



AUTOMATIC FISH ANALYSIS AND CLASSIFICATION USING AI APPROACH

¹Vidya A. Huse, ²Deepali G. Chaudhary, ³Pravin L. Yannawar, ⁴Bharti W. Gawali

¹Student, ²Student, ³Professor, ⁴Professor

¹Department of Computer Science & Information Technology,

Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS), India

Abstract: Automatic classification of fish and its analysis is a stimulating task due to the multifaceted textures, appearance, and diversity found in fishes. Automatic fish identification and classifications have enormous probable applications in expanded parks associated with fishery sciences and domestic commercial uses of fish. Traditionally, the approach of fish classification is founded on various morphological, geometrical, location features, and needs the support of taxonomists to identify fish classes. The dataset used in research work is collection of 500 fish images from Fishbase and Fish4Knowledge. This complex collection, makes the classification process challenging due to variance in data acquisition sensors, location, scale, etc. Since, the considered dataset is complex, presented research work presents a novel approach of Artificial Intelligence (AI) methods like Local Binary Patterns (LBP), Self-Organizing Maps (SOM), and Principal Component Analysis (PCA) exploiting computations of prominent morphological features of fins, gills resulting 97% accuracy.

Index Terms - Local Binary Pattern (LBP), Self-Organizing Map (SOM), Principal Component Analysis (PCA). Fin's, Gill's, Artificial Neural Network (ANN), Contrast Limited Adaptive Histogram Equalization (CLAHE).

I. INTRODUCTION

Recent advancement in electronic infrastructure, imaging and visualization techniques has generated opportunities in developing automatic classification system based on image processing methods. Numerous researchers and analysts have built up an assortment of the automatic ordering of fishes, fish acknowledgment frameworks for the purpose of the fishing business, and different AI-based frameworks are under development.

Artificial Intelligent is used to make an intelligent machine. It is a technique of getting machines to work and behave like a human. In the resent AI has been able to accomplish this by creating machines and robots that are being used in a wide range of fields including healthcare, robotics, marketing business analytics, finance, education, and many more. Machine learning is a subset of AI that focused on getting machines to make decisions by feeding them data on the other hand deep learning is the subset of machine learning that uses the concept of a neural network to solve complex problems so to sum it up in AI. Deep learning is a type of ML and AL that imitates the way how humans gain a certain type of knowledge.

AI is not only restricted to machine learning and deep learning but it also covers a vast domain of fields including natural language processing, computer vision, robotics, expert system, object detection, and so on.

The strategy for deciding fish freshness is through tactile valuation of the gills and eyes of the fish. Despite of the fact that these components are discernible by the unaided eye, a customary shopper can't precisely recognize the freshness of the fish similarly to some profoundly talented staff. [2-5]

We classify fish or detect objects using some steps like data or image acquisition, preprocessing of the image, extracting features from the image then detecting the image, and presenting results. On the basis of the previous studies done on the different types of methods and algorithms, we have used the canny edge detection technique for detecting a cleared boundary of images for better results. We have used the k-mean algorithm for classification and PCA for dimensionality reduction. LBP method was also used for the image detection process to get a better result.

Authors have proposed an imaging-based examination mechanism where an inimitable identifier (ID) was given to each fish under examination. Most accustomed methodologies utilize a fish picture wherein extraction is simple where the fish locale stands out from a white or uniform foundation. This exploration acquaints a methodology by giving a few components that focus on manual activity. In the proposed approach, we acknowledge the fish picture with a novel convoluted mechanism to explore proficient highlights for fish acknowledgment in pictures. The process helped in characterizing mathematical highlights, the sack of visual words (BoVW) models, and surface highlights. The proposed method was implied on 129 fish species samples of different photographic conditions and it was seen that a mix of mathematical highlights and BoVW models acquired the most elevated acknowledgment precision.

In our day-by-day voyages, we frequently experience an enormous numeral of common articles, from the study, it is found that 28000 fish species exist and are placed in 5 different classes as hagfish, lampreys, cartilaginous fish, ray-finned fish, and lobe-finned fish. The fishes can be differentiated on the basis of their structure [17]. For example, plants, fishes, creatures, and winged animals. With the rapid expansion in the numeral of versatile terminals (i.e. handsets and tabs), apprehending pictures of such characteristics,

articles become simple. In that capacity, there have been various efforts carried on the picture-based acknowledgment of common articles. Concerning photograph-based acknowledgment of common articles, numerous scientists have proposed classification of flower [1,2], leaf [3-5], and birds [6-7]. Nowadays most up-to-date records processing irrespective of its type and volume, researchers are imparting AI methods for training models and advanced computing networks that are designed and tested in a more diverse set of data. As a result, the capacity and complexity of such networks can exceed the needs of a small data set, which leads to a significant impact when training from the start. Taking into account, this proposed work the performance of various networks was examined and also compared the performance of each network in two situations vise the first, is about starting a pre-trained weight training and the second is a network beginning with random weights.

The problem statement is extracted from the previous studies when we used some problems that occurred in digital images like noise, distortion, lack of information, segmentation error, overlap, and occlusion of the object in a color image. Classification techniques and identification of images gained a lot of attention in the last few years. Fish classification and identification is the active area in the agriculture domain and consider as potential research in utilizing the existing technology for encouraging and publishing the agriculture researches ahead. Existing systems are still limited in their ability to detect or classify fish. Due to lack of knowledge or information about the fish many people go in danger of having some reaction on skin. Because fish have the same body structure the human eye is not able to identify [16].

II. RELATED WORK

The research work on the automatic classification of fish was reported by various authors. A picture-based fish species technique was proposed with only six species. The gauges were used to measure the various fish highlights, as observed by a camera opposite to the transport line. The highlights are widths and statures at different areas of fish. These highlights were utilized as info esteems for a neural organization [8]. A programmed shading leveling model was proposed on the shading adjustment strategy. The developed technique was applied on a fish submerged picture for the fragmentation of fish areas. The venture centers for building data frameworks for an aquarium. Different highlights are computed in this study including mathematical highlights, shading, surface, and movements. To kill pointless or excess highlights a component decrease measure was additionally applied. To characterize the fish in one educated animal group Quadratic Bayes classifier was used. In this study species of 12 fishes and 1346 examples were used [9].

A study proposed framework for perceiving disconnected fish examples. Here the fish picture is first trimmed for elimination ventral piece. A shading histogram is determined in the study and three highlights as energy, standard deviation, and energy were extracted. Further, one the dim level co-occurrence framework (GLCM); and two highlights of middle and fluctuation esteems were straightforwardly determined. For ordering the task, a multilayer feed forward neural organization model with a back-proliferation classifier was used. The species used for the study were 20 [10]. A study was proposed for live fish acknowledgment followed by various order leveled techniques. The data was recorded by submerged cameras in unlimited regular habitat. Fish fragmentation was done using Grab cut calculation. Here 66 highlights were separated, as comprised of a blend of shading, shape, and surface highlights from various fishes. The number of species in this study is ten and 3179 examples of fish pictures [11].

The study presents an intuitive framework made at an aquarium that perceives different species of fishes and displays moment data to clients. Sectioning of fishes was done using preprocessing method and Support Vector Machine (SVM) was used for the reorganization of 20 species of fishes on the basis of shading and shape [12]. A PC framework was build and was known as FIRS. These pictures of fishers were captured underneath of bright light bulb against the white plastic plate. Fish areas were separated by applying a threshold. Eight highlights were characterized, and a counterfeit neural organization was utilized for the acknowledgment cycle. The examination was made on 30 different fish species [13]. Shape coordination and geodesics for the acknowledgment were proposed and the developed technology was autonomous of interpretation, scale, turn, and beginning stage determination. A Shape Queries Using Image Databases (SQUID) was used for the study [14].

On the basis of previous studies and papers, we have observed some techniques and datasets with their results on the fish image. There is much research done on an underwater live fish image or video recorded on the collected data from the international levels. Some of the authors used the Fishery Department of the federal university of technology Akure(FUTA), some datasets with the name of QUT fish data.zip, Fish images are carried out by placing the individual fish sample on a clear white plane and photographing five times using cannon power A14000HD digital camera (Malaysia and Nigeria). The details are presented in table 1.

Table 1 Available dataset with result

Sr.no	Experiment conducted on	Methods and Technique	dataset	result	Refarances
1	30 Fish spices, 600 for training and 300 for testing	EDM,ANN	Collected from Thailand in Southeast Asia(Andaman Sea)recorded during 1991-1998	81.67	Chomtip Pornpanomchai, et.al, "Shape-and Texture-Based Fish Image Recognition System", 2016.
2	18 fish spices	K-NN,SVD	EBUTE River in Epe local govt area of lagos state, Nigeria (cannon power A14000HD digital camera used.)	94%	Daramola S. Adebayo, et. al, Nigeria, "Fish Classification Algorithm using Singal Value Decomposition", IJRSET.
3	369 fishes of various size of 30 species	Used the fish image segmentation algorithm designed for head, abdomen and tail	Yuan Ge, the author of 'QUT Fish data 'collected 3960 image of 468 species	87.5%	Amanullah Baloch, et.al, "Fish Image Segmentation Algorithm (FISA) for Improving the Performance of Image Retrieval System", IJACSA,2017.
4	Total 350 fish image divides into 2 datasets 257 training images and 93 testing images	ANN,NN,	Global information system on Fishes (Fish-base) and department of fisheries Malaysia ministry of agricultural and Agro-based industry in putrajaya.	75% min,97%max	Mutaseem Khalil Alsmadi, et.al, "Fish Recognition on Rebutst Feature Extraction From Size and Shape Measurement Using Neural Network", JCS, 2010.
5	320 images used. For each fish they used 14 images acquired from the live streaming and 18 images obtained by affine transformation.	1)Fish Affine Transformation 2)Texture Features 3)Boundary Features	Acquired from an ecological source in Taiwan and made accessible to marine biologists via the Ecogrid project coordinate by the national center for high performing computing(NCHC) Taiwan	Fish detection and tracking accuracy 85%	Concetto Spampinato, et.al, " Automatic Fish Classification for Underwater Species Behavior Understanding", ARTEMIS, 2010.
6	500 for training and 110 for testing	ANN, NN, Back Propagation	Global information system on Fishes (Fish-base) and department of fisheries Malaysia ministry of agricultural and Agro-based industry in putrajaya, Malaysia acquired on 22th august 2008.	84%	Mutaseem Khalil Alsmadi, et.al, "Fish Recognition Based On Robust Features Extraction From Color Texture Measurements using Back-Propagation Classifier", JATIT, 2010.
7	178 images	CNN	freshwater fish malaysia	60-80%	Siti.Nurulain Mohd Rum, et.al, "FishDeTech:A Fish Identification Application Using Image Recognition Approach",IJACSA,2021.

An enormous dataset was designed for dealing with the issue of fine-grained acknowledgment recognition of fish species obtained from pictures and recordings. A powerful nonparametric methodology was developed for programmed mark spread. The consequently named dataset was used for the fish assignment of LifeCLEF2014. Ten fish species with 31000 examples were used for the study. The proposed approach in this study can acknowledge fish pictures with confounded foundations, including against rough foundations. Further, to distinguish effective highlights for fish acknowledgment, we have characterized different highlights, including mathematical highlights, the sack of visual word (BoVW) models, and surface highlights. Besides, in related exploration, the quantity of species is reliably not exactly or equivalent to 30. Assessment tests utilizing such little scope datasets yield low dependability and are in this way not reasonable. With respect to the concept in this study, we have developed a huge scope dataset as contrasted and the connected examination and assessed our proposed strategy [15].

The work reported above has provided a stimulus that leads this work as a novel fish order system based on a composition of vigorous component determination, picture division, and mathematical boundary procedures utilizing Artificial Neural Network (ANN) and Decision Tree (DT). In contrast to existing works for fish order, where descriptors were proposed to avoid dissection of samples for their individual effects in the entire arrangement task. The approach did not blend the component choice, picture division, and mathematical boundary together. The proposed method simplifies an overall arrangement of highlights extraction, utilizing strong element determination, picture division, and mathematical boundary and their loads that should be utilized as from the earlier data by the classifier.

III. DATASET

It becomes extremely difficult to get marine information because of the inaccessibility of data. The dataset used in the study is taken from Fishbase (finfishes-based global biodiversity information system and provides the authentic details of fish if we know the name or family of fish) and Fish4knowlage Ground-Truth datasets (project explores big data) as a different type of fish families were available. The dataset consists of 500 fish images from different families and species. The fishes available here are of different types such as poison fish, non-poison fish, and aquarium fish. The size of the image used in this study is adaptable and hence any sized image can be matched. Here we are not specifying the size of fish because to find the name and information of respective fish from the system we are making use of the images present in the image gallery

IV. PROPOSED SYSTEM AND METHODOLOGY

The study is carried out to design a robust and scalable automatic fish classification system. The objective of this research work is to understand the classification of the Fish and non-Fish image samples from the set of the complex fish datasets. The AI methods particularly pattern recognition techniques were utilized for automatic classification of features and identification of fish and non-fish images. The abstract view of the proposed system as presented in figure (1) consists of two modules Training and Testing. The dataset used for the study is a set of around 500 samples. Figure (2) presents the form of a sample from the fish4knowledge dataset. The dataset is divided into a 70:30 ratios, where 70% samples from the dataset were used for the training purpose with known classes and 30% samples were reserved to test unknown samples.

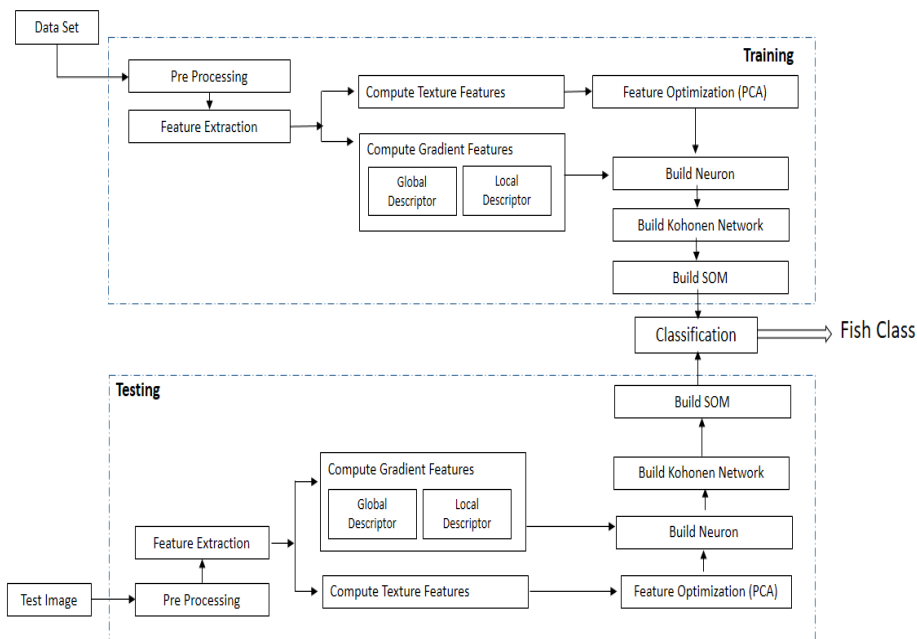


Figure 1 Architecture of proposed framework

The complex dataset of the fish sample was provided for the training module where every sample from the dataset was preprocessed and made suitable for the computation of features. The preprocessing phase involves methods that are used to enhance the dataset samples by improving comparison, blurring, and reduces the noise contained in the sample. The major activities like foreground, background separation, RGO channel separation, histogram equalization, Contrast Limited Adaptive Histogram Equalization (CLAHE), were performed with respect to each sample from the dataset shown in figure (3).



Figure 2 Sample from the dataset (fish4knowledge)

The phase is also evaluating parameters that are used to determine the quality of the sample like Peak Signal to Noise Ratio (PSNR). The samples having a high peak signal to noise ratio the sample is used for the segmentation and feature extraction phase. The Gaussian filters and Laplacian filters are used to enhance the sample whenever it was required. The enhanced sample is processed for isolation of Region on Interest (ROI) using the inherent segmentation process.

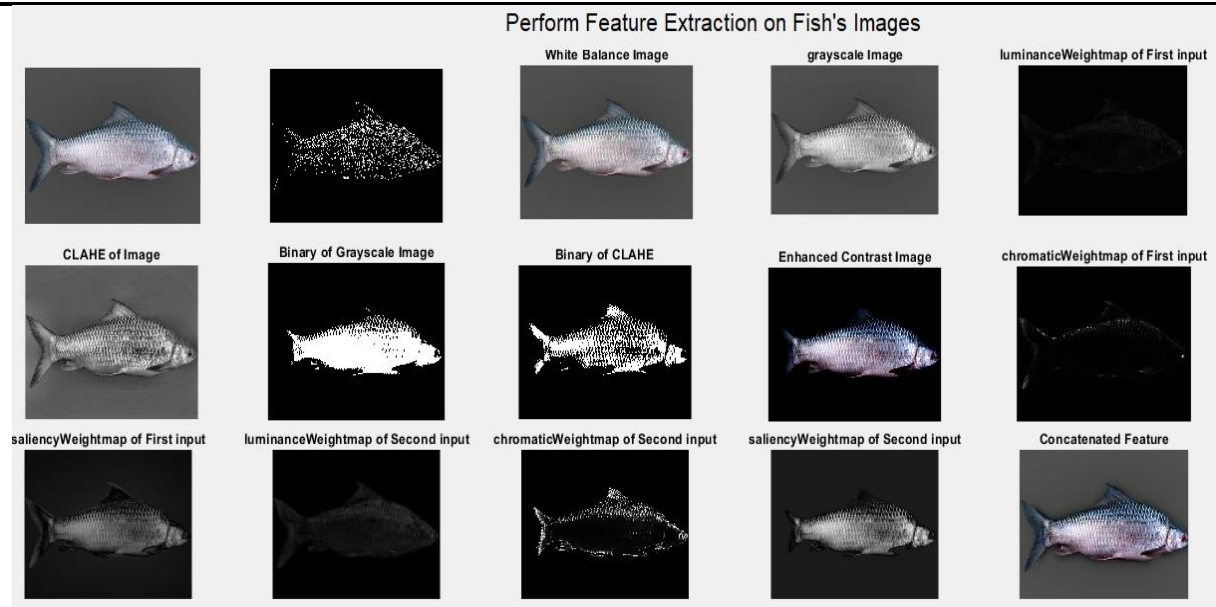


Figure 3 Perform Feature Extraction on Fish's Images

The process of isolation of ROI ends with highlighting and locating boundaries of the fish sample under observations. This ROI estimation help in the computation of textural and gradient features. The processed sample was passed to the feature extraction phase and textural and gradient (local and global) features of the sample were computed. The Local Binary Pattern (LBP) histograms were extracted from the Gabor map of fish samples for the identification of their morphological characteristics such as Finns and Gills. It was worth noting that, the basic version of the LBP operator uses the center pixel value as a threshold to the 3×3 neighbor pixel and when it makes use of 8 pixels then the number of histogram bins will either be 28 or equal to 256. This mechanism helps in creating binary patterns representing texture characteristics which are based on threshold operations.

An image is encoded by a descriptor in such a way that it allows the image to compare and match the other images [5-8]. The Global Descriptors Feature describes a whole image to generalize the entire object. Global features are used in low-level applications such as object detection and classification. Global features include contour representations, shape descriptors, and texture features, whereas Local Descriptors describe the image patches of an object (key points in the image). Multiple local descriptors are used to match an image and this is more robust as not all the descriptors need to match for the comparison that is to be made. Local Features used in the higher-level applications such as object recognition. Scale-invariant feature transform (SIFT), Speeded up robust feature (SURF), etc. are some examples of the local descriptor.

The prominent textural feature features were computed by using Principal Component Analysis (PCA) and the dimension of the feature set was reduced. The reduced prominent textural PCA features were combined with gradient features and neurons were built. The Kohonen network and Self Organizing Map (SOM) model were developed for the purpose of classification.

SOM was trained over all samples of the training dataset and saved on the feature space for further classification, similarly, the Kohonen network was also configured. Feature Space was used in the classification phase will help to predict the class of test sample from the training set of features so that the test sample will be classified. By the training module, the test sample was given as input to the test phase, where the given sample is preprocessed and its features textural, gradient features were computed. The computed feature of the test sample was optimized using Principal Component Analysis (PCA). The optimized features were used to build the Kohonen network, Self-Organizing Map (SOM), and submitted to the classification phase for prediction. Support Vector Machine (SVM) was used for the classification of test samples. The performance of the system was evaluated by using the ROC curve, precision, and recall. The classification algorithm used in this study is Expectation-Maximization (EM) algorithm and it is used as it is an iterative method for finding maximum likelihood. In the scenario when the data is missing or incomplete then the maximum-likelihood estimates model parameter. The performed Performance measurement was made by using the five-fold cross-validation schema and was adopted in the presented research work for performance measurement. The five-fold cross-validation process was performed by randomly dividing the dataset into five disjoint sets. From the divided sets four sets were used for training and the remaining one was used for the testing purpose. This process was repeated over a period of time in a process for testing each set. The terms recall precision, the specificity of each class, and average accuracy were measured for the performance of classifying tasks. According to the confusion matrix, these terms were calculated.

Classification prediction is referred by positive and negative terms. The true and false values term helped to analyze whether the prediction corresponds to the samples of the real class [18].

V. RESULT AND DISCUSSION

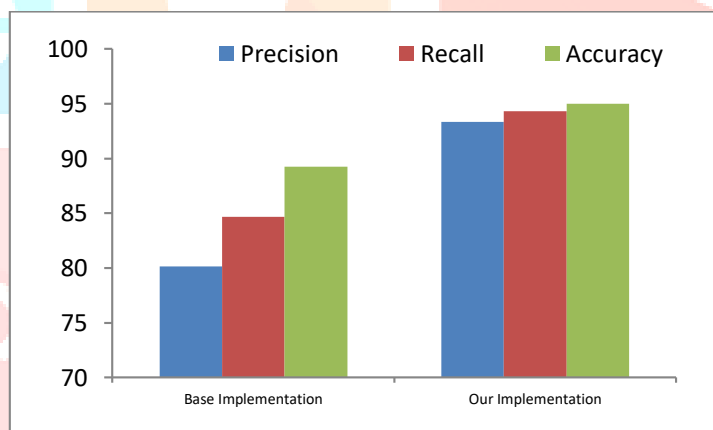
In this study, AI methods are discussed in section IV have been evaluated, the observations were analyzed and tabulated. The work examined 500 samples from the complex dataset of fish image samples out of which 350 samples were used for training purposes and 150 samples were reserved for testing. Based on the experiments, it was seen that the overall accuracy of this study was found to be 97%. Here we have focused on all types of fishes from the dataset and their evaluation was presented in Table 2 and Table 3. There are all types of fish available as poison and non-poison.

Table 2 Description of the Overall Accuracy of Training and Testing

Description	Result
Overall Training Accuracy	100%
Overall Testing Accuracy	96%

Table 3 Result of Fish Image

Sr. No	Fish Family	Fish Name	Type of fish	Accuracy	Precision	Recall
1	Cyprinidae	Cirrhinus Mrigala	Non Poison	95.91	94.73	94.73
2	Cyprinidae	Cyprinus Carps	Non Poison	90.00	94.73	94.73
3	Cyprinidae	Gold Fish	Non Poison	92.85	94.73	94.73
4	Bramidae	Pomfret	Non Poison	97.29	94.73	94.73
5	Tetraodontal	Puffer Fish	Poison fish	97.56	96.00	96.00
6	Eel family	Moray Eel	Poison fish	97.64	96.00	96.00
7	Alestidae	Tiger fish	Poison fish	97.67	96.00	96.00
8	Scorpaenidae	Red Lion fish	Poison fish	97.59	96.00	96.00
9	Serranidae	Groupers	Non Poison	97.53	96.00	96.00



Graph 1: Graphical Comparison with existing system

In the proposed work we have used different types of fish and their family to identify their name and property. We have also used edge detection to detect the boundary of the fish image to identify its fins, which can be one dorsal or two dorsal. Fish have many fins to swim in. In the fish images, we have found their gills, head, eye, and mouth which is the difference in all fishes. In this system, we have matched the fish fins and gills to identify them. For example, we took a Puffer fish and it was matched with our dataset and was found that it got matched and is a poison fish that belong to the Tetraodontidae family and the accuracy is calculated is 97%. The Moray Eel fish which belongs to the Eel family an accuracy obtained for it is 97.64%.

VI. CONCLUSION

This research, aimed to get the solution for accurate recognition and to get accurate information about the fishes using fish images. The study contributes to the determination and classification of fish families and species. The number of samples (fish family's species) in the proposed dataset was 500 including the different shapes of fishes. The image in the dataset is of multiple fish that were not segmented correctly because the proposed program works on a single whole fish image. The program that we have design shows correct segmentation on fish that were taken in the controlled environment and on the fish whose images were taken with a plain background in a natural situation. In this study, we have obtained results from the fish image and used some image processing techniques like edge detection and methods like Local Binary Pattern and PCA technique. The overall accuracy obtained in this study is 97%.

VII. ACKNOWLEDGMENT

Author would like to thank, Department of Computer science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) for all infrastructure support.

REFERENCES

- [1] Deepa S. Deshpande, Archana M. Rajurkar and Ramchandra R. Manthalkar L.C. Jain et al. (eds.), "Computational Intelligence in Data Mining - Volume 1", Smart Innovation, Systems and Technologies 31, DOI 10.1007/978-81-322-2205-7_37
- [2] Takeshi Saitoh, Toshiki Shibata, and Tsubasa Miyazono, " Feature Points based Fish Image Recognition," International Journal of Computer Information Systems and Industrial Management Applications, vol. 8, pp. 012-022, 2016.
- [3] Pradip Salve, Milind Sardesai, Ramesh Manza, Pravin Yannawar, "Identification of the plants based on leaf shape descriptors", Proceedings of the Second International Conference on Computer and Communication Technologies, 85-101, 2016, Springer, New Delhi
- [4] Pradip Salve, Pravin Yannawar, Milind Sardesai, Multimodal plant recognition through hybrid feature fusion technique using imaging and non-imaging hyper-spectral data, Journal of King Saud University - Computer and Information Sciences, 2018, ISSN 1319-1578
- [5] Pradip Salve, Milind Sardesai, Pravin Yannawar, Classification of plants using GIST and LBP score level fusion, International Symposium on Signal Processing and Intelligent Recognition Systems, pp. 15-29, 2018, Springer, Singapore
- [6] Miss Priyanka Pawar, P. P. Belagali, "Image Retrieval Technique Using Local Binary Pattern", International Journal of Science and Research, vol. 4, no. 7, 1440-1443,2015.
- [7] Gajendra Singh Chandel, Ravindra Kumar, Deepika Khare, Sumit Verma, "Analysis of Image Segmentation Algorithm Using MATLAB", International Journal OF Engineering and Research, vol. 1, no. 1, pp. 51-55, 2012.
- [8] Md. Abdur Rahim, Md. Najmul Hossain, Tanzillah Wahid and Md. Shafiul Azam, "Face Reorganization using Local Binary Pattern (LBP)", Global Journal of Computer Science and Technology, vol. 13, no. 4, pp. 1-8, 2013.
- [9] N. Dhungel, G. Carneiro, and A. P. Bradley, "Automated Mass Detection in Mammograms using Cascaded Deep Learning and Random Forests," in Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on, 2015, pp. 1-8: IEEE.
- [10] L. Liu, J. Li, and Y. Wang, "Breast mass detection with kernelized supervised hashing," in 2015 8th International Conference on Biomedical Engineering and Informatics (BMEI), 2015, pp. 79-84: IEEE.
- [11] P. Kang and S. Cho, "EUS SVMs: Ensemble of under-sampled SVMs for data imbalance problems," in International Conference on Neural Information Processing, 2006, pp. 837-846: Springer.
- [12] H. He and Y. Ma, Imbalanced learning: foundations, algorithms, and applications. John Wiley & Sons, 2013.
- [13] Bria, N. Karssemeijer, and F. Tortorella, "Learning from unbalanced data: a cascade-based approach for detecting clustered micro calcifications," Medical image analysis, vol. 18, no. 2, pp. 241-252, 2014.
- [14] E. Kozegar, M. Soryani, B. Minaei, and I. Domingues, "Assessment of a novel mass detection algorithm in mammograms," Journal of cancer research and therapeutics, vol. 9, no. 4, p. 592, 2013.
- [15] Y. Wang, "Hierarchical Masses Detection Algorithms Based on SVM in Mammograms," Master's Thesis, Xidian University, China, 2006.
- [16] M. Alsmadi, K. B. Omar, S. A. Noah and I. Almarashdeh, "A Hybrid Mementic Algorithm with Back-propagation Classifier for Fish Classification Based on Robust Feature Extraction from PLGF and Shape Measurements" Information Technology Journal, vol.10, no.5, pp.944-954,2011.
- [17] Vidya A. Huse, Deepali Chaudhari and Prof. Dr. Bharti W. Gawali, "Image Processing Approach for Fish Image Analysis-A Review" IJICS, vo. 6, no. 7, July 2019,