[1]Dr. Mohammad Ahmar Khan,

[2]Dr. Syed Ahsan Jamil,

[3]Sameera Vellanki,

[4]Dr. M. Latha,

[5]Kamred Udham Singh

[6]Dr. Mridula Batra

Creation of Machine Learning-Based Fish Classification Systems Based on Morphometric and Mathematical Transform Data

Journal of Electrical Systems

Abstract: - The accurate classification of fish species is crucial for various ecological, conservation, and fisheries management applications. Traditional methods relying solely on manual observation or morphological characteristics can be time-consuming and prone to errors. In recent years, advancements in machine learning (ML) techniques coupled with morphometric and mathematical transform data have shown promising results in automating and enhancing fish species classification processes. This paper presents a comprehensive review of the development and application of ML-based fish classification systems utilizing morphometric measurements and mathematical transforms. The core of this paper focuses on the integration of morphometric and mathematical transform data with ML algorithms for fish classification. In this review commonly employed ML algorithms, including but not limited to support vector machines (SVM), artificial neural networks (ANN), random forests (RF), and convolutional neural networks (CNN). Furthermore, this work discuss feature selection techniques to optimize classification performance and reduce dimensionality. These may include the incorporation of multimodal data sources, transfer learning approaches, and the development of user-friendly tools for field biologists and conservationists. Overall, this paper serves as a comprehensive guide for researchers and practitioners interested in leveraging ML techniques for accurate and efficient fish species classification.

Keywords: fish species classification, machine learning, morphometric measurements, mathematical transforms, support vector machines, artificial neural networks, random forests, convolutional neural networks, feature selection, multimodal data, transfer learning

Introduction

In terms of human health and trade, fish is an essential resource. Many fish species are used as food all around the globe due to their delicious flesh and high nutritional and protein content. More over one million metric

[1] Dept. of MIS, CCBA, Dhofar University, Salalah, Sultanate of Oman

mkhan@du.edu.om

[2]Dhofar University, Salalah, Sultanate of Oman

[3]Research Scholar, Department of English, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India

sameeravellanki1987@gmail.com

[4] Associate Professor, Department of English, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India Email: mlatha@kluniversity.in

[5]Assistant Professor, School of Computing, Graphic Era Hill University, Dehradun; Adjunct Professor, Graphic Era Deemed to be University, Dehradun, Uttarakhand-248002, India.

Mail id- kusingh@geu.ac.in

[6]Associate Professor Centre for Distance and Online Education, MRIIRS

E- mail Id adir@manavrachnaonline.edu.in

Orcid id 0000-0003-2472-1913"

tonnes of fish and fish products leave India every year, accounting for about 10% of the country's overall agricultural product exports. A significant portion of India's GDP comes from the export of fish species used for both decorative purposes and as food[1]. Sorting fish into different species is a necessary step in the fish trading industry for exporting both raw and processed fish products. Manually sorting fish is laborious and prone to mistakes. Thanks to advancements in machine learning and image processing, automatic methods for fish species separation have emerged. Using state-of-the-art methods in machine learning and image processing, the paper proposes many models for fish species categorization. From a business perspective, the research looks at how wrasse and tuna are classified as decorative and commercial food fish, respectively. Using an ELM classifier and a multidomain feature-based approach, wrasse fishes are grouped into genus and species[2]. Tuna and other commercially significant fish species are first separated from commercial food fishes using feature descriptors and models based on pre-trained convolutional neural networks (CNNs). We also use a standard transform-based technique and CNN architectures to analyse the categorization of economically significant tuna into species kinds. Ornamental Wrasse fishes are grouped using multidomain traits. This method employs the following tools: colour histograms, histograms of oriented gradients (HOGs), and statistical features derived from wavelet transform sub bands. A combination of dimension reduction methods is used to reduce features. Using a range of classifiers, we assess how well the combined and reduced feature sets perform in classifying genus and species[3].

Differentiating between species of tuna is crucial in the export business. Using both traditional approaches and CNN models, we assess the accuracy of bigeye, skipjack, and yellow fin tuna species categorization. For the purpose of species-level tuna fish classification, a wavelet-transform-based deep architecture is suggested. A lifting approach is used to learn a dual tree complex wavelet transform from a set of seed pictures. Features in the scattering architecture are extracted using this wavelet. Scattering networks are used to extract translation and deformation invariant features from pictures, which are then evaluated using machine learning methods[4][5]. Additionally, a region-based technique is used to assess how well deep learning architectures perform in tuna species categorization. Partitioning each picture into a head, belly, and tail is how the region-based technique works. Separate convolutional neural network (CNN) models are trained using each of these part pictures. The research suggests two separate pre-trained CNN models depending on regions. One model integrates pre-trained convolutional neural network (CNN) models using the sum rule, while the other employs a super learner technique to integrate pre-trained convolutional neural network (CNN)-grouped 2D-LBP models[6]. We look at a region-based split-octonion CNN model to enhance the tuna classification system. Recently, complex algebra has been applied to CNN convolutional layers, which developed the model. The model incorporates a new attention module called split-octonion channel. In order to enhance performance, CNN models make use of attention modules to draw attention to the most important parts of pictures. The final prediction is the result of applying region pictures to split-octonion models using the attention module and then merging the models. An assortment of performance criteria, including recall, accuracy, precision, and F-score, are used to evaluate each of the suggested models[7]. A confusion matrix is generated for every system so that its performance metrics may be assessed. On the basis of these performance metrics, the proposed systems are contrasted with the cutting-edge systems[8].

Fishes are gill-bearing marine animals that can be found in almost any body of water, from high mountain streams to the deepest ocean depths. They have the most species diversity of any group of vertebrates, with approximately 34300 species [1]. Marine fishes are categorised according to their dwelling place. Fishes that live in the pelagic zone of the sea, which is neither near the bottom nor near the shore, are known as pelagic fishes. Mackerel, tuna, dolphin, flatfish, and hagfish are examples of pelagic fishes [2]. Demersal fishes are those that live on or near the seafloor, while reef fishes are those that live near coral reefs. Coral fishes from the Labridae, Chaetodontidae, and Cichlidae families are commonly used as ornamental fishes because of their vibrant colours. Fishes play an important role in culture, diet, religion, and as subjects in books, sculpture, and films. They are a valuable asset for humans in terms of both health and commerce [3]. Fishes are delicious and are an excellent source of protein, vitamins, and minerals and it is a primary source of animal protein for around one billion people [4]. Many fish species are consumed as food all over the world. It is beneficial to include fish in one’s daily diet because it reduces the risk of lung cancer, stroke, and dementia. Whitefish, oily fish and shellfish are the three types of fishes that are consumed by humans [5]. Haddock and seer fishes belong to the category of whitefish, which has low-fat content. Oily fish contains 10-25% fat, vitamins A, D, E and K, and essential fatty acids. Shellfishes are an excellent source of zinc, which is essential for skin and muscle health. Fishes come in various shapes, colours, and sizes, with some species growing up to several metres in adulthood. They show a variety of body shapes, including fusiform, elongated, compressed, and depressed [6]. The fins are another distinguishing feature of fish species. Dorsal, anal, adipose, pectoral, pelvic, and caudal fins are the different types of fins found in fishes. It is extremely difficult and challenging to separate fishes into their species because of these variations. Identification of fish species is needed to determine the abundance of a specific species in an ecosystem[9].

The method of collecting and evaluating the abundance of a fish population to obtain information about the impact of fishing and other natural changes on the stock is known as stock assessment [7]. This assessment offers information about an ecosystem's health. Identification of fish species is critical for obtaining knowledge about an ecosystem and it would serve as an indicator of habitat on the verge of extinction. Ecologists may use stock analysis to determine the health of the habitat and prepare the appropriate measures to protect the environment. The fisheries industry is one of the most profitable and successful sectors, which can revitalize the economy of a country. India has a vast marine coastline of 7516 kilometres, with 3827 fishing villages and approximately 1900 fish landing stations [8]. With it's wide coastline, fishing is a major sector in India, which employs 13 million people. In many parts of the world, the growing popularity of keeping ornamental fishes in home aquaria has increased the demand for these fishes. With roughly 100 million hobbyists worldwide, ornamental fishkeeping is the second most popular hobby. India is ranked 31st in the world in terms of ornamental fish exports, bringing in a revenue of Rs. 8.40 crores [9]. Fishes and fish products exports are the main contributors to India's agricultural GDP, according to the economic survey for the year 2018-19 [10]. The export of commercial fishes, ornamental fishes, and different fishery products brings in a total of 7.2 billion dollars to the country. The most popular commercial fishes exported from India are mackerel, tuna, seer fish, promfret, cuttlefish, squid, and crustaceans [11]. In the export industry, fishes are separated into different categories in order to export raw fishes and processed fish products. The by-products of various fish species obtained from the fish processing industry are a good source of fat, minerals, and proteins. The types of byproducts derived from fishes vary depending on the species. Manual separation of fishes into species types is labourious. To expedite the process in these export industries, a quick and efficient system for classifying commercial and ornamental fishes is required

Fish is one of the world's most important commodities, accounting for around 10% of total agricultural exports. For exporting fishes and processed fish products, as well as for stock analysis, ornamental and commercial fishes must be classified. Categorization of fish species based on morphological features is the most realistic, fast, and low-cost process. Experienced people with knowledge gained from long-term observations separate the species based on morphological characteristics. The sorting of fishes is a labour-intensive process. It is a time-consuming and exhausting operation both on land and on commercial ships. Some export industries employ automated systems that sort fish according to their size or weight. Such existing methodologies of sorting may not yield fruitful results, and may even spoil the fish, thereby lowering its quality. The developments of image processing techniques have aided the advancement of morphometric methods for species identification [12]. The use of an image-based system allows for the automated separation of fish species, making the process more efficient and quicker, and causing less spoilage to the fish. Sorting fishes according to their species can help industries speed up their processes even more. The availability of various feature descriptors and machine learning algorithms has paved the way for the development of various automated fish species classification systems. Novel and more powerful feature descriptors and classifiers that can improve system performance have recently been presented in the literature [13]. The use of these descriptors and classifiers for fish classification has not been investigated much. The aim of this paper is to build better automated classification frameworks for fishes using newly developed image processing algorithms and machine learning classifiers. The research focuses on the classification of ornamental and commercial food fish, both of which are important in terms of export. Among the ornamental fish category, Wrasse fish species of the Lambridae family are considered for the study as they are useful for identifying the health of coral reefs and for export industries. While tuna is chosen from among the various commercial food fishes because it has a high economic value in comparison to other fishes

Advances in image feature descriptors have contributed to the development of automated image classification systems. For the automated image categorization process, two approaches are employed: the conventional method and the deep learning method. The conventional approach tries to extract information from the input images using various feature descriptors, and then trains a classifier to classify the data according to these features. The deep learning method, on the other hand, uses a layered architecture to learn the characteristics and classify the images. Fig. 1 shows the block diagram of the two approaches used in automated image classification Acquisition of images is the first step in the image classification scheme. The quality of the images has a profound effect on system performance, so it must be captured with great care. The acquired images are often affected by blur, noise, lighting fluctuations, and background. The images are subjected to several preprocessing processes in order to remove these effects and extract better features, thereby improving system performance

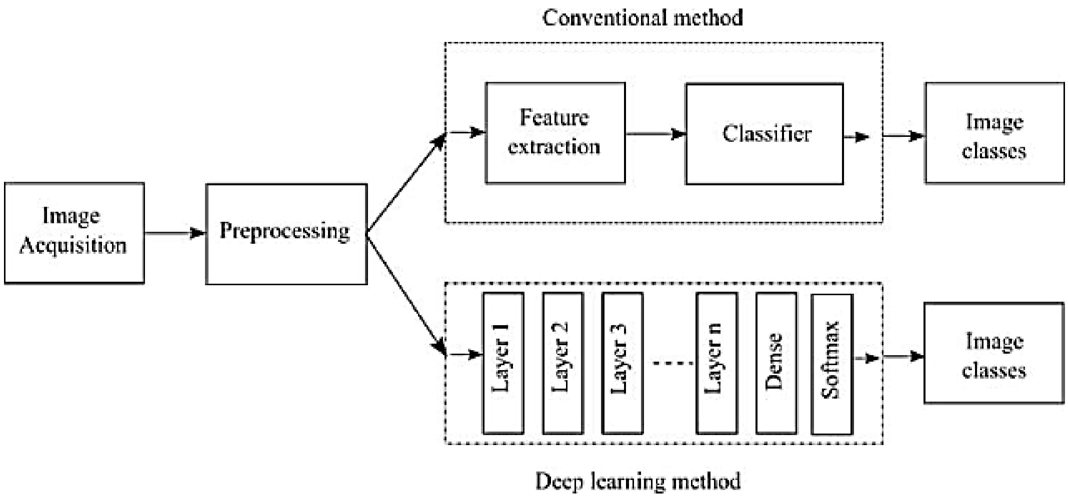

Figure 1.Simplified block diagram of image classification approaches.

In The Conventional Approach, Preprocessed Images Are Applied To Various Feature Descriptors To Extract Features. The Features Extracted From The Input Data Are Normalised And Are Applied To Different Classifiers. In The Deep Learning Approach, The Input Images Are Passed Through A Layered Architecture, With The Final Layer Forming The Classifier Section. The Layers Learn The Characteristics Of The Data And Classify Them Into Different Classes Simultaneously.

Deep multiresolution approach with learned complex wavelet transform for tuna classification

This study proposes a multiresolution method for taxonomy of economically significant tuna species. The categorization is done using a two-layer scattering network design that uses a signal-matched complicated wavelet transform. In order to create a wavelet that is compatible with the target signal, a lifting strategy is used[14]. A complicated wavelet transform that is signal-matched between two trees is then generated using the learnt wavelet. Features that are invariant under translation and deformation may be extracted from the scattering network using the complex wavelet that was constructed. Various classifiers are used to assess the acquired feature set, and their performance measurements are derived.

Region-based pre-trained CNN models for tuna classification

Two methods based on regions are suggested for classifying tuna species using CNN models that have already been trained. Separated into the three areas of the body—head, abdomen, and tail—are the input pictures. The first setup uses these area pictures to train four models: MobileNet, Xception, VGG16, and VGG19. The total rule takes the four models' scores for a given area and adds them together[15]. After that, we get the final image-wise forecast by adding together the sum rule scores for each area and then voting by majority. The second technique that has been suggested uses the activations of the top region model to extract textural information. In order to extract textural information from the model's activations, a new grouped 2D-LBP (G2DLBP) descriptor is suggested. In order to determine how well these attributes categorise a specific area picture, several classifiers are used. The ensembling of a region-based G2DLBP model yields the final forecast.

Multidomain Feature based System for Wrasse Fish Classification

The ecosystems of coral reefs are intricate and teem with many kinds of life. Coral reefs are home to more than four thousand different kinds of fish. Reef fishes are often kept as decorative fish due to their vibrant colours and patterns. As a hobby, keeping colourful ornamental fish has been around for a long time and is still quite popular today. Only around one percent of the ornamental fish traded worldwide comes from India. The majority of the fish life on tropical coral reefs are wrasses, which are members of the Labridae family. Marine ornamental fish from the wrasse family are much sought for in India. Stock analysis is used to determine the health of a coral reef by measuring the quantity of fishes in the reef. Of all the marine fish families, wrasses have the most members (about 500 species spread over 60 genera). So, you can tell how healthy a coral reef is by looking at the wrasse population. Consequently, coral fish species identification is vital for stock assessments and exports. Since it may expedite the process in the export business and stock analysis, an automated method for Wrasse fish categorization is more appealing. In this chapter, we provide a multi-stage automated system for genus and species level Wrasse fish classification utilising multi-domain characteristics extracted from fish photos. There is a hybrid approach to classification that makes use of both spatial domain information and characteristics extracted from wavelet sub-bands. The spatial domain characteristics include colour histograms, LBPs, and HOGs, whilst the frequency domain features are statistical features derived from sub-band pictures acquired using wavelet decomposition. Using the combined feature set, we can assess how well various classifiers perform. The performance of an ensemble dimension reduction approach is assessed while it generates a feature set with a decreased number of attributes. This is the first effort at multi-level Wrasse fish classification utilising a mix of wavelet, colour histogram, LBP, and HOG features.

Overview of the system

Species of wrasse are known for their vibrant colours and ability to blend in with their natural habitat. Recognising these fishes in their native habitat is no easy feat. At first, the suggested system uses landmark locations on the fish's body to separate it from its environment. Prior to being applied to various feature descriptors, the segmented pictures undergo enhancement and denoising. The colour histogram characteristics are derived from the statistical features found in each image's R, G, and B plane histograms. Applying the HOG description yields structural characteristics, whereas the LBP histogram provides textural features for each picture. To extract frequency-domain characteristics from images, the wavelet transform is used for decomposition. An ensemble dimension reduction approach is used to apply all the characteristics together. Using 10-fold cross validation, we test the condensed feature set with several well-known classifiers. The first step of the classification process is to group fishes into genus groups, and the second step is to group them into species groups. The suggested multistage Wrasse fish categorization system is shown schematically in Fig. 2.

Details of datasets

The work uses a dataset of 371 images of which 132 images are of genus L. Halicoeres, 141 images of genus L. T halassoma and 98 images of other genus. Images were collected from QUT database [64], FishBase database [1], and from different professional and academic repositories. 117 images were collected from QUT database, 58 from FishBase, and 196 from other repositories. The species H. margaritaceus, H. marginatus, and H. melanurus are used from the genus L. Halicoeres. T. lunare, T. lutescens, and T. quinquivittattum are the three species of L. T halassoma used in the dataset.

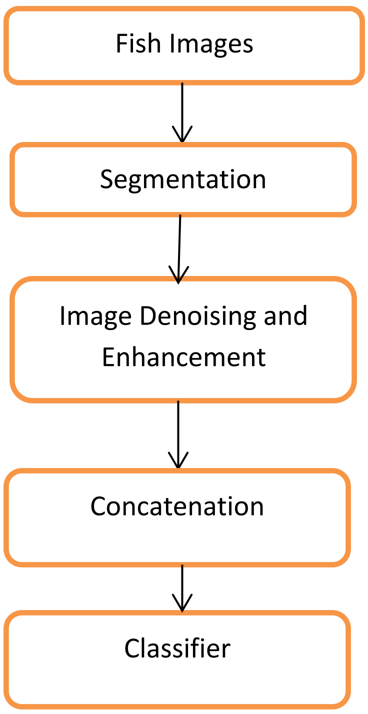

Figure 2.Schematic diagram of the multistage Wrasse fish classification system.

Feature reduction

After merging features from the frequency and spatial domains, a feature vector of size (1 x 190) is produced. Features derived from the data can be redundant due to the high number of correlations. By converting data from higher dimensions to lower ones, dimensionality reduction approaches aim to remove this unnecessary information from the feature set. Methods for reducing the number of dimensions include principal component analysis, MCML, large margin nearest neighbours, GDA, NCA, multi-dimensional scaling, Probabilistic principal component analysis (Prob PCA), and factor analysis.

PCA: This method maps a high dimensional data to low dimension such that the mapping minimises the squared error or maximises the variance [75]. Feature reduction is done by finding the eigen values and eigen vectors of the covariance matrix of the dataset. First k eigen vectors of the covariance matrix are used as the reduced feature set. This reduces the dimension of the original dataset to k-dimensional space.

MCML: This technique maps all the points in the same class to a single location and all other points to another in the feature space [76]. The points that are similar will be close to each other whereas those which are different will be far. Mahalanobis distance is used to generate similarity metric between two points in the feature space. Conditional distribution of a point over all other points is determined and based on the value, the entire datapoint is given a bi-level distribution. Hence, MCML tries to collapse each class to a single point in the feature space.

LMNN: This method tries to improve the classification performance of k-nearest neighbor classifier by learning a Mahalanobis metric [77]. Two types of neighborhoods are learned for an input in a given dataset. For each input data, a set of target neighbors are identified. The k nearest points based on Euclidean distance between input and target points are taken as target neighbors. These points are the actual neighbors of the input data point. Second type of neighbors are those which belong to a different class but are seen within the neighborhood of the input data. These neighbors are called Imposters. Objective function of LMNN tries to pull the input point and target points closer while it moves the imposters far away. Using the objective function, LMNN learns a projection of input dataset to a lower dimensional space

GDA: LDA is applied for linear classification problems whereas, GDA is designed for non-linear applications [78]. A kernel function is applied on the original feature space to transform it into a different higher dimensional space. GDA finds vectors that best discriminate between classes and gives independent features which describes the data.

NCA: This technique finds a distance metric which can improve performance of k-NN classifier [79]. A metric is learned by applying a linear transform on input space. In the transformed space, neighbors are assigned by using a stochastic selection rule. The rule gives the probability of a point to be correctly classified and the objective function tries to maximize the number of points classified correctly by the analysis.

MDS: It is a family of tools used for data analysis and dimension reduction [80]. Some non-linear transformations such as raw stress and Sammon cost functions are used to map high dimensional data to low dimension. A map is generated based on the relationship between items such that similar items are located closer and dissimilar far apart. Distance between items on the map provides information about the similarity between the items. Probabilistic

PCA: It is proposed by Tipping and Bishop [81], where a Gaussian noise model is used for dimension reduction. Conditional distribution of data obtained from these models are taken as the PCA scores. Prob PCA becomes classical PCA when the noise variance of the model is infinitesimally small. The principal components in Prob PCA are maximum likelihood estimates of the model parameters and Prob PCA shows better performance than classical PCA as it can handle missing data problems

Factor analysis: It is a dimensionality reduction technique that uses unsupervised learning [82]. FA reduces the dimension of the data by representing the original data matrix as the product of two smaller matrices added with some random noise. These matrices are the factor matrix and the factor loading matrix. A joint embedding of the dimension reduction methods gives better performance than any single method. The features obtained from spatial and frequency domains are reduced by an ensemble dimension reduction technique [83]. Each dimensionality reduction technique is applied on the feature set individually and the performance of the reduced feature set is analysed. Among the dimension reduction techniques, the methods that show best performance are used in the ensembling method. In the work, the feature set is reduced to a size of 75 by each of the dimension reduction method. The first two predictors from the best three techniques are combined to form the final reduced feature set. The performance of the reduced feature set is analysed using different classifiers.

Conclusion

A major step forward in aquatic ecology, conservation, and fisheries management has been the creation of machine learning-based fish categorization algorithms that use mathematical transform data and morphometric measurements. In this article, we have covered every angle of the methods, strategies, and procedures used in this field. The use of machine learning algorithms in conjunction with morphometric and mathematical transform data has shown encouraging results in the automation and improvement of fish species categorization operations. For complicated data representations, popular algorithms like convolutional neural networks, support vector machines, random forests, and artificial neural networks have shown to be effective in properly recognising fish species. For academics and industry professionals looking to classify fish species efficiently and accurately using machine learning, this study is a must-read. We can improve our knowledge of aquatic ecosystems and help implement better management and conservation plans by adopting novel approaches and multidisciplinary cooperation.

REFERENCES

1. M. A. Iqbal, Z. Wang, Z. A. Ali, and S. Riaz, "Automatic fish species classification using deep convolutional neural networks," Wirel. Pers. Commun., pp. 1–11, 2019
2. S. Guennouni, A. Mansouri, and A. Ahaitouf, "Biometric systems and their applications," in Visual Impairment and Blindness-What We Know and What We Have to Know. IntechOpen Limited, London, 2019
3. M. A. Islam, M. R. Howlader, U. Habiba, R. H. Faisal, and M. M. Rahman, "Indigenous fish classification of bangladesh using hybrid features with svm classifier," in Proc. of Int. Conf.on Comput., Commun., Chem., Materials and Electron. Engg. IEEE, Rajshahi, Bangladesh, July 11-12, 2019, pp. 1–4

4. U. Andayani, A. Wijaya, R. Rahmat, B. Siregar, and M. Syahputra, “Fish species classification using probabilistic neural network,” in Proc. of Int. Conf. on Comput. and Appl. Inform., vol. 1235, no. 1. IOP Publishing, Sumatera Utara, Indonesia, December 1-2, 2019, p. 012094
5. M. K. Alsmadi, M. Tayfour, R. A. Alkhasawneh, U. Badawi, I. Almarashdeh, and F. Haddad, “Robust feature extraction methods for general fish classification,” Int. J. Electr. Comput. Eng., vol. 9, no. 6, pp. 5192–5204, 2019
6. M. K. Alsmadi, “Hybrid genetic algorithm with tabu search with back-propagation algorithm for fish classification: Determining the appropriate feature set,” Int. J. Appl. Eng. Res., vol. 14, no. 23, pp. 4387–4396, 2019
7. N. Kiaee, E. Hashemizadeh, and N. Zarrinpanjeh, “Using GLCM features in haar wavelet transformed space for moving object classification,” IET Intell. Transp. Syst., vol. 13, no. 7, pp. 1148–1153, 2019
8. B. Rekha, G. Srinivasan, S. K. Reddy, D. Kakwani, and N. Bhattad, “Fish detection and classification using convolutional neural networks,” in Proc. of Int. Conf. on Comput. Vision and Bio Inspired Comput. Springer, Coimbatore, India, November 18-19, 2019, pp. 1221–1231.
9. A. A. dos Santos and W. N. Goncalves, “Improving pantanal fish species recognition through taxonomic ranks in convolutional neural networks,” Ecol. Inform., vol. 53, p. 100977, 2019.
10. H. S. Chhabra, A. K. Srivastava, and R. Nijhawan, “A hybrid deep learning approach for automatic fish classification,” in Proc. of Int. Conf. on Emerg. Trends in Inform. Techn. Springer, New Delhi, India, 2020, pp. 427–436.
11. Jalal, A. Salman, A. Mian, M. Shortis, and F. Shafait, “Fish detection and species classification in underwater environments using deep learning with temporal information,” Ecol. Inform., vol. 57, p. 101088, 2020.
12. X. Lan, J. Bai, M. Li, and J. Li, “Fish image classification using deep convolutional neural network,” in Proc. of Int. Conf. on Comput., Inform. Process. and Adv. Edu. IEEE, Ottawa, ON, Canada, October 16-18, 2020, pp. 18–22
13. J. Mei, J. N. Hwang, S. Romain, C. Rose, B. Moore, and K. Magrane, “Video-based hierarchical species classification for longline fishing monitoring,” in Proc. of Int. Conf. on Pattern Recog. Springer, CuricA, Chile, January 10-15, 2021, pp. 422–433.
14. X. Xu, W. Li, and Q. Duan, “Transfer learning and SE-ResNet152 networks-based for small-scale unbalanced fish species identification,” Comput. Electron. Agric., vol. 180, p. 105878, 2021
15. N. Abinaya, D. Susan, and R. Kumar, “Naive Bayesian fusion based deep learning networks for multisegmented classification of fishes in aquaculture industries,” Ecol. Inform., vol. 61, p. 101248, 2021.