Mango Plant Leaf Disease Detection and Recognition using Neural Networks and K-Means Clustering Techniques

Chandru Jathar[1], Dr. Janapati Venkata Krishna[2], Meghana G. R.[3]

[1]*Research Scholar, Department of CS&E,College of Engineering and Technology, Srinivas University, Mangalore, India & Assistant Professor Department of Master of Computer Applications, KLE-Technological University, Hubballi, Karnataka, India*

[2]*Associate Professor, Department of Computer Science and Engineering, Institute of Engineering and Technology, Srinivas University, Mukka, Mangalore, Karnataka, India*

[3]*Assistant Professor, Department of Information Science and Engineering, Jain Institute of Technology, Davanagere, Karnataka, India*

Abstract: Smart agriculture contributes a lot to the global economy. One of the serious problems that are faced by the agriculturist is identifying the plant disease. Many agricultural solutions are evolving to assist the farmers in improving their crop production. With the advancement of deep learning & neural network many models have emerged to identify the plant diseases. The proposed work presents identifying the illnesses in tapioca and mango leaves. The methodology divides the leaves and stem into two classes as infected and not infected. K-means clustering and Neural Networks are used to classify and group the plant diseases. The working model identifies the five plant related diseases such as early scorch, cottony mould, ashen mould, late scorch, and small whiteness. The outcomes of the experiment suggest that the proposed methodology is beneficial and may greatly aid in the precise identification of leaf diseases with minimal computational work by providing about 95% of efficiency.

Keywords: *Leaf disease, K-means clustering and Neural Networks, RGB, HIS.*

Introduction

For many countries the primary source of their national income is through agriculture. Crop diseases are extensively reducing the quality and quantity of the yield being produced. So it is very important and necessary to identify the plant diseases and take necessary measures to improve the yield [1]. The disease symptoms arise in different parts of the plant. Plant leaves are commonly used to identify the diseases occurred. Accurate and early detection of the diseases will significantly reduce the propagation of disease and leads to increase in the yield [2-5].

India stands in the second place for producing the horticultural crops like papaya, banana, mango, grapes, cashew, okra and potato. 104.17 million tons of horticultural crops are produced by India every year. Plants must be protected from these diseases in order to produce the quality food. There are various factors that affect the food quality like water supply quality, change in climate, plant diseases, degeneration in pollinators etc. To yield quality and value added products it is very necessary to identify these diseases at the early stages.[6-8] According to the reports of Food and Agriculture Organization (FAO) of the United Nations around 40% of the worldwide crops are destroyed every year because of pests and plant diseases. Therefore it is very important to minimize these plant diseases in the early stage to improve the overall quality of the food.[9]

Accurate and timely disease treatment is required for reducing the plant disease propagation. Earlier agricultural experts were involved in identifying the disease, but this method was time consuming. Recently Artificial intelligence (AI) and Convolution Neural Network (CNN) solutions are incorporated to classify the leaves based on the diseases [10]. CNN is used in Deep Learning (DL) methods where image related tasks can be handled easily with reduced computational complexity. CNN is efficiently used for image pattern recognition where training time is shortened and requires few neurons [11-12]. CNN can easily classify the image dataset by extracting features and processing it in different layers. Weights are updated from the hidden layer and images are classified with the activation function from the final layer. DL models can be trained to identify all the different types of plant diseases which may not be possible to detect from the farmers through naked eyes[13-14].

This work proposes a solution that can automatically recognize and categorize the plant leaf diseases. It also provides a suggestion to stop propagation of the plant disease based on individual plant condition. The remaining sections are organized as section 2 discusses the related work that is carried out by many other researchers. Section 3 discusses the system model. The proposed system is discussed in section 4. Results of the proposed system are discussed in the section 5 and finally the conclusion of the research work is mentioned in the section 6.

1 Related Work

A Learning Vector Quantization (LVQ) based algorithm is proposed by Melike Sardogan et.al [15] for tomato leaf disease detection and classification. The author has trained the model by taking 500 images dataset. RGB component model is used to apply filter to the three channels. LVQ is used for training the output features of the convolution network. The results in this system effectively recognize the four types of tomato leaf diseases.

Falaschetti et.al[16] proposes a low cost, low power real time plant classification system with the images provided. The ESCA-dataset and the PlantVillage- data set are used to train the CNN Model. Camera is used here for real time image possession and classification. Upon classification the real time results are displayed on the LCD display.

H. Ajra et.al[17] proposes a technique which uses a kaggle dataset to identify the unhealthy potato and tomato leaves. AlexNet and ResNet-50 models are used for feature extraction, classification of the unhealthy leaves and for detecting the diseases. The system also provides a graphical layout about the disease prevention measures. High computation and more parameters are required by the standard CNN models.

Hassan et.al [18] proposes a technique that replaces the standard CNN model with depth separable convolution where the number of parameters and computation cost is reduced. Color, segmented and gray images are used in this model to examine the robustness. The model when compared with other existing models achieved greater accuracy and reduced training time.

Khattab et.al [19] proposed an IoT based system where the sensors are deployed to collect the environment data of the place where the plants are led. Prediction algorithm is written which predicts the crop health based on the environment data collected. As the plant diseases are identified the artificial intelligent system sends a signal to the user about the occurrence of the plant disease. The system also identifies and signals the farmer about the correct amount of pesticides to be sprinkled on the disease causing plants.

Srdjan Sladojevic et.al. [20] Proposed a deep CNN model for leaf image classification and identification of the plant disease. The model is trained with various leaf images so that it can identify the different leaf features. CNN model is used to learn the different parts, shape and features of different leaves.

2 Proposed Architecture

The real time images of various leaves in our surrounding are collected. These collected real time images are then examined to understand their features and characteristics. Various analytical discriminating methods are then applied to categorize these images based on the issues faced.

In the initial phase of image acquisition, different leaves images that need to be classified are collected with the help of a camera. The images are prepared and preprocessed in the second phase. In the third phase k-means clustering algorithm is used to identify the different leaves segment. The diseased leaves feature extraction is done by extracting the unique characteristics and texture of each pixel from the image. In the later stage a statistical analysis is carried out in order to choose the feature that best describes the affected leaves. In the final stage CNN algorithms are applied to categorize these images as affected and not affected leaves.

Image classification and acquisition process details are depicted in figure 1. In the initial phase the RGB images of the leaf samples and affected leaves are captured. In the later stages the RGB image formats are then converted to the HIS format. The SGDM matrices are produced for every pixel in the H & S images. SGDM is used to measure the distance and orientation between the two pixels in the images. The system is trained with the test data so that it can carry out the classification task.

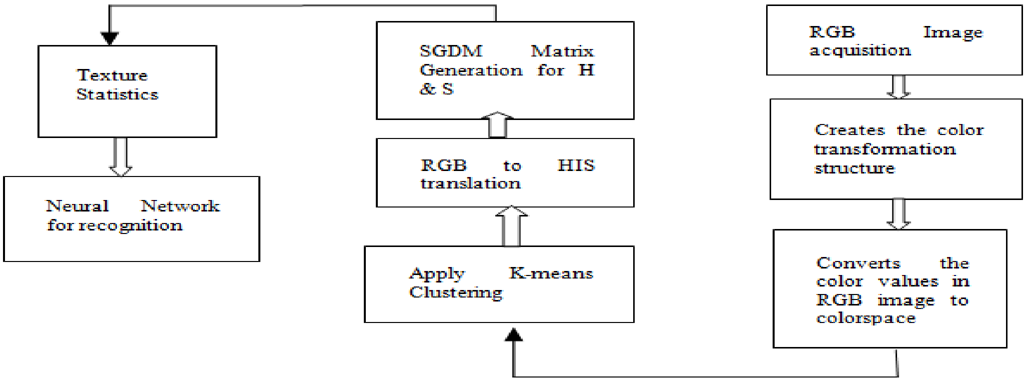

Figure 1: Image acquisition and classification model

An image processing based model is being proposed that is capable of identifying and classifying the different leaf diseases. The system is evaluated by considering the different plant illnesses such as cassava bacterial blight, cassava leaf spot, phoma blight, bacterial canker, red rust and sooty mould. The system also specifies the treatment required for the infected plants.

The model will first identify the type of sickness in the plant, identifies about the plant disease in the early stage. Then the affected leaves are categorized according to the different diseases. Finally the system also provides guidance to the disease so that early treatment of the disease can be initiated. The proposed system works in different phases. These phases are discussed in this section

3.1 Image Acquisition: In this work first stage is need to collect the various datasets of mango leaf infections such as Anthracnose, Bacterial black spot, Twig blight, Gummosis, Bark splitting from the digital camera.

2.1 Pre-Processing: In this phase the images are pre-processed and noise, lighting variations, climatic conditions, poor resolutions and unwanted background are removed so that the identifying of diseases can be done easily.

2.2 Color Transformation Structure: In this phase RGB images are converted to HIS Color space. After conversion only the H (hue) component is considered for further analysis and S (saturation) & I(intensity) components values are dropped.

2.3 Masking Green Pixels: In this phase Green pixel are identified and masked. If the green color pixel intensity is greater than the predefined threshold value than the RGB colored pixel values is assigned a 0 value. This masking is done because green color usually represents healthy part of the leaf so when it is masked it reduces the preprocessing of the healthy leaf part.

2.4 Removing the masked cells: This phase identifies the 0 marked RGB pixels and completely removes them so that accurate disease classification and identification can be performed.

2.5 Segmentation: This phase partitions the leaf into 4 clusters. K-means clustering algorithm is used to classify and the cluster the objects based on features of k classes. The classification is done by minimizing the sum of squares of distances between the objects and the corresponding cluster or class Centroid.

2.6 Feature Extraction: Color Co-occurrence method (CCM) is used for feature extraction. Here both the color and texture of an image is considered to arrive at the unique features.

2.7 Color Co-occurrence method for texture analysis: The CCM methodology consists of three major mathematical processes. First, the RGB images of leaves are converted into HSI color space representation. Once this process is completed, each pixel map is used to generate a color co-occurrence matrix, resulting in three CCM matrices, one for each H, S and I pixel maps. Color spaces can easily be transformed from one to another. Following equations can be used to transform the images from RGB to HSI.

$$Intensity\ (I) = \frac{R+G+B}{3} \quad \text{--------------------------------}(1)$$

$$Saturation\ (S) = 1 - \frac{3\min(R,G,B)}{(R+G+B)} \quad \text{--------------------------------}(2)$$

$$Hue\ (H) = 2 - ACOS\left\{\frac{[(R-G)+(R-B)]}{\sqrt[2]{(R-G)^2+(R-G)(G-B)}}\right\},\ B > G \quad \text{-----------------}(3)$$

$$Hue\ (H) = ACOS\left\{\frac{[(R-G)+(R-B)]}{\sqrt[2]{(R-G)^2+(R-G)(G-B)}}\right\},\ B \leq G \quad \text{------------------------}(4)$$

For the position operator p, we can define a matrix P_{ij} that counts the number of times a pixel with gray-level i occurs at position p from a pixel with gray-level j. SGDM's are generated for H image. The SGDM's are represented by the function P(i,j,d,θ) where i represent the gray level of the location (x, y) in the image I(x, y), and j represents the gray level of the pixel at a distance d from location (x, y) at an orientation angle of θ. The reference pixel at image position (x, y) is shown as an asterix. All the neighbors from 1 to 8 are numbered in a clockwise direction. Neighbors 1 and 5 are located on the same plane at a distance of 1 and an orientation of 0 degrees. In this research one pixel offset distance and a zero degree orientation angle is used. GLCM function in MatLab to create Gray-Level Co-Occurrence Matrix. The number of gray-levels is set to 8 and the symmetric value is set to true and finally offset is given a 0 value.

Algorithm used in this model:

When a new image inserted into the system that new data would be compared with trained data.

The steps of workflow are listed below as per the architecture is shown in Fig. 1.

Step 1: Normal and infected leaves with different symptom were collected.

Step 2: Normal and infected leaves were grouped apparently.

Step 3: MATLAB were used to read data from collected leaves.

Step 4: Color spaces were specified through HSV.

Step 5: After specifying HSV, color feature was extracted from images or regions.

Step 6: Texture feature was then applied for a wide range of image to eliminate.

Step 7: Images were filtered with shape features to encode simple geometrical forms.

Step 8: Images were segmented in different regions.

Step 9: Finally, diseases were detected by matching.

2.1 Normalizing the CCM Matrices:

The CCM matrices are then normalized using the equation below, where, $\boldsymbol{p(i,j,1,0)}$ represents the intensity co-occurrence matrix:

$$\mathbf{p(i,j)} = \frac{\mathbf{p(i\,j,1,0)}}{\sum_{i=0}^{N_g-1}\sum_{j=0}^{N_g-1}\mathbf{p(i,j,1,0)}} \text{------------------------}(5)$$

where N_g is the total number of intensity levels. Next is the marginal probability matrix:

$$\boldsymbol{p_x}(i) = \sum_{j=0}^{N_g-1}\boldsymbol{p(i,j)} \text{ ------------------------}(6)$$

Sum and differences matrices:

$\boldsymbol{p_{x+y}(k)} = \sum_{i=0}^{N_g^{-1}} \sum_{j=0}^{N_g^{-1}} \boldsymbol{p(i,j)}$ ------------------------*(7)*

where k = 1+j, for k = 0, 1, 2, ..., 2(N_g -1) and

$\boldsymbol{p_{x-y}(k)} = \sum_{i=0}^{N_g^{-1}} \sum_{j=0}^{N_g^{-1}} \boldsymbol{p(i,j)}$ ------------------------*(8)*

where k = |I-j|; for k= 0, 1, 2,..................,2(N_g -1) and p(i. j) is the image attribute matrix.

2.2 **Neural Network for Recognition:** The proposed system uses neural network model to for automatic detection of leaves disease. The model is trained and tested with the valid data set. The capability of ANN model to respond accurately was assured using the Mean Square Error (MSE) criterion to emphasis the model validity between the target and the network output.

3 Results and Discussion

In this section the results and screenshot of the plant disease detection system are discussed.

The figure 2 depicts the front screen of the system where the user are allowed to select the image for disease identification and then clustering, feature extraction etc can be performed.

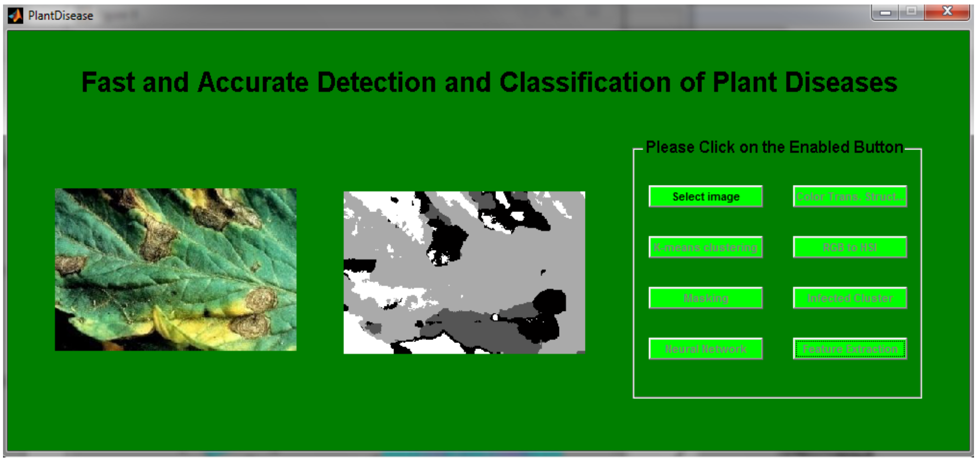

Figure 2: Front end of the system

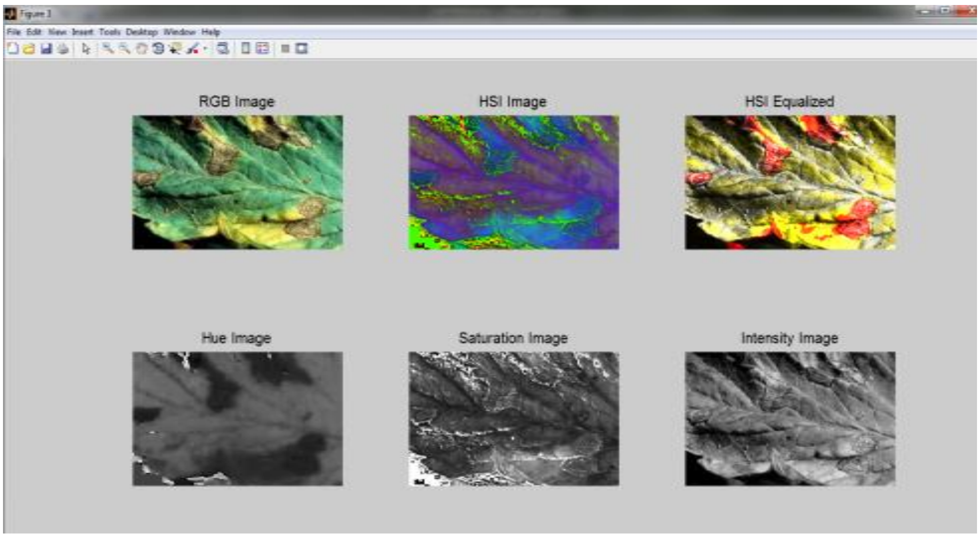

Figure 3: RGB to HIS images

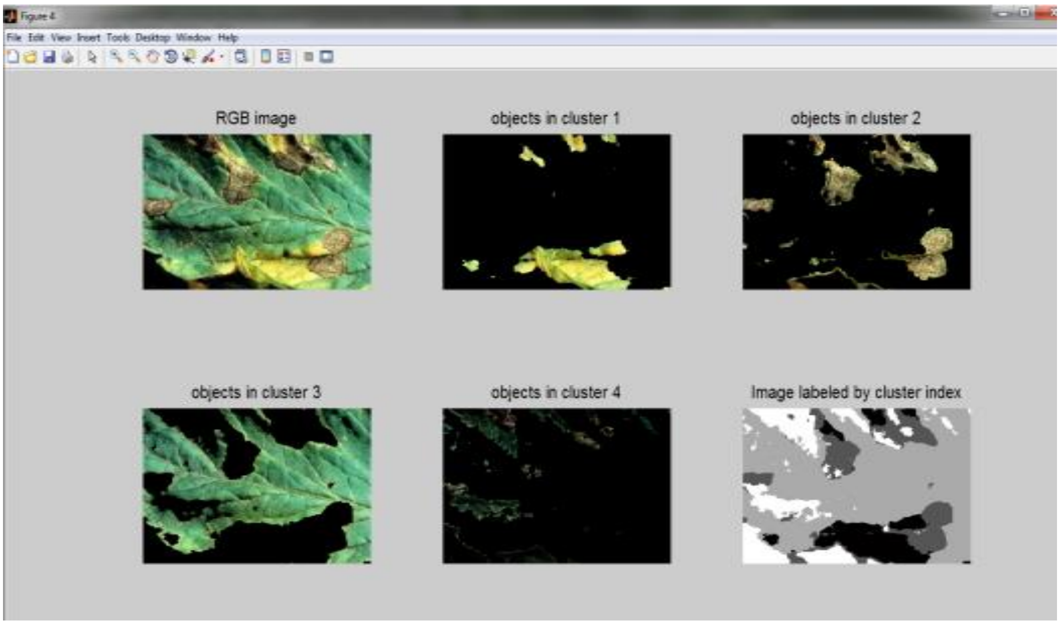

Figure 4: kmeans clustering

Figure 3 depicts the system where the RGB images are converted to the HIS images and the figure 4 shows the output when the kmeans clustering is applied to the images and the images are grouped into different clusters.

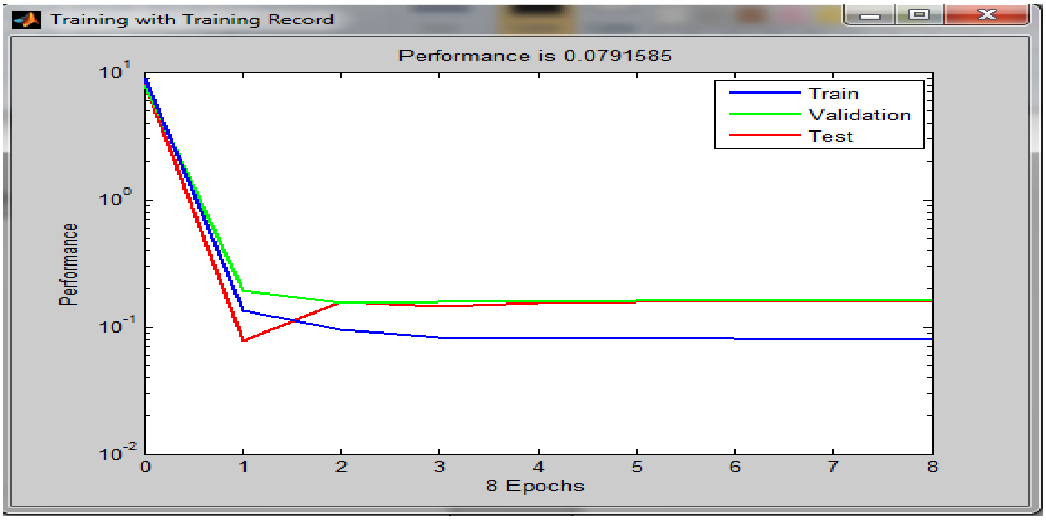

Figure 5: Performance analysis in terms of Train Validation and Test

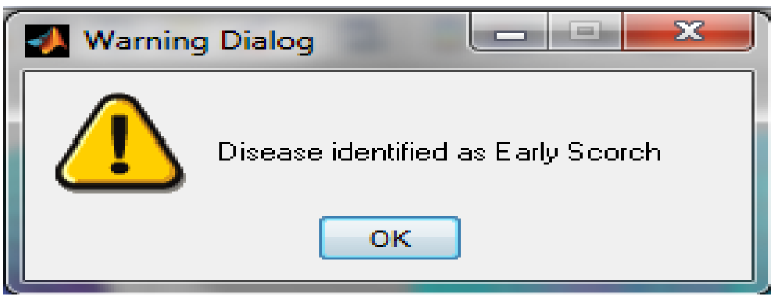

Figure 6: Warning Dialog shows the type of leaf disease for the test data

Figure 5 Shows the graphical illustrations of Performance analysis in terms of Train, Validation and Test. The final results of the type of leaf infection are illustrated on warning dialog box as shown in Figure 6. Figure 7 shows the comparison chart of the other algorithms with the proposed model.

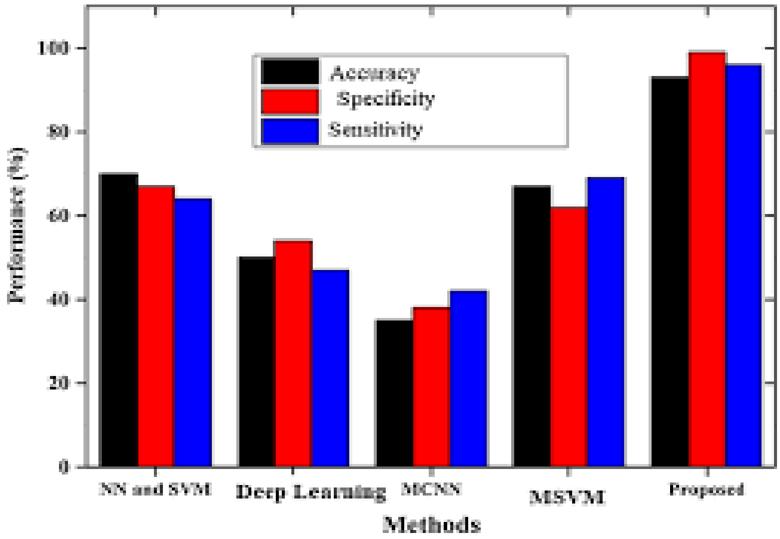

Figure 7 : Comparison chart of the other algorithms with the proposed model.

4 Conclusion

This research work demonstrates that the formulation of Neural Networks (NNs) and K-means clustering applications for the purpose of classifying and grouping plant leaf diseases. The major goal of the suggested technique is to identify the illness. The major mango leaf diseases such as Anthracnose, Bacterial black spot, Twig blight, Gummosis, Bark splitting etc were detected using this model. The experimental findings show that the suggested method is a useful one that may greatly aid in the precise and low-computational-effort identification of leaf diseases. Future work will concentrate on automatically estimating the severity of the detected disease. An extension of this work will focus on developing hybrid algorithms, such as genetic algorithms and NNs, to increase the recognition rate of the final classification process underscoring the advantages of hybrid algorithms.

Compliance with Ethical Standards:

Conflict of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding Information: There is no funding for this paper

Acknowledgements: The authors would like to thank all who are directly or indirectly involved, advised and coordinated in collecting data and software information for the support of this paper.

References

[1] Praveen Kumar Ramtekkar, Anjana Pandey & Mahesh Kumar Pawar, “Accurate detection of lung cancer using optimized feature selection based on deep learning techniques” - 2023 Springer

[2] Shubhangi Solanki, Uday Pratap Singh, Siddharth Singh Chouhan, and Sanjeev Jain, “Lung Cancer Detection and Classification Using Intelligence Techniques: An Overview” - 2023 IEEE

[3] Md Ishtyaq Mahmud,Muntasir Mamun, and Ahmed Abdelgawad, “A Deep Analysis of Lung Cancer Detection from MR Images Using Deep Learning Networks” - 2023 MDPI

[4] Aqsa Tahir, Muhammad Asif, Maaz Bin Ahmad, Toqeer Mahmood, Muhammad Adnan Khan, and Mushtaq Ali6, "Lung Cancer Detection using Decision-Based Fusion Empowered with Fuzzy Logic" – 2022 Hindawi

[5] Anjanayya S, V. M. Gayathri, and R. Pitchai,"Lung Cancer Segmentation and Survival Prediction using Multimodal MRI Scans with Deep learning Algorithms" – 2022 IEEE

[6] Yue Zhang, Benxiang Jiang, Jiong Wu, Dongcen Ji, Yilong Liu, Yifan Chen, Ed X. Wu, and Xiaoying Tang, “Deep Learning Initialized and Gradient Enhanced Level-Set Based Segmentation for Lung Tumor From CT Images” – 2020 IEEE

[7] Tao Lei, Risheng Wang, Yuxiao Zhang, Yong Wan, Chang Liu, and Asoke K. Nandi, “DefED-Net: Deformable Encoder-Decoder Network for Lung and Lung Tumor Segmentation”. – 2022 IEEE

[8] Xin Dong, Yizhao Zhou, Lantian Wang, Jingfeng Peng, Yanbo Lou, and Yiqun Fan, “Lung Cancer Detection Using Hybridized Fully Convolutional Neural Network Based on Deep Learning Framework”. – 2020 IEEE

[9] Simona Turco, Thodsawit Tiyarattanachai, Kambez Ebrahimkheil, John Eisenbrey, Aya Kamaya, Massimo Mischi, Senior Member, Andrej Lyshchik, and Ahmed El Kaffas, “Interpretable Machine Learning for Characterization of Focal Lung Lesions by Contrast-Enhanced Ultrasound”. – 2022 IEEE

[10] Sowjanya M. N,Stafford Michahia, Mahanthesha U, and Nagalakshmi T. S,"Lung Cancer Detection in Chest X - Ray Image". – 20 IEEE

[11] Dr. H S Mohan and Mahanthesha U, “Human action Recognition using STIP Techniques”, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-7, May 2020

[12] J. F. Allen, “Maintaining knowledge about temporal intervals,” Commun. ACM, vol. 26, no. 11, pp. 832–843, Nov. 1983.

[13] C. Fernandez, P. Baiget, X. Roca, and J. Gonzalez, “Interpretation of complex situations in a semantic-based surveillance framework,” Image Commun., vol. 23, no. 7, pp. 554–569, Aug. 2008.

[14] J. Candamo, M. Shreve, D. B. Goldgof, D. B. Sapper, and R. Kasturi, “Understanding transit scenes: A survey on human behavior-recognition algorithms,” IEEE Trans. Intell. Transp. Syst., vol. 11, no. 1, pp. 206–224, Mar. 2010.

[15] Sardogan, Melike, Adem Tuncer, and Yunus Ozen. "Plant leaf disease detection and classification based on CNN with LVQ algorithm." In 2018 3rd international conference on computer science and engineering (UBMK), pp. 382-385. IEEE, 2018.

[16] Falaschetti, Laura, Lorenzo Manoni, Denis Di Leo, Danilo Pau, Valeria Tomaselli, and Claudio Turchetti. "A CNN-based image detector for plant leaf diseases classification." HardwareX 12 (2022): e00363.

[17] H. Ajra, M. K. Nahar, L. Sarkar and M. S. Islam, "Disease Detection of Plant Leaf using Image Processing and CNN with Preventive Measures," 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE), Bangladesh, 2020, pp. 1-6, doi: 10.1109/ETCCE51779.2020.9350890.

[18] Hassan, Sk Mahmudul, Arnab Kumar Maji, Michał Jasiński, Zbigniew Leonowicz, and Elżbieta Jasińska. "Identification of plant-leaf diseases using CNN and transfer-learning approach." Electronics 10, no. 12 (2021): 1388.

[19] Khattab, A., Habib, S. E., Ismail, H., Zayan, S., Fahmy, Y., & Khairy, M. M. (2019). An IoT-based cognitive monitoring system for early plant disease forecast. Computers and Electronics in Agriculture, 166, 105028.

[20] Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. Computational intelligence and neuroscience.