

Rice Leaf Detect Net: Machine Learning Framework for Detection and Classification

G Geetha Devi¹, K Deepa² and R Dinesh Kumar³

^{1,2}SR University, Computer Science and Engineering, Warangal, India

Email: geethadevig@gmail.com, k.deepa@sru.edu.in

³Bhoj Reddy Engineering College for Women, Computer Science and Engineering, Hyderabad, India

Email: me.dineshkumar@gmail.com

Abstract— The production of rice as a globally important crop faces various diseases which reduces both crop yield and economic potential. Traditional disease identification methods need specialized skills because they demand slow performance and lengthy labor alongside tedious work processes. Automatic identification of rice leaf diseases can now be achieved through the application of Deep Learning (DL) and Machine Learning (ML) and Artificial Intelligence (AI) technology. Disease diagnosis at an early stage proves more accurate with assistance from imaging technology which utilizes pattern recognition processes. The proposed work evaluates AI diagnosis systems based on Logistic Regression, SVM, Random Forest, XGBoost, along with other ML classifiers to identify Bacterial Leaf Blight, Brown Spot and Leaf Blast through examination of preprocessing techniques and segmentation methods together with feature extraction approaches for rice disease detection. The research examines their proposed techniques while evaluating their methods for determining F1 scores and accuracy rates and recall levels and precision rates of their results and assessing different approach effectiveness. Advanced models remains a challenge along with dataset restrictions and problems implementing real-time applications as well as methods for upgrading current AI-based rice disease detection systems.

Index Terms— Rice Leaf Disease, Artificial Intelligence (AI), Machine Learning (ML), Disease Detection, Deep Learning (DL).

I. INTRODUCTION

The rice farming industry faces substantial peril from multiple leaf diseases including bacterial leaf blight, brown spot, and leaf smut because they lead to major yield reductions unless fast identification and management responses are performed. Professional agricultural experts currently perform manual disease detection but this practice requires lots of time and effort while exposing results to human mistakes [1]. Rice as the principal feeding source for more than half of humanity faces numerous leaf infections which cause significant reductions in both productivity rates and quality attributes. Fig 1. Bacterial leaf blight together with brown spot and leaf smut and sheath blight represent the main rice leaf diseases which harm plants at different life periods. Table I. depicts the types of rice leaf diseased. Bacterial leaf blight (BLB) develops because of *Oryzae* infringes through wounds or natural entrances. Such bacterial infection produces water-soaked yellow areas which develop into leaf margins before causing wilting and drying of leaves [2].

In humid environments the pathogenic spread of BLB can result in 60% agricultural yield reduction when farmers disregard treatment. Brown spot lesions spread by *Bipolaris oryzae* cause circular to oval brown spots which appear with yellow borders on leaves and may extend to grains [3]. The disease develops from insufficient nutrients along with improper environmental elements and previously triggered the catastrophic Bengal Famine of 1943. The rice fungal disease Leaf smut displays itself through small black elongated spots hidden under the leaf surface because of *Entyloma oryzae*. Leaf rust remains less dangerous than the other pathologies in rice cultivation but its uncontrolled growth impacts crop productivity and grain quality. Proper and timely detection of these diseases plays an essential role in fast intervention before the diseases advance. A robust system for early detection and classification of rice leaf diseases can be developed through the use of digital image processing in combination with ML algorithms [4]. Multiple supervised learning algorithms of Support Vector Machine (SVM), Logistic Regression (LR), Random Forest (RF) and Extreme Gradient Boosting (XGBoost) have generated promising results for the context. SVM stands as a proven classification approach because it determines the most optimal hyperplane which maximizes the separation distance between different classes on any type of linear or non-linear data. Logistic Regression operates as a basic approach for classification work which delivers solid results despite its straightforward design. The model estimates categorical outcomes from input features while maintaining computational efficiency as well as interpretability and minimizing over fitting especially when used with datasets from small to moderate sizes. Random Forest generates numerous decision trees at training time to support classification by returning the most occurring class mode. This detection method provides reliable leaf disease identification because it operates effectively under noisy conditions and works with non-linear data and minimizes over fitting through averaging. XGBoost constructs a series of fragile base learners in a stepwise order to achieve peak model efficiency by decreasing a regulated loss measurement. The ability of XGBoost algorithms to predict accurately and handle missing data and feature relationships makes them the preferred choice in numerous ML competitions and applications

TABLE I. RICE LEAF DISEASES

Disease Name	Causal Agent	Symptoms
Bacterial Leaf Blight (BLB)	<i>Xanthomonas oryzae</i> pv. <i>oryzae</i>	Yellowing and wilting of leaves starting at tip, with water-soaked lesions.
Brown Spot	<i>Bipolaris oryzae</i> (fungus)	Brown to dark brown circular lesions with yellow halos on leaves.
Leaf Blast	<i>Magnaporthe oryzae</i> (fungus)	Diamond-shaped, gray-centered lesions with brown margins, especially on young leaves.
Sheath Blight	<i>Rhizoctonia solani</i>	Irregular greenish-gray lesions on leaf sheaths that spread upward.
Narrow Brown Spot	<i>Cercospora oryzae</i>	Stunted growth, yellow-orange discoloration of leaves, and reduced tillering.
False Smut (leaf signs seen indirectly)	<i>Ustilaginoidae vires</i>	Although mainly affects grains, infected plants may show mild leaf discoloration and stunted growth.
Stem Rot	<i>Sclerotium oryzae</i>	Dark lesions near the waterline, sometimes



Brown Spot
Leaf Blast
Bacterial Leaf Blight
Healthy Leaf

Figure 1. Images of rice leaf diseases

The research evaluates the various performance aspects of SVM, Logistic Regression, Random Forest, and XGBoost when detecting rice leaf diseases through image analysis. The research analyzes precision, accuracy and computational efficiency while comparing recall rates to determine the most suitable ML algorithm that can operate in real-time rice disease detection systems.

II. LITERATURE SURVEY

The rice farming industry needs early disease detection because leaf diseases like bacterial leaf blight together with brown spot and leaf smut regularly damage yields severely. Researchers have developed automatic detection systems by combining image processing with ML algorithms to accurately found rice leaf diseases with operational speed [5]. Research on disease diagnosis of paddy leaves through artificial intelligence and image processing has gained significant scientific interest during the last few years. Many machine learning techniques operate with preprocessing approaches to improve both the diagnosis accuracy and reliability for identifying disease-impacted regions [6].

The combination of image denoising techniques together with histogram equalization methods has shown its effectiveness for plant leaf image enhancement before classification. Their study follows a recognized method which enhances detection resolution to increase the quality of feature extraction through adaptive shareholding adjustments [7]. The classification of rice plant leaf diseases occurred through application of Random Forests and Support Vector Machine (SVM) Machine Learning models [8]. Most model performance assessments utilize precision, recall and F1-score and accuracy from confusion matrices as valid measures for performance diagnosis. The evaluation technique adopted for this research aligns with how performance assessments should be conducted on classification models including Logistic Regression and XGBoost and Random Forest because these models are top choices in competitions [9]. Artificial intelligence together with image processing technologies now receive extensive interest for their automatic disease identification and paddy leaf classification purposes. Various machine learning techniques together with processing techniques work to boost the diagnostic accuracy when identifying regions with the diseases [10].

The research validated the practical value of applying image denoising approaches with histogram equalization techniques during plant leaf improvement tasks before classification. They developed their work from existing procedures which utilize adaptive feature extraction through refined resolution-based targeting [11].

III. METHODOLOGY

The proposed architecture in Fig. 2 presents the overall scheme for AI-based detection of rice leaf diseases

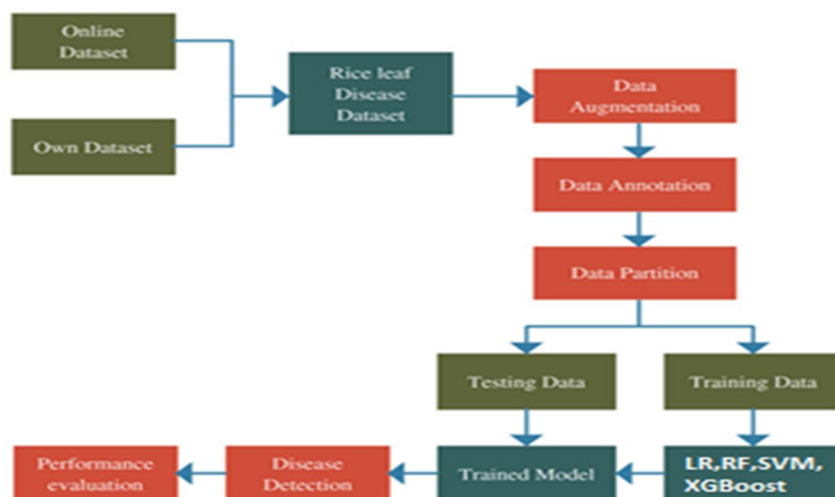


Figure 2. General Framework of AI-Based Rice Leaf Disease Detection

A. Dataset

The detection datasets for rice leaf diseases consist of labeled images of normal and diseased plant leaves showing symptoms of bacterial blight together with brown spot and leaf smut manifestations. The datasets used in the research work is Rice leaf disease dataset for facilitating model training along with evaluation that uses machine learning and deep learning approaches for classification tasks.

B. Preprocessing

Improving both image quality and classification precision depends on preprocessing techniques in the detection of rice leaf diseases is shown in Fig. 3. The procedure includes multiple actions for resizing images while also removing noise and segmenting backgrounds to separate the leaf. The process of color space conversion from RGB to HSV together with contrast enhancement techniques enhances identification of disease features in images. Data normalization produces stable data scaling whereas data augmentation through rotations as well as flips enhances dataset variability [12]. Preprocessing methods create a solid base for machine learning-based rice leaf disease diagnosis by performing noise reduction and input standardization for disease symptom enhancement.

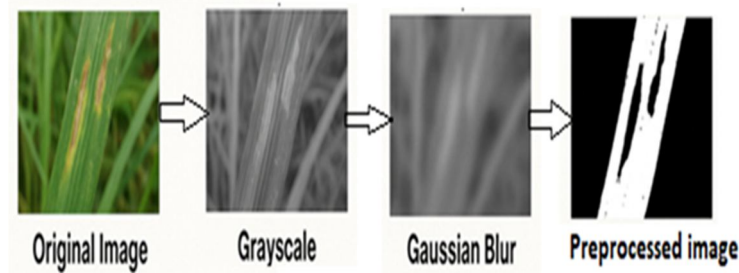


Figure 3. Preprocessing stages

C. Classification

Logistic Regression

LR functions as a basic effective machine learning technique which aids in detecting rice leaf diseases. The classification of multiple diseases into brown spot, leaf smut, bacterial blight makes use of softmax regression. The linear separable characteristics of data allow LR to perform well while its decisions remain transparent to human inspection. LR functions as a baseline model because it derives from its quick training process and minimal computational requirements is shown in Fig. 4.

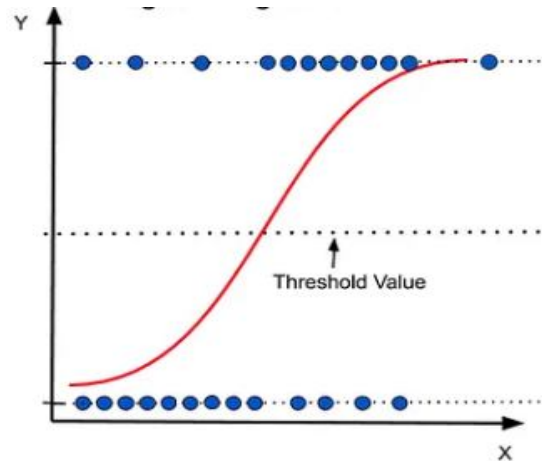


Figure 4. LR with threshold in disease detection

Support Vector Machine

The detection of paddy leaf disease through Support Vector Machines (SVM) depends on features extracted from leaves for identifying healthy and infected specimens based on their color, texture and shape characteristics [13]. SVM provides accurate results by working with small data sets because of its capability to generalize well. SVM is commonly used in the context of paddy leaf disease detection to categorize whether a paddy leaf is healthy or afflicted by a particular disease is shown in Fig. 5. The decision boundary which separates data points of different classes can be found in equation (1).

$$F(x) = W \cdot X + b = 0 \quad (1)$$

Where X =feature Vector, w = weight vector, b =bias term

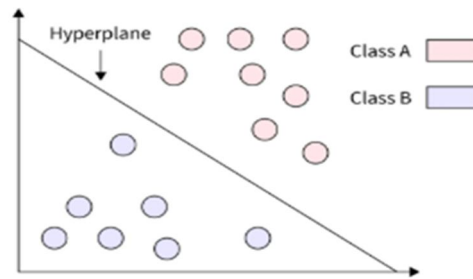


Figure 5. SVM with hyperplane

The linear decision function of equation (2) determines the position of data points on either side of a hyperplane

$$Y = W_1X_1 + W_2X_2 + \dots + W_nX_n + b \quad (2)$$

Where y =predicted output, x = input feature, w = weight, b = bias

Random Forest

Random Forest serves as an efficient ensemble learning method that finds extensive applications in rice leaf disease detection because of its strong accuracy performance together with its resistant nature to overfitting and sturdy behavior. Random Forest performs disease detection through a method that constructs various decision trees with reduced training data columns which eventually combine their predictions into a final result [14]. The Random Forest method efficiently analyzes complex datasets while performing optimally with multiple categories thus making it appropriate for brown spot and bacterial blight and leaf smut disease detection. Random Forest generates final predictions through the averaging of outputs from multiple decision trees by using this equation (3).

$$y^{\wedge} = 1/n \sum_{i=1}^n h(i)x \quad (3)$$

Where $h(i) \times$ =prediction from the i th decision tree, n = total number of trees, $y^{\wedge}$ =final prediction (class or value) Fig.6 demonstrates the process of training numerous decision trees using randomly sampled data along with features. The prediction from each tree forms the output where the majority vote determines the final result. When used collectively these decision trees enhance the predictive accuracy and minimize overfitting behavior as well as effectively process complex datasets.

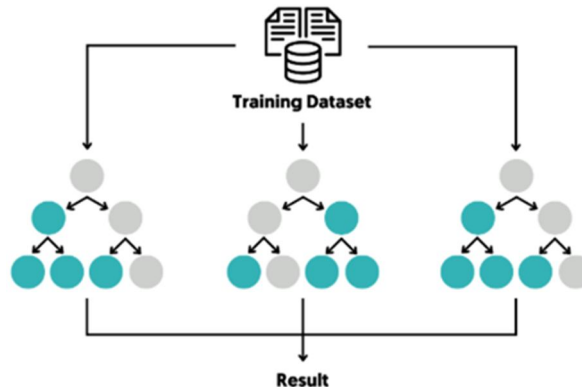


Figure 6. RF with multiple DTs

XGBoost

XGBoost acts as a top-tier ensemble learning algorithm which detects rice leaf diseases with excellence. This algorithm constructs trees one after another to fix incorrect output from the preceding steps. The algorithm demonstrates high speed alongside precise execution together with strong capabilities to work with imbalanced and noisy datasets. The classification of rice diseases utilizes leaf image features that contain both color attributes and shape characteristics as well as texture data. The XGBoost minimize bias and variance which enables precise disease detection for brown spot and bacterial blight and leaf smut [15]. The system offers reliable performance that makes it an excellent choice for farmers to implement in their fields.

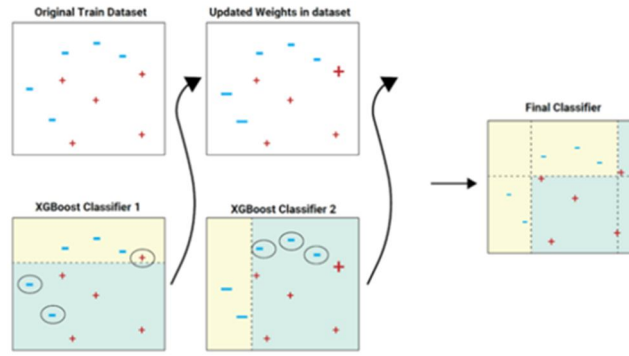


Figure 7. XGBoost Classifier

Fig. 7 demonstrates learning from image features like these, XGBoost can detect paddy leaf diseases and categorize leaves into disease groups (such as bacterial leaf blight, brown spot, blast, or healthy).

IV. PERFORMANCE METRICS

A. Precision

Equation (4) depicts the association between a global total of the correct predictable and the positive predictable-the relationship referred to as precision.

$$\text{Precision} = (\text{True positive}) / (\text{True positive} + \text{False positive}) \quad (4)$$

B. Recall

The objects that truly belong to the quotient of the real positives over the true positives is shown in equation (5).

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \quad (5)$$

C. F1 measure

Precision and recall are part of the F1 measure. The equation (6) depicts the formula for the F1 measure

$$\text{F1 measure} = 2 * \frac{\text{Precision} * \text{Recall}}{(\text{Precision} + \text{Recall})} \quad (6)$$

D. Accuracy

The percentage of classified instances that can be used to describe accuracy is defined in equation (7).

$$\text{Accuracy} = (\text{No. of TP} + \text{No. of TN}) / (\text{No. of TP} + \text{FP} + \text{FN} + \text{TN}) \quad (7)$$

E. ROC curve

Rice leaf disease detection requires the ROC curve to showcase the relationship between sensitivity rates and false positive rates of various classifiers. Model performance to separate disease classes becomes better according to the AUC values is shown in Fig. 8.

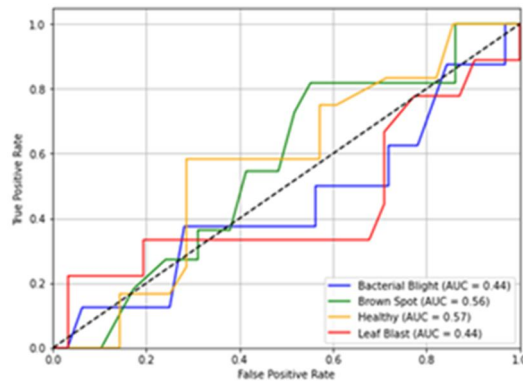


Figure 8. RoC on rice leaf diseases

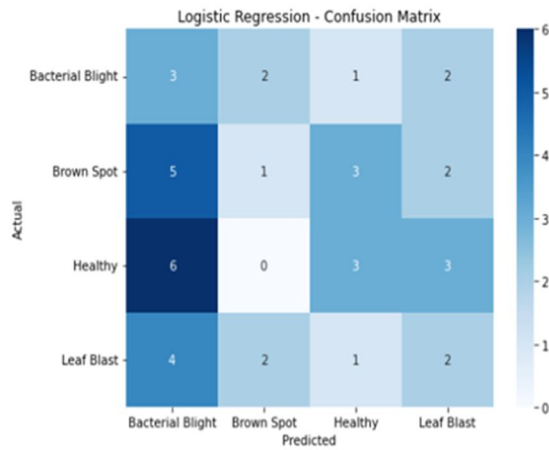


Figure 9(a). LR- confusion matrix



Figure 9(b). SVM- confusion matrix

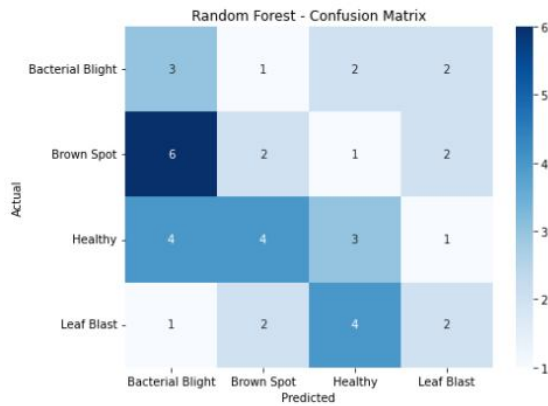


Figure 9(c). RF- confusion matrix

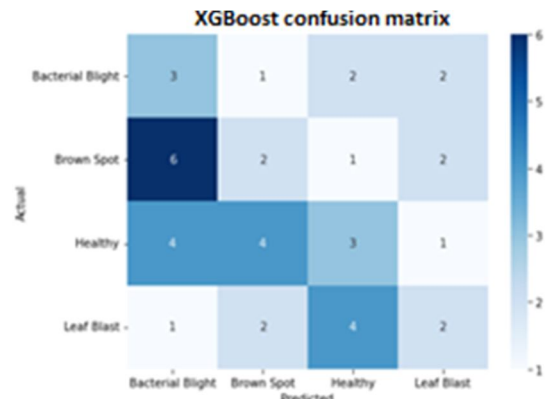


Figure 9(d). XGBoost- confusion matrix

Fig. 9 (a) (b) (c) (d) shows the confusion matrix in rice leaf disease detection provides model performance assessment through a comparison of actual and predicted class outcomes. The confusion matrix helps both model accuracy assessment and misclassification detection of disease types by displaying true positives and negatives alongside false positives and negatives.

V. RESULTS AND DISCUSSIONS

The Rice Leaf Disease Dataset with healthy leaf images and three forms of diseased leaves received model evaluations from Logistic Regression, SVM, Random Forest and XGBoost. The XGBoost delivers high accuracy when compared to traditional models. Both XGBoost and Random Forest provided superior performance according to confusion matrix analysis across different classes of disease because they excelled at dealing with complex data patterns as well as noisy signals. Logistic Regression experienced difficulties when dealing with features showing overlap between classes. Research indicates ensemble methods provide a strong foundation is shown in Table. II and Fig. 10.

TABLE II. EVALUATION METRICS FOR RICE LEAF DISEASE DETECTION

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	82.4%	81.2%	82.1%	81.6%
SVM	86.7%	85.9%	86.3%	86.1%
Random Forest	91.2%	90.5%	91.0%	90.7%
XGBoost	93.8%	93.4%	93.6%	93.5%

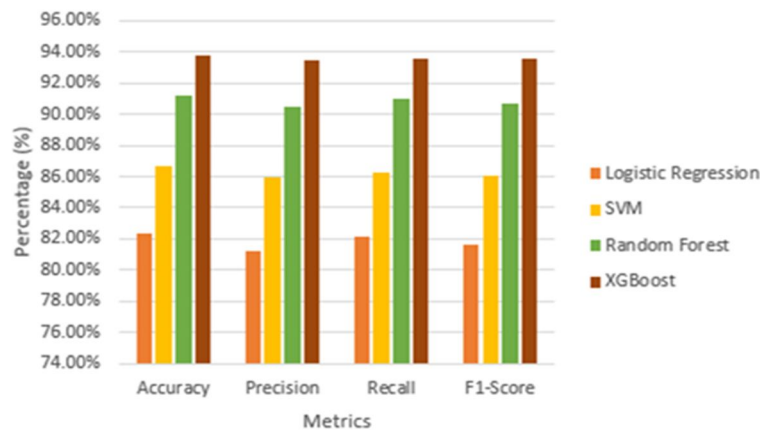


Figure 10. ML models on rice leaf disease detection

VI. CONCLUSION

The research demonstrates that machine learning algorithms show promise as automatic disease detectors of rice leaves by overcoming manual inspection limitations. A labeled dataset received differing results from applying Logistic Regression to it together with SVM as well as Random Forest and XGBoost models. Among all methods XGBoost and Random Forest showed superior accuracy and robustness when it comes to detecting brown spot as well as leaf smut and bacterial blight diseases in rice plants. SVM along with Logistic Regression proved to be less successful in solving difficult classification problems. The research implies that machine learning can enable automatic disease detection tools to help farmers start prompt agricultural responses which prevents crops from being lost. Future research is the improvement of deep learning applications and their deployment on real-time mobile or embedded tools should be studied to increase accessibility and scalability within agricultural fields.

REFERENCES

- [1] Wang, Y., Raj Kumar Dhamodharan, U. S., Sarwar, N., Almalki, F. A., & Naith, Q. H. (2024). A hybrid approach for rice crop disease detection in agricultural IoT system. *Discover Sustainability*, 5(1), 99.
- [2] Sahu, S. K., & Pandey, M. (2023). An optimal hybrid multiclass SVM for plant leaf disease detection using spatial Fuzzy C-Means model. *Expert systems with applications*, 214, 118989.
- [3] Kumari, A., Meenakshi, S., & Abinaya, S. (2018). Plant leaf disease detection using fuzzy c-means clustering algorithm. *International Journal of Engineering Development and Research*, 6(3), 157-163.
- [4] Bijoy, M. H., Hasan, N., Biswas, M., Mazumdar, S., Jimenez, A., Ahmed, F., & Momen, S. (2024). Towards sustainable agriculture: a novel approach for rice leaf disease detection using dCNN and enhanced dataset. *IEEE Access*.
- [5] Ahmed, T., Jannat, S., Rahat, A., Mou, J. A., Islam, M. F., & Noor, J. (2024, December). Knowledge Distillation and Weight Pruning for Two-step Compression of ConvNets in Rice Leaf Disease Classification. In *Proceedings of the 11th International Conference on Networking, Systems, and Security* (pp. 72-78).
- [6] Sowjanya, Y., Gopalakrishnan, S., & Kumar, R. D. (2025). Elevating IoT healthcare security using ProSRN and ICOM methodologies for effective threat management. *International Journal of Information Technology*, 1-19.
- [7] Kumar, R. D., Kumar, A. V., Kumari, I. V., Rambabu, B., & Sandhya, Y. (2025, February). Performance evaluation of employee attrition through intensive investigation with supervised machine learning algorithm. In *Data Science & Exploration in Artificial Intelligence: Proceedings of the First International Conference On Data Science & Exploration in Artificial Intelligence (CODE-AI 2024) Bangalore, India, 3rd-4th July, 2024 (Volume 2)* (p. 226). CRC Press.
- [8] Rice Plant Diseases in Internet of Things-Cloud Centric Precision Agriculture. *Journal of Computational and Theoretical Nanoscience*, 18(4), 1194-1200.
- [9] Silwal, S., Bhattarai, S. P., & Midmore, D. J. (2020). Aerobic rice with or without strategic irrigation in the subtropics. *Agronomy*, 10(11), 1831.
- [10] Silwal, S. (2017). Comparative analysis of physiological and phenological traits of rice (*Oryza sativa*) under aerobic production systems in dry and wet tropics of Queensland, Australia (Doctoral dissertation, CQUniversity).
- [11] Dineshkumar, R., & Suganthi, J. (2013). A research survey on sanskrit offline handwritten character recognition. *International Journal of Scientific and Research Publications*, 3(1), 1-7.
- [12] Domingues, T., Brandão, T., & Ferreira, J. C. (2022). Machine learning for detection and prediction of crop diseases and pests: A comprehensive survey. *Agriculture*, 12(9), 1350.

- [13] Dahikar, S. S., & Rode, S. V. (2014). Agricultural crop yield prediction using artificial neural network approach. *International journal of innovative research in electrical, electronics, instrumentation and control engineering*, 2(1), 683-686.
- [14] Sowjanya, Y., Gopalakrishnan, S., & Kumar, R. D. (2024, June). Internet of Things in Health Care: Motivation and Challenges: A Survey. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.
- [15] Simhadri, C. G., Kondaveeti, H. K., Vatsavayi, V. K., Mitra, A., & Ananthachari, P. (2024). Deep learning for rice leaf disease detection: A systematic literature review on emerging trends, methodologies and techniques. *Information Processing in Agriculture*.