A Novel Framework For Disease Severity Level Identification Of Cotton Plant Using Machine Learning Techniques

Aurangzeb Magsi, Riaz Ahmed Shaikh, Zulfiqar Ali Shar, Rafaqat Hussain Arain, Asad Ali Soomro

Abstract: The World is moving with technological revolution. Computers are considered as the principal object in almost all the fields of life. In this concern, needs of biotechnology applications is immensely required to solve complex problems. Cotton plant is an important sector in the field of agriculture. Disease to that particular plant may also cause a loss to the agriculture sector. This paper is aims at dealing with cotton disease and its time based severity. Cotton plant is among those imperative plants which grow majorly in Pakistan and has a huge impact on its economy. Yield amount and quality of cotton plant is compromised every year damaging by some highly harmful diseases. Since, in this paper we presents a methodology to identify the severity level of a common and complex disease namely Cotton leaf curl Disease (CLCuD) by using methods of image processing and machine learning techniques. Color and texture features are used to extract values of an input image while deep learning method is use for decision making purpose. For experimentation process, a dataset of 1600 images is set. A Deep Convolutional Neural Network (CNN) is use for the classification. Cumulatively, 89.4% accuracy is received with the proposed model in term of proper identification and classification. This research work will be beneficial for local as well international harvesters and can be used to take time based preventive measures in order to reduce loss percentage.

Keywords: Biotechnology, Disease Identification, Machine Learning, Features Extraction, Image Processing, Cotton Plant, Artificial Intelligence

1 INTRODUCTION

Agriculture is considered as the backbone of a country. Pakistan's economy is highly depend on its agriculture sector. About 26% of its GDP is rely on its agriculture sector which include major productions such as Cotton crop, Wheat crop, Rice crop, Date fruit, Sugarcane, Vegetables and Fruits. Round the globe, among various other important commodities, cotton is one such vital product has importance in world-wide trading market and is grown in several countries of the world. About 150 countries around the globe are engaged in the import and export of cotton. It is one of the fiber and cash crop grow in Pakistan which earns largest foreign exchange of country [1]. Plants diseases and infections have been continuously a big challenge in the agriculture sector which causes aplenty yield loses in term of quality and quantity as well. These diseases to the plants are caused by distinctive sorts of parasites, microbes, phytoplasma, viroids etc [2]. This paper is aims at improving modern farming techniques with the use of advanced biotechnology software applications. Current study may allow farmers to detect and categorized the severity level of disease and Take preventive measures accordingly to improve yield ratio.

The advancement in the computer vision based systems with respect to the field of agriculture will overcome manmade interventions by introducing software applications which can perform the disease identification and its classification tasks. Such inventions will help farming sector to make time-based unbiased decisions in order to stop or cure infectious plants within time limit for the sake of improving farming quality and quantity. We use CNN in this research work for classification task which is preferable by most of the researchers for Fruit Identification tasks [3]. CNN use image database for training and testing purpose and performed disease stage classification accordingly [4][5].

In this research paper we take Cotton leaf curl disease (CLCuD) into consideration. CLCuD is a severer rather dangerous disease transmitted in plants by whitefly (Bemisia tabaci Gennadius). Cotton yields in Pakistan experienced a great loss in term of quality and quantity due to this CLCuD. The disease has been reported as occurring in various parts of Sindh mostly in Central and Lower parts. CLCuD cause extensive damages to the cotton field which resulting produce low quality cotton variety not approved by the cultivation authorities [6]. Fig.1 show the images of CLCuD infected leaves of cotton plant at various stages.

- *Aurangzeb Magsi is currently pursuing Phd degree in Shah Abdul Latif University, Khairpur, Pakistan. E-mail: aurangzeb.magsi.sef@gmail.com*
- *Riaz Ahmed Shaikh is Associate Professor in Shah Abdul Latif University, Khairpur, Pakistan. E-mail: riaz.shaikh@salu.edu.pk*
- *Zulfiqar Ali Shar is currently pursuing Phd degree in Shah Abdul Latif University, Khairpur, Pakistan. E-mail: zulfiqarali.shar@muetkhp.edu.pk*
- *Rafaqat Hussain Arain is Associate Professor in Shah Abdul Latif University, Khairpur, Pakistan. E-mail: rafaqat.arain@salu.edu.pk*
- *Asad Ali Soomro is currently pursuing Phd degree in Shah Abdul Latif University, Khairpur, Pakistan. E-mail: asad.soomro@salu.edu.pk*

***Fig. 1**. Disease infected leaves at various stages.*

2 LITERATURE SURVEY

A huge amount of work is done in the field of plant disease identification on various plants which include cotton plant as well using image processing and artificial intelligence based techniques. An automatic classification of Cotton root rot (CRR) disease is performed using unmanned aerial vehicle (UAV) remote sensing images [7]. The authors used Supervised, unsupervised, and combined unsupervised classification methods with the incorporation of two new automated classification methods to differentiate healthy plants and CRR diseased plants. The statistical comparison shows that the received results were 8.89% better than the results achieved using conventional classification methods. 88.5% accuracy acquired by combining K-means segmentation along with the advanced methods for classification. Similarly another study involves development of unsupervised algorithm of Plant-by-Plant (PBP) classifier for the purpose of delineation of CRR-infested area using five-band multispectral data via UAV based remote sensing images [8]. K-means clustering and superpixel segmentation shaped PBP classifier. The authors received 77.48% accuracy with unsupervised two-class k-means classifier. However, semi-supervised method provide 84.12% and supervised learning provide 87.72% accuracy. Comparing these conventional regional classification methods, PBP provide 92.1% as overall accuracy. Various methods and concepts have been adopted by scholars to enhance the field of agriculture using computer vision and software computing techniques [9]. Lot of research done on Plant Pathology using computer vision based methods along with the coordination of machine learning techniques on the account of increase plant growth and quality production. Digital Image processing in this perception plays a vital role which extract the image information for computational measures. An attempt to identify three cotton plant leaf diseases was done by proposing pattern recognition system. Image segmentation process done using active contour model while the adaptive neuro-fuzzy inference system were trained using Hu's moments as the extracted features by the scholars. Delaying in the plant disease identification can cause excessive and poor use of various chemicals which directly increase the overall cost of production. Hence, a pilot dataset is produced using 3651 images of infected apple fruit with real-life symptoms of different disease [10]. Authors created pilot dataset for Cedar Apple Rust and Apple Scab along with images of healthy leaves with the help of local expert available at workshop Computer Vision and Pattern Recognition 2020. Using this data, an off-the-shelf Convolutional Neural Network (CNN) was created for the process of classification through which scholar attained accuracy of 97%. The significance and authentication of the developed method is measured form 1500 entries submitted to the developed pilot dataset so far with high accuracy rate. Furthermore, a Deep Learning based algorithm known as Few-Shot Learning (FSL) algorithm is proposed for the plant disease identification using a small datasets namely PlantVillage comprising of 54303 labeled images of 38 different plant leaf diseases [11]. Researchers defined proposed algorithms is classified in to two blocks where first block deals with the CNN based image classification which performed a leaf image map. While the second block comprised on a Support Vector Machine (SVM) classifier and deals with the separation of the classes of plant leaves. Each class contained 80 images of 6 different diseases infected leaves. With this data, the trained CNN attained 90% accuracy. Similarly, another pattern recognition system was proposed by Rothe. Author performed classification task on three cotton leaf diseases namely [12]. Authors captured the images from the fields of selected districts manually. For training purpose of neuro-fuzzy inference system, scholars encouraged the method of invariant moments and used Hu's moment in the form of extracted features along with Active contour model for segmentation process. Seven Hu's moments are defined in the proposed model which were, in term of central moments, invariant to the object's rotation, scale and translation and MTLAB programming language was used for calculations. Classification was done using Feed Forward Back Propagation method by updating its values of Weights and Bias. 5000, 10, 1x10-10, 0.001 and 1x1010 were set as a maximum number of epochs, maximum validation checks, minimum gradient, minimum Mu and maximum Mu respectively. The trained network provide 85% of accuracy in the classification. Moving forward, in the field of fruit disease detection, another researchers named Sanjeev [13] proposed a model of detection comprised on two well-known grape fruit disease i.e. Downy Mildew and Powdery Mildew. Researcher standardized the input image at 300x300 size. For the sake of preservation of chief information of an input image i.e. infected portion, scholar used five iterations of Anisotropic Diffusion in order to enhance the input image. Grey Level Co-occurrence Matrix

(GLCM) was used to extract features of an image while the segmentation process was performed using K-means clustering technique. 100% accuracy was achieved with the proposed model in training using Neural Network based Back Propagation method. Another research on the leaf disease identification was performed by Asma Akhtar [14]. Scholar used Rose leaf as an object and identified Anthracnose and Black Spots diseases. Regarding features extraction, researcher extracted eleven haralick texture features along with the DCT and DWT features. Thresholding technique was used for image segmentation process whose values was resolute using Otsu's algorithm and classification was performed by SVM which provide 94.45% accuracy as a whole. Alexandre A. also proposed a system for the cotton crop disease identification [15-17]. To extract the color feature of an input image, researcher decomposed image in to the color channels such as R, G, B, H, S, V, I3a, I3b, and grey levels. Likewise [18], DWT also applied in this research work on all the color channels of an image. To combine the feature vectors, the wavelet energy need to be computed for each sub-band individually. Acceptable results received by the scholar using SVM for classification. Similarly, Jie Tian developed an automatic system which can identify as well as classify disease infected leaves of apple fruit [19]. Researcher work on mosaic virus, rust and leaf spot diseases. Features extracted in the research work were color and texture features of an image of apple leaf. The classification accuracy received by researchers by implementing KPCA and GASVM model was found maximum in the comparison of the PCA and GA-SVM model.

3 RESEARCH METHODOLOGY

The identification and classification of CLCuD severity level is focused in this research which can be performed by doing some random operations. Image acquisition is the initial step in most of the image processing based systems which then processed as an input to the system. For image normalization, we perform some pre-processing operations after which extraction of features of the input image took place. Outcomes of feature extraction phase were used for statistical analysis which ultimately used for the task of identification and classification. Proposed methodology is depicted in Fig. 2.

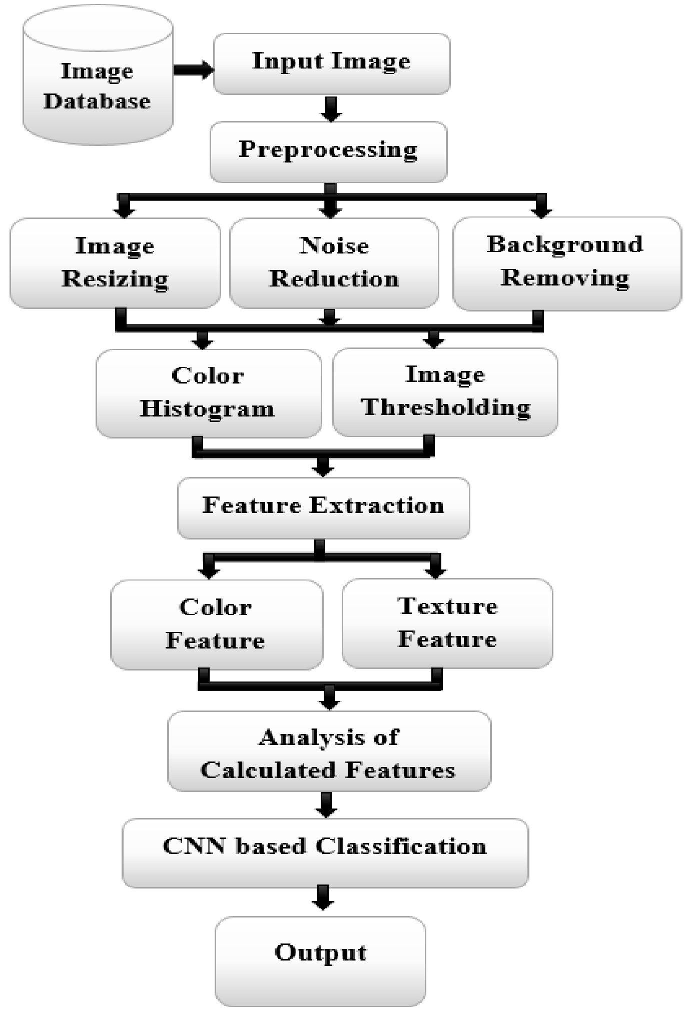


***Fig. 2**. Research Methodology*

Image input is the next step after image acquisition. All the selected images of diseased cotton leaves were set in the form of database. Some basic tasks performed on the input as a pre-processing operations which include noise reduction, image resizing and removal of background. These operations were performed on the input image in order to make image suitable for next processes. These preprocessing operations are performed in order to define standard protocols such as noise reduction operations remove the unwanted data appear in the image during image acquisition process. Similarly to maintain image uniformity, some image resizing operations performed on the input image to reduce processing time. While background removing operations deletes all the additional objects except one key object which is the origin of the research. In order to generate the color histogram, image Thresholding technique is used to separate region of interest from the input image. Values generated from features extraction and color histogram both will be used for classification process. Following preprocessing operations, features extraction is the next step of the research. Color, Size and texture features were extracted [20]. Color feature provides huge information in the image processing operations. Hence, using MATLAB built-in tools for image processing, HSV color space is sued to segment the infected part of the cotton leaf from the healthy part and generate color values of both parts with respect to the whole image [21]. Furthermore, Scale Invariant Feature Transformation method is used for the sake of infected leaf part calculation, measure the size of the yellowish/pale part of the leaf image with respect to the whole image. Texture feature with the additional support of morphological features were then used to extract the information like roughness and smoothness of both infected as well as healthy part of leaf image [22]. Statistical analysis made on the generated values form features extraction. Identification and classification process was done on the basis of the statistical analysis and associated CNN.

4 DATA COLLECTION

Pakistan is the 4th largest cotton producer in the world. This production takes place every year between the month of May and August. Hence, leaf images for experimentation collected manually at first hand during these months. Based on the disease stage, images of infected leaves were collected in various months which show overinflated disease level ranges from initial stage to the extreme stage. To achieve better results during experimentation, standard level images are required. Therefore, a Digital Single-Lens Reflex (DSLR) camera was used for image acquisition. To avoid luminosity problem, images of cotton leaves were captured in various timings such as some images were taken in morning time while some of them were taken at the mid of the day similarly some were taken in evening time. 120,000 lux luminance standard of the captured images was assured and maintained image quality at 300 dpi. A database was designed where 1600 captured images of best quality were stored. The dataset contained infected leaf images at various stages such as less infected leaves, moderately infected leaves and highly infected leaves. Among all the captured images, 400 images of each stage of diseased leaves were set for experimentation. Table 1 show the statistical information of the dataset.

TABLE 1.
FIGURES OF TOTAL CAPTURED IMAGES AND SELECTED IMAGES

Category	Captured	Selected
Stage–1	510	400
Stage–2	614	400
Stage–3	740	400
Stage–4	1136	400
Total	3000	1600

5 IMPLEMENTATION

1600 selected images of infected leaves were taken into experimentation process.

5.1 PREPROCESSING

Preprocessing is the first step based on three tasks such as noise reduction, background removing and image resizing. Background removing is the first among the three mentioned preprocessing operations where image of key object was separated from the whole image which is the mandatory part before taken images into further processing process. Fig. 3 shows the output image after applying background removal operation.

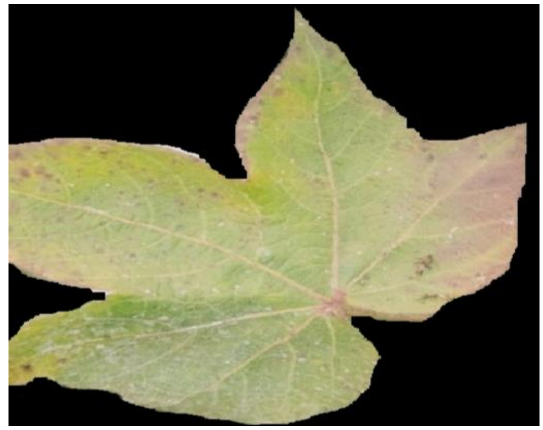
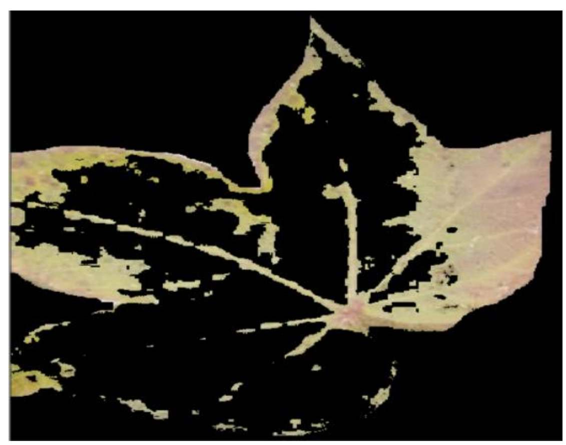


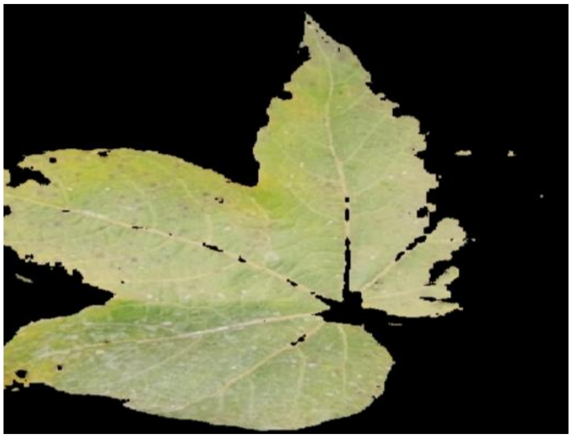
Fig. 3. *Image without background*

5.2 IMAGE THRESHOLDING

After applying background removing operation, image Thresholding was performed on the input image for color values extraction. In order to manifest infected part of leaf, green part of the leaf was extracted and calculated values of other left other colors. Values of other colors with respect of the value of entire image let us know about the infected part of leaf. To identify the disease severity level, we extracted yellowish part of the leaf and calculated the remaining dried part only through which infection percentage calculated and received results were stored in to the database for future use of classification task. “Fig. 4” is the depiction of Image Thresholding operation.



(a)



(b)

Fig. 4*.(a)(b) Image Thresholding process*

5.3 FEATURE EXTRACTION

Irrespective to the color feature extraction performed in Thresholding, using MATLAB built-in tool for features extraction, color values of leaf image was extracted. In this feature extraction, Red, Green and Blue color component values of each pixel of image was extracted and calculated ratio individually regardless of difference of the basic or secondary colors and calculated RGB ratio. "Fig. 5" is the representation of color feature extraction.

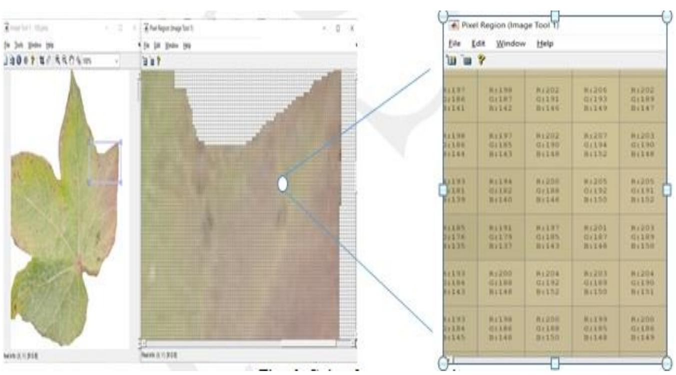


Fig 5*. Color feature extraction*

After the color feature extraction, we extracted the texture features of the input image which require a gray scaled image or this extraction, we use method of "grey level co-occurrence matrix approach (GLCM)" [8]. GLCM require a grey level image to be processed. Hence, an input image was converted in to its equal grey level and then processed. With the help of this feature, an area of infected part of leaf in the comparison of the healthy part was calculated using extracted edges of an image. Adding the results of texture feature to the color feature extraction increased the overall accuracy of the system. The extraction of the Texture feature of an image is represented in "Fig. 6".

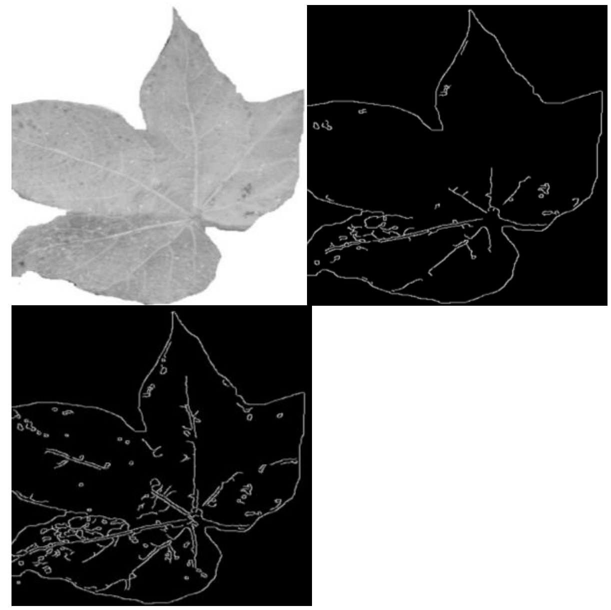


Fig.6. Texture feature extraction

6 IDENTIFICATION & CLASSIFICATION USING CNN

All the extracted values of color and texture features were used by CNN for computation and analysis. For identification and classification we used a predefined dataset. CNN, with respect to these predefined values, make identification and classification of disease stage. MATLAB provides facility with built-in tools for Neural Network (NN) instead of go through the complex and time taken programming. We use NN for identification and classification process. Therefore, all the extracted values undergo statistical analysis which then used as an input to the CNN [23]. The developed network were trained and tested on these values. For training the network, we use 863 images and for testing 337 images were used. The network was trained to work with the basic factors like average values R, G and B, Color histogram and texture values, the system generates the ultimate output in the form of identification and classification. Fig 7 is the depiction of performed experiments on developed network. While Fig.8 shows identification and classification process.

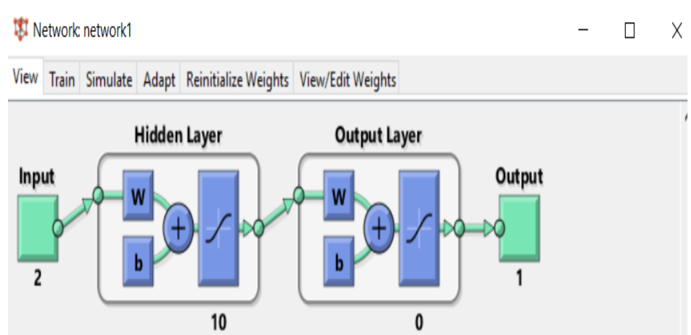


Fig.7*. Developed neural network*

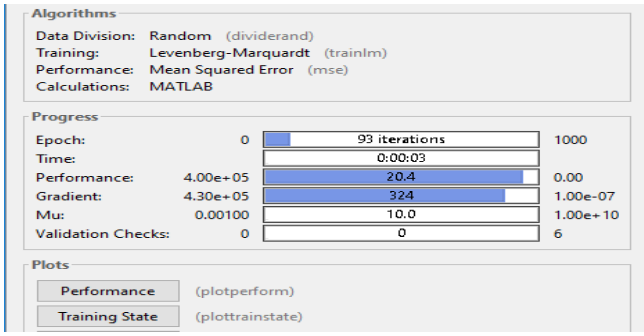


***Fig. 8**. Classification process using trained NN*

Keeping in view disease stage, images classified as stage 1, 2, 3 and 4 and all the selected images were taken in to experimentations. The outputs of the experiments show that disease at 1st stage is hard to be identified due to minute variances in color as well as texture values. Hence, it requires more research work in order to exact identification. Results received at Stage 2 and 3 are quit at satisfactory level while best results generated by the system are at Stage 4 disease. Best validation performance at stage 4 is depicted in "Fig. 9". The overall performance the system gives acceptable results as R=0.99 at 370th epoch of the iteration process and is displayed in "Fig. 10".

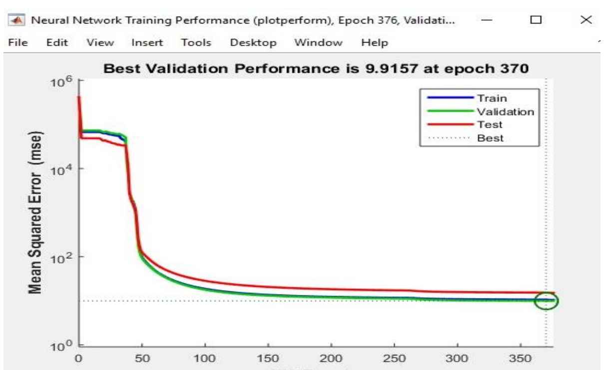


***Fig. 9**. Validation performance at stage 4*

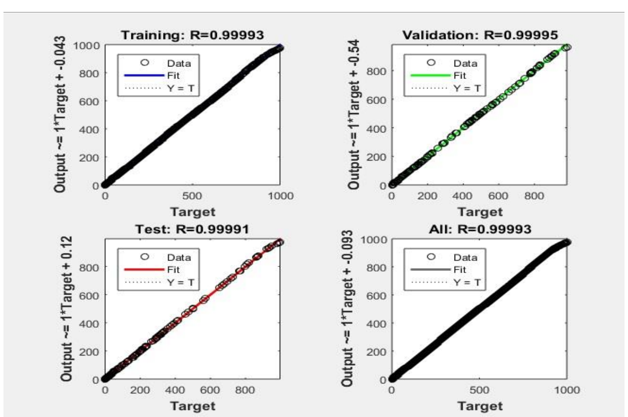


***Fig. 10**. Overall system performance*

Based on the number of disease stages, results were also classified in to the four stages. From the 400 selected leaf images of stage-1 disease, we received 72.3% valid results. It shows that out of 400 images, system successfully classified 289 images of stage-1 disease which is less than other stages but is quit acceptable. The system was unable to process remaining 111 images of stage-1 and the possible reason was noise and other compatibility issues. However, 364 images of stage-2 out of 400 were successfully identified and classified by the system which is 91% accuracy in results. Stage-3 provide 95.6% accuracy in results where system successfully identified and classified 382 images. High accuracy of 99% is gained with stage-4 images. The model successfully identified 396 images of stage-4 out of 400 selected images. "Fig. 11" is the complete depiction of results generated by the system. Using developed and trained network, identification and classification process was done and the overall accuracy in results generated by the system was 89.4% which is acceptable. However, results at stage-1 and stage-2 were less accurate as compared to the results at stage-3 and 4. Furthermore, number of features to be extracted can be increased as a future work to this research along with the addition of classification techniques may provide much accurate and quality based results.

7 DISCUSSION AND CONCLUSION

The main determination of this research was to propose an automatic system which identify the plant disease and classify its stage or severity level. For this reason we selected a worldwide affecting disease of cotton plants which causes huge loss to cotton yields. The proposed model achieved the said task with the help of some additional functions. The proposed model is entirely based on the combination of image processing and machine learning tools and techniques. To achieve the task of disease severity identification, we used feature extraction techniques and the process classification was done through deep learning methods. For experimentation, we captured 3000 images of infected leaves of cotton plant. Images of infected leaves were captured at various stages of disease. All the captured images were set into the database.

***Fig. 11**. Stage-wise results*

Initially, some preprocessing operations were performed on all the images such as noise reduction, image resizing and background removing. Those operations performed in order to make uniform image for experimentation. MATLAB R2015 version software was used for experiments. During the feature extraction phase, color, texture and morphological features were extracted from the input image. Statistical analysis were made on the outputs received from the feature extraction which with the collaboration of CNN used for decision making process such as identification of severity level of disease and its classification. At stage- 4 system provide maximum output in the form of accuracy as it specifically identified and classified 99% images. While disease at stage-3 was identified and classified by the system as 95.6% and stage-2 was 91%. However, disease at initial stage were difficult to be identified due to minute variances in features and system identified only 72.3% images of stage-1 disease. The proposed system is able to be executed at local and global level for the purpose of plants disease identification processes. However, in future system may be improved in term of accuracy by increasing the number of features and classification techniques. Also current system works only with single image of plucked leaf. Hence, in future system may able to work with bunch of images concurrently.

REFERENCES

[1] S. Abbas, “Climate change and cotton production: an empirical investigation of Pakistan”, Environmental Science and Pollution Research 27, 2020, p. 29580-29588.

[2] H. Zhao, C. Yang,.W. Guo, L. Zhang and D. Zhang, “Automatic estimation of crop disease severity levels based on vegetation index normalization”, Remote Sensing 12, no. 12, 2020, p.1930.

[3] W. Jia, Y. Tian, R. Luo, Z. Zhang, J. Lian, and Y. Zheng., “Detection and segmentation of overlapped fruits based on optimized mask R-CNN application in apple harvesting robot”, Computers and Electronics in Agriculture 172, 2020, p.105380.

[4] A.Johannes, A. Picon, A. Alvarez-Gila, J. Echazarra, S. Rodriguez-Vaamonde, A. D. Navajas, and A. Ortiz-Barredo., “Automatic plant disease diagnosis using mobile capture devices, applied on a wheat use case”, Computers and electronics in agriculture 138, ,2017, p. 200-209.

[5] K. Simonyan, and Z.Andrew, “Very deep convolutional networks for large-scale image recognition”, arXiv preprint arXiv 2014,p.1409.1556

[6] L. Amrao, S.Akhter, M. N. Tahir, I. Amin, R. W. Briddon, S. Mansoor., “Cotton leaf curl disease in Sindh province of Pakistan is associated with recombinant begomovirus components”, Virus research 153, no. 1, 2010, p. 161-165.

[7] T.Wang, J. A. Thomasson, C. Yang, T. Isakeit, and R. L. Nichols., “Automatic classification of cotton root rot disease based on UAV remote sensing”, Remote Sensing 12, no. 8, 2020, p. 1310.

[8] T.Wang, J. A. Thomasson, T. Isakeit, C. Yang, and R. L. Nichols., “A Plant-by-Plant Method to Identify and Treat Cotton Root Rot Based on UAV Remote Sensing”, Remote Sensing 12, no. 15 , 2020, p. 2453.

[9] D. Krisnandi, H. F. Pardede, R. S. Yuwana, V. Zilvan, “Diseases classification for tea plant using concatenated convolution neural network“, CommIT (Communication & Information Technology) Journal, vol. 13, no. 02, pp. 67–77, 2019.

[10] R. Thapa, , S. Noah, B. Serge, and K. Awais, “The Plant Pathology 2020 challenge dataset to classify foliar disease of apples”, arXiv preprint arXiv:2004., 2020, p.11958

[11] D. Argüeso, P. Artzai, I. Unai, M. Alfonso, M. G. San-Emeterio, B. Arantza, and A-Gila.Aitor Alvarez-Gila. "Few-Shot Learning approach for plant disease classification using images taken in the field”, Computers and Electronics in Agriculture 175, 2020 p.105542.

[12] Rothe, P. R., and R. V. Kshirsagar., “Cotton leaf disease identification using pattern recognition techniques”, In 2015 International conference on pervasive computing (ICPC), IEEE, 2015, p. 1-6.

[13] S. S Sannakki, V. S Rajpurohit, V B Nargund, P. Kulkarni4, “Diagnosis and classification of grape leaf diseases using neural networks”, In 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), IEEE, 2013. pp. 1-5.

[14] A. Akhtar, A. Khanum, S. A. Khan, A. Shaukat, “Automated plant disease analysis (APDA): performance comparison of machine learning techniques”, In 2013 11th International Conference on Frontiers of Information Technology, IEEE, 2013. pp. 60-65.

[15] A. Alexandre, Bernardes, G. Jonathan., Rogeri, B Roberta, Oliveira, N. Marranghello, A. S. Pereira, A. F. Araujo and R. João Manuel R, S. Tavares, “Identification of foliar diseases in cotton crop”, In Topics in Medical Image Processing and Computational Vision, Springer, Dordrecht, 2013. p. 67-85.

[16] J. TIAN, Q. HU, X. MA3; M. HAN, “An improved KPCA/GA-SVM classification model for plant leaf disease recognition”, Journal of Computational Information Systems 8, no. 18 , 2012, p. 7737-7745.

[17] S. P.Mohanty, D. P.Hughes. and M. Salathé, “Using deep learning for image-based plant disease detection”, Frontiers in plant science 7, 2016 p.1419.

[18] 15. Shaikh, R.A.; Memon, I.; Arain, R.H.; Maitlo, A.; and Shaikh, H. (2018). A Contemporary Approach for Object Recognition Using Spatial Layout and Low Level Features’ Integration”, Multimedia Tools and Applications

[19] P. R. Rothe, R. V. Kshirsagar, “Automated extraction of digital images features of three kinds of cotton leaf diseases”, In 2014 International Conference on Electronics, Communication and Computational Engineering (ICECCE) IEEE,, 2014, p. 67-71.

[20] Revathi, P., and M. Hemalatha, “Classification of cotton leaf spot diseases using image processing edge detection techniques”, In 2012 International Conference on Emerging Trends in Science, Engineering and Technology (INCOSET),. IEEE, 2012. p. 169-173

[21] Tian, Chunwei, Lunke Fei, Wenxian Zheng, Yong Xu, Wangmeng Zuo, and Chia-Wen Lin., “Deep learning on image denoising: An overview”, Neural Networks, 2020.

[22] Hussain, H.; Gao, H. and Shaikh, R.A. (2016). Segmentation of Connected Characters in Text-Based CAPTCHAs for Intelligent Character Recognition. Multimedia Tools and Applications

[23] L. D. Apostolopoulos, T. A. Mpesiana, “Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks”, Physical and Engineering Sciences in Medicine 43, no.2, 2020, p.635-640.