IDENTIFICATION OF LARGE CARDAMOM DISEASES USING MACHINE LEARNING AND DEEP LEARNING

A Project Report submitted in

Fulfilment of requirements for the degree of

Bachelor of Technology

In

Computer Engineering

By

Roshan Chettri (21CSEC10)

Pritam Kumar Bhagat (21CSEC14)

Under the supervision of

Ms. Dibya Thapa

Assistant Professor, Computer Engineering Department

DEPARTMENT OF COMPUTER ENGINEERING

SIKKIM INSTITUTE OF SCIENCE & TECHNOLOGY

(Affiliated to Sikkim University)

2024

DECLARATION

We, the undersigned, hereby declare that the work recorded in this project report entitled **"Identification of Large Cardamom Diseases Using Machine Learning And Deep Learning"** in partial fulfilment for the requirements of award of B. Tech (Computer Engineering) from Sikkim Institute of Science and Technology is a faithful and Bonafide project work carried out at "SIST, Chisopani" under the supervision and guidance of Ms. Dibya Thapa, Assistant Professor, Department of Computer Engineering, Sikkim Institute of Science and Technology.

The results of this investigation reported in this project have so far not been reported for any other Degree / Diploma or any other technical forum.

The assistance and help received during the course of the investigation have been duly acknowledged.

Roshan Chettri(21CSEC10): _________________________

Pritam Kumar Bhagat(21CSEC14): ___________________

CERTIFICATE BY SUPERVISOR

This is to certify that the work recorded in this project report entitled **"Identification of Large Cardamom Diseases Using Machine Learning and Deep Learning"** has been jointly carried out by Mr. Roshan Chettri (21CSEC10) and Mr. Pritam Kumar Bhagat(21CSEC14), of Computer Engineering Department of Sikkim Institute of Science and Technology in partial fulfilment of the requirements for the award of Bachelor of Technology in Computer Engineering. This report has been duly reviewed by the undersigned and recommended for final submission for Mini project presentation.

Project Guide	Head of the Department
Department of Computer Engineering	Department of Computer Engineering
Sikkim Institute of Science and Technology	Sikkim Institute of Science and Technology

Project Coordinator

Department of Computer Engineering

Sikkim Institute of Science and Technology

ACKNOWLEDGEMENT

Firstly, we would like to thank Prof. Dr. B.B Pradhan, Principal cum Director of Sikkim Institute of Science and Technology and the Computer Engineering Department for providing us such a wonderful opportunity to do Mini project on the **“Identification of Large Cardamom Diseases Using Machine Learning and Deep Learning”.**

We would like to express our special thanks of gratitude to our project guide Ms. Dibya Thapa (Assistant Professor, Computer Engineering Department) for her able guidance and support in the completion of our project successfully.

We would also like to extend our gratitude to the project coordinator Ms. Pragya Pradhan (Assistant Professor, Computer Engineering Department) and Head of Department, Mr. Prashant Pradhan (Assistant Professor, Computer Engineering Department) for their help, support and guidance.

Lastly, we would also like to thank our parents, friends and teammates for their contribution in the completion of this project successfully.

Mr. Roshan Chettri (21CSEC10)

Mr. Pritam Kumar Bhagat(21CSEC14)

Abstract

Large Cardamom is a high-value spice crop that is vulnerable *to Cardamom Bushy Dwarf Virus* and *Large Cardamom Chirkey Virus*. If these diseases are not treated promptly, they can result in big losses. Manual inspections and expert knowledge are traditionally used to identify diseases, but these methods are slow and inexact. The purpose of this project is to develop an automated system for classifying *Cardamom Bushy Dwarf Virus* and *Large Cardamom Chirkey Virus* diseases using Machine Learning and Deep Learning algorithms. As part of the pipeline, we will acquire images, preprocess them, extract features, and classify them. In this project, Convolutional Neural Networks (CNNs) will be used to perform image recognition tasks and compared to traditional machine learning algorithms, such as Support Vector Machines (SVMs) and Random Forests. The Models like CNN, ANN, and SVC performed exceptionally well, with CNN achieving the most balanced and reliable metrics across all evaluation parameters with Color Hog as feature extraction technique.

Keywords:

Cardamom Bushy Dwarf Virus, Large Cardamom Chirkey Virus, Disease, Machine learning, Deep Learning, Convolutional Neural Network.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction

The Large cardamom cultivation is the source of income and livelihood of almost 88% population of Sikkim, and income from this particular crop is significantly higher than other livelihood options [1].

The occurrence of *Large Cardamom Chirkey Virus* (LCCV) and *Foorkey* has a negative impact on production of Large Cardamom in Sikkim [1]. If these diseases are not discovered in time impact on the income of farmers and overall GDP of Sikkim is negative. Early detection is an effective prevention and control of plant diseases, and they play a vital role in the management and decision making of agricultural production [3].

Chirkey and Furkey disease shows pale streaks on the leaves [2] and sprouting and formation of bushy dwarf clumps at the base of the plants [2]. Traditionally, agricultural and forestry experts identify disease on-site based on characteristics of infected plant, individuals experience and visual inspection. This method is not only time consuming but also unreliable, laborious, and prone to human made errors making it difficult to achieve an accurate and consistent results which leads to use of unreliable methods to treat the plant, significantly impacting quality and quantity of the yield.

The goal of this project is to develop an AI driven mechanism that will automate the detection of these infected large cardamom plants. By training Machine learning and Deep Learning models on sufficient data to create a model that can increase accuracy in classifying Chirkey and Furkey infected plants. Which will lead to increase in the yield of cardamom and ultimately contributing to Sikkim's GDP.

To achieve this, we are focusing on collecting high quality data that contains the relevant features of these infected Large Cardamom plants. Preprocessing data will be a crucial step to ensure the model trains on accurate data.

The performance will be evaluated using accuracy, precision, recall and F1 score to ensure models reliability.

The National Informatics Centre and Spices Board of Sikkim has considered to work on creating an AI driven classification of Large Cardamom plant diseases [4].

Fig 1. Furkey effected Large cardamom [8].

Fig 2. Chirkey effected large cardamom plant [9].

1.1. Literature Survey

Table 1: Literature Survey 1

1.	Name of the paper	Curvularia eragrostidis, a new threat to large cardamom (Amomum subulatum Roxb.) causing leaf blight in Sikkim
	Author	Kabita Gurung, Khashtu Dasila, Anita Pandey and Niladri Bag
	Publication Details	J Biosci (2020) 45:113 Indian Academy of Sciences DOI: 10.1007/s12038-020-00086-7
	Research Findings	• The large cardamom in Sikkim faces leaf blight caused by Large Cardamom Chirkey virus. • This disease causes severe foliar damage impacting crop cultivation significantly
	Research Relevance	The paper shows us how the disease can be identified based on its features such as leaf blight.
	Research Gap	Lack of data on disease distribution and severity.

Table 2: Literature Survey 2

2.	Name of the paper	Status of Viral Diseases of Large Cardamom (Amomum subulatum Roxb.) and its Management in Sikkim and Darjeeling, West Bengal
	Author	AK Vijayan, BA Gudade, TN Deka and P Chhetri
	Publication Details	J Mycol Plant Pathol, Vol. 44, No.4, 2014
	Research Findings	• Describes Chirkey and Foorkey viral diseases affecting large cardamom cultivators. • Describes the symptoms, transmission and crop loss.
	Research Relevance	• Chirkey and Foorkey disease impact on the Yield. • Very little work has been done regarding management of viral diseases.
	Research Gap	Needs comprehensive study to explore various disease management approach.

Table 3: Literature Survey 3

3.	Name of the paper	Plant Disease Detection and Classification by Deep Learning—A Review
	Author	Lili Li, Shujuan Zhang, and Bin Wang
	Publication Details	08 April 2021 2169-3536 - 10.1109/ACCESS.2021.3069646 IEEE
	Research Findings	• Deep Learning has become an important tool in plant disease detection. • There is a lack of attention to recent development in visualization and early detection techniques. • Paper discusses how the ML and DL algorithms can help inexperienced farmers to identify diseases in crops.
	Research Relevance	• Deep learning improves plant disease recognition. • Transfer learning can address the curse of limited data. • The DL model shows the issue of robustness across diverse data.
	Research Gap	• Lacks robust models which can classify diverse diseases. • Needs real condition plant disease dataset.

Table 4: Literature Survey 4

4.	Name of the paper	Plant Disease Detection Using Image Processing
	Author	Sachin D. Khirade, and A. B. Patil.
	Publication Details	2015 International Conference on Computing Communication Control and Automation
	Research Findings	• Used various methods for plant disease detection using Image processing.
	Research Relevance	• RGB images are converted to grayscale images by using filters, and following with histogram equalization enhancing the visibility of affected areas of plants.
	Research Gap	• Lacks comparison with other ML/DL algorithms. • Lacks the comparative feature extraction techniques.

Table 5: Literature Survey 5

5.	Name of the paper	Plant Disease Detection Using Machine Learning
	Author	Shima Ramesh, Mr. Ramachandra Hebbar, P V Vinod
	Publication Details	2018 International Conference on Design Innovations for 3Cs Compute Communicate Control
	Research Findings	• Research highlights the effectiveness of machine learning classifiers. • Random forest classifier improves the accuracy significantly when trained on less data.
	Research Relevance	The Histogram of Oriented Gradients captures the intensity gradients of image.
	Research Gap	• Lacks comparison with other advance ML and Deep Learning architecture. • Uses less data for training the model which may lead to overfitting.

Table 6: Literature Survey 6

6.	Name of the paper	Tomato Plant Disease Identification via Deep Learning Technique.
	Author	Shrinivas Bhosale, Aditi Chhabria.
	Publication Details	29 Jun 2024 -International Journal of Image and Graphics
	Research Findings	Paper proposes a method for tomato disease identification using Deep learning techniques. Introduced a novel method for adjusting weights (WUDHOA).
	Research Relevance	Other Conventional ML techniques have lower specificity rate.
	Research Gap	The model increases the need of computational power by adding more computational parameters.

Table 7: Literature Survey 7

7.	Name of the paper	An efficient deep learning model for tomato disease detection
	Author	Xuewei Wang, Jun Liu
	Publication Details	09 May 2024 Plant Methods
	Research insights	• This paper introduced the Tomatonet model for precise Tomato disease detection.
	Research Findings	• Custom CNN design for specific problem
	Research Gap	• High False positive and False Negative rate in real world settings. • Author did not explore other Architecture/methods that could have reduced False positive and Negatives rate like ResNet and VGGNet.

Table 8: Literature Survey 8

8.	Name of the paper	Feature extraction using convolution neural networks (CNN) and deep learning
	Author	Jogin M, Mohana, Madhulika M, Divya G, Meghana R, Apoorva S
	Publication Details	2018 Institute of Electrical and Electronics Engineers Inc.
	Research insights	• Author of this paper has trained many models like (SVM, K- Nearest Neighbour etc..) to classify images containing 9 classes the highest accuracy received was from CNN (85.97%) followed by ANN (46.4%).
	Research Findings	• The results indicate Convo layer can be used for feature extraction.
	Research Gap	• The paper only talks about Sequential CNN model i.e. connecting to Fully connected layer • The paper fails to use Power of Convolution for feature extraction and then feeding it to other algorithms which may increase the accuracy

Table 9: Literature Survey 9

9.	Name of the paper	A survey of feature selection and feature extraction techniques in machine learning
	Author	Nasreen, Shamila, Khalil, Tehmina, Khalid, Samina
	Publication Details	2014 Institute of Electrical and Electronics Engineers Inc.
	Research insights	• The Author of this paper has analysed feature extraction and feature selection techniques
	Research Findings	• The paper summarizes feature extraction techniques.
	Research Gap	• The paper fails to compare feature extraction techniques with metrices like accuracy, R2, ROC.

1.2. Summary of Research Gaps based on the Review

Based on research articles reviewed on Classification of plant diseases using various Machine Learning and Deep Learning algorithms, the following points should be taken into account while designing suitable model.

1. Needs data on proper disease distribution and severity.

2. Needs comprehensive study to explore various disease management approach.

3. Need of proper robust Machine learning or Deep Learning models which can classify and predict the diseases.

4. Lack of comparative feature extraction techniques.

5. Need of exploration of other ML/DL algorithms that can possibly automate the disease detection in cardamom.

1.3. Problem Definition

i. The *Cardamom Bushy Dwarf Virus* (CBDV) and *Large Cardamom Chirkey Virus* (LCCV) combined has resulted in decrease in yield of Large Cardamom by 60% to 85% of total production in Sikkim which is comparatively higher than any other factors.

ii. These diseases have led to the significantly reduced production capacity and decrease in personal incomes of farmers from Sikkim. Besides large cardamom the LLCV (Chirkey virus) infects wheat, arrow root, ginger and the perennial weed.

iii. This makes the early identification of these diseases crucial. Therefore, there is a need for AI-driven solutions that can provide accurate and consistent detection of these diseases to lessen their impact and support the sustainability of agriculture in Sikkim.

1.4 Objective

Based on literature review and the mentioned issue, the research initiative is motivated towards the achievement of following objectives:

i. Collection of sufficient cardamom images for its disease detection.
ii. Comparison of different image Augmentation techniques using metrices like accuracy and to create more diverse data to prevent overfitting.
iii. Data Preprocessing normalizing and resizing.
iv. Comparison of different feature extraction techniques using metrics like accuracy.
v. Building and comparing different Machine Learning and Deep Learning classifier which can efficiently classify diseased or healthy plant.
vi. Testing of developed models on unseen images.

2.1. Methodology

To address the challenges possessed by *Cardamom Bushy Dwarf Virus* and *Large Cardamom Chirkey Virus (LCCV)* a multi-pronged strategy is required. The goal is to develop Machine Learning (ML) and Deep Learning (DL) models for the early and accurate detection of these diseases in Large Cardamom plants.

The solution strategy will be implemented in the following stages:

1) **Data Collection and Preparation:**
   a) **Data collection:** The first step involves gathering dataset that includes images of Large Cardamom plants, both healthy and diseased. The datasets will be collected physically from multiple sources, including local farms and agricultural research centres.
      b) **Data Labelling:** The collected data will be labelled to distinguish between healthy and those infected with Chirkey and Furkey diseases.
2) **Data preprocessing:**
   a) **Normalization:** Normalize the pixel values of image to ensure that the input has consistent scales.
      i) **Min - Max Scaling:** Every pixel ranges between [0 – 255] providing in consistent scale. Therefore, every pixel will be divided by the maximum value (i.e. 255) making the values ranging between [0 - 1] ensuring consistent scale.
   b) **Image Resizing:** The aspect ratio of images may not be consistent. Therefore, we will standardize the size and aspect ratio of images to 256, 256.
3) **Data Augmentation**
   It is used to artificially increase the size of dataset by applying various transformation techniques to the existing data. Data Augmentation technique in case of images are

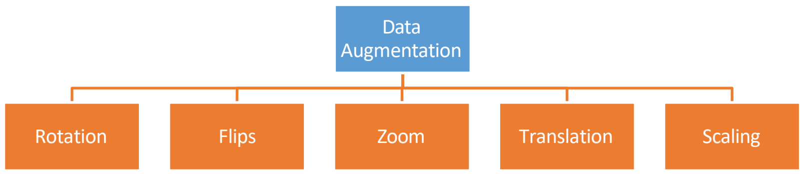


Figure 3: Data Augmentation Technique

   a) **Geometric Transformation:** Flipping 90° on X and Y axis, rotating images by a random degree, changing image size while keeping original dimensions
   b) **Color Transformation:** Increasing or decreasing the intensity of light in the image, Enhancing or reducing the difference between light and dark areas, Increasing or decreasing the intensity of colors. Altering the overall color tone of the image.
   c) **Noise Injection:** Adding random noise to image.

In this experiment Geometric Transformations is used to create more variation in Large Cardamom data set to improve the accuracy score of the machine learning and deep learning models without changing the important aspect of images like color or shapes.

4) **Normalization**

It is the process of adjusting the pixel values in an image to specific range. Normalizing the image makes it easier for algorithms to learn patterns and improve performance [13]. We employed Min-Max normalization (Equation 1) to normalize the cardamom images i.e,

$$x' = x/255$$

5) **Feature Extraction**

Involves transforming raw data into a set of attributes that are most relevant to distinguish between different categories. The feature extraction will be compared using metrics such as accuracy, ROC. The one which performs well be selected.

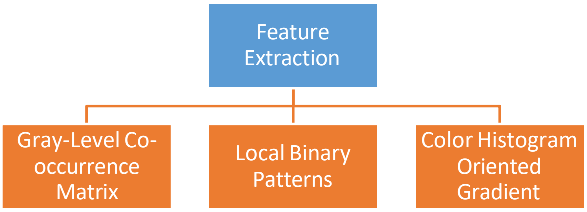


Figure 4: Feature Extraction Techniques

a) **Color HOG:** It focuses on capturing distribution of edge directions or gradient orientations in localized portion of image. The distribution of edge is different in Diseased and Healthy plant leaves.
b) **Gray-Level Co-occurrence Matrix (GLCM):** It captures spatial relationship between pixels in grayscale. It describes co-occurrence of physical intensity values at certain distance and angles.
c) **Local Binary Patterns:** It captures micropatterns in an image by comparing pixels intensities. LBT looks at neighbouring pixels within a specific radius and thresholds.

6) **Dataset split:**

a) **Training Dataset:** The training dataset is a collection of labelled data used to teach a Machine learning model how to make prediction and perform a specific task. Each example in training dataset consists of input features (attributes or characteristics) and their corresponding labels. The model learns from these data to learn pattern and relationships between the features and labels, enabling it to make accurate prediction on new unseen data.
b) **Testing and validation Dataset:** This dataset is portion of the overall dataset that is reserved specifically for evaluating performance of model. Consisting of collection of examples, each with input features and their corresponding target values or labels. This is used to assess how well model performs on new unseen-data.

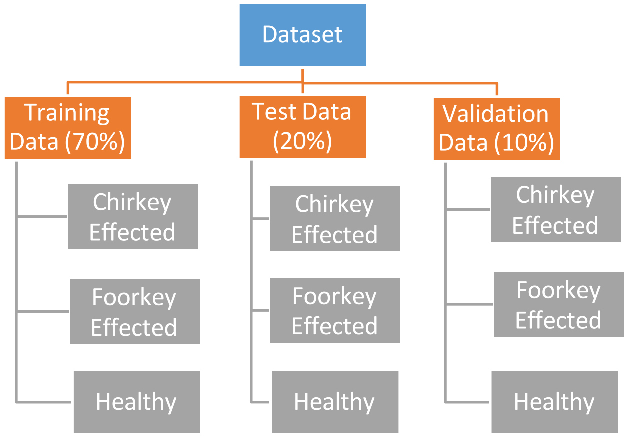


Figure 5: Data Split

7) **Model Development:**
   a) **Model Selection:** Explore and select the suitable ML and DL algorithm, such as SVM, KNN, Convolutional Neural Network (CNNs), for image classification tasks.
   b) **Model Training**: Train the selected models using the preprocessed and augmented dataset, allowing them to learn and recognize patterns that distinguish between healthy and diseased plants.
   c) **Model Optimization**: Optimize the models through hyperparameter tuning, regularization techniques to attain high accuracy.
8) **Validation and Testing:**
   a) **Model Evaluation:** Once the classification model is selected and trained, we will use Validation and Testing data to assess the performance using metrics such as accuracy, precision, recall and F1-score.
9) **Result:**
   a) After training and testing the results will be generated whether the plant is Chirkey, furkey infected, or the plant is healthy.

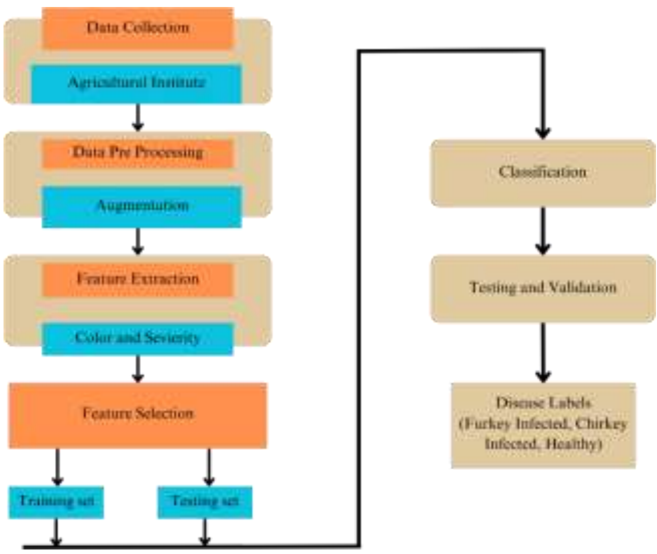


Fig 6: Block Diagram of Proposed Model.

1) Gantt Chart

Activities/Duration	August				September				October				November			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Problem Identification																
Feasibility Study																
Literature Survey																
Requirement Gathering and Specification																
Model Development and evaluation																
Testing and analyzing																
Documentation																

PROPOSED ACTIVITY	ACHIEVED ACTIVITY

Figure 7: Gantt Chart

Chapter 4. Results

The research is implemented using Python 3.11 programming language and executed using 11th Gen Intel(R) Core (TM) i5-1135G7 with Intel IRIS Xe Graphics card. The project evaluated multiple Machine Learning (ML) and Deep Learning (DL) models for the classification of Large Cardamom plant diseases. Performance metrics such as accuracy, precision, recall, and the confusion matrix were used to assess the effectiveness of these models across different feature extraction techniques.

Dataset Overview and Preprocessing

- **Dataset Size**: The dataset contained 2,222 images of healthy and diseased Large Cardamom plants. These were split into:
    - 80% for training (1,777 images),
    - 10% for validation (222 images), and
    - 10% for testing (222 images).
- **Image Preprocessing**:
    - Resized to 255×255 pixels.
    - Normalized pixel values to range [0, 1].
    - Augmentation included geometric transformations such as rotation, flipping, rescaling, and shifting.

Feature Extraction Techniques

Three feature extraction techniques were applied before training the models:

1. **Color HOG (Color Histogram of Oriented Gradients)**: Best performance across all models.
2. **Local Binary Pattern (LBP)**: Moderate performance.
3. **Gray Level Co-occurrence Matrix (GLCM)**: Lowest performance.

Table 10: Accuracy (in Percent)

Model	Gray level co-occurrence matrix	local binary pattern	Color HOG
Ada Boost	0.578261	0.582609	0.93913
Bagging	0.578261	0.66087	0.947826
Extra Trees	0.578261	0.673913	0.943478
Gaussian NB	0.578261	0.491304	0.947826
Gradient Boosting	0.578261	0.66087	0.947826
Hist GB	0.578261	0.634783	0.956522
Logistic	0.421739	0.591304	0.917391
Random Forest	0.578261	0.66087	0.952174
SVC	0.578261	0.591304	0.956522
XGB	0.578261	0.656522	0.956522
ANN	0.46511	0.76761	0.96047

- Color HOG significantly outperformed GLCM and LBP for all models.
- ANN, SVC, and XGBoost achieved the highest accuracy with Color HOG features.
- GLCM produced the lowest accuracy, indicating limited suitability for this dataset.

Table 11: Precision

Model	Gray level co-occurrence matrix	local binary pattern	Color HOG
Ada Boost	0.578261	0.605714	0.910345
Bagging	0.578261	0.682119	0.929078
Extra Trees	0.578261	0.664773	0.916667
Gaussian NB	0.578261	0.547059	0.929078
Gradient Boosting	0.578261	0.651934	0.929078
Hist GB	0.578261	0.646707	0.93617
Logistic	0	0.603175	0.890411
Random Forest	0.578261	0.65896	0.929577
SVC	0.578261	0.590698	0.942446
XGB	0.578261	0.658824	0.93617
ANN	0	0.76761	0.90149

Table 11 indicates the precision of models where higher precision indicates fewer False Positives (FP). Where SVC achieved the highest precision (94.24%), followed by XGB (93.61%) as compared to other models reducing misclassifications of healthy plants as diseased.

Table 12: Recall

Model	Gray level co-occurrence matrix	local binary pattern	Color HOG
Ada Boost	0.578261	0.605714	0.910345
Bagging	0.578261	0.682119	0.929078
Extra Trees	0.578261	0.664773	0.916667
Gaussian NB	0.578261	0.547059	0.929078
Gradient Boosting	0.578261	0.651934	0.929078
Hist GB	0.578261	0.646707	0.93617
Logistic	0	0.603175	0.890411
Random Forest	0.578261	0.65896	0.929577
SVC	0.578261	0.590698	0.942446
XGB	0.578261	0.658824	0.93617
ANN	0	0.76761	0.97906

Table 12 indicates recall where higher recall reflects lower False Negatives (FN). ANN achieved the highest recall (97.9%), followed by SVC (94.24%) on Color Hog feature effectively identifying diseased plants.

Table 13 and 14 compares the output before and after data augmentation

Table 13: Comparison of Accuracy, Precision and Recall of color HOG.

Before Augmentation			
Model	**Accuracy**	**Precision**	**Recall**
Ada Boost	0.93913	0.91034	0.91034
Bagging	0.94782	0.92907	0.92907
Extra Trees	0.94347	0.91666	0.91666
Gaussian NB	0.94782	0.92907	0.92907
Gradient Boosting	0.94782	0.92907	0.92907
Hist GB	0.95652	0.93617	0.93617
Logistic	0.91739	0.89041	0.89041
Random Fores	0.95217	0.92957	0.92957
SVC	0.95652	0.94244	0.94244
XGB	0.95652	0.93617	0.93617
ANN	0.96047	0.90149	0.97906
CNN	0.97959	0.97959	0.97959

Table 14: Comparison of Accuracy, Precision and Recall of color HOG.

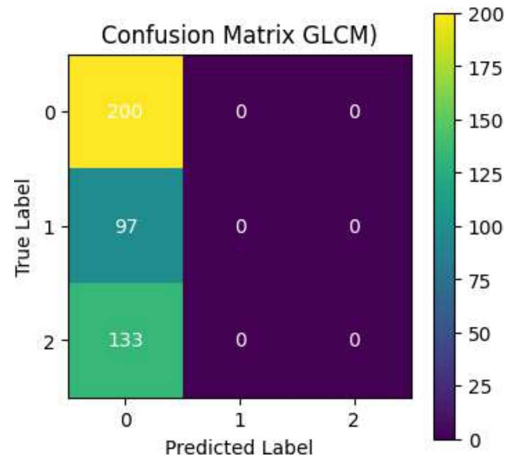
After Augmentation			
Model	**Accuracy**	**Precision**	**Recall**
Ada Boost	0.91861	0.91925	0.918605
Bagging	0.91395	0.91470	0.913953
Extra Trees	0.95116	0.95434	0.951163
Gaussian NB	0.91395	0.91411	0.913953
Gradient Boosting	0.94884	0.95112	0.948837
Hist GB	0.95349	0.95694	0.953488
Logistic	0.75581	0.81942	0.755814
Random Fores	0.93488	0.93906	0.934884
SVC	0.96744	0.96759	0.967442
XGB	0.94651	0.94917	0.946512
ANN	0.95581	0.66772	0.98139
CNN	0.9766	0.9766	0.9745

Where Augmentation improved metrics across all models.

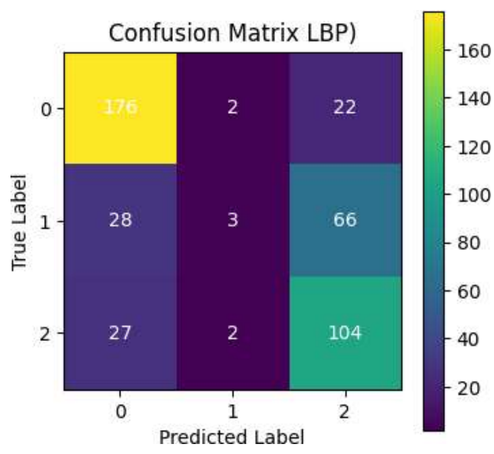
CNN achieved the highest overall performance across accuracy, precision, and recall metrics. Its hierarchical feature extraction capability enables it to identify intricate patterns in diseased leaves effectively.

ANN excelled in recall (97.9%), demonstrating its ability to identify diseased samples, reducing false negatives significantly.

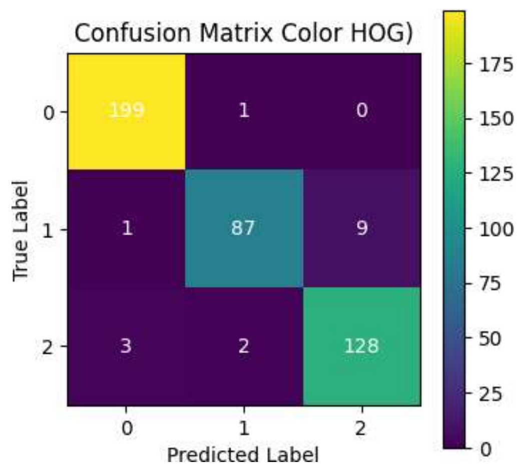
SVC balanced precision and recall well (94.2% for both), indicating its suitability for handling high-dimensional data generated from Color HOG.



a)



b)

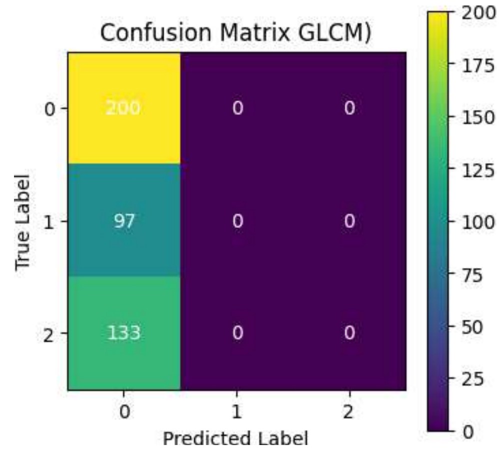


c)

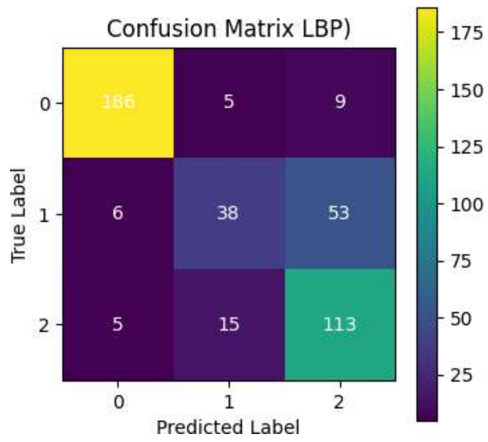
Figure 8: Confusion matrix of SVC across Feature extraction techniques

a) GLCM, b) LBP, c) Color HOG

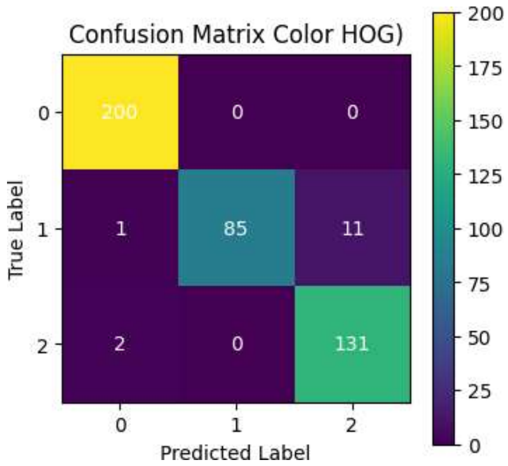
Figure 8 and Figure 9 shows that GLCM has high FP and FN rates which misclassifies a significant number of healthy plants as diseased leading to inflated FP and lower Recall and Precision while LBP has moderate FP and FN and Color Hog performing best across all metrics.



a)



b)



c)

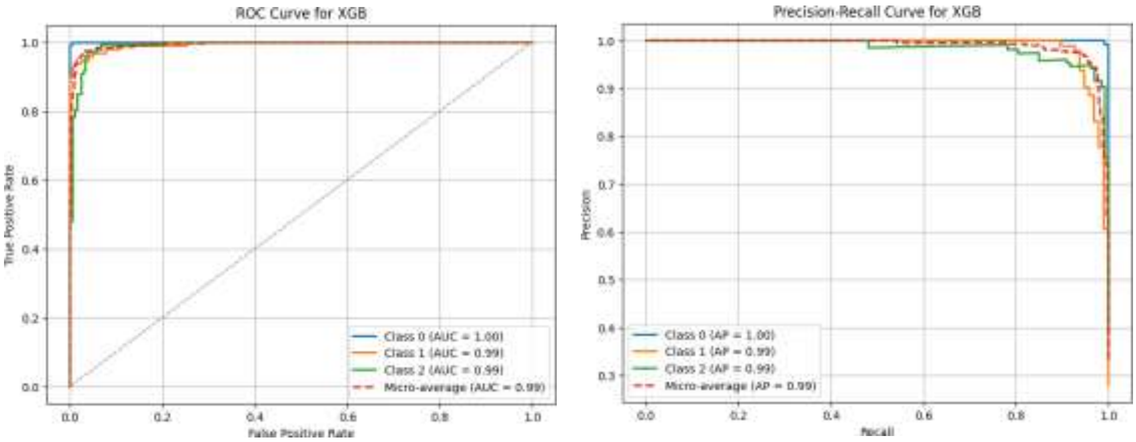
Figure 9: Confusion Matrix of XGBoost

a) GLCM, b) LBP, c) Color HOG

- **GLCM**:
  a) Poor performance due to its inability to capture fine textures and patterns.
  b) Suitable for simpler classification tasks but not for complex plant disease identification.
- **LBP**:
  a) Moderate improvement over GLCM.
  b) Captures texture-based features better but lacks the robustness of gradient-based methods.
- **Color HOG**:
  a) The most effective technique, consistently yielding high precision and recall across models.
  b) Its ability to capture edge and gradient information makes it ideal for detailed disease detection.

By using Color HOG, the models demonstrated reliable and accurate classification, significantly outperforming GLCM and LBP in all metrics. For practical applications, Color HOG is the recommended feature extraction method.

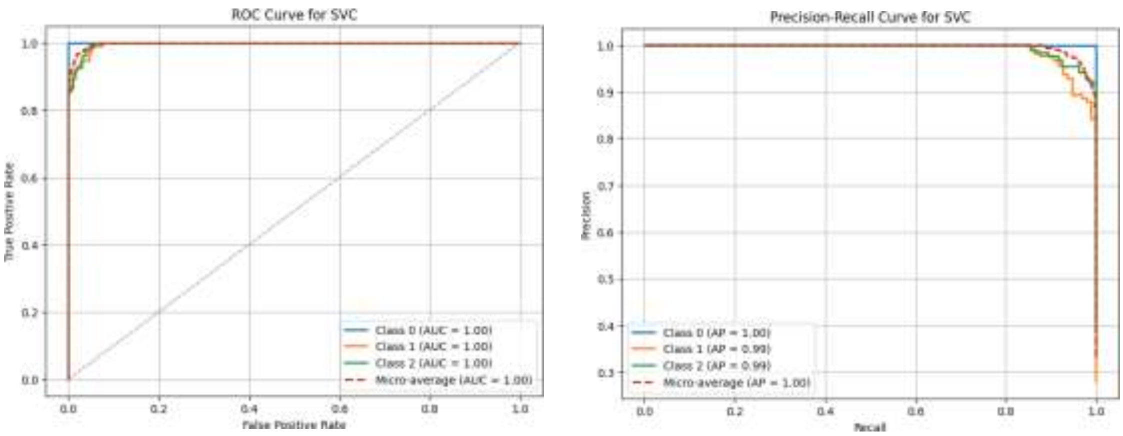
Figure 10, 11, and 12 denotes ROC (Receiver Operating Characteristics) and Precision-Recall curves which showcases TPR, Precision versus Recall.



a) b)

Figure 10: Roc and PR-Curve for XGB using Color HOG as Feature Extractor
a) ROC, b) PR-Curve

In Fig.6 XGB shows slightly dips for some classes with PR and AOC around 0.99 suggesting minor imperfection compared to SVC.



a) b)

Figure 11: Roc and PR-Curve for SVC using Color HOG as Feature Extractor
a) ROC, b) PR-Curve

In Fig.7 SVC shows nearly perfect AUC and PR values of 1.0 for all classes, meaning the model performs exceptionally well across both curves denoting less FPR and FNR.

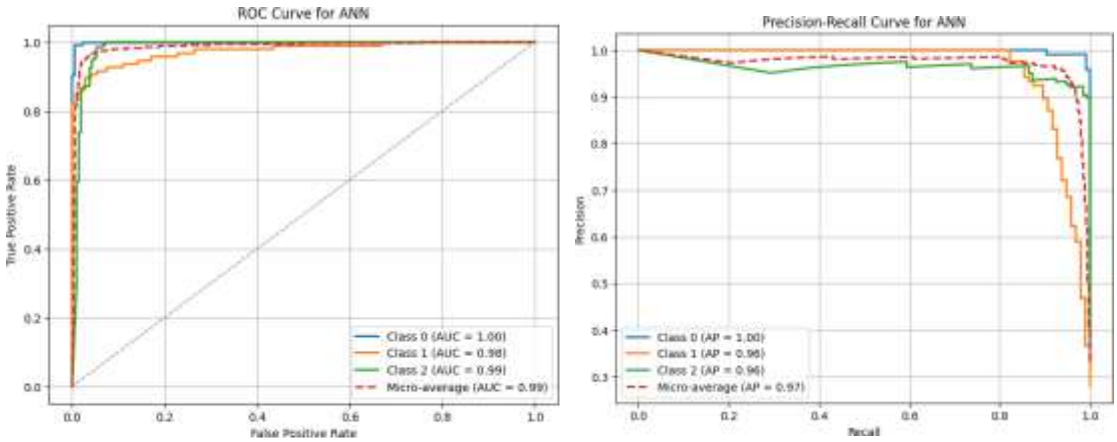


Figure 12: Roc and PR-Curve for ANN using Color HOG as Feature Extractor
a) ROC, b) PR-Curve

In Fig.12 ANN shows AUC of 0.99 denotes ANN performs very well in distinguishing between the classes with only minimal misclassification and AP of 0.97 confirms that ANN achieves high precision and recall effectively identifying most positives samples while keeping false positives low.

Conclusion

This project successfully developed and evaluated Machine Learning (ML) and Deep Learning (DL) models for classifying LCCV and LCBV in large cardamom plants. Among the tested feature extraction techniques, Color HOG consistently delivered superior results due to its ability to capture gradient and edge information critical for disease detection. Models like CNN, ANN, and SVC performed exceptionally well, with CNN achieving the most balanced and reliable metrics across all evaluation parameters.

The use of data augmentation played a pivotal role in enhancing model robustness and preventing overfitting. Techniques such as rotation, scaling, and flipping effectively diversified the dataset, leading to improved generalization.

Future work should focus on exploring advanced architectures like ResNet or Vision Transformers and expanding the dataset to include more diverse diseases with more samples. Additionally, further optimization of feature extraction and preprocessing techniques could enhance performance further.

This study demonstrated the feasibility of using Machine Learning as a solution for cardamom disease classification, setting a strong foundation for future advancements in agricultural technology.

References

[1] K. Gurung, K. Dasila, A. Pandey, and N. Bag, "Curvularia eragrostidis, a new threat to large cardamom (Amomum subulatum Roxb.) causing leaf blight in Sikkim," Journal of Biosciences, vol. 45, no. 113, 2020. [Online]. Available: https://doi.org/10.1007/s12038-020-00099-9

[2] A. K. Vijayan, B. A. Gudade, T. N. Deka, and P. Chhetri, "Status of viral diseases of large cardamom (Amomum subulatum Roxb.) and its management in Sikkim and Darjeeling, West Bengal," Journal of Mycology and Plant Pathology, vol. 44, no. 4, pp. 374-381, 2014.

[3] L. Li, S. Zhang, and B. Wang, "Plant disease detection and classification by deep learning," IEEE Access, 2021. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3067480

[4] S. D. Khirade and A. B. Patil, "Plant disease detection using image processing," in 2015 International Conference on Computing Communication Control and Automation, 2015, pp. 768-771. [Online]. Available: https://doi.org/10.1109/ICCUBEA.2015.153

[5] S. Bhosale and A. Chhabria, "Tomato plant disease identification via deep learning technique," International Journal of Image and Graphics, 2024. [Online]. Available: https://doi.org/10.1142/S0219467824500356

[6] X. Wang and J. Liu, "An efficient deep learning model for tomato disease detection," Plant Methods, vol. 20, 2024. [Online]. Available: https://doi.org/10.1186/s13007-024-01004-6

[7] Anonymous, "Chirkey and Foorkey disease of cardamom." [Online]. Available: https://www.slideshare.net/slideshow/chirkey-and-foorkey-disease-of-cardamom/137825827

[8] Anonymous, "Chirkey and Foorkey: The dangerous viral diseases of large cardamom in Sikkim," Sikkim Express. [Online]. Available: https://www.sikkimexpress.com/news-details/chirkey-and-foorkey-the-dangerous-viral-diseases-of-large-cardamom-in-sikkim

[9] M. Xu, S. Yoon, Y. Jeong, J. Lee, and D. S. Park, "Transfer learning with self-supervised vision transformer for large-scale plant identification," in Conference and Labs of the Evaluation Forum, Bologna, Italy, 2022. [Online]. Available: https://doi.org/10.1007/978-3-030-99836-1_37

[10] M. Jogin, M. Mohana, M. M. Madhulika, G. Divya, R. Meghana, and S. Apoorva, "Feature extraction using convolution neural networks (CNN) and deep learning," Institute of Electrical and Electronics Engineers Inc., 2018. [Online]. Available: https://doi.org/10.1109/SAI.2014.6918213

[11] S. Nasreen, T. Khalil, and S. Khalid, "A survey of feature selection and feature extraction techniques in machine learning," Institute of Electrical and Electronics Engineers Inc., 2014. [Online]. Available: https://doi.org/10.1109/SAI.2014.6918213.

[12] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, USA, 2005, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[13] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, "Normalization techniques in training DNNs: methodology, analysis and application," *arXiv*, 2020. [Online]. Available: https://arxiv.org/abs/2009.12836.

[14] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," *IEEE Access*, vol. 10, pp. 96451-96472, 2022, doi: 10.1109/ACCESS.2022.3207287.

[15] C. K. Sunil, C. D. Jaidhar, and N. Patil, “Cardamom Plant Disease Detection Approach Using EfficientNetV2,” *IEEE Access*, vol. 10, pp. 789–804, 2022, doi: 10.1109/ACCESS.2021.3138920.