International Journal of Advances in Scientific Research and Engineering (ijasre)

E-ISSN : 2454-8006

DOI: 10.31695/IJASRE.2024.11.2

Volume 10, Issue 11
November - 2024

Comparative Experiments for Generation of Banana Black Sigatoka Disease Stages as Labels for Multiclass Classification Tasks

Edwin Kambo*[1], Devotha Nyambo[2], Judith Leo[2], Mussa Ally[2]

Research Scholar[1], Senior Lecture[2]

[1,2]School of Computation and Communication Science and Engineering,

The Nelson Mandela African Institution of Science and Technology

(NM-AIST), P.O.BOX 447

Arusha - Tanzania

ABSTRACT

Early and accurate diagnosis of Black Sigatoka (BSD), a fungal disease affecting banana production, is crucial for minimizing crop losses. An empirical understanding of the disease stages is important in developing predictors with adequate recommendations to farmers. This paper explores the use of K-means clustering algorithms to classify BSD stages in banana leaf images without relying on manual labelling. We evaluate the effectiveness of various image features, including "infected area", "colour histograms", "statistical features", and "texture features". The results indicate that using solely "infected area" achieved a moderate cluster separation, as observed from a Silhouette score of 0.5374 compared to others whose Silhouette scores were 0.3299 and 0.1366 for "statistical feature" and "colour histogram feature" respectively. Combining features, particularly, infected-area with texture or statistics, offered a promising balance between cluster separation and within-cluster variation as their Silhouette scores ranged between 0.32 and 0.47. Further investigation is needed to confirm the robustness of combining all features.

This research lays the groundwork for developing automated BSD classification systems to aid farmers in early disease detection and improved crop management. Automated classification using machine learning algorithms can significantly reduce the time and effort required for disease monitoring. Additionally, the integration of multiple image features could enhance the accuracy and reliability of the classification system. Future work will focus on validating these findings with larger datasets and exploring advanced machine learning techniques to further improve classification performance. Ultimately, the implementation of such systems could lead to better-informed decision-making in banana cultivation, reducing the impact of BSD on global banana production.

Key Words: Black Sigatoka (BSD), Banana leaf images, Disease stages, Image features, K-means clustering.

---

1. INTRODUCTION

Bananas are a critical component of global food security, particularly in Asia, Latin America, and Africa, where they are widely produced and consumed [1]. Figure 1 illustrates the average production of bananas as of 2017, highlighting that India and China dominate production primarily for domestic consumption. In Africa, especially in countries like Tanzania, bananas are a dietary staple and a vital source of income for millions[2]. However, the global banana industry faces a significant threat from Black Sigatoka, a fungal disease caused by Mycosphaerella fijiensis, which affects over 50 million hectares of banana crops, leading to an estimated 3% annual reduction in production and $1.6 billion in losses [3, 4].

Effective management of Black Sigatoka requires early and accurate diagnosis to minimize losses[5]. Traditional visual inspection methods are prone to subjectivity and scalability issues, whereas automated, data-driven solutions offer promising alternatives[4, 6]. Automated solutions rely heavily on large datasets, which can be labelled or unlabelled, corresponding to supervised and unsupervised machine learning techniques, respectively. For classification tasks, supervised machine learning is generally preferred due to the availability of labelled data that facilitates model training [7]. Therefore, labelled data is essential for effective classification.

Numerous studies have focused on detecting Black Sigatoka in banana leaves [8-11], employing various methodologies to improve detection models. However, none of these studies have considered the stages of Black Sigatoka disease, a crucial aspect for early detection and effective management. Understanding disease stages is vital for timely intervention.

Currently, no public datasets categorize Black Sigatoka-infected banana leaves by disease stages. For instance, an existing dataset by [12] includes images of infected leaves but does not group them by disease stages. This paper addresses this gap by presenting comparative experiments using K-Means clustering algorithms with different feature extraction techniques to derive Black Sigatoka disease stages as class labels from an existing image dataset. The goal is to establish a robust and objective disease stages dataset to develop accurate multiclass classification models, thereby enhancing early detection and intervention strategies.

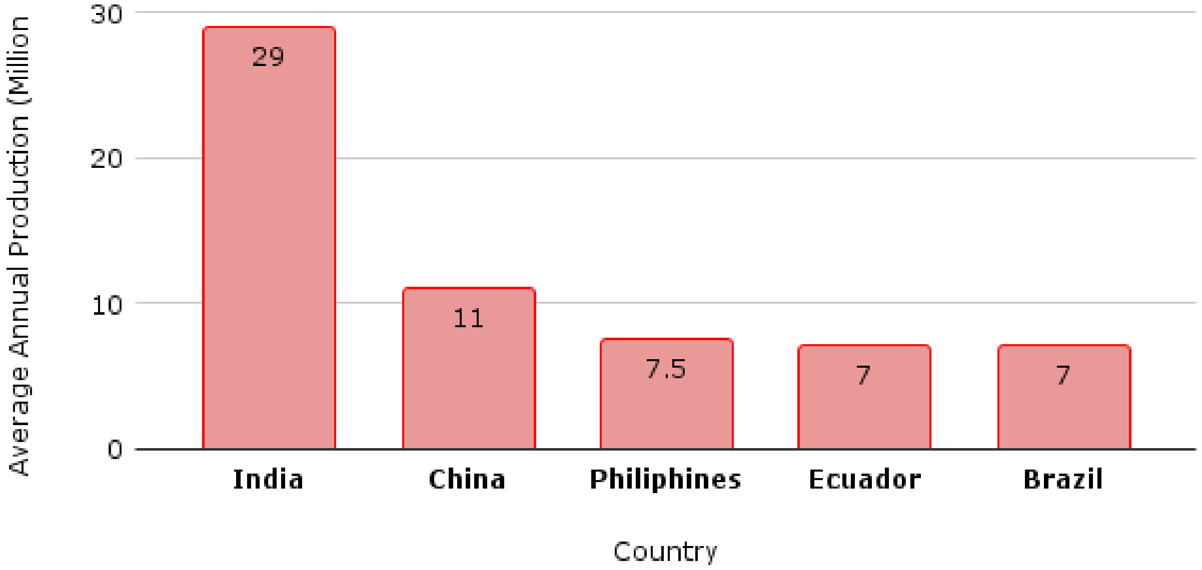

Figure 1: Top banana producers World Wide (2010-2017)

2. MATERIALS AND METHODS

The study area for this research was mainly from the regions of Kagera, Arusha, Dar es Salaam, Kilimanjaro and Mbeya where the main banana images for the dataset were obtained. Some other datasets were obtained from other parts of the world to complete the main dataset. This study is an experimental one where different experiments were carried out to obtain the required output.

2.1 Dataset description

2.1.1 Leveraging existing resources

This study utilised publicly available datasets containing banana leaf images with disease infestations, including Black Sigatoka. This strategic approach allowed us to reuse existing datasets from comparable locations.

2.1.2 Dataset selection

Three datasets were carefully chosen based on their relevance to our research goals. These datasets are described below:

Dataset of Banana Leave and Stem Images for Object Detection, Classification and Segmentation: A Case of Tanzania [12]: This dataset, characterized in Table 1, offers diverse banana leaf and stem images, including examples infected with Black Sigatoka.

Table 1: Facts about Dataset by [12]

Fact	Description
Dataset name	Dataset of banana leaf and stem images for object detection, classification and segmentation: A case of Tanzania
Location	Tanzania
Data type	Images
Content	Healthy Banana Leaves (5628), Black Sigatoka (5767) and Fusarium Wilt Race 1 (4697) infected banana leaves and stems
Collection method	Smartphones
Purpose	Develop machine learning models for early detection of banana diseases
Accessibility	Mendeley Data, https://doi.org/10.1016/j.dib.2023.109322

BananaLSD: A banana leaf images dataset for the classification of banana leaf diseases using machine learning by [13]. Table 2 provides annotated leaf images with various diseases, including Black Sigatoka and other diseases that affects the banana plant.

Table 2: BananaLSD description

Fact	Description
Dataset name	BananaLSD
Location	Gazipur, Bangladesh
Data type	Images
Content	Images of healthy and Pestalotiopsis (400), Sigatoka (400), Cordana (400) diseased banana leaves
Collection method	Smartphones
Purpose	Classification of banana leaf diseases
Accessibility	Mendeley Data, https://doi.org/10.1016/j.dib.2023.109608

PSFD-Musa: A dataset of banana plants, stem, fruit, leaf, and disease by [14]. Table 3 provides comprehensive metadata. PSFD-Musa offers data on various banana diseases, including Black Sigatoka. Notably, the authors addressed data imbalance through image augmentation.

Table 3: PSFD-Musa dataset description

Fact	**Description**
Dataset name	PSFD-Musa: A dataset of banana plant, stem, fruit, leaf, and disease
Location	BORTARI VILLAGE, Chaygaon, Kukurmara, District – Kamrup (Rural), Assam, India. HAJO VILLAGE, District – Kamrup (Rural), Assam, India.
Data type	Images
Content	An image dataset of varieties of banana plants and the diseases related to them: Bacterial Soft Rot (1078), Banana Fruit Scarring Beetle (150), Black Sigatoka (474), Yellow Sigatoka (264), Panama disease (102), Banana Aphids (366), and Pseudo-Stem Weevil (2736).
Collection method	Smartphones
Purpose	Machine learning related tasks
Accessibility	Mendeley Data, https://doi.org/10.1016/j.dib.2022.108427

Our study targeted images of banana leaves exhibiting solely Black Sigatoka disease across these diverse datasets. The process of combining images from multiple sources ensured data variety for model training. Initial visual inspection revealed that no single dataset provided sufficient images of Black Sigatoka stages due to differing initial data collection purposes and a potential lack of focus on stage classification. Therefore, incorporating images from various sources significantly increased the likelihood of capturing the complete spectrum of Black Sigatoka stages.

2.2 Data preprocessing

The study performed data preprocessing which is said to be a crucial step in the data analysis and machine learning pipeline that involves cleaning and transforming raw data into a format suitable for analysis or model training. The following is the description of the activities that were carried out in preprocessing the data.

2.2.1 Renaming the images

The images selected were all renamed to have one common structure. The structure of the name was a combination of a string and a number. For example, image_1, image_2, and image_3 with a sequential increase of numbers from 1 to n where n is the total number of all images of BSD-infected leaves. This process created uniformity and made it easier to manage the images and avoid unnecessary duplication.

2.2.2 Image resizing

Image resizing offers several advantages for deep learning applications, including computational efficiency, reduced memory usage, and consistent input size requirements. Convolutional neural networks (CNNs), for instance, necessitate consistent image dimensions across layers [15]. The optimal image size for a given task depends on the specific application. Studies have demonstrated improved performance with image sizes of 80x80, 120x120 or 224x224 for deep-learning tasks [16]. While larger images retain more information, they come at the expense of computational performance. In this study, all images were resized to 512x512 pixels to balance information retention and performance demands.

2.2.3 Data augmentation

To address the data imbalance found within the dataset, geometric transformations such as flipping, rotating, zooming, shifting, and shearing were applied via data augmentation techniques. The ImageDataGenerator class in Keras facilitated this augmentation, a well-established approach for enhancing image classification tasks, as demonstrated in prior research[17-19].

2.3 Experimental setup

2.3.1 BSD stages

The study aimed to classify Black Sigatoka (BSD) infection stages in banana leaves using unlabelled images. The study adopted a four-stage categorization system based on the work of [20] aligning with the specific needs of the research. This system groups the six stages described by Pérez-Vicente [21] and Fouré, Grisoni [22] into four broader categories:

- Healthy: No visible signs of BSD infection.
- Early: Encompasses stages I and II, characterized by faint reddish-brown specks and streaks.
- Medium: Combines stages III and IV, presenting larger, coalescing dark brown streaks and spots.
- Advanced: Represents stages V and VI, characterized by necrotic areas, yellow halos, and ascospore production.

Figure 2 provides visual examples of BSD progression across these stages, while Figure 3 depicts the updated categorization scheme. While the six-stage system offers finer detail, the four-stage approach aligns better to group unlabelled images into broader categories suitable for our chosen clustering algorithms.

DOI: 10.31695/IJASRE.2024.11.2

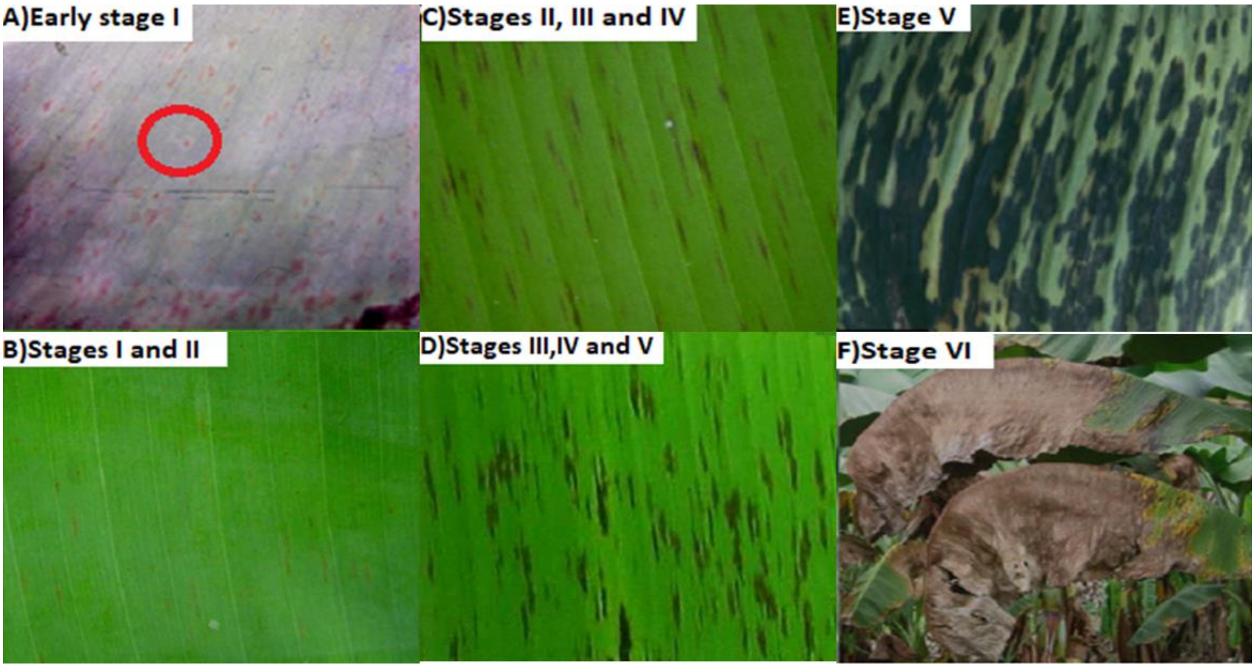

Figure 2: BSD at different stages[21]

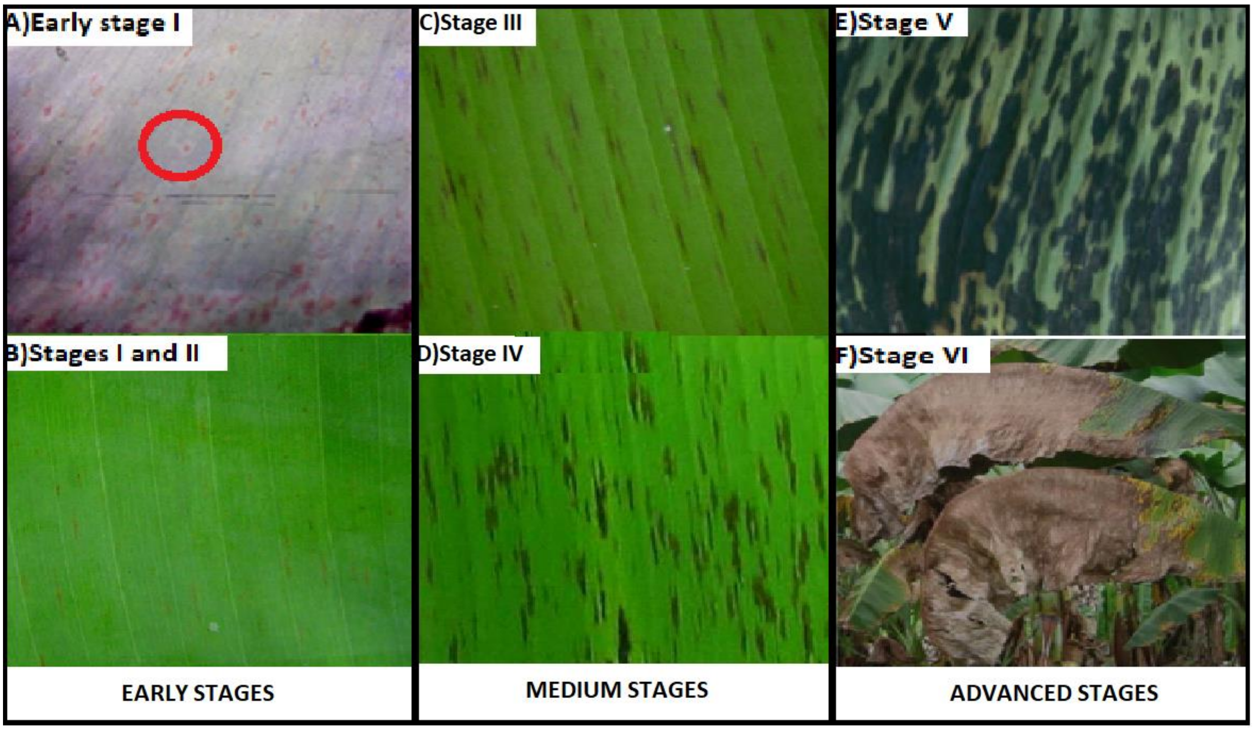

Figure 3: A four-stage approach for categorizing BSD

2.3.2 Rationale for four-stage system:

The decision to utilize the four-stage system stems from several key considerations: One is data Separability: As observed in Figure 3, particularly in sections C and D, differentiating early stages (I and II) can be challenging due to subtle visual differences. Grouping these stages into "Early" simplifies classification while capturing the essential distinction from healthy leaves. Two, clustering Algorithm Compatibility; many clustering algorithms struggle with large numbers of categories. The four-stage system reduces the number of classes, improving the suitability of chosen algorithms for our task. Lastly, it is because our primary goal is to develop a framework for grouping unlabelled images into meaningful categories. The four-stage system provides a practical balance between detailed information and feasibility for our chosen approach.

DOI: 10.31695/IJASRE.2024.11.2

2.3.3 Choice of clustering algorithm

The K-means clustering presents a valuable option for the task at hand due to its specific strengths and the nature of the data being analysed. The choice of K-means clustering algorithm for categorizing banana leaf images based on Black Sigatoka infection stages is well-suited due to several factors. The visual progression of the disease aligns with the core principle of K-means cluster formation, where data points are grouped based on their similarity. The algorithm will leverage features like the area of infection, colour histograms, and descriptive statistics to quantify the visual cues associated with each stage. The area of infection directly measures the extent of disease spread; colour histograms capture the chromatic shifts from healthy tissue to necrotic lesions; and statistics offer a numerical description of the distribution and nature of the diseased areas. By combining these features, K-means has the potential to effectively differentiate between the early, middle, and advanced stages of Black Sigatoka, facilitating the categorization of the image dataset.

2.3.4 Cluster evaluation metrics

To assess the quality and efficacy of the clustering results, the following evaluation metrics were used. Silhouette Score which measures how well data points are placed within their assigned clusters compared to how close they lie to points in neighbouring clusters. A higher Silhouette score indicates better-defined clusters[23]. Davies-Bouldin Index which calculates the average similarity between each cluster and its most similar cluster. A lower Davies-Bouldin index signifies better separation between clusters[24]. Calinski-Harabasz Index which evaluates cluster quality based on the ratio between inter-cluster dispersion (how spread-out clusters are from each other) and intra-cluster dispersion (how tightly packed points are within a cluster). A higher Calinski-Harabasz index implies better-defined clusters[25].

2.3.5 Experiment one: assessing infected leaf area as a feature for black sigatoka stage classification

This initial experiment investigated the suitability of the infected area as a feature for classifying Black Sigatoka (BSD) stages. Python programming language was employed for image processing and analysis. The OpenCV library provided functionalities for image manipulation, while Scikit-learn offered the K-means clustering algorithm and evaluation metrics.

Images were resized to 500x500 pixels for consistency. Conversion to grayscale was achieved using the cv2.COLOR_BGR2GRAY function, followed by thresholding via Otsu's method, is a widely recognized technique for automatic image segmentation based on intensity values[26]. This process effectively transformed the images into a binary format. The countNonZero function from OpenCV facilitated the calculation of the infected area ratio, which was subsequently considered as the feature representing the infected region.

The extracted area features for all images in the dataset were stored in a list, and this list served as input for the K-means clustering algorithm obtained from Scikit-learn. The elbow method was employed to determine the optimal number of clusters ('k'). The results depicted in Figure 4 guided the selection of k = 3 for the subsequent clustering process.

2.3.6 Experiment two: utilizing colour histograms for black sigatoka stage clustering

The second experiment explored the efficacy of colour histograms as features for classifying BSD stages. The study employed the calcHist function from the OpenCV library where colour histograms were computed for each image. These histograms were subsequently flattened using the flatten() function and normalized to ensure consistent feature representation. This process was repeated for all images in the dataset, and the extracted features were aggregated into a list for subsequent use with the K-means clustering algorithm. The chosen evaluation metrics were then computed, and the obtained results were stored for further analysis.

2.3.7 Experiment three: leveraging statistical features for Black Sigatoka stage classification

The third experiment investigated the potential of statistical features extracted from infected leaf regions for classifying BSD stages using K-means clustering. A Python function carried the following operation in achieving the task at hand. Colour Channel Splitting: The function begins by separating the input image into its individual blue (b), green (g), and red (r) channels using OpenCV's cv2.split function. Moment Calculation: For each channel, the function calculated various statistical moments, including:

- Mean: Represents the average intensity value within the channel.
- Standard Deviation: Captures the variability of intensity values around the mean.
- Skewness: Measures the asymmetry of the distribution of intensity values.
- Kurtosis: Characterizes the "peakedness" of the distribution compared to a normal distribution.

- Feature Vector Construction: The calculated features (means, standard deviations, skewness, and kurtosis) for all three channels are then combined into a single feature vector using np.array. This vector encapsulated various statistical properties of the infected region's colour distribution.

By applying this function to each image in the dataset, a collection of feature vectors was obtained. These vectors served as input to the K-means clustering algorithm, allowing it to group images based on the similarities and differences captured by the statistical features.

2.3.8 Experiment four: exploring texture features for Black Sigatoka stage classification

The fourth experiment delved into the realm of texture features, specifically investigating whether Local Binary Patterns (LBP) could aid in K-means clustering for BSD stage classification. This approach explored the possibility of leveraging spatial relationships between pixel intensities within the infected regions.

The experiment commenced with the familiar pre-processing steps: Image Loading and Resizing: Images were loaded and resized to ensure consistency in the analysis. Grayscale Conversion: Utilizing OpenCV's Otsu method, the images were converted to grayscale for effective LBP computation. Then the Local Binary Pattern (LBP) calculations were carried out through a defined function. The function utilized the local_binary_pattern function imported from “skimage.feature” library to calculate LBP features. LBP compares a central pixel's intensity with its surrounding neighbours and encodes the results as a binary pattern, capturing local textural variations. The function then constructed an LBP histogram using the np.histogram function. This histogram represented the frequency of different LBP patterns encountered in the image, providing a statistical summary of the image's texture.

Importantly, the histogram is normalized to ensure consistency and prevent bias towards features with higher frequencies. By applying this function to each pre-processed image, a collection of LBP histograms was obtained. These histograms served as texture-based features for the K-means clustering algorithm.

2.3.9 Experiment five: combining features for enhanced Black Sigatoka stage classification

Building upon the individual feature explorations conducted in the previous experiments, the fifth experiment investigated the potential benefits of combining various features for improved K-means clustering performance in classifying BSD stages. This approach capitalizes on the unique strengths of different features to overcome potential limitations and achieve more robust clustering results.

The following combinations were evaluated:

- Area and Colour Histogram: This combined the infected area size, capturing the extent of disease spread, with the colour distribution information from the colour histogram.
- Area and Statistical Features: This merged the area information with various statistical properties (mean, standard deviation, skewness, kurtosis) extracted from the infected regions, potentially providing a more comprehensive understanding of the infected area's characteristics.
- Area, Colour Histogram, and Texture Features: This combined all previously explored features, encompassing the area size, colour distribution, and spatial texture patterns represented by LBP histograms.
- All Features (Area, Colour Histogram, Texture, and Statistical Features): This experiment utilized the full spectrum of features extracted in the preceding experiments, providing the most comprehensive feature set for clustering.

The rationale behind combining features lies in the principle of complementarity. Each feature captures a specific aspect of the infected leaf region. By combining features, the clustering process can leverage the strengths of each feature to compensate for potential shortcomings in others, leading to a more refined and informative representation for clustering. This experiment aimed to determine if the combined feature sets can significantly improve the quality and accuracy of Black Sigatoka stage classification using K-means clustering.

DOI: 10.31695/IJASRE.2024.11.2

3. RESULT AND DISCUSSION

3.1 Results of experiments one to five

Figure 4 shows the result of the elbow method used to obtain the best k values to be used with the K-Means algorithm. From the figure, k was set to 3, representing the three stages of the disease besides the health leaves that form the fourth stage.

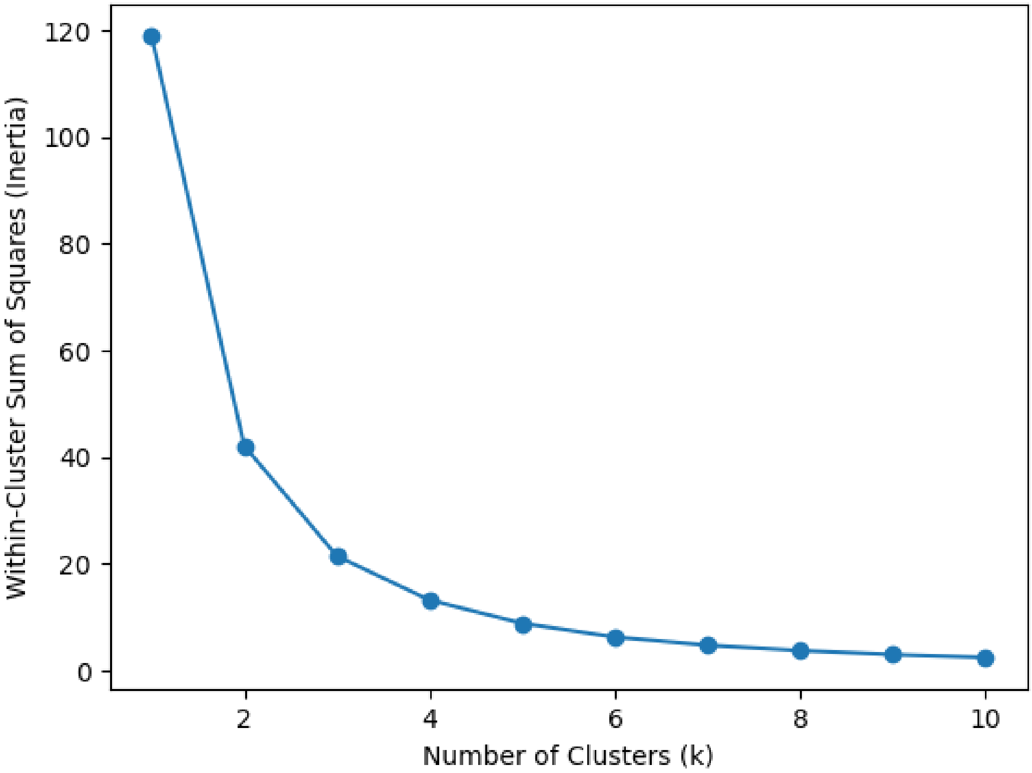

Figure 4: Results of the elbow method for optimal k value.

Table 4 shows the results of experiments one to five which involved the area feature, colour histogram feature, statistic feature, texture feature and a combination of different features.

Table 4: K-Means Clustering Results with Different Features

Feature Type	Cluster Assessment Criteria	Score
Area	Silhouette Score	0.5374
	Davies-Bouldin Index	0.5604
	Calinski-Harabasz Index	13112.89
Colour Histogram	Silhouette Score	0.1367
	Davies-Bouldin Index	2.3942
	Calinski-Harabasz Index	648.86
Statistics	Silhouette Score	0.3299
	Davies-Bouldin Index	1.0168
	Calinski-Harabasz Index	5136.79
Texture	Silhouette Score	0.4719
	Davies-Bouldin Index	0.6746
	Calinski-Harabasz Index	9741.91
Area + Colour Histogram	Silhouette Score:	0.1058
	Davies-Bouldin Index:	2.2318
	Calinski-Harabasz Index:	917.45
Area + Statistics	Silhouette Score:	0.3299

DOI: 10.31695/IJASRE.2024.11.2

	Davies-Bouldin Index:	1.0169
	Calinski-Harabasz Index:	5136.70
Area + Histogram + Texture	Silhouette Score:	0.1048
	Davies-Bouldin Index:	2.2464
	Calinski-Harabasz Index:	904.22
Area + Colour Histogram + Texture + Statistics	Silhouette Score:	0.3295
	Davies-Bouldin Index:	1.0186
	Calinski-Harabasz Index:	5136.53

3.2 Discussion of the results

In Table 4, the study evaluated the performance of different feature sets in clustering using three important metrics: Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index. These results offer valuable insights into how effectively each feature separates clusters.

Starting with the Silhouette Score, the “area” feature stood out with the highest score of 0.5374, indicating that it achieved a moderately good level of cluster separation. This suggests that using the area as a feature allows for clear and distinct clustering. On the other hand, the “Colour Histogram” feature had the lowest Silhouette Score of 0.1367, which highlights the difficulty of achieving effective clustering with this feature. Most of the other features and combinations, especially those including “Area” or “Texture” feature, scored between 0.32 and 0.47. This suggests a moderate level of separation, indicating that these features do contribute positively to clustering performance.

Looking at the Davies-Bouldin Index, the study found that a combination of “Area, Colour Histogram, Statistics, and Texture” had the lowest score of 1.0186. This suggests a potentially better separation between clusters compared to other features. In the context of this index, a lower score indicates less overlap between clusters and thus better performance. Once again, the “Colour Histogram” feature had the highest Davies-Bouldin Index score of 2.3942, reinforcing its weaker performance in separating clusters effectively. A similar trend was observed here as with the Silhouette Score, where lower scores generally correspond to better separation.

The Calinski-Harabasz Index further highlights the effectiveness of the “Area” feature, which achieved the highest score of 13112.89. This suggests significant cluster separation with consistent within-cluster variances, showing that “Area” is a robust feature for clustering. Conversely, the “Colour Histogram” feature had the lowest Calinski-Harabasz Index score of 648.86, implying limited separation and potentially high within-cluster variations. Feature combinations that include “Area” generally achieved scores between 5136 and 9742, indicating decent separation but potentially higher within-cluster variations compared to using “Area” alone.

Overall, the “Area” feature consistently showed strong performance across all metrics, making it the most effective feature for achieving good cluster separation. In contrast, the “Colour Histogram” feature consistently underperformed, revealing its limitations in clustering. However, combinations of features, particularly those including “Area” and “Texture”, demonstrated moderate to strong performance. This suggests that they can enhance clustering outcomes when used alongside the “Area” feature. Figure 5 provides a visualization of the comparisons of these experiments.

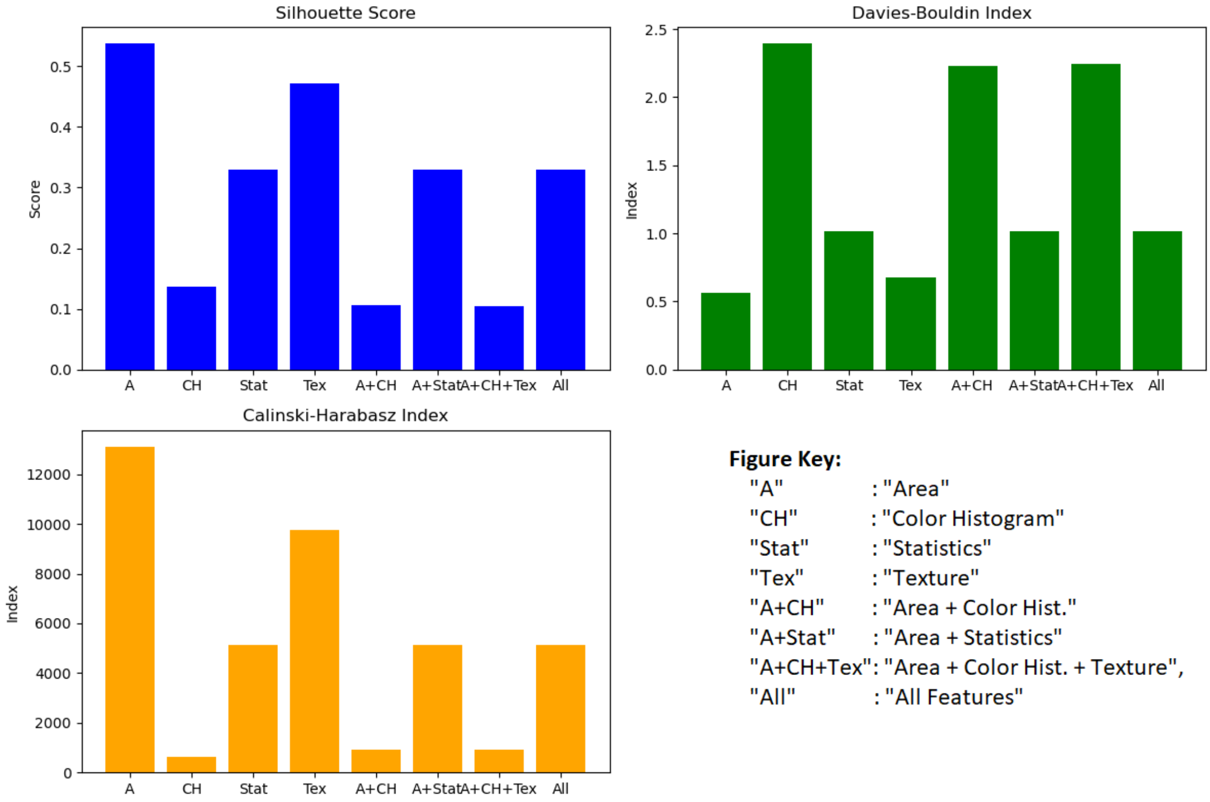

Figure 1: A comparison of the performance of different features in clustering BSD stages

From these results, while it can be observed that the “Area” feature alone achieved the highest Silhouette Score, it might not be the most robust choice due to potential limitations in capturing the full complexity of BSD stages as suggested by [27]. Further analysis and inspection of the clusters will be necessary to confirm. Numerous studies have explored the use of feature extraction for disease detection, employing the K-means algorithm and other classification techniques such as SVM. For instance, Chaudhari and Patil [28] evaluated the accuracy of the SVM model. Unlike our research, which emphasizes the features themselves for subsequent classification tasks, their study primarily assessed the model's accuracy.

4. CONCLUSION

This study investigated the feasibility of utilizing K-means clustering algorithms to classify Black Sigatoka (BSD) disease stages in banana leaves using unlabelled images. The exploration encompassed various image features, including infected leaf area, colour histograms, statistical features, and texture features.

The findings highlight the potential of K-means clustering for this task. However, the effectiveness of the clustering process hinges on the selection of appropriate features. Here, we observed that: Solely relying on area feature achieved a moderate level of cluster separation but might not capture the full complexity of BSD stages. Colour histograms alone exhibited poor performance in differentiating between disease stages. Combining features like area with texture or statistical features appeared to offer a promising balance between cluster separation and within-cluster variation. This suggests that leveraging complementary information from multiple features can enhance the clustering process. While the combination of all features yielded the lowest Davies-Bouldin Index, indicating potentially better separation, further investigation is necessary to confirm the robustness of these clusters.

For a more improved results domain knowledge needs to be incorporated into feature selection and cluster evaluation to ensure practical relevance for disease classification tasks. There is also a need to investigate alternative clustering algorithms that may be better suited for specific feature combinations or complex data structures. Validating the derived clusters using ground truth data from human experts needs to be carried out to assess the accuracy and effectiveness of the proposed approach for real-world applications. By addressing these limitations and exploring new avenues, this research can pave the way for developing robust and automated disease classification systems that empower farmers with early detection capabilities and contribute to improved crop management practices.

ACKNOWLEDGMENT

The authors express their sincere appreciation to The Nelson Mandela African Institution of Science and Technology and Mkwawa University College of Education for offering exceptional learning facilities and working environments.

REFERENCES

1. FAO. *Banana facts and figures*. 2018 [cited 2024 04/02/2024]; Available from: https://www.fao.org/economic/est/est-commodities/oilcrops/bananas/bananafacts/en/.
2. Lucas, S.S. and K.E. Jomanga, *The Status of Banana Production in Tanzania; a Review of Threats and Opportunities*. 2021.
3. Lopez, V. and L. Perez-Vicente, *FAO Response to the Black Sigatoka Disease Problem in the Caribbean*. 2013.
4. Luna-Moreno, D., et al., *Early detection of the fungal banana black sigatoka pathogen Pseudocercospora fijiensis by an SPR immunosensor method.* Sensors, 2019. **19**(3): p. 465.
5. Sinha, A. and R.S. Shekhawat, *Review of image processing approaches for detecting plant diseases.* IET Image Processing, 2020. **14**(8): p. 1427-1439.
6. Ugarte Fajardo, J., et al., *Effective Methods Based on Distinct Learning Principles for the Analysis of Hyperspectral Images to Detect Black Sigatoka Disease.* Plants, 2022. **11**(19): p. 2581.
7. Cocco, S., et al., *Supervised learning: classification with neural networks*, in *From Statistical Physics to Data-Driven Modelling: with Applications to Quantitative Biology*. 2022, Oxford University Press. p. 137 - 160.
8. Elinisa, C.A. and N. Mduma, *Mobile-Based Convolutional Neural Network Model for the Early Identification of Banana Diseases.* Smart Agricultural Technology, 2024: p. 100423.
9. Escudero, C.A., A.F. Calvo, and A. Bejarano, *Black Sigatoka Classification Using Convolutional Neural Networks.* International Journal of Machine Learning and Computing, 2021. **11**(4).
10. Tovar-Martínez, Y.T., A. Bejarano-Martínez, and A.F. Calvo-Salcedo, *Mobile application for the detection of black Sigatoka.* Visión electrónica, 2020. **14**(1): p. 111-118.
11. Yumang, A.N., et al. *Detection of Black Sigatoka Disease on Banana Leaves Using ShuffleNet V2 CNN Architecture in Comparison to SVM and KNN Techniques*. in *2023 15th International Conference on Computer and Automation Engineering (ICCAE)*. 2023. IEEE.
12. Mduma, N. and J. Leo, *Dataset of Banana Leave and Stem Images for Object Detection, Classification and Segmentation: A Case of Tanzania.* Data in Brief, 2023: p. 109322.
13. Arman, S.E., et al., *BananaLSD: A banana leaf images dataset for classification of banana leaf diseases using machine learning.* Data in Brief, 2023. **50**: p. 109608.
14. Medhi, E. and N. Deb, *PSFD-Musa: A dataset of banana plant, stem, fruit, leaf, and disease.* Data in Brief, 2022. **43**: p. 108427.
15. Rukundo, O., *Effects of image size on deep learning.* Electronics, 2023. **12**(4): p. 985.
16. Altabeiri, R., M. Alsafasfeh, and M. Alhasanat, *Image compression approach for improving deep learning applications.* International Journal of Electrical and Computer Engineering (IJECE), 2023. **13**(5): p. 5607-5616.
17. Ahmed, T.U., et al. *Facial expression recognition using convolutional neural network with data augmentation*. in *2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR)*. 2019. IEEE.
18. Soliman, A. and J. Terstriep. *Keras Spatial: Extending deep learning frameworks for preprocessing and on-the-fly augmentation of geospatial data*. in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. 2019.
19. Zak, J., et al., *Cell image augmentation for classification task using GANs on Pap smear dataset.* Biocybernetics and Biomedical Engineering, 2022. **42**(3): p. 995-1011.

20. Escudero, C.A., et al., *Development of a digital image classification system to support technical assistance for Black Sigatoka detection.* Revista Brasileira de Fruticultura, 2021. **43**.
21. Pérez-Vicente, L., *A holistic integrated management approach to control black Sigatoka disease of banana caused by Mycosphaerella fijiensis.* Food and Agriculture Organization of the United Nations, 2012: p. 30.
22. Fouré, E., M. Grisoni, and R. Zurfluh, *Les cercosporioses du bananier et leurs traitements. Comportement des variétés. Etude de la sensibilité variétale des bananiers et plantains à Mycosphaerella fijiensis Morelet et de quelques caractéristiques biologiques de la Maladie des raies noires au Gabon.* Fruits, 1984. **39**(6): p. 365-378.
23. İhsanoğlu, A. and M. Zaval. *Statistically Improving K-means Clustering Performance*. in *2024 32nd Signal Processing and Communications Applications Conference (SIU)*. 2024. IEEE.
24. Ros, F., R. Riad, and S. Guillaume, *PDBI: A partitioning Davies-Bouldin index for clustering evaluation.* Neurocomputing, 2023. **528**: p. 178-199.
25. Syahputri, Z. and M.A.P. Siregar, *Determining The Optimal Number of K-Means Clusters Using The Calinski Harabasz Index and Krzanowski and Lai Index Methods for Groupsing Flood Prone Areas In North Sumatra.* Sinkron: jurnal dan penelitian teknik informatika, 2024. **8**(1): p. 571-580.
26. Goh, T.Y., et al., *Performance analysis of image thresholding: Otsu technique.* Measurement, 2018. **114**: p. 298-307.
27. Chen, Y., et al., *Silhouette: Efficient cloud configuration exploration for large-scale analytics.* IEEE Transactions on Parallel and Distributed Systems, 2021. **32**(8): p. 2049-2061.
28. Chaudhari, V. and M. Patil. *Banana leaf disease detection using K-means clustering and Feature extraction techniques*. in *2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)*. 2020. IEEE.

DOI: 10.31695/IJASRE.2024.11.2