Rice Blast Disease Prediction Using Integrated SMOTE With Multilayer Perceptron

V.ARUN KUMAR[1], L. PRIYANKA[2], M. PRANATHI[2], M. SHIVANI[2], K. VAHINI[2]

[1]Assistant Professor, Department of Information Technology, Mallareddy Engineering College for Women, (UGC-Autonomous), Hyderabad, India, aruncse006@gmail.com.

[2]Student, Department of Information Technology, Mallareddy Engineering College for Women, (UGC-Autonomous), Hyderabad, India.

Abstract

In India, rice blast is a major concern due to favourable weather conditions during the crop season. Climate plays major role in the disease appearance, multiplication, and spread of the fungus. Along with climatic factors, the varieties of seeds also influence the occurrence of rice blasts, primarily the climate factors have a strong influence on the occurrence of blast disease even though a sufficient amount of nutrients are present in the plant. Thus, rice blast disease will occur and develop when certain weather conditions continue for the given period. Forecasting models that make predictions of possible blast disease occurrence may give important information to the producers of rice to manage the disease. Therefore, this project implements the rice blast disease prediction using data balancing technique based multilayer perceptron.

1. Introduction

Rice Blast Disease Prediction is a crucial area of research and application in the field of agriculture that focuses on the early detection and forecasting of one of the most devastating diseases affecting rice crops worldwide, known as rice blast [1]. This fungal disease, caused by the pathogen Magnaporthe oryzae, poses a significant threat to global food security, as rice is a staple crop for billions of people. Predictive models and systems have been developed to mitigate the impact of rice blast by providing farmers with early warnings and guidance for disease management. The prediction process typically involves a combination of various data sources and technologies. These may include remote sensing through satellites or drones to monitor field conditions, weather data to assess environmental factors conducive to disease development, genetic information about rice varieties to determine susceptibility [2], and historical disease outbreak data for trend analysis. Machine learning and data analytics play a pivotal role in processing and analyzing this wealth of information. These models are trained to recognize patterns and correlations between different variables, enabling them to make accurate predictions about the likelihood of rice blast outbreaks in specific regions or fields.

Early detection and prediction of rice blast are instrumental in implementing timely and targeted disease control measures. Farmers can receive alerts and recommendations on when to apply fungicides or adopt other preventive strategies to minimize crop damage and yield loss. Furthermore, this approach promotes more sustainable and environmentally friendly agricultural practices, as it allows for the precise and optimized use of resources [3]. So, Rice Blast Disease Prediction is a multifaceted and technology-driven endeavor that leverages various data sources and advanced analytical tools to anticipate and manage the threat of rice blast disease. By providing farmers with early warnings and actionable insights, it contributes to the protection of rice crops, ensuring food security for millions of people around the world while promoting sustainable farming practices.

2. Literature Survey

Varsha, M., B. Poornima, et al. [11] proposed study is to predict the severity of rice blast disease using the linear SVM model. Prediction of rice blast disease severity is divided into four classes: 0, 1, 2, and 3. Data imbalance is the most challenging problem in multi-class classification. This study has efficiently handled imbalanced data using k-means SMOTE and SMOTE oversampling techniques to balance training and testing data. Finally, cross-location and cross-year models are developed using a linear support vector machine and predict the severity of rice blast disease to the classes 0, 1, 2, 3, respectively. Cross-year and cross-location models are cross-validated using five-fold cross-validation.Sriwanna, Kittakorn, et al. [12] proposed ensemble features ranking method have a higher classification performance than the other methods. The ranking of the final features for all classifiers reveals that average visibility, amount of rainfall, hours of sun, maximum wind speed, and days of rain are the five most effective weather features for rice blast disease prediction. Moreover, various classification models achieve satisfactory performance, especially those that are combined with feature selection.O. V. Putra, N. Triangulum, et al. [13] proposed a method based on transfer learning to tackle such issues. Our method contains several steps. In the first step, the rice leaf is preprocessed. Second, due to data imbalance, balanced class weighting was employed. Third, to improve the network performance, three layers of convolution were added to the transfer learning model. The parameters in fully connected layers were optimized using bandit-based approach. In the last step, the leaf was classified into nine categories. They compare our method with the state-of-the-art (SOTA) works. Our model reaches the top in terms of accuracy with 98 % compared to the other SOTA.

Jiang, Zhencun, et al. [14] improve the Visual Geometry Group Network-16(VGG16) model based on the idea of multi-task learning and then use the pre-training model on ImageNet for transfer learning and alternating learning. The accuracy of such model is 97.22% for rice leaf diseases and 98.75% for wheat leaf diseases. Through comparative experiments, it is proved that the effects of this method are better than single-task model, reuse-model method in transfer learning, resnet50 model and densenet121 model. The experimental results show that the improved VGG16 model and multi-task transfer learning method proposed in this article can recognize rice leaf diseases and wheat leaf diseases at the same time, which provides a reliable method for recognizing leaf diseases of many plants.Bhatia, Anshul, et al. [15] implemented Extreme Learning Machine (ELM) algorithm for plant disease prediction based on a dataset collected in real time scenario namely Tomato Powdery Mildew Disease (TPMD) dataset. Since, the collected TPMD dataset was imbalanced thus; various resampling techniques namely Importance Sampling (IMPS), Synthetic Minority Over-sampling Technique (SMOTE), Random under Sampling (RUS), and Random over Sampling (ROS) have been used here for balancing the dataset before using it in the specified prediction model. ELM models have been developed for each of the balanced TPMD datasets obtained from these resampling techniques as well as for the imbalanced TPMD dataset. Ma, H.; Huang, W.; et al. [16] proposed approach incorporating both growth and environmental parameters of different crop periods could distinguish wheat powdery mildew and aphids at the regional scale. The bi-temporal growth indices and environmental factors-based SMOTE-BPNN, BPNN, and SVM models all had an overall accuracy high than 80%. Meanwhile, the SMOTE-BPNN method had the highest G-means among the three methods. These results revealed that the combination of bi-temporal crop growth and environmental parameters is essential for improving the accuracy of the crop disease and pest discriminating models. The combination of SMOTE and BPNN could effectively improve the discrimination accuracy of the minor disease or pest.

Gao, Qijuan, et al. [17] ensemble method combines different balanced data algorithms including Borderline SMOTE (BSMOTE), Adaptive Synthetic Sampling (ADSYN), SMOTE-Tomek, and SMOTE-ENN with the XGBoost model separately. The performances of the SMOTE-ENN-XGBoost model, which combined over-sampling and under-sampling algorithms with XGBoost, achieved higher predictive accuracy than the other balanced algorithms with XGBoost models. Thus, SMOTE-

ENN-XGBoost provides a theoretical basis for developing evaluation criteria for identifying orphan genes in unbalanced and biological datasets.Nettleton, D.F., Katsantonis, D., et al. [18] proposed four models for predicting rice blast disease, two operational process-based models (Yoshino and Water Accounting Rice Model (WARM)) and two approaches based on machine learning algorithms (M5Rules and Recurrent Neural Networks (RNN)), the former inducing a rule-based model and the latter building a neural network. In situ telemetry is important to obtain quality in-field data for predictive models and this was a key aspect of the RICE-GUARD project on which this study is based. According to the authors, this is the first-time process-based and machine learning modelling approaches for supporting plant disease management are compared.Luo, Yh., et al. [19] new method for rice blast grading based on sensitive bands was proposed. Then, the method of system clustering method, BP neural network and probabilistic neural network were used to establish the rice blast classification prediction model, respectively. Comparing the three models, the classification effect based on probabilistic neural network is the best. In the training samples, the logarithmic spectral classification accuracy is 97.8%. In the test samples, the logarithmic spectral classification accuracy is 75.5%.

Samudra, Ami Anggraini, et al. [20] aimed to develop a web-based system to predict the potential probability of blast disease occurrence by combining weather variable and cultural practices factor. ANN method was used to analyze the influence of weather variables on the occurrence of blast disease and to analyze the influence of cultural practices on the occurrence of blast disease executed by decision tree method. Results of this research indicate that proper cultural practices may inhibit the development of blast disease although weather variables support the development of the disease. As conclusion, a web-based system has been developed and can be used to predict the potential probability of blast occurrence.Das, Ankur, et al. [21] extracts different types of features from the disease portions (i.e., images) of the plants and identifies the most valuable features that can distinguish the disease types. To identify the most valuable features, initially, a weighted graph is constructed with extracted features as nodes and similarity between every pair of features as the weight of the corresponding edge. Based on the weights assigned to the edges, importance of each node of the graph is calculated. Finally, a graph-based clustering algorithm namely, Infomap clustering algorithm is applied on the graph to partition it into a set of connected subgraphs.Daniya, T., and S. Vigneshwari, et al. [22] notable ML and image processing concepts in detecting and classifying the plant diseases are discussed. Probabilistic Neural Network (PNN), Genetic Algorithms (GA), k-Nearest Neighbor Classifier (KNN) and Support Vector Machine (SVM) are the various classification techniques used in various applications in the agricultural research. Different input data yields varied quality of an outcome and so selecting a classification method is a critical task. Biological research, agriculture, etc. are the disparate fields where the plant leaf disease classifications are applied.

3. Proposed System Design

Activity diagram: Activity diagram is another important diagram in UML to describe the dynamic aspects of the system.

3.1 MLP Classifier

The MLP was developed to tackle this limitation. It is a neural network where the mapping between inputs and output is non-linear. A MLP has input and output layers, and one or more hidden layers with many neurons stacked together. And while in the Perceptron the neuron must have an activation function that imposes a threshold, like ReLU or sigmoid, neurons in a MLP can use any arbitrary activation function. MLP falls under the category of feedforward algorithms, because inputs are combined with the initial weights in a weighted sum and subjected to the activation function, just like in the Perceptron. But the difference is that each linear combination is propagated to the next layer. Each layer is feeding the next one with the result of their computation, their internal representation of the data. This goes all the way through the hidden layers to the output layer.

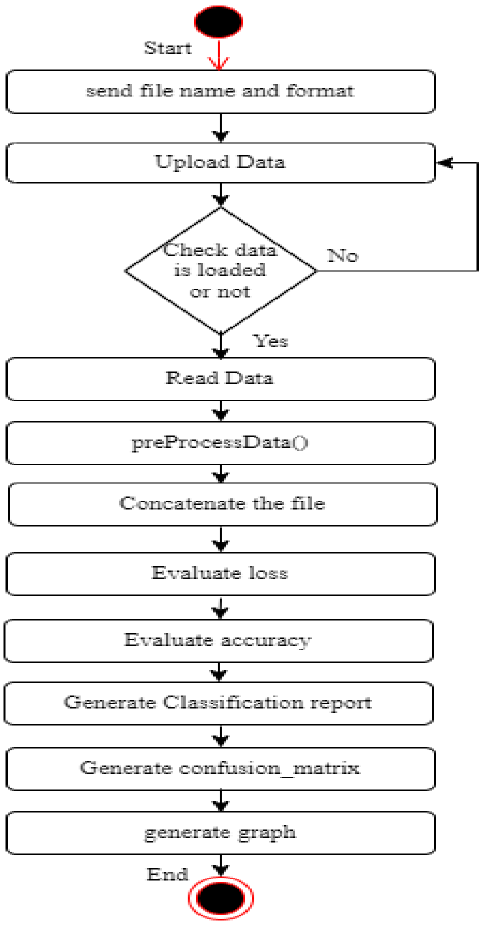

Figure 1: Proposed system design.

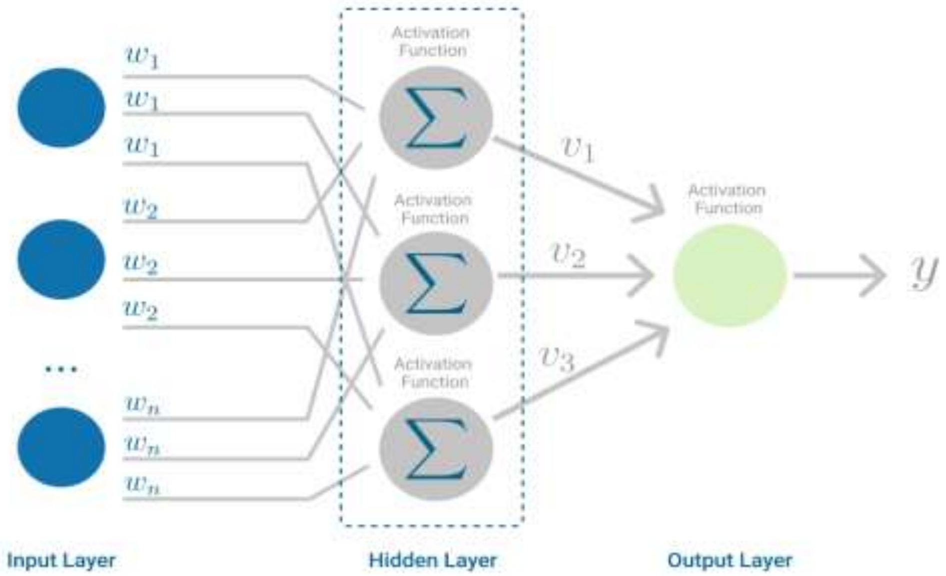

Figure 2: Architecture of MLP.

If the algorithm only computed the weighted sums in each neuron, propagated results to the output layer, and stopped there, it wouldn't be able to learn the weights that minimize the cost function. If the algorithm only computed one iteration, there would be no actual learning. This is where Backpropagation comes into play.

Backpropagation: Backpropagation is the learning mechanism that allows the MLP to iteratively adjust the weights in the network, with the goal of minimizing the cost function. There is one hard requirement for backpropagation to work properly. The function that combines inputs and weights in a neuron, for instance the weighted sum, and the threshold function, for instance ReLU, must be differentiable. These functions must have a bounded derivative because Gradient Descent is typically the optimization function used in MLP. In each iteration, after the weighted sums are forwarded through all layers, the gradient of the Mean Squared Error is computed across all input and output pairs. Then, to propagate it back, the weights of the first hidden layer are updated with the value of the gradient. That's how the weights are propagated back to the starting point of the neural network. One iteration of Gradient Descent is defined as follows:

$$\underbrace{\Delta_w(t)}_{\substack{\text{Gradient}\\ \text{Current Iteration}}} = -\overset{\text{Bias}}{\varepsilon}\,\frac{\overset{\text{Error}}{dE}}{\underbrace{dw_{(t)}}_{\text{Weight vector}}} + \overset{\text{Learning Rate}}{\alpha}\underbrace{\Delta_{w(t-1)}}_{\substack{\text{Gradient}\\ \text{Previous Iteration}}}$$

This process keeps going until gradient for each input-output pair has converged, meaning the newly computed gradient hasn't changed more than a specified convergence threshold, compared to the previous iteration.

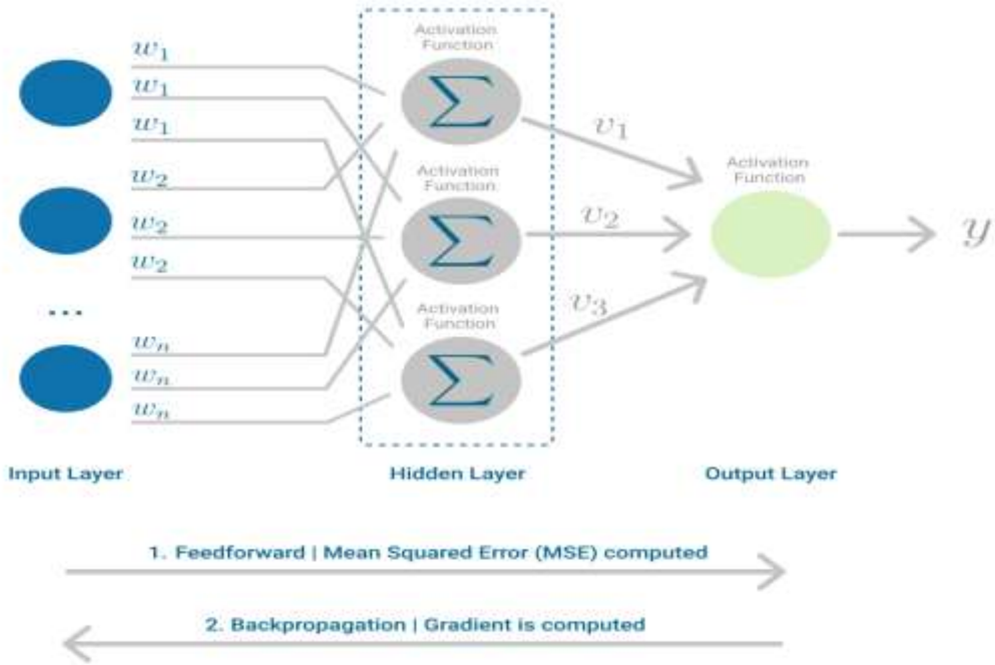

Figure. 3: MLP, highlighting the Feedforward and Backpropagation steps.

4. Results and Description

The figure below represents a portion of the dataset used for Rice Blast Disease prediction. It may include various columns representing features (independent variables) and the target variable (label).

	Unnamed: 0	label	StnPres	StnPresMax	StnPresMin	Temperature	T Max	T Min	RH	RHMin	...
0	1512	0	-1.365290	-0.923222	-1.339543	1.015785	0.978746	0.759838	0.660148	1.433364	...
1	1585	0	0.693535	0.428618	0.734359	0.198946	-0.342005	0.499561	1.124277	1.354817	...
2	1917	1	0.530605	0.340771	0.508779	0.640481	0.545067	0.808010	-0.984304	-0.608880	...
3	1865	1	0.500980	0.291968	0.450560	0.618404	0.111387	0.976736	0.079986	0.412242	...
4	2435	1	-0.098892	-0.064293	-0.073369	0.706711	0.623918	0.889977	0.079986	0.333695	...
...	...	...	...	...	...	...	...	...	...	...	...
3546	770	0	0.730564	0.428618	0.705250	0.044409	0.347940	-0.302961	-1.312401	-1.708550	...
3547	2479	1	-0.128517	-0.093573	-0.167972	0.066486	0.131100	-0.129443	-0.268110	-0.608880	...
3548	3365	1	-2.646505	-1.679667	-2.634822	0.750864	0.978746	0.868287	0.776180	1.040625	...
3549	31	0	0.360272	0.243165	0.370517	-1.169810	-0.913674	-1.148862	-1.544466	-1.551454	...
3550	2309	1	-1.209769	-0.810976	-1.143067	0.993708	0.564780	0.911667	-0.152078	-0.137593	...

3551 rows × 175 columns

Figure 4: sample dataset used for rice blast disease

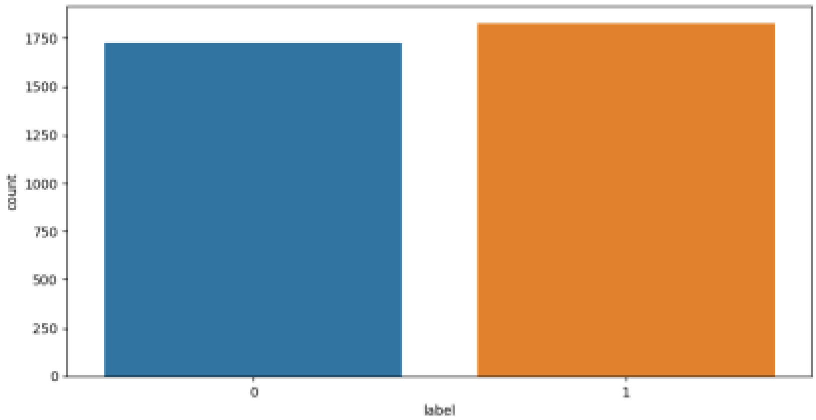

Figure 5: Count plot for label column of a dataset

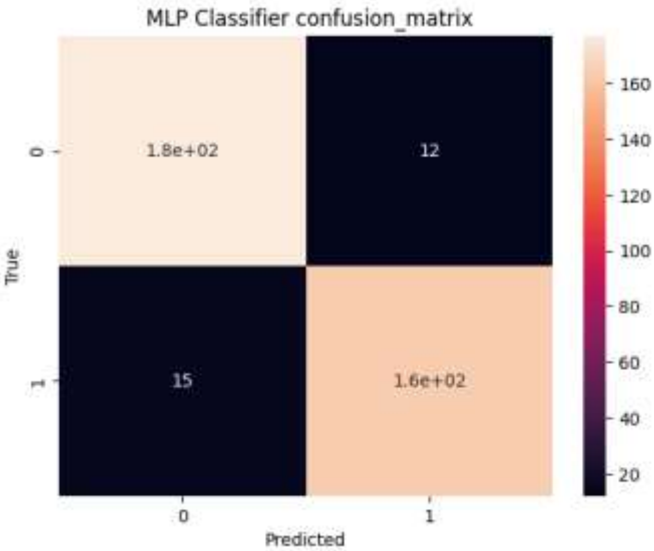

Figure 6: confusion matrix of MLP Classifier

5. Conclusion

For the majority of Indians, agriculture is their main source of revenue. India is the world's second-largest producer of wheat and rice, and agriculture contributes 17% of the country's overall GDP. A significant food product for many regions of India is rice (oryza sativa). Rice is an essential crop in India because it was grown there on the largest scale. Providing 20% of all energy and serving as the primary food source for more than 50% of the world's people, rice is a significant cereal crop. Recent advancements in crop production technology had put a strain on rice output and had an effect on disease occurrence as well Thus, intensive fertilization is a key component of crop management, while repetitive flooding exacerbates the disease issue and expanded rice monoculture aids in the spread of pathogens from one crop to another. Numerous pathogens had harmed India's rice harvest. The most destructive of the 36 rice illnesses was rice blast, which was brought on by Magnaporthe Oryae. The output of paddy crops throughout the nation was seriously threatened by this disease. Rice blast was still a perplexing issue in several rice-growing regions (tropical and temporal), where the pathogen spreads exponentially and was challenging for farmers to controlled, lowering paddy crop output in the field. Due to favourable weather during the crop season, rice blast in India was a major worry. The fungus that caused the illness manifests, multiplies, and spreads as a result of climate. Although the plant had enough nutrients, the climate factors had a significant impact on the frequency of rice blasts in addition to the climatic factors.

References

[1] Lamba, Shweta, et al. "A novel hybrid severity prediction model for blast paddy disease using machine learning." *Sustainability* 15.2 (2023): 1502.

[2] Mandal, Nandita, et al. "Spectral characterization and severity assessment of rice blast disease using univariate and multivariate models." *Frontiers in Plant Science* 14 (2023): 1067189.

[3] Pandit, Devanshi, et al. "Effect of weather parameters on the development and progression of rice blast disease in Jammu plains." *Indian Phytopathology* 76.1 (2023): 89-94.

[4] Das, Shubhajyoti, et al. "Deep Learning Analysis of Rice Blast Disease Using Remote Sensing Images." *IEEE Geoscience and Remote Sensing Letters* 20 (2023): 1-5.

[5] Daniya, Thavasilingam, and Vigneshwari Srinivasan. "Shuffled shepherd social optimization based deep learning for rice leaf disease classification and severity percentage prediction." *Concurrency and Computation: Practice and Experience* 35.4 (2023): e7523.

[6] Zhang, Nannan, et al. "Detection of Cotton Verticillium Wilt Disease Severity Based on Hyperspectrum and GWO-SVM." *Remote Sensing* 15.13 (2023): 3373.

[7] Sandeep, N., et al. "Rice Blast Forecasting Using Interval Valued Data at Coimbatore, India." *Int. J. Environ. Clim. Change* 13.10 (2023): 1882-1888.

[8] Stephen, Ancy, A. Punitha, and A. Chandrasekar. "Optimal deep generative adversarial network and convolutional neural network for rice leaf disease prediction." *The Visual Computer* (2023): 1-18.

[9] Singh, Gursewak, and Ranjit Singh. "Rice Leaf Disease Prediction: A Survey." *2023 International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2023.

[10] Mirandilla, Jean Rochielle F., et al. "Leaf Spectral Analysis for Detection and Differentiation of Three Major Rice Diseases in the Philippines." *Remote Sensing* 15.12 (2023): 3058.

[11] Varsha, M., B. Poornima, and Pavan Kumar. "A Machine Learning Technique for Rice Blast Disease Severity Prediction Using K-Means SMOTE Class Balancing." International Journal of Risk and Contingency Management (IJRCM) 11.1 (2022): 1-27. Singh, Gursewak, and Ranjit Singh. "Rice Leaf Disease Prediction: A Survey." *2023 International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2023.

[12] Sriwanna, Kittakorn. "Weather-based rice blast disease forecasting." Computers and Electronics in Agriculture 193 (2022): 106685.

[13] O. V. Putra, N. Trisnaningrum, N. S. Puspitasari, A. T. Wibowo and E. Rachmawaty, "An Optimized Rice Leaf Disease Classification using Transfer Learning and Balanced Class Weight Distribution based on Bandit Approach," 2022 5th International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2022, pp. 417-422, doi: 10.1109/ICOIACT55506.2022.9971878.

[14] Jiang, Zhencun, Wenping, Yuze. "Recognition of rice leaf diseases and wheat leaf diseases based on multi-task deep transfer learning." Computers and Electronics in Agriculture 186 (2021): 106184.

[15] Bhatia, Anshul, Anuradha Chug, and Amit Prakash Singh. "Application of extreme learning machine in plant disease prediction for highly imbalanced dataset." Journal of Statistics and Management Systems 23.6 (2020): 1059-1068.

[16] Ma, H.; Huang, W.; Jing, Y.; Yang, C.; Han, L.; Dong, Y.; Ye, H.; Shi, Y.; Zheng, Q.; Liu, L.; Ruan, C. Integrating Growth and Environmental Parameters to Discriminate Powdery Mildew and Aphid of Winter Wheat Using Bi-Temporal Landsat-8 Imagery. Remote Sens. 2019, 11, 846. https://doi.org/10.3390/rs11070846

[17] Gao, Qijuan, Jin, Xia. "Identification of orphan genes in unbalanced datasets based on ensemble learning." Frontiers in Genetics 11 (2020): 820.

[18] Nettleton, D.F., Katsantonis, D., Kalaitzidis, A. et al. Predicting rice blast disease: machine learning versus process-based models. BMC Bioinformatics 20, 514 (2019).

[19] Luo, Yh., Jiang, P., Xie, K. et al. Research on optimal predicting model for the grading detection of rice blast. Opt Rev 26, 118–123 (2019). https://doi.org/10.1007/s10043-018-0487-3

[20] Samudra, Ami Anggraini, Kudang Boro Seminar, and Widodo. "DEVELOPMENT OF WEB-BASED SYSTEM FOR BLAST DISEASE FORECASTING IN RICE PLANTATION." Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika 5.2 (2019): 17-28.

[21] Das, Ankur, Sunanda,Shampa. "Feature selection using graph-based clustering for rice disease prediction." Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2019. Springer Singapore, 2020.

[22] Daniya, T., and S. Vigneshwari. "A review on machine learning techniques for rice plant disease detection in agricultural research." system 28.13 (2019): 49-62.