A Novel Convolutional Neural Network Based Model for Recognition and Classification of Apple Leaf Diseases

Divakar Yadav*, Akanksha, Arun Kumar Yadav

Department of Computer Science and Engineering, National Institute of Technology, Hamirpur (HP) 177005, India

Corresponding Author Email: divakaryadav@nith.ac.in

ABSTRACT

Plants have a great role to play in biodiversity sustenance. These natural products not only push their demand for agricultural productivity, but also for the manufacturing of medical products, cosmetics and many more. Apple is one of the fruits that is known for its excellent nutritional properties and is therefore recommended for daily intake. However, due to various diseases in apple plants, farmers have to suffer from a huge loss. This not only causes severe effects on fruit's health, but also decreases its overall productivity, quantity, and quality. A novel convolutional neural network (CNN) based model for recognition and classification of apple leaf diseases is proposed in this paper. The proposed model applies contrast stretching based pre-processing technique and fuzzy c-means (FCM) clustering algorithm for the identification of plant diseases. These techniques help to improve the accuracy of CNN model even with lesser size of dataset. 400 image samples (200 healthy, 200 diseased) of apple leaves have been used to train and validate the performance of the proposed model. The proposed model achieved an accuracy of 98%. To achieve this accuracy, it uses lesser data-set size as compared to other existing models, without compromising with the performance, which become possible due to use of contrast stretching pre-processing combined with FCM clustering algorithm.

1. INTRODUCTION

The Indian economy is heavily dependent on efficient agriculture. The detection of diseases in plants therefore plays an important role in agriculture [1]. The use of automated disease detection techniques is advantageous for the fast identification of diseases in plants [2]. For instance, black rot is one of the most prevalent and serious diseases that plagues apple trees. They appear as brown spots, which expand in concentrated circles and finally turn black, decaying the fruits. Later, the disease spreads to the roots of the tree causing cancers that can ultimately kill the tree. Early-stage detection of these diseases in such situations could have been helpful.

To prevent large losses, different techniques for diagnosing diseases have been developed in the past. Techniques developed in microbiology and immunology offer correct recognition of the causative agents. Nonetheless, for many farmers, these approaches are inaccessible and require a thorough knowledge of the region or a large amount of money and energy to carry out. As per the United Nations Food and Agriculture Organization, most farms in the world are small and managed by families in developing countries such as India [3]. Such families grow food for a large proportion of the population of the country. Even so, hunger and food scarcity are not unusual and, market access and resources are constrained. For the above reasons, much work has been done in order to develop methods that are sufficiently reliable and available to the majority of farmers. The techniques of digital image processing increase the chance of early identification of diseases in plants, so that the required preventive steps can be taken [4].

While researchers have worked rigorously to identify plant diseases using different methods such as RNA/DNA, sensor techniques, etc. [5] but the field of machine vision to identify manifestations of fruit leaf diseases is still less examined. Apples are one of the widely consumed fruits, a great source of phytochemicals mostly expressing pertinent antioxidant abilities in vitro, and scientific studies have related apple ingestion to a lower chance of certain cancers, cardiovascular disease, asthma, and diabetes [6, 7]. The consolidated list of abbreviations used in the manuscript is as shown in Table 1.

Table 1. Acronyms used in the manuscript

Symbols	Abbreviations
CNN	convolutional neural network
FCM	Fuzzy C-means
SVM	Support Vector Machine
PSO	Particle Swarm Optimization
CFS	correlation-based feature selection
CLAHE	Contrast Limited Adaptive Histogram Equalization
GLCM	Gray Level Co-Occurrence Matrix
PHOG	Pyramid Histogram of Oriented Gradients
GA	genetic algorithm

The main contributions in this work are stated as follows:
1. A novel convolutional neural network based model for recognition and classification of apple leave disease is proposed. The proposed model utilizes contrast stretching based pre-processing and fuzzy c-means clustering for image segmentation. Both these approaches boost the performance of CNN classier even on lesser size of training data as compared to other state of the art methods.

2. A comprehensive discussion on the existing work is presented to elaborate the research gaps.

3. Extensive computer simulations are performed to determine the effectiveness of the proposed system. A benchmark dataset (Kaggle) which is composed of 4-types of apple leaves are used for simulations. Simulation result reveals that the proposed system showed competitive performance over the other state-of-the-art methods.

2. RELATED WORK

Traditional ways for identifying as well as analyzing the diseases in fruit leaves are manual. However, these manual processes take time, are cumbersome and also very subjective [8]. Several methods have been developed in recent years incorporating computer vision to detect and identify agricultural and horticultural crop diseases to address the manual techniques issues [9, 10]. Image collection, retrieval of features, filtering of features and classification analysis with parametric or non-parametric statistics are fundamental steps in those processes. Image processing techniques and classification mechanisms are the main concern for the efficient functioning of the computer vision system.

Research on the identification of plant disease using machine learning is on the rise. The main reason it may be that expert eye observation of scientists has often proved to be very impractical for such systems and moreover, constant surveillance is needed, which is very costly when dealing with large farms [11]. In some areas, farmers do not have appropriate facilities or even the knowledge that they can seek experts. Under these conditions, automatic disease detection by seeing the signs on the leaves of plants makes the system much faster, easier and cheaper. This also promotes machine vision to offer automated image based process control, examination and robot assistance [12].

Selvaraj et al. [13] suggested a four step scheme: first an RGB image color transformation system is generated for the input, the green pixels are then covered and replaced using different threshold values preceded by segmentation. For useful segments, texture statistics are calculated and the derived features are finally passed to the SVM classifier. Pujari et al. [14] proposed Support Vector Machine and Artificial Neural Network based identification and classification of fungal disease in cereals. The regions concerned are segmented using k-means segmentation. Color texture characteristics are extracted from affected regions and then used as classifier inputs. Vishnu et al. [15] used K-means clustering technique for leaf segmentation and then calculated the texture features for the segmented infected objects. Lastly, the derived features were processed through neural network model. Muthukannan and Latha [16] proposed a novel solution to image segmentation, called PSO. PSO is an efficient, self-regulating unsupervised algorithm that is used for improved segmentation and extraction of features. The hybrid characteristic coefficients were then obtained from the co-occurrence gray level matrices of various leaves. Chung et al. [17] have suggested an approach using support vector machine (SVM) classifiers to differentiate healthy and Bakanae-infected rice seedlings.

Zhang et al. [18] suggested a technique for the identification of cucumber disease based on decomposition of the global-local single value to increase the detection rate. The classification of the unidentified disease leaf picture was implemented by SVM classifier. Ashourloo et al. [19] utilized regression techniques for the identification of wheat plant rust disease. Later, they also evaluated the effect of data set on results. Ali et al. [20] used the ΔE color difference algorithm to isolate the affected area from the leaf as well as color histogram and compositional features to identify diseases. They applied principal components analysis for the features set dimension reduction and Bagged tree classifier for classification. Kaur et al. [21] used desegregated Particle Swarm Optimization (PSO) technique and support vector machine (SVM) for the identification and classification of plant leaf diseases. The prime motive of this study was to identify portion of the leaf, affected by the disease and a stable portion of the leaf. Ma et al. [22] employed comprehensive color function and its method of detection that can segment images of disease spots recorded under real field circumstances. The approach ensures reliable feedback into detection of CNN-based disease recognition.

Mondal et al. [23] used forty-three morphological characteristics of okra and bitter gourd leaves to identify symptoms of disease from their images. In this approach, feature set collection was rendered utilizing Pearson correlation coefficient and the entropy-based discretization was used to improve the classification success rate. Zhang et al. [24] utilized various useful features that were chosen by combining the merits of the genetic algorithm (GA) with correlation-based feature selection (CFS). Here, GA and CFS played a key role in decreasing the dissensions of the feature space. Lastly, SVM classifier was used for diseases identification. Singh et al. [25] described a method of image segmentation using GA and later classification was done by using SVM. B. Liu et al. [26] proposed a model for accurate identification of apple leaf disorder that requires producing many pathological representations. Finally, they developed a novel architecture of an AlexNet based deep CNN for disease detection. A computational network of CNNs was proposed by Dechant et al. [27] to tackle the limitations of restricted data and the countless variations that occur in field grown leaf images. Multiple CNNs were trained to identify small image areas, and their predictions were compiled into independent heat maps that are then fed into a final CNN trained model to identify the entire image as diseased or not.

Hanson et al. [28] proposed a new method for the detection of plant diseases using a deep convolutional neural network trained as well as fine-tuned to suit appropriately with plant leaves database that was independently collected for various plant diseases. Yao et al. [29] proposed an efficient three-layer discovery system for the characterization of different stages of growth of white-backed planthoppers on rice crops in paddy farmlands. Sethy et al. [30] proposed fuzzy Logic together with K-means segmentation method to measure the extent of the disease in rice crops. Fuentes et al. [31] proposed a deep-learning method to identify diseases as well as pests in tomato using photographs recorded in-place with various resolutions by camera devices. The research illustrated the efficiency of deep meta-architectures and characteristic extractors.

Sunny et al. [32] proposed two-stage solution to enhance the image clarity. The primary stage uses Contrast Limited Adaptive Histogram Equalization (CLAHE) to pre-process the leaf image, followed by segmentation using K-mean clustering and extraction of texture characteristics via statistical Gray Level Co-Occurrence Matrix (GLCM). The second stage utilizes support vector machine to identify the plant as healthy or ill. Zhang et al. [33] suggested clustering algorithm to split

the color-diseased image of the leaf into many small super-pixels and then used K-means clustering method to fragment the image of every super-pixel of the lesion. Finally, the Pyramid Histogram of Oriented Gradients (PHOG) functionality was derived from three color attributes of each fragmented lesion image as well as its grayscale image, and a vector was concatenated with four PHOG descriptors.

Alsuwaidi et al. [34] used a ground-breaking analytical classification system in which they incorporated adaptive feature collection, novelty identification and ensemble learning with the hyper spectral datasets. Singh et al. [35] presented an automated approach to differentiate between Neem and Bakain using the texture characteristics of its leaves and then they used tree classifier to separate them in separate classes. Brahimi et al. [36], suggested the employment of saliency maps for simulation to perceive and decode the CNN classification process. This process of simulation improves the clarity of deep learning frameworks and offers further knowledge about plant disease symptoms.

Yue et al. [37] proposed a super resolution model that relies on the residual deep recursive network and provides the state-of-the-arts performance as compared to the traditional methods. Iqbal et al. [38] presented a review on the different approaches of identification and classification of diseases specific to citrus plants.

Dhingra et al. [39] has outlined a neutrosophic approach based on computer vision for plant disease analysis. This system uses a fuzzy set extension technique based on neutrosophical logic segmentation to analyze the area of interest and then, new feature subset is assessed on the basis of segmented area to classify the basil leaf as healthy or diseased. Picon et al. [40] used the Deep Residual Neural Network-based algorithm to detect plant diseases under specific acquisition circumstances where various adaptations have been suggested for early disease discovery. This research analyzes the early identification success of three related European endemic diseases of wheat: Septoria, Tan Spot and Rust.

Wu et al. [41] proposed a new solution to crop disease detection focused on multi-functional sparse constrain system that mainly includes three phases: segmentation of lesions, extraction of features and disease detection. They applied this model for diseased cucumber images and achieved accuracy of 88.05% Waheed et al. [42] proposed DenseNet, a dense convolutional neural network based model for recognition and identification of diseases in corn leaves. They claimed that their proposed method uses significantly lesser parameters as compared to other state-of-the-arts methods.

3. MODELS AND METHODS

This section discusses in detail about the proposed method of apple leaves disease classification and identification including dataset collection, pre-processing, segmentation, feature extractions, training, testing etc. It is organized in different subsections.

3.1 Data set

The dataset is composed of four types of apple leaves downloaded from Kaggle [43]. These are healthy and unhealthy apple leaves. Among the unhealthy apple leaves, it consists of three types of infections namely: apple scab, black rot and apple rust. Figure 1 shows healthy and unhealthy sample of images in the dataset collection used in this article.
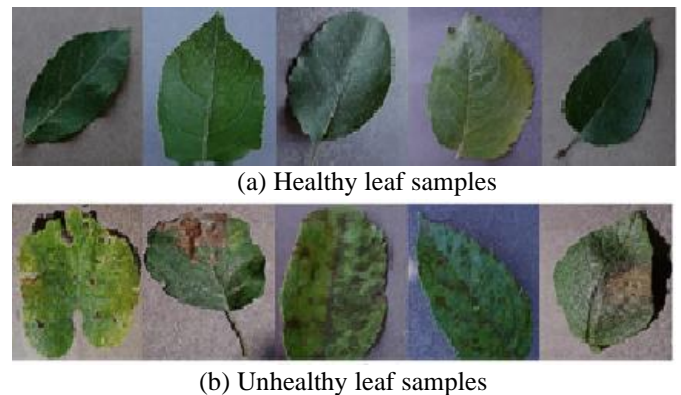

(a) Healthy leaf samples

(b) Unhealthy leaf samples

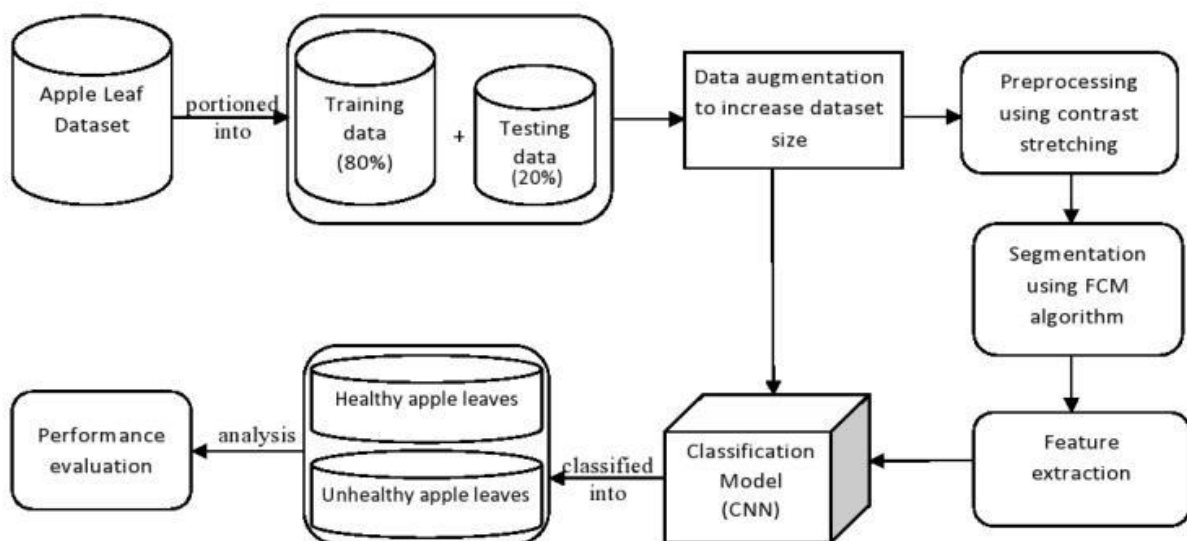**Figure 1.** Dataset collection of apple leaf images (a) Healthy (b) Unhealthy

Figure 2. Flow diagram of basic system model for proposed approach

3.2 Basic model design

Following are fundamental steps followed to design the proposed systems:
1. *Pre-processing of data:* The main objective of pre-processing is to highlight the information which is concealed by the contrast stretching [44] approach for improved contrast.
2. *Segmentation:* The concerned area is identified after pre-processing using FCM clustering segmentation.
3. *Extraction of features:* The classification model used in the proposed work is convolutional neural networks [45] which is also used for extraction of features.
4. *Classification:* The methodology used for the identification of leaf disease is CNN.

The steps mentioned above are addressed in more depth in the subsequent parts. The flow diagram given in Figure 2 presents the proposed approach. The primary purpose of this work is to enable the system to learn the characteristics that differentiate one class from another. In order to achieve this, dataset can be increased by using augmented images to improve the network's probability of learning the correct features [46]. Figure 3 shows the result of augmentation on unhealthy apple leaf images.
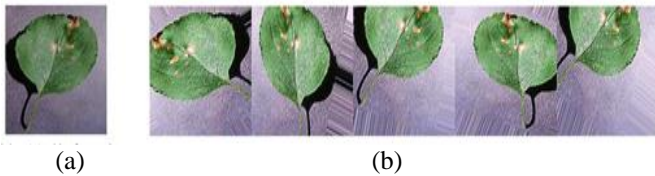

(a)   (b)

Figure 3. (a) Original (b) augmented samples of unhealthy apple leaf

3.3 Preprocessing

Image preprocessing is an important step in examination and manipulation of a digital photo [47], particularly with a view to highlight problem regions i.e., diseased region after data gathering is done. However, visual enhancement is one of the complexities of image processing [48], and is also very unique to a specific task. In this work, the contrast stretching algorithm is used to enhance the image. The contrast enhancement methods are used to extend the range of brightness levels in an image, so that the image can be viewed effectively in the way the analyst wishes. The contrast level in an image can differ due to poor lighting or inappropriate setting in the calibration device.

The method of contrast stretching is to examine the spread of pixel concentrations in a picture and then dynamically resize the picture so as to include all levels of intensity falling within 2nd and 98th percentiles. The formula for global contrast stretching is governed by the following equation

$$out_{RGB}(x, y) = 255 * \left[\frac{(in_{RGB} - min_{RGB})}{min_{RGB} - min_{RGB}}\right] \qquad (1)$$

where, $in_{RGB}(x, y)$ is the original RGB value of the pixel, $out_{RGB}(x, y)$ is the new RGB value of the pixel, min_{RGB} is minimum value between the components, and max_{RGB} is maximum value between the RGB components (red, green, and blue) of the original image.

3.4 Segmentation technique

Segmentation of images is an arduous process because of the intricacy and variety of images [49]. Factors like lighting [50], contrast, interference etc. affect the outcome of segmentation. The segmentation aim is to identify the areas of concern to determine the disease. We have advocated FCM clustering approach for segmentation. FCM is a clustering technique that enables a piece of information to be a member of more than one cluster and therefore, it belongs to the class of soft segmentation technique. These techniques are popularly used for image segmentation since much better details from the main image could be retrieved in comparison to hard segmentation approaches. FCM is a technique of clustering that allocate pixels to unlabelled clusters with different membership rates. FCM clustering segmentation algorithm is outlined in algorithm 1 [51]. Unlike other clustering methods, where data point must exclusively belong to one cluster center, in FCM data clustering, point is assigned membership to each cluster center as a result of which data point may belong to more than one cluster center. Figure 4 shows the result on unhealthy leaf sample after applying pre-processing and segmentation steps.
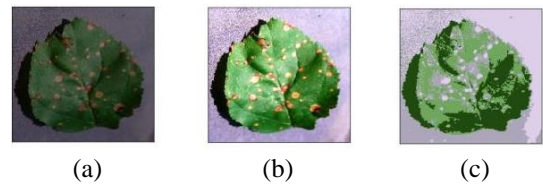

(a)   (b)   (c)

Figure 4. Result of pre-processing and segmentation on unhealthy leaf sample, (a) Original sample, (b) Contrast stretched sample, (c) FCM segmented sample

Algorithm 1: FCM Clustering segmentation
1. Select principal centroids C_i.
2. Use Eq. (2) to evaluate the disparity between the vectors and the centroid.

$$disp_a = \sqrt{(d_x - p_x)^2 + (d_y - p_y)^2} \qquad (2)$$

where, $disp_a$ is Euclidean distance between data points/clusters center (d_x, d_y) and (p_x, p_y).
3. Evaluate the membership coefficient for vector belonging to a particular cluster using Eq. (3).

$$mem_b(d_a) = \frac{\left[\frac{1}{disp_{ba}}\right]^{\frac{1}{f-1}}}{\sum_{n=1}^{N}\left[\frac{1}{disp_{na}}\right]^{\frac{1}{f-1}}} \qquad (3)$$

where, f represents fuzziness coefficient. This value of membership has to be calculated for all the clusters; n is the number of data points belonging to the cluster.
4. Determine new centroid using Eq. (4) and upgrade membership coefficient from step 3.

$$P_b^{new} = \frac{\sum_a [mem_b(d_a)]^f d_a}{\sum_a [mem_b(d_a)]^f} \qquad (4)$$

5. If the centroids do not alter, stop. Else go to step 4.

3.5 Feature extraction and classification

The remarkable improvement in performance, achieved for different tasks using deep neural networks inspired us to employ it for the purpose of image classification in the present work. We used CNN to extract features and classify images. The output of the segmentation is given to the CNN to classify whether an image is healthy or not. A CNN is a kind of feed-forward network and is an end-to-end pipeline methodology that can inevitably uncover the discriminatory characteristics for image classification. Since the features in CNN are not extracted and implemented on the basis of human knowledge, they are very less prone to artificial feature extraction.

CNN consists of layers; however, these layers are not fully interconnected. They have filtering process, which includes series of cube like shape weights applied all through the picture. Each two-dimensional filtration portion is called "kernel". Major component of CNN is the convolutional layer. A convolution functions between the higher feature maps of the present layer and the convolution kernels decide the resulting feature map for every convolution layer. The output characteristic map can be given by Eq. (5) [26].

$$In_b^l = \sum_{a \in X_b} In_a^{l-1} * c_{ab}^l + B_b^l \tag{5}$$

where, l is the lth layer, B_b denotes bias, c_{ab} is the convolutional kernel and X_b is set of input feature maps. The neural model's learning capacity is affected by the ReLU activation function and it has a quick convergence speed. This approach is therefore used for the output of each convolutional layer. Mathematically it can be expressed as given in Eq. (6) [26]:

$$f(z) = maximum(0, z) \tag{6}$$

```
Layer (type)                 Output Shape
=================================================
conv2d_1 (Conv2D)            (None, 128, 128, 32)
_____
activation_1 (Activation)    (None, 128, 128, 32)
_____
conv2d_2 (Conv2D)            (None, 126, 126, 32)
_____
activation_2 (Activation)    (None, 126, 126, 32)
_____
max_pooling2d_1 (MaxPooling2  (None, 63, 63, 32)
_____
dropout_1 (Dropout)          (None, 63, 63, 32)
_____
conv2d_3 (Conv2D)            (None, 61, 61, 64)
_____
activation_3 (Activation)    (None, 61, 61, 64)
_____
max_pooling2d_2 (MaxPooling2  (None, 30, 30, 64)
_____
dropout_2 (Dropout)          (None, 30, 30, 64)
_____
flatten_1 (Flatten)          (None, 57600)
_____
dense_1 (Dense)              (None, 64)
_____
activation_4 (Activation)    (None, 64)
_____
dropout_3 (Dropout)          (None, 64)
_____
dense_2 (Dense)              (None, 2)
_____
activation_5 (Activation)    (None, 2)
=================================================
```

Figure 5. Parameters of CNN model

Pooling is another crucial layer of CNN. It is a type of nonlinear down sampling. The size of the feature maps obtained from convolutional layers could be reduced using this layer to accomplish spatial invariance. Completely connected layers are inserted before a CNN's classification output, and used before classification to straighten the result. The end-product of the final pooling or convolutionary layer is the entry to the totally linked layer. Figure 5 represents the CNN model parameters used in this work.

4. EXPERIMENTAL OBSERVATION

4.1 Dataset of apple leaf images

The dataset collection consists of 400 images, out of which 200 images are healthy leaf samples and the rest consists of various categories of diseased leaf samples that includes apple scab, black rot, and apple rust. All the experiments are performed employing "Keras framework on top of tensorflow". The dataset is expanded by utilizing "ImageDataGenerator class" of tensorflow.

4.2 Evaluation metrices for system model

The efficiency of the proposed system was assessed using different evaluation parameters [52, 53]. A brief overview about these metrices is provided below.

1.   *Accuracy(A):* The accuracy metric is defined as the ratio of the number of correctly classified examples to the total number of classified examples.

$$A = (TP + FN)/(TP + FP + TN + FN) \tag{7}$$

2.   *Precision(P):* Precision is used to quantify the positive patterns correctly predicted in a positive class from total predicted patterns.

$$P = TP/(TP + FP) \tag{8}$$

3.   *Recall(R):* Recall is used to calculate the proportion of positive occurrences correctly categorized.

$$R = TP/(TP + TN) \tag{9}$$

4.   *F1-score(F):* F1-score or F-measure metric represents the harmonic mean between recall and precision values [52].

$$F = 2 * P * R/(P + R) \tag{10}$$

where, TP = True Positive, FP = False Positive, TN = True Negative and FN = False Negative.

4.3 System model evaluation (Training-Validation observation)

In this section, based on the training-testing dataset listed within Table 2 and Table 3, we examined the predictive outcome of CNN model. Data distribution for this model is set at 80% and 20% respectively in the training-validation analysis.

Table 2. Data partitioning of original image samples for training and validation

Data set	Total samples	Training samples	Validation samples
Healthy	200	158	42
Unhealthy	200	162	38

Table 3. Data partitioning of augmented image samples for training and validation

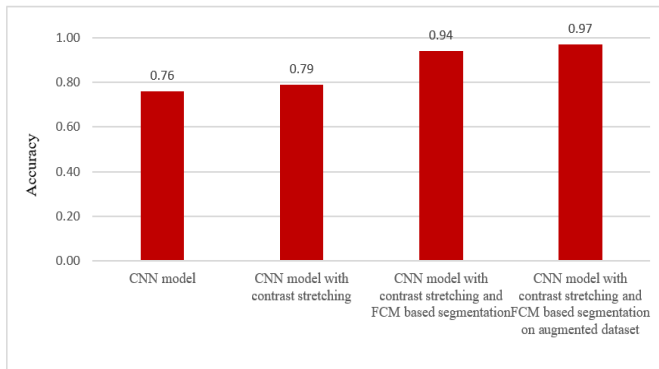| Augmented Dataset	Total samples	Training samples	Validation samples
Healthy	952	757	195
Unhealthy	944	759	185

Figure 6. Comparison of CNN model on adding each proposed step (in terms of accuracy)
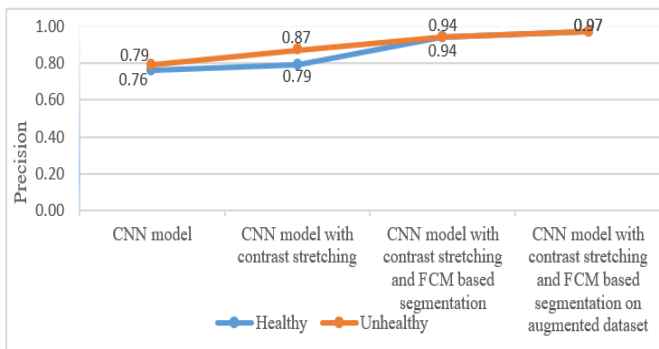

Figure 7. Comparison of CNN model on adding each proposed step (in terms of precision)
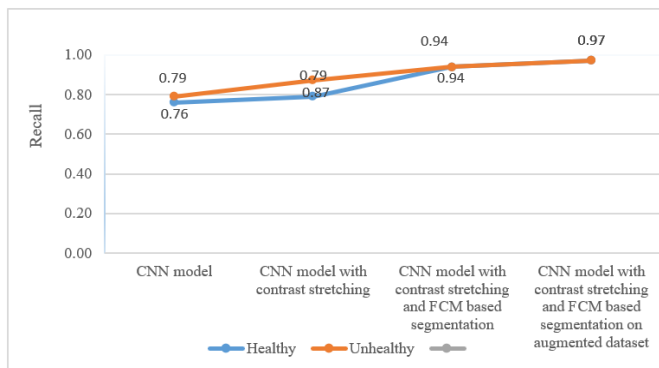

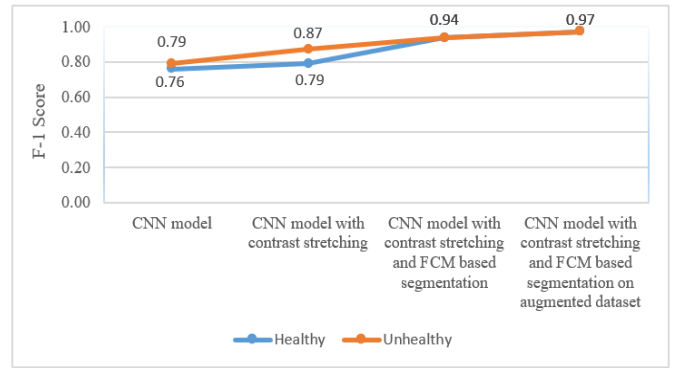**Figure 8.** Comparison of CNN model on adding each proposed step (in terms of recall)

Figure 9. Comparison of CNN model on adding each proposed step (in terms of F1-score)

Figures 6-9 demonstrates the CNN model's classification accuracy by adding each proposed step with respect to different assessment parameters, defined in section 4.2. Since on training the model with bigger dataset, we can expect the outcome to be more accurate on test data. Therefore, the assessment parameters are also evaluated on the augmented dataset. Augmented dataset is the improved version of existing dataset in terms of its size and variety, without the need to explicitly gather new data. It is clear from Figure 6 that the proposed model achieves 94% accuracy with lower dataset and this has been further improved to 98% with the use of augmented dataset.

4.4 Analysis on various Training-Validation partitions

In addition, the accuracy is measured on various training-validation divisions i.e. 50-50, 60-40, 70-30, 80-20 and 90-10, as shown in Tables 4 and 5 for original and augmented dataset respectively.

Table 4. Performance analysis on various training-validation divisions in the original dataset

Training and validation partitions	Training samples	Validation samples	Accuracy rate
50-50%	200	200	85%
60-40%	240	160	88%
70-30%	280	120	89%
80-20%	320	80	94%
90-10%	360	40	95%

Table 5. Performance analysis on various training-validation divisions in the augmented dataset

Training and validation partitions	Training samples	Validation samples	Accuracy rate
50-50%	948	948	91%
60-40%	1137	759	94%
70-30%	1327	569	95%
80-20%	1516	380	97%
90-10%	1706	190	98%

Accuracy rate is evaluated on both original as well as augmented dataset to check the robustness of the proposed model. The accuracy of CNN model depends strongly on its training dataset size. But the presented system achieves sufficiently high accuracy even with the small training dataset. Results in Table 4 show that the presented method can achieve accuracy of 95% when 360 training and 40 validation image

samples were used. Also, it is evident from Table 5 that with an improved data set and higher training ratio, the proposed system can achieve accuracy of up to 98%.

4.5 Comparison with the existing systems

In this experiment, we measure our proposed model's classification accuracy against the conventional methods [26, 28]. As outlined in Table 6, the proposed approach provides superior performance than other methods with far less augmented dataset size. The comparison with these systems is made on the grounds of common use of augmented dataset and to show how the proposed model performs better even the augmented set is much lower. The main idea of Liu et al. [26] is to generate ample pathological representation of apple images and build an innovative model of AlexNet based deep CNN for disease identification. Waheed et al. [28] also achieved good results by collecting more dataset and tuning of system variables. Waheed et al. [42] proposed optimized DenseNet model corn leaf and achieved an accuracy of 98.06%. In this work, the prime motive was to enhance the system accuracy with minimal dataset which is acquired by using contrast stretching as pre-processing method and FCM clustering algorithm for segmentation.

Table 6. Comparison of proposed model with the existing models

	Liu [26]	Hanson [28]	Waheed [42]	Proposed method
Original dataset size	1053	>200	12,332	400
Augmented dataset size	13689	33469	Not available	1896
Accuracy rate	97.62%	95%	98.06%	98%

5. CONCLUSION

In this work, we proposed a novel approach to efficiently use contrast stretching based preprocessing and fuzzy c-means segmentation, together with CNN to identify disease in apple leaves. The entire operation was outlined from collecting images to segmentation and eventually, feature extraction and classification by CNN. Based on the results of preprocessing followed by segmentation, a new deep convolutional neural network model has been developed that discover distinctive features automatically, and also determine apple leaf diseases accurately. The study proposed has been compared with existing state-of-the-art and the results were found quite impressive. The developed system works better with 98% accuracy rate and that too, with much lesser dataset size.

Future research could focus on extending suggested work to characterize each category of diseases separately and approximate the seriousness of the diseases identified. An undiscovered combination of the extraction, collection of features and learning approaches can also be analysed to improve the effectiveness of disease diagnosis and identification models.

REFERENCES

[1] Gavhale, M.K.R., Gawande, U. (2014). An overview of the research on plant leaves disease detection using image processing techniques. IOSR Journal of Computer Engineering, 16(1): 10-16. https://doi.org/10.9790/0661-16151016

[2] Kulkarni, A.H., R.K., Ashwin Patil. (2012). Applying image processing technique to detect plant diseases. International Journal of Modern Engineering Research, 2(5): 3661-3664.

[3] Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., Stefanovic, D. (2019). Solving current limitation of deep learning based approaches for plant disease detection. Symmetry, 11(7): 939. https://doi.org/10.3390/sym11070939

[4] Mainkar, P.M., Ghorpade, S., Adawadkar, M. (2015). Plant leaf disease detection and classification using image processing techniques. International Journal of Innovative and Emerging Research in Engineering, 2(4): 139-144.

[5] Fang, Y., Ramasamy, R.P. (2015). Current and prospective methods for plant disease detection. Biosensors, 5(3): 537-561. https://doi.org/10.3390/bios5030537

[6] Hyson, D.A. (2011). A comprehensive review of apples and apple components and their relationship to human health. Advances in Nutrition, 2(5): 408-420. https://doi.org/10.3945/an.111.000513

[7] Koutsos, A., Tuohy, K.M., Lovegrove, J.A. (2015). Apples and cardiovascular health - Is the gut microbiota a core consideration? Nutrients, 7: 3959-3998. https://doi.org/10.3390/nu7063959

[8] Tadeusiewicz, R., Tylek, P., Adamczyk, F., Kielbasa, P., Jablonski, M., Pawlik, P., Pilat, A., Walczyk, J., Szczepaniak, J., Juliszewski, T., Szaroleta, M. (2017). Automation of the acorn scarification process as a contribution to sustainable forest management, case study: Common oak. Sustainability, 9(12): 2276. https://doi.org/10.3390/su9122276

[9] Barbedo, J.G.A. (2013). Digital image processing techniques for detecting, quantifying and classifying plant diseases. SpringerPlus, 2(1): 1-12. https://doi.org/10.1186/2193-1801-2-660

[10] Dhingra, G., Kumar, V., Joshi, H.D. (2018). Study of digital image processing techniques for leaf disease detection and classification. Multimedia Tools Appl, 77(15): 19951-20000. https://doi.org/10.1007/s11042-017-5445-8

[11] Ghaiwat, S.N., Arora, P. (2014). Detection and classification of plant leaf diseases using image processing techniques: A review. Int J Recent Adv Eng Technol, 2(3): 2347-2812.

[12] Dhaygude, S.B., Kumbhar, N.P. (2013). Agricultural plant leaf disease detection using image processing. Int J Adv Res Electr Electron Instrum Eng, 2(1).

[13] Selvaraj, A., Shebiah, N., Ananthi, S., Varthini, S.V. (2013). Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. Agric Eng Int CIGR, 15(1): 211-217.

[14] Pujari, J.D., Yakkundimath, R., Byadgi, A.S. (2013). Classification of fungal disease symptoms affected on cereals using color texture features. Int. J. Signal Process. 6(6): 321-330.

[15] Vishnu, S., Ram, A.R. (2015). Plant disease detection using leaf pattern: A review. International Journal of Innovative Science, Engineering and Technology, 2(6): 774-780.

[16] Muthukannan, K., Latha, P. (2015). A PSO model for disease pattern detection on leaf surfaces. Image Analysis and Stereology, 34(3): 209-216. https://doi.org/10.5566/ias.1227

[17] Chung, C.L., Huang, K.J., Chen, S.Y., Lai, M.H., Chen, Y.C., Kuo, Y.F. (2016). Detecting Bakanae disease in rice seedlings by machine vision. Computers and Electronics in Agric, 121: 404-411. https://doi.org/10.1016/j.compag.2016.01.008

[18] Zhang, S.W., Wang, Z. (2016). Cucumber disease recognition based on Global-Local Singular value decomposition. Neurocomputing, 205: 341-348. https://doi.org/10.1016/j.neucom.2016.04.034

[19] Ashourloo, D., Aghighi, H., Matkan, A.A., Mobasheri, M.R., Rad, A.M. (2016). An investigation into machine learning regression techniques for leaf rust disease detection using hyper spectral measurement. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(9): 4344-4351. https://doi.org/10.1109/JSTARS.2016.2575360

[20] Ali, H., Lali, M.I., Nawaz, M.Z., Sharif, M., Saleem, B.A. (2017). Symptom based auto-mated detection of citrus diseases using color histogram and textural descriptors. Computers and Electronics in Agric, 138: 92-104. https://doi.org/10.1016/j.compag.2017.04.008

[21] Kaur, P., Singla, S., Singh, S. (2017). Detection and classification of leaf diseases using integrated approach of support vector machine and particle swarm optimization. Int. J. Adv. Appl. Sci., 4(8): 79-83.

[22] Ma, J.C., Du, K.M., Zhang, L.X., Zheng, F.X., Chu, J.X., Sun, Z.F. (2017). A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing. Computers and Electronics in Agriculture, 142: 110-117. https://doi.org/10.1016/j.compag.2017.08.023

[23] Mondal, D., Kole, D.K., Roy, K. (2017). Gradation of yellow mosaic virus disease of okra and bitter gourd based on entropy based binning and Naive Bayes classifier after identification of leaves. Computers and Electronics in Agriculture, 142: 485-493. https://doi.org/10.1016/j.compag.2017.11.024

[24] Zhang, C.L., Zhang, S.W., Yang, J.C., Shi, Y.C., Chen, J. (2017). Apple leaf disease identification using genetic algorithm and correlation based feature selection method. International Journal of Agricultural and Biological Engineering, 10(2): 74-83. https://doi.org/10.3965/j.ijabe.20171002.2166

[25] Singh, V., Mishra, A.K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. Information Processing in Agriculture, 4(1): 41-49. https://doi.org/10.1016/j.inpa.2016.10.005

[26] Liu, B., Zhang, Y., He, D.J., Li, Y. (2017). Identification of apple leaf diseases based on deep convolutional neural networks. Symmetry, 10(1): 11. https://doi.org/10.3390/sym10010011

[27] Dechant, C., Wiesner-Hanks, T., Chen, S., Stewart, E.L., Yosinski, J., Gore, M.A., Nelson, R.J., Lipson, H. (2017). Automated identification of northern leaf blight-infected maize plants from field imagery using deep learning. Phytopathology, 107(11): 1426-1432. https://doi.org/10.1094/PHYTO-11-16-0417-R

[28] Hanson, A.J., Joy, A., Francis, J.N. (2017). Plant leaf disease detection using deep learning and convolutional neural network. International Journal of Engineering Science and Computing, 7(3): 5324-5328.

[29] Yao, Q., Chen, G.T., Wang, Z., Zhang, C., Yang, B.J., Tang, J. (2017). Automated detection and identification of white-backed planthoppers in paddy fields using image processing. Journal of Integrative Agriculture, 16(7): 1547-1557. https://doi.org/10.1016/S2095-3119(16)61497-1

[30] Sethy, P.K., Negi, B., Barpanda, N.K., Behera, S.K., Rath, A.K. (2017). Measurement of disease severity of rice crop using machine learning and computational intelligence. Cognitive Science and Artificial Intelligence, Springer, Singapore, pp 1-11. https://doi.org/10.1007/978-981-10-6698-6_1

[31] Fuentes, A.F., Yoon, S., Lee, J., Park, D.S. (2018). High-performance deep neural network based tomato plant diseases and pests diagnosis system with refinement filter bank. Frontiers in Plant Science, 9: 1162. https://doi.org/10.3389/fpls.2018.01162

[32] Sunny, S., Gandhi, M.P.I. (2018). An efficient citrus canker detection method based on contrast limited adaptive histogram equalization enhancement. International Journal of Applied Engineering Research, 13(1): 809-815.

[33] Zhang, S., Wang, H., Huang, W., You, Z. (2018). Plant diseased leaf segmentation and recognition by fusion of superpixel K-means and PHOG. Optik, 157: 866-872. https://doi.org/10.1016/j.ijleo.2017.11.190

[34] Alsuwaidi, A., Grieve, B., Yin, H. (2018). Feature-ensemble based novelty detection for analyzing plant hyperspectral data. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11(4): 1041-1055. https://doi.org/10.1109/JSTARS.2017.2788426

[35] Singh, N.P., Nagahama, T., Yadav, P., Yadav, D. (2018). Feature based leaf identification. 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON-2018), pp. 1159-1165.

[36] Brahimi, M., Arsenovic, M., Laraba, S., Sladojevic, S., Boukhalfa, K., Moussaoui, A. (2018). Deep learning for plant diseases: detection and saliency map visualization. Human and Machine Learning Human–Computer Interaction Series, pp. 93-117.

[37] Yue, Y., Cheng, X., Zhang, D., Wu, Y.Z., Zhao, Y., Chen, Y.Q., Fan, G.H., Zhang, Y.H. (2018). Deep recursive super resolution network with Laplacian Pyramid for better agricultural pest surveillance and detection. Computers and Electronics in Agriculture, 150: 26-32. https://doi.org/10.1016/j.compag.2018.04.004

[38] Iqbal, Z., Khan, M.A., Sharif, M., Shah, J.H., Rehman, M.H., Javed, K. (2018). An automated detection and classification of citrus plant diseases using image processing techniques: A review. Computers and Electronics in Agriculture, 153: 12-32. https://doi.org/10.1016/j.compag.2018.07.032

[39] Dhingra, G., Kumar, V., Joshi, H.T. (2019). A novel computer vision based neutrosophic approach for leaf disease identification and classification. Measurement, 135: 782-794. https://doi.org/10.1016/j.measurement.2018.12.027

[40] Picon, A., Alvarez-Gila, A.., Seitz, M., Ortiz-Barredo, A., Echazarra., J., Johannes, A. (2019). Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. Computers and

Electronics in Agriculture, 161: 280- 290. https://doi.org/10.1016/j.compag.2018.04.002

[41] Wu, Y.R., Li, J.H. (2019). Multi-feature sparse constrain model for crop disease recognition. Applied Ecology and Environmental Research, 17(4): 9229-9245. https://doi.org/10.15666/aeer/1704_92299245

[42] Waheed, A., Goyal, M., Gupta, D., Khanna, A., Hassanien, A.E., Pandey, H.M. (2020). An optimized dense convolutional neural network model for disease recognition and classification in corn leaf. Computers and Electronics in Agriculture, 175: 105456. https://doi.org/10.1016/j.compag.2020.105456

[43] Bhattarai, S. New Plant Diseases Dataset https://www.kaggle.com/vipooool/new-plant-diseases-dataset, accessed on 12 July 2020.

[44] Perumal, S., Velmurugan, T. (2018). Preprocessing by contrast enhancement techniques for medical images. International Journal of Pure and Applied Mathematics, 118(18): 3681-3688.

[45] Bodapati, J.D., Veeranjaneyulu, N. (2019). Feature extraction and classification using deep convolutional neural networks. Journal of Cyber Security and Mobility, 8(2): 261-276. https://doi.org/10.13052/jcsm2245-1439.825

[46] Parente, A.P., de Souza, M.B., Valdman, A., Folly, R.O.M. (2019). Data augmentation applied to machine learning-based monitoring of a pulp and paper process. Processes, 7(12): 958. https://doi.org/10.3390/pr7120958

[47] Muthuselvi, S., Prabhu, P. (2016). Digital image processing techniques - A survey. International Multidisciplinary Research Journal, 5(11).

[48] Shukla, K.N., Potnis, A., Dwivedi, P. (2017). A review on image enhancement techniques. International Journal of Engineering and Applied Computer Science (IJEACS), 2(7): 232-235. https://doi.org/10.24032/ijeacs/0207/05

[49] Wang, Z.B., Wang, K.Y., Yang, F., Pan, S.H., Han. Y.Y. (2018). Image segmentation of overlapping leaves based on Chan-Vese model and Sobel operator. Information Processing in Agriculture, 5(1): 1-10. https://doi.org/10.1016/j.inpa.2017.09.005

[50] Pratiwi, D., Kartowisastro, I.H. (2015). Object segmentation under varying illumination effects. NewTrends in Intelligent Information and Database Systems Studies in Computational Intelligence, pp. 13-21. https://doi.org/10.1007/978-3-319-16211-9_2

[51] Kumari, A., Meenakshi, S., Abinaya, S. (2018). Plant leaf disease detection using fuzzy c-means clustering algorithm. International Journal of Engineering Research and Technology, 6(3): 157-163.

[52] Hossin, M., Sulaiman, M.N. (2015). A review on evaluation metrices for data classification evaluations. International Journal of Data Mining and Knowledge Management Process, 5(2): 1-11. https://doi.org/10.5121/ijdkp.2015.5201

[53] Novakovic, J.D., Veljovic, A., Ilic, S.S., Papic, Z., Tomovic, M. (2017). Evaluation of classification models in machine learning. Theory and Applications of Mathematics and Computer Science, 7(1): 39-46.