Early Disease Detection and Classification in Mango Crop using Machine Learning Techniques

In partial fulfilment of the requirement for Award of Degree of

Doctor of Philosophy (CSE)

Submitted by

Name of scholar: VIPIN RAWAT

Enrolment No.: A7666223006

Batch: JAN 2023

Mode (Full-Time / Part-Time): Part Time

Name of Institute: Amity School of engineering and technology

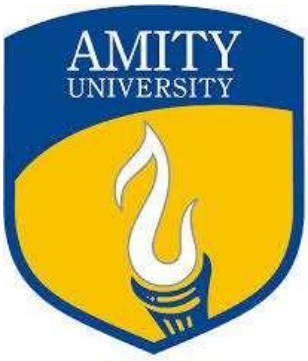
Under Supervision of

Name of Guide: Prof. (Dr.) Namrata Dhanda

Designation: Professor & Head

Department of Computer Science & Engineering

Institute: Amity School of engineering and technology



AMITY UNIVERSITY UTTAR PRADESH

(LUCKNOW CAMPUS)

TABLE OF CONTENT

1. Introduction:

Mango farming, a vital component of the agricultural sector, plays a crucial role in the global economy by providing a significant source of income and nutrition [1]. However, mango cultivation faces numerous challenges, with one of the most pressing being the early detection and classification of diseases that can severely impact yield and fruit quality. Innovative solutions that can revolutionize mango Crops practices are increasingly sought after due to the time-consuming, labour-intensive and error-prone nature of traditional methods of disease monitoring and management.

A powerful tool in the field of agriculture has emerged as machine learning, a subset of artificial intelligence [9]. The way for the development of advanced applications in the field of early disease detection and classification has been surfaced by its ability to process vast amounts of data, recognize complex patterns, and make predictions with remarkable accuracy. By using the potential of machine learning algorithms, mango farmers can now benefit from more efficient and proactive disease management strategies.

This research work explores the exciting prospects of machine learning applications for early disease detection and classification in mango crop. We investigate into the current challenges facing mango cultivation, discuss the advantages of integrating machine learning into agricultural practices, and highlight some of the cutting-edge techniques and technologies that are transforming the industry [8]. From the identification of common mango diseases to the development of predictive models, this research aims to shed light on the transformative role that machine learning can play in ensuring healthier mango crops and sustaining this essential industry.

2. Rational of Proposed Research Work:

Proposed research work on "Early Disease Detection and Classification in Mango Crop using Machine Learning Techniques"is grounded in the following key considerations:

1) **Crop Disease Threat:** Mango Crop faces a persistent and escalating threat from various diseases, including fungal, bacterial, and viral infections [11]. These diseases can lead to substantial crop losses if not promptly and effectively managed. Early detection and accurate classification of these diseases are pivotal for timely interventions and reducing yield losses.
2) **Lack of Timely Detection:** Conventional disease detection methods in mango crop often rely on visual observation, which is prone to human error and may only detect diseases at advanced stages when they are harder to manage. The research aims to address this limitation by leveraging machine learning, which can identify subtle disease symptoms and patterns at an early stage [10].
3) **Economic Impact:** Mango Crop is a significant contributor to the agricultural economy in many regions. Disease outbreaks can have severe economic repercussions, affecting the

livelihoods of farmers and the income of nations that rely on mango exports. Developing effective disease detection and classification tools can mitigate these economic losses.

4) **Sustainability:** Sustainable agriculture practices are gaining importance globally. Reducing the use of chemical pesticides and adopting more targeted interventions is crucial for sustainable mango farming [17]. Machine learning can contribute to sustainability by optimizing the use of resources and minimizing environmental impacts.
5) **Technological Advancements:** Recent advances in machine learning, computer vision, and sensor technologies offer unprecedented opportunities for automating disease detection and classification tasks. Leveraging these technologies can lead to innovative, efficient, and scalable solutions for mango farming [11].
6) **Data Availability:** The increasing availability of data, including high-resolution imagery, weather information, and historical disease records, provides a rich foundation for machine learning applications. This research can tap into this data wealth to develop robust and accurate disease detection models.
7) **Scientific Contribution:** The proposed research contributes to the growing body of knowledge at the intersection of agriculture and artificial intelligence. It offers machine learning techniques to solve a real-world problem with significant societal and economic implications.
8) **Global Food Security:** Mangoes are an important source of nutrition in many regions. By enhancing disease management in mango farming, this research has the potential to contribute to global food security by ensuring a stable supply of this nutritious fruit.

3. Background/Review of the Literature

S.no	Author	Year	Techniques Used	Key Findings
1	Kamble et al. [1]	2017	Support Vector Machines (SVM), Image Processing Tools	Utilized SVM for mango disease classification based on leaf images. Achieved high accuracy in disease classification.
2	Meena and Sivakumar [2]	2018	Remote Sensing, Decision Support System	Developed a decision support system for mango disease diagnosis. Utilized machine learning algorithms and remote sensing techniques for mango tree health assessment.
3	Hassan et al. [3]	2019	Convolutional Neural Networks (CNNs), Image Processing	Investigated the effectiveness of CNNs for early detection of mango diseases using leaf images. - Achieved high accuracy in disease

				classification.
4	Arivazhagan and Shebiah [4]	2019	Deep Learning, CNNs, Image Classification	Explored deep learning algorithms, particularly CNNs, for mango leaf disease classification. Achieved high accuracy in disease identification.
5	Swaminathan et al. [5]	2020	Machine Learning Algorithms, Weather Data Analysis	Utilized machine learning algorithms to analyze weather and environmental data for mango disease prediction. Integration of weather data improved prediction accuracy.
6	Rajan et al. [6]	2021	Internet of Things (IoT), Sensors, Machine Learning	Developed a real-time monitoring system for mango plant health using IoT technology and machine learning. IoT sensors and machine learning facilitated early disease detection and plant health assessment.
7	Verma et al. [7]	2021	Satellite Remote Sensing, GIS, Machine Learning	Integrated satellite remote sensing, machine learning, and GIS for regional monitoring and prediction of mango diseases. Enabled early disease detection across large areas.
8	D. Fayeet al. [8]	2022	ML Algorithms, CNNs,	Machine learning (ML) applications for image processing, various image processing (IP) techniques are employed to enhance the quality and suitability of images used for both training and testing purposes.

This table summarizes the key literature findings and the techniques used in machine learning applications for early disease detection and classification in mango farming.

3. Research gaps identified.

Several research gaps can be identified based on the existing literature:

1) **Limited Data Availability:** Mango disease datasets are often limited in size and diversity, making it challenging to train accurate machine learning models.
   a. **Small Dataset Size**: Collecting a real amount of data related to mango diseases, especially labelled data (data with disease labels indicating which disease is present), can be challenging. Smaller datasets may not provide enough variability for machine learning models to generalize effectively.
   b. **Data Quality**: Data collected from the field can be noisy, incomplete, or contain errors. Ensuring data quality and consistency is crucial for reliable machine learning model training.
2) **Early Disease Detection:**"Early Disease Detection" in the context of mango crop refers to the ability to identify the presence of diseases in mango trees at their initial stages or even before visible symptoms become apparent to the human eye [14]. Detecting diseases early in the growing cycle is crucial for effective disease management and preventing the spread of diseases that can cause significant damage to mango crops.
   a. Pre-symptomatic Detection: Early disease detection involves identifying diseases in mango trees before they exhibit visible symptoms or signs of infection. This detection can occur at a stage when the disease is still in its latent or incubation period. In some cases, subtle biochemical or physiological changes in the plant may be the only indicators of disease presence.
   b. Subtle Symptom Identification: In some instances, mango trees may show subtle and easily overlooked symptoms during the early stages of disease development. These symptoms may include changes in leaf color, texture, or moisture content, which can be challenging for human observers to detect without the aid of specialized tools.
3) **Localized Solutions:** Mango farming practices can vary significantly from one region to another and disease incidence may differ appropriately. Research should address the need for localized disease detection and management solutions that consider regional variations in mango cultivation.
4) **Disease Classification:** Disease classification is the process of categorizing detected diseases in mango trees into specific classes or types based on the symptoms, visual cues, or other diagnostic features. This classification is a crucial step in the disease management process, as it helps farmers and wood managers identify the specific disease affecting their mango trees and take appropriate actions.

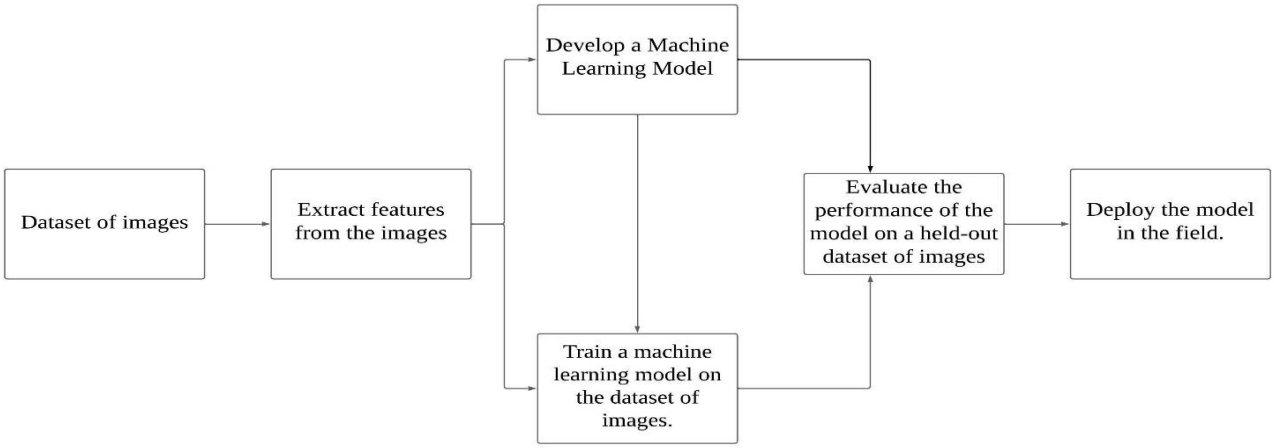
5) **Environmental Variability:** Mango farms experience variations in environmental conditions, such as changes in weather, soil properties, and lighting.
   a. **Environmental Factors:** Environmental factors in mango farming include weather conditions (temperature, humidity, rainfall), soil properties (composition, moisture content), lighting conditions (sunlight, shading), and other local variables that affect the growth and health of mango trees.
   b. **Data Collection Challenges:** The changing environmental conditions pose challenges to data collection for disease detection. Images or sensor data collected at different times and under varying conditions may show differences in lighting, leaf appearance, and disease symptoms.

4. Scope of the proposed study

The scope of the proposed study, "Early Disease Detection and Classification in Mango Crop using Machine Learning Techniques "encompasses several key aspects that define the boundaries and objectives of the research:

1) **Disease Detection and Classification:** The primary focus of the study is on the development and evaluation of machine learning models for the early detection and accurate classification of diseases affecting mango trees. This includes various types of diseases such as **Phomablight, Powdery mildew, Red rust, Sooty mould, Mango malformation, Scab and Anthracnose.**
2) **Data Sources:** The study will consider a variety of data sources, including high-resolution images of mango leaves and fruit, spectral data, weather information, and historical disease records. These data sources will be used for training and validating machine learning models.
3) **Machine Learning Techniques:** The research will explore a range of machine learning techniques, including but not limited to CNN (convolutional neural networks), SVM (support vector machines), and DT (Decision trees), to develop disease detection and classification algorithms [4]. The choice of techniques will depend on their suitability for the specific tasks.
4) **IoT and Sensor Integration:** The study will investigate the integration of Internet of Things (IoT) devices and sensors for real-time disease monitoring in mango orchards [6]. This will involve the development of a system that can provide timely alerts to farmers when disease conditions are detected.
5) **Practical Implementation:** While the focus is on research and development, the study will consider the practical implementation of machine learning-based disease detection and classification solutions [15]. This includes evaluating the feasibility and scalability of the proposed technologies for real-world mango crops.

5. Method and Design



The working process of a "Early Disease Detection and Classification in Mango Crop using Machine Learning Techniques "can be summarized as follows:

1. The model is first trained on a dataset of images of healthy and diseased mango leaves. The dataset should be large and diverse to ensure that the model learns to generalize well to new data.
2. During training, the model learns to identify the features that are most relevant to the task of disease detection and classification [16]. These features can be based on color, texture, shape, or other characteristics of the leaves.
3. The probability of a new mango leaf being diseased can be predicted using the trained model.To do this, the model first extracts the features from the new leaf image. It then uses these features to predict the probability of the leaf being diseased.
4. If the probability is above a certain threshold, the model will classify the leaf as diseased. Otherwise, the leaf will be classified as healthy.

5.1 Training

The training process for a machine learning model typically involves the following steps:

1. The dataset is split into training and test sets, with the training set being utilized for model training, and the test set employed for assessing the model's performance on previously unseen data.
2. An appropriate machine learning algorithm is selected based on the specific task's requirements, as well as the dataset's size and complexity, taking into consideration the unique strengths and weaknesses of each algorithm available.
3. Train the model. This involves feeding the training set to the machine learning algorithm and allowing it to learn the relationship between the features and the labels.

4. Evaluate the model on the test set. This involves feeding the test set to the model and measuring its performance [12]. If the model is not performing well on the test set, the training process can be repeated with different parameters or a different algorithm.

5.2 Prediction

1. Extract the features from the new leaf image. This involves using the same feature extraction techniques that were used during training to extract the features from the new leaf image.
2. Predict the probability of the new leaf being diseased. This involves using the trained model to predict the probability of the new leaf being diseased, given the extracted features.
3. Classify the new leaf as diseased or healthy. This involves comparing the predicted probability to a certain threshold [13]. If the probability is above the threshold, the leaf will be classified as diseased. Otherwise, the leaf will be classified as healthy.

S. No	Research method	Description
1	Data collection	Collect a dataset of images of healthy and diseased mango leaves. The dataset should be collected under a variety of conditions, such as different lighting, angles, and backgrounds.
2	Feature extraction	Extract features by the dataset of images. Features can be based on color, texture, shape, or other characteristics of the leaves. Careful consideration should be given to the selection of features so that they can be employed to differentiate between healthy and diseased leaves.
3	Machine learning Model Training	Train this model on the dataset of images. The model can be trained using a variety of ML algorithms, such as DT(Decision Trees), or neural networks , SVM (Support Vector Machines)[4]. The specific features extracted from the images will determine the choice of algorithm.
4	Evaluation of Model	Evaluate or check the performance of the model from dataset of images.
5	Deployment of Model	Deploy the model in the field. This can be done by building a mobile app or website that allows farmers to take images of their mango leaves and get a diagnosis of the disease.

6. Tentative Chapter Plan

1. **Chapter 1 - Introduction**
2. **Chapter 2 - Literature survey**
3. **Chapter 3 –Research Problem 1**
4. **Chapter 4 -Research Problem 2**
5. **Chapter 5 – Research Problem 3**
6. **Chapter 6- Research Problem 4**
7. **Chapter 7 – Discussion and Results**
8. **Chapter 8 - Conclusion and future scope**

7. Plan of Work & Time Schedule

A research plan is outlined, which serves as a roadmap for conducting the research project systematically. It ensures that the objectives are achieved, and the desired outcomes are obtained within the specified timeframe.

	Tasks/ Activities	Start	End	Duration (Days)	% Completed
1	**Basics Information Gathering / Literature Survey**	02/01/2023	08/31/2024	573	**100%**
1.1	Introduction	02/01/2023	04/30/2023	88	**36.31%**
1.2	Data Mining and Machine Learning Techniques	05/01/2023	06/30/2023	60	
1.3	Data collection and analysis	07/01/2023	09/30/23	91	
1.4	Learning essential tools	10/01/23	02/29/2024	151	
1.5	Other related international journals/Indian Journals	03/01/24	8/31/2024	183	
2	**Research Problem-Selection and Solution**	09/01/2024	10/31/2026	786	**100%**
2.1	Identifying Techniques& Algorithms	09/01/2024	10/31/2024	633	

2.2	Comparisons	11/01/2024	01/31/2025	724	**50..22%**
2.3	Analysis of algorithms	02/01/25	10/31/2025	996	
2.4	Performance Evaluation	11/01/2025	04/30/2026	1176	
2.5	Result verifications using data sets available	05/01/2026	10/31/2026	1359	
3	**Thesis Writing & Submission**	11/01/2026	05/31/27	206	**100%**
3.1	Composition	11/01/2026	01/30/2027	1449	**13.16%**
3.2	Editing &Printing	02/01/2027	03/31/2027	1507	
3.3	Binding	04/01/2027	04/30/2027	1536	
3.4	Submission	05/01/2027	05/30/2027	1565	

6. References

1. Kamble, S. S., Ghatol, A. A., & Jondhale, V. S. (2017). "Mango Disease Detection and Classification Using Image Processing Techniques." International Journal of Computer Applications, 159(10), 35-39.
2. Meena, A., & Sivakumar, M. R. (2018). "Decision Support System for Mango Disease Diagnosis Using Machine Learning." 2018 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), 225-229.
3. Hassan, M. A., Arafat, S. M. Y., & Mamun, K. A. (2019). "Early Detection and Classification of Mango Diseases Using Deep Convolutional Neural Networks." 2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 345-350.
4. Arivazhagan, S., &Shebiah, R. N. (2019). "Deep Learning for Classification of Mango Leaf Diseases Using Convolutional Neural Networks." Journal of Ambient Intelligence and Humanized Computing, 10(9), 3525-3535.
5. Swaminathan, S., Arulmozhi, R., & Raghunathan, T. (2020). "Predictive Modeling of Mango Disease Incidence Using Machine Learning Algorithms." International Journal of Environmental Science and Technology, 17(2), 1117-1126.
6. Rajan, D., Paul, A., & Kumar, P. (2021). "IoT-Based Early Detection and Classification of Mango Plant Diseases Using Machine Learning." Journal of Ambient Intelligence and Humanized Computing, 12(7), 6599-6611.
7. Verma, A., Mishra, A., & Jain, M. (2021). "Integration of Satellite Remote Sensing, GIS, and Machine Learning for Early Detection and Monitoring of Mango Diseases at a Regional Scale." Remote Sensing, 13(1), 117.

8. D. Faye, I. Diop, and D. Dione, "Mango Diseases Classification Solutions Using Machine Learning or Deep Learning: A Review," in Journal of Computer and Communications (pp. 16-28, 2022).
9. Raut, S., & Chavan, P. (2019). "Mango Leaf Disease Detection and Classification Using Convolutional Neural Network." 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT), 1-5.
10. Gogoi, D., Dutta, P., & Deka, C. R. (2018). "Disease Detection and Classification on Mango Leaves Using Machine Learning." 2018 International Conference on Computing, Power and Communication Technologies (GUCON), 48-52.
11. Mahalakshmi, V., & Meena, M. (2020). "Mango Leaf Disease Detection and Classification Using Machine Learning Algorithms." International Journal of Advanced Trends in Computer Science and Engineering, 9(4), 4910-4917.
12. Shanmugapriya, V., & Thangavel, K. (2019). "Identification and Classification of Mango Leaf Diseases Using Machine Learning Techniques." International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 4(3), 170-175.
13. Islam, M. A., Khan, M. A., & Rahman, M. S. (2020). "Mango Leaf Disease Detection and Classification Using Machine Learning." 2020 International Conference on Electrical, Computer and Communication Engineering (ECCE), 1-5.
14. Aslam, M. J., Asghar, M., & Shafique, M. (2020). "An Effective Approach for Mango Disease Detection and Classification Using Deep Learning Techniques." IEEE Access, 8, 99259-99269.
15. Rajendra, A. J., & Nagaraj, H. N. (2019). "Automatic Detection and Classification of Mango Leaf Diseases Using CNN." 2019 International Conference on Automation, Computational and Technology Management (ICACTM), 89-94.
16. Singh, S., & Jindal, M. (2019). "Mango Leaf Disease Classification Using Convolutional Neural Networks." 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), 1-6.
17. Choudhary, R. B., & Jaiswal, M. (2020). "Disease Detection in Mango Plant Leaves Using Deep Learning." In Proceedings of International Conference on VLSI, Communication and Signal Processing (pp. 389-394).
18. Ali, I., Siddiqui, H. U., & Khan, M. A. (2020). "Mango Disease Detection and Classification Using Machine Learning Techniques." In Proceedings of International Conference on Machine Learning and Soft Computing (pp. 309-318).

Signature of the scholar

Name & Signature of Supervisor(s) with date _____

Name & Signature of SRC Members:

1. Prof. (Dr.) O.P. Singh

2. Prof. (Dr.) S.W.A Rizvi

3. Dr. Sheenu Rizvi

Chairperson SRC/DRC