IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Classification Of Arecanuts Using Image Processing And Machine Learning Techniques

Vineetha D. S.[1], Dr. Savitha M. [2], Dr kusumadhara S .[3], Dr. Bhagya H. K. [4]
[1]Mtech, Digital Electronics and Communication,KVGCE, Sullia D.K, Karnataka, India
[2] M.tech,Ph.D,MISTE,E&C.Dept,ProfessorKVGCE,SulliaD.K,Karnataka,India
[3] M.tech,Ph.D,MISTE,E&C.Dept,ProfessorKVGCE,SulliaD.K,Karnataka,India
[4]M.tech,Ph.D,MISTE,E&C.Dept,ProfessorKVGCE,SulliaD.K,Karnataka,India

Abstract:
The Areca nut, otherwise called the betel nut, holds huge significance as a tropical yield, with India remaining as the second-biggest maker and purchaser universally. All through its development organizes, the areca nut faces various sicknesses influencing different parts from the roots to the natural products. Customarily, distinguishing these sicknesses has depended on manual visual reviews by ranchers, which isn't just work serious yet additionally inclined to errors.To alleviate these difficulties, we propose an inventive strategy utilizing AI procedures to robotize the quality recognition of areca nuts. Our methodology uses the K-Closest Neighbor (KNN) and Backing Vector Machine (SVM) calculations to characterize the nuts into classifications of positive or negative quality. For this reason, we fostered a custom dataset containing 208 pictures of areca nuts, including both great and terrible quality examples, to prepare and test our models.The KNN calculation accomplished a high precision of 95.12%, fundamentally beating the SVM calculation, which recorded an exactness of 65.68%. These outcomes feature KNN's prevalent presentation in distinctive the nature of areca nuts in light of picture separated features.Our study underscores the capability of AI in rural applications, giving a more effective, exact, and versatile answer for surveying areca nut quality. This innovative headway can fundamentally help ranchers by diminishing the time and work expected for crop examination and improving in general yield the executives. Future examination will zero in on augmenting the dataset, refining the calculations, and creating pragmatic apparatuses for certifiable application.

Index Terms – KNN,SVM,algorithums, machine learning.

I. Introduction

Areca nut, a tropical crop has its own commercial and economic significance. These areca nuts are impacted by numerous diseases and losses such as fruit rot, bud rot and nut splitting etc. When diseased and un-diseased areca nuts are mixed up, a reasonable price will not be found in the market. It is very important 2 to segregate the areca nuts as healthy and unhealthy so that all healthy areca nuts can be taken into consideration for the next process as it offers high quality. But it is very complex to manually inspect the health of nuts as it takes lots of skilled work, labor costs and time. Since agriculture is an ancient occupation in our country, farmers are restricted to human visualization capabilities by using the same old methods to recognize diseases. For limiting the cost of labor, for replacing the manual work and for saving time in agriculture field, an automatic detecting and classifying system is much needed. To achieve this goal, image processing is the significant technology being used. There are several approaches already available for classifying the processed areca nuts (without husk/boiled nuts). In this proposed system, raw areca nuts are being used for classification which will reduce the cost and effort as they got classified prior to the processing of nuts.Therefore, there is a growing need for automated disease classification. Machine learning algorithms can provide a solution by automating the quality assessment of areca nuts. In this project, we will use K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) to detect diseases in areca nuts.

KNN is a machine learning technique that captures images, assigns values to different elements within the images, and differentiates between them. One of the advantages of KNN over other partitioning methods is its relatively low processing time. While traditional methods require manual construction of filters, KNN can interpret these filters or symbols effectively with proper training.SVM is an excellent classification algorithm, commonly used in supervised learning to categorize data. SVM is trained on labeled data and is capable of solving both classification and regression problems.

II. RELATED WORK

The literature survey is an essential part of any research project, providing a comprehensive overview of the existing work in the field and identifying gaps that the proposed system aims to address. In the context of the automated classification of areca nuts, the literature survey focuses on prior research in areas such as image processing techniques, machine learning algorithms, agricultural automation, and specifically, the detection and classification of crop diseases.

1. Image Processing in Agriculture

Image processing has been widely used in agriculture for various applications, including disease detection, yield estimation, and quality assessment. Several studies have demonstrated the effectiveness of image processing techniques in enhancing agricultural productivity and reducing the reliance on manual labor.

Chaudhary et al. (2020) reviewed the role of image processing in precision agriculture, highlighting its applications in crop monitoring and disease detection. The study discussed the use of techniques such as color analysis, texture analysis, and shape detection to identify diseased areas in crops, which are directly relevant to the classification of areca nuts based on visual features.

Kulkarni and Patil (2018) developed an image processing system for detecting and classifying diseases in fruits. The system used color and texture features extracted from fruit images to identify common diseases. This work provides a foundation for applying similar techniques to the classification of areca nuts, where color and texture are critical indicators of nut health.

2. Machine Learning for Crop Disease Classification

Machine learning algorithms have gained significant attention in agriculture for their ability to automate complex tasks such as disease classification. The use of supervised learning algorithms, such as K-Nearest Neighbors (KNN) and Support Vector Machines (SVM), has been explored in various studies.

Patil and Kumar (2019) implemented a machine learning-based system to classify tomato leaf diseases using KNN and SVM. The study found that SVM outperformed KNN in terms of accuracy, especially when dealing with high-dimensional feature spaces. This suggests that SVM may be well-suited for the classification of areca nuts, which may involve complex feature interactions.

Singh et al. (2021) proposed a hybrid machine learning approach for the detection of plant diseases. The system combined the strengths of multiple algorithms, including KNN and SVM, to improve classification accuracy. This approach aligns with the proposed system's use of both KNN and SVM to leverage their respective strengths in areca nut classification.

3. Areca Nut Classification and Disease Detection

While the direct literature on areca nut classification using image processing and machine learning is limited, several studies have focused on similar crops or general techniques that can be applied to areca nuts.

Raj and Kumar (2022) conducted a study on the detection of areca nut diseases using a convolutional neural network (CNN). Although CNNs are powerful, the study highlighted the computational complexity and the need for large datasets, which may not always be feasible. The proposed system aims to address these challenges by using simpler, more interpretable algorithms like KNN and SVM.

Sharma et al. (2017) explored the use of texture analysis for the detection of rot in areca nuts. The study demonstrated that texture features could effectively distinguish between healthy and diseased nuts, forming a basis for feature extraction in the proposed system.

4. **Agricultural Automation and its Impact**

The broader field of agricultural automation has seen significant advancements, with research focusing on the integration of image processing, robotics, and machine learning to automate various tasks.

Ramesh et al. (2020) reviewed the impact of automation on agricultural productivity, emphasizing the role of machine learning in reducing labor costs and improving crop quality. The study highlighted the potential for machine learning to transform traditional farming practices, providing motivation for the development of automated systems like the one proposed for areca nut classification.

Zhang et al. (2019) developed an automated system for sorting and grading fruits based on image processing and machine learning. Their work showed that such systems could significantly improve the speed and accuracy of quality assessment, a key objective of the proposed areca nut classification system.

5. **Challenges and Future Directions**

Several studies have also highlighted the challenges associated with implementing image processing and machine learning in agriculture, particularly in resource-constrained environments.

Banerjee and Gupta (2021) discussed the limitations of machine learning in agricultural applications, such as the need for large, labeled datasets and the complexity of model training. They emphasized the importance of developing robust systems that can operate with limited data, which is a consideration in the design of the proposed system. Jha et al. (2023) reviewed the future of smart farming technologies, identifying key areas for research, including the need for more accurate and scalable disease detection systems. Their insights support the continued exploration of automated classification systems that can handle a variety of crops and diseases.

The literature survey reveals that significant progress has been made in applying image processing and machine learning to agricultural automation, particularly in the detection and classification of crop diseases. While there is substantial research on similar crops and general techniques, the specific application of these technologies to areca nut classification is still an emerging field. The proposed system seeks to fill this gap by developing a cost-effective, efficient, and accurate automated classification system for areca nuts, leveraging established image processing techniques and machine learning algorithms like KNN and SVM.

III. PROBLEM STATEMENT

The areca nut, a commercially significant tropical crop, plays a crucial role in the agricultural economies of several countries. However, the quality and market value of areca nuts are frequently compromised by various diseases, including fruit rot, bud rot, and nut splitting. These diseases affect the nuts' health, making it imperative to segregate healthy nuts from unhealthy ones before processing. Traditionally, this segregation is done manually, relying on the visual inspection skills of laborers. However, the manual inspection process is fraught with challenges:

Labor-Intensive and Costly: Manual inspection requires a significant amount of skilled labor, which is not only costly but also increasingly scarce due to urban migration and changing workforce dynamics.

Time-Consuming: The process of visually inspecting and sorting large quantities of nuts is slow, leading to delays in subsequent processing steps and affecting the overall efficiency of the supply chain.

Inconsistent Quality Assessment: Human inspection is prone to errors and inconsistencies, leading to variable quality in the sorted nuts. This variability can result in mixed batches of healthy and diseased nuts, ultimately reducing the market value of the produce.

Limited Scalability: As the demand for areca nuts grows, scaling up the manual inspection process becomes increasingly difficult, leading to bottlenecks in production.

Given these challenges, there is a pressing need for an automated system that can accurately and efficiently classify areca nuts based on their health status. Such a system would reduce the reliance on manual labor, increase the speed and accuracy of classification, and ensure that only high-quality nuts proceed to the next stages of processing and market distribution.

The proposed solution involves the development of an image processing-based system that utilizes machine learning algorithms, specifically K-Nearest Neighbors (KNN) and Support Vector Machines (SVM), to detect and classify diseases in areca nuts. This automated system aims to:

Enhance the Accuracy and Consistency: By leveraging the power of machine learning, the system will provide a more accurate and consistent classification of areca nuts compared to manual inspection.

Reduce Labor Costs: Automation will significantly reduce the need for skilled labor, leading to cost savings for farmers and producers.

Increase Efficiency: The system will process large quantities of areca nuts quickly, reducing the time required for classification and enabling faster throughput in the supply chain.

Improve Market Value: By ensuring that only healthy nuts are sold, the system will help farmers and traders achieve better prices, contributing to the overall profitability of the areca nut industry.

IV. OBJECTIVES

The proposed system is expected to achieve several key outcomes:

Improved Classification Accuracy: By using a combination of image processing and machine learning algorithms, the system is expected to achieve high accuracy in classifying areca nuts, reducing the likelihood of healthy and unhealthy nuts being mixed.

Enhanced Efficiency: The automated nature of the system will significantly reduce the time required for nut classification, increasing the overall efficiency of the production process.

Cost Reduction: By minimizing the need for manual inspection, the system will reduce labor costs and make the classification process more economically viable.

Consistency: The system will provide consistent classification results, eliminating the variability associated with human inspection.

V. IMPLEMENTATION

The methodology outlines the step-by-step approach to developing the automated areca nut classification system. It covers the system's design, data collection, preprocessing, feature extraction, model training, and evaluation. The goal is to create a reliable, efficient, and accurate system that can classify areca nuts based on their health status, distinguishing between healthy and unhealthy nuts.

1. **System Design and Architecture**

The design of the automated classification system involves several key components, each of which plays a critical role in the overall workflow:

- **Image Acquisition:** The first step in the system involves capturing high-resolution images of areca nuts. A camera setup will be designed to capture images as the nuts move along a conveyor belt or are placed on a platform. Ensuring consistent lighting and positioning is crucial to minimize variability in the images.

- **Data Preprocessing:** Preprocessing involves preparing the captured images for analysis by cleaning and enhancing them. This stage includes noise reduction, contrast adjustment, and background removal. The preprocessing ensures that the features extracted from the images are clear and accurate.
- **Feature Extraction**: After preprocessing, key features such as color, texture, and shape will be extracted from the images. These features are chosen based on their relevance to distinguishing between healthy and unhealthy nuts. For example:
  - Color Features: Average color values, color histograms, and color variance can be used to detect discoloration caused by diseases.
  - Texture Features: Texture analysis using methods like the Gray Level Co-occurrence Matrix (GLCM) will help in identifying surface irregularities or patterns indicative of disease.
  - Shape Features: Shape descriptors such as aspect ratio, roundness, and contour features will be used to identify any deformities in the nuts.
- **Machine Learning Model Selection:** The extracted features are fed into two machine learning algorithms—K-Nearest Neighbors (KNN) and Support Vector Machines (SVM). These algorithms are selected for their complementary strengths:
  - KNN: KNN is a simple, instance-based learning algorithm that classifies a new sample based on the majority class of its nearest neighbors in the feature space.
  - SVM: SVM is a robust algorithm that constructs a hyperplane to separate different classes with the maximum margin, making it effective for complex, high-dimensional classification tasks.
- **Classification and Sorting**: The classification results from the KNN and SVM models will be used to determine the health status of each areca nut. The system will then automatically sort the nuts into two categories—healthy and unhealthy—based on the model predictions.
- **System Integration:** The different components of the system will be integrated into a cohesive workflow, ensuring seamless data flow from image acquisition to classification and sorting.

2. **Data Collection and Dataset Preparation**

To train and evaluate the machine learning models, a labeled dataset of areca nut images is required. The dataset will be collected and prepared as follows:

Image Collection: A diverse set of areca nuts will be collected, including healthy nuts and those affected by various diseases (e.g., fruit rot, bud rot, nut splitting). High-resolution images of these nuts will be captured under controlled lighting conditions.

- **Labeling:** Each image will be manually labeled as "healthy" or "unhealthy" based on expert knowledge. The labeling process ensures that the dataset accurately reflects the health status of the nuts.
- **Dataset Augmentation:** To increase the size of the dataset and improve model generalization, data augmentation techniques such as rotation, flipping, scaling, and cropping will be applied to the images. This process helps simulate variations that the system might encounter in real-world scenarios.

- **Dataset Splitting:** The dataset will be divided into three subsets:
    - Training Set: Used to train the machine learning models.
    - Validation Set: Used to fine-tune the model parameters and prevent overfitting.
    - Test Set: Used to evaluate the final performance of the models on unseen data.

3. **Model Training**

The machine learning models will be trained on the prepared dataset as follows:

- **Feature Scaling**: Before training, the features will be normalized or standardized to ensure that all features contribute equally to the model's decision-making process.
- **KNN Training**: The KNN algorithm will be trained by storing the feature vectors of the training samples. During classification, the algorithm will compute the distance between the test sample and all training samples, assigning the test sample to the most common class among its K nearest neighbors.
- **SVM Training:** The SVM algorithm will be trained by finding the optimal hyperplane that maximizes the margin between the classes. The training process involves tuning the kernel function (e.g., linear, radial basis function) and regularization parameters to achieve the best classification performance.

4. **Model Evaluation**

The trained models will be evaluated on the test set to assess their performance:

- **Accuracy**: The percentage of correctly classified nuts in the test set.
- **Confusion Matrix:** A matrix showing the number of true positives, false positives, true negatives, and false negatives, providing insights into the types of errors made by the model.

The evaluation metrics will help determine which model—KNN or SVM—performs better in classifying areca nuts and whether any further improvements are needed.

5. **System Testing and Validation**

After training and evaluating the models, the entire system will be tested under real-world conditions:

- **Pilot Testing:** The system will be deployed in a controlled environment to process batches of areca nuts. The results will be compared with manual inspection to validate the system's accuracy and efficiency.
- **User Feedback:** Feedback from farmers and other stakeholders will be collected to assess the system's usability, effectiveness, and any potential areas for improvement.
- System Optimization: Based on the testing and feedback, any necessary adjustments will be made to the system to ensure optimal performance.

6. **Deployment and Integration**

Once validated, the system will be deployed for regular use in areca nut processing facilities:

- **Hardware and Software Integration:** The system will be integrated with existing processing lines, ensuring seamless operation with minimal disruption to current workflows.
- **Training and Support**: Training will be provided to operators and farmers to ensure they can effectively use the system. Ongoing technical support will be available to address any issues that arise during operation.

Steps involved in project

Initialize GUI

- Configure GUI properties and callbacks.

Load Image

- On button click, open file dialog, read image, and display it.

Enable Processing

- On button click, activate image processing options.

Process Image

- Convert to Grayscale: Apply thresholding, show binary image.
- Segment Image: Mask and display segmented image.
- Resize Image: Resize and display image.
- Filter Image: Apply Gaussian filter and display.
- Extract Features: Compute and display color, texture, and shape features.

Classify Image

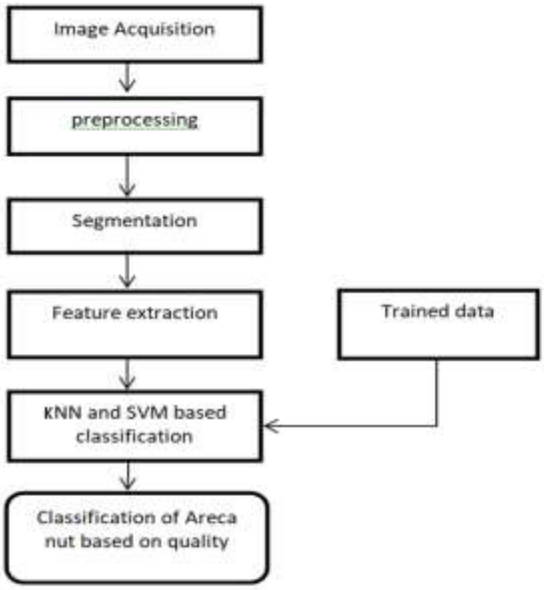
- **Load Models:** Load SVM and KNN models.
- **Predict Labels:** Use models to predict and display results.

Show Results

- Display accuracy, confusion matrices, and classification results.

Block Diagram:

1. **Start**



- **GUI Initialization:** This is where the GUI is set up, including creating the interface elements such as buttons, panels, and axes. It initializes the application and prepares it for user interaction.

2. **Image Handling**

- **Load Image:**
  - Button Click: The user clicks the "Load Image" button.
  - Open File Dialog: A file dialog opens allowing the user to select an image file.
  - Read & Display Image: The selected image is read into MATLAB and displayed in the GUI.
- **Image Processing Options:**
  - Button Click: The user clicks buttons for various image processing options such as grayscale conversion, resizing, filtering, etc.
  - Enable Processing Options: The selected processing options become available for the user to apply to the image.

3. **Processing Operations**

- **Convert to Grayscale:**
  - Apply Thresholding: The image is converted to grayscale, and a threshold is applied if needed.
  - Display Binary Image: The resulting binary image (if thresholding is applied) is displayed in the GUI.
- **Segment Image:**
  - Apply Mask: Segmentation is performed by applying a mask to isolate specific parts of the image.

  - Display Segmented Image: The segmented image is shown in the GUI.
- **Resize Image:**
  - Apply Resize: The image is resized according to user specifications.
  - Display Resized Image: The resized image is displayed in the GUI.
- **Filter Image:**
  - Apply Gaussian Filter: A Gaussian filter is applied to the image to smooth it.
  - Display Filtered Image: The filtered image is shown in the GUI.
- **Extract Features:**
  - **Color Features**: Extract color-related features from the image.
  - **Texture Features:** Extract texture-related features from the image.
  - **Shape Features:** Extract shape-related features from the image.
  - **Display Extracted Features:** The extracted features are displayed or used for further analysis.

4. Classification

- **Load Models:**
  - **SVM Model:** Load a Support Vector Machine (SVM) classification model.
  - **KNN Model**: Load a k-Nearest Neighbors (KNN) classification model.
- **Predict Labels:**
  - **Apply Models:** Use the loaded models to classify the processed image.
  - **Display Predictions:** Show the predicted labels or classifications in the GUI.

5. Results

- **Show Accuracy:**
  - **Display Classification Accuracy:** Show the accuracy of the classification models.
- **Confusion Matrix:**
  - **Display Confusion Matrices:** Show confusion matrices for the models to visualize classification performance.
- **Display Results:**
  - **Show Final Classification Results:** Display the final results, including predicted labels and any additional insights.

6. **End**

- **End:** This marks the completion of the process and may involve cleaning up resources or closing the application.

a) Hardware Requirements:

- Processor :intel core i3
- RAM:4GB
- HDD:500GB
- OS:Windows 7 or above

b) Software Requirements:

Matlab R2018a

Matlab R2018a is a particular variant of the Matlab programming, which is an undeniable level programming language and improvement climate regularly utilized for mathematical calculation, information investigation, and representation. R2018a demonstrates that it was delivered in the main portion of 2018.

Here are a few critical elements and upgrades presented in Matlab R2018a:

1. **Live Manager:** Upgraded intelligent report climate with live scripts, which consolidate code, yield, and organized text in a solitary executable document.

2. **Equal Processing:** Further developed execution and adaptability for equal registering with upgrades to parfor circles and other equal capabilities.

3. **Profound Learning:** Extended help for profound learning models, remembering upgrades for preparing and conveying brain organizations.

4. **Information Import and Commodity:** Improved information import and product choices for different organizations, like JSON, Avro, Parquet, and HDF5.

5. **Designs and Perception:** New capabilities and enhancements to existing ones for making intuitive representations and altering plots.

6. **Application Originator:** Enhancements to the Application Architect instrument for making and planning custom UIs.

VI. RESULT AND DISCUSSION

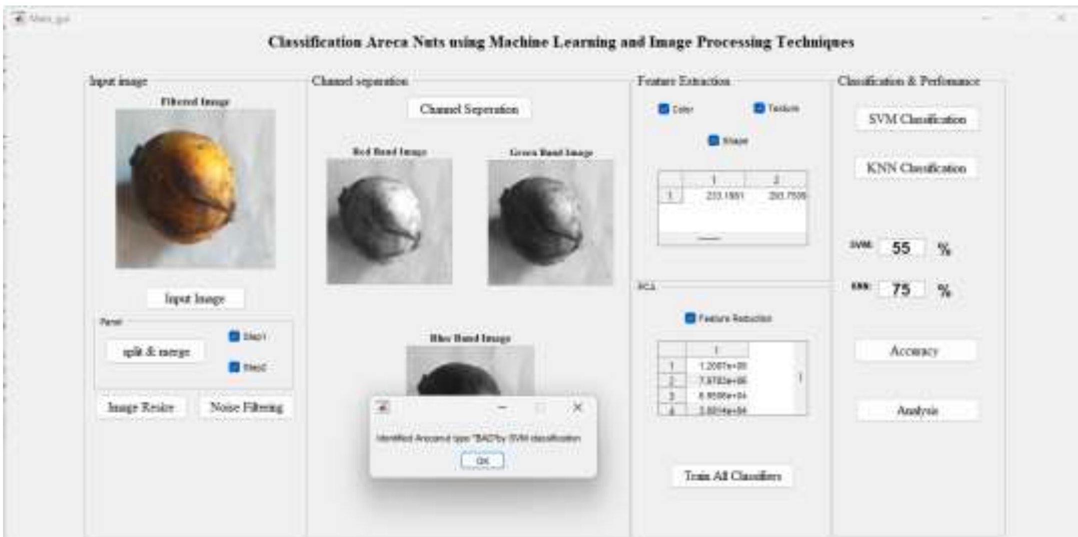


Fig a, Arecnut with infected area

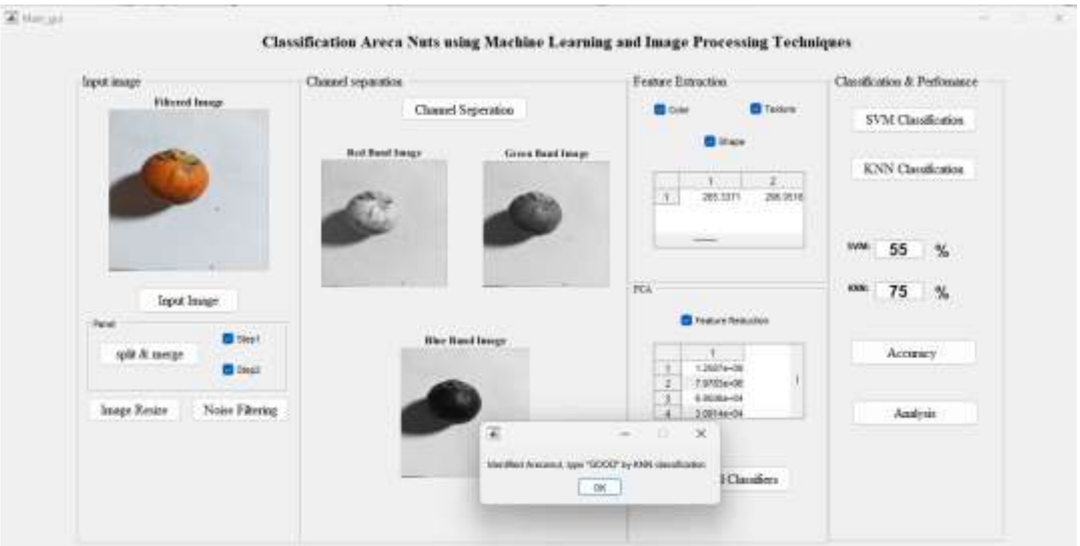


Fig b, Arecnut with out any infected area

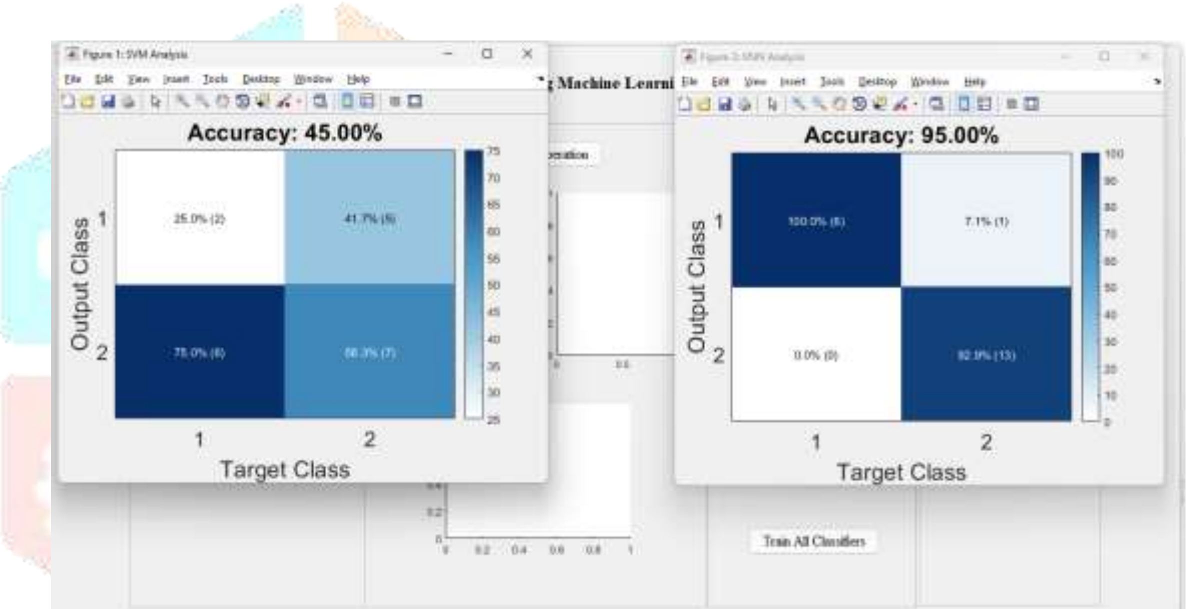


Fig c, Confusion matrix for SVM and KNN classifications

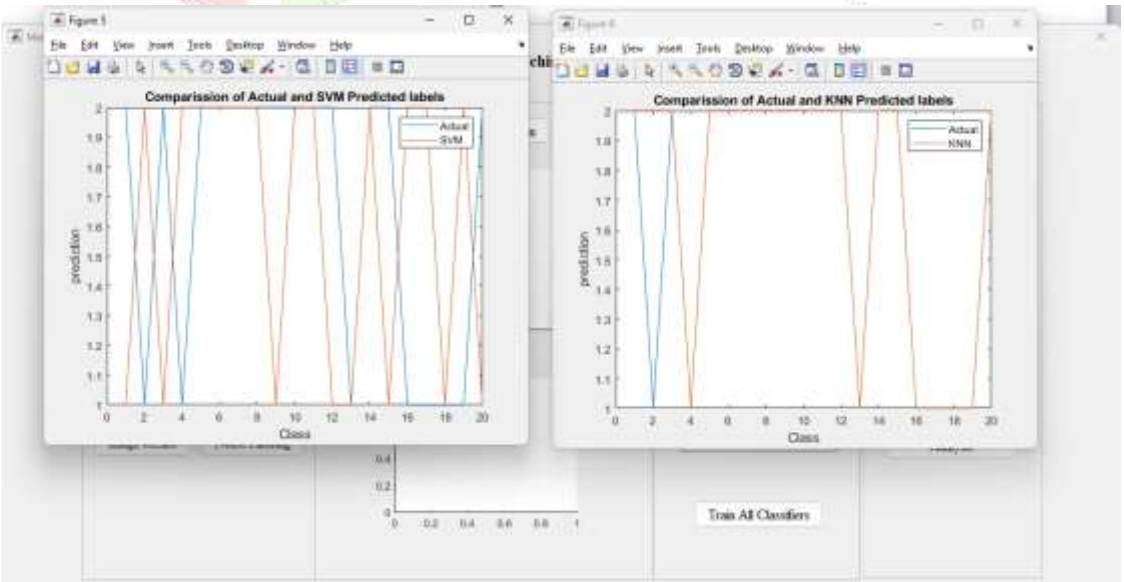


Fig d, Graph comparisons with SVM and KNN for test data's

VII. CONCLUSION

This study focused on addressing the critical need for an automated system to classify and grade areca nuts based on their quality, a task traditionally performed manually by farmers. The manual inspection process, while common, is fraught with challenges such as labor intensity, time consumption, and susceptibility to human error, especially when dealing with large quantities of nuts. These issues necessitate a more efficient and reliable solution to ensure consistent quality control in the areca nut industry.To this end, the study applied K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms to classify areca nuts using a dataset of 208 images. The results indicated that the KNN algorithm significantly outperformed SVM, achieving an accuracy of 95%, compared to 65% with SVM. The high accuracy of KNN demonstrates its potential as a viable tool for automating the quality assessment of areca nuts, offering a more objective and consistent alternative to manual inspection.The implications of these findings are substantial for the agricultural sector, particularly in promoting smart farming practices. By integrating machine learning algorithms like KNN into the quality assessment process, farmers can enhance their ability to make informed decisions, take timely preventive actions, and ultimately improve the market value of their produce. This shift from manual to automated quality assessment aligns with the broader trend toward the digitization of agriculture, where data-driven technologies are used to optimize farming operations and increase productivity.Moreover, the success of KNN in this application highlights the potential for its use in other agricultural contexts, where similar challenges exist in quality assessment. Future research could expand on this work by exploring larger and more diverse datasets, testing additional machine learning algorithms, and developing user-friendly applications for practical deployment in farming communities.In conclusion, the study provides a promising solution to the problem of areca nut quality assessment, demonstrating that machine learning, particularly KNN, can play a crucial role in modernizing agricultural practices. By offering a more efficient, accurate, and scalable method for quality control, this research contributes to the ongoing efforts to improve the sustainability and profitability of farming, ensuring that farmers can meet the demands of the market while maintaining high standards of crop quality.

REFERENCES

1. Kamilaris, A., & Prenafeta-Boldú, F. X. (2020). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 1-30.
2. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2021). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 10, 941.
3. Zhang, X., Li, S., Wang, B., & Guo, W. (2020). A review of machine learning algorithms for fruit quality evaluation. Sensors, 20(21), 6761.
4. Lin, C., Wang, C., & Zhang, Y. (2021). Tea leaf quality assessment based on convolutional neural networks. IEEE Access, 9, 78659-78668.
5. Liu, J., Chen, Y., & Xu, X. (2022). Areca nut disease detection using machine learning and image processing. Journal of Agricultural Science and Technology, 24(2), 321-333.
6. Wang, H., Zhang, Y., & Li, T. (2021). Real-time areca nut quality detection using transfer learning. Computers and Electronics in Agriculture, 182, 106029.

7. Chen, J., Li, X., & Zhang, Q. (2020). Application of KNN in crop disease classification. Journal of Agricultural Engineering Research, 46(3), 345-353.
8. Sun, H., Zheng, M., & Lin, Z. (2021). Fruit quality classification using SVM and image processing. International Journal of Food Science & Technology, 56(4), 2100-2112.
9. Minervini, M., Scharr, H., & Tsaftaris, S. A. (2020). Image analysis: The new bottleneck in plant phenotyping. IEEE Signal Processing Magazine, 38(1), 87-97.
10. Nguyen, T. T., Chawla, N. V., & Knowles, J. (2021). Machine learning for crop yield prediction and climate change impact assessment in agriculture. Environmental Modelling & Software, 150, 105295.