IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

DISEASE DETECTION ON MANGO LEAVES USING MACHINE LEARNING ALGORITHMS

Mr.CHEEKATI SUDHEER[1], Mr.M.RAMJEE[2]

1. DEPARTMENT OF COMPUTER SCIENCE AND SYSTEMS ENGINEERING (A), ANDHRA UNIVERSITY, COLLEGE OF ENGINEERING VISAKHAPATNAM-53 003.

2. PROFESSOR IN DEPARTMENT OF COMPUTER SCIENCE AND SYSTEMS ENGINEERING (A), ANDHRA UNIVERSITY, COLLEGE OF ENGINEERING VISAKHAPATNAM-530 003.

ABSTRACT

Agriculture has been the foundation of our nation for many generations. However, we're increasingly worried about a troubling drop in crop production, which is a concern for the future. Among the many reasons for this decline, diseases are a major culprit. This study focuses on two specific plant diseases, Anthracnose and Black Sooty Mold, both of which harm mango leaves and reduce overall mango production. Our main goal in this research is to introduce a computer-based approach, using machine learning, to detect these diseases in mango leaves. First, we use a technique called "k-means clustering" to analyze pictures of mango leaves. Then, we extract various features like patterns, shapes, and colors from these images, including GLCM, HOG, SIFT, and the Statistical Feature Matrix. We then use different computer programs, such as Support Vector Machine, K Nearest Neighbors, Decision Tree, Naive Bayes, and Random Forest, to help us find these diseases.Our experiments showed that the Random Forest computer model worked better than the others, especially when we used image analysis and the Statistical Feature Matrix. This means our method is good at finding diseases in mango leaves early, which is important for farmers.

In summary, our research could help farmers manage diseases in their mango trees better, leading to increased mango production. This advancement has the potential to improve farming strategies in the agriculture sector.

Keywords: Anthracnose Disease, Black Sooty Mold, Image Segmentation, Feature Extraction, Machine Learning Techniques.

INTRODUCTION

Mango, a fruit crop that is widely cultivated, possesses significant value due to its delectable taste, nutritional richness, and substantial economic contribution. However, the production of mangoes is frequently threatened by various diseases that result in significant yield losses and a deterioration in fruit quality. The emergence of these diseases can be attributed to multiple factors, including fungi, bacteria, viruses, and environmental stresses. According to a recent survey conducted by the Food and Agriculture Organization (FAO), these plant diseases are responsible for an estimated 10-16% reduction in global crop yields [18]. The timely identification and precise diagnosis of these ailments are crucial prerequisites for effective management and control strategies. Traditional methods of detecting diseases in mango trees involve manual examination by experts, a process that There are various types of diseases that impact mango trees, including anthracnose (shown in figure 1), black sooty mold (shown in figure 3), gall midge (shown in figure 4), bacterial cancer (shown in figure 2), and others [19, 20]. These diseases can lead to a range of symptoms, such as leaf spots, stem cankers, fruit rot, and deformities in the fruit.

Fig 1: Anthracnose

Fig 2: Bacterial Cancer

Fig 3: Black sooty mold

Fig 3: Gall Midge

Among the diseases affecting mango trees, anthracnose and black sooty mold are the most prevalent. Anthracnose is caused by the fungus Colletotrichum gloeosporioides, which triggers leaf spots, stem cankers, and fruit rot. These illnesses can significantly impact the production and quality of mango fruit. For instance, a study conducted by Singh et al. (2017) discovered that anthracnose can result in a fruit yield reduction of up to 50% . Similarly, black sooty mold can lower fruit quality, rendering it unsuitable for sale in the market.

In contemporary times, the advent of machine learning methodologies has presented a promising avenue for the detection of diseases in plants, including mango trees. These methodologies have the capacity to discern intricate patterns and scrutinize vast datasets that often pose challenges for human comprehension. By virtue of their capabilities, they facilitate prompt and precise diagnostic outcomes for plant diseases. Machine learning techniques facilitate the training of expansive image datasets that encompass both healthy and afflicted mango leaves. Through exposure to such datasets, these techniques attain proficiency in discriminating between various disease types, leveraging distinctive attributes such as color, texture, shape, and size. Following the training phase, these algorithms are primed to classify novel images of mango leaves, affording categorization into healthy or diseased states. Furthermore, these algorithms furnish a quantifiable gauge of disease severity, thereby enhancing the analytical depth of the diagnosis process.

Numerous investigations have substantiated the efficacy of machine learning implementations in the context of disease identification within the realm of plants, encompassing mango trees as well. To exemplify, Kumar et al.employed a fusion of machine learning methodologies and image processing techniques to discern mango leaf diseases, relying upon salient attributes pertaining to leaf color and texture. In a parallel vein, Singh et al. harnessed machine learning algorithms to classify mango leaves, stratifying them into categories denoting vitality and affliction, utilizing texture attributes as discriminative markers.

The scholarly landscape also reveals explorations into the deployment of diverse machine learning modalities, such as support vector machines (SVMs), decision trees (DTs), and k-nearest neighbors (KNNs), as vehicles to facilitate disease detection within the domain of plants.

This paper outlines an investigation into the realm of disease detection within mango leaves, employing machine learning methodologies, while eschewing the utilization of neural networks or deep learning techniques. The study encompasses an assessment of the performance exhibited by several widely recognized machine learning algorithms across two distinct datasets, both in the presence and absence of segmentation. Our paper not only contributes empirical insights to the field but also offers valuable implications for forthcoming research endeavors in this domain, owing to our discerning observations concerning the effectiveness of diverse machine learning applications in the context of disease detection within mango leaves. The employment of machine learning techniques in the identification of diseases within mango trees holds the potential to enhance disease detection capabilities and diagnostic efficiency, consequently fostering advancements in mango production and mitigating losses attributed to disease-related issues. The paper's structure unfolds as follows: Section 2 delves into a comprehensive review of pertinent literature, followed by Section 3, which expounds upon the methodology encompassing segmentation, feature extraction, and classification. Subsequently, Section 4 is dedicated to presenting the conducted experiments and an in-depth analysis of the obtained outcomes. Finally, the concluding remarks are encapsulated in Section 5.

2. LITERATURE REVIEW

Researchers have worked on how to accurately identify the disease in an early stage using various computer vision and machine learning approaches. We have surveyed the papers related to our work here.

Tajul Rosli Razak et al., worked on input images by applying various image processing techniques such as noise removal, contrast and color enhancement, and segmentation. In the second stage, the authors extract features from the preprocessed images using texture analysis techniques as gray-level co-occurrence matrix (GLCM) and local binary pattern (LBP) and used a support vector machine (SVM) classifier to classify the extracted features into the three disease classes. The SVM classifier is trained on the extracted features and tested on a separate set of images to evaluate its performance . Shripad S.Veling et al., developed a method where contrast is enhanced. Fuzzy C means is used for image segmentation and the features extracted are Contrast, Homogeneity, Cluster prominence, shade, Correlation, Energy, Entropy, Variance and Dissimilarity. SVM is used to classify the data and it is tested to evaluate the performance of the model generated .

Rabia Saleem et al., proposed a segmentation approach by considering the leaf's vein pattern. Texture features are analyzed and color space models are also used to compute the color features and the statistical metrics to form a feature vector. They are fused with canonical correlation analysis fusion and SVM is used for identification of the disease leaves .

Ooi Wei Herng et al., focused on image segmentation on mango leaves using fast k-means clustering after the color component extraction using RGB and HSV color spaces of the image. quantitative analysis is done to evaluate the segmentation based on accuracy, specificity and sensitivity [8].

Eftekhar Hossain et al., worked on identifying anthracnose, canker, leaf spot, and blight of various species in plants. The texture features, like coarseness, granularity, roughness, are used and KNN is used to classify the feature vectors .

M.P.Vaishnnave et al., worked on disease detection in groundnut leaves where binary masking of the image is done and it is converted to HSV color space. GLCM is computed for each of the HSV spaces to extract the features. SVM and KNN are used for classification .

Qing Gu et al., used hyper spectral imaging by using a genetic algorithm, successive projections algorithm, boosted regression tree, and three wavelength methods. The classification techniques used are support vector machine (SVM), boosted regression tree, random forest (RF) and classification and regression trees (CART). Wilt Virus, at an early stage, is detected using this approach .

Kshyanaprava Panda Panigrahi et al., tried to detect disease in Maize leaf. Cercospora leaf spot, leaf blight and common rust diseases are focused in the work. Naive Bayes (NB), K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM) are used for classification .

Sa'ed Abed et al., proposed a flow using Support Vector Machine (SVM) as the classification technique to predict recognized disturbances. The testing and training images for the proposed flow were collected from a public database and were initially designed for the detection of two different forms of disruptions in beans. These photos are then enhanced and transformed from the RGB color model to the HIS using linear contrast stretch. By combining Euclidean distance and K-means clustering, the H element will be partitioned. The segmented image is transformed to a gray-scale image using the color co-occurrence matrix in order to extract features. Features of test photos are provided to the trained SVM model in order to categorize diagnosed diseases.

Sharath D M et al., proposed a word focusing at establishment of a system for identification of diseases. The proposal is divided into two phases, first is Training phase where acquisition of images, pre-processing, feature extraction from images post-processed, image segmentation, classification and calculation of diseased area is done. Grab cut segmentation (GCS) is utilized to segment the image, and canny edge detection is employed to identify the diseased part of the fruit. The shape of the region containing the stiff fruit is then examined in terms of pixels. The percentage of fruit contamination is determined by the number of pixels counted, and a precautionary estimate is given based on the recognized disease.

3.3.1Flowchart

Figure 11 represents the step-by-step procedure of the proposed system. It clearly explains us our complete view of the proposed system. That is where we have taken the data and what methods and techniques are applied etc.In the below flowchart,can be explained by identifying the problems to selecting the dataset in the UCI repository to pre-processing of the data can be cleaned and transformations to classifying the different machine learning algorithms (naïve Bayes, k-nearest neighbor,Random forest, support vector machine, decision tree) to evaluate the results.

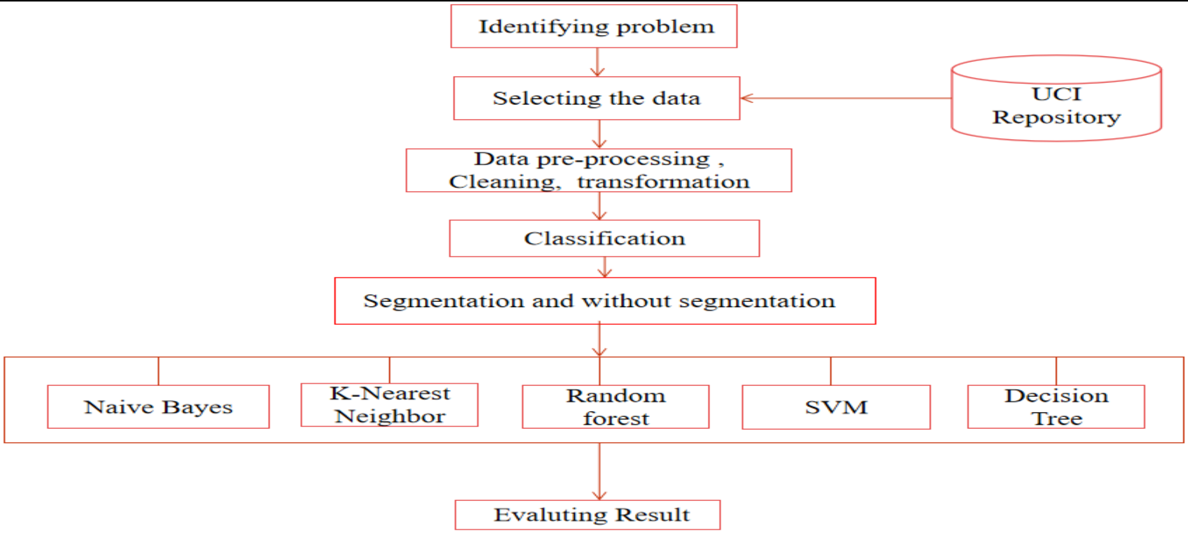

Figure 5:Architecture of the Proposed Model

This section describes the proposed model for the detection of diseased mango leaves. Steps involved in image classification include dataset acquisition, image pre-processing, segmentation, feature extraction and classification.

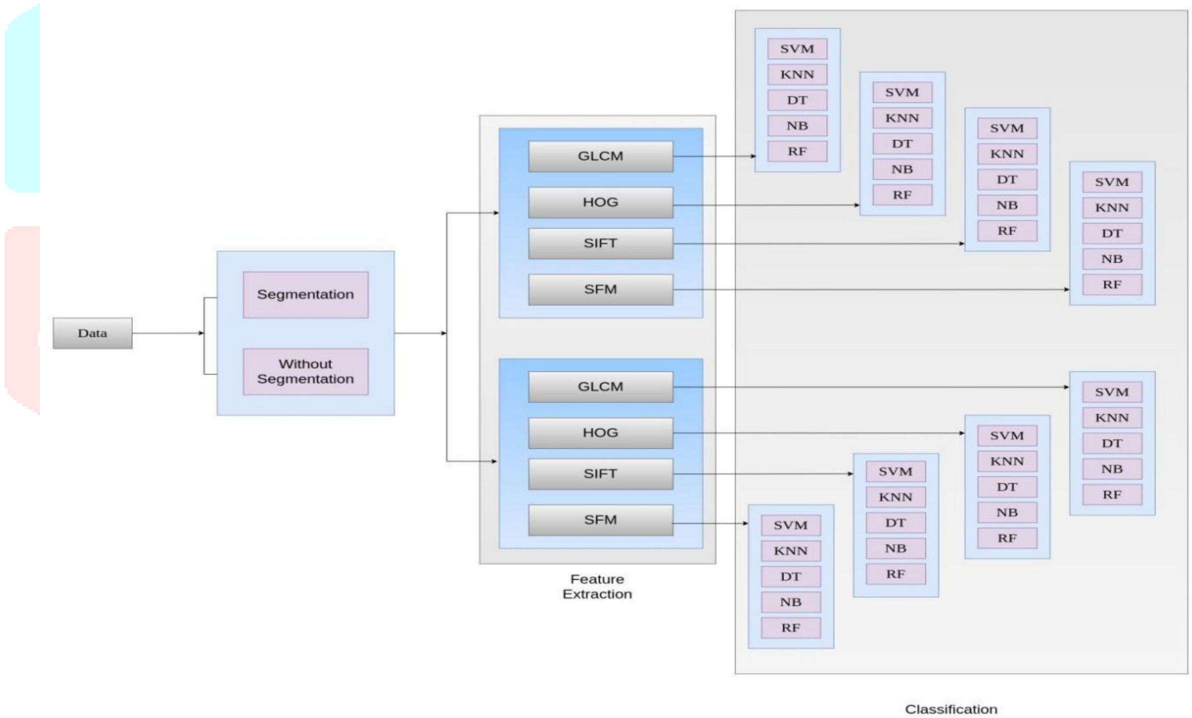

Figure 6: Architecture of the Proposed Model

Figure 12 shows the proposed model. The modules present in the model are explained in the below sub sections.

3.4 Dataset Description

The primary step in classifying the image data is the data collection. We have taken two datasets from Kaggle [4,5]. Harumani Dateset includes three categories i.e., healthy, anthracnose and black sooty mold. The total images in the dataset is 1405. The leaves' images with Anthracnose are 525, the ones with Black sooty mold are 656 and the healthy leaves are 224.

Mango LeafBD dataset contains 4000 images which is post augmented set out of which 1800 are distinct image leaves and the rest are generated using augmentation techniques like rotations, flps and zooming in. The images are captured using a mobile camera. The resolution of the images is 240x320. The categories present in the dateset are Anthracnose, Black sooty mold, Bacterial Canker, Cutting Weevil, Die Back, Powdery Mildew, and Gall Midge.

3.5 Image Preprocessing

Image pre-processing in machine learning involves a series of essential steps undertaken to enhance the quality and utility of raw images before they are utilized for analysis or classification using machine learning algorithms. The primary objective of image pre-processing is to refine images, making them more conducive to subsequent analysis and improving the efficiency and accuracy of machine learning models.

During image pre-processing, various tasks are executed to optimize images for further processing. These tasks include resizing and scaling images to a standardized format, which aids in streamlining subsequent operations. Noise reduction techniques are employed to mitigate unwanted artifacts caused by factors like lighting conditions or sensor limitations. Additionally, contrast enhancement is applied to improve visibility by adjusting image contrast, while color correction standardizes color representations across images to rectify variations in lighting or color balance.

Furthermore, image registration techniques ensure proper alignment of multiple images when comparison or alignment is required. For meaningful analysis, segmentation is employed to segment images into distinct regions, isolating objects of interest. Extracting relevant features, such as edges, corners, or texture patterns, from images is also part of pre-processing, providing valuable input for machine learning algorithms.

Normalization of pixel values to a consistent range minimizes the influence of diverse lighting conditions and aids in the convergence of machine learning algorithms. To bolster the diversity of training data, data augmentation techniques like rotation, flipping, and cropping are utilized to generate variations of original images. Additionally, pre-processing involves addressing undesirable artifacts like scratches, dust, or imperfections to ensure cleaner and more relevant input data.

The selection of specific pre-processing techniques is contingent on factors such as the problem at hand, image characteristics, and the requirements of the chosen machine learning algorithm. Effective image pre-processing substantially elevates the performance of machine learning models by furnishing them with refined and pertinent input data, ultimately leading to more accurate and robust outcomes.

3.6. Image Segmentation

Image segmentation is a fundamental process within the realm of machine learning, involving the partitioning of an image into distinct and meaningful regions. This technique is utilized to identify and isolate specific objects or features of interest within an image. By segmenting an image, intricate structures and boundaries within the visual data can be delineated, facilitating subsequent analysis and interpretation. Image segmentation plays a pivotal role in various applications, such as object recognition, medical image analysis, autonomous driving, and more. Through the identification of regions of interest, machine learning models can then focus on analyzing the segmented portions with greater precision and accuracy, enabling tasks like object detection, classification, or measurement. Techniques for image segmentation encompass both traditional methods, like thresholding and region-based approaches, as well as advanced methodologies such as convolutional neural networks (CNNs) designed for semantic segmentation. Successful image segmentation is a critical step towards unlocking deeper insights from visual data and enabling more refined and targeted machine learning analyses.

3.7 Feature Extraction

Feature extraction in machine learning is the process of transforming raw data into a more concise and informative representation, known as features. Features are specific attributes or characteristics extracted from the original data that capture relevant patterns and information. This transformation is essential for improving the performance of machine learning algorithms, as it reduces the dimensionality of the data and focuses on the most important aspects.

In various data types, such as images, text, or sensor measurements, raw data can be complex and contain noise or redundant information. Feature extraction involves selecting, combining, or transforming these raw data points to create a set of features that retains the essential information while discarding less important details. These features serve as input for machine learning models, making them more capable of understanding the underlying relationships in the data and making accurate predictions or classifications.

in image processing, features might include edges, textures, or color histograms, while in text analysis, features could involve word frequencies or semantic embedding. The goal is to capture the distinctive characteristics of the data that are relevant to the task while removing noise and redundancy.

Effective feature extraction is crucial because it simplifies data representation, accelerates computation, and aids in generalization to new, unseen data. It also enables machine learning models to perform better on tasks like classification, regression, clustering, and more.

3.7.1 Gray-Level Co-occurrence Matrix

The Gray-Level Co-occurrence Matrix (GLCM) is a computational technique widely utilized in machine learning and image analysis to extract essential texture information from images. This method is particularly effective in capturing intricate spatial relationships between pixel values within an image. By quantifying the occurrences of pixel value pairs at varying distances and directions, GLCM constructs a matrix that encodes the distribution of these relationships. This matrix can then be analyzed to derive various statistical measures, such as contrast, energy, homogeneity, and entropy, which collectively characterize the texture properties of the image. GLCM finds applications in diverse fields, including medical imaging, satellite remote sensing, and industrial quality control. Its ability to unveil complex textures empowers machine learning algorithms to discern subtle patterns and structures within images, leading to enhanced performance in tasks like classification, segmentation, and anomaly detection

Texture feature	Equation
Contrast	$\sum_{i=1}^{N}\sum_{j=1}^{N}(i-j)^2P(i,j)$
Entropy	$-\sum_{i=1}^{N}\sum_{j=1}^{N}P(i,j)\,lgP(i,j)$
Correlation	$\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(i-\bar{x})(j-\bar{y})P(i,j)}{\sigma_x\sigma_y}$
Energy	$\sum_{i=1}^{N}\sum_{j=1}^{N}P(i,j)^2$

3.7.2 Histogram of Oriented Gradients

The Histogram of Oriented Gradients (HOG) is a prominent feature extraction technique used in machine learning, primarily in computer vision tasks such as object detection and image classification. HOG computes and analyzes the distribution of local gradient orientations within an image to capture the underlying shape and texture information. By dividing the image into small regions and calculating gradient magnitudes and orientations within each region, HOG constructs histograms that summarize the frequency of gradients' directions. These histograms effectively represent the distinctive features of an object's edges and texture patterns. HOG features are particularly robust against changes in lighting conditions and variations in object pose, making them suitable for object recognition in diverse scenarios. In machine learning applications, HOG features serve as informative inputs for classifiers like support vector machines (SVMs) or neural networks. This technique's ability to encode crucial visual cues in a manner that is resistant to various environmental factors underscores its significance in bolstering the accuracy and robustness of computer vision algorithms.

3.7.3 Scale-Invariant Feature Transform

Scale-Invariant Feature Transform (SIFT) is a pivotal algorithm in computer vision and machine learning that facilitates the extraction and matching of distinctive features from images. Developed by David Lowe, SIFT addresses the challenge of identifying features regardless of their scale, rotation, or affine transformations. SIFT detects key points in an image by analyzing the distribution of gradients and creating descriptors that encapsulate the local appearance and orientation of these key points. What sets SIFT apart is its ability to recognize features invariant to changes in size and orientation, rendering it exceptionally effective for tasks like object recognition, image stitching, and 3D reconstruction. In machine learning applications, SIFT features serve as robust inputs for algorithms such as support vector machines (SVMs) or neural networks. The versatility and resilience of SIFT features make them a cornerstone in the realm of computer vision, enabling machines to perceive and understand visual content with heightened accuracy and adaptability.

3.7.4 Statistical Feature Matrix

A statistical feature matrix in machine learning is like a table where each row represents a data point, and each column contains different statistical measures (like averages, totals, or other calculations) based on the data attributes. It's a way to summarize and transform raw data into a structured format that machine learning models can use for analysis and prediction.

3.8. Classifier Selection

This section details the different classifiers which are used in the work. This work compares the two different flows with and without image segmentation on two different datasets.

3.8.1 Support Vector Machine

Support Vector Machine (SVM) is a machine learning algorithm used for classification and regression tasks. In SVM, the algorithm finds a "hyperplane" (a line for 2D data or a plane for higher dimensions) that best separates different classes of data points. This hyperplane is positioned in a way that maximizes the margin, which is the distance between the hyperplane and the nearest data points from each class. SVM aims to find this optimal hyperplane to make accurate predictions. It's effective for tasks where you want to find a clear boundary between different groups of data points.

3.8.2 K Nearest Neighbor

K-Nearest Neighbors (K-NN) is a simple and intuitive machine learning algorithm used for classification and regression tasks. In K-NN, when you want to make a prediction for a new data point, it looks at the 'K' nearest data points in the training set and assigns the majority class (for classification) or the average value (for regression) among those 'K' neighbors as the prediction. It's like asking your nearest neighbors for advice, where 'K' represents how many neighbors you consult. K-NN is easy to understand but can be sensitive to the choice of 'K' and the distance metric used to measure closeness.
L-

3.8.3 Decision Tree

A Decision Tree is a machine learning algorithm used for both classification and regression tasks. It works by breaking down a dataset into smaller subsets based on the values of input features. At each step, it makes a decision based on a feature to split the data, creating a tree-like structure of decisions. These decisions lead to the prediction of a class label (for classification) or a numerical value (for regression) for a given input. Think of it as a series of if-else questions that guide the algorithm to make predictions based on the input features. Decision Trees are easy to understand and interpret, making them a popular choice in machine learning.

3.8.4 Naive Bayes

Naive Bayes is a simple and effective classification algorithm used in machine learning. It works by calculating the probability that a data point belongs to a particular category based on the features it has. It assumes that all the features are independent of each other, even though this assumption is often not true in real-world data. Despite this simplification, Naive Bayes is commonly used for tasks like spam email detection, sentiment analysis, and text classification due to its speed and decent performance.

3.8.5 Random Forest

Random Forest is a machine learning algorithm that combines the predictions of multiple decision trees to make more accurate and robust predictions. It works by creating many decision trees, each trained on a different subset of the data, and then averaging their predictions for classification or taking their average for regression. This ensemble approach improves accuracy and helps prevent over fitting, making it a popular choice for various machine learning tasks.

4. EXPERIMENT AND RESULT ANALYSIS

This section explains the experimental findings based on the proposed framework model. We used one segmentation technique, four feature extraction techniques and five classifiers.

For Segmentation, K-means clustering technique is used with the number of clusters to be formed in an image being k=9. The following are the results of the K-means clustering.

LeafBD Dataset						
Segmentation	**Feature Extraction Techniques**	**SVM**	**KNN**	**DT**	**NB**	**RF**
With Segmentation (K-Means Clustering)	**GLCM**	74.8	58	81.2	66	87.4
	HOG	85.6	77.2	67.6	70.2	83.2
	SIFT	81.88	78.16	74.41	64.06	82.69
	SFM	87.2	89.4	86.2	75.8	91.8
Without Segmentation	**GLCM**	71.6	65.4	88.6	67.4	89.2
	HOG	86.6	81.4	68.2	67.8	84.2
	SIFT	84.82	81.94	78.03	64.63	85.64
	SFM	90.2	93	92.8	77.6	95.4

Table 1Accuracies of LeafBD Set

Harumani Dataset						
Segmentation	**Feature Extraction Techniques**	**SVM**	**KNN**	**DT**	**NB**	**RF**
With Segmentation (K-Means Clustering)	**GLCM**	63.5	55.49	74.78	62.91	82.49
	HOG	87.83	90.8	74.18	71.81	85.76
	SIFT	76.06	74.53	69.35	54.79	79.98
	SFM	81.11	89.63	88.15	76.67	91.11
Without Segmentation	**GLCM**	67.06	64.99	86.65	69.73	92.88
	HOG	95.25	88.13	79.82	75.37	90.5
	SIFT	77.69	78.16	72.56	53.01	83.19
	SFM	77.41	89.26	90	80.74	95.19

Table 2 Accuracies of Harumani Dataset

The accuracy of the models are observed and tabulated in Tables 1 and 2. The results are analyzed below.

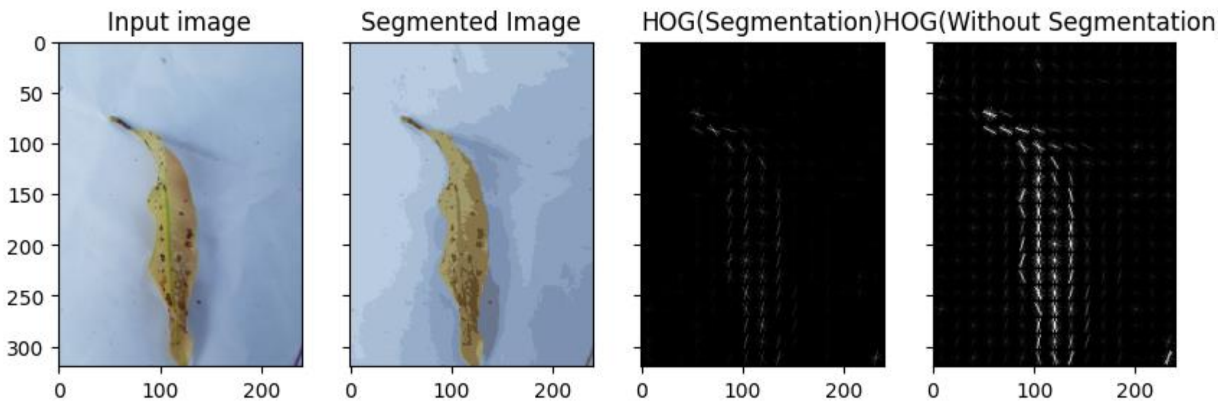

Figure 7: HOG with and without segmentation

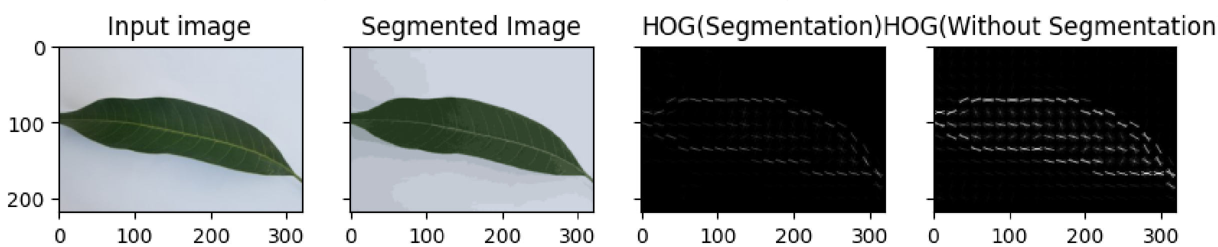

Figure 8: HOG with and without Segmentation for healthy leaves

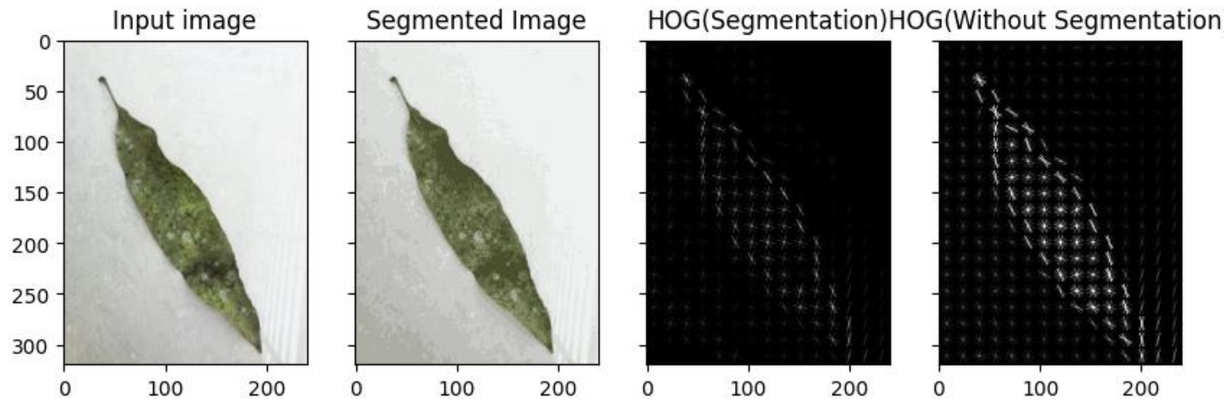

Figure 9: HOG With and without Segmentation for Black Sooty Mold

From Figures 7,8 and 9, it is evident that when HOG is used after segmentation, the features slightly tend to decrease. So, it can be inferred that segmentation might reduce the quality of the feature vector. But still segmentation is done and compared with other feature extraction techniques to prove if it reduces the quality of feature set.

For Feature Extraction, GLCM, HOG, SIFT and SFM are used. All of them are used with the raw images after resizing and the segmented images to analyze the accuracies of the models.

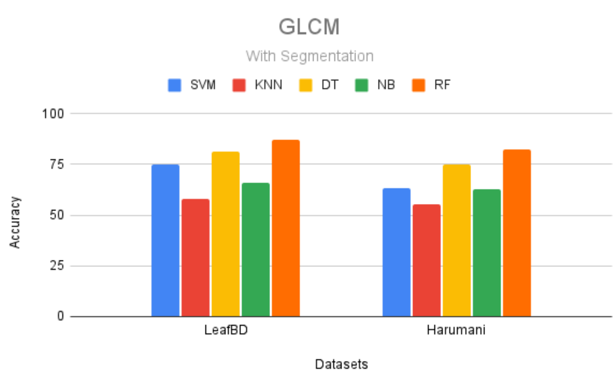

Figure 10: GLCM With Segmentation

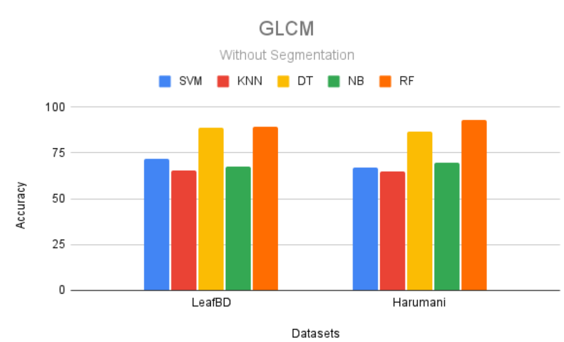

Figure 11: GLCM Without Segmentation

From figures 10 and 11 can be observed that the feature extraction technique, GLCM, when used without segmentation performs better than with segmentation. In both, leafBD and Harumani, datasets GLCM works better when Random Forest classifier is used for classification of the disease leaves.

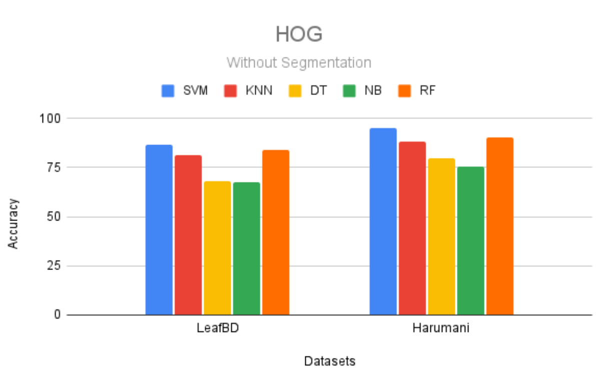

Figure 12: HOG With Segmentation

Figure 13: Hog Without Segmentation

The bar graphs, figures 12 and 13 show us the performance of the classifiers when used with HOG feature extraction technique. HOG with SVM performs the best in both the datasets, except on the Harumani dataset and with segmentation, where KNN performs better than other classifiers with an accuracy of 90.8.

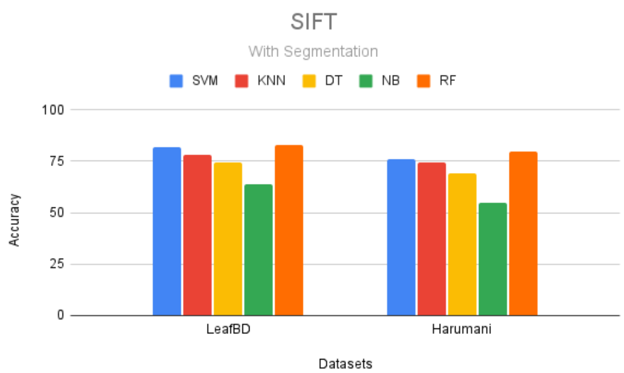

Figure 14: SIFT With Segmentation

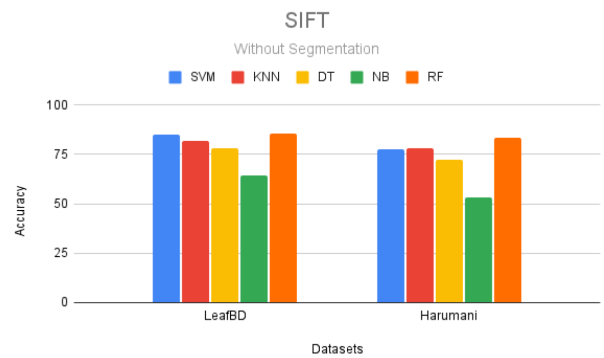

Figure 15: SIFT Without Segmentation

By looking at the bar plots, figures 14 and 15, it can be inferred that, SIFT performs the best with Forest Classifier, but on an average, it is the feature extraction technique with least performing accuracy.

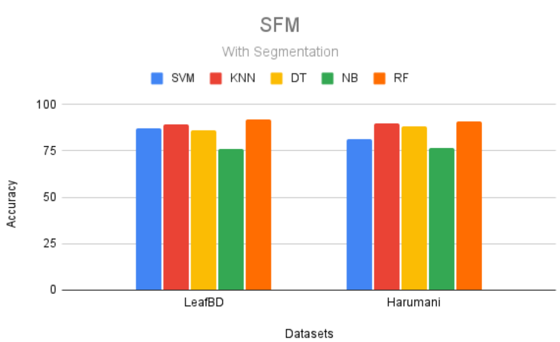

Figure 16: SFM With Segmentation

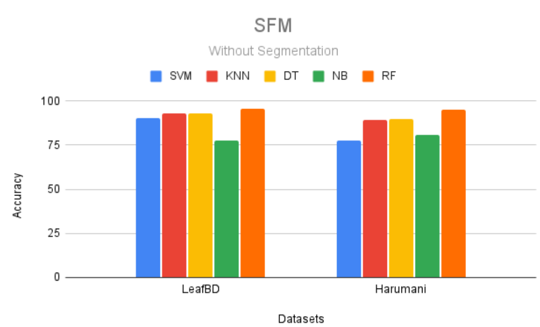

Figure 17: SFM Without Segmentation

The figures 16 and 17 show the performance of the classifiers when SFM is used as a feature extraction technique. It is the best yielding feature extraction when coupled with Random Forest classifier. Out of the sixteen cases, twelve times, Random Forest Classifier proved to be the best classifier with SFM as the feature extraction mechanism.

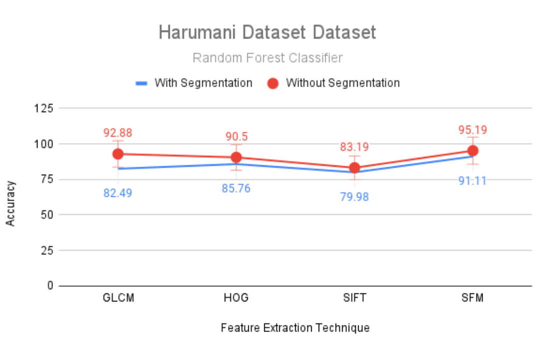

Figure 18: LeafBD Dataset with Random Forest

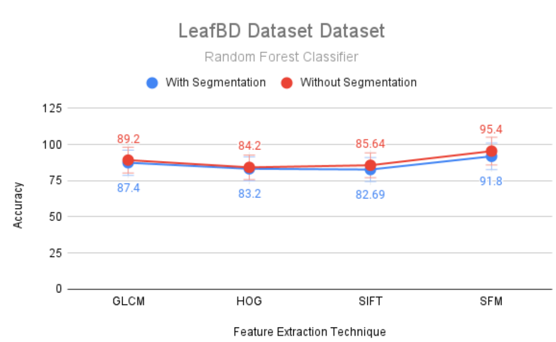

Figure 19: Harumani Dataset with Random Forest

As the Random Forest Classifier, when coupled with SFM, has performed the best, plotting of the bar graphs is done to check if it works the best with segmentation or now on both the datasets. For the leafBD dataset Random Forest classifier gives the accuracy of 95.4 while for the Harumani dataset we get 95.19

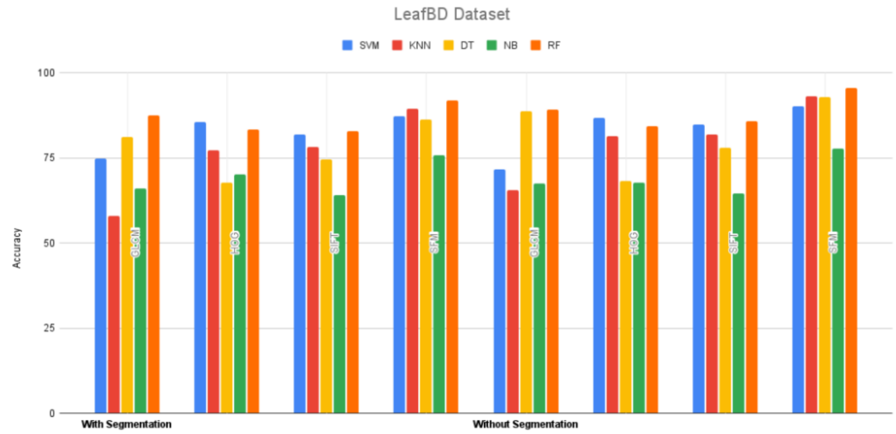

Figure 20: LeafBD Comparison

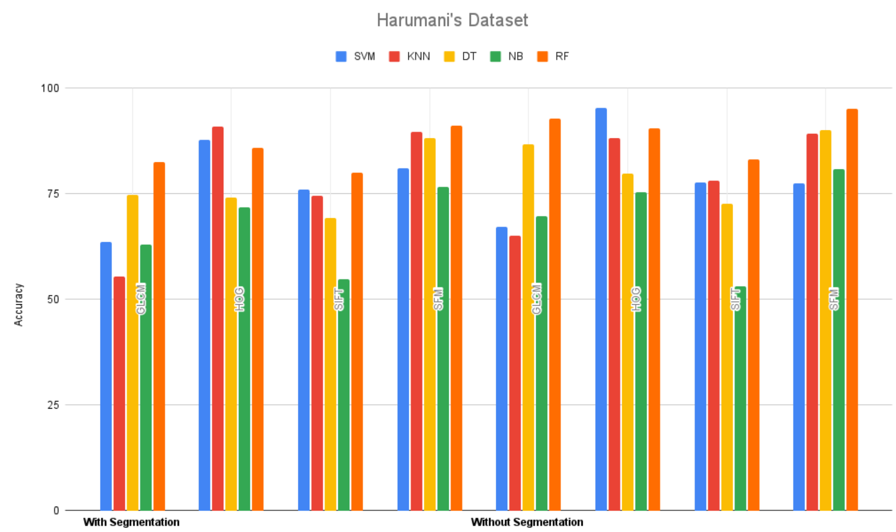

Figure 21 Harumani Dataset Comparison

both the datasets, on an average Random Forest classifier outperforms the other classifiers and the next highest being the support vector machine. Overall, SIFT has the lowest accuracy and Naïve Bayes classifier is not able to classify the data better using any of the feature extraction techniques.

5 CONCLUSION AND FUTURE WORK

This study has conducted an in-depth investigation into various methodologies for classifying mango leaves as healthy or diseased. Initially, K-means segmentation was explored as a means of delineating regions of interest, but its efficacy was limited when applied to the datasets. Instead, a shift was made towards employing feature extraction techniques as an alternative to segmentation, yielding more promising outcomes. Specifically, techniques like Gray-Level Co-occurrence Matrix (GLCM), Histogram of Oriented Gradients (HOG), Scale-Invariant Feature Transform (SIFT), and Statistical Feature Matrix (SFM) were evaluated, among which SFM demonstrated superior performance in capturing relevant characteristics.

Furthermore, five distinct classifiers were employed to evaluate their effectiveness in categorizing mango leaves. Notably, the Random Forest classifier emerged as the frontrunner, showcasing remarkable accuracy rates. The model achieved an impressive accuracy of 95.15% for the Harumani dataset and 95.4% for the LeafBD dataset, both without relying on segmentation.

Looking ahead, the research trajectory involves the integration of multiclass classifiers, incorporating a broader range of disease categories. Additionally, feature selection methods will be explored to identify the most discriminative attributes for classification. Furthermore, the strategy of combining multiple feature extraction techniques will be pursued to enhance the classification process. By building upon these insights, the aim is to refine and expand the capabilities of mango leaf disease classification, fostering advancements in plant health monitoring and contributing to improved agricultural practic

REFERENCES

[1] A. Singh, A. Kumar, and S. Singh, "Mango leaf disease detection using texture features and machine learning techniques," Computers and Electronics in Agriculture, 2021, vol. 184, p. 106003

[2] Ghosal, S., Bhowmik, P., & Das, A. A comparative study of machine learning algorithms for plant disease detection. 2nd International Conference on Inventive Communication and Computational Technologies, 2018 (pp. 1-5).

[3] Zhang, Y., Zhang, Y., & Wang, Y. A comparative study of machine learning algorithms for plant disease identification. In 2019 2nd International Conference on Computer Science and Software Engineering, 2019 (pp. 1-5). IEEE.

[4] Sa'ed Abed and Anwar Ali Esmaeel, "A Novel Approach to Classify and Detect Bean Diseases based on Image Processing", 2018.

[5] Saleem, Rabia, Jamal Hussain Shah, Muhammad Sharif, Mussarat Yasmin, Hwan-Seung Yong, and Jaehyuk Cha. "Mango leaf disease recognition and classification using novel segmentation and vein pattern technique." Applied Sciences 11, no. 24 (2021): 11901.

[6] Herng, Ooi Wei, Aimi Salihah Abdul Nasir, Ong Boon Chin, and Erdy Sulino Mohd Muslim Tan. "Harumani Mango Leaves Image Segmentation on RGB and HSV Colour Spaces using Fast k-Means Clustering." In Journal of Physics: Conference Series, vol. 2107, no. 1, p. 012068. IOP Publishing, 2021.

[7] Panigrahi, Kshyanaprava Panda, Himansu Das, Abhaya Kumar Sahoo, and Suresh Chandra Moharana. "Maize leaf disease detection and classification using machine learning algorithms." In Progress in Computing, Analytics and Networking: Proceedings of ICCAN 2019, pp. 659-669. Springer Singapore, 2020.

[8] Sumit Nema, Aarju Dixit, "Wheat Leaf Detection and Prevention Using Support Vector Machine", 2018 IEEE International Conference on Circuits and Systems in Digital Enterprise Technology

[9] Kaur, Navneet. "Plant leaf disease detection using ensemble classification and feature extraction." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12, no. 11 (2021): 2339-2352.

[10] Dissanayake, M. L. K., & Gunasekera, C. P. (2019). Mango diseases and their management. In Mango Diseases (pp. 1-22). Springer, Singapore.

[11] Singh, A., & Singh, S. (2018). A review on plant disease detection using image processing techniques. Journal of Advances in Biology & Biotechnology, 17(1), 1-10.

[12] Akhtar, M. S., & Siddiqui, M. W. (2019). Mango anthracnose: a review of the disease and its management. Journal of Plant Protection Research, 59(3), 225-233.

[13] Singh, A., & Singh, S. (2017). A review on anthracnose of mango. International Journal of Current Microbiology and Applied Sciences, 6(6), 1-8.