

AN LSTM BASED CNN MODEL FOR POMEGRANATE FRUIT CLASSIFICATION WITH WEIGHT OPTIMIZATION USING DRAGONFLY TECHNIQUE

M.T.Vasumathi

Research Scholar, Department of Computer & Information Sciencet,
Annamalai University, Annamalai Nagar, Chidambaram, Tamilnadu, India
vasumathi_mt@yahoo.com

M. Kamarasan

Assistant Professor, Department of Computer & Information Science,
Animalia University, Annamalai Nagar, Chidambaram, Tamilnadu, India
smkrasan@yahoo.com

Abstract - Pomegranate is a widely grown plant in India. This highly beneficial fruit is infected by multiple pests and diseases which cause great economical losses. Different forms of pathogen diseases on leaf, stem and the fruits are present. Some of the diseases that affect pomegranate fruits are anthracnose, cercospora, heart rot and bacterial blight. There is a need for disease control strategies to incorporate timely action on the developed diseases. Thus, there is a need for intelligent and self-learning recognition systems to detect these diseases on time. This study is aiming to classify pomegranate fruits into two classes normal and abnormal using CNN LSTM technique. This research work uses a hybrid CNN-LSTM technique to detect four types of diseases present in the pomegranate fruits and classify them into four classes. The results obtained using CNN LSTM are then optimized using dragonfly algorithm. The features like colour, texture and shape of the fruits are collected and fed into the hybrid CNN-LSTM. The dataset for the classifier is given as an excel file which is initially pre-processed using map reduce technique and dimensionality reduction carried using Principal Component Analysis and Discriminant analysis. The CNN LSTM classifier identifies the 4 types of diseases and normal fruit. The classification is further optimized using dragonfly algorithm. The optimized weight and cost function has further explored to support the multi-class disease detection process. Experimental results have shown an accuracy of 92% in classification using CNN-LSTM technique and optimization using dragonfly techniques shows an improved classification accuracy of 97.1%.

Keywords: Hybrid LSTM-CNN, Dragonfly algorithm, Optimization, multi-class classification.

1. Introduction

Pomegranate (*Punica granatum*) is a medically high-valued fruit. It is found that pomegranates are helpful in prevention of cancer and treats heart diseases. It is a very good antioxidant. However, these fruits develop many diseases due to various [1] pests during the course of the growth of the tree and fruit as well. Pomegranates are affected by diseases like anthracnose, bacterial blight, cercospora, heart rot, wilt and so on. The diseases in the fruits can be clearly identified by certain abnormalities appearing on the external features of the fruit. The colour, shape, weight, spots on the fruit allows us to detect the type of the disease present in the fruit. The present study concentrates on detecting four classes of diseases present in the fruit namely anthracnose, bacterial blight, cercospora, heart rot [2]. This paper describes the accuracy obtained in detecting and classification of diseases using machine based techniques. The five classes include four diseases mentioned above and fifth class includes normal healthy fruits. The symptoms of the four mentioned diseases are described below.

Anthracnose

This disease shows symptoms like small, regular to irregular black spots on leaves, calyx region and fruits which turn later on as dark brown depressed spots. Infected leaves turn yellow and drop off.

Bacterial Blight

One to several small water soaked, dark coloured irregular spots appear on leaves resulting in premature defoliation under severe cases is the major symptom of this disease. The pathogen also infects stem and branches causing girdling and cracking symptoms. Spots on fruits were dark brown irregular slightly raised with oily appearance, which split open with L-shaped cracks under severe cases.

Cercospora

Affected plants show small, dark brown spots on fruits/leaves that are initially circular but eventually become irregular as they grow. On the leaves, the lesions are dark, reddish brown to almost black and show a faint halo. The spots on fruit resemble bacterial spot, but they are darker, of various sizes, without cracks, and no stickiness. Infected twigs produce raised, black lesions, and the twigs dry out and die. More severe infections cause the leaves to turn yellow and fall prematurely. Spots on leaves decrease yield of fruit via decreased rate of photosynthesis.

Heart Rot

Diseased fruit generally are lighter in weight and the rind may be pale red or brownish-red. In later stages arils show a brown decay and black sporulation can be seen inside the fruit. In advanced stages of internal decay the fruit's exterior show some shriveling. The images of the fruits acquired by the above diseases are shown in figure 1.

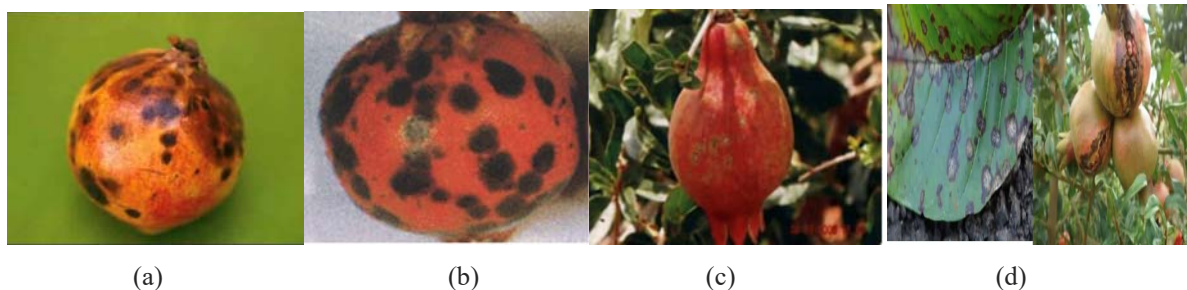


Figure 1. (A) Anthracnose (b) Bacterial Blight (c) Cercospora (d) Heart Rot

Owing to the detection of diseases, several techniques are established and also addressed the problem of recognizing at an earlier stage. In order to detect and classify pomegranate fruit diseases an application of Machine Learning Techniques (MLT) [3] is employed. Mostly machine learning techniques related to the classification of the diseases are easily designed. It helps to design a classifier by making use of relevant features that abides the prediction of diseases. It also assists in understanding the accuracies of the detection process.

The research paper is organized into 6 sections, Section 2 presents the Related work; Section 3 presents the Research methodology; Section 4 depicts the Fruit disease identification using LSTM based CNN Model; Section 5 presents the Optimization of the proposed model; Section 6 presents the Results and discussions.

2. Literature Review:

This section includes the study of various existing techniques on fruit disease identification and provides a brief overview of significant research conducted in the realm in fruit disease identification and improving the efficiency of the fruit disease identification processes. There are not many studies or research on pomegranate disease detection using machine learning techniques so far. The pomegranate fruits [4] are classified as infected and uninfected using SVM classifier by Manish Bhange and Hingoli Wala. Features such as color histogram, coherence and local binary patterns were employed on the pomegranate images. Then, the extracted features were given to k-mean clustering techniques which effectively classified the detected diseased apples. System wasn't able to handle multi-diseases detection. They obtained an accuracy of 82%. In another paper pomegranate disease classification is done using Ada-boost Ensemble classifier by Pooja Kantale and Shubada Thakare [5]. The researchers have identified and classified three diseases namely anthracnose fruit rot, anthracnose and bacterial blight. The processing time of the Ada-Boost classifier for training is 14.15 sec. The classification accuracy obtained using Ada-boost classifier is 92.9%, 90.6% sensitivity and 89.83% f-score. In another paper by Shaath D M [6] and et al. three diseases of pomegranate fruits such as cercospora, bacterial blight and pomegranate borers are detected using image processing techniques. It gives a detection accuracy of 85%. Using artificial neural networks Mrunmayee Dhakate and Ingole A. B. [7] tries to diagnose diseases in pomegranates and in this research work feature extraction is done using GLCM method. The ANN classification using Back Propagation algorithm has resulted in an overall accuracy of 90%. Prof. Sona Pawara, Dnyanesh Nawale [8], Kunal Patil, Rakesh Mahajan says early detection of pomegranate disease using machine learning and Internet of Things can be done and also suggested a working model for the same.

As there is not much work done on multi-class detection of pomegranate fruits disease, we have done literature surveys on other common fruits like apple, grapes, peach, cherry and pear.

In [9], they discussed a novel segmentation on fruits by extracting the information from the colours. The segmented information was given as inputs to the k-mean clustering technique. The segmentation process was carried into two stages: a) apple image pixels were clustered on their colour and spatial features. Then, the clustered pixels were then combined into specified regions. Experimental analysis was conducted on the apple fruit that detected the defected apples by its colour. The detected accuracy was higher than the existing systems. The false positive rate was higher due to the similarity colours.

Features such as colour histogram, coherence and local binary patterns were employed on the pomegranate images. Then, the extracted features were given to k-mean clustering techniques which effectively classified the detected diseased apples. System wasn't able to handle multi-diseases detection. Author in [10] presented the recognition model for wheat leaf disease using learning vector optimization based neural networks. Here, the multiple parameters were taken as the feature input. Based on the colour of the diseased image part was segmented using k-mean hard clustering. To begin the process, lifting wavelength transform and the pulse coupled neural network were employed to improve the image segmentation process. The combination analysis of the features has recognized the wheat leaf diseases with the lowered FP rate. However, the computational steps for feature vector analysis are complex.

In [11], they discussed the effects of ultrasound and the SAR image processing analysis of segmentation was studied. Segmentation process is a complex task as well as it determines the classification efficiency. It comprises partitioned images to extract the useful information from the partitioned images. Several general-purpose algorithms were designed to refine the image segmentation process. In [12], they discussed a unified approach that makes use of many features and classifiers. The objective of the study was to eliminate the high time consumed during the training process. In order to deal with the pattern recognition issues, several feature descriptor problems were resolved. By doing so, the accuracy of the classification system was improved. The fusion of features were creating different issues in the classification results. Some features were not normalized and pre-processed. Thus, a multi-class approach under a semi-controlled environment was studied. The designed multi-class model [13] has increased the classification accuracy with simpler computational steps. Backpropagation networks and radial basis functions were designed to recognize the fruit diseases images. It was implemented in the grape fruits. The colour, shape and texture features were obtained from the diseased grape plant. It has increased the false positive rate. In developing countries, the detection of plant diseases has consumed more time and is also very expensive. Before fed into the neural networks, k-mean clustering was used to obtain the disease parameters [14].

With the help of SVM classifiers, the segmentation method was improved using refining threshold strategies for the extracted textural features [15]. This has reduced the false positive rate even for multi-class detection approaches. Classification techniques on different diseases were studied using genetic algorithms. The images were segmented using genetic algorithms that described the textural and shape features information. Then, the segmented images are fed into the PSO and SVM to detect and classify the grape fruit diseases [16]. However, some climate oriented image features are ignored in this study. Colour based segmentation was improved to recognize the diseases in pomegranate leaves [17] by extracting the information from colour, morphology and textures. It effectively classified the diseased leaves of pomegranate fruit. Regardless, the rate of false positive rate and the false negative rate. Most studies presented the improvisation of classification results via different machine learning classifiers. This was further extended by the neural network classifiers during the climatic parametric monitoring of plants. With the help of the IoT system, the plant leaves diseases using textural features. The GLCM matrix was developed from the obtained textural features [18]. This has reduced the time taken for training classifiers. In [19], the author has introduced SVM classifiers by comparing the different kernel functions such as Cauchy kernel, Invmult Kernel and Laplacian Kernel. It was implemented and tested on the tomato leaf. The constant weight estimation on different kernels has displayed the highest rate of RMSE errors.

Pomegranate leaves were classified using SVM classifiers [20] by improvising the segmentation process. Here, statistical features were collected using k-mean clustering. Though it reduced the false positive rate, yet the multi-class issue degraded the system's performance. In [21], the authors have introduced the decision support systems by developing a segmentation approach. Here, it has improved the thresholding for detecting the two types of diseases in classification. It was tested on the grape fruit images which has downy mildew and the powdery mildew. Finally, backpropagation neural networks were employed to classify the two class diseases. This resolved the multi-class issue, yet the computation complexity is high. Advances in smartphones technology [22] was introduced by classifying the diseases based on the severity levels. In [23] SVM and ANN classifier are used for the automatic citrus fruit disease detection and obtained an accuracy of 93.12 and 88.16 respectively. However, the challenges like class imbalance and decision support systems have to be addressed

further. Our work is identifying and classifying four diseases of pomegranate using multi-class classification using CNN LSTM model which has not been suggested or proposed until now.

3. Proposed Methodology

We have used a pomegranate fruit dataset in a form of excel file as input for this system which detects four types of diseases. This machine learning technique allows anyone to easily detect and classify the fruit under a particular disease. The input file consists of data records of healthy and disease affected fruits. The proposed methodology framework is shown in figure.2. First of all 18 features of 6520 pomegranate fruits are observed and recorded in a excel file and given as input to the proposed framework. The dataset is pre-processed and dimensionality reduction and feature extraction techniques are applied to the features. Thus obtained dataset is then fed into the CNN LSTM algorithm for classification of fruits under 5 classes. The accuracy obtained LSTM CNN based classification is further optimized using invasive weed dragonfly algorithm.

3.1 Dataset Preparation

Pomegranate fruits are infected by many diseases such as anthracnose, bacterial blight cercospora, heart rot and so on. These diseases are visually identified by the experienced farmers with certain symptoms on the external features of the fruit like colour, morphology, weight, spots or cracks on the rind of the fruit. To apply machine learning techniques for identification and classification of diseases, the external features of the fruit has been observed and recorded. As there is no benchmark dataset currently available that has recorded these features of the fruit and hence we have used the images available in Kaggle dataset and Cofilab dataset. The images contained fruits with normal and abnormal fruits exhibiting symptoms required for identifying the above said diseases. With the downloaded images of pomegranate fruits with different grade quality and diseases we approached the respectable professor of Department of Horticulture, Annamalai University, Tamilnadu. From the valuable information obtained from the domain expert and by using image processing techniques, the necessary features are observed from the images and recorded as values and stored in an excel file. As the data was still inadequate we visited few pomegranate orchards in and around Bangalore and recorded the fruit features by observing the fruits visually. A set of 6519 sample images of dataset with various diseases fruits were observed and recorded. Actually, 18 features of the data were observed and recorded, but later realised only 9 features are relevant for the detection and classification of the mentioned 4 diseases. This excel file is given as input to the CNN LSTM model for identification and classification of the disease. The sample images of the data set is shown in figure 3. The feature of the fruit is organised in an Excel file and it is shown in Table 1.

2 Principal component analysis: It is one of the dimensionality reduction techniques [24] that eliminates the presence of irrelevant data. It helps to enhance the efficiency of the classifiers. The goal of this study is to transform the data features into a set of new correlated attributes which is known as principal components. It resolves the high dimensional and overfitting data. Let M be the matrix that contain $p * q$ matrix. In $p * q$ matrix, p -vector contains features set and q -vector contains records. It transforms the inputs m_1, m_2, m_3 into the column vectors such as c_1, c_2, \dots, c_n where c has the information about the original data. The principal components reduce the loss of information and also removes the collinearity prevailing in the data. The matrix P and the variables C are given as:

$$C = M P \quad (1)$$

$$P = C M^C \quad (2)$$

Where P is the loadings matrix. This matrix is further analysed by estimating the eigenvectors of the matrix. The eigenvectors values are used for selecting the features.

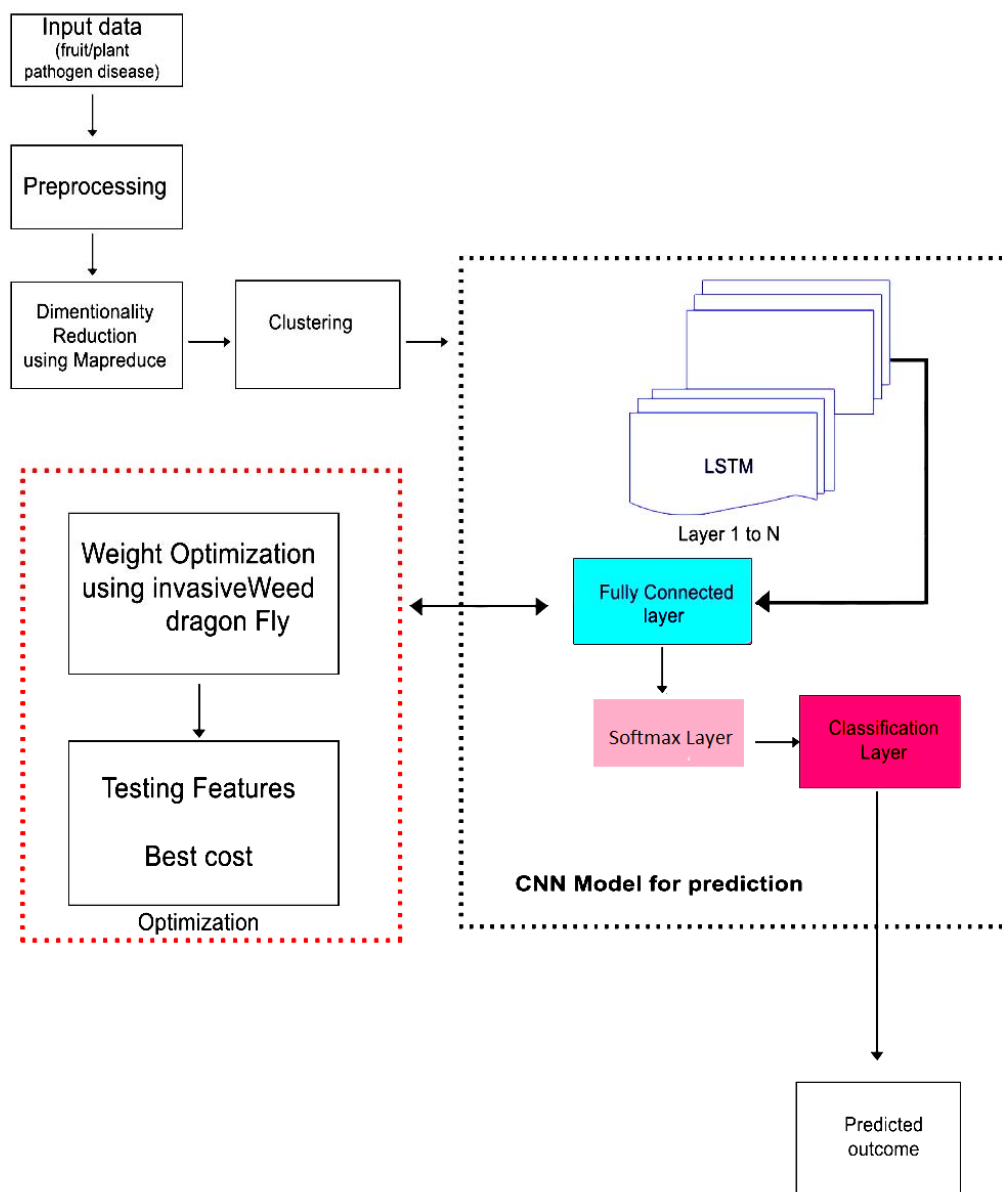


Figure 2.Flow Diagram of Proposed Methodology



Figure 3.Sample Images of Dataset with various Diseases

Table 1.Dataset Features

No	Fruit Spot	Fruit Colour	Leaf Appearance	Leaf Colour	Defoliation	Fruit Shape	Weight	Plant Stand	Leaf Spot Size
0	18	Pale Yellow	Irregular	Yellow	Less	Regular	373	Normal	None
1	0	Dark Orange	Regular	Green	Less	Regular	353	Normal	None
2	1	Pale Yellow	Regular	Yellow	Heavy	Regular	417	Less than Normal	Large Hollow
3	10	Orange	Regular	Yellow	Less	Regular	313	Normal	Small
4	2	Pale Yellow	Irregular	Green			227	Normal	Small
....
6516	23	Dark Orange	Regular	Yellow	Less	Regular	208	Normal	Small
6517	1	Pale Yellow	Irregular	Green	Less	Regular	268	Normal	Small
6518	10	Pale Yellow	Irregular	Yellow	Less	Regular	328	Normal	None
6519	15	Dark Orange	Irregular	Yellow	Less	Regular	237	Normal	None

3.2.1 Discriminant classification analysis [25]: It is a well-known approach that helps to discriminative features transformations. It resolves the multi-class classification issues. The objective of the LDA is to discover the linear transformation among the classes. It is done by the Euclidean distance. It is a kind of data reduction technique that resolves the class imbalance on the lower dimensional space. For a given data sample with high dimensional space D , then records x decides over class C_0 (or) C_1 . Since the two classes are chosen randomly that decides the class based on the available information. The count of linear combination is denoted by the L , where $L < D$. The defined problem is viewed in the linear vector form M to perform the classification and thus, it is given as,

$$M = qD \quad (3)$$

Where,

$Q \in R^{q \times m} \rightarrow$ Linear discriminant

$m \in R^l \rightarrow$ vector that collects the l linear combinations

$D \in R^n \rightarrow$ data to be classified.

Here, the Q value contains the linear discriminant that can be formulated to resolve optimization problems. In such case, the objective measure for the class separability as $Q \in \max(R^{q \times m})$

By defining the optimized rules and the cost function, this closed form expression will yield a better solution to multi-class problems.

3.3 Data Pre-processing: This is the second step of the study that helps to pre-process the data by eliminating the missing and redundant data. In this study, MapReduce technique [26] is used to organize the data for an effective data classification system. MapReduce technique consists three phase, namely,

- Map phase: It takes the input data via key-value pairs.
- Shuffle and sort phase: According to the assigned key, the inputs are organized.
- Reducer phase: The sorted key is processed and the redundant data are reduced.

The arranged files are then divided into blocks where each block contains the similar set of data. There are two nodes, namely, master nodes and the slave nodes. Master node is the main entity that allocates the resources to the slave nodes. All information related to the attributes are stored and retrieved at the time of classification. From the organized data and attributes, the required features are extracted.

4. Fruit disease identification using LSTM based CNN Model

This section elaborates the implementation process and workflow of the proposed model along with different steps employed in this research.

4.1 Convolutional neural network: Its performance is similar to the Artificial Neural Networks (ANNs) [27] that consist of several neurons. Each neuron communicates directly to the convoluted network layer by possessing self-learning quality. The neuron receives an input and data operations, so as to find out their relevant classes. The weight score is taken as an objective function to learn the features and classify the diseases. The final layer of the CNN presents the error functions connected to the classes. CNN architecture [28] is shown in the figure 4. CNN consists of three layers, namely, convolutional layers; pooling layers and the fully-connected layers.

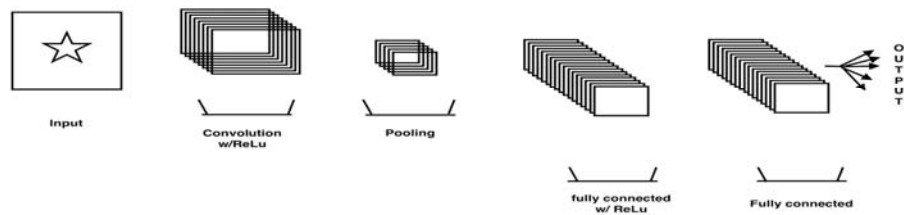


Figure.4 CNN Architecture

The fundamental steps of the CNN are:

- Input layer:** Here the pixel information is given as the input.
- Convolution layer:** It determines the output of the neurons. The textural properties of the input images are evaluated by their weight function. The rectified linear unit (ReLU) acts as an activation function. It learns the pixels of input information, so as to connect with the upcoming layers.
- Pooling layer:** It performs a down sampling process with the spatial information of the given inputs. It tries to reduce the dimensionality issue developed by activation function.
- Fully-connected layers:** The given information is analysed until the pixels input reaches their relevant classes.

4.2 Long short-term memory (LSTM): It performs similar to the Recurrent Neural Network (RNNs) [29]. Additionally, it learns the feature vectors in long-term dependencies and also preserves the learnt information as a default. It helps to reduce the computational complexity. It works in a chain module, regardless, each module has a different form. It consists of four interactive layers which is shown in figure 6.

Initially, the required information is considered by the LSTM [30] networks to apply the sigmoid function. The irrelevant data are identified and eliminated by this network. Each cell unit maintains a timestamp to process the information. With the estimated weight function, sigmoid layers decide the new information and it is being updated incessantly until the new cell states are updated. Finally, the sigmoid layer evaluates the output based on the RMSE value of each cell. Cells with least RMSE are considered as the output layer.

4.2.1 Hybrid LSTM-CNN [31]: In our study, the main novelty lies in hybridizing the LSTM with CNN for fruit diseases classification. The ReLU unit of CNN combines with the LSTM networks. Since we deal with the different types of diseases, the deployment of LSTM network will help the activation unit of CNN to easily compute the multiple classes. The dependencies nature of LSTM model performs more effectively in formatting the input image data for the input layer. Here, sequential steps of vector data with specified time steps have fastened the pooling layer of the CNN architecture. It explores the hidden data patterns from the observed training module with the limited time steps. CNN is used for feature extraction on input data combined with LSTMs to support sequence prediction. The LSTM architecture is shown in figure 5.

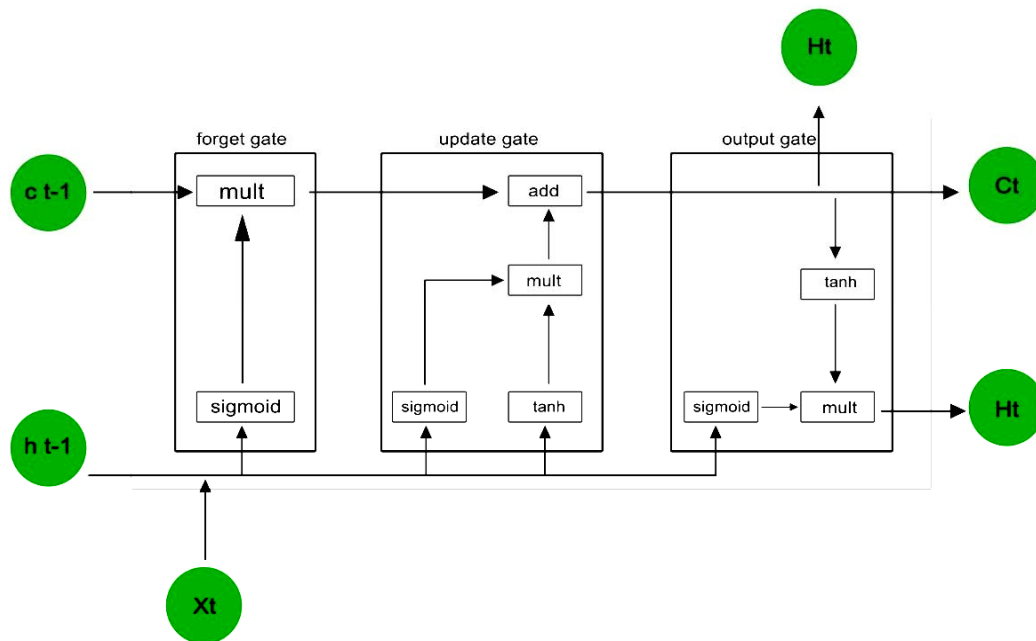


Figure.5 LSTM Architecture

4.3 Optimization of the proposed model: This section discusses the optimization process of the proposed fruit disease identification model using weed dragonfly optimization algorithm [30]. This study attempts to find the optimized data patterns, so as to assist the future datasets to be classified with more accuracy.

4.3.1 Dragonfly Algorithm (DA)

Dragonfly Algorithm [32] was first proposed by Seyedali Mirjalili in 2015. This algorithm is an inspiration of static and dynamic behaviour of dragonflies. A is an optimization algorithm used in the synaptic weights optimization in neural networks. The swarm consists of N dragonflies as search agents. The location and step vector of a search agent is X and ΔX respectively, with $1 \leq i \leq N$. The movement of a search agent is modelled after five individual behaviours in swarm [33], with explanation as follows. The DA phases [34] are shown in figure.6.

Separation Principle This principle is to avoid collision of one dragonfly with the other dragonflies close to its position.

$$S_i = - \sum X_i - X_j \quad (4)$$

Alignment Principle Here dragonflies match its velocity to other dragon flies close to its position.

$$A_i = 1/J (\sum \Delta X_j) \quad (5)$$

Cohesion Principle It is the tendency of the dragonfly moving towards the centre of the space that contains other dragonflies close to its position.

$$C_i = [(1/J) \sum X_j] - X_i \quad (6)$$

Attraction to food Principle All dragonflies move towards food sources for its survival.

$$F_i = F_{loc} - X_i \quad (7)$$

Distraction from enemies Principle This happens when all dragonflies move away from the enemies for its survival.

$$E_i = E_{loc} + X_i \quad (8)$$

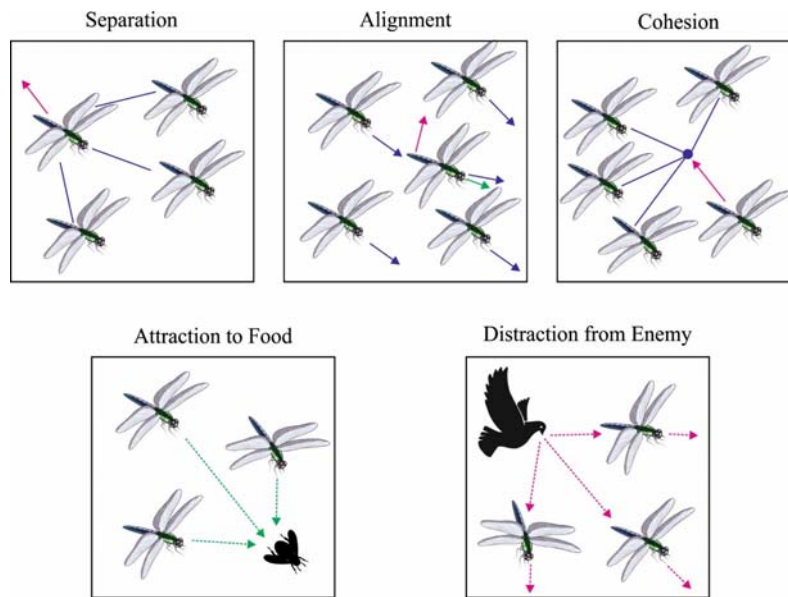


Figure 6.DA Phases

The DA algorithm is given as follows.

Algorithm Pseudocode of DA

1. Initialize size of swarm N .
2. Initialize food source fitness F_{fit} and F_{loc} .
3. Initialize enemy fitness and enemy location E_{fit} and E_{loc} .
4. Initialize number of neighbor J .
5. Initialize maximum iterations k_{max} .
6. Initialize objective function $func(x)$.
7. for search agent I to N do
 8. initialize search agent location X_i
 9. Initialize search agent step vector ΔX_i
10. end for
11. for iteration $k=1$ to k_{max} do
 12. define w, s, a, c, f, e
 13. for each search agent i do
 14. calculate fitness value $func(X_i)$
 15. if $func(X_i)$ is better than F_{fit} then
 16. $F_{fit} \leftarrow func(X_i)$
 17. $F_{loc} \leftarrow X_i$
 18. end if
 19. if $func(X_i)$ is worse than E_{fit} then
 20. $E_{fit} \leftarrow func(X_i)$
 21. $E_{loc} \leftarrow X_i$
 22. end if
 23. end for
 24. for each search agent X_i do
 25. calculate J
 26. calculate S_i, A_i, C_i, F_i, E_i using equation(4-9)
 27. If $J > 0$ then

```

28.    update  $\Delta X_i$ 
29.    update  $X_i$ 
30. else
31.    update  $X_i$ 
32. end if
33. end for
34. end for

```

5. Results and Discussions

The proposed framework is examined using Python, a high-level programming language. It provides a tremendous amount of mathematical functions for linear algebra, filtering, optimization and differential equations. The 9 features such as number of fruit spots, fruit colour, fruit shape, weight, defoliation, plant stand, leaf appearance, leaf colour, leaf spots are recorded accurately to detect the four diseases and the normal fruit. In the total dataset collected and pre-processed 80% of the data has been used as training data set and remaining 20% has been used as testing data. The hybrid CNN-LSTM Model has generated a 5-class classification accuracy of 92.14% with the number of epochs 10 and with a learning rate of 0.001. The same size training and testing data set fed into the weed dragon fly algorithm has produced an accuracy of 97.1% with the number of epochs 10 and learning rate 0.001.

5.1 Accuracy Graph

The two models have been implemented for 10 epochs to obtain the best results. The optimizer used is Weed Dragonfly that optimizes the weight, loss and cost function which resolves the multi-class classification problem. Learning rate is an important parameter that determines the optimal solutions between the local and global optima. The CNN LSTM model has classified the fruit diseases with the accuracy of 92.01% and the hybrid of CNN LSTM with dragonfly has improved the accuracy of 97.1%. The accuracy of test and training data of CNN LSTM and the optimized classifier is shown in figure 7.

Premises

Our study consists of a 5-class classification problem. We have a dataset that has five class labels anthracnose (0), bacterial blight (1), cercospora fruit spot (2), heart rot (3) and normal (4) fruit.

True Positive (TP) - No. Of samples correctly identified as given disease.

False Positive (FP) - No. Of samples incorrectly identified as a given disease.

True Negative (TN) - No. Of samples correctly identified as not the given disease.

False Negative (FN) - No. Of samples incorrectly identified as not the given disease.

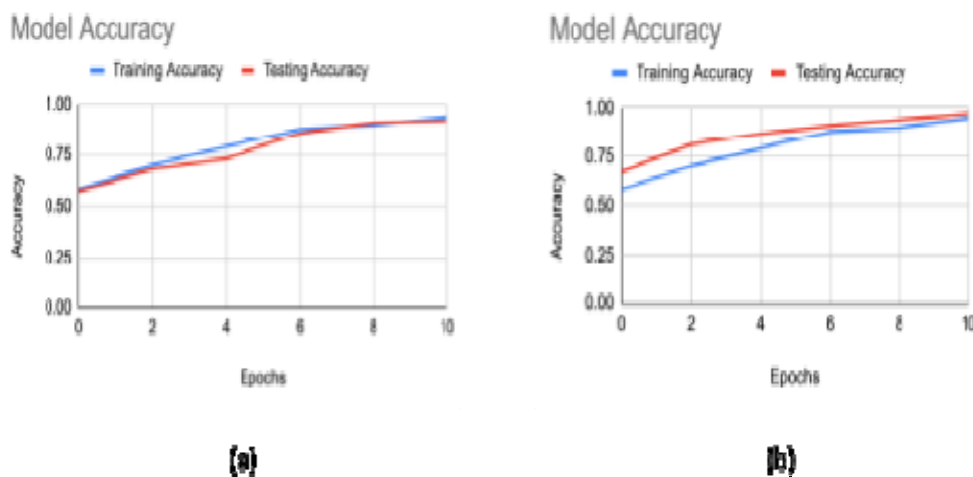


Figure 7 Accuracy Model Curve (a) Existing CNN and LSTM Model (b) Proposed hybrid CNN-LSTM

Sensitivity: It is defined as the proportion of real positive cases that are correctly predicted positive. It is given as:

$$\text{Sensitivity} = \frac{TP}{TP+FN}$$

Precision: It is defined as the proportion of predicted positives that are correctly real positives. It is given as:

$$\text{Precision} = \frac{TP}{TP + FP}$$

Detection Accuracy: It is the most institutional performance that measures the correctly predicted observation to the total observations. It defines the ability of distinguishing normal and abnormal cases. It is given as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

The confusion matrix [35] for the 5-class classification using CNN LSTM and Hybrid CNN LSTM Model is shown in the figure 8.

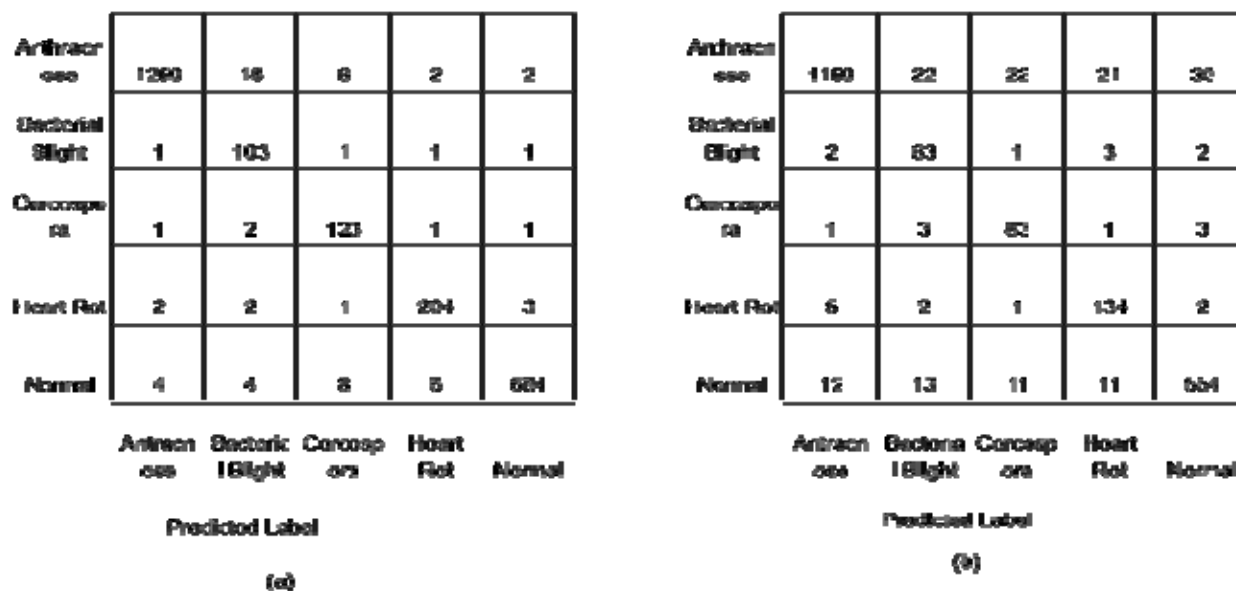


Figure 8. Disease Based Confusion Matrix (a) CNN LSTM (b) Optimised Classifier

The TP, FP, TN and FN of each disease using CNN LSTM Classifier is given in Table 2.

Table 2. TP, FP, TN, FN of CNN LSTM

Disease Name	TP	FP	TN	FN
Anthracnose	1150	95	887	20
Bacterial Blight	63	95	2041	40
Cercospora	83	8	1967	35
Heart Rot	134	10	1972	36
Normal Fruit	554	47	393	37

The TP, FP, TN and FN of each disease enhanced using Weed Dragonfly Classifier is given in Table 3

Table 3. TP, FP, TN, FN of CNN LSTM

Disease Name	TP	FP	TN	FN
Anthracnose	1250	26	837	39
Bacterial Blight	103	3	2043	3
Cercospora	123	2	2027	1
Heart Rot	204	8	1933	7
Normal Fruit	684	23	1434	21

The model loss for the existing and optimized model is shown in the figure 9.

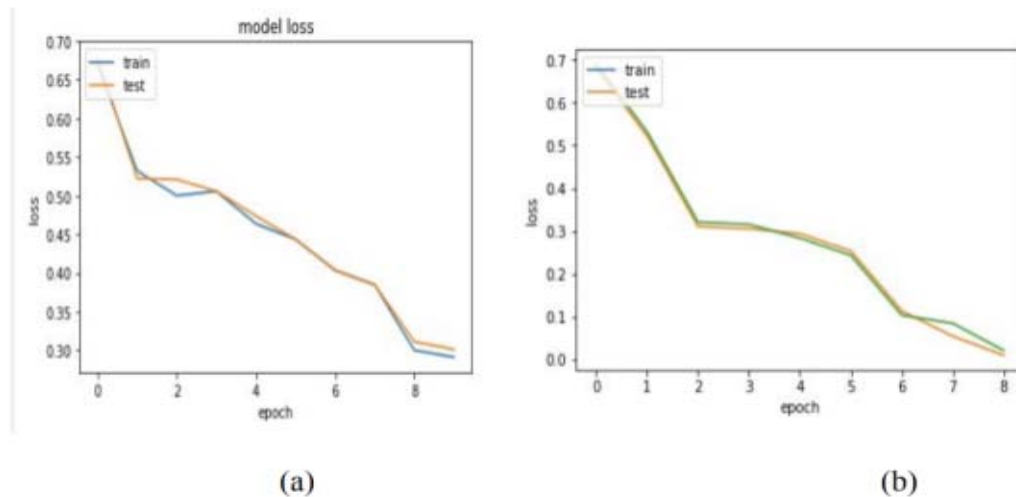


Figure 9. Model Loss (a) Existing Model (b) Optimized Model

Finally, the ROC curve for the given model is shown in fig.10.

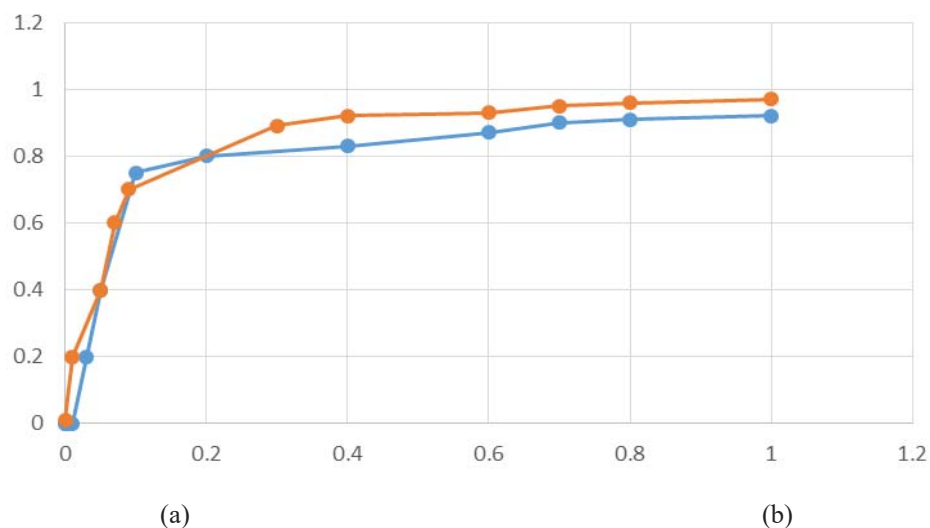


Figure 10. ROC Curve for the CNN LSTM and Optimized Model

The performance of the model is further explored by the computation of precision, recall and F-score. These values are shown in the classification report Table 4.

Table 4 Performance Evaluation of CNN LSTM classifier and Optimized Model

Disease ID	Precision		Recall		F1-Score	
	CNN-LSTM	Optimized	CNN-LSTM	Optimized	CNN-LSTM	Optimized
0	98	98	92	97	95	97
1	61	97	89	97	71	97
s2	70	98	91	99	79	98
3	79	96	93	97	85	96
4	94	97	92	97	93	97

The comparative analysis of various models in pomegranate fruit disease detection and classification is shown in Table.5.

Table: 5. Comparative Results of Various Classifier Models in Pomegranate Fruit Disease Detection

SNO	RESEARCH WORK	DISEASES DETECTED	CLASSIFIER MODEL	ACCURACY
1	Current	Anthracnose, Bacterial Blight,Cercospora,Heart Rot	Hybrid CNN LSTM	97.16
2	Manish Bhange, H A Hingoliwala	Bacterial Blight	SVM	85
3	Pooja Kantale and Shubada Thakare	Anthracnose Fruit Rot,Anthracnose,Bacterial Blight	AdaBoost Ensemble Algorithm	92.9
4	Sharath D M et al	Cercospora,Bacterial Blight,Pomegranate Borers	Neural Networks	85
5	Mrunmayee Dhakate and Ingole A. B.	Bacterial Blight,Fruit Spot,Fruit Rot	ANN Back Propagation Algorithm	90

Conclusion

In this paper, an intelligent and self-learning recognition model is designed to recognize the diseases in pomegranate fruits using hybrid LSTM-CNN. Initially, the dataset is collected by visiting pomegranate orchards around Bangalore. The input records are organized into an excel file and MapReduce technique that eliminates the irrelevant and duplicate records is used. Then, an improved segmentation approach via Principal Component Analysis and Discriminant analysis are done on the input by collecting the colour, texture and shape features of the fruits. These extracted features are then fed into the hybrid CNN- LSTM. It explores the hidden data patterns from the observed training module with the limited time steps. CNN is used for feature extraction on input data combined with LSTMs to support sequence prediction. Here, sequential steps of vector data with specified time steps have fastened the pooling layer of the CNN architecture. Experimental results have shown that the optimization of the classifier using dragonfly technique has efficiently increased the accuracy rate with the least computation time and effort.

References

- [1] Ravindar Kaur,Sonia Kaushal, "Antimicrobial and antioxidant potential of pomegranate (*Punica granatum* L.) peel",International Journal of Chemical Studies 2018; pp 3441-3449 22.
- [2] Pala, H. & Tatli, A. & Yilmaz, Cenap & Özgüven, A.I... (2009). "Important Diseases Of Pomegranate Fruit and Control Possibilities in Turkey" *Acta Horticulturae*. 285-290. 10.17660/ActaHortic.2009.818.42.
- [3] O. Obulesu, M. Mahendra and M. ThirlokReddy, "Machine Learning Techniques and Tools: A Survey," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, 2018, pp. 605-611, doi: 10.1109/ICIRCA.2018.8597302.
- [4] Manisha Bhange, H.A. Hingoli Wala,Smart Farming: Pomegranate Disease Detection Using Image Processing,Procedia Computer Science,Volume 58,2015,Pages 280-288.
- [5] Pooja Kantale , Shubhada Thakare, 2020, Pomegranate Disease Classification using Ada-Boost Ensemble Algorithm, International Journal of Engineering Research and Technology(IJERT) Volume 09, Issue 09 doi:092020
- [6] M, Sharath & Akhilesh, & G, Rohan & Kumar, S & C, Prathap. (2020). Disease Detection in Pomegranate using Image Processing. 994-999. 10.1109/ICOEI48184.2020.9142972.
- [7] Dhakate, Mrunmayee & Ingole, Atul. (2015). Diagnosis of pomegranate plant diseases using neural networks. 1-4. 10.1109/NCVPRIPG.2015.7490056.
- [8] S. Pawara, D. Nawale, K. Patil and R. Mahajan, "Early Detection of Pomegranate Disease Using Machine Learning and Internet of Things," 2018 3rd International Conference for Convergence in Technology (I2CT), Pune, 2018, pp. 1-4, doi: 10.1109/I2CT.2018.8529583.
- [9] Shiv Ram Dubey, Pushkar Dixit, Nishant Singh, Jay Prakash Gupta, " Infected Fruit Part Detection using K-Means Clustering Segmentation Technique", International Journal of Artificial Intelligence and Interactive Multimedia, Vol.2, 2013, Page(s): 65-72 .
- [10] Feiyun Zhang, "Image Recognition of Wheat Leaf Disease based on Learning Vector Quantization Neural Network and Combination Feature Parameters", Journal of Convergence Information Technology, April 2013 Vol.8 No.8, Page(s):106- 112.
- [11] Rajeshwar Dass, Priyanka, Swapna Devi, "Image Segmentation Techniques", International Journal of Electronics and Communication Technology, Vol. 3, Issue 1, Jan.-March 2012.

- [12] A. Rocha, C. Hauagge, J. Wainer, and D. Siome, "Automatic fruit and vegetable classification from images," *Computers and Electronics in Agriculture*, Elsevier, vol. 70, 2010. Page(s): 96-104.
- [13] Haiguang Wang, Guanlin Li, Zhan Hong Ma, Xiao long Li, "Application of Neural Networks to Image Recognition of Plant Diseases", 2012 International Conference on Systems and Information, 2012 IEEE, Page(s) : 2159-2164.
- [14] Dheeb Al Bashish, Malik Braik, and Sulieman Bani-Ahmad, "A Framework for Detection and classification of Plant Leaf and Stem Diseases", 2010 International Conference on Signal and Image Processing, 2010 IEEE, Page(s): 113-118.
- [15] Agrawal, G.H., Galande, S.G., Londhe, S.R.: Leaf disease detection and climatic parameter monitoring of plants using iot 4, 9927 – 9932 (2015).
- [16] Al-hiary, H., Bani-ahmad, S., Reyat, M., Braik, M., Alrahamneh, Z.: Fast and accurate detection and classification of plant diseases. *International Journal of Computer Applications* 17(1) (2011).
- [17] G.PremRishiKranth, HemaLatha, LaharikaBasava, Anjali Mathur s: Plant disease prediction using machine learning algorithms. *International Journal of Computer Applications* 18(2) (2018).
- [18] H.Sabrol, K.Satish: Tomato plant disease classification in digital images using classification tree. In: 2016 [14] International Conference on Communication and Signal Processing (ICCS), pp. 1242–1246 (2016). \
- [19] Khot.S.T, Supriya, P., Gitanjali, M., Vidya, L.: Pomegranate disease detection using image processing techniques 5(1), 2248 – 2251 (2016).
- [20] Mokhtar, U., Ali, M.A.S., Hassenian, A.E., Hefny, H.: Tomato leaves disease detection approach based on support vector machines. In: 2015 11th International Computer Engineering Conference (ICENCO), pp. 246–250 (2015).
- [21] Mwebaze, E., Owomugisha, G.: Machine learning for plant disease incidence and severity measurements from leaf images. In: 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 158–163 (2016).
- [22] Nandhini, A., Hemalatha, Radha, Indumathi: Web enabled plant disease detection system for agricultural applications using wmsn. *Wireless Personal Communications* 102(2), 725–740 (2018).
- [23] A S M Farhan Al Haque, Rubaiya Hafiz, Md. Azizul Hakim, G. M. Rasiqul Islam. A Computer Vision System for Guava Disease Detection and Recommend Curative Solution Using Deep Learning Approach. 22nd International Conference on Computer and Information Technology (ICCIT), 2019.
- [24] Doh, Benjamin & Zhang, Duo & Shen, Yue & Hussain, Fida & Doh, Ronky & Ayepah, Kwaku. (2019). Automatic Citrus Fruit Disease Detection by Phenotyping Using Machine Learning. 1-5. 10.23919/IconAC.2019.8895102.
- [25] L.Palmerini, S. Mellone, L. Rocchi and L. Chiari, "Dimensionality reduction for the quantitative evaluation of a smartphone-based Timed Up and Go test," 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Boston, MA, 2011, pp. 7179-7182, doi: 10.1109/IEMBS.2011.6091814.
- [26] M. R. Wasef and N. Rafla, "HLS Implementation of Linear Discriminant Analysis Classifier," 2020 IEEE International Symposium on Circuits and Systems (ISCAS), Sevilla, 2020, pp. 1–4, doi: 10.1109/ISCAS45731.2020.9181270.
- [27] J. S. Manjaly and T. Subbulakshmi, "Various approaches to improve MapReduce performance in Hadoop," 2018 3rd International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2018, pp. 778-782, doi: 10.1109/ICICT43934.2018.9034419.
- [28] T. Liu, T. Wu, M. Wang, M. Fu, J. Kang and H. Zhang, "Recurrent Neural Networks based on LSTM for Predicting Geomagnetic Field," 2018 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES), Bali, 2018, pp. 1-5, doi: 10.1109/ICARES.2018.8547087.
- [29] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," 2017 International Conference on Engineering and Technology (ICET), Antalya, 2017, pp. 1-6, doi: 10.1109/ICEngTechnol.2017.8308186.
- [30] T. Liu, T. Wu, M. Wang, M. Fu, J. Kang and H. Zhang, "Recurrent Neural Networks based on LSTM for Predicting Geomagnetic Field," 2018 IEEE International Conference on Aerospace Electronics and Remote Sensing Technology (ICARES), Bali, 2018, pp. 1-5, doi: 10.1109/ICARES.2018.8547087.
- [31] Yu Wang, "A new concept using LSTM Neural Networks for dynamic system identification," 2017 American Control Conference (ACC), Seattle, WA, 2017, pp. 5324-5329, doi: 10.23919/ACC.2017.7963782.
- [32] J. Zhu, H. Chen and W. Ye, "A Hybrid CNN–LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar," in *IEEE Access*, vol. 8, pp. 24713-24720, 2020, doi: 10.1109/ACCESS.2020.2971064.
- [33] Mehrabian, A.R. & Lucas, Cameron. (2006). A novel numerical optimization algorithm inspired from weed colonization. *Ecological Informatics*. 1. 355-366. 10.1016/j.ecoinf.2006.07.003.
- [34] M. M. Mafarja, D. Eleyan, I. Jaber, A. Hammouri and S. Mirjalili, "Binary Dragonfly Algorithm for Feature Selection," 2017 International Conference on New Trends in Computing Sciences (ICTCS), Amman, 2017, pp. 12-17, doi: 10.1109/ICTCS.2017.43.
- [35] JOUR Lawrynczuk, Maciej, Rahman, Chnoor M., Rashid, Tarik A.2019/12/06 Dragonfly Algorithm and Its Applications in Applied Science Survey 1687-5265 <https://doi.org/10.1155/2019/9293617> 10.1155/2019/9293617
- [36] B. P. Salmon, W. Kleynhans, C. P. Schwegmann and J. C. Olivier, "Proper comparison among methods using a confusion matrix," 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, 2015, pp. 3057-3060, doi: 10.1109/IGARSS.2015.7326461.