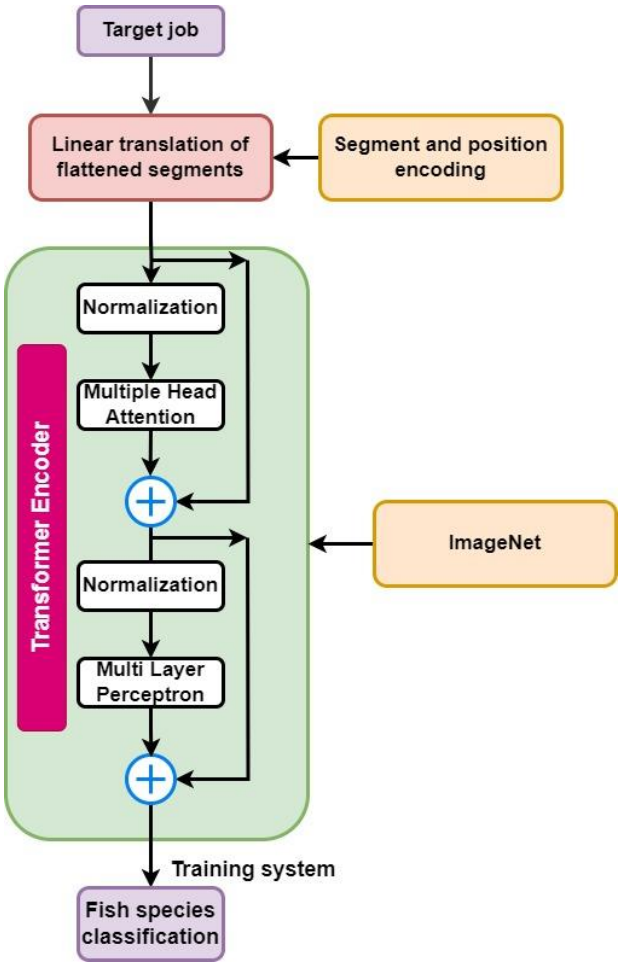


Figure 2. Architecture of VT for FSC

Fig. 2 depicts the architecture of VT for FSC. After undergoing location encoding and linear translation, the image is first divided into many segments input into the TE. The encoder utilizes weights previously trained and derived from the ImageNet dataset. The MLP ultimately generates the anticipated category. This research proposes a FSC approach using transfer learning and a VT model. Training the neural network from inception in traditional DL models is extremely time-consuming and requires significant data (Ahmed, 2024). To minimize computational expenses and data consumption, the VT framework has been previously trained on the ImageNet database for refinement, which facilitates the extraction of specialized characteristics from fish images and further trains the MLP for classification purposes.

Given the substantial data requirements for training a VT model from scratch, the mathematical resource requirements are excessive, and the current fish population is limited; therefore, it has been proposed to utilize transfer learning to enhance the model's feature mining capabilities for fish information. Consequently, we use the VT framework learned on the ImageNet information set, and upon acquiring the model's trained weights, we refine them for feature extraction in this work. The precise procedure is shown in Fig. 2. The VT has been trained in advance using the dataset to get the model parameters. The model that has been trained is then applied to the desired database, namely the fish databases (Fish4Knowledge). The encoder level of the VT is fixed, and only the MLP level gets training for the categorization of fish species images.

## Results and Discussion

The study uses the TensorFlow DL structure in conjunction with the library provided by Keras. The tasks were executed in the cloud GPU service using an Ubuntu Linux workstation with an NVIDIA A100 GPU. The model that was learned is subjected to thorough evaluation to determine its reliability and scalability on an independent test database. The model's effectiveness on Fish4Knowledge dataset is assessed using measures including accuracy, precision, recall, and F1 score for different CNN models. The impact of optimizer selection on the quality of classification is also examined.

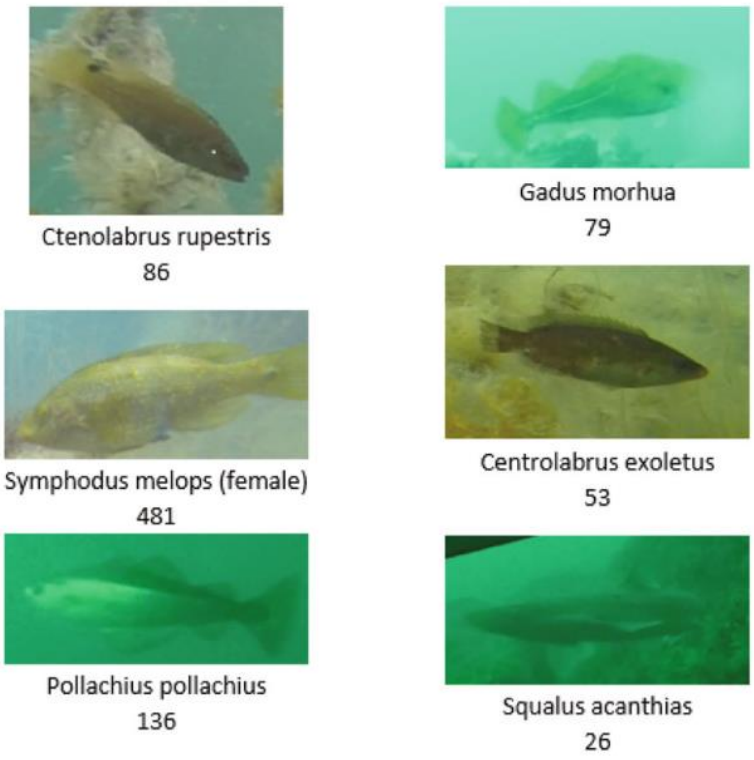


Figure 3. Sample fish images from the Fish4Knowledge database

This comprehensive assessment enables a detailed analysis of the model's efficacy in accurately identifying fish species. These varied evaluation metrics provide comprehension of the model's potential and appropriateness for the specified FSC according to distinct performance standards. Figure 3 illustrates the sample fish images from the Fish4Knowledge database used in the analysis and FSC.

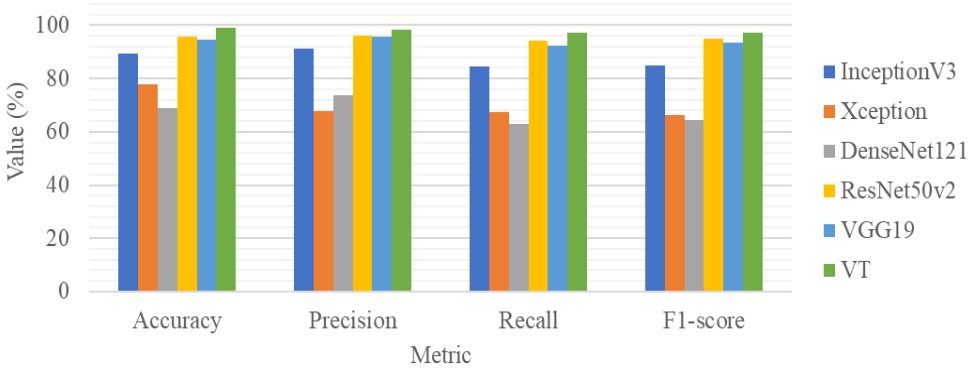


Figure 4. Performance comparison for various CNN models with the proposed VT for FSC

Fig. 4 compares various CNN models' performance with the proposed VT for FSC. Among the models, VT has the superior overall performance, with an accuracy of 99.14%, precision of 98.2%, recall of 97.32%, and F1-score of 97.12%. This is markedly superior to the other models. For example, ResNet50v2 and VGG19, the subsequent top performers, get accuracy ratings of 95.79% and 94.69%, respectively; however, they remain inferior to VT's measures across all categories. DenseNet121 has the worst performance across all criteria, with 69.06% accuracy (Lekunberri et al., 2022). The results indicate that VT's architecture is particularly adept at high-precision jobs in FSC, surpassing conventional CNN-based models in this domain.

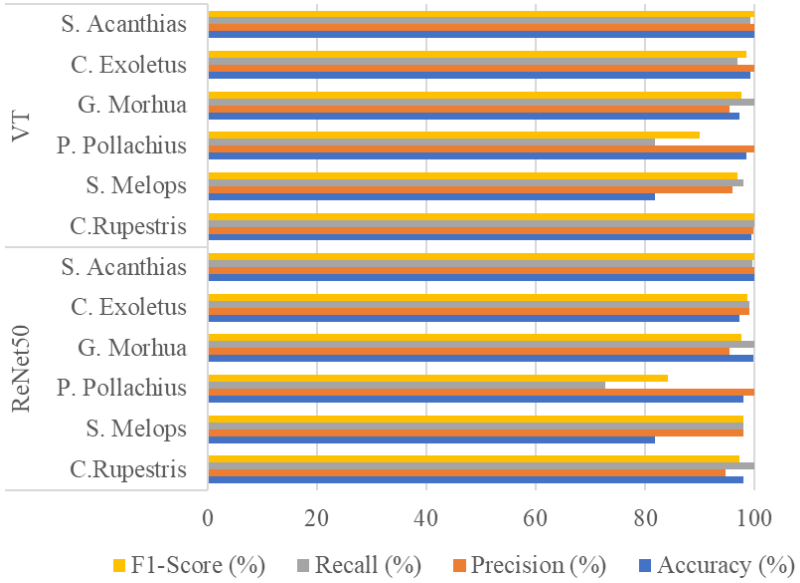


Figure 5. Comparative analysis of several categorization techniques for each category of fish images from the Fish4Knowledge database.

Fig. 5 illustrates the comparative analysis of several categorization techniques for each category of fish images from the Fish4Knowledge database. Both models had near-perfect results in the fish category S. Acanthias, achieving 100% accuracy, precision, recall, and F1-Score. In most areas, VT surpasses ReNet50, particularly in accuracy and recall, underscoring its dependability in properly categorizing fish photos. For example, VT achieved 99.89% accuracy and 100% recall for C. Rupestris, demonstrating its exceptional proficiency in properly recognizing this species. Overall, VT demonstrates superior consistency across categories, establishing it as a more successful model for fish categorization inside the Fish4Knowledge dataset.

## Conclusion

This work examines DL methods in conjunction with appearance-based feature selection for the automated and accurate identification of fish species from photographs. The study employs an extensive compilation of marine fish photos, including photos of various species, sizes, and biological environments. Conventional DL models fail to capture long-term relationships and need constant input sizes, making them less versatile for processing pictures of different dimensions. The Vision Transformer (VT) addresses these constraints using the Self-Attention Mechanisms (SAM) inherent in the transformer paradigm. This work utilizes a VT to tackle the FSC issue and offers Intelligent Automatic Detection and FSC in the Marine Environment (IAD-FSC-ME). The effectiveness of VT is assessed against pre-trained CNN models. The research used an open dataset (Fish4Knowledge), in which both object detection and classification methods are augmented with subtropical fish species of interest. Among the models, VT has the superior overall performance, with an accuracy of 99.14%, precision of 98.2%, recall of 97.32%, and F1-score of 97.12%. Both models had near-perfect results in the fish category S. Acanthias, achieving 100% accuracy, precision, recall, and F1-Score.

## Author Contributions

All Authors contributed equally.

## Conflict of Interest

The authors declared that no conflict of interest.

## References

Agnes Pravina, X., Radhika, R., & Ramesh Palappan, R. (2024). Financial Inclusiveness and Literacy Awareness of Fisherfolk in Kanyakumari District: An Empirical Study. *Indian Journal of Information Sources and Services, 14*(3), 265–269. https://doi.org/10.51983/ijiss-2024.14.3.34

Ahmed, I.M. (2024). Optimum Design of Reinforced Concrete Beams with Large Opening Using Neural Network Algorithm. *International Academic Journal of Science and Engineering, 11*(1), 138-152. http://doi.org/10.9756/IAJSE/V11I1/IAJSE1117

Freire, K. M. F., Belhabib, D., Espedido, J. C., Hood, L., Kleisner, K. M., Lam, V. W., ... & Pauly, D. (2020). Estimating global catches of marine recreational fisheries. *Frontiers in Marine Science, 7*, 12. https://doi.org/10.3389/fmars.2020.00012

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., ... & Tao, D. (2022). A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence, 45*(1), 87-110. https://doi.org/10.1109/TPAMI.2022.3152247

Harris, D., Johnston, D., & Yeoh, D. (2021). More for less: Citizen science supporting the management of small-scale recreational fisheries. *Regional Studies in Marine Science, 48*, 102047. https://doi.org/10.1016/j.rsma.2021.102047

Honarmand Ebrahimi, S., Ossewaarde, M., & Need, A. (2021). Smart fishery: a systematic review and research agenda for sustainable fisheries in the age of AI. *Sustainability, 13*(11), 6037. https://doi.org/10.3390/su13116037

Hyder, K., Maravelias, C. D., Kraan, M., Radford, Z., & Prellezo, R. (2020). Marine recreational fisheries—current state and future opportunities. *ICES Journal of Marine Science, 77*(6), 2171-2180. https://doi.org/10.1093/icesjms/fsaa147

Iqbal, M. A., Wang, Z., Ali, Z. A., & Riaz, S. (2021). Automatic fish species classification using deep convolutional neural networks. *Wireless Personal Communications, 116*, 1043-1053. https://doi.org/10.1007/s11277-019-06634-1

Jahanbakht, M., Xiang, W., Waltham, N. J., & Azghadi, M. R. (2022). Distributed deep learning and energy-efficient real-time image processing at the edge for fish segmentation in underwater videos. *IEEE Access, 10*, 117796-117807. https://doi.org/10.1109/ACCESS.2022.3202975

Kim, K., Ko, E., Kim, J., & Yi, J. H. (2019). Intelligent Malware Detection Based on Hybrid Learning of API and ACG on Android. *Journal of Internet Services and Information Security, 9*(4), 39-48. https://doi.org/10.22667/JISIS.2019.11.30.039

Larkin, K. E., Marsan, A. A., Tonné, N., Van Isacker, N., Collart, T., Delaney, C., ... & Calewaert, J. B. (2022). Connecting marine data to society. In *Ocean Science Data* (pp. 283-317). Elsevier. https://doi.org/10.1016/B978-0-12-823427-3.00003-7

Lekunberri, X., Ruiz, J., Quincoces, I., Dornaika, F., Arganda-Carreras, I., & Fernandes, J. A. (2022). Identification and measurement of tropical tuna species in purse seiner catches using computer vision and deep learning. *Ecological Informatics, 67*, 101495. https://doi.org/10.1016/j.ecoinf.2021.101495

Li, S., Li, P., He, S., Kuai, Z., Gu, Y., Liu, H., ... & Lin, Y. (2024). An Automatic Detection and Statistical Method for Underwater Fish Based on Foreground Region Convolution Network (FR-CNN). *Journal of Marine Science and Engineering, 12*(8), 1343. https://doi.org/10.3390/jmse12081343

Liu, S., Li, X., Gao, M., Cai, Y., Nian, R., Li, P., ... & Lendasse, A. (2018, October). Embedded online fish detection and tracking system via YOLOv3 and parallel correlation filter. In *Oceans 2018 Mts/Ieee Charleston* (pp. 1-6). IEEE. https://doi.org/10.1109/OCEANS.2018.8604658

Pedersen, M., Bruslund Haurum, J., Gade, R., & Moeslund, T. B. (2019). Detection of marine animals in a new underwater dataset with varying visibility. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 18-26).

Qu, H., Wang, G. G., Li, Y., Qi, X., & Zhang, M. (2024). ConvFishNet: An efficient backbone for fish classification from composited underwater images. *Information Sciences*, 121078. https://doi.org/10.1016/j.ins.2024.121078

Scherrer, K., & Galbraith, E. (2020). Regulation strength and technology creep play key roles in global long-term projections of wild capture fisheries. *ICES Journal of Marine Science, 77*(7-8), 2518-2528. https://doi.org/10.1093/icesjms/fsaa109

Silva, C. N., Dainys, J., Simmons, S., Vienožinskis, V., & Audzijonyte, A. (2022). A scalable open-source framework for machine learning-based image collection, annotation and classification: a case study for automatic fish species identification. *Sustainability, 14*(21), 14324. https://doi.org/10.3390/su142114324

Teng, B., & Zhao, H. (2020). Underwater target recognition methods based on the framework of deep learning: A survey. *International Journal of Advanced Robotic Systems, 17*(6), 1729881420976307. https://doi.org/10.1177/1729881420976307

Trivedi, J., Devi, M. S., & Solanki, B. (2023). Step Towards Intelligent Transportation System with Vehicle Classification and Recognition Using Speeded-up Robust Features. *Archives for Technical Sciences, 1*(28), 39-56. https://doi.org/10.59456/afts.2023.1528.039J

Uyan, A. (2022). A Review on the Potential Usage of Lionfishes (Pterois spp.) in Biomedical and Bioinspired Applications. *Natural and Engineering Sciences, 7*(2), 214-227. http://doi.org/10.28978/nesciences.1159313

Wei, X. S., Song, Y. Z., Mac Aodha, O., Wu, J., Peng, Y., Tang, J., ... & Belongie, S. (2021). Fine-grained image analysis with deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence, 44*(12), 8927-8948. https://doi.org/10.1109/TPAMI.2021.3126648

Xue, M. (2024). Assessing the Recreational Fishers and their Catches based on Social Media Platforms: Privacy and Ethical Data Analysis Considerations. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 15*(3), 521-542. https://doi.org/10.58346/JOWUA.2024.I3.033

Zhang, X., Huang, B., Chen, G., Radenkovic, M., & Hou, G. (2023). WildFishNet: Open set wild fish recognition deep neural network with fusion activation pattern. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. https://doi.org/10.1109/JSTARS.2023.3299703