International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 08 | Aug 2022 www.irjet.net p-ISSN: 2395-0072

FRUIT DISEASE DETECTION AND CLASSIFICATION USING ARTIFICIAL INTELLIGENCE

Shaik Naziya [1], Dr. C Sivaraj Ph.D.[2]

[1]*Student, Department of Computer Applications, Madanapalle institute of technology and science, India*
[2]*Asst. Professor, Department of Computer Applications, Madanapalle institute of technology and science, India*

Abstract

In the field of agriculture Fruits are affected by different climatic conditions. For the farmers it is very difficult to identify the fruits that are having disease or not. This will affect the global economy. By manual testing of fruits does not give the accurate output. In this method the fruit image is take to analyse the disease. The disease must be identified using colour, shape and size of the fruit. It will give the accurate result for farmers than the manual analysis of the fruits. Banana, apple and orange fruits are taken in this method. The system is already trained with different types of diseases that fruits are having. Input image given by the user undergoes with different kinds pre-processing steps. First the fruit image is resized in the required manner and the type of the fruit identified by the system. The fruits related to input image in the trained dataset must be analysed. And then features such as colour, shape and size are extracted from the image and image cluster is identified using k-means clustering algorithm. Next convolutional neural network(CNN) is used to find the fruit is infected or not infected. Experimental evaluation of the proposed approach is effective and 95% accurate to identify banana, apple and orange fruits disease.

Keywords: Features: Colour, Size, Shape; K-Means clustering, convolutional neural network.

I. Introduction

India is the green land by producing 44.04 million tons of fruit and it is a second-greatest producer of fruits. India contributes 10% to the world's fruit creation. Mostly Indians are dependent on agriculture. So, the constantly changing climatic conditions and different disease have a high impact on crops and they are leading to less crop yield. And India stands second in the list of highly populated countries and it is still increasing. on account of that food consumption will automatically increase. India produces food crops and also fruits. Analyzing the good and bad fruits is necessary.

In this project I have taken banana, apple and orange fruits into consideration. The following fig.1, fig.2 and fig.3 represents fruits that are affected by diseases.

Fig.1: Banana diseased fruit

Fig.2: Apple diseased fruit

Fig.3: Orange diseased fruit

In India, the livelihood of 58% of the Indian population is based on agriculture. So, the constantly changing climatic conditions and also some diseases have a high effect on crops and are leading to less crop yield. And India stands second in the list of highly populated countries and it is still increasing. On account of that, food consumption will automatically increase and this will lead us to the situation where people have to produce more food. India not only produces and exports food crops but also fruits. Here the classification of good and bad fruits is barely done manually in most of the places. This leads to more errors in the grading of fruits while exporting. So, to overcome the faults that happen during the manual classification, researchers have proposed an image detection method to classify the diseased fruits from good fruits to improve the quality of classification while exporting fruits. Here this approach is using CNN (Convolutional Neural Networks) which detects the quality of the fruit, layer by layer.

II. Literature Survey

Recently, Pantech solution [1] have done the research on fruit disease detection by using Convolutional neural network. This is used to detect the

fruit is having disease or not. Apple fruit is taken by the author for performing the experiment. For this model, the accuracy will be 97%. Since they are using only CNN algorithm the data takes more time to produce the output.

Author shiv Ram Dubey [2] considers the fruit is apple. The algorithm used by the author is image processing technique for fruit disease detection and identification of fruit disease. Some of The disease taken by the author is apple rot, apple blotch for execution. The author also used the K-means Clustering algorithm for clustering the dataset according to their disease. For disease detection support vector machine is used.

The author laria Pertot [3] suggested the multilingual web based tool for plant disease detection. In this the author taken the strawberry fruit. In this the user needs to provide the dimension, features and colour of the image as input. By using it, the software will analyse the disease of the fruit and It will be displayed to user. This software is not suitable if the user/farmer is illiterate.

Monica Jhuria [4] uses the technology called image processing and back propagation. grapes, apples and mangoes are selected for conducting the experiment. For disease detection and weight calculation of fruit, image processing techniques are used and for weight adjustment of images that are stored in learning database.

Author Manisha Bhange [5] uses the technologies called image processing, support vector machine and K-means algorithm for fruit disease detection. The author taken the pomegranate fruit. Feature Colour, morphology, color coherence vectors are considered for feature extraction. SVM is used for classification to classify the image is infected or non-infected. CCV and clustering is done using k-means algorithm. This software gives the 82% accurate to identify the pomegranate disease.

Author Dharmasiri, S.B.D.H [6] has done the research on passion fruit detection. The detection of fruit disease contains the main steps. they are, image acquisition, image preprocessing, image segmentation, feature extraction, dataset preparation, training and testing. They used two types of passion fruit diseases namely passion fruit scab and woodiness images, for this approach. K-means clustering is used for segmentation. passion fruit diseases can be identified in the average accuracy of 79% and its' stage can be identified in average accuracy 66%.

III.Pre-Requisites:

1. Convolutional Neural Network(CNN):

CNN are widely used in computer vision and have become the state of the art for many visual applications such as image classification, and have also found success in natural language processing for text classification.

CNN are very good at picking up on patterns in the input image, such as lines, gradients, circles, or even eyes and faces. Convolutional neural networks can operate directly on a raw image and do not need any pre-processing.

CNN is a feed forward neural network. CNN contains up to 20-30 layers. Convolutional layer is special kind of layer which makes the CNN more powerful. The key building block in a convolutional neural network is the convolutional layer. We can visualize a convolutional layer as many small square templates, called convolutional kernels, which slide over the image and look for patterns. Where that part of the image matches the kernel's pattern, the kernel returns a large positive value, and when there is no match, the kernel returns zero or a smaller value. CNN has three main types of layers. They are:

1. Convolutional layer
2. Pooling layer
3. Fully-connected layer

Fig.4 represents the architecture of Convolutional neural network.

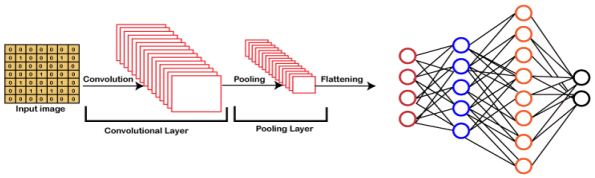

Fig.4 Architecture of CNN

i.Convolutional Layer:

. The key building block in a convolutional neural network is the convolutional layer. The Input image is a fruit, which is made up of a matrix of pixels in 3D. It means the image contains dimensions' height, width and depth. We also have feature detector it moves across the respective fields of the image, checking the feature is present.

ii. Pooling Layer:

Pooling layer is used for dimensionality reduction and for reducing the number parameters in the input image. This layer filters across the entire input. In this an input fruit image is filtered to get the accurate results. There are two types pooling:

i. Max pooling
ii. Average pooling

iii. Fully-Connected Layer:

The name of the full-connected layer aptly describes itself. The pixel values of the input image are not directly connected to the output layer in partially connected layers. However, in the fully-connected layer, each node in the output layer connects directly to a node in the previous layer. This layer performs the task of classification based on the features extracted through the previous layers and their different filters.

2. K-Means Clustering:

K-means clustering is unsupervised learning algorithm. K-means clustering technique is used for partitioning the training dataset according to their features. When we deal with larger dataset, K-means clustering algorithm gives greater efficiency. Fruit dataset is partitioned into fruits based on the type and also based on the fruit is having disease or not. This will increase the efficiency.

Defining K value using Elbow method:

The Elbow method is one of the most popular ways to find the optimal number of clusters. This method uses the concept of WCSS value. **WCSS** stands for **Within Cluster Sum of Squares**, which defines the total variations within a cluster. The formula to calculate the value of WCSS (for 3 clusters) is given below:

$$\mathbf{WCSS= \sum_{Pi\ in\ Cluster1} distance\ (P_i\ C_1)^2 + \sum_{Pi\ in\ Cluster2} distance\ (P_i\ C_2)^2 + \sum_{Pi\ in\ CLuster3} distance\ (P_i\ C_3)^2}$$

Where,
$\sum_{Pi\ in\ Cluster1}$ distance$(P_i\ C_1)^2$: It is the sum of the square of the distances between each data point and its centroid within a cluster1 and the same for the other two terms.

IV. Proposed Algorithm:

In this proposed method, CNN, image processing technique and K-means clustering techniques are used for fruit disease detection. Apple, orange and banana are considered and diseases are considered as fruit name and fruit is having disease or not. Fruit size, colour and shape feature vectors are chosen for feature extraction. CNN is used for classification of images. CNN gives better performance on real time data. The accuracy by using the CNN is more than 90%.

Image processing technique is used for disease detection and weight calculation of fruit. Feature extraction and image segmentation is done by using image processing technique. Re-sizing of image is also done by using image processing technique. K-Means clustering is used for partitioning the given dataset according to their features. Based on Fruit type and Disease, the training dataset is categorized into different clusters. This technique saves the time while giving the output to user.

Image Pre-processing:

we will start with importing the libraries, data preprocessing followed by building a CNN, training the CNN. Tensorflow library is used to train with large amount of data. By using this library, it is easy to build neural network.

```
from tensorflow.keras.utils  import to_categorical
import tensorflow as tf
import keras
from keras.layers import Dense,Dropout, Conv2D,MaxPo
oling2D , Activation, Flatten, BatchNormalization, Separa
bleConv2D

from keras.models import Sequential
```

Activation Function:

Relu:

The ReLU function is the Rectified linear unit. It is the most widely used activation function. It is defined as:

$$f(x)=\max(0, x)$$

The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time. In ReLU function if the input is negative it will convert it into zero and the neuron does not get activated.

Sigmoid:

Sigmoid function range from 0-1 and having "S" shape. It is widely used activation function. Sigmoid function is defined as follows:

$$\frac{1}{1+e^{-x}}$$

Convolutional Layer:

Convolutional layer is the first layer in CNN. It can be called by using add class. Convolutional layer contains the parameters. They are, filter, activation function(Relu), kernel size and input shape. It can be defined in python as shown below:

```
model.add(Conv2D(32, (3, 3), kernel_initializer='he_uniform', padding='same', activation='relu', input_shape=(100,100,3)))
```

Pooling Layer:

Pooling layer is next layer after the convolutional layer. It uses the Max Pooling. To add pooling layer in python we call the add method followed by two parameters, called pass pool size and strides.

```
model. add(MaxPooling2D((2, 2)))
```

Flattering:

We go with flattering with the result of convolution and pooling layer into one dimensional vector. The result of flattering is the input for fully connected layer. In python we use the flattering is as follows.:

```
model.add(Flatten())
```

Fully Connected Layer:

Now we are adding the fully connected layer to flatten layer. this layer should be added by using the add class. It contains the Dense layer followed by units, it defines the number of hidden neurons we want to have in fully connected layer and activation function(Relu) parameter.

```
model.add(Dense(128, activation='relu', kernel_initializer='he_uniform'))
```

Output layer:

Output layer is the final layer, it predicts the final output. output layer contains the dense class followed by the parameters units and activation function(sigmoid).

```
model.add(Dense(1, activation='sigmoid'))
```

V. Architecture:

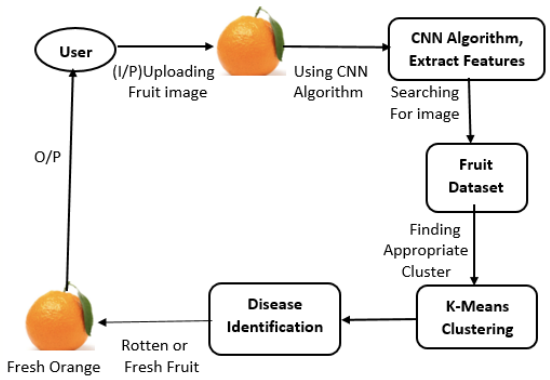

Fig.5: Architecture of Fruit Disease Detection

The architecture of Fruit disease detection displays the how image is processed using algorithms and how user get the accurate output.

Step by step process in this project is as shown below:

Step-1: Uploading the image of the fruit by user.

Step-2: Initially, by using CNN Algorithm the image is resized according to the requirements and then Features are extracted from the image.

Step-3: Features like fruit name, shape, size and colour of the image is extracted. And then the image is added to the dataset for the future analysis.

Step-4: By using K-Means clustering the image is mapped to the related clusters by using feature extracted and then the disease will be identified.

Step-5: the added cluster defines the status of the fruit also the disease is identified that fruit is rotten of fresh.

Step-6: Finally, the status of the fruit is identified and it passed to user as output.

VI. Experimental Analysis and Result:

The experiments are conducted using Google Colab. The dataset which is used for this analysis called “Fruits-Fresh-and-rotten-for-classification” is taken from kaggle Website. The dataset should be directly uploaded from kaggle website to Google Colab notebook. The dataset contains Two types of data. They are training data and testing data. The training and testing data contains the number of images as shown in Fig.6 and Fig.7.

Fruit image category	No.of images
Fresh banana	1581
Rotten Banana	2224
Fresh Apple	1693
Rotten Apple	2342
Fresh Orange	1466
Rotten Orange	1595

Fig.6. Training dataset

Fruit image category	No.of images
Fresh banana	381
Rotten Banana	530
Fresh Apple	395
Rotten Apple	601
Fresh Orange	388
Rotten Orange	403

Fig.7. Testing dataset

1.Dataset Dimensions:

The dataset contains the 3D (three dimension) image. Since it calculates the length, depth and width of the image. The input image is divided into small pixels and for each pixel the disease should be calculated using CNN algorithm. The pixels (part of image) 2D images. After that the output of pixel images is converted into matrix form as 0's and 1's. If the disease is present in the pixel the value will be one and otherwise the value will be 0.

2. Result Discussion:

Fig.8.1. represents the training loss and validation loss in the dataset. In x-axis the epochs and in y-axis the loss will be defined. In the below diagram the blue line defines the training loss.in this the less data is lossed. the orange colour line defines the validation loss it means the image is not clear. Much data is loosed in this phase.

Fig.8.2 Represents the training accuracy and testing accuracy. The blue line represents the training accuracy and the accuracy is more in this phase. The orange line represents the validation accuracy and accuracy for validation is satisfied.

Fig. 9. Shows the input images that are affected by disease or not.in this the input image is randomly taken and the disease must be identified. Below diagram representing, in testing phase the fruit are having disease or not and fruit name will be displayed as output. while executing different fruits are taken using random forest library.

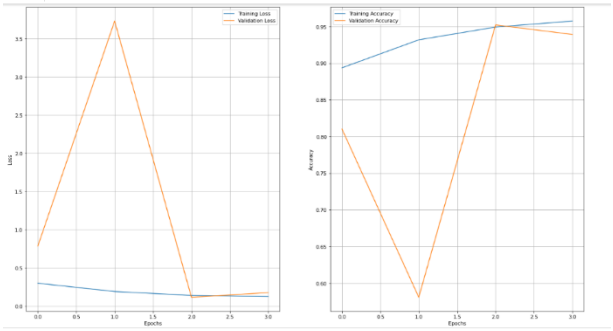

Fig.8.1: Training loss and validation Loss

Fig.8.2: Training accuracy and validation accuracy

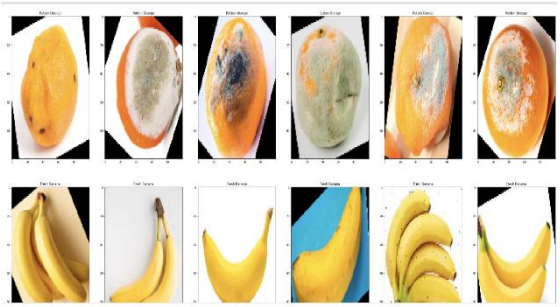

Fig.9: Classification of Data

VII. Conclusion:

Fruit disease detection and classification could be a little tedious process because it is very hard to recognize characteristics of fruits. The characteristics could vary between different fruits. These kinds of problems can be handled by feature vector. Before extracting feature, image pre-processing is done on every fruit. CNN model is constructed for CSV file which was created along with the fruits. Finally, the count of the fresh fruits and the rotten fruits is evaluated. The above said method was classifying fruits Fresh or Rotten. The proposed model has gone through pre-processing stage, feature generation stage and classifiers learning stage. The statically evaluation of the above proposed model is done in terms of precision, recall and accuracy.

In this the fruit image is taken as input. By using Convolutional neural network fruit name and features called colour, shape and size will be extracted.

Then the output will be passed to K-means clustering, based on the features extracted the fruit is mapped to different clusters. And then the output will be given as the input fruit is rotten or fresh fruit.

References:

1. Milos Ilic, Sinisa Ilic [1], Srdjan Jovic, Stefan Panic," Early fruit pathogen disease detection based on data mining prediction", Computers and Electronics in Agriculture © 2018.

2. Dubey, Shiv Ram [2], and Anand Singh Jalal. "Detection and classification of apple fruit diseases using complete local binary patterns." *2012 Third International Conference on Computer and Communication Technology*. IEEE, 2012.

3. Ilaria Pertot [3], Tsvi Kuflik, Igor Gordon, Stanley Freeman, Yigal Elad, Identificator: A web-based tool for visual plant disease identification, a proof of concept with a case study on strawberry, Computers and Electronics in Agriculture, Elsevier, 2012, Vol.88, p.144-154.

4. Monika Jhuria [5], Ashwani Kumar, Rushikesh Borse, Image Processing For Smart Farming: Detection of Disease and Fruit Grading, IEEE Proceedings of the 2013 IEEE Second International Conference on Image Information Processing, 2013, p.521-526.

5. Bhange, Manisha [6], and H. A. Hingoliwala. "Smart farming: Pomegranate disease detection using image processing." *Procedia computer science* 58 (2015): 280-288.

6. Dharmasiri [7], S. B. D. H., and S. Jayalal. "Passion Fruit Disease Detection using Image Processing." *2019 International Research Conference on Smart Computing and Systems Engineering (SCSE)*. IEEE, 2019.

7. Md. Rasel Howlader, Umme Habiba, Rahat Hossain Faisal and Md. Mostafijur Rahman," Automatic Recognition of Guava Fruit Diseases using Deep Convolution Neural Network", 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), 7-9 February, 2019.

8. Youssef Es-saady, Ismail El Massi, Mostafa El Y assa, Driss Mammass, Abdeslam Benazoun," Automatic recognition of fruit diseases based on serial combination of two SVM classifiers", 2nd International Conference on Electrical and Information Technologies ICEIT'2016.

9. Anne-Katrin Mahlein & Erich-Christian Oerke & Ulrike Steiner & HeinzWilhelm Dehn," Recent advances in sensing fruit diseases for precision crop protection", Eur J Plant Pathol, 2012.

10. Dr. N. Sasirekha, N. Swetha," An Identification of Variety of fruit Diseases Using Various Data Mining Techniques", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 10, October 2015.

11. Samiksha Bhor, Shubha Kotian, Aishwarya Shetty, Prashant Sawant," Developing an Agricultural Portal for Fruit Disease Prediction", IJIRST – International Journal for Innovative Research in Science & Technology, Vol 3, Issue 11, April 2017.

12. Dr. Kamaljit Kaur, Manpreet Kaur," Prediction of Fruit Disease from Weather Forecasting using Data Mining", International Journal on Future Revolution in Computer Science & Communication Engineering April 2018.