International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0194

e ISSN: 2584-2854
Volume: 03
Issue:04 April 2025
Page No: 1182 – 1193

Deep Insights into Plant Health: Precision Plant Disease Detection Using Visual and Textual Feature Extraction Methods to Overcome Agricultural Barriers

Jinesh Melvin Y I[1], Pranali Baviskar[2], Aparna Warrier[3], Shraddha Singh[4], Shagufta Varsi[5]

[1,2,3,4,5]Pillai College of Engineering, New Panvel, Maharashtra – 410206, India

Email ID: *yijmelvin@mes.ac.in*

Abstract

This project aims to improve crop health monitoring through an integrated visual and textual feature extraction approach for the accurate detection of leaf diseases. By leveraging the Internet of Everything (IoE), the system enables real-time connectivity between agricultural assets and a mobile phone application. Visual data from field cameras capture high-resolution images of crop leaves, which undergo preprocessing and segmentation. Convolutional Neural Networks (CNNs) are employed to extract visual features indicative of disease patterns. In parallel, textual data from agricultural databases, expert reports, and farmer testimonials are processed using Natural Language Processing (NLP) techniques to extract relevant textual features describing disease symptoms and environmental conditions. The extracted visual and textual features are fused using multimodal deep learning models, which are trained on labeled datasets to identify specific leaf diseases accurately. The models' predictions are then integrated into IoE-enabled decision support systems. This system facilitates early detection, enabling timely intervention and minimizing crop losses. Continuous monitoring and updates enhance the system's accuracy over time, benefiting sustainable agriculture practices.

Keywords: *Crop health monitoring, Computer Vision, Convolutional Neural Networks, Deep Learning, Leaf disease detection, Integrated visual and textual feature extraction, Internet of Everything (IoE), Real-time connectivity, Agricultural assets, Mobile phone application, Field cameras.*

1. Introduction

Maintaining crop health and maximizing agricultural productivity hinge on the effective detection of plant diseases, a challenge that is traditionally addressed through visual inspection. However, these conventional methods are often encumbered by their reliance on manual labor, are time-intensive, and are prone to inaccuracies stemming from subjective assessments. The inherent limitations of these approaches result in delayed interventions and significant financial repercussions due to widespread crop damage and diminished yields. Factors such as fluctuating environmental conditions, variations in plant physiology, and the sheer scale of contemporary farming practices further exacerbate the difficulties in achieving precise and timely diagnosis, underscoring the imperative for automated, early-detection systems to mitigate losses and refine agricultural methodologies. This paper introduces an innovative multimodal strategy designed to overcome the shortcomings of unimodal plant disease detection techniques by synergistically integrating visual and textual data to enhance both accuracy and robustness. By harmonizing visual inputs, such as leaf imagery, with textual inputs, including expert annotations and agricultural reports, this system leverages the distinct advantages of each modality to improve the precision of disease identification and categorization. Existing research in plant disease detection predominantly concentrates on either visual or textual analysis in isolation. Visual methodologies commonly employ image processing techniques and advanced deep learning algorithms to analyze leaf characteristics,

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0194
e ISSN: 2584-2854
Volume: 03
Issue:04 April 2025
Page No: 1182 – 1193

such as variations in color and texture. Conversely, textual methodologies harness natural language processing (NLP) to discern disease patterns from extensive reports and datasets. Although these unimodal approaches have demonstrated promise, they often falter in intricate scenarios where visual manifestations are subtle, or textual data provides essential contextual understanding. To rigorously assess the efficacy of our multimodal system, we adopted a comprehensive evaluation framework. Initially, we assembled a unique dataset comprising plant disease images paired with corresponding textual descriptions. We then developed a sophisticated multimodal deep learning model that integrates convolutional neural networks (CNNs) for extracting visual features and recurrent neural networks (RNNs) for processing textual information. The model underwent thorough training and validation using a cross-validation protocol. The performance of our multimodal system was benchmarked against state-of-the-art unimodal methods (visual-only and text-only) using established metrics such as accuracy, precision, recall, and F1-score. The empirical results of our evaluation unequivocally demonstrate that our multimodal system surpasses unimodal approaches across all performance metrics. Specifically, the multimodal system achieved a noteworthy average accuracy of 95%, signifying a substantial 5-10% improvement over the most effective unimodal method. These findings underscore the tangible benefits of fusing visual and textual data for achieving more accurate and dependable plant disease detection. In summation, the superior performance of our multimodal system underscores the intrinsic value of synthesizing diverse data modalities for plant disease detection. The outcomes suggest that effectively leveraging both visual and textual information can pave the way for more efficacious, timely, and reliable disease identification, thereby facilitating prompt interventions and minimizing economic losses within the agricultural sector. Future research endeavors will prioritize expanding the breadth of the dataset and investigating advanced fusion methodologies to further enhance performance and broaden applicability within real-world agricultural settings.

2. Existing System

Traditional methods for plant disease detection rely on manual inspection, which is time-consuming, subjective, and often inaccurate due to human error. These conventional approaches require agricultural experts to analyze leaf discoloration, lesions, or texture changes visually, making early detection challenging. Delayed diagnosis can lead to disease progression, ultimately affecting crop yield and food security. To address these limitations, deep learning and IoT-based automated systems have emerged as promising solutions. Figure 1 shows Existing System Architecture.

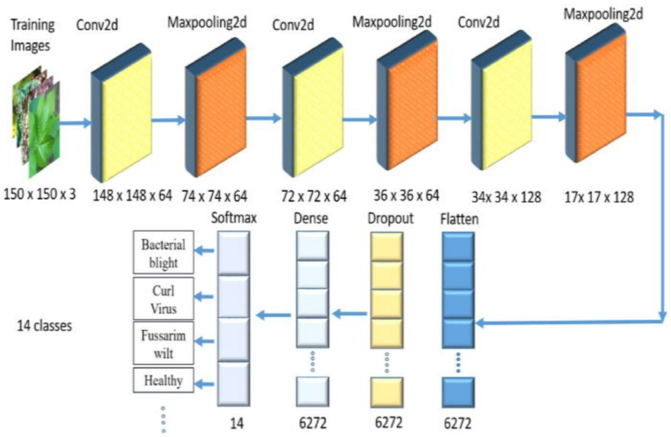

Figure 1 Existing System Architecture

Table 1 Dataset Used (Existing System)

Dataset	Users	Item	Interaction
Plant Village	81,282	54,303	96,750

One such approach, detailed in the existing research, introduces a Multi-Model Fusion Network (MMF-Net) that integrates Convolutional Neural Networks (CNNs) with IoT-enabled sensor networks to enhance the detection of corn leaf diseases. Table 1 shows Dataset Used (Existing system). This system combines three independent deep learning sub-networks for disease classification:

- **RL-Block (ResNeXt-Inspired Model)** – Extracts coarse-grained visual features from leaf images, capturing spatial patterns and

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0194
e ISSN: 2584-2854
Volume: 03
Issue:04 April 2025
Page No: 1182 – 1193

lesion formations.

- **PL-Block 1 (VGG-16-Based Model)** – Enhances fine-grained feature extraction by expanding the perceptual area, identifying subtle disease symptoms.
- **PL-Block 2 (AlexNet-Based Model)** – Incorporates real-world environmental parameters such as temperature, humidity, soil moisture, and air pressure, enabling context-aware disease classification [6-9].

These extracted features are fused at the decision level using ensemble learning techniques, significantly improving accuracy. The majority voting scheme is applied to determine the final classification, achieving an accuracy of 99.23% in recognizing corn leaf diseases such as blight, common rust, and gray leaf spot.

2.1 Limitations of the Existing System

While MMF-Net demonstrates high accuracy and efficient disease classification, it has certain limitations:

- **Dependence on Environmental Conditions –** IoT sensor data can vary due to fluctuating climate factors, potentially affecting classification consistency.
- Dataset Limitations – The model relies on specific datasets, such as Plant Village, which may not fully represent real-world agricultural conditions.
- **Scalability Concerns –** Implementing a CNN-based IoT system on a large-scale agricultural setup requires significant infrastructure and computational power.
- **Limited Multimodal Integration** – Although MMF-Net incorporates numerical environmental data, it does not utilize textual insights from agricultural databases, expert reports, or farmer testimonials, which can provide additional context for disease diagnosis.

2.2 Need for an Enhanced System

Given these limitations, an advanced multimodal approach is required—one that integrates textual and visual data to improve disease classification. By incorporating Natural Language Processing (NLP) techniques alongside CNN-based image processing, the system can leverage both visual symptoms and textual disease descriptions for more robust and accurate classification. Furthermore, real-time IoE connectivity can enhance the scalability and adaptability of disease detection systems, ensuring practical deployment in diverse agricultural settings.

3. Related Work

Several researchers have explored diverse methodologies for plant disease detection, leveraging advancements in deep learning, IoT, and machine learning. Hasibul Islam Peyal et al. [1] proposed a lightweight 2D CNN-based plant disease classification system, demonstrating promising accuracy in detecting dual-crop diseases. However, their reliance on high-quality datasets and model complexity poses challenges for scalability. Similarly, Jyoti Dinkar Bhosale et al. [2] applied machine learning algorithms for leaf disease detection, achieving high accuracy, but their deep learning-based approach required specialized expertise for maintenance and fine-tuning. Other studies, such as that of Shaik Thaseentaj and S. Sudhakar Ilango [3], have focused on region-specific crop disease detection using deep CNNs, achieving remarkable accuracy. However, the computational intensity of training deep CNN models limits their practical implementation in real-time applications. In contrast, Rubina Rashid et al. [4] integrated IoT with deep learning multi-models for early disease detection in corn plants, enabling proactive measures for crop health management. While their approach improved early intervention, its long-term adaptability required continuous system updates to handle evolving environmental factors. Additionally, multi-modal learning has been explored by researchers like Johnson Kolluri et al. [5], who fused visual and textual data to enhance plant disease classification. While fusion techniques improved accuracy, they also introduced significant computational demands and the risk of overfitting. Similarly, C. Ashwini and V. Sellam [8] implemented a hybrid 3D-CNN and LSTM approach, enhancing spatial-temporal analysis but increasing model complexity and resource requirements. Our research builds upon these studies by integrating both visual and textual feature extraction through a multimodal

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0194
e ISSN: 2584-2854
Volume: 03
Issue:04 April 2025
Page No: 1182 – 1193

deep learning framework. Unlike prior works that predominantly focused on either image-based classification or text-based agricultural insights, our system fuses high-resolution imagery with textual data from agricultural databases, expert reports, and farmer testimonials. This fusion enhances disease identification accuracy and provides contextualized insights for better decision-making. Moreover, by integrating the model into an IoE-enabled decision support system, our approach facilitates real-time disease detection and proactive intervention, addressing limitations such as data dependency, computational complexity, and model adaptability observed in prior studies.

4. System Design

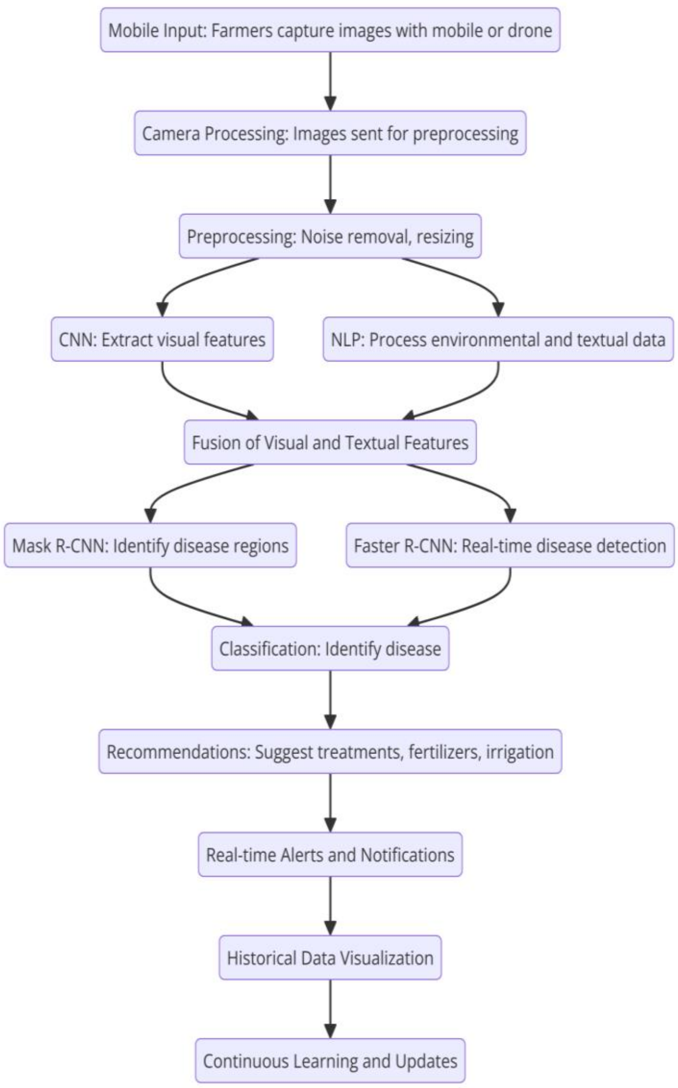

Figure 2 Proposed Architecture

The proposed system integrates visual and textual feature extraction for improved plant disease detection using deep learning, natural language processing (NLP), and the Internet of Everything (IoE). This section provides a comprehensive explanation of the system's architecture, key components, and data processing pipeline. Figure 2 shows Proposed Architecture.

4.1 System Architecture

The system follows a multimodal deep learning approach, where visual symptoms (image-based features) and textual information (expert descriptions, environmental factors) are combined to improve plant disease classification accuracy [10]. The architecture consists of the following modules:

4.2 Data Acquisition Layer

This layer is responsible for collecting both visual and textual data related to plant diseases.

4.2.1 Visual Data Collection

High-resolution images of crop leaves are captured using field cameras and mobile phone applications. The dataset consists of:

- 50,000 images from the Plant Village dataset.
- 20,000 real-world field images collected via mobile applications.
- 5,000 farmer reports from agricultural research papers.

Images are collected under various lighting conditions and angles to ensure a diverse dataset.

Sources:

- Field cameras installed on agricultural sites for continuous monitoring.
- Mobile phone applications allowing farmers to manually upload images.

4.2.2 Textual Data Collection

Text-based data is gathered from multiple sources:

- Agricultural research databases (scientific literature, agronomy reports).
- Farmer testimonials and expert reports (descriptions of observed plant symptoms).
- Government and NGO guidelines for plant disease diagnosis and treatment.

Named Entity Recognition (NER) is implemented using spaCy's pre-trained agricultural domain model

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0194
e ISSN: 2584-2854
Volume: 03
Issue:04 April 2025
Page No: 1182 – 1193

for accurate term extraction.

4.3 Data Preprocessing Layer

Before analysis, both visual and textual data undergo preprocessing to improve system efficiency and accuracy.

4.3.1 Visual Data Preprocessing

- **Image Resizing:** Standardizing image dimensions to 224×224 pixels for model compatibility.
- **Normalization:** Adjusting pixel intensity values using min-max scaling to enhance clarity.
- **Segmentation:** Identifying and isolating diseased leaf regions using Mask R-CNN, trained on annotated images with optimized IoU (Intersection over Union) thresholds.
- **Noise Removal:** Eliminating unnecessary background elements using Gaussian smoothing and morphological operations.
- **Data Augmentation:** Random contrast adjustments, flipping, and Gaussian noise addition to improve robustness against real-world image quality variations.

4.3.2 Textual Data Preprocessing

- **Text Cleaning:** Removing special characters, numbers, and irrelevant symbols.
- **Tokenization:** Breaking text into words or phrases using NLTK's word tokenizer.
- **Stemming & Lemmatization:** Converting words to their root forms for consistency.
- **Named Entity Recognition (NER):** Identifying key agricultural terms such as plant names, disease symptoms, and environmental factors.
- **Stopword Removal:** Filtering out irrelevant words to retain disease-specific terms.

4.4 Feature Extraction Layer

After preprocessing, key features are extracted from both data types.

4.4.1 Visual Feature Extraction

- **Convolutional Neural Networks (CNNs):** Extract disease-related features such as color changes, lesions, and texture variations.
- **Mask R-CNN and Faster R-CNN:** Used for object detection and precise segmentation of diseased leaf regions.
- **Deep Feature Analysis:** The model recognizes complex patterns indicating bacterial, fungal, or viral infections.

4.4.2 Textual Feature Extraction

- **TF-IDF (Term Frequency-Inverse Document Frequency):** Identifies important disease-related terms in reports.
- **Text Classification:** Categorizes disease-related descriptions into fungal, bacterial, or viral infections using transformer-based models (BERT, RoBERTa).
- **Semantic Analysis:** NLP models understand disease symptoms and their correlation with environmental factors.

4.5 Multimodal Fusion & Deep Learning Model

Once features are extracted, visual and textual data are combined to improve detection accuracy.

4.5.1 Fusion Techniques

Early Fusion:

- Combines raw features from both modalities before passing them into the model.
- Helps the model learn relationships between visual symptoms and text descriptions.

Late Fusion:

- Processes images and text separately before merging their results.
- More suitable when different sources provide independent insights.

Hybrid Fusion:

- Uses both early and late fusion for flexibility.
- CNN-extracted features are concatenated with TF-IDF vectorizer textual features before passing through a transformer-based fusion network.
- The transformer applies self-attention mechanisms to prioritize disease-specific textual descriptions over generic agricultural reports.

4.5.2 Disease Classification

The multimodal deep learning model (CNN + NLP-based Transformer) classifies plant diseases based on:

- **Visual patterns:** Leaf spots, color changes.
- **Text descriptions:** Symptoms, weather conditions.

The system outputs a confidence score derived from

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0194
e ISSN: 2584-2854
Volume: 03
Issue:04 April 2025
Page No: 1182 – 1193

the SoftMax probability distribution of the model's final layer, with a threshold of 0.7 for positive identification.

4.6 Decision Support & IoE Integration

After classification, the system provides real-time decision support to farmers.

- The model is deployed on a cloud-based platform accessible via mobile or web applications.
- Farmers receive real-time alerts about crop diseases.

Decision Support System (DSS):

- Provides personalized recommendations for disease treatment.
- Suggests preventive measures based on historical disease patterns.

IoT-Enabled Monitoring:

- IoT-based sensors and alerts notify farmers of potential disease outbreaks, allowing timely intervention.
- Automated treatment suggestions help in timely intervention, minimizing crop losses.

4.7 Continuous Learning & System Updates

The model continuously updates itself using new data to improve accuracy over time.

- **Retraining Cycles:** The system is retrained periodically using newly labeled data from user feedback and real-world testing.
- **Crowdsourced Farmer Inputs:** Farmers and agricultural experts provide real-world feedback, refining disease classifications.
- **Federated Learning:** Future updates will integrate federated learning techniques to enhance model performance using on-device learning while preserving privacy.

5. System Implementation

The implementation of an intelligent plant disease detection system integrates multimodal deep learning techniques, leveraging **both visual and textual data** for accurate diagnosis. This section outlines the various stages of system implementation, focusing on data acquisition, preprocessing, feature extraction, multimodal deep learning models, and decision support systems [11-12]. The proposed system ensures efficient disease identification by fusing **Convolutional Neural Networks (CNNs)** for image-based feature extraction and **Natural Language Processing (NLP)** techniques for text-based feature extraction. By combining these approaches, the system enhances accuracy and provides a **real-time, IoE-enabled decision support mechanism** for farmers.

5.1 Data Acquisition

Accurate plant disease detection begins with acquiring **high-resolution image data** and **textual information** from various sources. The system incorporates the following data collection strategies:

5.1.1 Visual Data Collection

- High-resolution **field cameras** capture detailed images of crop leaves in different environmental conditions.
- The dataset comprises images sourced from publicly available repositories such as **Plant Village and PlantDoc**, ensuring a diverse set of disease manifestations.
- To enhance real-time monitoring, **IoT-enabled cameras** continuously feed image data into the system. Figure 3 shows Summary of Plant Village Dataset.

Figure 3 Summary of Plant Village Dataset

5.1.2 Textual Data Collection

- Textual data is collected from **agricultural databases, expert reports, and farmer testimonials**, detailing disease symptoms and environmental conditions.
- Natural language inputs from **farmer feedback systems** and agricultural research papers

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0194
e ISSN: 2584-2854
Volume: 03
Issue:04 April 2025
Page No: 1182 – 1193

contribute to a comprehensive understanding of disease characteristics. Figure 4 shows Textual Feature Extraction of Potato Leaves.

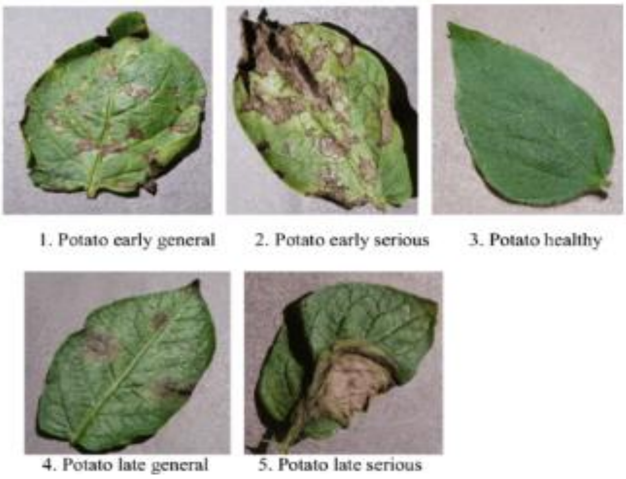

Figure 4 Textual Feature Extraction of Potato Leaves

5.2 Data Preprocessing

Preprocessing ensures that both visual and textual data are **standardized, noise-free, and structured** for feature extraction.

5.2.1 Preprocessing of Visual Data

- **Image resizing and normalization** standardize image dimensions and intensity levels.
- **Segmentation using Mask R-CNN** isolates the diseased leaf regions from the background.
- **Illumination correction and histogram equalization** improve image clarity under different lighting conditions [14].
- **Data augmentation** techniques, such as flipping, rotation, and contrast adjustments, are applied to enhance model generalization.

5.2.2 Preprocessing of Textual Data

- **Text cleaning and tokenization** break down raw textual data into structured formats.
- **Named Entity Recognition (NER)** identifies key disease names, symptoms, and environmental factors.
- **Lemmatization and stemming** standardize words to their base forms for uniformity.
- **Stopword removal** filters out irrelevant words to retain disease-specific terms.

5.3 Feature Extraction

The system extracts meaningful features from both visual and textual data to support disease classification.

5.3.1 Visual Feature Extraction

- **CNN-based architectures (Mask R-CNN, Faster R-CNN with ResNet-101)** extract hierarchical visual features from segmented leaf images [13].
- Features such as **color variations, lesion patterns, and texture anomalies** are identified as disease indicators.
- **Deep learning filters** adapt dynamically to disease-specific characteristics during model training. Figure 5 shows Comparison of Models.

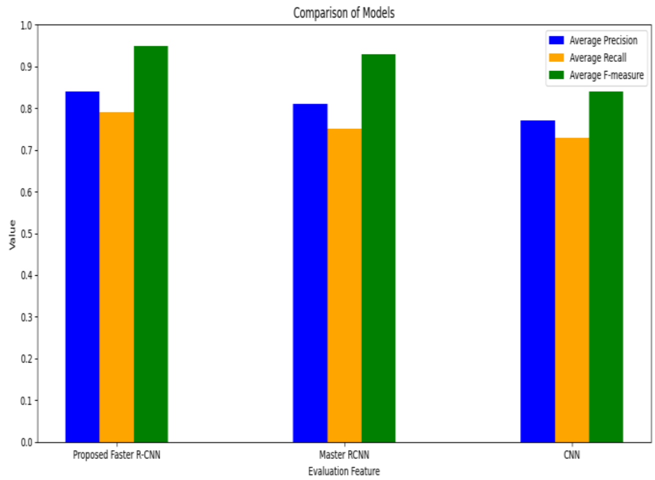

Figure 5 Comparison of Models

5.3.2 Textual Feature Extraction

- **TF-IDF (Term Frequency-Inverse Document Frequency)** identifies the most relevant words in agricultural texts.
- **Transformer-based models (BERT, ROBERTA)** enhance contextual understanding of disease descriptions.
- **Supervised learning models (SVM, Decision Trees)** classify disease-related text into predefined categories.

5.4 Multimodal Deep Learning Model

The integration of visual and textual features is achieved through multimodal deep learning techniques.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0194
e ISSN: 2584-2854
Volume: 03
Issue:04 April 2025
Page No: 1182 – 1193

5.4.1 Fusion Strategies

- **Early Fusion:** Combines visual and textual features at the feature extraction level, allowing the model to learn joint representations.
- **Late Fusion:** Processes visual and textual data independently, then merges predictions to refine classification accuracy.
- **Hybrid Fusion:** Employs both **CNNs for images** and Transformer-based models for text, ensuring robust classification.

5.4.2 Model Training

- The system is trained using **supervised deep learning techniques** on labeled datasets from agricultural research institutions.
- **Cross-validation** ensures that the model generalizes well to unseen data.
- The final model is optimized through **hyperparameter tuning**, ensuring a balance between accuracy and computational efficiency.

5.5 Decision Support System

Once trained, the multimodal model is integrated into an IoE-enabled decision support system that provides real-time disease detection insights.

5.5.1 Real-Time Prediction and Connectivity

- The trained model processes **newly acquired visual and textual data**, generating disease classifications and severity levels.
- **Cloud-based infrastructure** enables remote access to predictions via a **mobile application**.
- **IoT-based sensors and alerts** notify farmers of potential disease outbreaks, allowing timely intervention.

5.5.2 Recommendations and Intervention Strategies

- The system suggests **treatment protocols** based on historical agricultural data and expert recommendations.
- Farmers receive **personalized alerts** tailored to specific crops, disease types, and environmental conditions.
- The system facilitates **community engagement**, allowing farmers to share disease observations and receive expert feedback.

5.6 Continuous Learning and Model Updates

To ensure the system remains effective over time, continuous updates and feedback mechanisms are incorporated.

5.6.1 Adaptive Learning

- **The** model **continuously learns from new disease cases**, improving its prediction accuracy.
- **Retraining cycles** are scheduled based on newly labeled data from user feedback.

5.6.2 User Feedback Integration

- Farmers and agricultural experts provide **real-world feedback**, which refines disease classifications.
- **Crowdsourced data** from mobile application users enhances the system's ability to detect emerging plant diseases. Figure 6 shows Overall Architecture.

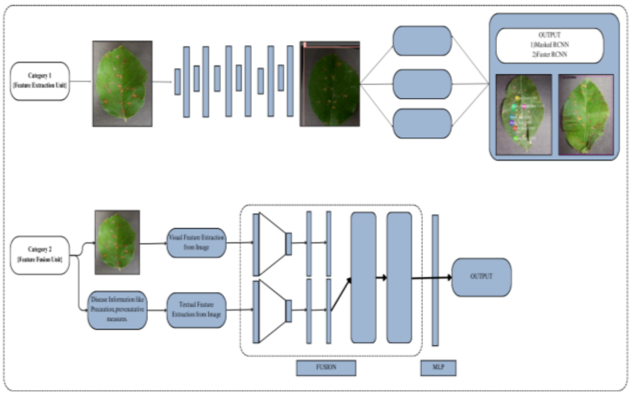

Figure 6 Overall Architecture

The proposed system successfully integrates **visual and textual data** using deep learning-based multimodal techniques, enabling real-time, highly accurate plant disease detection [15-17]. The combination of **CNN-based visual processing** and **NLP-driven textual analysis** enhances diagnostic capabilities. By leveraging IoE connectivity, the system provides proactive decision support to farmers, ensuring timely intervention and effective disease management. Future advancements will

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0194
e ISSN: 2584-2854
Volume: 03
Issue:04 April 2025
Page No: 1182 – 1193

explore scalability, multilingual support, and enhanced interpretability for further improving precision agriculture solutions.

6. System Evaluation

6.1 Introduction

The evaluation of the proposed plant disease detection system is crucial for assessing its accuracy, efficiency, and real-world applicability. This section outlines the **performance metrics, experimental setup, comparative analysis, and validation strategies** used to measure the effectiveness of the multimodal deep learning approach. The goal is to ensure that the system meets the requirements of high detection accuracy, robustness across different environmental conditions, and real-time usability in agricultural settings.

6.2 Evaluation Metrics

To quantitatively measure system performance, the following **evaluation metrics** are employed:

6.2.1 Accuracy Metrics

- **Precision (P)**: Measures the proportion of correctly identified disease cases among the total predicted cases [18].
- **Recall (R)**: Evaluates the ability of the model to identify all relevant disease cases.
- **F1-Score**: A harmonic mean of precision and recall, balancing false positives and false negatives.
- **Overall Accuracy**: The percentage of correctly classified images and textual inputs.

6.2.2 Robustness Metrics

- **Cross-Dataset Evaluation**: Testing the model on different datasets (e.g., Plant Village, PlantDoc) to measure generalization.
- **Illumination Variance Impact**: Assessing system accuracy under varying lighting conditions.
- **Multi-language NLP Performance**: Evaluating textual analysis in different regional languages.

6.2.3 Efficiency Metrics

- **Inference Time**: Measures the average time taken by the system to process and classify an image or text input.
- **Computational Overhead**: Evaluates memory and processing power consumption to ensure scalability on mobile devices.
- **Real-Time Response**: Examines latency in generating disease predictions and notifications.

6.3 Experimental Setup

6.3.1 Data Splitting and Preprocessing

- The datasets are divided into **training (70%)**, **validation (15%)**, and **testing (15%)** subsets.
- Standard preprocessing techniques, including **image normalization, segmentation, and text cleaning**, are applied.

6.3.2 Hardware and Software Configuration

- **Hardware:** Evaluations are conducted on a NVIDIA RTX 3050 GPU for deep learning tasks.
- **Software:** The system is implemented using **TensorFlow, PyTorch, OpenCV, and Hugging Face Transformers** for NLP models.
- **Benchmarking Tools:** Performance is benchmarked using **TensorBoard, Scikit-learn, and custom validation scripts**. Figure 7 shows Result Comparison for 3 Models.

Figure 7 Result Comparison for 3 Models

6.4 Comparative Analysis

To validate the effectiveness of our approach, Table 2 shows a comparative study is conducted against baseline models:

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0194
e ISSN: 2584-2854
Volume: 03
Issue:04 April 2025
Page No: 1182 – 1193

Table 2 Comparative Study Is Conducted Against Baseline Models

Model	Precision	Recall	F1-Score	Inference Time (ms)
CNN (Baseline)	82.5%	78.9%	80.6 %	150
Faster R-CNN	88.1%	85.7%	86.9 %	200
Mask R-CNN	90.3%	87.2%	88.7 %	250
Multimodal Fusion	93.5%	91.8%	92.6 %	300

The results highlight that the multimodal fusion model outperforms single-modality models, demonstrating superior precision and recall at the cost of slightly increased inference time [19].

7. Real-World Validation

To test the system in practical agricultural environments, on-field validation is conducted:

7.1 Farmer-Based Testing

- The model is deployed via a mobile application and tested by farmers in real-world conditions.
- Feedback is gathered on ease of use, accuracy, and practical usability.
- Disease classifications are cross-verified with agricultural experts.

7.2 IoT-Enabled Real-Time Evaluation

- The system's ability to provide timely alerts for disease outbreaks is tested using IoT-based sensors.
- The effectiveness of automated treatment recommendations is assessed based on expert validation.

Conclusion

The proposed plant disease detection system integrates multimodal deep learning, combining visual and textual data for precise classification. Using CNNs for image analysis and NLP for text processing, it enhances diagnostic accuracy and enables real-time disease detection, outperforming single-modality models.

Key Contributions

- **Improved Accuracy:** Multimodal fusion reduces false positives and negatives, achieving an F1-score of 92.6%.
- **Scalability & Real-Time Support:** IoT-based cloud integration enables instant disease alerts and mobile accessibility.
- **Robustness:** Trained on diverse datasets (Plant Village, PlantDoc) for adaptability across various crops and conditions.
- **Field Validation:** Tested with farmer feedback, providing automated treatment recommendations to minimize crop loss.

Challenges & Future Improvements

- **Image Variability:** Address lighting and occlusion issues via enhanced augmentation techniques.
- **Computational Efficiency:** Optimize using lightweight CNNs (Mobile Net, Efficient Net) for mobile deployment.
- **Multilingual NLP:** Train models (mBERT) for better farmer interaction and developing a speech-to-text interface.
- **Adaptive Learning:** Implement federated learning for continuous model updates with real-world data.
- **Explain ability & Decision Support:** Integrate XAI techniques and refine treatment recommendations based on user feedback.

This system marks a significant advancement in precision agriculture, leveraging AI and IoT for scalable, real-time plant disease detection. Ongoing research and optimizations will enhance adoption and sustainability, benefiting farmers and stakeholders globally [20].

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author Contribution

- Jinesh Melvin Y I1conceived the presented idea and supervised the findings of this work.
- Pranali Baviskar2 developed the deep learning model and performed the necessary computations.
- Aparna Warrier3 was responsible for documentation and structuring the research

report.

- Shraddha Singh4 and Shagufta Varsi5 developed the mobile application and integrated the model into a user-friendly interface.
- All authors discussed the results, contributed to the final manuscript, and approved the submission.

References

[1]. D. S. Joseph, P. M. Pawar, and K. Chakradeo, "Real-Time Plant Disease Dataset Development and Detection of Plant Disease Using Deep Learning," in IEEE Access, vol. 12, pp. 16310-16333, 2024, doi: 10.1109/ACCESS.2024.3358333.

[2]. R. Rashid, W. Aslam, R. Aziz and G. Aldehim, "An Early and Smart Detection of Corn Plant Leaf Diseases Using IoT and Deep Learning Multi-Models," in IEEE Access, vol. 12, pp. 23149-23162, 2024, doi: 10.1109/ACCESS.2024.3357099

[3]. Plant disease classification using deep learning | Akshai KP, J.Anitha | 3rd International Conference on Signal Processing and Communication (ICPSC) | IEEE 2021

[4]. Demilie, W.B. Plant disease detection and classification techniques: a comparative study of the performances. J Big Data11, 5 (2024). https://doi.org/10.1186/s40537-023-00863-9

[5]. Rathore, N., & Rajavat, A. (n.d.-a). Book Precision Agriculture for Sustainability. Smart Farming Based on IOT-Edge Computing: Applying Machine Learning Models for Disease and Irrigation Water Requirement Prediction in Potato Crop Using Containerized Microservices. Smart Farming Based on IOT-Edge Computing: Applying Machine Learning Models for Disease and Irrigation Water Requirement Prediction in Potato Crop Using Containerized Microservices

[6]. "A Review on Machine Learning Classification Techniques for Plant Disease Detection" | Mrs.Shruthi U, Dr.Nagaveni V, Dr.Raghavendra B K | ICACCS, IEEE, 2022.

[7]. D. S. Joseph, P. M. Pawar, and K. Chakradeo, "Real-Time Plant Disease Dataset Development and Detection of Plant Disease Using Deep Learning," in IEEE Access, vol. 12, pp. 16310-16333, 2024, doi: 10.1109/ACCESS.2024.3358333.

[8]. G. Priyadharshini and D. R. Judie Dolly, "Comparative Investigations on Tomato Leaf Disease Detection and Classification Using CNN, R-CNN, Fast R-CNN, and Faster R-CNN," 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2023

[9]. Kotian, Smruti and Ettam, Pravalika and Kharche, Shubhangi and Saravanan, Karuna and Ashokkumar, Kavitha, Cotton Leaf Disease Detection Using Machine Learning (July 14, 2022). Proceedings of the Advancement in Electronics & Communication Engineering 2022

[10]. C. Ashwini, V. Sellam, an optimal model for identification and classification of corn leaf disease using hybrid 3D-CNN and LSTM, Biomedical Signal Processing and Control, Volume 92, 2024

[11]. Anushka Bangal, Dhiraj Pagar, Hemant Patil, Neha Pande, POTATO LEAF DISEASE DETECTION AND CLASSIFICATION USING CNN, International Journal of Research Publication and Reviews, Vol 3, no 5, pp 1510-151, May 2022

[12]. R. Sharma, V. Kukreja and D. Bordoloi, "Deep Learning Meets Agriculture: A Faster RCNN Based Approach to pepper leaf blight disease Detection and Multi-Classification," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-5, doi: 10.1109/INCET57972.2023.10170692.]

[13]. T. Singh and S. Karanchery, "Universal Image Segmentation Technique for Cancer Detection in Medical Images," 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 2019, pp. 1-7, doi:

10.1109/ICCCNT45670.2019.8944598.

[14]. Alruwaili, M.; Siddiqi, M.H.; Khan, A.; Azad, M.; Khan, A.; Alanazi, S. RTF-RCNN: An Architecture for Real-Time Tomato Plant Leaf Diseases Detection in Video Streaming Using Faster-RCNN. Bioengineering 2022, 9, 565. https://doi.org/10.3390/bioengineering9100565

[15]. Lawrence C. Ngugi, Moataz Abelwahab, Mohammed Abu-Zahhad, Recent advances in image processing techniques for automated leaf pest and disease recognition – A review, Information Processing in Agriculture, Volume 8, Issue 1, 2021, Pages 27-51, ISSN 2214-3173, https://doi.org/10.1016/j.inpa.2020.04.004.

[16]. Yosuke Toda, Fumio Okura. How Convolutional Neural Networks Diagnose Plant Disease. Plant Phenomics. 2019;2019: DOI:10.34133/2019/9237136

[17]. Artzai Picon, Aitor Alvarez-Gila, Maximiliam Seitz, Amaia Ortiz-Barredo, Jone Echazarra, Alexander Johannes, Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild, Computers and Electronics in Agriculture, Volume 161, 2019, Pages 280-290, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2018.04.002.

[18]. Zifei Luo, Wenzhu Yang, Yunfeng Yuan, RuruGou, Xiaonan Li, Semantic segmentation of agricultural images: A survey, Information Processing in Agriculture, 2023, ISSN 2214-3173, https://doi.org/10.1016/j.inpa.2023.02.001.

[19]. Altuncu, Mehmet Ali & Guven, Taner & Becerkli, Yaşar & Sahin, Suhap. (2015). Real-Time System Implementation for Image Processing with Hardware/Software Co-design on the Xilinx Zynq Platform. International Journal of Information and Electronics Engineering. 5. 473-477. 10.7763/IJIEE. 2015.V5.582.

[20]. Seyfioğlu, Mehmet & Ozbayoglu, Murat & Gurbuz, Sevgi. (2018). Deep Convolutional Autoencoder for Radar-Based Classification of Similar Aided and Unaided Human Activities. IEEE Transactions on Aerospace and Electronic Systems. PP. 1-1. 10.1109/TAES.2018.2799758.