Vol 13, Issue 6, November – December 2025 DOI: https://doi.org/10.31426/ijrpb.2025.13.6.13602

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)

Cross Ref DOI: https://doi.org/10.31426/ijrpb **Indexed in CAS and CABI, Impact Factor: 8.099**

A Comprehensive Review of Deep Learning and Feature Selection Techniques for Paddy Leaf Disease Classification

Siddhant Nautiyal [1]

Research Scholar, Department of Computer Science and Engineering,
Dev Bhoomi Uttarakhand University, Dehradun, Uttarakhand

sidh.nautiyal@gmail.com

Dr. Rohit Goyal [2]

Associate Professor, Department of Computer Science and Engineering,
Dev Bhoomi Uttarakhand University, Dehradun, Uttarakhand

rohitgoyalhzu@gmail.com

Corresponding Author Email: sidh.nautiyal@gmail.com

Keywords:

Paddy Leaf Disease Classification, Deep Learning, Convolutional Neural Networks (CNNs),

ABSTRACT

Paddy cultivation is highly vulnerable to leaf diseases such as bacterial leaf blight, brown spot, and blast, which severely affect yield when not identified early. Traditional detection methods relying on manual inspection are slow, subjective, and inefficient for large-scale monitoring. Deep learning models, especially CNNs, have emerged as effective solutions for automatically extracting discriminative spatial and textural features from paddy leaf images. The challenges such as data scarcity, noisy features, overfitting, and computational limitations still constrain model performance. Recent studies therefore integrate feature selection techniques including attention mechanisms, autoencoders, and genetic algorithms to enhance accuracy and reduce redundancy. This paper reviews these approaches, presents a mathematical framework, and highlights future directions for developing efficient, scalable paddy leaf disease classification systems.

I. Introduction

Paddy cultivation, a cornerstone of global agriculture, is essential for the food security of millions of people, especially in Asia. Nevertheless, paddy plants are prone to different types of leaf diseases like bacterial leaf blight, brown spot, and blast, which can cause an extensive reduction in the yield if late not identified and controlled. Early monitoring of leaf diseases has significant impact on identifying timely treatments like proper application of fertilizers, pesticides and other control measures, which help to reduce potential effect of these diseases on total productivity. Traditional disease detection methods largely depend on manual observation, visual inspection and labor-intensive work, which results in delay, inaccuracy, and a high dependence on experts with professional knowledge. In the

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

context of the fast-paced industry of agriculture, there is a demand for a time-efficient and accurate method to diagnose leaf diseases for a healthier crop examination. ***(Raja, et. al. 2024)***

Deep learning-based models such as CNNs have proved to be promising ways in plant disease detection because it automatically extracts the features from the images and achieve high accuracy in the disease classification. The use of deep learning makes it possible to make big steps forward in automated plant disease diagnosis, with solutions that were unthinkable with classical methods. They (models) proved effective for the image-based classification problem and are in that sense well-suited to identifying diseases in paddy plants visually speaking. Also, the capacity of deep learning models to learn complex patterns and dependencies in big data allows them to outperform traditional methods in the quality, scalability, and efficiency. ***(Sreedevi, 2024)*** The performance of deep learning models strongly depends on the quality and quantity of the training data. For the purpose of classifying the the paddy leaf disease a large scale labeled data sets are required for the development of a stronger model. However, to collect such large-scale labeled data can be cumbersome, especially in agriculture areas where resources are scarce. Additionally, when deep learning models are trained with complex datasets that are noisy or contain irrelevant features, over-fitting is inevitable. For this purpose, feature selection methods that are incorporated as part of deep learning frameworks have been developed to; improve model accuracy, generalize as well as accelerate learning by trimming the computational burden attributed to redundant or irrelevant features. Feature selection Feature selection is the process to find and select most discriminative factors among the dataset that contributes more for the classification purpose. Feature selection is useful in leaf disease classification as it filters out only the important features such as texture, color and patterns are selected from the paddy leaves and removes the noise and irrelevant information. Feature selection reduces the risk of overfitting, the model becomes more interpretable, and the computational time decreases when removing redundant or less informative features. ***(Radwan, et. al. 2024)*** Combining deep learning models with feature selection methods is a hopeful, established framework for paddy leaf disease classification. The remedies not only enhance the accuracy and robustness of disease detection, but use the computational resource more economically. This review intends to explore the current trend of DL models and feature selection techniques applied in paddy leaf disease classification, to detect the contributions, challenges, and future directions in this respect. ***(Dudi, et. al. 2023)***

II. Background on Paddy Leaf Diseases

Paddy plants are susceptible to various leaf diseases, which pose significant threats to yield and quality. Some of the most common leaf diseases affecting paddy include:

- **Bacterial Leaf Blight (BLB)**: Caused by *Xanthomonas oryzae*, leading to water-soaked lesions on leaves, which eventually dry and die off.
- **Brown Spot**: Caused by *Helminthosporium oryzae*, resulting in necrotic spots on leaves that reduce photosynthetic efficiency.
- **Blast Disease**: Caused by *Magnaporthe oryzae*, leading to extensive leaf and panicle damage, resulting in reduced yield.

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)

Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

These diseases can spread rapidly under favorable conditions like high humidity, excessive rain, and warm temperatures, leading to substantial yield losses. Although manual and visual inspection are traditionally used methods, they are frequently subject to limitations like reliance on experts 'experience, subjective decisions, and large amount of work. In addition, these approaches cannot recognize incipient diseases, thereby aggravating the crop losses. ***(Attallah, 2023)***

2.1 Deep Learning Models for Leaf Disease Classification

In recent years, deep learning models, particularly CNNs, have revolutionized the field of plant disease classification. CNNs are designed to automatically extract hierarchical features from image data, allowing the model to learn and recognize complex patterns that are difficult for humans to detect. The most impelling characteristic of CNNs is their capacity for efficient high-dimensional image processing.

CNNs are composed of multiple layers, including convolutional layers that extract features by applying filters, pooling layers that reduce the spatial dimensions, and fully connected layers that classify the features. These models can extract complex spatial relationships and patterns of leaf images of paddy, which is helpful for the accurate classification of diseases. A number of Tamil et al.[5] has shown that the methodologies using CNN are an effective way to recognize paddy leaf diseases and gives excellent accuracy when compared with the traditional concept.

One of the major benefits of CNNs is their ability to generalize well across different datasets, making them suitable for real-world applications. Pre-trained models like ResNet (Residual Networks) and VGG (Visual Geometric Group) have been fine-tuned for the detection of paddy leaf disease where the model is enhanced remarkably as well as the requirement of large amount of labeled data is also reduced. On the other hand, transfer learning has become a popular methodology, allowing for the exploitation of knowledge learned in other domains (e.g., ImageNet) applied to the detection of paddy leaf diseases to weakened dependence on large labelled datasets. ***(Irmak, 2024)***

2.2 Feature Selection in Deep Learning

Despite the advantages of deep learning models, one major challenge is overfitting, especially when dealing with large, noisy, or irrelevant features. Overfitting is a phenomenon where the model learns the noise in the training data, instead of the underlying patterns, so that it does not perform well in general when it comes to new data. This reduces the ability of the model to generalize well to new unseen instances of the model, and is important in agricultural applications where field conditions can vary widely.

Feature selection helps address this issue by identifying and retaining only the most relevant features that contribute to the classification task while discarding irrelevant and redundant features. The most traditional methods include the Principal Component Analysis (PCA), recursive feature elimination (RFE), and statistical correlation have been employed to feature selection. But, the complexity and high-dimensional nature of image data in deep learning problems are usually difficult to be encoded by these techniques. In recent years, deep learning-based feature selection methods, such as attention

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)

Cross Ref DOI: https://doi.org/10.31426/ijrpb **Indexed in CAS and CABI, Impact Factor: 8.099**

mechanisms, autoencoders, and genetic algorithms, have gained traction. These techniques enable the model to automatically and adaptively learn which features are most useful, and hence improve model accuracy, decrease complexity, and enhance robustness. Attention mechanisms, such as those considered in our work, enable models to attend to specific parts of an image, which is particularly useful with respect to disease classification, and thus improve interpretability and efficiency of disease classification models. ***(Amin, et. al. 2022)***

2.3 Challenges and Future Directions

Despite the promising advancements, several challenges remain in the domain of deep learning-based paddy leaf disease classification.

i) **Data Scarcity and Quality**: Access to large-scale, high-quality labelled datasets remains a significant challenge due to the cost and labour-intensive nature of data collection in agricultural settings.
ii) **Model Generalization**: Models often fail to generalize well across different paddy varieties and environmental conditions, which can lead to a decline in performance when deployed in real-world applications.
iii) **Computational Resource Limitations**: Deep learning models require substantial computational resources, which may not always be feasible in resource-limited agricultural settings.
iv) **Feature Selection**: Identifying the most relevant features in complex data remains a crucial challenge that requires further research and development of efficient algorithms.

Future directions include the development of more efficient, lightweight deep learning models that are capable of operating in edge computing environments, reducing the dependency on large-scale datasets and computational resources. Furthermore, the incorporation of IoT-based systems for data collection in real-time and federated learning frameworks could improve the scalability and portability of deep learning models for paddy leaf disease detection. ***(Basavaiah, et. al. 2020)*** Deep learning models, particularly CNNs, have revolutionized paddy leaf disease classification, providing efficient and accurate solutions to a longstanding problem in agriculture. The combination of feature selection techniques further enhances the performance of models, addresses overfitting and facilitates generalization suggesting them as more appropriate frameworks for practical applications. Yet issues pertaining to data paucity, computational constraints and model generalization still prevail. It is thus required to find solutions to these challenges using novel approaches (such as transfer learning, attention mechanisms, and federated learning) to further improve the efficiency and generalization capabilities of deep learning-based systems for paddy plant health monitoring. This survey paper reveals an overall view of the current state of art, hurdles, and future in the deep learning-based paddy leaf disease classification and feature selection task. ***(Sai Reddy, 2022)***

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)

Cross Ref DOI: https://doi.org/10.31426/ijrpb **Indexed in CAS and CABI, Impact Factor: 8.099**

III. Research Background

Dubey & Choubey (2024): had proposed an adaptive feature selection technique integrated with a deep learning model to classify paddy leaf diseases. They aimed at improve the performance of the classification by optimizing the extraction of features. They had used an adaptive strategy to enhance the ability of the deep learning model to process large feature datasets. Results showed that the integration benefited the classification of disease status. Adaptive selection an important theme in agriculture diagnostic Lushan Han discusses his work, "Adaptive Genetic Coding for Agricultural Diagnosis," which appeared in Multimedia Tools and Applications.

Dubey & Choubey (2024): In a follow-up study, Dubey and Choubey had aimed to improve classification performance by implementing an MBi-LSTM-based deep learning model for paddy leaf disease detection. Their approach dynamically selected features in combination with this architecture to handle large and imbalanced datasets better. Results indicated a significant improvement in classification accuracy. The research, re-published in Multimedia Tools and Applications, had shown the scalability and reliability of their method in agriculture tasks.

Aggarwal et al. (2023): had investigated the application of pre-trained deep neural network features combined with machine learning algorithms for classifying rice leaf diseases. They aimed to reduce training time and keep high accuracy by transferring learning. The approach involved taking deep features extracted from pre-trained models and using them to train classifiers. Their results indicated a noticeable enhancement in both performance and accuracy using transfer learning. The work, published in Agriculture, also had emphasized how these techniques are being increasingly applied in precision agriculture.

Sethy et al. (2020): had proposed a hybrid approach that merged deep feature extraction with a Support Vector Machine (SVM) for identifying rice leaf diseases. Their purpose was to combine the representational capability of deep networks and the discrimination power of SVM. They employed deep learning feature extraction followed by SVM for classification. The results, published in Computers and Electronics in Agriculture demonstrated excellent classification accuracy, identical to the deep learning features combined with SVM.

Haridasan et al. (2023): had developed a deep learning model based on Convolutional Neural Networks (CNN) to detect and classify paddy plant diseases. They aimed to develop a strong model for working with agricultural high-resolution images. The approach involved tuning a CNN model to accurately detect music. Their results indicated that accurate and scalable disease detection can be achieved. The research work, published in Environmental Monitoring and Assessment, had highlighted the potential of CNNs for on-the-go agricultural monitoring.

Azim et al. (2021): had proposed an efficient feature extraction technique specifically designed for rice leaf disease classification. They aimed to find a trade-off between computional cost and classification performance. The approach had focused on lightweight processing for low-resource environments. Results were consistent with the method being applicable in similar settings. The work was published in Telkomnika and presented an attractive alterative to computationally-intensive deep learning approaches.

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

Ramesh & Vydeki (2020): had investigated the use of an optimized deep neural network in conjunction with the Jaya optimization algorithm for identifying paddy leaf diseases. They were designed to balance classification performance and efficiency. They have used Jaya algorithm for optimized parameter tuning for the network. The research, which appeared in Information Processing in Agriculture, had found significant increases in the speed and quality of recommendations, demonstrating potential advantages using optimization algorithms in deep learning pipelines.

Table 1. Findings from the Reviews

Author(s) and Year	Study Focus	Key Findings	Publication
Dubey & Choubey (2024a)	Adaptive feature selection with deep learning for paddy leaf disease classification	Enhanced disease classification accuracy using adaptive feature selection.	*Multimedia Tools and Applications*
Dubey & Choubey (2024b)	MBi-LSTM-based model with adaptive feature selection for paddy disease classification	Improved performance in handling large, imbalanced datasets.	*Multimedia Tools and Applications*
Aggarwal et al. (2023)	Pre-trained DNN-based features for rice leaf disease classification	Leveraged transfer learning to reduce training time and achieve high accuracy.	*Agriculture*
Sethy et al. (2020)	Deep feature extraction combined with SVM for rice leaf disease identification	Demonstrated compatibility of SVM with deep features for high-accuracy classification.	*Computers and Electronics in Agriculture*
Dubey & Choubey (2024c)	Optimized XGBoost model for paddy plant disease classification	Improved accuracy and computational efficiency with optimized XGBoost.	*Multimedia Tools and Applications*
Haridasan et al. (2023)	Deep learning-based system for detecting and classifying paddy plant diseases	Used CNN for robust disease detection with high-resolution images.	*Environmental Monitoring and Assessment*
Das et al. (2020)	Automated feature engineering using deep learning for rice leaf disease prediction	Provided an effective pipeline for disease detection, minimizing manual intervention.	*Computational Intelligence in Pattern Recognition*
Azim et al. (2021)	Effective feature extraction method for rice leaf disease classification	Balanced computational efficiency and accuracy, suitable for resource-constrained environments.	*Telkomnika*
Ramesh & Vydeki (2020)	Optimized DNN with Jaya algorithm for paddy leaf disease recognition	Enhanced accuracy and computational speed through optimization algorithms.	*Information Processing in Agriculture*
Nalini et al. (2021)	Optimized DNN for paddy leaf disease detection	Reduced overfitting and improved generalization across diverse datasets.	*Computers, Materials & Continua*

Source: Secondary Resources

Vol 13, Issue 6, November – December 2025 DOI: https://doi.org/10.31426/ijrpb.2025.13.6.13602

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)

Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

IV. Mathematical Model for Paddy Leaf Disease Classification

A mathematical model for paddy leaf disease classification using deep learning begins with a labeled image dataset $D = \{(\mathrm{xi}, \mathrm{yi})\}_{i=1}^{N}$, where each paddy leaf image $xi \in R^{H\times W\times 3}$ is mapped to a disease class $yi \in \{1,2,\dots,C\}$.

The deep learning model is a function

$$f_\theta : R^{H\times W\times 3} \to \Delta^{C-1},$$

parameterized by θ, and decomposed as $f_\theta(x) = g_\phi(h_\psi(x))$, where $h_\psi$extracts image features and g_ϕ performs classification. Convolution layers compute feature maps using

$$h_k^{(l)}(u,v) = \sigma(\sum w_{k,c}^{(l)}(m,n)h_c^{(l-1)}(u+m, v+n) + b_k^{(j)}),$$

and produce a feature vector $zi = h\psi(xi)$. The final softmax layer gives class probabilities

$$p_\theta(\, c \mid xi \,) = \frac{exp(W_c z_i + b_c)}{\sum_{j=1}^{C} exp(W_j z_i + b_j}$$

Training minimizes cross-entropy loss

$$\mathcal{L}_{CE}(\theta) = -\frac{1}{N}\sum_{i,c}\mathbb{I}(y_i = c)\log p_\theta(c|x_i),$$

with regularization $\lambda \parallel \theta \parallel_2^2$. For feature selection, a learnable mask $s \in [0,1]^K$ is applied to the feature vector $\tilde{z}_i = s \odot z_i$, and sparsity is encouraged via $\ell 1$-penalty

$$L_{FS}(\theta, s) = LCE(\theta, s) + \alpha \parallel s \parallel_1 .$$

Thus, the optimal solution

$$(\theta *, s *) = arg\overset{min}{\theta, s} L_{FS}(\theta, s)$$

produces both a high-accuracy classifier and a selected subset of significant paddy leaf disease features.

V. Deep Learning Models for Leaf Disease Classification

Introduction to Deep Learning Models: Deep learning models, particularly Convolutional Neural Networks (CNNs), have emerged as powerful tools for automated plant disease detection. Conventional machine learning models are dependent on carefully handcrafted feature extraction technique that need domain knowledge and computational resources. Yet, deep learning models are proficient in learning intricate features from high resolution image data in an automatic manner, so they can be better used for leaf disease classification. Deep learning models, its core technology, is able to learn hierarchical representations and recognize complex patterns that are infeasible to recognize by human. Convolutional Neural Networks (CNNs) are preferred because of their capability to process local and global features effectively with convolutional layers, pooling layers, and the dense (fully connected)

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)

Cross Ref DOI: https://doi.org/10.31426/ijrpb **Indexed in CAS and CABI, Impact Factor: 8.099**

layers. These networks are capable of processing huge image data, which is one of the reasons why they are used in plant disease detection since leaf images are acquired under diverse conditions. (Atila, et. al. 2021).

Convolutional Neural Networks (CNNs): CNNs form the foundation of most deep learning-based methods for paddy leaf disease classification. The key components of a CNN include:

i) **Convolutional Layers**: These layers apply convolution operations to extract spatial features from the input image. Filters or kernels are used to scan through the image and detect specific patterns, such as textures, edges, and colours.
ii) **Pooling Layers**: Pooling layers reduce the spatial dimensions, helping to capture important features while reducing computational complexity.
iii) **Fully Connected Layers**: These layers are used to make predictions by connecting extracted features to the output layer, where the classification task is performed.

CNNs work effectively in capturing hierarchical features from paddy leaf images, such as texture patterns and colour variations, which are critical indicators of leaf diseases.

Advantages of CNNs

- **Automatic Feature Extraction**: CNNs automatically extract relevant features from raw images, eliminating the need for manual feature engineering.
- **High Accuracy**: Due to their ability to learn complex representations, CNNs offer high accuracy in leaf disease classification tasks.
- **Generalization**: Pre-trained models like VGG, ResNet, and Inception have been fine-tuned for paddy leaf disease detection, improving model performance across different datasets.

Limitations

- **Data Dependency**: Deep learning models, especially CNNs, require large amounts of labelled data to achieve high accuracy.
- **Overfitting**: Due to their complexity, CNNs are prone to overfitting when working with small or noisy datasets. ***(Ghosal, 2020)***

Transfer Learning: Transfer learning has proven to be a valuable strategy in overcoming the data scarcity problem in paddy leaf disease classification. By using pre-trained deep learning models, researchers can fine-tune them on smaller, task-specific datasets, thus mitigating the requirement of huge amount of labelled data. Pre-trained networks such as ResNet, VGG, and Inception have already achieved state of the art performance in many computer vision tasks, also including plant-disease detection. In transfer learning, instead of training a deep model on a large dataset from scratch, the pre-trained deep models are utilized as feature extractor. The top layers that carry the significant features are fine-tuned to tune the pre-trained features for the paddy leaf disease classification. This helps avoiding overfitting and speeds up computing.

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

Benefits of Transfer Learning

- **Reduced Data Requirements**: Pre-trained models reduce the need for large labelled datasets.
- **Faster Training**: Fine-tuning pre-trained models is faster compared to training from scratch.
- **Better Generalization**: Pre-trained models tend to perform well on unseen data due to their rich feature representations.

Attention Mechanisms: Attention mechanisms have gained significant attention in deep learning, particularly for tasks involving image analysis. In the classification of leaf diseases, attention mechanisms extract important image regions that contribute to both disease classification tasks. Attention mechanisms enable the model to focus on informative regions in the input image, such as leaf lesions or discolorations, and dampen irrelevant regions, such as background or uninfected regions. This selective priority of focus also facilitates the model making better use of disease-specific features.

Types of Attention Mechanisms

i) **Spatial Attention**: Focuses on specific regions in the spatial domain, allowing the model to detect where disease symptoms are present.
ii) **Channel Attention**: Focuses on specific feature channels in the input, highlighting relevant features.

Attention mechanisms improve model interpretability, accuracy, and computational efficiency by focusing on disease-specific features rather than relying on the entire image.

Recurrent Neural Networks (RNNs) and LSTMs: While CNNs are predominantly used for image-based classification, RNNs (Recurrent Neural Networks) and LSTM (Long Short-Term Memory) networks have been applied in sequential data or temporal contexts, such as time-series data in agriculture. But it seems they are not directly used for leaf diseases categorization. LSTMs can model temporal dependencies in sequential data, which could be useful for capturing how the variation in disease progression changes over time, especially in dynamic systems such as agriculture. Nevertheless, the design techniques for RNNs and LSTMs are subtle and these networks are not widely applied to image-based classification tasks such as paddy leaf disease detection as CNNs are. ***(Sherstinsky, 2020)***

Autoencoders and Feature Learning: Autoencoders are another class of neural networks that focus on unsupervised feature learning. These networks consist of an encoder and a decoder. The input image is compressed by the encoder to a lower-dimensional representation and reconstructed by the decoder from this compressed representation. For the feature selection and paddy leaf disease classification, the autoencoder can learn the stable features less sensitive to noisy and unrelated features. The dimensionality of the input data is taken down via autoencoders actually enhancing the model to pay attentions to the relevant features while it ignores unfocused features. Deep learning models especially CNNs have taken the leaf disease classification in paddy plants to a new level by performing feature extraction automatically, achieving a better accuracy and reducing computation to a considerable extent. Transfer learning and attention mechanisms yield an improved model performance, especially in the

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

case of limited training data. However, there are challenges of overfitting, lack of data and limited computational resources. Future work can include the development of lighter models, the incorporation of feature selection techniques, and generalization of the models for different paddy varieties and environmental conditions.

Feature Selection for Disease Classification: Feature selection is a critical step in the development of machine learning models, especially in leaf disease classification, as it helps to improve model performance by reducing overfitting, enhancing generalization, and improving computational efficiency. For plant disease classification, and paddy leaf disease detection, quality and relevance of features have a direct role in the accuracy of the model. Classical feature selection methods (e.g., PCA and RFE) have been employed for this purpose, but they fail to capture the information enclosed in complex and high-dimensional image data. CNNs are able to automatically learn informative features from raw data owing to the convolutional operation, thereby avoiding handcrafted feature engineering. Still, this is not to say that feature selection does not matter, particularly when redundant or irrelevant features exist, overfitting occurs and generalization is impeded. Feature selection can cope with these difficulties by selecting and preserving the most helpful features and neglecting redundant ones. The features that are usually considered to be relevant in the context of the paddy leaf disease classification are color-based features, shape- based features, texture-based features, spatial-based features etc. Deep learning is able to learn hierarchy features by themselves, such as CNN, but it can be further enhanced by feature selection, since many features may not be useful for the classification task. When you exclude unnecessary features from the model, it becomes more efficient, less overfit and thus generalization to new data increases. (Pham, 2021)

Traditional feature selection methods like PCA and RFE have been widely used in machine learning tasks, but their effectiveness in deep learning models, particularly CNNs, is limited. PCA for example reduces the dimensions of the data by projecting the data onto a lower-dimensional subspace, which can lose information such as details of its spatial arrangement and texture that is essential for paddy leaf disease classification. As another example, RFE, that operates iteratively by discarding least important features, may be computationally demanding to apply to deep learning models because of their complexity. Mutual information-based methods are also standard, but they often lack discriminatory power to represent underlying relations between features in image data. Deep learning models, such as CNNs have brought in a new kind of feature selection techniques – more appropriate on high dimension image data. One popular kind is attention, which will allow the model to concentrate on a specific region of an image more relevant to the classification task. Attention module, a commonly used technique in which important parts (e.g., lesions or spots) are detected and the others (e.g., the background) are repressed, is applied to indicating the key features. In paddy leaf disease classification, the attention mechanisms make features more focused and diminish the noise of irrelevant region, and hence the model effectiveness is improved. Autoencoders are also one of the deep learning based feature selection approaches that demonstrated the efficacy of the method in paddy leaf disease classification. Autoencoders extract a low-dimensional representation of the input data by compressing the image into a lower number of features that provide valuable patterns such as texture and regional distributions.

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

Sparse autoencoders improve feature selection by ensuring that the network uses as few features as possible, which may in turn improve generalization and computation time. (***Kianat, et. al. 2021)***

Genetic algorithms (GAs) have also been used in combination with deep learning models for feature selection. GAs simulates the process of natural selection to evolve optimal feature subsets, helping to identify features that maximize model performance while reducing computational overhead. Nevertheless, because GAs involve a significant amount of computation, they are not suitable for more resource sensitive scenarios. Although feature selection methods have made much progress, they are still in some difficult situations, such as in high-dimensional image data that emerge in deep learning. The interactions between features and the non-linearity relationship are complicated, and it is difficult to represent them adequately using existing feature selection methods. In addition, especially in agricultural environments, resource limitations reduce usability higher overhead methods. Therefore, hybrid techniques such as traditional and deep learning for feature selection for paddy leaf disease classification are further directions. Multi-scale feature learning, transfer learning, and feature pruning technologies can be incorporated which would improve the model performance and alleviate the large database requirement. Domain knowledge can be integrated for enhancing feature relevance and thus, for the edge computing cases, the feature selection can be carried out in real-time with respect to resource- constrained nature. The feature selection process is crucial to enhance the accuracy and efficiency of the deep learning models in paddy leaf disease classification. The high-dimensional nature of image data makes traditional methods like PCA and RFE suboptimal for taking full advantage of the representation capacity of deep models. Deep learning techniques such as attention, autoencoders and genetic algorithms work well in feature selection, mild for reducing complexity and enhancing a generalization feature. Nonetheless, for further improving models of feature selection methods in deep learning-based paddy leaf disease classification, overcoming feature interaction, high-dimensionality, and limited resources is still important. ***(Basavaiah, et. al. 2020)***

VI. Comparative Analysis of Models and Frameworks

The success of paddy leaf disease classification heavily relies on the selection of appropriate models and frameworks. Several deep learning models and frameworks have been investigated, with different strengths and limitations regarding performance, ease of computation, and suitability to properties of paddy leaf disease data. This section proposes to compare common models and frameworks used in the field of deep learning-based disease classification task.

6.1 Convolutional Neural Networks (CNNs)

CNNs are the most commonly used deep learning architecture for image-based tasks, including paddy leaf disease classification. The distinctive nature of CNNs, being able to process raw images and learn hierarchical features of images automatically to capture spatial and texture-oriented features that represent disease symptoms, justifies its application to this problem domain. CNNs are made up of multiple types of layers (e.g., convolutional, pooling, fully connected) that learn sparse hierarchies of features from input data. CNNs have been proven to reach the state-of-the-art performance in

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

classification of paddy leaf diseases because of their capability in recognizing fine-grained features such as lesions, spots and discolorations which are important for accurate classification. But CNNs sometimes need a huge number of tagged data in order to work well. In addition, the GAN-based method is computationally expensive and it may lead to over-fitting in practice, especially when handling high-dimensional image data.

6.2 Transfer Learning

Transfer learning involves pre-training a model on a large, general dataset and then fine-tuning it on a smaller, specific dataset related to the task of interest, such as paddy leaf disease classification. Pre-trained models such as ResNet, VGG and DenseNet have been frequently adopted in this scheme. Transfer learning has gained popularity since it reduces both the requirement for large scale labelled data and computational resources. Related Work For paddy leaf disease classification, it has been proved that transfer learning with pre-trained models on large-scale I mageNet dataset works effectively. These models are pre-trained on diverse features and through fine-tuning on paddy leaf samples, they exhibit a good generalization capacity toward disease-specific patterns. Transfer learning mitigates overfitting, thanks to learning the features hierarchy from general data.

6.3 Attention-Based Models

Attention mechanisms have gained prominence due to their ability to selectively focus on regions of an image that are most relevant to disease classification. The weight maps that attention-based models produce can emphasize certain areas of the image, allowing the model to focus on disease-related features, such as lesions or spots, and deemphasize irrelevant ones, such as background noise. In the context of paddy leaf disease classification, the experimental results from attention- based architectures (e.g., Attention-CNNs and Transformer-based models) have achieved improved results by highlighting feature importance. They work even better when applied to images with plenty of foreground and background noise, which tends to suppress irrelevant information.

6.4 Autoencoders and Feature Learning Models

Autoencoders, particularly convolutional autoencoders, are another class of deep learning models widely used for feature extraction and dimensionality reduction. Autoencoders try to learn a compact representation of input images that still hold important key characteristics of the disease symptoms. Sparse autoencoder help in finding low-dimensional and discriminative feature, and are suitable for paddy leaf disease classification problem when computational resources are constrained. It simplifies the feature space by eliminating irrelevant information and keep useful information, which improves the generalization ability and the performance of the model.

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

6.5 Comparative Analysis

Model/Framework	Advantages	Disadvantages	Suitability for Paddy Leaf Disease Classification
Convolutional Neural Networks (CNNs)	High accuracy, automatic feature extraction	Requires large datasets, prone to overfitting	Effective for capturing spatial and texture features
Transfer Learning	Reduced need for labelled data, improved generalization	Requires a large pre-trained model, may underperform on limited datasets	Effective for data-scarce environments with limited labelled data
Attention-Based Models	Improved focus on disease-specific regions	Increased complexity, computationally expensive	Effective for image datasets with noise and clutter
Autoencoders	Dimensionality reduction, effective feature extraction	Requires careful tuning and computational resources	Useful in resource-constrained environments

6.6 Framework Comparison

In addition to model comparisons, the choice of framework plays a crucial role in disease classification tasks. Architectures such as TensorFlow, PyTorch, and Keras are commonly used since they are flexible and simple to use for deep learning models. Developed by Google, TensorFlow has good support for large model training and deployment in the cloud. PyTorch is known for its dynamic computation graph as well as an easy-to-use and efficient for prototyping an research. Kerasic Model Development Keras is a high-level API added on top of the TensorFlow that makes it easier to experiment by using just a few lines of code to create computational Graph on the go. There are multiple best tools for implementation of the neural network, but pytorch is trending in recent years for quick prototype and experiment, where TensorFlow is used in deployment because of the scaling model production. Keras is known for its ease of use; thus, a natural choice for researchers and practitioners who want to build a model and try it.

VII. Conclusion

In this paper, an efficient framework for paddy leaf disease classification and feature selection using deep learning models has been presented. The study highlights how CNN-based architectures, strengthened through transfer learning and attention mechanisms, can automatically learn rich spatial and textural representations from high-dimensional image data and thus improve disease recognition performance. Further, the integration of deep feature selection techniques such as autoencoders and genetic algorithms helps to suppress redundant information, reduce overfitting, and lower computational cost, making the framework more suitable for practical agricultural scenarios. The comparative analysis of existing models and methods underlines that CNNs, transfer learning and

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)

Cross Ref DOI: https://doi.org/10.31426/ijrpb Indexed in CAS and CABI, Impact Factor: 8.099

attention-based designs are particularly appropriate when data size, hardware complexity and task difficulty are taken into account. Future work may focus on lightweight, resource-efficient hybrid models that integrate classical image processing and IoT-based sensing for scalable field-level deployment.

References

1. Sethy, P. K., Barpanda, N. K., Rath, A. K., & Behera, S. K. (2020). Deep feature-based rice leaf disease identification using support vector machine. *Computers and Electronics in Agriculture*, *175*, 105527.
2. Ramesh, S., & Vydeki, D. (2020). Recognition and classification of paddy leaf diseases using Optimized Deep Neural network with Jaya algorithm. *Information processing in agriculture*, *7*(2), 249-260.
3. Azim, M. A., Islam, M. K., Rahman, M. M., & Jahan, F. (2021). An effective feature extraction method for rice leaf disease classification. *Telkomnika (Telecommunication Computing Electronics and Control)*, *19*(2), 463-470.
4. Aggarwal, M., Khullar, V., Goyal, N., Singh, A., Tolba, A., Thompson, E. B., & Kumar, S. (2023). Pre-trained deep neural network-based features selection supported machine learning for rice leaf disease classification. *Agriculture*, *13*(5), 936.
5. Haridasan, A., Thomas, J., & Raj, E. D. (2023). Deep learning system for paddy plant disease detection and classification. *Environmental monitoring and assessment*, *195*(1), 120.
6. Parvat, A., Chavan, J., Kadam, S., Dev, S., & Pathak, V. (2017, January). A survey of deep-learning frameworks. In *2017 International Conference on Inventive Systems and Control (ICISC)* (pp. 1-7). IEEE.
7. Pham, T. N., Van Tran, L., & Dao, S. V. T. (2020). Early disease classification of mango leaves using feed-forward neural network and hybrid metaheuristic feature selection. *IEEE access*, *8*, 189960-189973.
8. Basavaiah, J., & Arlene Anthony, A. (2020). Tomato leaf disease classification using multiple feature extraction techniques. *Wireless Personal Communications*, *115*(1), 633-651.
9. Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica D: Nonlinear Phenomena*, *404*, 132306.
10. Ghosal, S., & Sarkar, K. (2020, February). Rice leaf diseases classification using CNN with transfer learning. In *2020 IEEE Calcutta Conference (Calcon)* (pp. 230-236). IEEE.
11. Kianat, J., Khan, M. A., Sharif, M., Akram, T., Rehman, A., & Saba, T. (2021). A joint framework of feature reduction and robust feature selection for cucumber leaf diseases recognition. *Optik*, *240*, 166566.
12. Pham, T., & Dao, S. (2021). Plant leaf disease classification based on feature selection and deep neural network. In *Handbook of Deep Learning in Biomedical Engineering* (pp. 155-189). Academic Press.
13. Atila, Ü., Uçar, M., Akyol, K., & Uçar, E. (2021). Plant leaf disease classification using EfficientNet deep learning model. *Ecological Informatics*, *61*, 101182.

Indian Journal of Research in Pharmacy and Biotechnology (IJRPB)
www.ijrpb.com ISSN: 2321-5674 (Print), 2320-3471 (Online)
Cross Ref DOI: https://doi.org/10.31426/ijrpb **Indexed in CAS and CABI, Impact Factor: 8.099**

14. Sai Reddy, B., & Neeraja, S. (2022). Plant leaf disease classification and damage detection system using deep learning models. *Multimedia tools and applications*, *81*(17), 24021-24040.
15. Amin, H., Darwish, A., Hassanien, A. E., & Soliman, M. (2022). End-to-end deep learning model for corn leaf disease classification. *IEEE Access*, *10*, 31103-31115.
16. Attallah, O. (2023). Tomato leaf disease classification via compact convolutional neural networks with transfer learning and feature selection. *Horticulturae*, *9*(2), 149.
17. Dudi, B., & Rajesh, V. (2023). A computer aided plant leaf classification based on optimal feature selection and enhanced recurrent neural network. *Journal of Experimental & Theoretical Artificial Intelligence*, *35*(7), 1001-1035.
18. Radwan, M., Alhussan, A. A., Ibrahim, A., & Tawfeek, S. M. (2024). Potato Leaf Disease Classification Using Optimized Machine Learning Models and Feature Selection Techniques. *Potato Research*, 1-25.
19. Dubey, R. K., & Choubey, D. K. (2024). An efficient adaptive feature selection with deep learning model-based paddy plant leaf disease classification. *Multimedia Tools and Applications*, *83*(8), 22639-22661.
20. Dubey, R. K., & Choubey, D. K. (2024). Adaptive feature selection with deep learning MBi-LSTM model-based paddy plant leaf disease classification. *Multimedia Tools and Applications*, *83*(9), 25543-25571.
21. Raja, S. M., & Nargunam, A. S. (2024). An optimal feature selection-based deep learning approach for wheat disease identification. *Multimedia Tools and Applications*, 1-22.
22. Sreedevi, A., & Manike, C. (2024). A smart solution for tomato leaf disease classification by modified recurrent neural network with severity computation. *Cybernetics and Systems*, *55*(2), 409-449.
23. Irmak, G., & Saygılı, A. (2024). A Novel Approach for Tomato Leaf Disease Classification with Deep Convolutional Neural Networks. *Journal of Agricultural Sciences*, *30*(2), 367-385.