Grenze International Journal of Engineering and Technology, June Issue

Fully Connected Deep Learning for Tomato Disease Detection and Classification using Grey Co-Occurrence Matrix and Haar

Vidya Vasant Waykule[1] and Dattatraya Shankarrao Bormane[2]
[1,2]Aissms College of Engineering, Savitribai Phule University, Pune, Maharashtra, India.
Email: {vvwaykule, principal} @aissmscoe.com

***Abstract*—Tomato crops are vulnerable to a variety of diseases, which can result in severe financial losses for producers. Traditional methodologies for disease detection and classification are labor-intensive and necessitate specialized knowledge. In the recent years, advancements in deep learning models have shown favorable results in automating this process. In this study, we introduce a Fully Connected Neural Network deep learning architecture aimed at the identification and categorization of tomato diseases. The implementation of image segmentation via K-means clustering utilizing HSV color space is employed to enhance the quality of the outcomes. The suggested methodology employs the Grey Level Co-occurrence Matrix (GLCM) and Haar to facilitate the extraction of significant image features. It also compares the result of both the feature extraction techniques using Decision tree classifier, Random Forest, Principal Component Analysis and KNN. We utilize the tomato-leaf-multiple-sources dataset from Kaggle along with Real Image dataset to systematically train and assess the efficacy of our model. Our suggested model exhibits superior performance compared to existing state-of-the-art techniques in the diagnosis and categorization of tomato diseases using Fully Connected Neural Network.**

***Index Terms*— Tomato diseases, deep learning, GLCM, Haar, disease classification, hyper parameter, crop yield.**

I. INTRODUCTION

The tomato is an important crop globally, providing farmers with both sustenance and revenue. But tomato plants are subject to a range of illnesses, which can result in considerable output losses and economic harm. The ability to identify and classify these illnesses is critical for optimal management and control. There has been an increase in interest in employing molecular approaches to diagnose and categorise plant diseases, notably those harming tomato harvests, in recent years. These approaches can give prompt and precise diagnosis, allowing farmers to take necessary actions to limit disease spread and crop losses.

.In this study, we proposed a deep learning model for the detection and classification of tomato diseases that combines a fully connected neural network for illness classification with a GLCM and Haar for image feature extraction. By assisting farmers in identifying and treating tomato diseases sooner, our proposed methodology may boost crop productivity and reduce financial losses.

Grenze ID: 01.GIJET.11.2.203_20

II. RELATED WORK

In this study, we observed that deep learning models have done remarkably well in the automation process of identification and classification of various plant diseases. Machine learning and deep learning field have shown great potential in increasing agricultural productivity. Various observations have been taken in the realm of image segmentation utilizing clustering methodologies. The most widely adopted technique is the K-means clustering algorithm, which functions as an unsupervised algorithm utilized to delineate the area of interest from the background in expanded images. Subtractive clustering is utilized to determine the initial centroid [1]. In order to make the categorization better of tumour images in medical image analysis, Sekar, R., Harikrishnan, P. R., & Acharjya, K. [2] talk about the application of image segmentation techniques, especially boundary-based segmentation. It emphasizes how GLCM characteristics can be used for efficient classification. Water microbiological picture segmentation is the main emphasis of Kristanto, S. P., Hakim, L. N., & Yusuf, D. [3], who employ K-means clustering in conjunction with colour and texture feature extraction. By combining colour data with texture analysis using a Gabor filter, it highlights difficulties of segmenting tiny and challenging-to-identify microbiological organisms. The proposed image segmentation algorithm by B. He, Qiao, F., and Weijun, C. [4] successfully addresses the problem of edge information loss that arises during the convolutional process.

Manju C. C., & Jose, M. V. [5] presents a method for effective classification of Human Epithelial-2 (HEp-2) cell images, which is crucial for diagnosing various autoimmune diseases. Das, M., Gifford, H. C., and Andrade D. [6] explains benefits of applying multi-modality features more especially, grey level co-occurrence matrix texture features to improve image classification and cancer diagnosis in MRI and X-ray tomography images. Hasan, N., Ahmed, M. F., Nasif, M. A., et al. [7] use a combination of various feature extraction technique to accurately categorize brain cancers. The proposed approach is more effective, with a 98% accuracy rate. To identify early plant disease, a new technique which includes colour analysis and K-Means clustering is presented by Sharma, K., Jayanthi, J., Kumar, A., Jain, S., Jain, G., & Vishwakarma [8]. The technique used in this paper has demonstrated the ability to enhance plant health monitoring in contemporary agriculture by reaching a high accuracy rate of 97.2% in recognizing and identifying diseased spots. Through the different features extracted as well as shape recognition metrics and eccentricity values, Anggraini, R. A., Wati, F. F., Shidiq, M. J., Suryadi, A., Fatah, H., & Kholifah, D. N [9] focuses on the identification of herbal plants using image analysis techniques.

Prasad S. J., Mounika, G., and Kumari C. U. [10] By employing image processing techniques to identify leaf diseases, particularly leaf spot, early on, the project aims to increase agricultural productivity in India. Dzikri, M. A., Rijati, N., Septianingrum, S. A., & [Other Authors] [11] employs a dataset of 275 distinct leaf kinds and a 5-step iteration optimization method to improve classification accuracy in order to categorize different leaf types using a Neural Network and a momentum methodology. With certain settings such as 100 neural training cycles, a learning rate of 0.5, of 0.6, the model reached an accuracy of 92.8% after 4751 iterations. The K-Means algorithm is used as a pre-processing step in Fotiadis G. [12]'s research on plant leaf segmentation.477 plant leaf samples make up the dataset, which divides dataset into training and testing to highlight how crucial precise segmentation before classification. Gia L. etc used a machine learning model to detect bacterial wilt [13] in tomato plants. The model was trained using photos of bacterial wilted tomato plants as well as healthy tomato plants. The model identified bacterial wilt in tomato plants and has given success rate of 92.3%. Ghosal, S., & Sarkar [14] used in one study to identify tomato illnesses and healthy plants. The CNN was trained on a huge dataset of tomato photos and obtained a 99.14% accuracy rate in recognizing the various tomato illnesses. Li W. and Tang, W [15] offer an innovative technique that improves the segmentation of Regions of Interest (ROI) in a variety of medical pictures by combining Neural Networks (NN) with Grey Level Co-Occurrence Matrix (GLCM) features. This method successfully delineates the ROI borders by extracting GLCM properties such energy, homogeneity, contrast, and correlation using texture analysis. Ferentinos K.P., Sharma, R., & Mishra A. [16]- [17] compared deep learning techniques like CNN, RNN and transfer learning for their effectiveness in classifying diseases in plants based on the symptoms of the diseases.

III. PROPOSED METHODOLOGY

This text outlines a comprehensive workflow for automated image classification. It incorporates image pre-processing, segmentation using K-means clustering based on Hue values.

As depicted in Fig. 1, classification is performed using Fully Connected Neural Network (FCNN) after extracting Grey-Level Co-occurrence Matrix (GLCM) and Haar features.

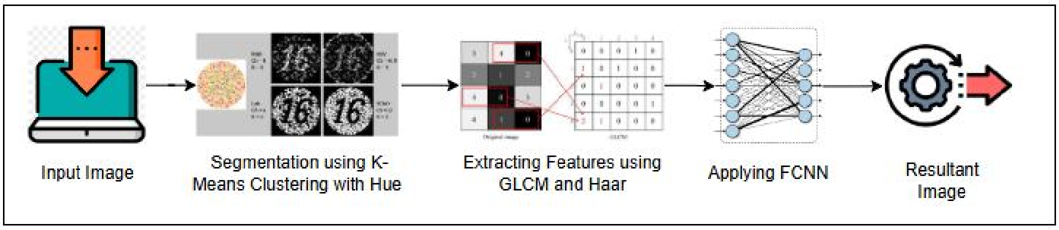

Figure 1. Classification Workflow

The process begins by acquiring an input image, which serves as the raw data for analysis. Ensuring the image is of high quality is essential for accurate analysis. As Pre-processing enhances the image by mitigating noise and standardizing input dimensions we applied image resizing to a dimension of [227x227 (rows and columns)]. A new mixed segmentation method is applied which combines the improvement of K-means clustering with HSV colour transformation. This method is particularly beneficial in cases where colour differentiation is very important.

After segmentation of the image, feature extraction is executed to extract numerical representations or patterns that encapsulate significant features of the image like shape descriptive, colour histograms, texture patterns etc. During the characteristic extraction phase, we implement GLCM and Haar feature techniques to capture valuable attributes from segmented images.

This work uses a fully connected neural network (FCNN) classification to find out the effectiveness of properties obtained from various extraction methods. The FCNN architecture consists of an i/p layer, two hidden layers containing 256 and 128 neurons, each with the ReLU activation function, and an output layer with a softmax activation function to facilitate the classification of multi-class labels. The model was carefully compiled using the Adam optimizer in combination with the loss of categorical cross-entropy, and it was run for 500 periods with a batch size fixed at 32. The data set was systematically divided into training and testing subsets, granting 80% of the data for training purposes and reserving 20% for testing purpose to ensure that the model is trained in a significant part of available data and maintaining a separate data set for impartial evaluation. This FCNN classifier is used not only for the GLCM features, but also for Haar feature extraction techniques as shown in the Fig.2. For each method of extracting features, we conducted performance assessments that included accuracy, precision, recall, F1 score, specificity and sensitivity.

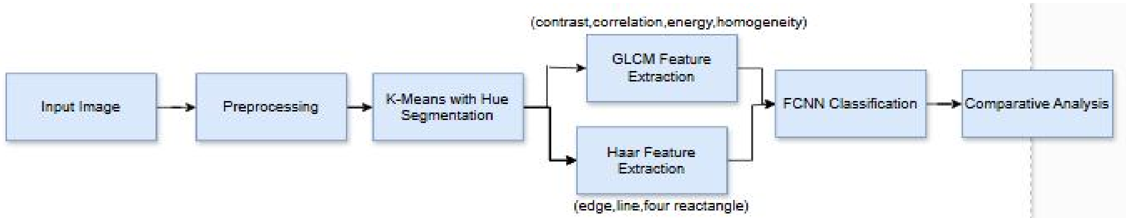

Figure 2. Working Model for Comparative Analysis

IV. Experimental Verification and Result Analysis

This text outlines a comprehensive workflow for automated image classification. It incorporates image pre-processing, segmentation using K-means clustering based on Hue values. As depicted in Fig. 1, classification is performed using Fully Connected Neural Network (FCNN) after extracting Grey-Level Co-occurrence Matrix (GLCM) and Haar features.

A. Dataset and Experimental Settings

The data collected during the research were collected through the use of a Samsung Galaxy A50 phone equipped with an Android version 11 camera, generating 412 photos. Within this dataset, 4382 images were obtained from the tomato-leaf-multiple-sources dataset available on Kaggle. In order to deal with the challenges posed by limited training samples and the sensitivity to over-installation resulting from various modifications applied to the resultant image, various techniques are systematically used. These methodologies included mirror and left-right rotation, the use of CLAHE (Contrast Limited Adaptive Histogram Equalization) to enhance, sharpen, and use media filters to reduce noise, refinements, image scaling, rotation, cutting, and image completion. Dataset contains each image with a dimension of 227 × 227.

B. Result and Analysis

After getting relevant information of the images, we conduct through study to measure the performance of each model. After extracting features from the images, we perform a comprehensive analysis to evaluate the performance of each method. The evaluation is based on several parameters, including PCA variance (to assess dimensionality reduction effectiveness), runtime (to measure computational efficiency) and the classification accuracy of Random Forest, K-Nearest Neighbors (KNN), and Decision Trees (DT) as depicted in Fig. 3.

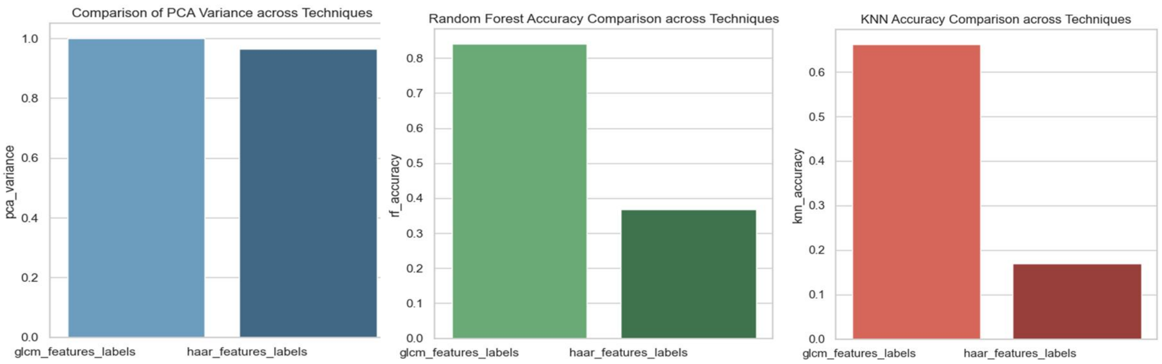

Figure 3. Accuracy of Feature Extraction Techniques using PCA, RF, KNN

From the above analysis, it has been demonstrated that the Grey-Level Co-occurrence Matrix (GLCM) provides superior results in terms feature discrimination compared to other method. Additionally, the GLCM's ability to capture spatial relationships and texture information enables for more precise classification in various imaging applications which makes it a preferred choice among researchers and practitioners. Fig. 4 compares the classification using GLCM and Haar using Accuracy, Precision, Recall, F1 Score, Sensitivity and Specificity. The results indicate that GLCM consistently outperforms Haar features across all metrics, underscoring its effectiveness in extracting meaningful patterns from images. This superiority is particularly evident in complex datasets where texture plays a crucial role, demonstrating the GLCM's robustness in handling variations that Haar features may overlook.

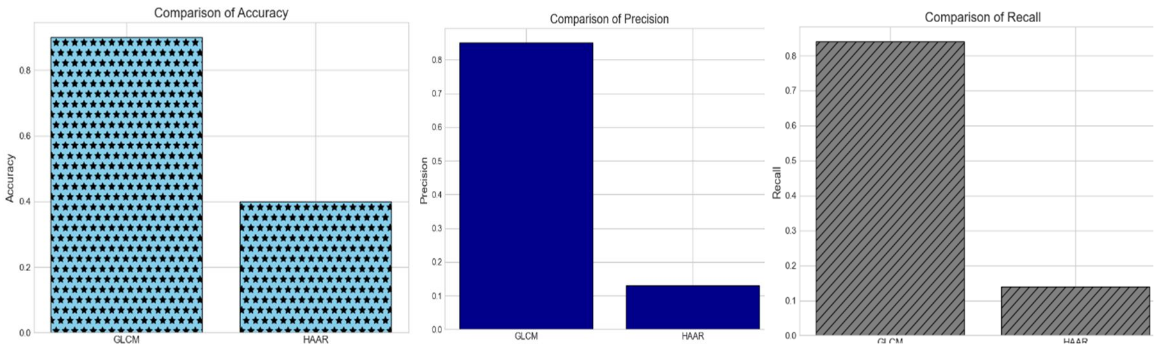

Figure 4 Accuracy, Precision and Recall of classification

V. CONCLUSION

We are able to classify the feature extraction techniques in accordance with their performance metrics using the received results. Based on it, we finally concluded that GLCM features exceeded the performance of alternatives. This finding underscores the importance of selecting appropriate feature extraction methods in image processing tasks, as GLCM features provide superior texture analysis capabilities that can enhance classification accuracy. Results obtained from this study will open the door for additional research into optimizing these hyper parameters which ultimately leads to more resilient models that can adapt to various agriculture challenges Continuing investigation in this field, it will not only improve the effectiveness of disease management strategies but also contribute to sustainable farming practices that guarantees food security in an ever-changing climate. We aim to create a comprehensive system that empowers farmers with timely information and actionable insights by combing real time data collection and advanced machine learning techniques. This system will facilitate proactive decision-making, enabling farmers to implement targeted interventions that minimize the impact of diseases on their crops while optimizing resource use and promoting environmental sustainability. The anticipated outcomes of this research include increased crop yields, reduced chemical inputs, and improved resilience against emerging agricultural threats, ultimately fostering a more sustainable food production system.

REFERENCES

[1] Nameirakpam, D., Khumanthem, M., & Yambem, J. C., "Image Segmentation using K- means Clustering Algorithm and Subtractive Clustering Algorithm", Procedia Computer Science, 54, 764–771, Dec 2015.

[2] Sekar, R., Harikrishnan, P. R., & Acharjya, K., "Brain Image Segmentation via GLCM Features and CNN Classification for Improved Image Retrieval Using Machine Learning", IEEE International conference on ICT in Business Industry and Government, Dec 2023

[3] Kristanto, S. P., Hakim, L. N., & Yusuf, D., "K-means Clustering Segmentation on Water Microbial Image with Color and Texture Feature Extraction. Building of Informatics", Technology and Science (BITS), 2022.

[4] B He, Qiao, F., Weijun, C., & others. "Fully convolution neural network combined with K-means clustering algorithm for image segmentation", Tenth International Conference on Digital Image Processing, Proceeding Volume 10806, May 2018.

[5] Manju, C. C., & Jose, M. V., "HEP-2 specimen image segmentation and classification using GLCM and DCT-based feature extraction with CNN classifier", Advances in Communication Systems and Networks Book, 2020

[6] Andrade D., Gifford, H. C., & Das, M., "Multi-modality GLCM image texture feature for segmentation and tissue classification", Proceedings of SPIE Medical Imaging, 2023.

[7] Hasan, N., Ahmed, M. F., Nasif, M. A., & others. "Hybrid Feature Extraction Approach for Robust Brain Tumor Classification: HOG, GLCM, and Artificial Neural Network", Proceedings of 6th International Conference of Electrical Engineering and Information & Communication Technology, May 2024.

[8] K. Sharma, J. J, A. Kumar, S. Jain, G. Jain and H. Vishwakarma, "Computer Vision Based Smart Leaf Disease Detection System," 2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, 2024, pp. 1-9.

[9] Anggraini, R. A., Wati, F. F., Shidiq, M. J., Suryadi, A., Fatah, H., & Kholifah, D. N., "Identification of herbal plant based on leaf image using GLCM feature and K-means", Journal ofComputing and Information Technology, Mar 2020, Vol. 17 No 1.

[10] Kumari, C. U., Prasad, S. J., & Mounika, G., "Leaf disease detection: Feature extraction with K-means clustering and classification with ANN", 3rd International Conference on Computing Methodologies and Communication, Mar 2019.

[11] M. A. Dzikri, S. A. Septianingrum, N. Rijati, P. Pujiono and M. A. Soeleman, "Artificial Neural Network for image based Leaf Type Classification with Feature Extraction GLCM Method," 2023 International Seminar on Application for Technology of Information and Communication, Semarang, Indonesia, 2023., pp. 35-39.

[12] Fotiadis, G.,"Color space based methodology for leaf segmentation", https://doi.org/10.31219/osf.io/m3aj6, Jun 2023.

[13] Gao, L., Lin, S., & Gao, H., "A novel deep learning framework for early detection and diagnosis of tomato bacterial wilt", Computers and Electronics in Agricultur73, 105384,2020.

[14] Ghosal, S., & Sarkar, K., "Rice Leaf Diseases Classification Using CNN With Transfer Learning.", IEEE Calcutta Conference (CALCON), https://doi.org/10.1109/CALCON49167.2020.9106423,2020.

[15] Li W., Li Q., Tang W., & Li, X., "Tomato Leaf Disease Recognition Based on Convolutional Neural Network", 8th International Conference on Agro- Geoinformatics (Agro-Geoinformatics) (pp. 1-5) IEEE, 2019.

[16] Ferentinos K. P. (2021). Deep learning models for plant disease detection and diagnosis: a comprehensive review. Neural Computing and Applications, 33(6), 2251-2285.

[17] Sharma, R., & Mishra, A, "Deep learning-based disease detection in tomato plants: A comprehensive review", Bio systems Engineering, 212, 92-115, 2021.