Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.2, June : 2024

DEEP LEARNING BASED DISEASE DETECTION IN ORANGE FRUIT

Mr. Jyotsana Goyal, Research Scholar, Dept. of Computer Science and Engineering, SAGE University Indore.
Dr. Rajat Bhandari, Professor, Dept. of Computer Science and Engineering, SAGE University Indore.

ABSTRACT
One of the main factors contributing to output and financial losses in agricultural businesses worldwide is fruit illness. Numerous natural factors, such as a deficiency of plant sources specific to the condition, Plant growth is currently being hampered by a lack of sunlight, unfavourable plant weather, and a lack of knowledge about proper pesticide application techniques. Fruit wilt is most frequently caused by diseases. Through early detection, image processing can help stop the spread of fruit illnesses. This study compares the performance of many machine learning algorithms, including Support Vector Machine (SVM), Convolutional Neural Network (CNN), and KNN (K-Nearest Neighbours), in order to forecast six orange fruit diseases and determine which model is most appropriate. Based on the results, it can be concluded that the CNN methodology performs 15% and 25% more accurately than the SVM and KNN methods, respectively. The CNN model can be used to rapidly identify fruit diseases and the procedure can be automated because of its 90% overall accuracy.

Keywords*:*
Deep Learning, Orange disease, K-Nearest Neighbors, Support Vector Machine, Convolution Neural Network

I. INTRODUCTION
In India More than 70% of people are dependent on the production of agricultural crops, plants, and fruits. India is primarily an agricultural country. Fruit illnesses affect the environment, which makes them a serious problem for humans. Fruit infections have the potential to significantly reduce fruit yield and quality. Algorithms and technology are important in many industries, but traditional agricultural methods are still used in India. Machine learning methods like CNN, SVM, and KNN are applied to predict diseases in oranges. Precision, specificity, recall, sensitivity, and f1 score are compared to determine the best model. In order to control the spread of infections, early detection and detection of diseases are essential.

A. DISEASES
Bacteria, fungi, insects, and pests pose a threat to the health of orange fruit. Numerous diseases have the capacity to damage both the oranges and the farmers' ability to cultivate them.

- **Scab**

The lesion on the fruits, as depicted in Figure 1, is caused by the fungus Elsinoe australis, commonly known as Scab (Elsino faucetti). This disease spreads through water, overhead irrigation, and splashing. To prevent it, spraying copper hydroxide is recommended.

Figure 2. Image of Scab

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.2, June : 2024

- **Melanose**

Melanose, caused by Diaporthe citri as depicted in Figure 2, is a significant affliction affecting orange fruit. The disease manifests as a small spot on the fruit's surface, making it easily recognizable. Application of fungicides is an effective measure to prevent its occurrence.

Figure 2. Image of Melanose

- **Green Mold**

Penicillium digitatum, or green mold, is a fungus that can cause up to 90% of orange fruit to develop this disease (Figure 3). Cleaning the fruit with a cloth dipped in vinegar or hot water can stop it

Figure 3. Image of Green Mold

- **Citrus Canker**

A bacterial disease called citrus canker, depicted in Figure 4, damages the leaves and fruits of citrus plants. The illness affects the health of citrus trees that are afflicted, even though it poses little threat to humans. There is no recognized therapy for citrus canker. Avoiding citrus canker is the best course of action. Canker causes a reduction in the health and fruit production of citrus trees, eventually leading to no fruit at all.

Figure 4. Image of Citrus canker

II. LITERATURE SURVEY

The literature review for the proposed research on early disease detection in orange trees using deep learning techniques encompasses a comprehensive examination of existing studies and developments related to citrus disease detection, deep learning models, and optimization methods. Below is a summarized literature review:

Huanglongbing (HLB) is one of the most threatening diseases for citrus production and it has caused significant economic damage worldwide. Hence, computer-vision systems that are based on convolutional neural networks (CNNs) can detect HLB accurately. Moreover, the detection system should be able to discriminate between HLB and other citrus abnormalities to ensure that any treatments are effective. Besides, the causal pathogen of HLB is usually detected and diagnosed by the quantitative real-time polymerase chain reaction (qPCR) test, which is costly. Consequently, it is difficult to collect large datasets to train CNN-based systems. In this case, transfer learning from pre-trained CNNs is a solution for building an HLB-detection system using small-sized datasets. This paper evaluates two kinds of CNN architectures: series network (represented by AlexNet, VGG16, and VGG19 models) and directed acyclic graph (DAG) network (represented by ResNet18, GoogLeNet, and Inception-V3 models). These pre-trained CNNs are fine-tuned to distinguish HLB, healthy cases, and 10 kinds of abnormalities of the Citrus sinensis species, which is commonly known as sweet orange. The dataset includes 953 color images, where the leaf samples were collected from orange groves in north Mexico. The 10-fold cross-validation results show that all the CNNs present a 95% or higher HLB sensitivity. However, the number of trainable parameters impacts HLB detection more than the network's depth. Specifically, VGG19, with 19 layers and 144 M parameters, reached a perfect sensitivity for all cross-validation experiments; whereas Inception-V3, with 48 layers and 24 M parameters, reached 95% sensitivity to HLB detection. This outcome happens because a higher number of parameters compensates for the limited number of HLB cases, so VGG19 can successfully transfer the learned characteristics to new cases. This study gives guidance when choosing an adequate CNN to efficiently detect HLB and other orange abnormal[10].

India is a country that is highly dependent upon agriculture but the sad truth is many plants are being wasted due to an inexpert approach towards their growth. The various diseases in plants lead to their insufficient growth. Thus, it is very important to protect the plants at the right time so that the farmers do not face losses. Farmers usually analyze the damaged yield very late, thus failing to prevent it. They rush to domain experts which results in delay as they are not present in all regions. Thus a tool needs to be generated approaching to smart farming which will help farmers protect their crops and reduce losses with less human efforts. This can be made using various models of machine learning. In this paper, a review of various existing approaches is done in order to find which algorithm works better and provides higher accuracy. This paper consists of research in disease detection of various plants and vegetables such as potato, tomato, bell pepper, apple, grapes, cherry, strawberry, orange etc[11].

In order to solve the current fruit surface disease detection algorithm's problems of low accuracy, slow speed and heavy workload of quality classification, this paper takes apple, peach, orange, and pear as the research objects and proposes a model based on Mask R-CNN for detecting disease spots on the surface of fruits which accurately detects the defects on the surface of the fruit after the picking robot recognizes and locates the fruit. By adding a bottom-up horizontal connection path, the feature pyramid (FPN) structure of Mask R-CNN is improved to enhance the fusion of high and low-level features. Experimental research shows that the improved Mask R-CNN algorithm has a detection accuracy of more than 95% for the four kinds of fruit surface lesions, and the detection speed reaches 2.6 frames per second when using GPU, which is significantly better than Fast R-CNN and SSD algorithms and has good detection performance and robust ness[12].

Several deep learning-based object detection techniques in medical imaging have been proposed. Chest X-rays are widely used for detecting thorax diseases due to the convenience and low radiation dose compared to Computed Tomography (CT). However, the research on rib fracture detection in chest X-rays is still inadequate. Most of the research primarily focused on frontal CXR and some on lateral CXR. No study of rib fracture detection on oblique view CXR has been previously proposed. Due to

the overlapping characteristic of human ribs, the oblique view can help radiologists to recognize the fractured ribs that are blocked in the frontal view. In this paper, we employed a YOLOv5 model along with the techniques of data augmentation and image enhancement for rib fracture detection. We trained and evaluated on E-DA dataset, a private dataset collected from E-DA Cancer Hospital containing frontal and oblique chest X-rays. The developed model can detect fractured ribs in both projection views of CXR[13].
Numerous diseases, including lemon canker, melanose, orange black spot, etc., harm the citrus fruit plants. Citrus canker is a contagious ailment of citrus trees that causes yellow halo lesions and scabs on the leaves or fruits. Extreme diseases have the potential to kill a person as well as damage fruit. The cancer bacteria can easily and quickly spread to plants, birds, and humans through cloths and other contaminated items. The proposed method uses effective CNN models to recognize and categories citrus illnesses from citrus images. The automatic detection and early diagnosis of illnesses in citrus fruits for enhancing growth and quality are greatly facilitated by computational methodologies. Pre-processing, image segmentation, features extraction, and classification are the image processing methods utilized to create a diagnosis system for citrus plant diseases. In most cases, users collected images of citrus fruit and leaves from the field using sensors and cameras. The pre-processing procedures improve the images by removing noise from the ones that are captured. Analytically, CNN separates features from raw inputs. For categorization, the characteristics with highest likelihood value are picked. The sick regions' features are extracted using feature extraction techniques. The categorization approaches are then used to identify the diseases of citrus fruit[14].
Food contains essential nutrients for human beings to grow and develop. Out of so many type of food, vegetables and fruits are important for humans' daily healthy diet as they provide all the nutrients that helps human to prevent diseases. However, fruit will get rotten easily if not store properly due to the spread of bacteria. Therefore, it is important for food industry to perform inspection on fruits before selling to the consumers. The problem encountered in the human inspection is lower in consistency and accuracy as the manual inspection by humans' eye will consume time and energy. To solve this problem, the proposed method is to apply the deep learning technique which is Convolutional Neural Networks (CNNs) for feature extraction and classification of rotten fruits. The types of fruits that will be detected and classified in this paper are banana, apple and orange. The validation accuracy obtained in this paper is 98.89%. The total duration of training stage is 212.13 minutes. Hence, the required time to classify single fruit image is approximately 0.2 second [15].
In the existing literature, several studies have addressed the detection and prevention of diseases in citrus fruits, particularly focusing on citrus canker and other abnormalities. However, there is still a research gap in achieving highly accurate and efficient detection of various citrus diseases with limited datasets. Many existing approaches heavily rely on manual data collection and lack a comprehensive evaluation of different deep learning architectures for citrus disease classification. Furthermore, while some research has explored disease detection in citrus leaves, there is limited exploration of detecting diseases directly on the fruit's surface, which is crucial for early disease identification.

III. RELATED WORK

The use of digital image processing to identify the six main illnesses that affect orange fruit has been extensively studied [1]. It requires more skill and is more difficult to visually identify diseases on leaves. Research on automated plant disease identification is highly sought after due to its advantages in the world of agriculture. The authors of [2] suggested utilizing SVM classifiers for disease detection in grape leaves. The objective of this study is to identify and classify grape leaf diseases using the SVM classification algorithm. The sick region is first identified using K-means clustering segmentation, and then its color and texture information are retrieved. Finally, the type of leaf disease is determined using a categorization technique. With an accuracy of 88.89%, the model can correctly identify and categorize the tested disease. SVM classifier was employed by the authors [3] to detect orange fruit illness. This work presents an image preparation-based method for detecting and evaluating bacterial spot defects and orange size. It is likely to be applied in areas where orange fruit quality is assessed and determined. This paper discusses the benefits and drawbacks of several features and categorization systems. In the paper [4], fuzzy logic is used to compute the orange disease severity with success. Multi-class SVM with k-means clustering is introduced for disease classification with a 90% accuracy rate. In this paper, a machine learning system that can recognize orange flaws and categorize them is presented. The paper [5] provided AI (artificial intelligence) algorithms for the automatic detection and classification of sugarcane leaf diseases using image processing techniques. This is achieved by extracting the leaf area from the damaged area and utilizing the K-means clustering approach to retrieve the diseased fraction of a leaf. Lastly, diseases are categorized using the GLCM (Gray-Level Co-occurrence Matrix) textural feature of the afflicted area. An effective k-mean clustering method for identifying infected leaf regions is suggested by the authors of [6]. To partition the diseased zone into segments and assign each section to the relevant group, a color-based segmentation approach is applied. It shows what proportion of the leaf is affected. In [7] authors have used RSC (Relative Sub-Images Coefficients) features and SVM Classifier to determine automated leaf recognition system. This method is very useful in forest research, biology researchers and agricultural needs to study and identify various biological or botanical gardens. It's used to find medicinal plants so that medicines can be made. It has a lower time complexity and is easier to execute with higher efficiency.

IV. METHODOLOGY

The project's main objective is to investigate CNN, SVM, and KNN for orange fruit disease detection and compare the outcomes. Obtaining the photos from the dataset will be the first step in the detection of illnesses in orange fruit. After that, in order to train, the photos in each category will be loaded and resized. [8] Duplicate data is eliminated after resizing since it could influence the model's performance. To train the model, SVM, CNN, and KNN are all utilized. To train the SVM model, several values of c, gamma, and kernel are employed. Multiple layers, including convolution, pooling, and fully linked, are used to teach CNN. The KNN model is trained using the appropriate value for K. The model's performance metrics include f1-score, accuracy, sensitivity, precision, recall, and specificity. It classifies illnesses, including green mold, melanose, scab, thrips, citrus canker, and leaf miner. The flowchart in Fig. 5 shows how each image in the collection is resized to a specific size. Using extracted features, the KNN, SVM (which employs SVC-Support Vector Classifier), and CNN [9] algorithms are used to train the model on the preprocessed images. An algorithm looks at these characteristics and studies them to categorize the diseases. The needed image is then provided as input in jpeg or png format, and these steps are followed to use the chosen algorithm to forecast the diseases

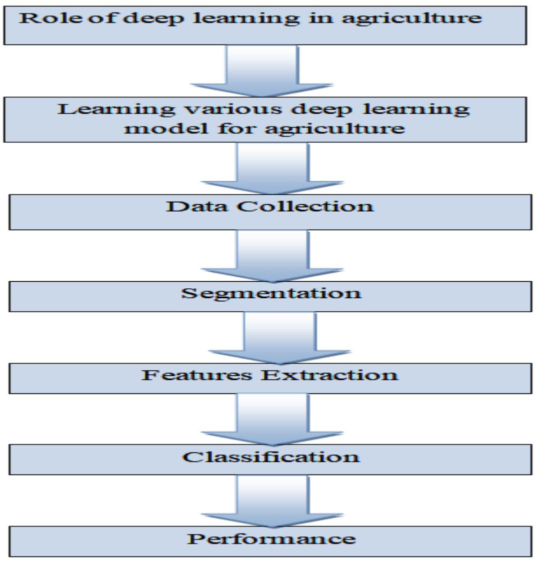

Figure 5. Workflow of project using CNN

A. K-NERAEST NEIGHBOUR

As a supervised machine learning tool, the KNN algorithmic technique [10] is used to solve issues involving regression predicting and classification. KNN aims to predict the correct class for the test data by calculating the distance or gap between the test data and all of the training points. When the model is being trained using the training images, the value of k is ascertained.

B.SUPPORT VECTOR MACHINE

The most common machine learning method for categorization jobs is support vector machines (SVM). [11] By creating a decision boundary between various disease classifications, the SVM algorithm operates. Fig. 6 shows the classification of three classes using an SVM classifier. By matching the color and texture features of the given image with the features of each class that is divided by the border, the given image is appropriately classified. We refer to this decision boundary as the hyper-plane.The quantities and kind of parameters used in the SVM classifier determine the efficiency of the SVM model. The project's parameters are c, which has a value of 10, gamma, which has a value of 0.0001, and the kernel, which is RBF[12]. Permissioned

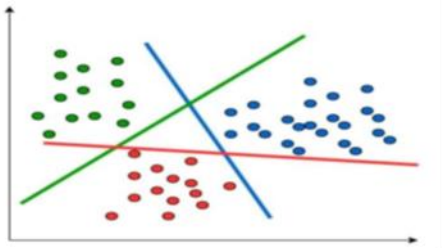

Figure 6. Multiclass classification using SVM

C. CONVOLUTION NEURAL NETWORK

An artificial neural network that is utilized for image recognition and process is called a CNN. It consists of multiple layers of artificial neurons built with the purpose of processing picture pixels. CNN is mostly used for image recognition and to cluster images based on similarity [12]. It is composed of a number of neurons that have learning weights. A significant quantity of data is supplied into each neuron, and it uses similarity to build a weighted total. This sum is then passed through an activation function, and the output is anticipated. CNN is a deep learning system that can recognize differences between multiple features in an image by using layers to assign varying degrees of priority to each feature. The ability of CNN to identify important characteristics without requiring human interpretation gives it a considerable advantage over other algorithms.

There are three layers in CNN, which are as follows:

- Convolution layers
- pooling layer
- fully connected layer

Fig. 7 displays the convolution, pooling, and fully connected layer configurations. By increasing the amount of training data, image augmentation helps to increase accuracy.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 6, No.2, June : 2024

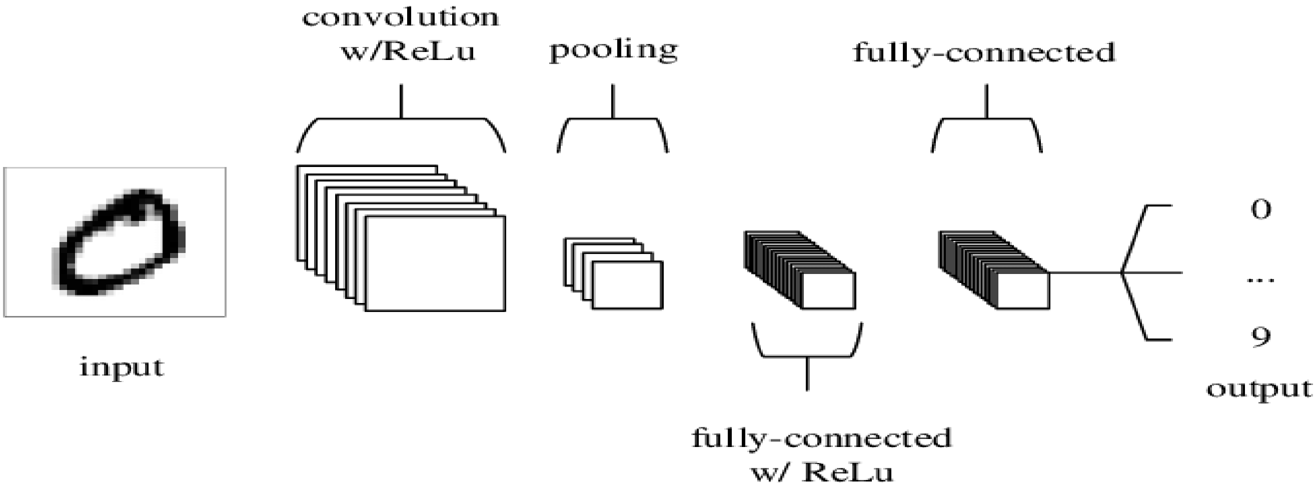

Fig.7. Layers of CNN

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this work, CNN, SVM, and KNN are used to predict six diseases in orange, and the outcomes are compared to identify the most appropriate model. Thus, the outcomes obtained by employing the KNN, SVM, and CNN approaches are also displayed and contrasted below, contingent upon four crucial factors. A confusion matrix, which is a square matrix that shows the true positives, true negatives, false positives, and false negatives, is used to extract these traits.

Figure 8. shows the KNN confusion matrix for six classes (diseases), using 136 test images and 494 training photos. The parameters are assessed to ascertain the model's efficacy.

$$\begin{bmatrix} 33 & 4 & 0 & 3 & 6 & 5 \\ 0 & 54 & 2 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 & 0 \\ 5 & 0 & 0 & 2 & 3 & 1 \\ 1 & 1 & 5 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Fig.8.Confusion matrix of KNN

TABLE I Classification report of KNN

PARAMETERS DISEASES	PRECISION	RECALL	F1 SCORE	SUPPORT
CITRUS CANKER	0.65	0.82	0.73	40
GREEN MOLD	0.96	0.92	0.94	59
LEAF MINER	0.75	0.30	0.43	10
MELANOSE	0.18	0.40	0.25	5
SCAB	0.11	0.10	0.11	10
THRIPS	1.00	0.42	0.59	12
ACCURACY		0.72		136

Figure 9 displays the SVM confusion matrix for six classes (diseases), with 105 test photos and 525 training images. The parameters are assessed to ascertain the model's efficacy.

$$\begin{bmatrix} 36 & 1 & 0 & 0 & 0 & 1 \\ 1 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 24 & 1 & 6 & 1 \\ 1 & 0 & 0 & 5 & 1 & 0 \\ 0 & 2 & 4 & 2 & 7 & 1 \\ 0 & 0 & 0 & 0 & 1 & 7 \end{bmatrix}$$

Figure 9. Confusion matrix of SVM

TABLE II Classification report of SVM

PARAMETERS DISEASES	PRECISION	RECALL	F1 SCORE	SUPPORT
GREEN MOLD	0.95	0.95	0.95	38
LEAF MINER	0.50	0.75	0.60	4
CITRUS CANKER	0.86	0.75	0.80	32
MELANOSE	0.62	0.71	0.67	7
SCAB	0.47	0.44	0.45	16
THRIPS	0.70	0.88	0.78	8
ACCURACY		0.78		106

Fig. 10 displays the CNN confusion matrix for six classes (diseases), with a total of 136 test images and 675 training photos. The parameters are assessed to ascertain the model's efficacy.

$$\begin{bmatrix} 33 & 4 & 0 & 3 & 6 & 5 \\ 0 & 54 & 2 & 0 & 0 & 0 \\ 1 & 0 & 3 & 0 & 0 & 0 \\ 5 & 0 & 0 & 2 & 3 & 1 \\ 1 & 1 & 5 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 5 \end{bmatrix}$$

Figure 10. Confusion matrix of CNN

TABLE III Classification report of CNN

PARAMETERS DISEASES	PRECISION	RECALL	F1 SCORE	SUPPORT
GREEN MOLD	0.98	0.96	0.97	52
LEAF MINER	1.00	0.50	0.67	8
CITRUS CANKER	0.80	1.00	0.89	36
MELANOSE	1.00	0.90	0.95	10
SCAB	00.82	0.88	0.85	16
THRIPS	1.00	0.71	0.83	14
ACCURACY		0.90		136

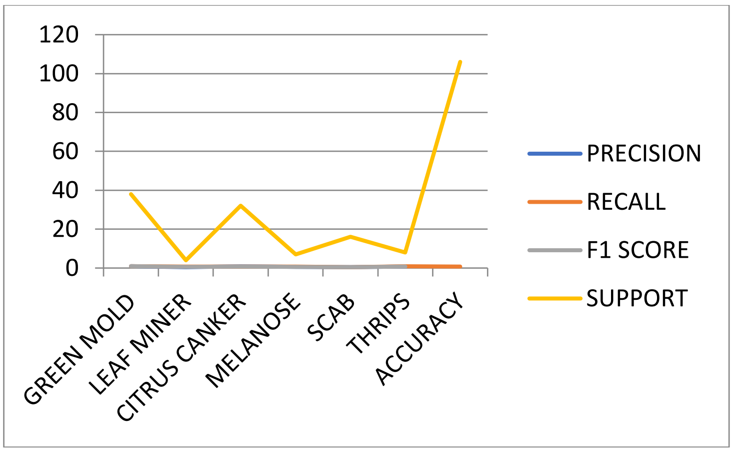

Figure 11. Comparison of KNN,SVM and CNN models for Green Mold

The CNN, SVM, and KNN models all predict green mold illness similarly (Fig. 11).

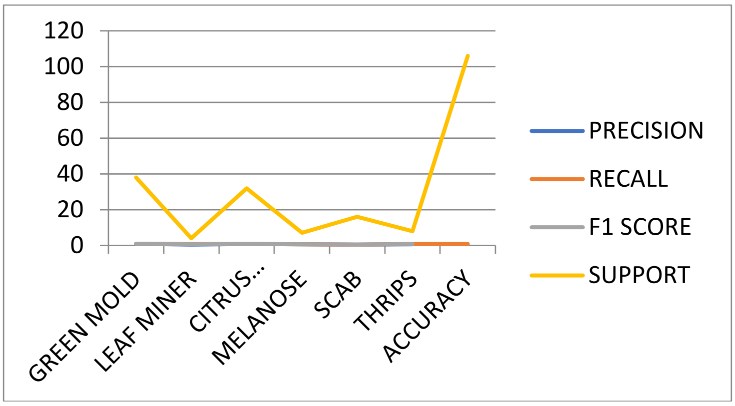

Figure 12.Comparison of KNN, SVM and CNN models for Scab

Based on a comparison of parameters with the SVM and KNN models, the CNN model is the best effective at predicting disease as Scab (Fig. 12).

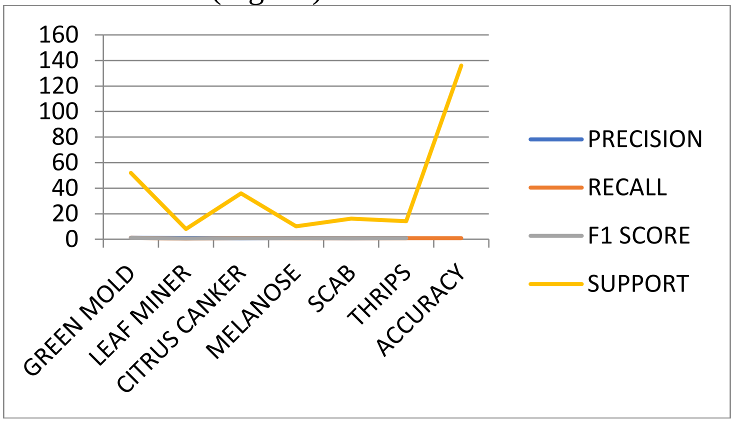

Fig.13. Comparison of KNN, SVM and CNN models for Citrus Canker

Based on a comparison of parameters, the CNN model outperforms the KNN and SVM models in terms of illness prediction for Citrus Canker (Figure 13).

5. CONCLUSION

Machine learning techniques like CNN, SVM, and KNN were used in this study to identify illnesses in oranges. The dataset comprises 680 images for the CNN and KNN models and 525 photos for the SVM model; 20% of the dataset is used for test images and the remaining 80% is used for training images for the models. The SVM model's initial accuracy was only 68%. Increasing the size of the dataset used to train the model, fixing the dataset, and eliminating duplicate data are some methods used to improve accuracy. The model was trained with an accuracy of 78% using the SVM technique, and 72% using the KNN algorithm, with a K value of 9. 90% accuracy was attained when the CNN method was used to train the model. The output from the CNN model is more efficient than that from the KNN and SVM models when other characteristics such as accuracy, recall, specificity, and the f1 score of each disease are compared for all models. Additionally, it was discovered by looking at the classification reports for all the models that the CNN model is trained effectively because the majority of its values are above 0.82. Nonetheless, the majority of the values below 0.75 were discovered in the SVM classification report.

REFERENCES

[1] Shweta Marandil, Prof. Chandrani Chakravorty, “Apple Fruit Disease Detection Using Image Processing”, Journal of Emerging Tech. and Innovative Research (JETIR), vol.8,no.6,2021.
[2] Pranjali. B. Padol and Prof. Anjali A. Yadav, “SVM classifier based grape leaf disease detection”, Conference on Advances in signal proc.,pp. 175-179,2016.

[3] Hardik Patel, Rashmin Prajapati and Milin Patel, “Detection of Quality in Orange Fruit Image using SVM Classifier”, The Third International Conference on Trends in Electronics and Informatics, pp. 74-78,2019.
[4] Santi Kumari Behera, Lipsarani Jena, Amiya Kumar Rath and Prabira Kumar Sethy, “Disease Classification and Grading of Orange using Machine Learning and Fuzzy Logic”, International conference on communication and signal processing, pp. 0678-0682,2018.
[5]K. Renugambal, B. Senthilraja, “Application of Image Processing Techniques in Plant Disease Recognition”, IJERT, Volume 4, No. 3, 2015.
[6] Saradhambal.G, Dhivya.R, Latha.S, R.Rajesh, “Plant disease detection and its solution using image classification”, International Journal of Pure and Applied Mathematics, Volume 119, No. 14, 2018.
[7] Shitala.P, Krishna.M.K, Tripathi.R.C, “Relative sub-image based features for leaf recognition using support vector machine”, International Conference on Communication, Computing & Security, pp. 343-346,2011.
[8] Dhiya Mahdi, Asriny, Septia, Rani, Ahmad, Fathan Hidayatullah, “Orange Fruit Images Classification using Convolutional Neural Networks”, International Conference on Information Technology and Digital Applications,2019.
[9] Mehenag Khatun, Forhad Ali, Nakib Julker Nine, Pritom Sarker, Fruits classification using Convolutional Neural Network”, Global research and development journal for engineering, vol.5, no.8,2020.
[10] Prabira Kumar Sethy, Baishalee Negi, Nilamani Bhoi, “Detection of Healthy and Defected Diseased Leaf of Rice Crop using K-Means Clustering Technique”, International Journal of Computer Applications,2017.
[11] K.Thilagavathi, K.Kavitha, R.Dhivya Praba, S.V.Arockia Joseph Arina and R.C.Sahana, 'Detection of Diseases in Sugarcane Using Image Processing Techniques', Bioscience Biotechnology Research Communications. Special Issue vol. 13 no. 11,pp-109-115,2020.
[12] Thilagavathi, K & Vasuki, A, ‘A Novel Hyperspectral Image Classification Technique Using Deep Multi-Dimensional Recurrent Neural Network’, Applied Mathematics & Information Sciences, vol. 13, no. 6, pp. 955-963.ISSN: 2325-0399,2019.
[13] Hsin-Chun Tsai;Yang-Yen Qu;Chuan-Han Lin;Nan-Han Lu;Kuo-Ying Liu;Jhing-Fa Wang Automatic Rib Fracture Detection and Localization from Frontal and Oblique Chest X-Rays2022 10th International Conference on Orange Technology (ICOT) 2022
[14] Sadhana T S;Anil Kumar R J;Bhavani S D;Santosh Kumar B N Fruit Quality Identification Using Deep LearningTechniques 2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT) 2022
[15] Chai C. Foong;Goh K. Meng;Lim L. Tze Convolutional Neural Network based Rotten Fruit Detection using ResNet50 2021 IEEE 12th Control and System Graduate Research Colloquium (ICSGRC) 2021