Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

An in-depth automated approach for fish disease recognition

Md. Jueal Mia [a,*], Rafat Bin Mahmud [a], Md. Safein Sadad [a], Hafiz Al Asad [a], Rafat Hossain [b]

[a] Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh
[b] Faculty of Fisheries, Hajee Mohammad Danesh Science and Technology University, Dinajpur, Rangpur, Bangladesh

A B S T R A C T

Fish plays a significant role in food and nutritional security in our country as well as the whole world. Owing to this reason, it becomes essential to increase the production of fish. But it is diminishing due to numerous diseases which can deteriorate the national economy. It is a fact that there is no single effective research work that has been done in regards to fish disease due to a lack of data and a high level of expertise. Consequently, our aim is to recognize the fish disease effectively that can help the remote farmers who need proper support for fish farming. Recognition of disease-attacked fish at an early stage can help us take necessary steps to prevent from spreading of the disease. In this work, we have performed an in-depth analysis of expert systems that can continue with an image captured with the help of smartphones and identifies the disease. Two set of features is selected then a segmentation algorithm is employed to detect the disease attacked portion from the disease-free portion. Furthermore, eight prominent classification algorithms are implemented accordingly to measure the performance using performance evaluation matrices. The achieved accuracy of Random forest 88.87% which is promising enough.

1. Introduction

Bangladesh is one of the significant inland fishing countries. The fisheries sector plays a vital role in our national economy. According to the Yearbook of Fisheries Statistics of Bangladesh 2018–2019, this sector contributes 3.5 percent to our national GDP (Yearbook of Fisheries Statistics of Bangladesh, 2019). There are many people involved in this sector for their livelihood. Almost 12 percent people of the total population of Bangladesh directly or indirectly engaged with this sector (Yearbook of Fisheries Statistics of Bangladesh, 2019). 1.2 million people are doing jobs in inland fishing sectors whereas the other 0.3 million in marine fishing sectors. Fish provide 55% of animal protein intake in our country (Fishery and aquaculture country profiles, 2021). According to the Yearbook of Fisheries Statistics of Bangladesh 2018–2019, the total production of fish is 43.84 lakh MT in the fiscal year

2018–2019 that proves that Bangladesh is one of the leading countries in fish cultivation. The average fisheries growth is 12 percent in the last 12 years. Bangladesh's fish production has raised six times more than the previous production in the last three decades (7.54 Lakh MT in 1983–84 to 43.84 Lakh MT in 2018–19) (Aftabuddin et al., 2016). Bangladesh obtains 8th position in fish production according to the report of Prothom Alo (Bangladesh ranks in fish production, 2021). Fish firming in fisheries of Bangladesh by the amateur peoples shown in Fig. 1.

Bangladesh acquires a vast amount of foreign currency from exporting fish and fish-related products. According to the Yearbook of fisheries statistics of Bangladesh 2018–2019, the country earns BDT 4,25031.00 lakh (as per EPB) by exporting almost 73.17 thousand MT of fish and fishery products (Yearbook of Fisheries Statistics of Bangladesh, 2019). Infected diseases damage the production of fish as well as diminish its quality and quantity. Approximately 15% of the production of fish are lessened per year for the reason of infected diseases when these fishes are ready in the kitchen, cultivators are acutely realized their investment paralyzing in developing countries. Without appropriate steps, the disease will decrease the production of fish which can deteriorate the national economy. Bacteria, viruses, and fungi are mainly responsible for fish diseases. There are many usual diseases of fish, for instance, Epizootic Ulcerative Syndrome (EUS), Ulcer, Skin Eosin, Gill Damage, Fin and Tail Rot, etc. We have mainly conducted our

Fig. 1. Fish Firming in Bangladesh.

research on Epizootic Ulcerative Syndrome (EUS), Fin and Tail Rot. Besides, we also work with fresh fish with no disease.

Traditional assessment is very difficult and does not provide a standard picture of the result. It has many disadvantages. It kills time and money. So technological identification is a timely step for reducing the losses of money and time. Furthermore, it increases the production of fish and ensured the protein in the human body. The computer vision approach is one of the popular approaches in modern times. It is widely used for the identification of diseases. In this research, we have implemented a depth computer vision approach for automated fish disease recognition. The proposed expert system for our research work has been shown in Fig. 2. Firstly, digital images are captured with the help of digital cameras or smartphones. Then captured images upload to the mobile application to process the results. Images go through some steps for processing on the expert system. Expert systems at first identify the diseased region of images from disease free region. Furthermore, recognizes the disease based on the features of the images. After that expert systems sent back results to the user.

2. Related works

The failures in the production of fish are occurred due to redundant and unplanned steps that can lead to various diseases. Manual identification of fish disease is not an error-free task by amateur farmers. Therefore, a computer vision-based automated approach can be a considerable solution for the reduction of disease. Fish disease recognition is divided into two problems, namely detection of diseased region and classification of disease. Only two works have been conducted to recognizes one fish disease. As there are very few works in this context. So, we reviewed some relevant works on agricultural fruit disease recognitions.

Malik et al. (2017) stated a model for automated fish identification using image processing techniques. They have identified the Epizootic Ulcerative Syndrome(EUS) disease which is caused by a fungal pathogen. This disease is always misidentified by people because it looks like an ulcer. In this work, image segmentation was done by histogram equalization method, after applying this they applied Canny's edge detection technique. They applied two feature descriptor techniques HOG and FAST for extracted features of images. By applying two Neural Network and K-NN (Nearest Neighbor) algorithm classifications had been done in this work. Chakravorty et al. (2015) stated a model for detection of fish disease from the images of fish with help of the PCA method and where *K*-means clustering is used for segmentation based on the color feature. Achieving better accuracy in disease identification, they have introduced HSV images and Morphological operations. Four fish images are used which are affected with Epizootic Ulcerative Syndrome (EUS) in this system. PCA method is used after extracting the feature to form the feature vector and segmented then according to Euclidian distance. Very few datasets are used for their experiments which is not enough to ensure the performance of the proposed system. López-Cortés et al. (2017) stated a model to detect the transmissible pathogens which are responsible for salmonid's mortality. mass spectrometry and machine learning technique are combined to build an automated model which is applied to classify and predict the disease. The achieved accuracy of their model is equal to or greater than 92%. Zhang et al. (2017) illustrated machine vision-based techniques for cucumber disease recognition. They have dealt with the seven types of disease with 420 images. *K*-means clustering is employed to extract features from the images. Combined shape and color features are extracted from lesion information and then classified using sparse representation (SR). The achieved accuracy of their model is 85.7%. Habib et al. (2020) have performed an in-depth investigation on Jackfruit disease recognition. They have used the *K*-means clustering algorithm as segmentation to extract the features from captured images. In order to evaluate the performance, nine classifiers had been adopted accordingly. Among all the classifiers Random forest outperforms in regards to the performance evaluation matrices. The achieved accuracy of their system is 89.59% and has also some errors which are 10.41%. The steps they followed are image acquisition, resizing, segmentation of images, extraction of feature, classification, output generation. Their dataset consists of 480 images with 5 classes for both disease and
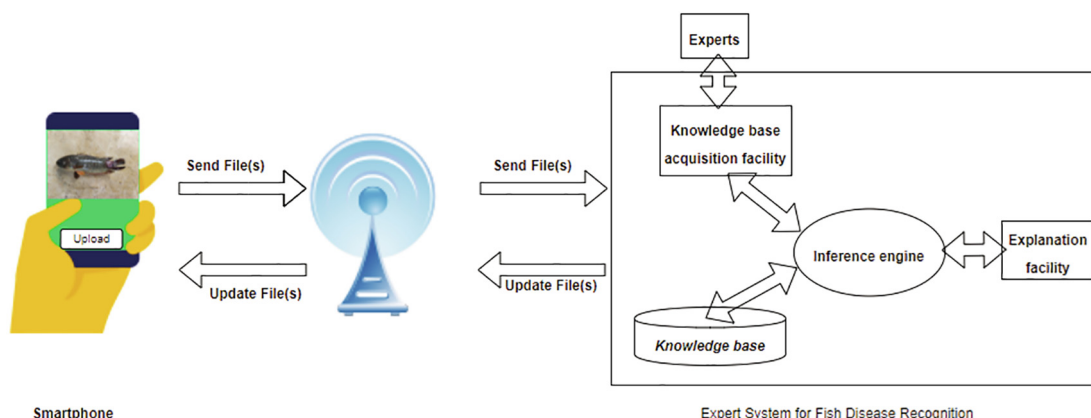

Fig. 2. Expert system to recognize fish disease (EUS, Fin and Tail Rot).

non-disease images. Ghyar and Birajdar (2017) proposed a model for rice leaf disease recognition using texture and color features. They have used SVM and ANN as a classification algorithm which provides respectively 92.5% and 87.5% accuracy. 120 image datasets were used in their work. For image segmentation, they used the thresholding method and extracted 13 features from the rice leaf images. Aftabuddin et al. (2016) have done a survey on the status of fish disease and remedial measurements taken by the fish farmers. They have taken interviews sessions as well as Focus Group Discussion (PGD) of 196 farmers regarding fish diseases. Study shows that disease can detriment the freshwater aquaculture due to a lack of knowledge and veterinary backstopping services. Habib et al. (2019) stated an automated computer vision approach for Jackfruit disease identification and classification. They have also used K-means clustering as a segmentation algorithm to extract the features from captured images. They have used three hundred and sixty data for carrying out the research. Transformation of the image from color space to RGB was done by applying K-means clustering. They have used two features set which is consists of ten features. The achieved accuracy of their model was 88.67% by the SVM classifier. Wei et al. (2018) stated a model for skin disease recognition using image processing techniques. grey-level co-occurrence matrix (GLCM) was introduced in their model to segment the images. Features from the segmented images are extracted accurately. They have used SVM as a classification algorithm and achieved accuracy is 95%. Samajpati and Degadwala (2016) worked on apple fruit recognition using a traditional machine learning approach. They have used 80 images to train the model. K-means clustering is used as a segmentation algorithm to extract the features from images. thirteen features are extracted with the help of the segmentation algorithm which is used to train the model. The random forest classifier provides better accuracy for their proposed systems. Mia et al. (2019) stated an agro medical expert system for various local fruit recognition and classification. They have dealt with the six types of fruit with two hundred and forty images. They have also used K-means clustering is used as a segmentation algorithm to extract the features from images. SVM classifier provides 90.79% accuracy which is promising enough for their proposed systems.

To summarize, the main contributions of our proposed expert system are given below:.

- We came to know that no work has been done for the recognition of two diseases as well as healthy fish. Our proposed system will work as a contributing automated system for the production of fish that can enrich our national economy. So, for the very first time, we have conducted this research in order to recognize the fish disease.
- Extracted features of our system worked as deciding factors for classifying the disease that will form a strong base for further classification in this context.
- Performed in-depth analysis of the classification model which will enhance the performance of the proposed expert system.
- After experimental analysis, it can be seen that the achieved accuracy of our system is promising enough.

3. Research methodology

3.1. Method applied

Fig. 3 shows the step-by-step approach for automated fish disease recognition. From the figure it can be seen that the adopted expert system starts with the color image of diseased and healthy fish.

After the acquisition of the image, it has been resized with a fixed size with the help of bicubic interpolation (Keys, 1981).

Considering that we found the intensity values as I, and differential coefficients are the I_x, I_y, and I_{xy}. Equation of the interpolated intensity surface is written as:.

$$s(x, y) = \sum_{m=0}^{3} \sum_{n=0}^{3} c_{mn} x^m y^n,$$ (1)

where c_{mn}'s are coefficients.

Then image contrast is enhanced with the help of histogram equalization. By letting, n and m are the column-wise and row-wise pixel frequencies. I_k is the color intensity, c_k is the frequency in the pixels and l be the legitimate level of color intensity in the image. Every pixel of images are comprising color intensity I_k is mapped in a similar type of pixel enduring color intensity $I'k$ with the help of a technique called histogram equalization. Thus, the color-mapped image is built as follows (Gonzales and Woods, 2008).

$$I'_k = T(I_k) = \frac{l-1}{mn} \sum_{i=0}^{k} c_i$$ (2)

where k = 0, 1, …, $l - 1$.

It can be difficult to obtain good results by relying solely on raw images. Owing to this reason, data processing such as resizing, contrast enhancement is performed. After contrast enhancement of digital images, color conversion is done from RGB to $L^*a^*b^*$. The segmentation algorithm K-means clustering is employing for $L^*a^*b^*$ color space where L is luminosity layer 'L^*' and a^*b^* is chromaticity-layer. Firstly, the transformation of images is done from $L^*a^*b^*$ to CIE (International Commission on Illumination) named a color space XYZ (Color Conversion Algorithms, 2019). The 'a^*' and 'b^*' layers contain all of the color information, and K-means clustering is used to label each pixel and segment of the image by color in the 'a^*b^*' space. As a result, images are split into different sections with the help of K-means clustering algorithms. The algorithms aim to group similar and dissimilar objects. Furthermore, the complex background of the images is subtracted from the original image. Thus, we got the lucid images with the diseased region as an output of segmentation that has been proven by Burney (Burney and Tariq, June 2014). There are some other segmentation algorithms namely, Otsu's method, fuzzy C-means clustering, and region growing. But K-means clustering works outstanding than any other segmentation algorithm. By implementing this algorithm diseased regions are separated from disease-free regions. We have found 2 feature set consist of 10 features namely co-occurrence matrix and statistical. Furthermore, eight prominent classification algorithms were employed to classify the disease of the fish. Accordingly, in-depth experiments have been done to find out the best classifier using performance evaluation matrices. Implemented classification models are Logistic Regression, BPN, CPN, Gradient Boosting, SVM, Random Forest, K-Star, KNN.

Accuracy cannot be only the performance evaluation matrices to precisely measure the actual classification results since might have not been completely fitted for estimating the performance of the adopted classification model with unlabeled data set. There are some other evaluation matrices except accuracy to measure the performance of the adopted expert systems as presented in (Tan et al., 2006) and (Han et al., 2012). For the binary confusion matrix, the number of true positives (TP's), and false negatives (FN's), true negatives (TN's), and false positives (FP's) is calculated to measure the performance. But in multiclass, there have been more than 2 classes that can be calculated using the following equations (3).

$$\mathbf{C} = [e_{ij}]_{n \times n}.$$ (3)

There are no direct techniques for multiclass to calculate the number of TP's, FP's, FN's, and TN's. It can be easily seen that according to the (Habib et al., 2020) confusion matrix C for the
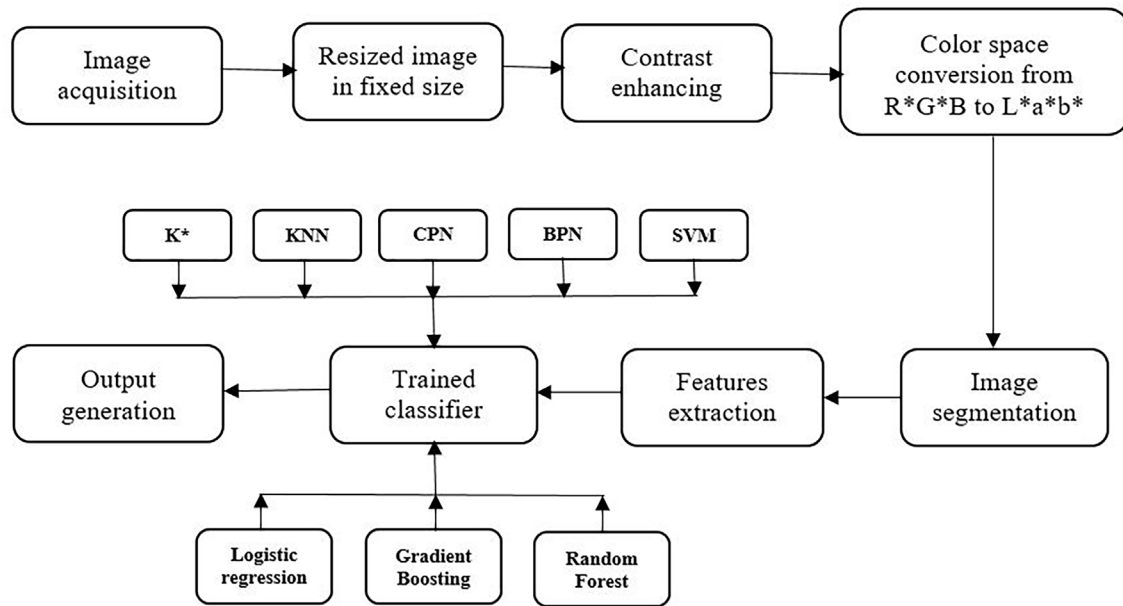
Fig. 3. Working process of automated fish disease recognition.

multiclass that is $n \times n$ matrix (where $n > 2$). In the $n \times n$ matrix, n is the number of rows and columns respectively. We have found n^2 records from the $n \times n$ matrix. *TP's*, *FP's*, *FN's*, and *TN's* are calculated from the confusion matrix C as follows:.

$$TP_c = e_{cc}. \tag{4}$$

$$FP_i = \sum_{\substack{j=1, \\ j \neq i}}^{n} e_{ji}. \tag{5}$$

$$FN_i = \sum_{\substack{j=1, \\ j \neq i}}^{n} e_{ij}. \tag{6}$$

$$TN_i = \sum_{\substack{j=1, \\ j \neq i}}^{n} \sum_{\substack{k=1, \\ k \neq i}}^{n} e_{jk}. \tag{7}$$

By employing the adopted techniques, we can build up the binary two-dimensional 2×2 square matrices which consist of the average results of the confusion matrix for individual classes. To conclude, performance evaluation matrices namely accuracy, sensitivity or true positive rate, specificity or true negative rate (TNR), *F1*-score, precision, false-positive rate (FPR), and false-negative rate (FNR) used for evaluating the performance of the adopted expert systems for fish disease recognition. All the evaluation matrices are calculated as follows:.

$$Accuracy = \left(\frac{TP + TN}{TP + FP + FN + TN} \times 100 \right)\%. \tag{8}$$

$$Sensitivity = \left(\frac{TP}{TP + FN} \times 100 \right)\%. \tag{9}$$

$$Specificity = \left(\frac{TN}{FP + TN} \times 100 \right)\%. \tag{10}$$

$$Precision = \left(\frac{TP}{TP + FP} \times 100 \right)\%. \tag{11}$$

$$F_1 \; score = \left(\frac{2 \times TP}{2 \times TP + FP + FN} \times 100 \right)\%. \tag{12}$$

$$FPR = \left(\frac{FP}{FP + TN} \times 100 \right)\%. \tag{13}$$

$$FNR = \left(\frac{FN}{TP + FN} \times 100 \right)\%. \tag{14}$$

The 10 fold cross-validation techniques are applied, as stated in (Habib et al., 2020) and (Tan et al., 2006) to estimate the overall performance of the employed classifier accordingly using the equations (8)–(14). Finally, we have found the most desirable classifier to classify the fish disease recognition.

3.2. Description of fish diseases and features

In-depth investigation of disease is playing a most significant role to develop a system to classify accurately. In this task, image acquisition is the initial step where a number of the image are restored for training and testing. Digital and mobile cameras are used for capturing these images. We have not only collected diseased fish images but also images of healthy fish. Captured images are of different quality. The picture is in RGB form. A couple of images of Epizootic Ulcerative Syndrome (EUS) and Fin and Tail Rot disease affected fish are shown in Fig. 4.

- **Epizootic ulcerative syndrome:** EUS is a common fish disease. It's mainly known as red spot disease which is caused by oomycete fungi. EUS causes hideous lesions in diseased fish. Lesions can be seen most often on the adjacent surface but also found in any part of the body in diseased fish. It spread rapidly in the water (Disease Guide, 2019).
- **Fin and Tail Rot:** Tail and fin rot is the most well-known disease in tropical fishes. It is caused by gram-negative bacteria such as Vibrio, Aeromonas, and Pseudomonas. The disease-affected fish has lesions and erosion in the fin and tails. The fin or tail edges have turned white most of the time but it can be black and brown in some cases. A part of the Fin and Tail have rotted and fallen (Rahman et al., 2010).
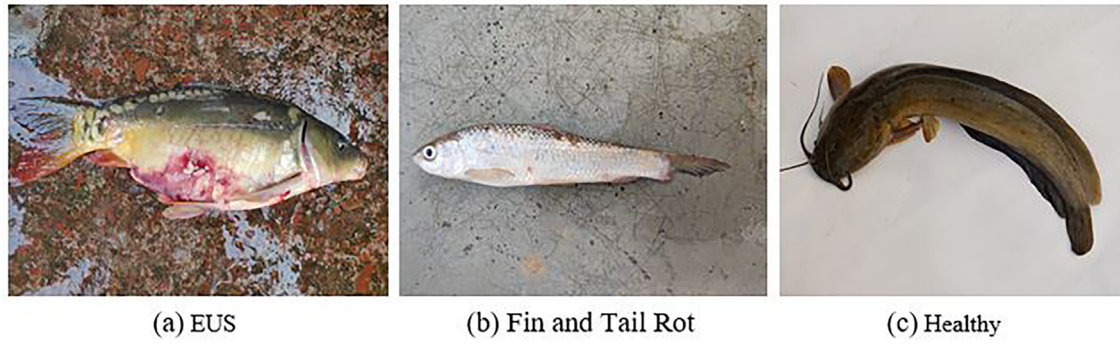
Fig. 4. Common fish disease: (a) Epizootic Ulcerative Syndrome (EUS) (b) Fin and Tail Rot (c) Healthy.

- **Healthy:** Fish body with no spot and also tail and fins are all okay. Full fish should be fresh.

We have used two types of feature set to carry out our research namely statistical and gray-level co-occurrence matrix (GLCM). We have selected 10 features from both features set. The adopted classification model worked based on the extracted features of the input images to classify the diseases correctly. There is some effective research in computer vision in another context. Habib et al. (2020) presented an approach to jackfruit disease recognition using the mentioned features. As a result, we have selected the features that already has been used by Habib et al. (2020). Standard deviation (σ), variance ($\sigma2$), mean (μ), kurtosis (κ), and skewness (γ) are the selected statistical features in our research. If I is the gray-scale intensity of the pixel and $I_\sigma, \bar{I}, I_M$ standard deviation, mean, and mode of the gray-scale intensity for all the pixel particularly. Equation of the mentioned features shown as follows:

$$\text{Standard deviation}\,(\sigma)\,:\,\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(I_i - \bar{I}\right)^2}{n}}. \tag{15}$$

$$\text{Mean}\,(\mu)\,:\,\mu = \frac{1}{n}\sum_{i=1}^{n}I_i \tag{16}$$

$$\text{Variance}\,(\sigma^2)\,:\,\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(I_i - \bar{I}\right)^2. \tag{17}$$

$$\text{Skewness}\,(\gamma)\,:\,\gamma = \frac{\bar{I} - I_M}{I_\sigma}. \tag{18}$$

$$\text{Kurtosis}\,(\kappa)\,:\,\kappa = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(I_i - \bar{I}\right)^4}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(I_i - \bar{I}\right)^2\right)^2} - 3 \tag{19}$$

Haralick et al. (Haralick et al., 1973) stated experiments using GLCM features in this field. They have examined that five GLCM features play a significant role to classify the disease from images. If $g(x, y)$ is a binary grey-scale image of size $K \times L$ pixels with lg grey levels. Let us assume some variables namely *(p1, q1)* and
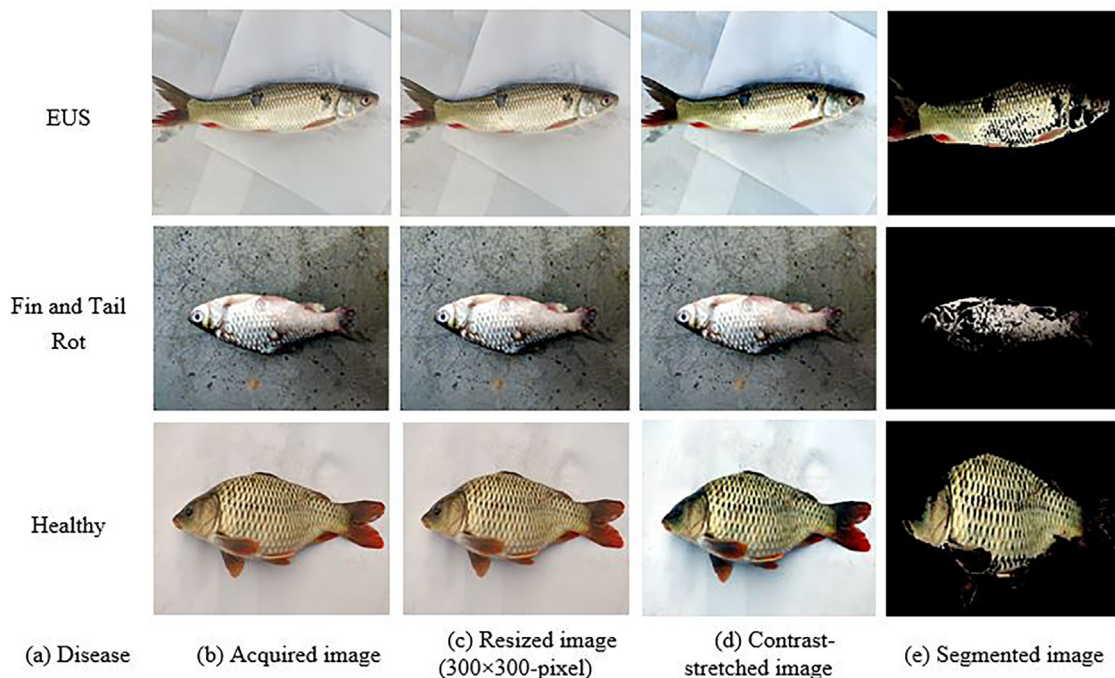

Fig. 5. Stepwise changes of images. (a) Disease. (b) Acquired image. (c) Resized image (300 × 300 pixels). (d) Contrast-stretched image. (e) Disease-affected segmented areas.

(p2, q2) are two pixels in *g(x, y)*, *r* is the distance and α is the angle between two and the ordinate. According to (Naikwadiand and Amoda, 2013), a GLCM P*(i, j, r, α)* equation as follows:

$$P(i, j, r, \alpha) = |(p_1, q_1), (p_2, q_2) \in K \times L: r, \alpha, g(p_1, q_1) = i, g(p_2, q_2) = j|. \tag{20}$$

Five types of GLCM features have used in our research namely, contrast (*C*), entropy (*S*), energy (*E*), correlation (ρ), and homogeneity (*H*) in our context. The equations of all the features are presented as follows, where *P (i, j)* is the (*i, j*)-th entry of the estimated GLCM: *lg* is the total amount of grey levels in an image;

and *sx*, *sy*, and *mx*, *my*, are the standard deviations and means of the row and column sums of the GLCM. Equation of GLCM features are shown as follows:.

$$\text{Contrast} (C) : C = \sum_{i=0}^{l_g-1} \sum_{j=0}^{l_g-1} (i-j)^2 P(i,j) \tag{21}$$

$$\text{Correlation} (\rho) : \rho = \frac{\sum_{i=0}^{l_g-1} \sum_{j=0}^{l_g-1} ijP(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y} \tag{22}$$

i). Actual Class	ii). Acquired Image	iii). *K*-means clustering segmented image	iv). Extracted Features (C), (ρ), (E), (H), (μ), (σ), (S), (σ²), (κ), (γ)	v). Recognized Class
Fin and Tail Rot			(0.64, 0.75, 0.85, 0.97, 9.46, 39.07, 0.83, 1452.22, 19.59, 4.20)	Epizootic Ulcerative Syndrome (EUS)
Fin and Tail Rot			(0.41, 0.77, 0.92, 0.98, 5.82, 32.92, 0.46, 1030.52, 35.44, 5.77)	Fin and Tail Rot
Epizootic Ulcerative Syndrome (EUS)			(0.29, 0.92, 0.82, 0.97, 13.85, 47.70, 1.25, 2183.69, 15.88, 3.68)	Epizootic Ulcerative Syndrome (EUS)
Epizootic Ulcerative Syndrome (EUS)			(1.14, 0.75, 0.46, 0.90, 30.44, 52.99, 2.91, 2265.35, 4.69, 1.60)	Fin and Tail Rot
Healthy			(0.49, 0.84, 0.82, 0.96, 11.64, 42.69, 1.07, 1702.94, 17.13, 3.85)	Healthy
Healthy			(0.07, 0.82, 0.90, 0.99, 3.19, 16.72, 0.56, 268.08, 38.85, 5.84)	Epizootic Ulcerative Syndrome (EUS)

Fig. 6. Representation of extracted features which get from a pair of three different class's images where each image of the pairs is correctly recognized and other image of the pair is incorrectly recognized.

Table 1
Class wise collected data set.

Class Name	Number of Images
EUS	211
Tail and Fin rot	99
Healthy	175
Total	485

Table 2
Detailed specification of classifier.

Classifier	Specifications
Random Forest	Size of each bag = Percentage of training data size = 100%. Number of iterations = the number of trees in the random forest = 100 The maximum depth of the tree = 0
Gradient Boosting	The loss function to be optimized. 'deviance' refers to deviance Number of boosting stages = n_estimators (int type) = 100 (default)
BPN	Network topology: Fully connected two-layered feedforward (10–15–5) Transfer function: Sigmoid function, $s_c(x) = (1 + e^{-\alpha x})^{-1}$ Learning rate, $\gamma = 0.27$. Momentum rate, $\alpha = 0.21$.
CPN	Network topology: Fully connected two-layered feedforward (10–11–5) Layer 1|2: Learning rate, $\gamma 1\|2 = 0.35$. Layer 2|3: Learning rate, $\gamma 2\|3 = 0.25$.
KNN	*Manhattan Distance,* $L_M = \sum_{i=1}^{10}[x_i - f_i]$. *Weight vector, W =* $(w_1, w_2, w_3, \ldots \ldots \ldots \ldots \ldots \ldots \ldots w_{10}) = (1)_{1 \times n}$. *k = 3.*
SVM	*Linear function used as Kernel Value of C is 250043.*
K*	*Manual blend setting = 20% Enable entropic auto-blend setting = Only symbolic class Specify the missing value in treatment mode*
Logistic Regression	*Multinomial Model Logistic function:* $logit(q) = \ln(q) - \ln(1 - q)$

$$\text{Entropy } (S) : S = -\sum_{i=0}^{l_g-1}\sum_{j=0}^{l_g-1} P(i,j)\log P(i,j) \qquad (23)$$

$$\text{Energy } (E) : E = \sum_{i=0}^{l_g-1}\sum_{j=0}^{l_g-1} P(i,j)^2 \qquad (24)$$

$$\text{Homogeneity } (H) : H = \sum_{i=0}^{l_g-1}\sum_{j=0}^{l_g-1} \frac{P(i,j)}{1+(i-j)^2} \qquad (25)$$

4. Experimental evaluation

Computer vision-based agro-medical expert system for fish disease recognition and classification is portrayed in Fig. 2. This figure just shows the hypothesis that fisheries farmers who need to identify the fish disease at an early stage to stop the spreading of disease can capture images through smartphones or other digital devices. Different devices have different screen sizes that can affect the accuracy of the system. Consequently, the images of different sizes are converted into a fixed size of 300 × 300 pixels. Image is captured from different angles and perspectives. Furthermore, the contrast of the image is enhanced with the help of histogram equalization techniques. Segmentation techniques K-means clustering employed on the acquired image after contrast enhancement for image feature extractions. Using mentioned segmentation algorithm diseased regions are extracted from disease-free regions. Fig. 5 shows the pictorial diagram of image segmentation.

K-means clustering performs outstanding as a segmentation algorithm than any others algorithm already stated by Rozario et al., 2016. These are the most significant steps for image recognition. To evaluate the performance of the segmentation algorithms ground truth images are labeled manually. After completing the training of the data we have tested using the systems which illustrate in Fig. 6. We have used six images of fish including four diseases as well as two healthy images of fish. From all the images three are recognized correctly and the other three are recognized incorrectly presented in the Fig. 6.We have used a total of 485 images for three classes including two disease classes (i.e: EUS, Fin and Tail Rot) and healthy fish images. All the images are collected locally from the fish market and fisheries. The class-wise image data is already depicted in Table 1. The cross-validation (Han et al., 2012; Tan et al., 2006) techniques are used to evaluate our expert system. 10-fold cross-validation techniques are applied in our adopted model. Following this method, nine parts are used for training purposes, and one part is used for test purposes. After completion of the training, we have calculated the performance evaluation matrices with the help of the confusion matrix of the test data. The acquired confusion matrix of our model is three binary (2 classes) that are already explained by Habib et al. (2020). After getting the matrix, step by step experimental evaluation done according to the classifier which shown in Table 2.

During the performance measurement of the system's number of the classifier is employed. We have used eight asserted classifiers namely *Logistic Regression, BPN, CPN, Gradient Boosting, SVM, Random Forest, K-Star, KNN*. They have not been adopted whimsically, rather consciously to include a broader range of all the classifiers. The values of all the variables are calculated through the training process. Calculated variables are used for evaluating the performance. Remarkable several performance matrices are namely i.e. accuracy, sensitivity, specificity, precision, *F1*-score,

Table 3
Performance evaluation matrices of all classifier.

Classifier	Metric							
	Accuracy	Error rate	Sensitivity	Specificity	Precision	F_1-Score	FPR	FNR
Logistic Regression	83.64%	16.36%	69.94%	87.01%	71.71%	70.31%	12.99%	30.06%
BPN	85.02%	14.98%	72.37%	88.00%	74.84%	72.95%	12.00%.	27.63%
CPN	84.74%	15.26%	73.56%	88.59%	76.60%	74.28%	11.41%	26.44%
Random Forest	88.87%	11.13%	80.01%	91.08%	82.40%	80.91%	08.92%	19.99%
SVM	77.04%	22.96%	59.21%	81.56%	60.02%	67.87%	18.44%	40.79%
Gradient Boosting	86.67%	13.33%	74.93%	89.22%	78.24%	75.76%	10.78%	25.07%
K-Star	79.24%	20.76%	64.43%	83.88%	64.42%	64.35%	16.12%	35.57%
KNN	81.03%	18.97%	34.99%	84.58%	67.45%	65.11%	15.42%	34.99%

Table 4
Class wise performance evaluation matrices of all classifier.

| Classifier	Disease	Metric							
		Accuracy	Error rate	Sensitivity	Specificity	Precision	F1 score	FPR	FNR
Logistic Regression	EUS	79.79%	20.21%	81.99%	78.10%	74.25%	77.93%	21.90%	18.01%
	Fin and Tail rot	81.44%	18.56%	40.40%	91.97%	56.34%	47.06%	08.03%	59.60%
	Healthy	89.69%	10.31%	87.43%	90.97%	84.53%	85.96%	09.03%	12.57%
BPN	EUS	80.83%	19.17%	82.94%	79.20%	75.43%	79.01%	20.80%	17.06%
	Fin and Tail rot	83.51%	16.49%	44.44%	93.52%	63.77%	52.38%	06.48%	55.56%
	Healthy	90.72%	09.28%	89.71%	91.29%	85.33%	87.47%	08.71%	10.29%
CPN	EUS	81.44%	18.56%	84.36%	79.20%	75.75%	79.82%	20.80%	15.64%
	Fin and Tail rot	84.33%	15.67%	45.45%	94.30%	67.16%	54.21%	05.70%	54.55%
	Healthy	91.75%	08.25%	90.86%	92.26%	86.89%	88.83%	07.74%	09.14%
Random Forest	EUS	88.66%	11.34%	62.63%	95.34%	77.50%	69.28%	04.66%	37.37%
	Fin and Tail rot	85.98%	14.02%	87.68%	84.67%	81.50%	84.48%	15.33%	12.32%
	Healthy	91.96%	08.04%	89.71%	93.23%	88.22%	88.96%	06.77%	10.29%
SVM	EUS	76.70%	23.30%	24.24%	90.16%	38.71%	29.81%	09.84%	75.76%
	Fin and Tail rot	68.66%	31.34%	71.09%	66.79%	62.24%	66.37%	33.21%	28.91%
	Healthy	85.77%	14.23%	82.29%	87.74%	79.12%	80.67%	12.26%	17.71%
Gradient Boosting	EUS	85.15%	14.85%	47.47%	94.82%	70.15%	56.62%	05.18%	52.53%
	Fin and Tail rot	82.27%	17.73%	85.31%	79.93%	76.60%	80.72%	20.07%	14.69%
	Healthy	92.58%	07.42%	92.00%	92.90%	87.98%	89.94%	07.10%	08.00%
K*	EUS	77.11%	22.89%	38.38%	87.05%	43.18%	40.64%	12.95%	61.62%
	Fin and Tail rot	72.78%	27.22%	69.19%	75.55%	68.54%	68.86%	24.45%	30.81%
	Healthy	87.84%	12.16%	85.71%	89.03%	81.52%	83.56%	10.97%	14.29%
KNN	EUS	79.79%	20.21%	29.29%	92.75%	50.88%	37.18%	07.25%	70.71%
	Fin and Tail rot	73.20%	26.80%	77.73%	69.71%	66.40%	71.62%	30.29%	22.27%
	Healthy	90.10%	09.90%	88.00%	91.29%	85.08%	86.52%	08.71%	02.00%
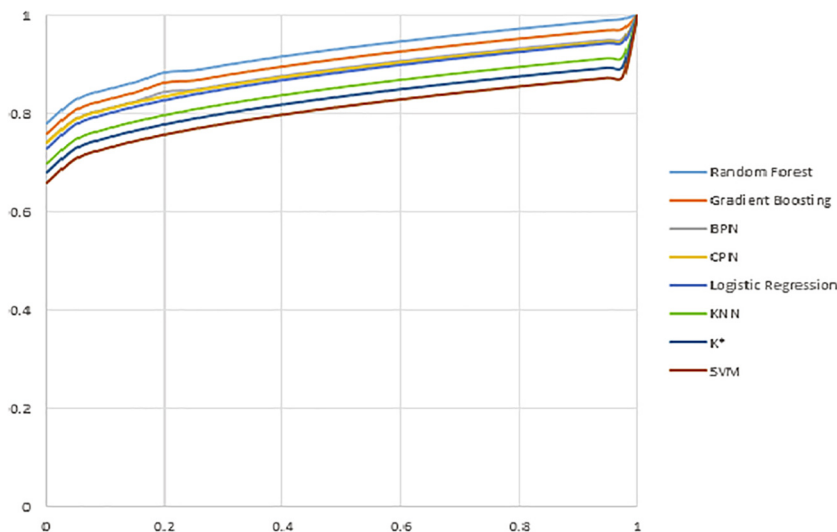
Fig. 7. ROC curve of all the implemented classifiers.

FPR, and FNR, accordingly shown in Table 3. Class-wise performance evaluation matrices of all the classifiers are also shown in Table 4.

It can be seen from Table 3 that the Random forest classifier becomes notable among all the classifiers in the context of perfor-

Table 5
AUC value of all the adopted classifier.

Classifier	AUC
Random Forest	0.8971
Gradient Boosting	0.8776
BPN	0.8581
CPN	0.8553
Logistic Regression	0.8508
KNN	0.8215
K*	0.8020
SVM	0.7825

mance. In the contrast, SVM and K^* provide the poorest accuracy in parallel. Even Random forest provides the highest accuracy, there are some other prominent classifiers like Gradient boosting which also provides good accuracy compared to Random Forest. The accuracy of Random forest and Gradient Boosting is 88.87% and 86.67% respectively which is close enough. We also notice from Table 4 that it visualizes the class-wise performance evaluation matrices of all the classifiers. It remarked that the highest accuracy of individual classes is achieved by the Random forest 88.66% for EUS and 85.98% for Fin and Tail rot. We have found better accuracy in the non-disease or healthy class is of 92.58% by the Gradient Boosting Classifier. Taking all these observations into account as well as analyzing all eight metrics, we can state that the Random forest performs the finest among all of the nine classifiers used concerning the recognition of fish diseases.

We have used another approach for evaluating the performance of all the classifiers. The adopted approach name is a receiver oper-

Table 6
Comparative analysis between relevant research and our work.

Research work	Dealt Objects	Number of class	Data size	Technique Used	Segmentation Algorithm	Feature Set	Best Classifier	Accuracy
This Work	Fish Disease	3	485	Machine Learning	K-Means Clustering	10	Random Forest	88.87%
Malik et al. (2017)	Fish Disease	2	NM	Machine Learning	Histogram Equalization	NM	Neural Network and K-NN	NM
Chakravorty et al. (2015)	Fish Disease	2	NM	Machine Learning	K-means clustering	NM	PCA	NM
Zhang et al. (2017)	Cucumber	7	420	Machine Learning	K-means Clustering	NM	Sparse Representation (SR)	85.70%
Wei et al. (2018)	Skin	3	NA	Machine Learning	K-means Clustering	NM	SVM	90.00%
Habib et al. (2020)	Jackfruit	5	480	Machine Learning	K-means Clustering	10	Random Forest	89.59%
Samajpati and Degadwala (2016)	Apple	3	80	Machine Learning	K-means Clustering	13	Random Forest	(60–100)%
Mia et al. (2019)	Fruit	6	240	Machine Learning	K-means Clustering	10	SVM	90.79%
Rozario et al. (2016)	Fruit	3	63	Machine Learning	K-means Clustering, Modified K-means Clustering, Otsu	NM	NM	NM

NM* =Not Mentioned.

ating characteristic (ROC) curve that is used to visualize the comparative performance of several classifiers. As a result, we have used the ROC curve to measure the performance of classifiers accordingly. Furthermore, if the area under the ROC curve is equal to 1 then the performance of the classifier is considered to be promising enough.

It can be seen from Fig. 7 that the Random forest outperforms among all the classifiers. SVM performs lowest compared to Random forest. AUC values of all the classifiers are shown in Table 5.

5. Comparative analysis

As I have already said that there is no single effective work that has been done for fish disease recognition that dealt with the two fish disease as well as a healthy fish. In order to evaluate the recognition rate of the proposed expert systems, needs to compare with published relevant research work. Only two works have been done for fish disease recognition but they carried their research work only for one disease which is limited compared to data size and accuracy. In addition, most of the implemented work has been done in another context and limited to the perceptions of the computer vision approach for recognition of fruits and vegetable disease shown in Table 6. It can be seen that their effort was good in the recognition but meticulous performance evaluation based on practical evidence is still insufficient. As a result, we have tried hard to compare the various research paper in the context of disease recognition to investigate the performance accurately. Furthermore, we have measured the performance of our system compared to other existing work which is promising. Comparative analysis of different relevant research work with our work has been shown in Table 6. The achieved accuracy of our systems is nearly 89% which claimed that the recognition rate of our proposed expert system is good enough.

6. Conclusion and future work

In this paper, the necessity of computer vision-based fish disease recognition has been investigated intensely. Though fish disease recognition is arduous work, we have done our work perfectly. Two types of features set have been used to carry out this research. In order to extract the features, the *K*-means clustering algorithm has been used. Adopted segmentation algorithm implemented on the digital image which is captured by the digital camera or smart-

phones. Extracted features have been used as input in the eight unique classifiers accordingly. Furthermore, performance evaluation matrices are used to evaluate performance. Among all the classifiers Random forest perform outstandingly in terms of matrices. To be conclude, our proposed systems achieving accuracy is 88.87% that is promising enough. However, there is still some need for improvement. Therefore, future works could be to increase datasets and work on a much broader variety of fish diseases.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Yearbook of Fisheries Statistics of Bangladesh 2018-2019. Available online: http://fisheries.portal.gov.bd/ [Last accessed on May 17, 2021].

Fishery and aquaculture country profiles. Available online: http://www.fao.org/fishery/facp/BGD/en [Last accessed on April 30, 2021].

Aftabuddin, S., Nurul Islam, M., Bhuyain, M.A.B., Abdul Mannan, M., Mahbub Alam, M., 2016. Fish diseases and strategies taken by the farmers in freshwater aquaculture at southwestern Bangladesh. Bangladesh J. Zool. 44 (1), 111–122.

Bangladesh ranks in fish production, Available online: https://en.prothomalo.com/bangladesh/Bangladesh-ranks-eighth-in-fish-production [Last accessed on April 30, 2021]

Malik, Shaveta, Tapas Kumar, Sahoo, A.K., "Image processing techniques for identification of fish disease." In 2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP), pp. 55-59. IEEE, 2017.

Chakravorty, H., Paul, R., Das, P., 2015. Image processing technique to detect fish disease. Int. J. Comput. Sci. Security (IJCSS) 9 (2), 121–131.

López-Cortés, X.A., Nachtigall, F.M., Olate, V.R., Araya, M., Oyanedel, S., Diaz, V., Jakob, E., Ríos-Momberg, M., Santos, L.S., 2017. Fast detection of pathogens in salmon farming industry. Aquaculture 470, 17–24.

Zhang, S., Wu, X., You, Z., Zhang, L., 2017. Leaf image based cucumber disease recognition using sparse representation classification. Comput. Electron. Agric. 134, 135–141.

Habib, Md Tarek, Md Jueal Mia, Mohammad Shorif Uddin, and Farruk Ahmed. "An in-depth exploration of automated jackfruit disease recognition." Journal of King Saud University-Computer and Information Sciences (2020).

Bhagyashri S. Ghyar and Gajanan K. Birajdar, "Computer Vision Based Approach to Detect Rice Leaf Diseases using Texture and Color Descriptors", Proceedings of the In-ternational Conference on Inventive Computing and Informatics (ICICI 2017) IEEE Xplore Compliant - Part Number: CFP17L34-ART, ISBN: 978-1-5386-4031- 9.

Md Tarek Habib, Md Robel Mia, Md Jueal Mia, Mohammad Shorif Uddin, and Farruk Ahmed. "A Computer Vision Approach for Jackfruit Disease Recognition." In International Joint Conference on Computational Intelligence, pp. 343-353. Springer, Singapore, 2019.

Wei, L.-S., Gan, Q., Ji, T., 2018. Skin disease recognition method based on image color and texture features. Comput. Math. Methods Med. 2018.

Samajpati, B.J., Degadwala, S.D., 2016. Hybrid approach for apple fruit diseases detection and classification using random forest classifier. In: 2016 International conference on communication and signal processing (ICCSP), pp. 1015–1019.

Mia, Md Robel, Md Jueal Mia, Anup Majumder, Soummo Supriya, and Md Tarek Habib. "Computer Vision Based Local Fruit Recognition." Int. J. Eng. Adv. Technol. 9, no. 1 (2019): 2810-2820.

Keys, R., 1981. Cubic convolution interpolation for digital image processing. IEEE Trans. Acoust. Speech Signal Process. 29 (6), 1153–1160.

Gonzales, R.C., Woods, R.E., "Digital Image Processing," Pearson, 2008.

Color Conversion Algorithms, Available online: http://cs.haifa.ac.il/hagit/courses/ist/Lectures/Demos/ColorApplet2/t_convert.html#RGB%20to%20XYZ%20&%20XYZ%20to%20RGB [Last accessed on October 12, 2019]

Burney, S.M.A., Tariq, H., June 2014. K-means cluster analysis for image segmentation. Int. J. Comput. Applications 96 (4), 1–8.

Tan, P.-N., Steinbach, M., Kumar, V., 2006. Introduction to Data Mining. Addison-Wesley.

Han, J., Kamber, M., Pei, J., 2012. Data Mining Concepts and Techniques. Elsevier.

DISEASE GUIDE Epizootic Ulcerative Syndrome (Red Spot Disease), Available Online: https://thefishsite.com/disease-guide/epizootic-ulcerative-syndrome-red-spot-disease [Last accessed on October 12, 2019].

Rahman, M.M., Ferdowsy, H., Kashem, M.A., Foysal, M.J., 2010. Tail and fin rot disease of Indian major carp and climbing perch in Bangladesh. J. Biol. Sci. 10 (8), 800–804.

Haralick, R.M., Shanmugam, K., Dinstein, I., "Textural Features for Image Classification," IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-3, issue 6, pp. 610–621, Nov. 1973.

Naikwadiand, S., Amoda, N., "Advances in image processing for detection of plant diseases," International journal of application or innovation in engineering & management, pp. 168-175, Nov. 2013.

Rozario, L.J., Rahman, T., Uddin, M.S., 2016. Segmentation of the region of defects in fruits and vegetables. Int. J. Comput. Sci. Information Security 14 (5), 399.