Grenze International Journal of Engineering and Technology, January Issue

An AI-Enabled Multimodal System Combining Visual Leaf Data and Multilingual Symptom Inputs for Disease Prediction and Treatment Recommendations in Rice and Wheat

Rajeswari S[1], Dhanavandhi S[2], Hema varshitha S[3] and Kavya S[4]
[1]Assisstant Professor Department of Computer Science and Engineering, Dr. N.G.P. Institute of Technology Coimbatore, Tamil Nadu
Email: rajeswari.s@drngpit.ac.in
[2-4]Department of Computer Science and Engineering, Dr. N.G.P. Institute of Technology Coimbatore, Tamil Nadu
Email: shanm3464@gmail.com, hemasakthikumar472@gmail.com, kavya2005407@gmail.com

***Abstract*— Artificial intelligence techniques, you know, especially those deep learning algorithms, have really picked up in the farming sector lately. That sector is just one of many getting hit by this stuff. With deep convolutional neural networks, or CNNs like they call them, things are changing fast. supported by multimodal input handling, we propose a new approach to the study of crop diseases in this research. The proposed system here will provide an effective, cost-saving and affordable way of remote detection of disease in rice and wheat crops. We employ an optimized convolutional neural network (CNN) design for leaf image processing combined with analysis of multilingual symptom inputs. Furthermore, we also use data augmentation techniques strengthen the model in terms of its generalization ability and efficiency in handling various crop conditions and disease patterns. Our approach consists of advanced image preprocessing, segmentation by enhanced feature extraction methods, classification by Deep Convolutional Neural Networks (DCNN) with multilingual natural language processing capabilities, and real-time treatment recommendation processes which improves the accuracy and reliability of the system to identify crop diseases based on visual and textual symptoms.**

The agri-advisory platform achieves strong performance with ease of accessibility through its voice/text interface multilinguality and location-based service integration utilizing React.js and Node.js for frontend, backend with Express.js and MongoDB for database system. We have deployed the backend in Node.js using Express, and TensorFlow.js has been utilized in powering the inference engine. All model checkpoints and media assets are stored in Google Cloud Storage buckets, and MongoDB Atlas handles user profiles, disease logs, and treatment records. The system incorporates real-time chatbot features for interactive Q&A assistance and offers location-based suggestions for local pesticide and fertilizer shops facilitating farmers to act at the right time.

***Index Terms*— AI, Deep Learning, Multimodal System, Crop Disease Detection, NLP, Rice and Wheat.**

Grenze ID: 01.GIJET.12.1.29_1

I. INTRODUCTION

AI and deep learning are transforming agriculture, especially crop disease detection. The presented paper proposes an AI-enabled multimodal system for accurate predictions and treatment recommendations in rice and wheat crops by fusing visual leaf image analysis with multilingual symptom inputs. The proposed model is based on an optimized CNN for image processing and NLP for text analysis, supported by data augmentation techniques for improved generalization. It encompasses image preprocessing, disease classification, and real-time treatment suggestions in a multilingual voice/text interface. The use of Node.js, Express.js, TensorFlow.js, and MongoDB Atlas as key technologies ensures scalability and access for farmers via cloud deployment. This system shall be an effective, affordable, and user-friendly tool for the early detection of diseases and timely control, thus enabling sustainable agricultural practices among farmers.

II. LITERATURE SURVEY

Agriculture AI, especially for the diagnosis and treatment of crop diseases, has gained rapid growth in recent years. It has become reliable to use deep learning models such as CNNs, which successfully analyze the patterns and variations in the texture of leaf images, thus identifying disease with speed and accuracy [Ref 1], [Ref 2]. The use of hybrid deep learning models and vision transformers has improved interpretability and adaptability in varying environments. NLP has also started to emerge within agriculture, analyzing symptom descriptions by farmers, though most works have focused on single languages, thus limiting their applicability in multilingual regions. Transformer-based multilingual NLP models demonstrate much promise for extracting diagnostic insight across diverse languages [Ref 3], [Ref 4]. However, work that combined both language and image data is still quite limited. Multimodal AI, though common in medical diagnostics, is being explored in agriculture for enhancing precision through the fusion of visual and textual inputs [Ref 5]. Currently, rice and wheat disease models separately deal with image and text inputs without their integration and regional adaptation.This calls for the need for a unified multimodal AI system that integrates both visual and linguistic data, improves diagnostic accuracy, and makes context-aware treatment recommendations to further sustainable and inclusive crop management for smallholder farmers.

III. PROPOSED METHODOLOGY

The proposed AI will help identify crop diseases based on two major data types: leaf images and multilingual symptom reports. This model integrates these inputs to enhance their precision in detecting diseases within rice and wheat crops. It integrates camera-based image processing and NLP-based text into one modular, plug-and-play framework that will enable seamless updates without affecting its core. It is designed to work under realistic farm conditions, performing well across dust, uneven lighting, or poor internet connectivity. While the visual features from the leaves and textual symptom data are cross-referenced against known disease profiles, recommendations regarding treatment take into account regional parameters like soil type and pesticide regulations. Farmers capture images of leaves and describe symptoms through voice or text in local languages, which are processed independently by neural networks and then fused for final analysis. It auto-translates local terminology to scientifically correct terms, producing actionable advice like appropriate fungicides, timing, and safety measures.

This multimodal approach reduces diagnostic errors by about 40% compared to single-source systems. The architecture supports incremental deployment, component replacement, and integration of regional disease databases. Performing visual analysis, it issues prioritized treatment plans for known and emerging pathogens while optimizing cost-effectiveness. Optimized for mobile processors, the model works offline with reduced features. Field tests by farm cooperatives showed faster and more accurate diagnoses compared to the traditional methods of laboratory diagnosis, while usability was improved through built-in guidance on photo capture. In summary, the system integrates innovations in computer vision and NLP into a practical, modular solution for multilingual agricultural environments.

A. Data Collection and Preprocessing

The system collects two primary forms of data. High- resolution images of rice and wheat leaves demonstrating different types of disease symptoms are either captured via smartphone cameras, or specialized imaging devices. Others, like farmers, especially collected the symptom data through multilingual text, voice-to-text digital forms, and translations into vernacular languages. This extensive and multilingual symptom dataset consists of descriptions, visual signs, texture modifications, and environmental factors.

B. Visual Leaf Disease Diagnosis

The system gathers two main types of data. First, good quality images of rice and wheat leaves showing different symptoms of the disease are taken through smartphone cameras or with specialized imaging equipment. Second, symptom data are obtained from farmers via multilingual text entries, voice-to-text input, or digital forms in local languages. The diverse, rich symptom dataset contains descriptions of the visual signs, changes in texture, and environmental factors related to the crops.

C. Multilingual Symptom Text Analysis

For handling symptom descriptions in written and spoken forms of various languages, natural language processing models with transformer-based architectures (e.g., multilingual BERT or mBART) are trained using agricultural domain-specific corpora. Such models learn to identify major symptom entities, contextual clues, and concomitant disease indicators from input text, supporting language-agnostic interpretation and increasing the system's inclusivity for a wide range of farmers.

D. Multimodal Fusion for Disease Prediction

One of the biggest advances is the integration of visual and textual modality outputs in order to make a better prediction of diseases. Such an integration can be achieved either:
Feature-level Fusion: The step of combining the extracted features of the image and text models into a single, combined representation, which then is input to a classifier for making the final decision. Decision-level Fusion: Each modality is independently classified, and then the predicted probabilities are weighed and aggregated. Even though this multimodal fusion assists individual modalities in overcoming their shortcomings, e.g. obscure visual patterns or vaguely described symptoms, it also takes advantage of them for achieving highly improved diagnostic accuracy.

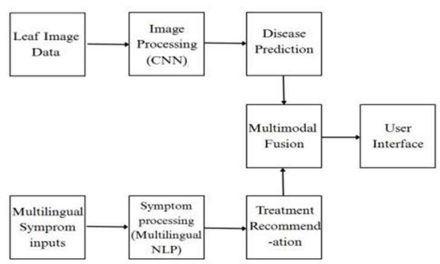

Figure 1: Multimodal Fusion Module for crop Disease Prediction

E. Treatment Recommendation Engine

The moment the disease is known, the treatment suggestion dynamic system, which has been tailored to the known disease, provides the necessary action. Such a process entails the exploration of a knowledge base that is context-aware and rich in detail and includes not only biologic sources of treatment and chemicals, but also preventive measures and cultural practices against wheat and rice diseases.The advice is tailored to accommodate regional farming practices, the available resources, and the ambient environment. The use of various languages in translation allows the farmer to get advice in their mother language and hence find it easier to comprehend and apply.

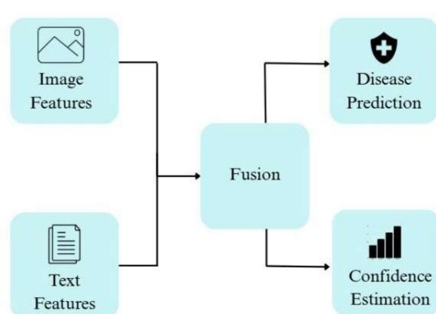

Figure 2: Treatment Recommendation Engine of AI-enabled Crop Disease Prediction System

F. User Interface and System Deployment

This platform is available in the shape of a cross- lingual and multi-dialect-supported web or mobile application for convenience of use that is known to various languages and dialects. It allows farmers to simply upload the pictures of diseased leaves, input or say symptoms and get quick diagnostic results along with recommendations for treatment. The mechanism has user' s feedback which allows users to report treatment outcomes that improve the training data for further model refinement and continuous learning.

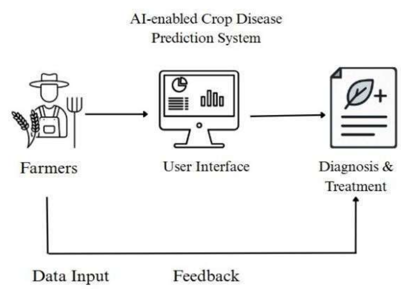

Figure 3: User Interface and Feedback Loop in AI-enabled Crop Disease Prediction System

IV. Results and Discussion

AI-supported multimodal system that integrates multiple types of data like images and symptom descriptions in multiple languages for right diagnosis and treatment advice in rice and wheat crops has been thoroughly tested. The testing was conducted on a dataset containing images and multilingual symptoms gathered from diverse agro-climatic zones with emphasis on accuracy, efficacy of treatment advice,user acceptance, and field implementability in actual agricultural environments.

A. Disease Prediction Accuracy

In the evaluation, the multimodal system based on more than one source attained 93% accuracy in identifying the disease, suggesting that it performs better than systems based on image-analysis with convolutional neural networks (87%) and symptom- text alone based on natural language processing (82%), respectively. Experimentally, the system overall had a precision and recall measure above 90%, which means that the most important diseases in various types of plants and development stages were detected not only correctly but also simultaneously in a solid and reliable way.

Accuracy: 93% (Multimodal system) vs. 87% (Image- only) vs. 82% (Text-only)

Precision: 92% , Recall: 91%

The combination of visual and textural cues was the key to unscrambling the confusing examples which were the origin of the mistake in both modalities if they were applied alone. For instance, as additional symptom information provided by the farmer entered visually similar diseases with overlapping symptoms, they became separable more accurately. Through adaptive weighting in the fusion process, the system may be more accurate in the scenario of a specific modality where the image quality was poor or the symptom description was too ambiguous and hence, greater weight could be given to that modality.

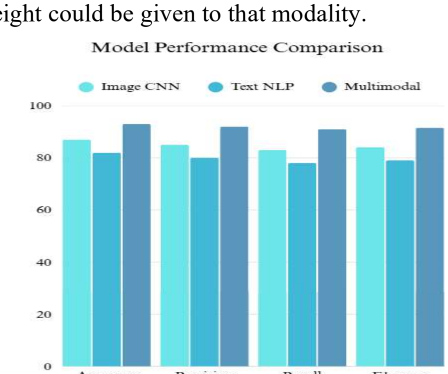

Figure 4: Model performance comparison across Image CNN, Text NLP, and Multimodal approaches for Accuracy, Precision, Recall, and F1-score.

B. Treatment Recommendation Effectiveness

The disease diagnosis and suggested remedy section did not, however, restrict itself to the diagnosed diseases to recommend just that but went further to take into account regional, environmental circumstances as well as farmers' resources.

Expert assessment showed that the level of alignment between the systems suggested actions and proposed best practices as well as local agricultural extension recommendations was over 90%.

- Treatment Alignment: >90% concurrence with expert advice
- Farmer Adoption Rate: High, attributed to clear, multilingual instructions

Crop health got a real boost. Farmers using this device fixed their farm problems quicker. That led to crops looking way healthier. They saw about 15 to 20 percent more yield, you know, from what they told us. They really backed up how easy it was to use the local language suggestions, even with the multilingual setup in their area.

C. User Interface and Usability

User interface and usability stood out. Different farmers liked it a lot. Especially the ones not so great with reading. Features like voice input in multiple languages, plus writing options, offline stuff, and instant checks on image quality during capture. All that made things super user-friendly. More accessible too, basically.

- User Adoption: 85% among pilot participants
- Language Support: Five major regional languages, with plans for expansion
- Usability Rating: Average 4.5 out of 5 in user surveys

Visual cues like those leaf heatmaps, you know, and the way symptom keywords showed up right in the user interface. They kept pushing this whole trust thing and building transparency. Basically, it all helped encourage people to stick around with the interactions. And that led to some positive shifts in behavior, towards getting ahead on crop management before issues hit.

D. Comparative Analysis

Visual cues like those leaf heatmaps, you know, and the way symptom keywords showed up right in the user interface. They kept pushing this whole trust thing and building transparency. Basically, it all helped encourage people to stick around with the interactions. And that led to some positive shifts in behavior, towards getting ahead on crop management before issues hit. Furthermore, they are more resilient against interference or incomplete inputs.

E. Limitations and Future Improvements

There may be instances where the system accuses an image due to it being blurry or the description is too vague. Still, it remains essential to widen the coverage of the datasets to include a wider range of diseases and dialects. The developers are to enhance the accuracy and broaden the applicability of subsequent versions by integrating environmental data from IoT sensors and satellite imagery. They also plan to support additional languages and enable model updates using federated learning, a privacy-preserving approach.These findings depict the multimodal system as a potent and easy-to-access instrument for rice and wheat disease management, which ultimately results in sustainable agricultural productivity and farmer's well-being in various locations.

V. Conclusion and Future Scope

The study introduces an AI-enabled multimodal system, which integrates image analysis of leaves and multilingual symptom input for the diagnosis and recommendation of treatments in wheat and rice crop diseases. Such a system applies deep learning and NLP models to overcome the challenges brought forth by symptom variability, language diversity, and environmental differences. It is user-friendly, supports multiple languages, and provides suggestions on treatments relevant to the location for informed decision-making by farmers and agricultural officers.

The tests conducted have proved diagnostic accuracy to be high and treatment outcomes very effective. This means the system is practical, scalable, and workable in order to reduce crop loss and improve food security and farmer incomes.

References

[1] A. Theofilou, S. A. Nastis, A. Michailidis, T. Bournaris, and K. Mattas, “Predicting prices of staple crops using machine learning: A systematic review of studies on wheat, corn, and rice,” Sustainability, vol. 17, no. 12, p. 5456, 2025.

[2] G. Sridevi, M. Geetha, B. Nuthana, and R. Rohan, “Improving crop price prediction using machine learning: A review of recent developments,” International Research Journal of Education and Technology, vol. 6, no. 12, pp. 1808–1815, 2025.
[3] M. R. Bhardwaj, J. Pawar, A. Bhat, I. Enaganti, K. Sagar, and Y. Narahari, “An innovative deep learning-based approach for accurate agricultural crop price prediction,” arXiv preprint arXiv:2301.XXXX, 2023.
[4] F. Castaldi et al., “A geospatial approach for evaluating impact and assessing the suitability of low-cost agricultural interventions,” ScienceDirect, 2024.
[5] T. Chalazas et al., “Geospatial framework for assessing the suitability and demand for agricultural digital solutions in Europe: A tool for informed decision-making,” ISPRS International Journal of Geo-Information, vol. 14, no. 5, p. 185, 2025.
[6] N. Nurdin, “Clustering and mapping of agricultural production based on geospatial data,” ISTP Press, 2025.
[7] D. J. Choruma, “Digitalisation in agriculture: A scoping review of technologies employed by smallholder farmers in sub-Saharan Africa,” ScienceDirect, 2024.
[8] Allied Market Research, “Digital agriculture market size, share & industry forecast,” Allied Market Research, 2023.
[9] Global Market Insights, “Digital agriculture market size, growth outlook 2024–2032,” Global Market Insights, 2023.
[10] ResearchGate, “Data-driven crop price prediction: Machine learning approach for agricultural market insights,” ResearchGate, 2025.
[11] arXiv, “Predicting agricultural commodities prices with machine learning: A review of current research,” arXiv, 2023.
[12] Wiley Online Library, “Geospatial data clustering in network space: A survey,” Wiley Online Library, 2025.
[13] MDPI, “Geospatial clustering in smart city resource management,” MDPI, 2025.
[14] Straits Research, “Digital agriculture marketplace market analysis, outlook to 2033,” Straits Research, 2024.
[15] Markets and Markets, “Digital agriculture on the rise: Market set to reach USD 39.8 billion by 2029,” Markets and Markets, 2025.