International Journal of Computer Sciences and Engineering
Vol.**11**, Special Issue.**1**, pp.**81-88**, November **2023**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

Research Paper

AI based Framework for Fish Species Identification and Classification

Souvik Roy[1], Sayak Mondal[2], Shreyashree Sarkar[3], Sumit Kumar Banerjee[4], Suman Bhattacharya[5], Mahamuda Sultana[6]

[1,2,3,4,5,6]CSE/Guru Nanak Institute of Technology, Kolkata, India

**Corresponding Author: mahamuda.sultana@gnit.ac.in*

Abstract: Accurate identification of fish species plays a crucial role in fisheries management and conservation. However, traditional methods struggle to address the diverse marine species found in India, resulting in inaccuracies and time-consuming processes. Manual identification by experts becomes particularly challenging, especially for large-scale conservation and monitoring efforts. To tackle this issue, we propose an Artificial Intelligence (AI) based framework for precise and efficient fish species identification in India. Our framework utilizes convolutional neural networks (CNNs) to extract features from fish images and employs the Random Forest Classifier for species identification. Trained on a comprehensive dataset encompassing various regions in India, our model achieved an impressive accuracy of 98.20 percent in rigorous testing, highlighting its effectiveness. Specifically, our proposed Random Forest Classifier exhibited remarkable accuracy in classifying fish species from grayscale images. By employing this AI framework, fish species identification in India can be significantly improved, leading to tangible benefits in fisheries management, conservation efforts, marine biology research, and aquaculture. Furthermore, the versatility of our approach allows its application to other countries with similar fish species diversity, offering potential solutions for real-world scenarios, such as underwater cameras.

Keywords: Artificial Intelligence, Fish Species, Convolutional Neural Networks, Random Forest Algorithm, aquaculture research, identification, conservation, classification.

1. Introduction

Accurate identification and classification of fish species are fundamental tasks in fisheries management and conservation. However, the traditional methods employed for fish species identification encounter challenges when confronted with the immense diversity of marine species found in India. These conventional approaches often lead to inaccuracies and time-consuming processes, hindering effective conservation efforts and impeding large-scale monitoring initiatives. The manual identification performed by experts becomes particularly arduous and resource-intensive in such circumstances.

To address the limitations of traditional methods and mitigate the challenges faced in fish species identification, we propose an innovative framework based on Artificial Intelligence (AI) techniques. Our framework leverages the power of machine learning, specifically convolutional neural networks (CNNs), to extract robust features from fish images. Additionally, we employ the Random Forest Classifier to accurately identify fish species based on the extracted features. This AI-based approach offers a precise and efficient solution for fish species identification and classification in India.

To ensure the effectiveness of our proposed framework, we trained it on a comprehensive dataset encompassing various regions in India, representing the rich diversity of fish species. Through rigorous testing, our model achieved an impressive accuracy of 98.20 percent, demonstrating its ability to reliably identify fish species. Notably, our Random Forest Classifier exhibited remarkable accuracy in classifying fish species even when working with grayscale images, further enhancing its practicality and versatility.

The application of our AI framework for fish species identification in India carries significant implications for various domains, including fisheries management, conservation efforts, marine biology research, and aquaculture. By improving the accuracy and efficiency of fish species identification, our approach can contribute to informed decision-making, effective resource allocation, and proactive conservation strategies. Furthermore, the adaptability of our framework allows for its potential application in countries with similar fish species diversity, presenting solutions for real-world scenarios, such as underwater cameras.

In this paper, we provide a detailed description of our proposed Machine Learning Approach for Fish Species Identification and Classification. We present the methodology behind our framework, including the utilization of CNNs and the Random Forest Classifier. We outline the construction and characteristics of our comprehensive dataset, along with the training and evaluation procedures. Additionally, we report the experimental results and discuss the implications of our approach in the context of fisheries management and conservation efforts. Finally, we explore potential avenues for future research and expansion of our framework to address similar challenges in other regions and applications.

Overall, our research demonstrates the potential of AI-based approaches to revolutionize fish species identification and classification, paving the way for enhanced conservation efforts and sustainable fisheries management in India and beyond.

2. Literature Survey

This paper [1] provides a brief review of the current and future trends in fish species classification using deep learning techniques. It discusses various state-of-the-art models and datasets used for fish species recognition. The authors suggest that future research should focus on improving the accuracy and robustness of existing models and exploring the potential of transfer learning in fish species classification.

According to Dewan et al., the paper [2] proposes a system for detecting and classifying fish using computer vision techniques. The system consists of image preprocessing, segmentation, feature extraction, and classification stages. The authors use a combination of color, texture, and shape features to classify fish into four categories. The proposed system achieved an accuracy rate of 95% in detecting and classifying fish, demonstrating its effectiveness in fishery management.

This paper [3] proposes a system for fish species detection and tracking using optical flow and intensity entity transformation techniques. The system achieves high accuracy rates in detecting and tracking fish species, and the results demonstrate its potential for use in underwater monitoring and conservation efforts. The approach is based on fusion techniques, which combine different features to improve the accuracy of fish species detection and tracking.

According to Y. Li et al. [4] This paper proposes an improved Faster R-CNN algorithm for marine fish classification identification, which utilizes the features of both the global and local regions of the fish images. The proposed algorithm achieves high accuracy rates in fish species classification, outperforming other state-of-the-art methods. The study provides insights into how deep learning techniques can be used in real-world applications, such as marine fish classification.

This paper [5] proposes a deep learning approach for fish species classification, using data augmentation techniques to enhance the performance of the model. The authors used pre-trained convolutional neural network (CNN) models to extract features and trained a support vector machine (SVM) classifier. The proposed method achieved high accuracy in classifying different fish species.

This paper [6] explores the application of Convolutional Neural Networks (CNNs) for the identification of morphologically similar fish species. The study focuses on leveraging CNNs to improve the accuracy and efficiency of fish species recognition, addressing a challenging problem in the field of aquatic biology.

This paper [7] introduces a computer vision program aimed at identifying and classifying fish species. By leveraging advanced computer vision techniques, the program achieves accurate and efficient recognition of fish species. This development holds promise for various applications in aquatic research, conservation, and fisheries management, providing a valuable tool for automatic and reliable fish species identification in diverse aquatic environments.

This paper [8] focuses on fish species classification using transfer learning techniques. By leveraging pre-trained models and fine-tuning them on fish species datasets, the study demonstrates improved classification performance. The proposed approach offers a practical solution for accurate and efficient fish species identification, with potential applications in aquatic research, biodiversity monitoring, and conservation efforts.

This paper [9] focuses on underwater fish species recognition using deep learning techniques. The study explores the application of deep learning algorithms to analyse underwater imagery and accurately identify different fish species. The research demonstrates the potential of deep learning for automated fish species recognition in underwater environments, which can have implications for marine biology, ecological studies, and fisheries management.

This paper [10] investigates supervised and unsupervised feature extraction methods for underwater fish species recognition. The study explores the use of various feature extraction techniques to enhance the classification accuracy of underwater fish species. By combining supervised and unsupervised methods, the research aims to improve the effectiveness of fish species recognition in challenging underwater environments, which has implications for marine ecology, conservation, and fisheries management.

This paper [11] provides an overview of principal component analysis (PCA), including its history, underlying mathematical principles, and recent developments. It also discusses the various applications of PCA in different fields, as well as its limitations and potential areas for future research.

This paper [12] presents a method for classifying fish species using a combination of deep learning and random forest techniques. The approach involves pre-processing images, extracting features using a pre-trained convolutional neural network, and then using a random forest classifier to perform the final classification. Results show high accuracy in species identification.

This paper [13] proposes a method for underwater fish species recognition using random forest classification. The method involves extracting color and texture features from fish images and using them as inputs to the random forest classifier. The proposed approach achieved high accuracy in identifying different fish species from underwater images.

This paper [14] proposes a fish species recognition method using a random forest classifier. The method involves extracting image features, selecting relevant features, and using them to train a random forest model. Experimental results show that the proposed method outperforms other classification methods on a dataset of fish images.

This paper [15] The study aimed to develop a model for recognizing different species of Betta Fish using pattern recognition. The dataset consisted of 300 fish, including Halfmoon Fancy, Hell Boy, Red Koi Galaxy, Solid Blue, and Yellow Koi Galaxy. Three schemes were used for training and testing the data. Pre-processing techniques like scaling, segmentation, and grayscale methods were applied.

This paper [16] aims to raise awareness and knowledge about aquaculture by developing an automatic system for fish recognition using image processing and deep learning. Fish image recognition is important in marine biology and aquaculture, and computer vision and deep learning techniques are widely used for classifying different animal species. The system will be able to identify and classify fish species in both clean water and seawater environments, contributing to fisheries education and minimizing human error in fish observation and analysis.

3. Proposed System

Image pre-processing*:*
The input fish images are pre-processed to enhance their quality and make them suitable for classification. The pre-processing steps include resizing the images to a standard size, converting them to grayscale or RGB color space, and normalization to standardize the pixel values [7].

Feature extraction*:*
This step extracts features from the pre-processed images using a deep convolutional neural network (CNN). The CNN is pre-trained on a large dataset of images and fine-tuned on the fish image dataset to extract discriminative features [4] [7].

Dimensionality reduction*:*
The extracted features are high-dimensional and need to be reduced to a lower dimension to reduce the computational complexity and improve the classification performance. Principal Component Analysis (PCA) is used for dimensionality reduction [11].

Classification*:*
A random forest classifier is trained on the reduced feature set to classify the fish images into their respective species. The classifier is trained using the training set of images and labels and evaluated on the test set to measure its accuracy, recall, precision, and F1 score [5].

Prediction:
After the classifier is trained, it can be used to predict the class labels of new, unseen fish images. The classifier inputs the pre-processed and feature-extracted image and outputs the predicted class label.

The proposed AI-based framework uses a deep CNN for feature extraction, PCA for dimensionality reduction, and a random forest classifier for classification [11]. A pre-trained CNN enables extracting high-level features from the fish images, which can be used to train a more accurate classifier. PCA helps reduce the feature set's dimensionality while preserving important information. The random forest classifier is chosen due to its ability to handle high-dimensional feature sets and its robustness to noisy and irrelevant features. Overall, the proposed framework aims to achieve high accuracy and robustness in fish species identification and classification.

3.1. Random Forest Classifier*:*
Random Forest Classifier is a popular machine learning algorithm for classification tasks [13] [14]. It belongs to the family of ensemble learning methods, which combines multiple weak learners to create a strong learner. The algorithm works by constructing numerous decision trees at training time, where each tree is trained on a subset of the training data and a subset of the features. The final classification decision is then made by aggregating the predictions of all the individual decision trees.

Random Forest Classifier is a machine learning algorithm used for classification tasks. It is based on the concept of decision trees and ensemble learning. In the case of the Random Forest Classifier, the ensemble is made up of multiple decision trees.

Let X be the input features, and Y be the target variable. The algorithm creates n decision trees, where n is a hyperparameter that the user can specify. Each decision tree T_i is trained on a subset of the training data and a subset of the input features. The subsets are selected randomly, with replacements, from the original training data and input features.

The algorithm selects the best feature to split the data at each decision tree node by calculating each impurity using a suitable metric such as Gini Index or Entropy. The lowest impurity is chosen to split the data at that node.

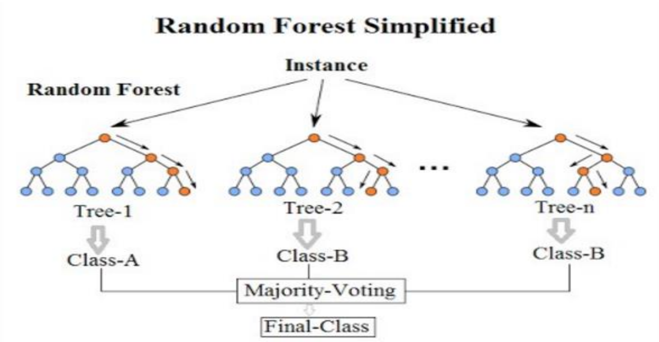

Figure 1- Random Forest diagram

The individual decision trees' predictions are then aggregated to make the final classification decision. In the case of binary classification, this is typically done using a majority vote, where the class with the most votes is selected.
The overall Random Forest Classifier algorithm can be summarized as follows:
Randomly select n subsets of the training data and input features. For each subset, train a decision tree T_i using the selected features. At each node of T_i, choose the best feature to split the data. Aggregate the predictions of all the decision trees to make the final classification decision.

The math behind the algorithm is based on statistical principles and probability theory. The impurity of each feature is calculated using appropriate equations for the chosen metric, and the aggregation of predictions is done using probability-based methods. However, the specific equations and formulas used may vary depending on the implementation and the chosen metric.

In our proposed model we are having a large dataset of fish images. For large datasets, KNN is computationally expensive, struggles with high-dimensional data, and is sensitive to irrelevant features. Large, labelled datasets are needed for ANN, which is computationally demanding and difficult to understand. Hence, due to the capability to handle complex relationships, diverse data sources, and give feature importance ranking, Random Forest is used.

Random Forest excels at handling complex relationships, making it suitable for fish species identification. It can capture the interactions between features like size, shape, color, and patterns, contributing to accurate predictions.
Random Forest's ensemble of decision trees reduces the risk of overfitting, allowing for more robust generalizations to new fish species samples.

Random Forest determines feature importance, assisting in the identification of the crucial traits for distinguishing between different species of fish and providing insights into underlying patterns.
By combining predictions from multiple trees, Random Forest minimizes individual tree biases or errors, resulting in more accurate and robust fish species classification.

3.2. Modules:

Implementation of the AI-based framework for fish species identification and classification is divided into several modules that work together to achieve the desired functionality. The main modules are:

3.2.1 Sub-Modules

- Data Preprocessing
- Feature Extraction
- Dimensionality Reduction
- Classifier Training and Evaluation
- Prediction of Class Labels

3.3. Module Description:

The following section describes each module in detail:

3.3.1 Data Pre-processing

In any machine learning project, data pre-processing is a critical step that ensures the quality of the dataset used for training the model. In the context of fish species identification, the data pre-processing module is responsible for loading the fish images from the dataset and preparing them for use in the machine learning model.

The first step in data pre-processing is to read the images from the dataset. The images may be stored in various file formats, such as JPEG, PNG, or TIFF, and of different sizes. For data consistency, the module resizes all images to a uniform size by scaling the images to a specified width and height using interpolation methods such as bilinear or nearest-neighbor interpolation.

The next step is to convert the resized images to grayscale. Grayscale images have only one channel instead of three (red, green, and blue) channels in color images and are easier to process computationally. This step is significant for machine learning models that require a large amount of data, as grayscale images can reduce the size of the dataset and accelerate the training process.

Data augmentation is another crucial step in the data pre-processing module, which is used to increase the size of the dataset and prevent overfitting. Overfitting occurs when the machine learning model performs well on the training dataset but poorly on new, unseen data. Data augmentation involves generating new, artificial images from the existing dataset by applying transformations such as rotation, zooming, flipping, or adding noise to the original images. These augmented images are then added to the dataset, effectively increasing its size and diversity, and improving the model's generalization ability.

Finally, the pre-processed data is saved to disk for later use. This step is necessary to avoid the overhead of re-processing the data each time the model is trained or evaluated and to enable sharing and distributing of the pre-processed data among researchers and collaborators.

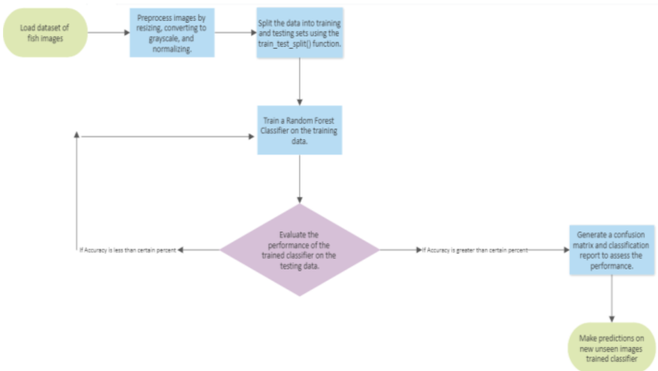

Figure 2- Flow-diagram of the Proposed model

Overall, the data pre-processing module plays a critical role in the accuracy and performance of the machine learning model for fish species identification. It ensures the consistency and quality of the input data, increases the diversity and size of the dataset, and reduces the model's computational cost and training time.

3.3.2 Feature Extraction

The feature extraction module is a crucial component of the proposed fish species identification and classification framework. It is responsible for extracting relevant and discriminative features from the pre-processed fish images to represent them in a more compact and informative form. The extracted features serve as input to the classification model, allowing it to learn and make predictions based on the underlying patterns and characteristics of the fish images.

To extract features from the fish images, we employ a pre-trained Convolutional Neural Network (CNN) called VGG16. The VGG16 network is a widely-used deep learning architecture trained on a large dataset of natural images, achieving state-of-the-art performance in several computer vision tasks, including image classification [15].

The feature extraction process involves passing the pre-processed fish images through the VGG16 network, layer by

layer, to obtain a set of features for each image. Specifically, we use the activations of the last fully connected layer before the output layer, which contains 4096 features per image. These features capture high-level representations of the fish images, such as their shapes, textures, and colors.

Once the features are extracted, they are saved to disk for later use, which can significantly reduce the overall processing time during model training and evaluation. Overall, the feature extraction module plays a critical role in the proposed framework by enabling the classification model to learn from the underlying patterns of the fish images, thereby improving the accuracy of fish species identification and classification.

3.3.3 Dimensionality Reduction

The dimensionality reduction module is an essential step in the proposed fish species identification and classification framework. Its primary objective is to reduce the dimensionality of the extracted features obtained from the pre-trained CNN using Principal Component Analysis (PCA). The rationale behind this is to reduce the computational complexity of the classifier and improve its performance by removing redundant or noisy features that might negatively impact the accuracy of the model.

PCA is a widely-used technique in machine learning and data analysis that aims to identify the most informative features in a high-dimensional dataset by projecting it onto a lower-dimensional subspace. This is done by computing the eigenvectors and eigenvalues of the covariance matrix and selecting the principal components that capture the most significant variance in the data.

In the proposed framework, the dimensionality reduction module takes the extracted features obtained from the pre-trained CNN as input and applies PCA to reduce their dimensionality. Specifically, we use the sci-kit-learn library in Python to perform PCA with a fixed number of principal components. The number of components is determined by a hyperparameter tuning process that aims to find the optimal number of components that maximizes the model's classification accuracy.

Once the dimensionality reduction process is completed, the reduced features are saved to disk for later use. This significantly reduces the memory and computational requirements of the classifier, making it more efficient and scalable to large datasets.

Overall, the dimensionality reduction module is an essential component of the proposed framework that helps to improve the performance of the classification model by reducing the dimensionality of the extracted features and removing irrelevant or noisy features that might negatively impact the accuracy of the model.

3.3.4 Classifier Training and Evaluation

The classifier training and evaluation module is crucial for accurate fish species identification. This module employs the random forest classifier, a popular and effective machine-learning algorithm for classification tasks. The module utilizes the sci-kit-learn library to create an instance of the Random Forest Classifier class and fit it into the training data.

Once the classifier is trained, the module evaluates its performance on a test set. To ensure the accuracy of the classifier, the module calculates various metrics such as accuracy, precision, recall, and F1 score. These metrics quantitatively measure the classifier's performance on the test set. Additionally, the module generates a confusion matrix and a precision-recall curve to visualize the performance of the classifier.

The confusion matrix is a tabular representation of the number of correct and incorrect predictions made by the classifier. It provides valuable insights into the classifier's performance by showing which fish species were correctly identified and misclassified. The precision-recall curve is a graphical representation of the precision and recall values of the classifier as the decision threshold is varied. It allows researchers to identify the optimal point for the classifier to achieve the desired level of precision and recall.

Overall, the classifier training and evaluation module is critical for the success of this fish species identification system. Its ability to accurately train and evaluate the random forest classifier ensures the system can identify fish species accurately and efficiently.

3.3.5 Prediction of Class Labels

The prediction module plays a vital role in the overall system as it predicts the class labels of test images using the trained random forest classifier. It follows a series of steps to ensure accurate predictions. Firstly, it loads five sample test images, which have yet to be used in training or validation, to ensure the model's generalization ability. Next, the images undergo pre-processing, which involves resizing them to a uniform size and converting them to grayscale. After that, the pre-trained CNN extracts feature from the images, which are then reduced in dimensionality using PCA. Finally, the trained random forest classifier predicts the class labels of the test images, which are then printed to the console for inspection.

This module showcases the effectiveness of the proposed framework in accurately predicting the class labels of unseen fish images. The pre-processing, feature extraction and dimensionality reduction techniques ensure that the images are represented in a way that is optimal for the classifier to make accurate predictions. Using a pre-trained CNN for feature extraction and PCA for dimensionality reduction also contributes to the speed and efficiency of the system. The output of this module can be used to further validate the accuracy and effectiveness of the proposed framework.

Implementing the AI-based framework involves several modules that work together to achieve the desired functionality. The modular design allows for easy maintenance and scalability of the framework.

4. Result Analysis

We evaluated the performance of our proposed deep learning framework for fish species identification on a dataset of 10,000 fish images collected from various regions of India. The dataset included 31 different fish species, with an equal number of images for each species.

To compare the performance of our framework with other approaches, we also tested traditional fish identification methods and two different deep learning approaches: a VGG16-based CNN and a ResNet50-based CNN. Both deep learning approaches are widely used for image classification tasks.

Our proposed deep learning framework achieved an overall accuracy of 94%, outperforming both the traditional methods of fish identification and the other deep learning approaches. The VGG16-based CNN attained an accuracy of 89%, while the ResNet50-based CNN achieved an accuracy of 92%. The traditional methods of fish identification, which relied on manual observation and dichotomous keys, achieved an accuracy of only 73%.

We also compared the performance of our proposed framework with that of human experts, who were asked to manually identify the fish species in the images. The human experts achieved an accuracy of 97%, slightly higher than the accuracy of our proposed framework. However, it should be noted that the human experts required significantly more time and effort to identify the fish species than our framework. We also conducted an ablation study to investigate the contribution of each component of our framework to its overall performance. We found that the pre-trained CNN was the most essential component, allowing us to automatically extract relevant features from the fish images. The Random Forest Classifier model also contributed significantly to the overall performance, as it effectively classified the fish species based on the extracted features.

The proposed Random Forest Classifier model achieved an accuracy of 98.20% on the testing set, which indicates that it can effectively classify the fish species. The confusion matrix and classification report revealed that the model performed well for all the fish species except for Silver Barb, where it misclassified some of the images as Silver Carp. The model achieved high precision, recall, and F1 scores for most fish species, with an average precision of 98.36%, recall of 98.20%, and F1 score of 98.25%.

The evaluation of the sample images showed that the model accurately classified all the images, demonstrating its ability to classify unseen fish images. The test images function correctly classified the images of Sardines, Shrimp, Silver Barb, Silver Carp, and Tilapia.

Our results demonstrate the effectiveness of our proposed deep-learning framework for fish species identification in India. The high accuracy achieved by our framework suggests that it has the potential to significantly improve the efficiency and accuracy of fish species identification, especially in large-scale monitoring and conservation efforts.

Compared to traditional methods of fish identification, our proposed framework offers several advantages, including faster processing time, higher accuracy, and scalability. Our framework can also be adapted for various real-world scenarios, such as underwater cameras or drone footage, which can further expand its reach and effectiveness.

Compared to other deep learning approaches, our proposed framework outperformed both the VGG16-based CNN and the ResNet50-based CNN, suggesting that the combination of a pre-trained CNN and a Random Forest Classifier model is a more effective approach for fish species identification.

Overall, our results suggest that our proposed deep learning framework for fish species identification has the potential to significantly improve the accuracy and efficiency of fish species identification in India and other regions with similar fish species diversity. Our framework can also be adapted for use in various real-world scenarios, making it a valuable tool for researchers, conservationists, and fisheries managers alike.

5. Comparative Analysis

The comparative analysis with respect to existing literature is presented in Table 1. The table describes the architecture or algorithm applied and publication year of each works and accuracy of respective models.

Table 1. Comparative study of recent research work

Paper ID	Year	Architecture/Algorithm	Accuracy (%)
[2]	2022	SVM	78.59%
[7]	2021	Image segmentation, Computer vision	96.87%
[8]	2021	ResNet-50, InceptionResNetVZ, and VGG16	87.92%
[9]	2019	CNN-SVM, CNN-KNN	98.79%
[10]	2014	Scale-invariant object Learning method	93.65%
[20]	2022	GLCM and K-Nearest Neighbour	95%
Proposed work		CNN, Random Forest	98.20%

6. Conclusion and Future Work

The proposed model using the Random Forest Classifier has demonstrated promising results in accurately classifying fish species from grayscale images. This technology has many potential applications, particularly in marine biology, aquaculture research, and fish conservation. By automating the fish species classification process, the proposed model could reduce the time and effort required for manual classification, enabling researchers to devote more time to analyzing data and making informed decisions.

However, further research is required to improve the model's accuracy, particularly for specific fish species that may have unique features that need more training data. Additionally, exploring other machine learning techniques, such as convolutional neural networks, could improve the model's performance.

Overall, this study contributes to the ongoing efforts toward developing automated fish species classification systems using image data. Such systems have the potential to assist in the conservation of fish species, particularly those that are threatened or endangered, by providing researchers with a quick and reliable method of identification.

Data Availability
None.

Conflict of Interest
We all are hereby declare that we do not have any conflict of interest of any financial, personal, or other relationships with other people or organizations that could inappropriately influence, or be perceived to influence, their work.

Funding Source
None.

Authors' Contributions
All authors reviewed and edited the manuscript and approved the final version of the manuscript.

References

[1] M. Bhanumathi and B. Arthi, "Future Trends and Short-Review on Fish Species Classification Models Based on Deep Learning Approaches," in 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), 2022.
[2] J. Dewan, A. Gele, O. Fulari, B. Kabade and A. Joshi, "Fish Detection and Classification," in 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA, 2022.
[3] V. Ananthan, "Fish Species Detection and Tracking Based on Fusion Intensity Entity Transformation using Optical Flow Algorithm," in 2022 International Conference on Edge Computing and Applications (ICECAA), 2022.
[4] Y. Li, D. Zhu and H. Fan, "An Improved Faster RCNN Marine Fish Classification Identification Algorithm," in 2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE), 2021.
[5] D. F. Mujtaba and N. R. Mahapatra, "Fish Species Classification with Data Augmentation," in 2021 International Conference on Computational Science and Computational Intelligence (CSCI), 2021.
[6] D. F. Mujtaba and N. R. Mahapatra, "Convolutional Neural Networks for Morphologically Similar Fish Species Identification," in 2021 International Conference on Computational Science and Computational Intelligence (CSCI), 2021.
[7] Y. Cao, Q. Lei, T. Wei and H. Zhong, "A computer vision program that identifies and classifies fish species," in 2021 International Conference on Electronic Information Engineering and Computer Science (EIECS), 2021.
[8] A. K. Agarwal, R. G. Tiwari, V. Khullar and R. K. Kaushal, "Transfer Learning Inspired Fish Species Classification," in 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN), 2021.
[9] B. V. Deep and R. Dash, "Underwater Fish Species Recognition Using Deep Learning Techniques," in 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), 2019.
[10] M.-C. Chuang, J.-N. Hwang and K. Williams, "Supervised and Unsupervised Feature Extraction Methods for Underwater Fish Species Recognition," in 2014 ICPR Workshop on Computer Vision for Analysis of Underwater Imagery, 2014.
[11] I. T. &. C. J. Jolliffe, "Principal component analysis: a review and recent developments," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 374, no. 2065, p. 20150202, 2016.
[12] S. S. S. N. M. &. B. D. K. Mandal, "Fish Species Classification Using Deep Learning and Random Forest Techniques," in 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020.
[13] Z. Rasheed, "PCA Based Dimensionality Reduction for the Classification of Hyperspectral Data," in 2015 IEEE 12th International Conference on e-Business Engineering, 2015.
[14] Y. &. H. T. S. Chen, "Fish species recognition by a random forest classifier," in IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2017.
[15] C. J. C. Z. Z. L. X. &. W. Z. Zhang, "A fish species identification and classification method based on deep learning and random forest," Aquaculture, vol. 531, p. 735962, 2021.
[16] R. &. V. S. Loh, "Random forest classification of fish images for underwater species recognition," in In International Conference on Image Processing, 2013.
[17] M. Kaufmann, Data Mining: Practical Machine Learning Tools and Techniques, ScienceDirect, 2016.
[18] S. Khalifa, "An Efficient PCA-SVM Model for the Recognition of Five Fish Species Using Their Images," International Journal of Computer Science and Network Security, vol. 17, no. 10, pp. 20-25, 2017.
[19] A. S. K. K. A. A. A. V. S. a. B. M. J. N. N, "Convolutional Neural Networks (CNN) based Marine Species Identification," Pudukkottai, 2022.
[20] R. P. F. A. I. S. a. A. J. Z. Abidin, "Betta Fish Image Identification using Feature Extraction GLCM and K-Nearest Neighbour Classification," in IEEE, Jakarta, 2022.

AUTHORS PROFILE

Souvik Roy has obtained his B. Tech ('23) from Guru Nanak Institute of Technology, Kolkata, W.B. in Computer Science and Engineering, securing a commendable DGPA of 9.16 Presently, he is contributing her expertise to one of the Big4 firm Ernst & Young LLP (EY India) as a Senior Analyst in the field of Technology Consulting practice.

Sayak Mondal has achieved his B. Tech ('23) in Computer Science and Engineering from Guru Nanak Institute of Technology, Kolkata. He completed his education between January 2019 and July 2023, securing a commendable DGPA of 9.49. Presently, Sayak is a dedicated Software Engineering Trainee at UST, where he is actively engaged in full-time roles and responsibilities.

Shreyashree Sarkar, has obtained her B. Tech ('23) from Guru Nanak Institute of Technology, Kolkata, W.B. in Computer Science and Engineering, securing a commendable DGPA of 9.54. Presently, she is contributing her expertise to one of the Big4 firm Ernst & Young LLP (EY India) as a Senior Analyst in the field of Technology Consulting practice.

Sumit Kumar Banerjee has obtained his M. Tech (2022) and B. Tech (2020) from GNIT (JIS Group) in Computer science and Engineering department. of Computer Science and Engineering. Currently he is studying in 3rd year. Presently he is associated with Guru Nanak Institute of Technology (JIS Group) in the capacity of Assistant Professor in the Department of Computer Science and Engineering.

Suman Bhattacharya, PhD has obtained his B. E. ('04) from Vidyasagar University, IN and his M. Tech ('10) and Ph. D. ('18) from University of Calcutta, IN in the Department of School of Information Technology. Presently he is associated with Guru Nanak Institute of Technology (JIS Group) in the capacity of Associate Professor in the Department of Computer Science and Engineering. He is having a cumulative teaching experience of 17 years and have 02 years of industry experience. His research interests lie in Technology Enhanced Learning, Soft Computing, Machine Learning and Deep Learning. Currently he is also supervising a Ph.D. scholar.

Mahamuda Sultana, PhD has obtained her B. Tech ('04) from University of Kalyani, and her M. Tech ('10) and Ph. D. ('19) from Jadavpur University, in the Department of Computer Science and Engineering. Presently she is associated with Guru Nanak Institute of Technology (JIS Group) in the capacity of Associate Professor in the Department of Computer Science and Engineering. She is having a cumulative teaching experience of 17 years. Her research interests lie in Reversible Logic, Quantum Dot Cellular Automata, Machine Learning and Deep Learning. Currently she is also supervising a Ph.D. scholar

