A Framework for Fish Species Identification, Freshness Assessment, and Formalin Detection

B. Pavan Kalyan
Electronics & Telecommunication Department, Don Bosco College of Engineering, Fatorda-403602, Goa, India

Siddhant Kunde
Electronics & Telecommunication Department, Don Bosco College of Engineering, Fatorda-403602, Goa, India

Shreyas Simu
Electronics & Telecommunication Department, Don Bosco College of Engineering, Fatorda-403602, Goa, India

Varsha Turkar
Electronics & Telecommunication Department, Vidyalankar Institute of Technology, Mumbai, India.

Vaishab Jalmi
Electronics & Telecommunication Department, Don Bosco College of Engineering, Fatorda-403602, Goa, India

ABSTRACT

Fish is most popularly consumed in coastal areas of India and some hinterlands as well, with demand increasing every year. To cater to this demand, fish must be transported to many metro cities and tourism-dominated areas. As a result, fish identification and fish freshness have become critical issues. Fish traders often store fish in cold storage with unapproved chemicals, such as formaldehyde, to extend storage life and prevent aging. However, this practice causes several health issues, including cancer. Additionally, the similarity in appearance of different fish species makes it difficult for consumers to identify the type of species. To confront this problem, this paper outlines the development of a fish identification and quality assessment system using Image Processing and Machine Learning (ML). The system can identify different fish species and detect fish freshness based on the number of days since the catch. Various feature extraction and machine learning classifier methods are applied to fish images for species identification and freshness assessment. Moreover, the paper introduces an approach to detect the presence of artificial formalin added to fish. The system, designed to be implemented on a mobile phone, combines software techniques with a hardware interface, including an HCHO sensor, to detect artificial formalin. This complete system is user-friendly and aims to enhance consumer safety by addressing both fish identification and the detection of harmful chemicals.

General Terms

Image processing, Machine learning.

Keywords

Fish Identification, Fish Freshness, Formalin Detection, Non-invasive.

1. INTRODUCTION

India's fishing industry is the second-largest in aquatic sciences and the third-largest in fishing production. Fish is an essential source of minerals such as iron, zinc, calcium, iodine, phosphorus, magnesium and potassium, and are an essential component of global food. Fisheries contribute 1.07 percent to India's GDP, while freshwater sources make up 55 percent of the country's total fish production. The richness of India's water and natural resources provides significant opportunities for aquaculture growth [1]. The Indian fish market, divided by type of fish into inland fish, marine fish, shrimp and scampi, reached almost 1,232 billion INR in 2020. States like Goa, West Bengal, Kerala, and Maharashtra are the largest fish-producing states. In Goa alone, approximately 6.57 million metric tons of various fish are produced annually to meet the high demand among the population, which heavily relies on fish intake. The state’s favorable climatic conditions for fish farming enable it to primarily supply sea fish. However, fish are highly contaminated and the globalization of food trade makes quality control more difficult [2]. Fish products are susceptible to rejection because of their quality, especially when the raw materials in their original form are lower than the standard despite technological advances in fish production [2]. Controlling fish product quality is difficult due to dependency on fishing conditions, seasons, and quotas. Fresh fish with the necessary shelf life for supermarket distribution is hard to find, leading to discounted sales near the end of their shelf life, which reduces retailer profits. To avoid such losses, retailers often seek methods to extend product shelf life, enabling them to sell at full price and maximize profits [3]. Fishmongers use ice storage to preserve fish for longer periods, a practice also applied during interstate export. However, this can result in a loss of nutritional value and taste. Popular and expensive fish like Pomfret are often adulterated, and increased demand can drive up prices significantly. Traditionally, fish freshness is detected through visual inspection of gill color, texture dehydration, and eye clarity. However, fish kept in cold storage pose a challenge for freshness prediction at the time of purchase. Additionally, the wide variety of species can make it difficult for non-native consumers to identify fish by their local or base names, as many species look similar in color and shape. In Goa alone, there are 35 kinds of seawater fish and around 18 different kinds of freshwater fish.

Recent research includes information on invasive and non-invasive techniques used to determine fish freshness. One method is by visually inspecting fish (eyes, glands, smells, size, skin tissues) and cutting them into pieces to analyze them. The determination of fish quality based on visual images and classification was carried out using fuzzy logic to detect if the fish is in good or bad condition [4]. Support Vector Machine (SVM) is used to eliminate limitations of some existing techniques and improve fish classification. The authors used features based on shape and texture. The data set was created for selected species. A total of 150 images were used (76 for training and 74 for testing), achieving a classification accuracy of 78.59 percent, which is significantly higher than that of ANN, KNN, and K-means clustering algorithms. [5]. The classification of tuna fish using a decision tree was carried out by extracting characteristics from frozen tuna images [6]. Another classification model was developed using back propagation classifiers (BP) to extract robust features from the color signature and recognize the isolated patterns of interest (fish). Color signatures were derived from the RGB color space, color histograms, and gray-level co-occurrence matrices. The system was tested across 20 different fish families, each containing various species, with a sample comprising 610 distinct fish images. The overall accuracy of the test data was 84% [7]. For fish classification, hybrid genetic algorithms (GAs) and reverse propagation have been proposed. The feature set is based on shape, color, character, and texture. Fish images are classified into different categories, such as predatory, toxic, garden and food fish families. 24 fish families were used, each with different species. BP and hybrid GA classifiers have improved classification accuracy from 82.1% to 87.7% compared to the latest methods. Color and texture characteristics are used to classify fish species using Random Forest (RF), Support Vector Machine (SVM), Logistic Regression (LR) and the K-Nearest Neighbors (k-NN). The SVM using the radial-based kernel achieves superior classification performance, with an accuracy rate of 82% (CCR) and a Kappa coefficient of 0.65. [8]. An invasive method was proposed to determine the freshness using electrodes applied to the muscles and skin of fish [9]. The conductivity and ability of the fish muscles are used to predict the destruction of the fish. The technique of extracting features was studied for future research, including GLCM, shape variant features [10]. A comparison study of fish freshness meters and RGB color index quantification was conducted to determine fish freshness [11]. The sensors used in this study are Torrymeter, which measures three species and the quantification of RGB color focuses on fish eyes and tail.

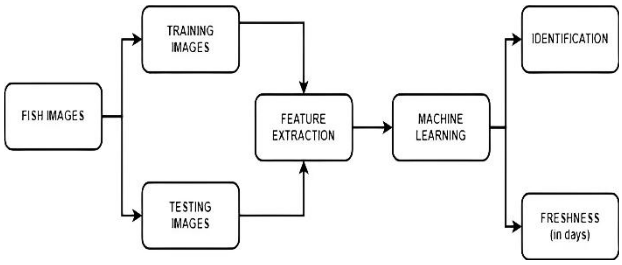
After an extensive literature survey following points were concluded and this research aims to mitigate the gaps:

1. As India is a diverse country with different cultures and languages spoken, many times a non-local person is unable to identify the type of fish. Also, they are unable to gauge the freshness of fish in terms of days from the catch. This research paper involves design of system that identifies the type/class/family of fish and also, it's freshness.
2. Most of the methods used for detection of formalin on fish are invasive system, where fish is cut into pieces for testing and electrodes are inserted for concentration detection. This research involves designing a non-invasive system.
3. Certain testing methods make use of chemical reagent for formalin detection which creates a problem for common people to handle chemicals. This research avoids this drawback as no chemicals are involved.
4. Finally, there is a requirement for a complete system which will perform all above analysis and give a one stop solution to the fish consumers and retailers.

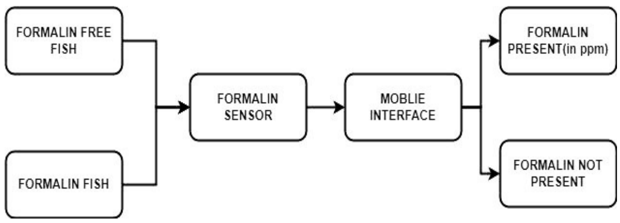
This paper develops a system for identifying fish and determining its freshness. Comparison analysis using various features extraction and machine learning techniques is done to determine fish freshness and to classify fish into their respective families. Furthermore, the paper discusses use of formalin sensors and mobile phone interfaces to detect formalin sprayed on fish. Section 2 outlines the materials used and methods employed in this research. Section 3 presents the results, and Section 4 concludes the research work.

2. MATERIALS AND METHODS

A real-time system is developed for fish identification and freshness detection from Day 0 to Day 5. Also to detect the presence of artificially added formalin on fish using a formalin sensor and display the concentration of formalin present on an android application via an USB interface. The proposed systems for fish identification and estimation of fish freshness are described below, in Figure 1.



a)

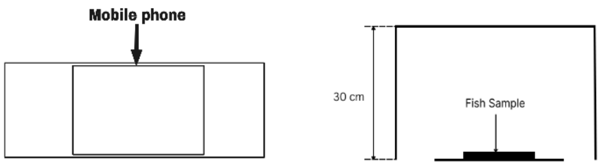


b)

Figure 1: Block diagrams a) Fish Identification and freshness using Machine learning and Image Processing,b) Fish Freshness using Formalin Sensor.

2.1 Fish Image Dataset

The first stage includes capturing the fish images which act as the input to the system. For this study, the fish dataset was created for fish identification using 4 different fish families. Also, for fish freshness, a dataset was created using one fish family. For accurate detection, the dataset was generated at room temperature with a 48 megapixels mobile camera which was fixed perpendicular to fish on a stand at 30 cm as shown in Figure 2. Also, the illumination condition and background were kept the same for all the images. Fish family used were local fish namely Green Chromide (Calandor), Crescent Grunter (Kharkhare), Gray Mullet (Shevto), Jibra (Local name) from brackish water for identification of fish whereas Indian Mackerel (Bangda) from seawater for experimental analysis of fish freshness from day 0 to day 5. A detailed description of the dataset is given below Table1.

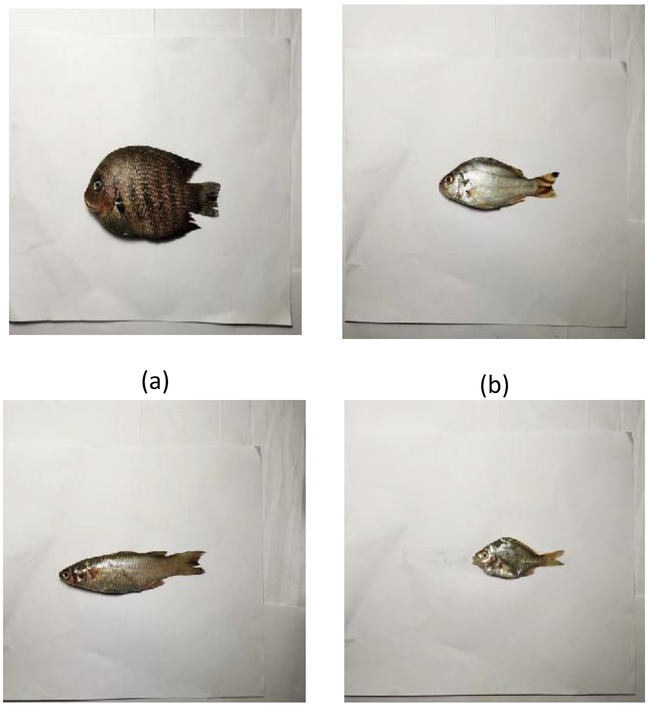


(a) (b)

Figure 2: Setup for dataset generation. a. Top View, b. Front View.

Table 1. Dataset description for fish identification

Fish Family	Total Number of Images
Green Chromide	70
Crescent Grunter	70
Gray Mullet	70
Jibra	70



(c) (d)

Figure 3: Images from the dataset for fish identification. a. Green Chromide, b. Crescent Grunter, c. Gray Mullet, d. Jibra (Local name)

A total 280 images containing four different species as shown in Figure3 and indicated in Table 1 were generated and used for further processing. The dataset shown in Table 1 is used for fish identification process.Figure 4 shows few images of Indian Mackerel for 6 consecutive days. A total of 600 images were generated of Indian Mackerel as indicated in Table 2 having variations such as angular rotation and shearing which were obtained using data augmentation. This dataset represented in Table 2 is used for fish freshness detection process. Data augmentation is a data analysis technique used to increase data by recreating/modifying the original data to create samples having slight variation. Data augmentation helps overcome the problem of overfitting while training a machine learning model.

Table 2. Dataset description for fish identification

Fish Family	Total images per day	Number of days
Indian Mackrel	100	6

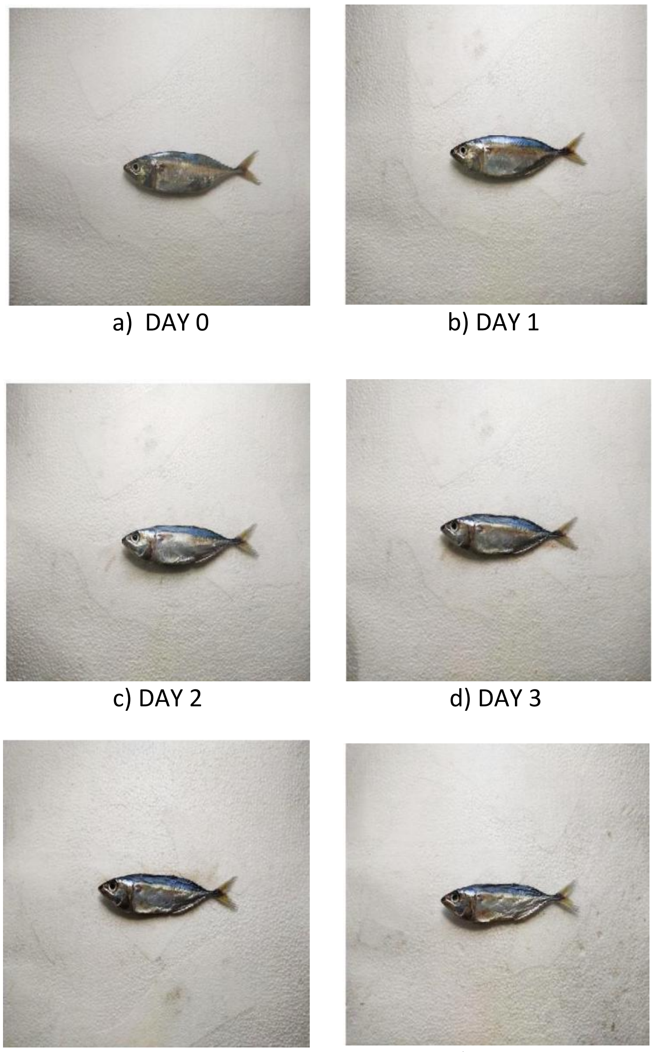


Figure 4: Images from the dataset for fish freshness detection (Indian Mackerel)

2.2 Feature Extraction Methods

In the feature extraction stage images from the fish dataset were converted into grayscale images. Later these images were subjected to various image processing techniques to find and extract the unique features for accurate detection and classification. Techniques such as Gray Level Co-occurrences Matrix (GLCM), Histogram of oriented Gradient (HOG), Speed Up Robust Feature (SURF), Discrete Wavelet Transform (DWT) were performed. Further on these features were saved in an excel file.

2.2.1. Gray Level Co-occurrence Matrix(GLCM)

The Gray Level Co-occurrence Matrix (GLCM) is a statistical method utilized to analyze texture by studying the spatial relationships among pixels. GLCM evaluates image texture by tallying occurrences of pixel pairs with defined values and spatial configurations within the image. From this matrix, statistical measures are derived to quantify texture characteristics [12].

2.2.2. Discrete Wavelet Transform(DWT)

The Discrete Wavelet Transform (DWT) functions as a signal processing tool that decomposes signals into wavelets, facilitating simultaneous time and frequency analysis of signals. Wavelets' energy is concentrated in time while maintaining wave-like properties, making them ideal for this purpose. DWT enables the decomposition of an image into various levels of resolution, resulting in a new feature set based on wavelet coefficient analysis. The model feature vector was trained using first-order statistical moments, including the mean and standard deviation of different image levels, as detailed in equations 1 and 2.

$$m_f = \frac{1}{N_{m.n}} \sum f(m,n) \tag{1}$$

$$S_f = \sqrt{\frac{1}{N-1_{m,n}} \sum f(m,n)} \tag{2}$$

2.2.3. Histogram of Oriented Gradient(HOG)

This feature descriptor is derived from image data. It is generated based on the gradient and orientation of pixel values within the image. HOG emphasizes the structure and shape of objects by capturing edge directions, achieved through the extraction of gradient magnitudes and their corresponding directions. The HOG character descriptor counts the gradient direction of images located in the location [13]. This represents the distribution of the intensity fluctuation in different directions.

2.2.4. Speed-up Robust Feature (SURF)

The SURF algorithm approximates the Laplacian of Gaussian (LoG) with a box filter. This is a second-order filter that is sensitive to noise. To address this issue, the image is often pre-processed with Gaussian smoothing before applying the Laplacian filter. This helps to reduce the high-frequency component before taking the differentiation step. This process can be parallelized for different scales. The equation for LoG is given below where I stands for the image.

$$L(x,y) = \frac{\partial^2}{\partial x} I(x,y) + \frac{\partial^2}{\partial y} I(x,y) \tag{3}$$

2.3 Machine Learning Algorithms

There are three primary types of learning: unsupervised learning, supervised learning, and reinforcement learning. In unsupervised learning, networks are trained solely based on input data without corresponding output labels. In contrast, supervised learning involves training a network with a dataset consisting of paired inputs and outputs, allowing the network to learn from these examples. Reinforcement learning is employed when input-output patterns cannot be explicitly defined. In this approach, the system enhances its performance by interpreting feedback—positive or negative signals related to its actions—and adjusting its parameters accordingly. The datasets used for training the network are referred to as learning or training sets. The algorithms discussed in this paper are supervised learning classifiers, as the training data consists of an input-output model, with image features serving as inputs and family/days as labels.

2.3.1 K-Nearest Neighbor classifier

This classifier is a straightforward, yet powerful classification model that leverages the nearest neighbor algorithm to categorize a specific data point [14]. This model identifies the K closest data points to determine the category of the input data, assigning a class based on the majority of its corresponding feature vectors. By evaluating the feature similarity, the classifier selects the data point's class from the K data points with the most similar features. The selection of nearest neighbors is based on the Euclidean distance between the objects, with a larger distance resulting in a lower Euclidean score, indicating dissimilarity. Therefore, the Euclidean distance and score are inversely related. This process defines the boundaries of each class for a given K value, facilitating the segregation of data within each class.

When the training and validation are segregated from the original dataset, an optimal value of K can be determined. The similarity between two points is evaluated based on their distance in the given space using a suitable metric. To predict the class for a new observation, the K-Nearest Neighbors algorithm identifies the k closest data points to the observation and then selects the most common class among these points. This key characteristic is what gives the algorithm its name, K-Nearest Neighbors [15]. Results generated through KNN are straightforward to interpret, making it applicable to a wide range of datasets. Many datasets exhibit irregular data point distributions, suggesting that employing different values of k for various data points could be beneficial [16].

2.3.2. Random Forest

Random Forest represents a specific case within a collection of items (multiple classifiers) utilized as an ensemble rather than individual learning. In this approach, Decision Trees are employed to build individual models, comprising the ensemble of decision trees used for output value prediction. [17]. The Random Forest algorithm functions in two stages. In the first stage, the random forest is constructed by combining n decision trees. The second stage involves making predictions for each tree using new data points and categorizing them based on the majority vote.

2.3.3. Naïve Bayes

The Naïve Bayes technique employs Bayes' theorem to develop classifiers [18]. This classifier is grounded in Bayes' theorem, which calculates the probability of an event occurring by considering various related conditions. The formula of Bayes theorem is given by the following equation.

$$P(A|B) = \frac{P\left(\frac{B}{A}\right)p(A)}{P(B)} \tag{4}$$

where P(A|B) is the posterior probability, P(B|A) is the likelihood probability, P(A) is the prior probability, and P(B) is the marginal probability. Naïve Bayes classifiers are constructed by assigning class labels to problem instances that are represented as feature value vectors. The assumption behind this is that the value of any one feature does not depend on the value of any other feature, hence the name Naïve. The classifier believes that each feature contributes independently to the outcome. Three types of Naïve Bayes models exist, and this paper utilizes the Gaussian Naïve Bayes classifier. This type of classifier assumes that the values associated with each class follow a Gaussian distribution, which can be expressed as follows:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} e^{\left(\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right)} \quad (5)$$

where σ is the standard deviation and μ_y is the mean value.

2.3.4. Artificial Neural Network

Artificial Neural Networks (ANNs) are computational systems designed to mimic the information processing methods of the human brain. These networks are composed of interconnected neuron nodes arranged in a web-like structure, with processing units that include both input and output units. ANNs have many interconnected units, also referred to as nodes or neurons, which operate in parallel and communicate with each other through connection links. Each link has weights that determine the input signal's influence on the neurons. The activation signal reflects the internal state of each neuron, whereas the output signals, generated by applying an activation rule to the input signals, can be sent to other units. The training process for a neural network involves comparing the network's processed output (often a prediction) to the target output and adjusting the weights accordingly. ANNs can be arranged in layers (multi-layer networks) or have a specific connection topology.

2.4 Fish dataset for formalin detection

Fish samples of Indian Mackerel were used for formalin detection experiment. The fish samples were sprinkled with formalin solutions of different concentrations and stored in containers for a period of four days. The fish samples as shown in Figure 5, were collected from seawater. The testing was conducted using three different concentrations of formalin solution: 5 ppm, 10 ppm, and 20 ppm.

Figure 5: Fish sample used for formalin detection process

2.5 Formalin Sensor

The Formalin sensor is built around a semiconductor-based volatile organic compound (VOC) gas sensor known as WSP2110. This component's conductivity alters in response to changes in the concentration of VOC gases in the air, which can then be converted into an output signal that corresponds to the gas concentration. This sensor is capable of detecting gases with concentrations ranging from 1 ppm to 50 ppm, making it particularly useful for detecting formaldehyde. Using potentiometer, the VOC sensor was set to detect formaldehyde. The equation that converts the voltage to ppm is:

$$ppm = \frac{e^{\left(10.0\left[\frac{R_S}{R_0}\right]-0.0827\right)}}{(-0.4807)} \quad (6)$$

where, R0 – Value of resistance in air.
Rs – Value of resistances of concentration detected.

2.6 Formalin solution preparation

Formalin solution with different concentration was prepared using standard calculation given by the following equationwhere,

$$M_1V_1 = M_2V_2 \quad (7)$$

M1 = concentration of stock formalin solution.
V1 = volume of the stock solution.
M2 = concentration of the solution to be prepared.
V2 = volume of the solution to be prepared.

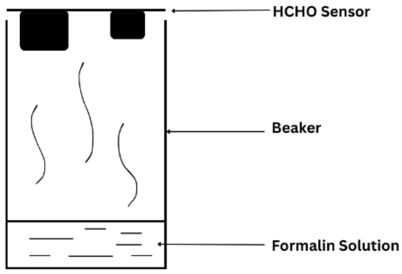


Figure 6: Sensor values read from formalin solution.

Using equation 7, different formalin solutions were prepared which were 5 ppm, 10 ppm and 20 ppm. Prepared solution was directly taken into the beaker as shown in Figure 6 and formalin concentration was measured using the HCHO sensor.

2.7 Mobile Interface Design

In this stage the data received by the sensor is transferred using an USB cable which is used as a communication medium to transmit data onto the mobile application. The value of the formalin concentration is displayed in ppm value onto the mobile application. The flow diagram is shown in Figure 7.

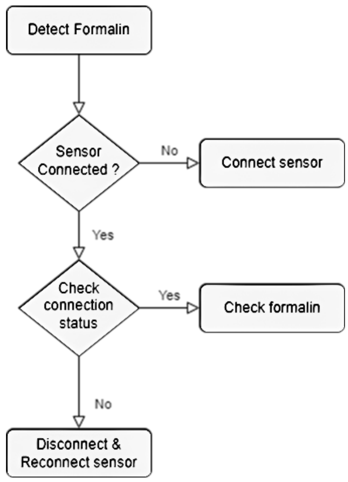


Figure 7: Flow diagram for formalin detection using sensor on mobile.

2.8 Hardware and Software Requirements

All types of features were extracted from the fish images using MATLAB 2013. Machine Learning algorithms were modeled using python on Google Colab IDE. For formalin detection volatile organic compound (VOC) sensor was used which was tuned to detect formalin and the data values from sensor was transmitted to the computer through Arduino microcontroller. The mobile app was designed using MIT App Inventor software.

3. RESULTS AND DISCUSSION

This section discusses the results obtained from features extraction and subsequent application of machine learning algorithms on datasets for fish identification and fish freshness. Further we will discuss the results obtained from formalin detection system and design of the corresponding app for display of the results.

3.1 Results of Feature Extraction

Images from the dataset were subject to various feature extraction methods discussed in previous section to extract unique features. The detailed description of the feature vector is given in the Table 3. Figure. 8 shows the direction of gradient as a feature in the localized portion. Gradients of an image are useful because the magnitude of the gradient is large around edges and corners (which result in giving a lot of data). HOG would generate a histogram for each of these regions separately. Figures 9 and 10 show the application of DWT and SURF on the fish images. As indicated in Table 3, GLCM and DWT features and combined to make a single feature vector of size 12. HOG feature is of length 100 and SURF of 64 respectively. These features act as input to the machine learning section where various ML algorithms were applied on it for the study.

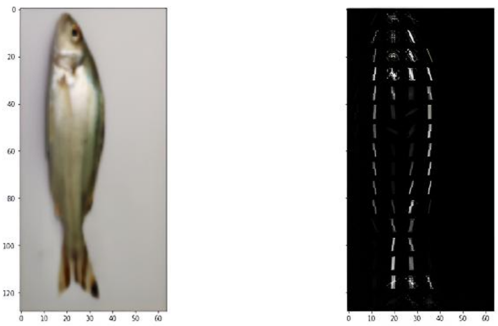


Figure 8: HOG applied on a fish image

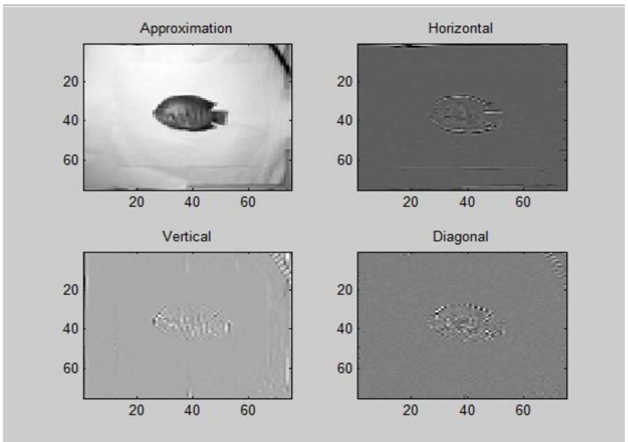


Figure 9: DWT applied on a fish image

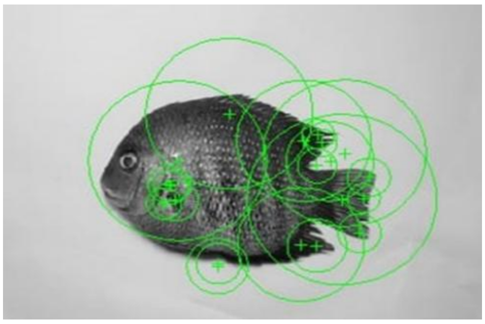


Figure 10: SURF applied on a fish image

Table 3. Table showing feature description

Extraction Technique	Feature Vector Size
GLCM & DWT	1*12
HOG	1*100
SURF	1*64

3.2 Fish Identification and Freshness

Predictive models were generated, for fish identification, the dataset had 4 different fish families. Similarly, for fish freshness dataset included 6 different days of the same fish with different variations including angular rotation, flipping, and shearing. The features extracted from dataset of fish images were textual information (using GLCM), structural information (using HOG), analysis of image in both time and frequency (using DWT) and second order features (using SURF). These features were used separately to train and test the various machine learning models such as KNN classifier, Random Forest classifier, Naïve Bayes classifier and ANN classifier. These numerical features are grouped as training sets with 80% of features and remaining 20% as a testing set and are used to train the model. K-fold cross-validation was implemented with k=5 and the mean accuracy achieved using various classifiers is tabulated in Table 4 and 5.

Table 4 shows that fish family identification with SURF features and Random Forest gives the best results among all the feature extraction and machine learning technique combinations. Followed by SURF feature and Naïve Bayes classifier and HOG features with ANN classifier. But clearly SURF feature has given better information for fish identification. This results as SURF extracts interest points that can be used to construct a 3D representation of the image.

Table 4. Classifier accuracy for fish identification

Feature Extraction	**Classifier accuracy for fish identification**			
	KNN	Random Forest	Naïve Bayes	ANN
GLCM & Discrete Wavelet Transform	72%	86%	73%	77%
Histogram of Oriented Gradient	83%	87%	80%	99.64%
Speeded-up Robust Feature	97%	**99.89%**	99.81%	97.5%

Table 5 shows that fish freshness detection using SURF features and ANN classifier gives the best result among all feature extraction and machine learning combinations. GLCM & DWT features gives better accuracy with Random Forest classifier whereas Histogram of Oriented Gradient features give better accuracy with Artificial Neural Network and least accuracy with Naïve Bayes classifier. But Overall SURF features along with Artificial Neural Network gives the best result for fish freshness as compared to other feature extraction and classifiers.

Table 5. Classifier accuracy for fish freshness

Feature Extraction	Classifier accuracy for fish freshness			
	KNN	Random Forest	Naïve Bayes	ANN
GLCM & Discrete Wavelet Transform	59%	89%	75%	78.83%
Histogram of Oriented Gradient	69%	75%	45%	88.17%
Speeded-up Robust Feature	78%	83%	78%	**91%**

3.3 Formalin Detection

Formalin contaminated fish was used for detection of formalin using HCHO sensor. The entire process has been carried out for a period of 4 days. Formalin concentration used was 10ppm prepared from standard 40-37% weight/volume formalin solution. The setup for detecting the presence of formalin is shown in Figure. 11 and the readings obtained for each day are tabulated in Table 6.

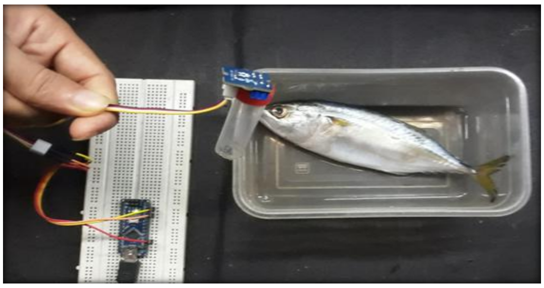


Figure 11: Formalin detection on fish using HCHO sensor.

Table 6. Sensor values of fish laced with formalin

Formalin Concentration (ppm)	Sensor Values (ppm)			
	Day 1	**Day 2**	**Day 3**	**Day 4**
10	9	7	6	3

Table 7. Sensor values detected from formalin solution

Formalin Concentration (ppm)	Sensor Value (ppm)
5	4-6
10	9-11
20	19-21

Table 6 shows the sensor values obtained using HCHO sensor on each day which shows that formalin, a volatile component has reduced slowly but not completely vanished. Also, formalin concentration of different ppm levels was detected using the sensor which is tabulated in Table 7. This table shows that the accuracy of sensor in detecting the concentration levels.

3.4 Mobile Application

A mobile application developed using MIT App Inventor incorporates feature extraction and machine learning algorithms, which are applied on images taken through the mobile app. Figure 12 shows designs of the mobile interface.

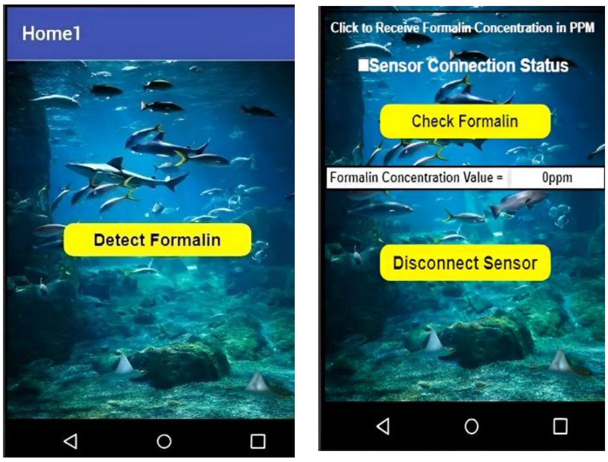


Figure 12: Images of designed User Interface

4. CONCLUSION

Various classifier models were compared to achieve the main objectives of identifying fish species and detecting freshness in terms of days after the catch. Four fish species were used for fish identification and one of them was used for freshness detection. The same fish samples were used over a period of six days, helping gather unique features for accurate freshness detection. Feature extraction from fish images employed GLCM, DWT, HOG, and SURF techniques. Classification was carried out using KNN, Random Forest, Naïve Bayes, and Multilayer Neural Networks. SURF features combined with Random Forest yielded the best results for fish identification, while SURF features paired with Artificial Neural Networks provided the best results for freshness detection. SURF features proved dominant in both cases due to their good stability and better distinctiveness. The developed system can be further improved by incorporating more fish family types with appropriate feature selection.

Formalin solution was prepared at concentrations of 5 ppm, 10 ppm, and 20 ppm and tested with a sensor to determine the concentration with and without fish samples. It was observed that formalin loses its concentration over time. The sensor is fairly accurate with an error of ±1 ppm. Additionally, the formalin sensor can be improved to reduce the error margin.

In this paper the authors have presented a system for comprehensive fish quality assessment using feature extraction and machine learning techniques. The overall system enables accurate identification of fish species, assessment of fish freshness and presence of formalin on fish. By developing a mobile application, the authors have demonstrated the practical applicability in real-world. This work will help users to take informed decisions and contribute towards food safety.

The proposed system can be further enhanced by incorporating a wider variety of fish species to improve its generalizability and robustness. To keep the mobile application light and for faster results machine learning techniques were implemented although advanced deep learning models, such as Convolutional Neural Networks (CNNs), could be explored to further improve feature extraction and classification accuracy.Enhancements in sensor technology could reduce the error margin in formalin detection. Finally, expanding the mobile application with cloud-based data storage and analytics could enable large-scale deployment and facilitate centralized food quality monitoring.

5. ACKNOWLEDGMENTS

This work was supported by Goa State Innovation Council.[Unique Registration Number: 308]

6. REFERENCES

[1] Abotaleb, Shiwani Tiwari, Amr Badr, and Ritisha Balloo. "Estimation of fish production in India using ARIMA, Holt's Linear, BATS and TBATS models." Indian Journal of Ecology 48.5 (2021): 1254-1261.

[2] Alasalvar, Cesarettin, Tony Garthwaite, and Abdullah Öksüz. "Practical evaluation of fish quality." Seafoods—Quality, Technology and Nutraceutical Applications. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002. 17-31.

[3] Sutton, Oliver. "Introduction to k nearest neighbour classification and condensed nearest neighbour data reduction." University lectures, University of Leicester 1 (2012).

[4] Permata, Endi, Ri Munarto, and Reza ZembarYupri. "The identification of fish eyes imagery to determine the quality of fish meat based on fuzzy logic method." journal of electrical engineering and computer sciences, Vol 1 number 1, June 2016 1.1 (2017).

[5] Ogunlana, S. O., et al. "Fish classification using support vector machine." African Journal of Computing & ICT 8.2 (2015): 75-82.

[6] Khotimah, Wijayanti Nurul, et al. "Tuna fish classification using decision tree algorithm and image processing method." 2015 International Conference on Computer, Control, Informatics and its Applications (IC3INA). IEEE, 2015.

[7] Alsmadi, Mutasem K., Khairuddin B. Omar, and Shahrul AM Noah. "Fish classification based on robust features extraction from color signature using back-propagation classifier." Journal of Computer Science 7.1 (2011): 52.

[8] Saberioon, Mohammadmehdi, et al. "Comparative performance analysis of support vector machine, random forest, logistic regression and k-nearest neighbours in rainbow trout (oncorhynchus mykiss) classification using image-based features." Sensors 18.4 (2018): 1027.

[9] Gil, Luis, et al. "Analysis of fish freshness by using metallic potentiometric electrodes." 2007 IEEE International Symposium on Industrial Electronics. IEEE, 2007.

[10] Alsmadi, Mutasem K., and Ibrahim Almarashdeh. "A survey on fish classification techniques." Journal of King Saud University-Computer and Information Sciences (2020)

[11] Jarmin, Roziah, et al. "A comparison on fish freshness determination method." 2012 International Conference on System Engineering and Technology (ICSET). IEEE, 2012.

[12] Mohanaiah, P., P. Sathyanarayana, and L. GuruKumar. "Image texture feature extraction using GLCM approach." International journal of scientific and research publications 3.5 (2013): 1-5.

[13] Singh, Aishwarya. "Feature Engineering for Images: A Valuable Introduction to the HOG Feature Descriptor." Medium, Analytics Vidhya 4 (2019).

[14] Mehanović, Dželila, and Jasmin Kevrić. "Phishing Website Detection Using Machine Learning Classifiers Optimized by Feature Selection." Traitement du Signal 37.4 (2020).

[15] Chen, Ming-Syan, Jiawei Han, and Philip S. Yu. "Data mining: an overview from a database perspective." IEEE Transactions on Knowledge and data Engineering 8.6 (1996): 866-883.

[16] "India's Blue Economy net getting biger". Financial Express. 14 February 2020. Retrieved 18 July 2020.

[17] Yiu, Tony. "Understanding random forest." Towardsdatascience. com. June 12 (2019).

[18] Ray, Sunil. "easy steps to learn naive bayes algorithm (with codes in python and r),”." Analytics Vidhya (6).

[19] Zurada, Jacek. Introduction to artificial neural systems. West Publishing Co., 1992.

[20] Mehrotra, Kishan, Chilukuri K. Mohan, and Sanjay Ranka. Elements of artificial neural networks. MIT press, 1997.

[21] Sivanandam, S. N., and S. N. Deepa. Principles of soft computing (with CD). John Wiley & Sons, 2007.