ISSN 1840-4855
e-ISSN 2233-0046

Original scientific article
http://dx.doi.org/10.70102/afts.2026.1835.027

RICE LEAF DISEASE DIAGNOSIS THROUGH DEEP LEARNING: AN INCEPTIONV3 APPROACH WITH SPATIAL ATTENTION FOR SUSTAINABLE AGRICULTURE AND FOOD SECURITY

R. Dhanya[1*], Dr.S. Mythili[2]

[1]Research Scholar, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India. e-mail: dhanyar23@gmail.com, orcid: https://orcid.org/0009-0004-3886-8638*
[2]Professor and Head, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India. e-mail: smythili78@gmail.com, orcid: https://orcid.org/0000-0003-3196-6257

Received: December 23, 2025; Revised: February 11, 2026; Accepted: March 26, 2026; Published: May 29, 2026

SUMMARY

This research introduces a sophisticated deep learning framework for automatically identifying diseases in rice leaves by combining the Inception V3 architecture with spatial attention mechanisms. Rice is one of the most important foods in the world, in terms of food security and agricultural economics, so the establishment of efficient disease surveillance systems is now necessary to support sustainable farming activities. This is particularly important in achieving SDG 2 (Zero Hunger), which aims to ensure access to sufficient food, and SDG 12 (Responsible Consumption and Production), which promotes sustainable farming practices. Convolutional neural networks have been successfully used to classify plant diseases in the past, but traditional convolutional neural network models are sometimes not capable of ranking the most diagnostically relevant features. The attention mechanisms used in this methodology defeat this challenge by integrating the Inception V3 framework. In particular, the spatial attention aspect guides the model to areas that are of disease-specific features. The study used a dataset of 18,160 images of the rice leaf, which included nine separate disease types and controls that were selected through the Kaggle and Plant Village datasets. Results showed that the attention-enhanced hybrid model reached 92.98% accuracy in classification tasks, surpassing the standard Inception V3 baseline in fewer than 12 training epochs. Significant improvements were also noticed in the cases of distinguishing diseases having similar visual appearance, especially in early and late blight conditions. The model showed strong performance across all ten rice disease categories that were tested, reaching its best validation accuracy of 92.98% during the 12th training epoch. These findings indicate that there is indeed a benefit of incorporating attention processes into the InceptionV3 architecture for the task of disease detection in agricultural crops.

Key words*: deep learning, convolutional neural networks, inceptionv3, spatial attention, rice leaf disease, plant disease detection, sdg 2 (zero hunger), sdg 12 (responsible consumption and production), sdg 15 (life on land)*

INTRODUCTION

The cultivation of rice is repeatedly having a war against the plant diseases that may destroy the whole harvest, killing millions of farmers who are relying on this essential crop to earn a living. Conventional methods of disease identification are extensive and dependent on the manual efforts of agronomists who conduct manual field studies, a model that finds it difficult to match the intensity of the contemporary rice growth, and most of the problems are detected when they are already at advanced stages [16]. This study deals with these shortcomings by constructing an intelligent diagnostic system that will use two strong technology elements together in order to detect rice leaf diseases using photographs automatically. Fundamentally, the InceptionV3 structure is the basis of the multi-scale visual features extraction based on the advanced 48-layer deep neural network. The architecture uses a set of inception modules, which process visual inputs at different scales in parallel based on parallel convolutional pathways with different sizes (1×1, 3×3, and 5×5) of filters to respond to fine-grained and broad contextual features [17]. This multi-resolution method provides a successful method by which the network can learn to identify the sophisticated morphological patterns to identify various manifestations of disease in various crops and under different environmental conditions.

In addition to this powerful feature extraction property, the spatial attention mechanism provides a directed attention to disease-related areas in the leaf image [18]. Instead of using a static approach that only considers the spatial locations as equal, this mechanism is dynamic in the sense that it calculates weights of attention, which represent the specific disease-affected areas, basically training the model on where to focus its analytical processes. The attention module attains this by using learned transformations to produce spatial probability maps and enables the network to reduce the background noise and irrelevant foliar structure and enhance the areas of the image that have pathological symptoms of lesions, chlorotic pattern, or necrotic tissue.

The combination of hierarchical feature extraction of InceptionV3 and the intelligent region prioritization of the spatial attention mechanism forms a diagnostically powerful system that goes beyond traditional classification methods [19]. The working principle of this integration is a well-structured architectural pipeline: deep convolutional features initially extract the identifying characteristic disease markers, and the attention mechanism establishes the wear and identifies the exact spatial position of them [20]. This dual-pathway processing resembles that of professional plant pathologists who scan a whole specimen, but pay attention to the analysis of the symptomatic area.

In addition to offering better diagnostic accuracy, the architecture also offers important interpretability in the form of attention visualizations [21]. The system instills necessary confidence with the agricultural practitioners and the plant pathologists since it creates saliency maps, which precisely clarify the areas of the leaf that contribute to the diagnostic decision. Such visualizations allow experts in the domain to have confidence in the model in that the model's conclusions are in agreement with known pathological information and known symptom patterns, as opposed to basing them on incoherent algorithmic decision-making. This transparency is especially useful in farm environments, where the confidence of stakeholders directs the use of technology and its real implementation.

Essentially, the effectiveness that the system brings lies in its end-to-end combined training paradigm, where the backpropagation of the feature extraction backbone to the attention mechanism together is learned. By means of this single-minded optimization, it is guaranteed that the attention mechanism picks out really significant diagnostic clues - those properties that predict the real pathology of the disease as opposed to mere visual correlations or artifacts in the data. By refining the attention weights through training epochs, the attention weights will collectively focus on biologically relevant areas to form a system whose learned representations are in line with domain knowledge and may also identify subtle diagnostic patterns, which may not be noticed by humans.

Section 1: Introduction - The establishment of the foundation and motivation of the research. Section 2: Literature Review, Section 3: Methodology - Detailed experimental framework with: Dataset composition (13,632 images, 10 disease categories with descriptions table), Data preprocessing steps, Model architecture (InceptionV3 + Spatial Attention), Training configuration, Evaluation metrics.

Section 4: Results - Performance measures such as: Peak validation accuracy, training accuracy, learning curve analysis, and minimal overfitting demonstration. Section 5: Conclusion and Future Work.

LITERATURE REVIEW

The sphere of artificial intelligence application in the detection of diseases in plants has gone through an unprecedented path of development throughout the last few years, and scholars worldwide keep expanding upon one another and their findings. Every research has provided important fragments to the knowledge, and it is important to note these contributions by their names, since behind every progress there are real individuals who have dedicated their profession to finding solutions to such issues.

The study conducted by al-Gaashani and others (2023) [1] used ResNet50 architecture together with a kernel attention mechanism in the diagnosis of rice diseases and showed that attention-based adjustments to the residual networks could increase the accuracy of the detection. This is because the approach of using the kernel attention model enabled the model to focus on the diagnostically prominent features and to eliminate the background noise. In Zhang et al. [2] proposed ResViT-Rice that combines transformer encoders with residual modules to detect rice disease. This hybrid design makes use of local property extraction capacities of residual networks and the global contextual knowledge of transformers. In a study carried out by Jiang et al. (2023) [3], the proposed study methodology was a combination of the attention mechanisms and dense deep networks in identifying rice disease. Their method focused on feature reuse by using dense connections, but taking into consideration attention to emphasize disease-specific patterns. Bhuyan et al. (2023) [4] proposed SE_SPnet, which is a stacked parallel convolutional neural network to predict the rice leaf disease with the help of squeeze-and-excitation blocks. The parallel architecture allows the run of multi-scale feature extraction, and in the process, the SE blocks recalibrate the channel-wise response of features. A special ResNet architecture created by Stephen et al. (2023) [5] was specifically designed to perform the classification of rice leaf disease based on self-attention. In their model, self-attention layers were implemented to have significant improvements in dynamically changing the importance of the features used during the classification process. The identification network that Yang et al. (2023) [6] proposed is lightweight, includes attention mechanisms, and dynamic convolution to identify the disease in rice. The dynamic convolution optimizes parameters of the kernels with respect to the properties of the input and ensures computer performance at the computational cost otherwise prohibited by resource-intensive settings. Yang et al. (2023) [7] also conducted another study that used GoogLeNet with residual connections and attention mechanisms to detect diseases in rice leaves. This mixture saved the multi-scale processing feature of GoogLeNet and introduced residual pathways to help the gradient flow and attention modules target the discriminative localities. Ni et al. (2023) [8] were interested in classifying standard rice pests/diseases with the ECA (Efficient Channel Attention) mechanism. The ECA module offers channel attention that has lower complexity in computation than the conventional methods; thus, it is feasible in real-world applications. The authors of Lu and Chen (2024) [9] proposed a parallel structure of residual, which has coordinated attention and is applied to classify rice disease. The coordinate attention mechanism is a horizontal as well as vertical way of encoding spatial information that is important in recognizing patterns of disease and capturing positional relationships with the disease patterns. Wang et al. (2024) [10] investigated the issue of ensemble learning by stacking separate deep convolutional neural networks to classify rice disease. Predictions of various architectures are added together in an ensemble method, making the system more robust and general to different presentations of disorders. The system proposed by Memon et al. (2024) [11] to classify the rice leaf disease is an effective approach based on a MobileNet architecture with attention modules. This architecture is computationally efficient, and hence it is especially suitable when used in agricultural contexts as a mobile and edge computing solution. Thakur and co-authors (2023) [12] developed the explainable vision transformer named PlantXViT that consists of a convolutional neural network and the vision transformer used to identify plant diseases. Attention visualizations are interpretable results offered by the model to enable users to know the image regions that were used in making classification decisions. At [13], Chakrabarty et al. introduced an interpretable fusion model that incorporates lightweight CNN models and components of transformers to identify rice leaf disease. This design combines computational effectiveness with the capacity of the transformer to extract long-range dependencies of image features. Deep multi-net, the multi-scale based on multi-scale depth-wise separable convolution, was suggested by Hu et al. (2025) [14] as a method of

identifying rice leaf disease. The architecture also minimizes the number of parameters without compromising on classification accuracy due to a selected set of feature extraction in multiple resolutions. In [15], Veerasamy and Fredrik developed an intelligence system using deep learning and image processing methods to detect weeds in crop and vegetable plantations. Their method indicates the generalizability of computer vision techniques used in agricultural management, besides disease detection. Another intelligent farming system that was recently launched by Veerasamy and Fredrik (2023) is an uncertainty expert system and a butterfly optimization algorithm to recommend crops. The article demonstrates that optimization algorithms can be used to optimize the decision-making process in precision agriculture applications.

Based on the related papers, it is evident that Inception v3 is a decent architecture when it comes to image classification, and spatial attention focuses on the area of locus concentration. Hence, the joint technique of both methods (figure 1) in the detection of the rice leaf disease will give a superior validation accuracy that facilitates the early detection of the disease that will allow crop protection at an early stage.

METHODOLOGY

Dataset Acquisition and Preparation

Dataset Composition

Table 1. Dataset composition (Kaggle and Plantvillage)

Disease Name	Scientific Name	Type of Disease	No. of Images Used
Bacterial_Leaf_Blight	Xanthomonas oryzae pv. Oryzae	Bacteria	1386
Brown_Spot	Helminthosporium oryzae	Fungus	1480
Leaf_Blast	Magnaporthe oryzae	Fungus	1801
Leaf_Scald	Monographella albescens	Fungus	1670
Neck_Blast	Magnaporthe grisea	Fungus	1000
Rice_Hispa	Dicladispa armigera	Fungus	1461
Shealth_Blight	Rhizoctonia solani Kühn	Fungus	1578
Tungro	Rice tungro bacilliform virus	Virus	1740
Narrow Brown Spot	Cercospora oryzae	Fungus	1416
Healthy	-	-	1100

The study used a detailed dataset on rice leaf disease (Table 1) images of 10 types of classification. figure 1-based data set was obtained in Kaggle (https://www.kaggle.com/datasets/vbookshelf/rice-leaf-diseases) and the Plant Village dataset that is freely available and includes images in diverse lighting conditions, backgrounds, and development phases. This is because such diversity guarantees that the model will be applicable to the real-life conditions in the fields where the environment changes considerably.

1. Neck Blast - This is also due to fungus, and it is characterized by greyish spots on the neck of the rice panicle.
2. Rice Hispa - Damage by insects which form typical linear feeding spots.
3. Sheath Blight- Fungus the irregular lesions on the sheath of leaves.
4. Tungro - Viral disease that is yellowing and retarded growth.
5. Bacterial Leaf Blight- Bacterial infection causes water-soaked lesions.
6. Brown Spot- Fungal disease and brown spots are oval and occur in pairs.
7. Healthy -Normal leaves with no disease symptoms
8. Leaf Blast - The lesions are of a diamond or spindle form and are of the fungi type.
9. Leaf Scald - This is a condition with appearances of scalding and lesions.
10. Narrow Brown Spot- Fungal disease that has narrow brown lesions.

Figure 1. Dataset

Data Preprocessing

Image Preparation and Preprocessing

The preprocessing phase of rice leaf image data can be regarded as one of the initial stages in the development of a deep learning model, even though it may appear to be rather simple. The quality of data preparation has a direct impact on the model performance no matter the sophistication in the architecture. This rule reflects on the culinary preparation better recipes that need well-prepared ingredients to attain the best outcomes.

Image Resizing: Dimensional Standardization

All pictures are standardized to 299x299 pixels, which is the size of the input of the InceptionV3 that is tailor-made and optimized. This dimension compromises adequacy; detail capture and computational efficiency so that the multi-scale parallel pathways in the network may be in optimum functioning. Field photographs, however, seldom have square aspect ratios, landscape and portrait orientation is frequent, and simple reduction to square size would distort such disease features as circular lesions or diamond-shaped patterns of blast. To overcome this challenge, the adaptive bilinear interpolation was implemented but instead of just duplicating or eliminating adjacent pixels, the new pixel values are computed by taking weighted averages of adjacent pixels. In nearly square images (e.g. 300320) the algorithm uses simple uniform resizing with minimum distortion. The occasional images of moderate aspect (e.g. 600 x 400) are resized and proportions are maintained then very carefully to arrive at 299 299. Extremely high or low aspect ratios (e.g. 800300) of the image are first centre-cropped to the focus on the middle zones where the leaves are usually located, and then resized. This adaptive mechanism continues the morphological integrity of disease characteristics that make circular lesions circular and geometry characteristic lesions intact that is critical to proper diagnosis because these shapes form the diagnostic features that differentiate different diseases.

Normalization: Optimization of Neural Network

The channel-wise pixel intensity normalization by ImageNet statistics was used. ImageNet is an enormous collection of images in excess of 14 million images of thousands of categories, including animals, vehicles, and common objects. In the case of InceptionV3, when it was initially trained on ImageNet images were normalized with the help of special mean and standard deviation values computed on such a vast and heterogeneous data. This process of normalization will have three parts that are tightly interrelated and each of them is crucially important:

ImageNet statistics allows normalization, which has three important advantages to the model. To begin with, it more quickly converges to a solution by producing a homogenous optimization landscape which gradient descent can empirically take shorter paths to parameter optimality. Second, it mitigates internal covariate shift by maintaining consistent statistical distributions as data flows between layers, which allows each layer to learn more efficiently rather than constantly adapting to changing input characteristics. Third, and most important for transfer learning, normalizing rice leaf images with ImageNet statistics (mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]) ensures compatibility with the pretrained InceptionV3 features that were originally learned on identically normalized data this alignment allows the model's early layers to effectively recognize edges, textures, and shapes in the agricultural images without requiring extensive retraining to adapt to a different statistical distribution.

Noise Reduction: Preserving Signal, Removing Interference

The final preprocessing step addresses digital noise inherent in field photographs random pixel variations from sensor limitations, JPEG compression, or challenging lighting conditions that can obscure genuine disease features or be mistaken for diagnostic patterns. Applied the adaptive Gaussian filtering to reduce this high-frequency noise while preserving the lesion boundaries and texture patterns critical for disease identification. The filtering uses a 3×3 kernel with sigma (σ) = 0.5, parameters chosen to create gentle smoothing that removes sensor noise without blurring the edges that distinguish different disease types for instance, sharp lesion borders versus diffuse boundaries often indicate different pathologies. The adaptive approach proved particularly valuable because noise levels vary significantly across images depending on capture conditions: high-quality photographs taken in good lighting receive minimal filtering to avoid unnecessary blur, while noisier smartphone images captured in dim conditions receive slightly stronger smoothing (σ up to 0.8). The algorithm estimates noise levels by analyzing pixel variation in regions that should be uniform, such as healthy leaf tissue, then adjusts filtering strength accordingly. During filtering, each pixel's new value comes from a weighted average where the centre pixel contributes roughly 40% and neighbours contribute proportionally less based on distance, following a Gaussian distribution this means a lesion pixel with value [120, 80, 60] and similar neighbours ranging [115-125, 75-85, 55-65] becomes [120, 81, 61], maintaining its character while reducing random variation. This balanced approach successfully reduces noise interference without compromising the diagnostic features that enable accurate classification.

The Integrated Preprocessing Pipeline

The preprocessing pipeline transforms raw field photographs into model-ready inputs through three essential steps that work together to optimize data quality. First, the dimension of each image is downscaled to 299x299 pixels without bilinear interpolation with special consideration to the aspect ratio with due care to retain the disease lesions and symptoms in its original form. Then, pixel values are normalized with ImageNet statistics, which replaces the standard 0-255 RGB range with zero-centered distributions, which are usually in the range of approximately -3 to +3 such normalization causes the data to match the pretrained InceptionV3 model, and it accelerates learning. Lastly, adaptive Gaussian filtering is used to minimize sensor noise and image artifacts but (intentionally) does not restore the edges and textures necessary to differentiate the various disease types. This entire sequence executes in roughly 200-300 milliseconds on standard hardware, fast enough for real-time field deployment, and the careful attention to each preprocessing step directly contributes to the model's final validation accuracy of 92.98% across ten rice disease categories.

Data Augmentation

Once the separate images are meticulously prepared by resizing, normalizing and filtering them another important hurdle can be seen with the construction of a powerful disease detection system; It should not memorize certain training images but what it should do is to learn to detect diseases under any number of variations in which they may appear in the real world under conditions. It is at this point that data augmentation becomes extremely important.

The Overfitting Problem: When Models Memorize Instead of Learn

Rice leaf disease datasets, while comprehensive, still represent a finite sample of the virtually infinite variety of ways diseases can manifest in field conditions. A leaf photographed in bright midday sun looks dramatically different from one photographed in the soft light of late afternoon. A leaf held perfectly flat versus one slightly curved toward the camera presents different perspectives. A leaf in early stages of infection shows subtle symptoms that look quite different from advanced disease progression. A model may be trained to identify certain images in training sets that are beautiful without taking into consideration this natural variation where it will fail when faced with the inevitable variations it will experience in the real world.

Data Augmentation: Generating Diversity to Develop Resilience

To resolve this problem, the group used data augmentation a procedure that creates systematic differences of training images, which increases the exposure of the model to varying visual displays. It is a method which is comparable to the study of diseased leaf samples not just with 100 fixed images of the samples but with images of the same samples which have been taken under different light conditions, different angles and in different disease development stages. This extensive exposure permits the model to discover underlying pathological attributes as opposed to memorizing the superficial attributes of individual photographs. The study implemented a comprehensive augmentation framework incorporating several transformation techniques. Each transformation was deliberately selected to replicate the variability encountered in practical field conditions. This augmentation strategies served to enhance the model's generalization capabilities by introducing controlled variations that mirror real-world imaging scenarios, including fluctuations in ambient lighting, camera angles, image resolution, and specimen positioning within the frame.

Geometric Transformations: Mimicking Camera Perspectives

The first category of augmentations addressed how leaves might be oriented or positioned when photographed in the field. When farmers pull out smartphones to photograph suspicious leaves, they're not setting up professional photo shoots with controlled positioning. The leaf might be to their left or right, the camera might be tilted slightly, or they might need to photograph from an awkward angle because of where the affected plant is located in the field. Random horizontal flipping was applied with a 50% probability to each training image. This means that half the time, an image would be mirrored left-to-right, as if the leaf had been photographed from the opposite side. This makes perfect sense for plant leaves, which don't have an inherent "correct" orientation a leaf photographed from the left side and one photographed from the right side both contain the same diagnostic information. By including both orientations in training, models recognize diseases regardless of which direction the farmer approached the plant from.

Random rotation of up to ±10 degrees addressed another aspect of real-world variability. When standing in rice fields photographing leaves, especially on plants swaying slightly in the breeze or when reaching to photograph a leaf in an awkward position, perfect horizontal alignment is nearly impossible. The leaf might be tilted slightly clockwise or counter clockwise relative to the camera. The choice of ±10 degrees specifically represent a sweet spot. This range is large enough to encompass typical angular variations seen in casual field photography small tilts and rotations that happen naturally when taking handheld photos. But it's conservative enough to avoid extreme rotations that might create unrealistic appearances

or crop away important parts of the leaf when the image is rotated and then cropped back to its original rectangular dimensions.

Colour Space Transformations: Accounting for Lighting and Environmental Variability

The second major category of augmentations addressed something intuitively understood by outdoor photographers but that computer vision models must explicitly learn to handle: the dramatic impact of lighting conditions on how subjects appear in photographs. Brightness adjustment of ±20% simulated the range of lighting conditions from early morning or late afternoon (dimmer, requiring brightness increase) to harsh midday sun (brighter, potentially requiring brightness reduction for optimal visualization). This seemingly simple transformation has profound implications for disease detection.

Consider a brown spot disease lesion on a rice leaf. Under bright, direct sunlight, the lesion might appear as a darker brown against a brightly illuminated green background. Under overcast conditions or in shade, the same lesion appears as a medium brown against a darker green background. The absolute pixel values are completely different what measured as RGB [140, 100, 60] in bright light might measure as [100, 70, 40] in shade but it's the same lesion on the same leaf. Without brightness augmentation, a model might learn to associate specific absolute brightness values with diseases, making it unreliable across different lighting conditions.

By randomly adjusting brightness up or down by 20% during training, models were exposed to the same disease appearing at different overall illumination levels. This forced learning of brightness-invariant features characteristics of diseases that remain consistent regardless of lighting. The spatial pattern of lesions, the relative colour differences between diseased and healthy tissue, the texture and boundary characteristics these remain diagnostically valuable even when absolute brightness varies. Contrast variation of ±20% addressed a related but distinct aspect of image appearance. Contrast refers to the difference between the lightest and darkest parts of an image. High contrast images have very bright highlights and very dark shadows with sharp transitions between them. Low contrast images have a more compressed range, appearing somewhat washed out or flat.

Different cameras, different camera settings, and different atmospheric conditions (hazy versus clear air) all affect image contrast. A budget smartphone with basic image processing might produce lower contrast images than a high-end camera with sophisticated processing algorithms. Morning fog or haze reduces contrast by scattering light. Contrast augmentation simulated these variations. From a disease detection perspective, contrast affects how clearly lesion boundaries appear. A bacterial leaf blight lesion with a naturally sharp boundary between diseased brown tissue and healthy green tissue will show this boundary very distinctly in high-contrast images the colour transition is abrupt and obvious. In low-contrast images, the same boundary appears softer, with a more gradual transition. Both images show the same disease, but they look different. By training with varied contrast, models learned to recognize diseases across this spectrum of appearances.

Saturation modification of ±20% was perhaps the most subtle but still important colour transformation. Saturation refers to the intensity or purity of colours highly saturated colours are vivid and intense, while desaturated colours appear washed out or greyish. This matters particularly for plant disease detection because disease symptoms often involve colour changes: healthy green tissue turning yellow, brown, or grey. But the exact appearance of these colours depends on saturation. Under certain lighting conditions or with certain camera colour profiles, a diseased yellow area might appear as a vivid, saturated yellow. Under other conditions, it might appear as a pale, desaturated yellowish-tan. The hue (the basic colour family) is similar, but the saturation differs dramatically.

The Combined Effect: Simulating Real-World Diversity

What makes this augmentation strategy particularly powerful is how these transformations combine and interact. During training, images didn't receive just one augmentation—they might be horizontally flipped AND rotated 8 degrees counter clockwise AND have brightness increased by 15% AND contrast reduced by 10% AND saturation increased by 12%, all in the same augmented version. This creates an

exponential explosion of variety. With 1,000 original images of bacterial leaf blight, and each capable of being flipped or not (2 options), rotated to any angle in a range (effectively dozens of options), and varied in brightness, contrast, and saturation (dozens more options for each), the model is exposed to essentially millions of subtle variations during training. The model encounters bacterial leaf blight bright and dim, high contrast and low, saturated and desaturated, from left and right perspectives, at various tilts. This forces models to extract and learn invariant features—characteristics that remain constant across all these variations. For bacterial leaf blight, that might be the pattern of water-soaked lesions starting from leaf tips, the yellowish halo around lesions, and the characteristic progression pattern. These features are present whether the leaf is photographed in morning light or afternoon sun, from the left or right, with expensive cameras or budget smartphones.

Practical Implementation: Real-Time Augmentation

The augmentation implementation approach affects both training efficiency and model robustness. "Online" or "on-the-fly" augmentation was used rather than "offline" augmentation. Offline augmentation would mean taking the original dataset, applying all these transformations to create augmented versions, and saving them all as separate image files. If five different augmentation combinations were applied to each original image, there would be six times as much data stored (original plus five augmented versions). For large datasets, this could easily consume hundreds of gigabytes of storage. Instead, online augmentation was employed. Each time a training image was loaded during the training process, the augmentation pipeline would randomly apply these transformations on-the-fly in the moment, in memory, without saving the augmented version. This meant every time the model encountered a particular training image throughout the many training epochs, it saw a different randomly augmented version.

Dataset Splitting: Honest Performance Evaluation

All the image preparation and augmentation would be meaningless without proper evaluation of model performance. This introduces a critical aspect of machine learning methodology: dataset splitting and the importance of held-out test data. The complete dataset was divided into three distinct portions using a technique called stratified sampling:

Training set (70%): The largest portion, used for the actual learning process—adjusting the millions of parameters in the neural network based on the patterns the model discovers in these images.

Validation set (15%): A separate set never used for training, but examined during the training process to monitor how well the model generalizes to data it hasn't been trained on. This portion tunes hyperparameters (high-level settings like learning rates) and determines when to stop training.

Test set (15%): This is an entirely held-out set, never looked at in any way, until the very end, as only the final performance will be assessed to get an objective measure of the future model performance on indeed novel data.

The 70/15/15 division is the standard and rather reasonable distribution of the datasets of medium size. The training set must be large enough to enable the model to understand various pattern of all types of diseases- the 70% majority. Large validation and test sets are also required to have confidence in estimating the generalization performance hence significant 15% portions per each instead of 90/5/5 which could give invalid results due to too small statistically unacceptable validation and test sets.

Stratified Sampling: Maintaining Balance

The stratified nature of the sampling is critical and it is better to elaborate that. The dataset has a total of 10 categories of disease, yet as in many datasets in the real world, these categories are not balanced. Maybe 800 have been taken of bacterial leaf blight (a commonly occurring disease farmers take snapshots of regularly) but there are only 200 of a disease that is not very prevalent. Without stratified random splitting, there could be unfavourable imbalances. Suppose that one decides to randomly divide

70 % of data into training, 15 % into validation and 15 % into test not taking into account disease categories. The split could happen by mere coincidence resulting in the 75% of the bacterial blight images training, 20 % validation and justify 5 % test. Or 80 % of that infrequent disease it is training and only 10 % in validation and test. Such imbalances would distort the performance measures and it would be hard to effectively examine the performance of the model on each category of diseases. The stratified sampling will avert this through proportional representation. In case bacterial leaf blight constitutes 20% of the entire data, stratified sampling will assure that it constitutes about 20% of training, 20% of validation and 20% of test. Suppose that that rare disease is 5 % of the overall dataset, it is in turn 5 % of every split.

This is mathematically done by individually dividing each category of disease in terms of the 70/15/15, and combining the divisions. It would have 800 images of bacterial blight, 560 of these were going to be used in training (70%), 120 in validation (15%), and 120 in tests (15%). The rare disease will have 200 images of which 140 would be used in training, 30 in validation and 30 in testing. This will make sure each category of disease is included in each split proportionally.

A training rate of 85% and testing of the remaining 15 % without a distinct validation set would put one in a challenging situation during hyperparameter tuning. Each time alternative hyperparameter settings were checked and the most successful ones selected according to the results on the test set, implicit optimization on that particular test set would be achieved. With a large number of repetitions of such a process, hyperparameters that are particularly successful on the manners and peculiarities of the test set would be chosen, which is in effect a subtle kind of overfitting to the test set even though the test set was not directly trained. A report of 92.98% accuracy is on the results of the validation set when under training (at epoch 12). Upon ultimate testing on the held-out test set, results validated either the fact that this performance was generalized or that overfitting to the validation set had been experienced. The fact that the test performance was similar proved that indeed the model learned strong disease recognition instead of memorizing the peculiarities of the training or validation data.

The Complete Training Data Pipeline

Bringing all these pieces together pre-processing, augmentation, and stratified splitting—a robust training data pipeline was created that served the model throughout the learning process:

- Raw images enter the system in their original diverse formats and sizes
- Pre-processing standardizes them: resizing to 299×299, normalization, noise filtering
- Stratified splitting assigns each pre-processed image to training, validation, or test set
- During training, images from the training set are loaded in batches
- Online augmentation randomly transforms each training image on-the-fly
- The model learns from these augmented training images
- Periodically, the model is evaluated on validation images (pre-processed but not augmented)
- After training completes, final performance is measured on test images (also pre-processed but not augmented)

This pipeline ensured the model was exposed to rich, diverse data during training while being evaluated honestly on consistent, standardized images during validation and testing. The augmentation provided robustness to real-world variations. The stratified splitting provided honest performance estimates. Together, they form a methodologically sound foundation for training a reliable, deployable disease detection system.

Model Architecture

The inceptionV3 is used as a framework of extracting features due to its good balance in representational potency and computational performance. Its architectural achievements render it specifically fit well when it comes to fine-grained visual tasks like plant disease recognition.

CLASS InceptionV3WithAttention:

FUNCTION __init__(num_classes):
1.Load pre-trained InceptionV3:
- Weights trained on ImageNet (1.2M images, 1000 classes)
- All layers initialized with learned features
2. Modify final layer:
Original: FC(2048 → 1000)
Modified: FC(2048 → num_classes)
3. Add spatial attention:
spatial_attention = SpatialAttention (kernel_size=7)
4. Disable auxiliary classifier:
aux_logits = False

Factorized Convolutions

In the InceptionV3 architecture, instead of using a single $n \times n$ convolution, the operation is factorized into sequential $1 \times n$and $n \times 1$convolutions. This factorization improves computational efficiency while maintaining the capacity of the model.

Formally, the standard $n \times n$convolution operation (Equation 1) is:

$$F_{\text{out}} = \text{ReLU}\left(\sum_{m=0}^{n-1} W_{n\times 1}(m) \cdot F_{1\times n}(i+m, j) + b_{n\times 1}\right) \quad (1)$$

This is replaced by the factorized $n \times 1$followed by $1 \times n$convolution as shown in equation 2:

$$\sum_{m=0}^{n-1}\sum_{k=0}^{n-1} W(m,k) \cdot X(i+m, j+k) \rightarrow \sum_{m=0}^{n-1} W_{n\times 1}(m) \cdot \left(\sum_{k=0}^{n-1} W_{1\times n}(k) \cdot X(i+m, j+k)\right) \quad (2)$$

Equation 2 shows decomposition reduces the number of parameters and floating-point operations by approximately one-third, while preserving the expressive capacity of the network. As a result, deeper architectures can be constructed without incurring excessive computational cost.

Parallel Multi-Scale Feature Extraction

A defining characteristic of Inception modules is their ability to process information at multiple spatial scales simultaneously. Each module contains parallel convolutional paths with different receptive fields:

- convolutions emphasize pixel-level variations, such as color discoloration or chlorosis.
- 3×3 convolutions capture localized textural patterns, like streaks or fungal growth.
- 5×5 receptive fields, implemented as stacked 3×3 convolutions, model broader contextual patterns, such as lesions or necrotic regions.

Given an input feature map X, the output of an inception block (Equation 3) can be expressed as:

$$F_{\text{inception}} = \text{Concat}\left(F_{1\times 1}, F_{3\times 3}, F_{5\times 5}, F_{\text{pool}}\right) \quad (3)$$

This multi-scale representation is especially effective for rice disease detection, where symptoms may range from microscopic fungal spores to large, visually prominent lesions.

Auxiliary Classifiers

To facilitate stable training of deep networks, InceptionV3 includes auxiliary classifiers at intermediate depths. These auxiliary outputs (Equation 4) introduce additional loss terms during training:

$$L_{\text{total}} = -\sum_{c=1}^{C} y^c \log(\hat{y}^c) + \alpha \sum_{k=1}^{K} \left(-\sum_{c=1}^{C} y^c \log(\hat{y}_{\text{aux}}^c(k)) \right) \quad (4)$$

where $\alpha$is a weighting factor. This mechanism mitigates the vanishing gradient problem and encourages earlier layers to learn discriminative features.

Feature Extraction Layer Selection

Features are extracted from the Mixed_7 layer (using equation 5), which produces feature maps of size:

$$FM7 = \bigoplus_{k=1}^{K} \text{ReLU}\left(W_k^{M7} * FM6 + b_k^{M7}\right) \in \mathbb{R}^{H \times W \times C} \quad (5)$$

Where $FM6$is the feature map from the previous layer, $W_k^{M7}$is the weight matrix, and $b_k^{M7}$is the bias term for the k-th filter.

This layer provides an optimal trade-off between spatial resolution and semantic abstraction. With an effective receptive field of approximately 267×267267 \times 267267×267 pixels, it captures sufficient contextual information while retaining spatial detail critical for accurate disease localization.

Spatial Attention Module

To further enhance disease-relevant regions, integrate a spatial attention mechanism that guides the model to focus on informative areas of the leaf image.

Channel Pooling

Given the feature tensor $F \in \mathbb{R}^{C \times H \times W}$, two spatial descriptors are generated by pooling along the channel dimension (using Equation 6):

$$D_{\text{sp}} = \left[P_{\text{avg-ch}}(F); P_{\text{max-ch}}(F)\right] \in \mathbb{R}^{H \times W \times 2} \quad (6)$$

Both pooled maps lie in $\mathbb{R}^{1 \times H \times W}$and capture complementary statistical cues.

Concatenation and Convolution

The pooled features are concatenated (using equation 7) to form:

$$F_{\text{concat}} = \text{Concat}\left(\sum_{i=1}^{H} \sum_{j=1}^{W} F(i,j), \max F(i,j) \right) \quad (7)$$

A convolutional layer with a 7×7kernel is then applied, followed by a sigmoid activation (Equation 8):

$$A = \sigma\left(\sum_{m=0}^{6} \sum_{n=0}^{6} W(m,n) \cdot X(i+m, j+n) + b \right) \quad (8)$$

The larger kernel size allows the attention mechanism to model broader contextual relationships without significant computational overhead.

Feature Refinement

Finally, the attention map is applied to the original features using element-wise multiplication (Equation 9):

$$F_{\text{out}} = F \odot \sigma\left(W_{7\times7} * \left[F_{\text{avg-sp}} \parallel F_{\text{max-sp}}\right] + b_{7\times7}\right) \quad (9)$$

This operation amplifies disease-relevant regions while suppressing background noise, leading to more discriminative feature representations.

Integration and Classification

The refined feature maps $F_{\text{attended}} \in \mathbb{R}^{17\times17\times768}$ are compressed using global average pooling (Equation 10):

$$F_{\text{gap}} \in \mathbb{R}^{1\times1\times768} = \frac{1}{289}\sum_{i=1}^{17}\sum_{j=1}^{17} F_{\text{attended}}(i,j,c), \forall c \in \{1,2,\ldots,768\} \quad (10)$$

This representation is translation-invariant and aggregates discriminative information across the entire image. To reduce overfitting, dropout with a rate of 0.5 is applied, followed by a fully connected layer and SoftMax activation (Equation 11):

$$y = \text{Softmax}(W_{\text{fc}} \cdot \text{Dropout}(\text{GAP}(F_{\text{attended}}), p = 0.5) + b_{\text{fc}}) \quad (11)$$

3.3 Training Procedure

Transfer Learning Strategy

To effectively adapt a deep convolutional network to the rice disease classification task while avoiding overfitting, adopted a two-phase transfer learning strategy. This staged training approach enables stable convergence and progressive task-specific refinement.

Phase I: Selective Parameter Optimization

In the first phase, the InceptionV3 backbone was initialized with ImageNet-pretrained weights, leveraging its ability to capture general visual patterns such as edges, textures, and shapes. All backbone parameters (approximately 21.8 million) were frozen, and training was restricted to the newly introduced components:

- **Spatial attention module:** 76,864 parameters
- **Classification head:** 7,690 parameters

By limiting learning to these layers, the model is encouraged to specialize in identifying disease-relevant regions and class boundaries, while preserving the robust and transferable representations learned from large-scale natural image data. This phase was conducted for 30 epochs, allowing the attention mechanism to converge and reliably highlight symptomatic regions of rice leaves.

Training Configuration:

- **Optimizer: Adam**

$$\beta_1 = 0.9, \beta_2 = 0.999, \epsilon = 10^{-8} \quad (12)$$

Algorithm: Adam Optimizer

Initialize:

- $m_0 = 0$# First moment vector (mean of gradients)
- $v_0 = 0$# Second moment vector (variance of gradients)
- $t = 0$# Time step

For each training iteration:

1. $t = t + 1$
2. Compute gradient:

$$g_t = \nabla_\theta L(\theta_{t-1}) \quad (13)$$

(gradient of loss with respect to parameters)

3. Update biased first moment estimate:

$$m_t = \beta_1 \times m_{t-1} + (1 - \beta_1) \times g_t \quad (14)$$

(exponential moving average of gradients)

4. Update biased second raw moment estimate:

$$v_t = \beta_2 \times v_{t-1} + (1 - \beta_2) \times g_t^2 \quad (15)$$

(exponential moving average of squared gradients)

5. Compute bias-corrected first moment:

$$\widehat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (16)$$

6. Compute bias-corrected second moment:

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (17)$$

7. Update parameters:

$$\theta_t = \theta_{t-1} - \alpha \times \frac{\widehat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (18)$$

Training Settings:

- Learning rate: 1×10^{-3}
- Batch size: 32
- Loss function: Categorical cross-entropy with label smoothing ($\epsilon = 0.1$)

Label smoothing was employed to reduce overconfidence in predictions and improve generalization, particularly in visually similar disease classes.

The Adam optimizer is a type of adaptive learning rate optimization algorithm which is a combination of both momentum and RMSprop. It has two moment estimates of each parameter, the first-moment (mean of gradients) and the second moment (variance of gradients). These moment estimates are corrected and exponential moving averages updated in such a way that they give proper estimates even in the initial stages of training. These bias-corrected estimates of the moment are then updated with the learning rate with the parameters. Adam is a popular choice of deep learning models because of its performance and the capacity to work with big datasets at varying learning rates. The most important equations are the calculation of the gradient (Equation 13), the moment change (Equations 14 and 15), the bias corrections (Equations 16 and 17) and the parameter change as a final result (Equation 18).

Phase II: Discriminative Fine-Tuning

Once the attention module and classifier stabilized, proceeded to discriminative fine-tuning. In this phase, deeper layers of InceptionV3 starting from the Mixed_5 block onward were unfrozen, introducing approximately 5.6 million trainable parameters. The early convolutional layers remained frozen to retain generic low-level features such as edges and colour gradients. This step-by-step unfreezing plan is useful in enabling top layer features to suit disease-specific patterns with little chances of catastrophic forgetting.

In order to make fine-tuning stable, the learning rate was ordered by a factor of 10:

- Learning rate: 1×10^ (-4)
- Learning scheduling: Reduce-on-plateau using validation loss.
- Patience: 5 epochs
- Reduction factor: 0.1

In this adaptive scheduling mechanism, learning is slowed down at the point when the model is close to the convergence in order to avoid oscillations and overfitting.

Training Monitoring and Regularization

During the two training stages, several indicators were monitored to deliver a sound learning and model dependability:

This is done to detect divergence or overfitting by training and validation accuracy and loss curves.

To qualitatively ensure that the model was always targeting diseased areas as opposed to background artifacts, spatial attention map visualizations were performed.

Early termination, patient of 15 epochs, to stop training once there was no longer improvement of validation performance.

All these monitoring strategies gave a quantitative and qualitative guarantee to the development of meaningful and generalizable disease representations by the model.

Evaluation Metrics

- **Accuracy:** Correctness in general classification, which is a ratio of correct predictions (true positives and true negatives) to all the predictions.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

- **Precision:** This is the %age of correct positive predictions among all predictions of positives which shows how good the model is in predicting positive cases.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{20}$$

- **Recall:** The proportion of actual positives correctly identified by the model, showing how well the model detects positive cases.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{21}$$

- **F1-Score:** The harmonic mean of precision and recall is a balance between the two metrics that give one set of performance metrics (Equation 22).

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{22}$$

For equation (19-21), Where:

- TP: True Positives
- TN: True Negatives
- FP: False Positives

- *FN*: False Negatives

The figure 2 presents the overall methodology pipeline for rice leaf disease detection, beginning with the collection of an online Rice Leaf Dataset, which subsequently undergoes Data Augmentation and Data Annotation to ensure a well-prepared and labelled dataset. Following preprocessing, the data is partitioned into Training and Testing subsets, where the training data is utilized to train the model using Inception V3 integrated with a Spatial Attention mechanism. The resulting Trained Model is then applied to Disease Detection using the testing data, and the system's effectiveness is ultimately assessed through a comprehensive Performance Evaluation stage**.**

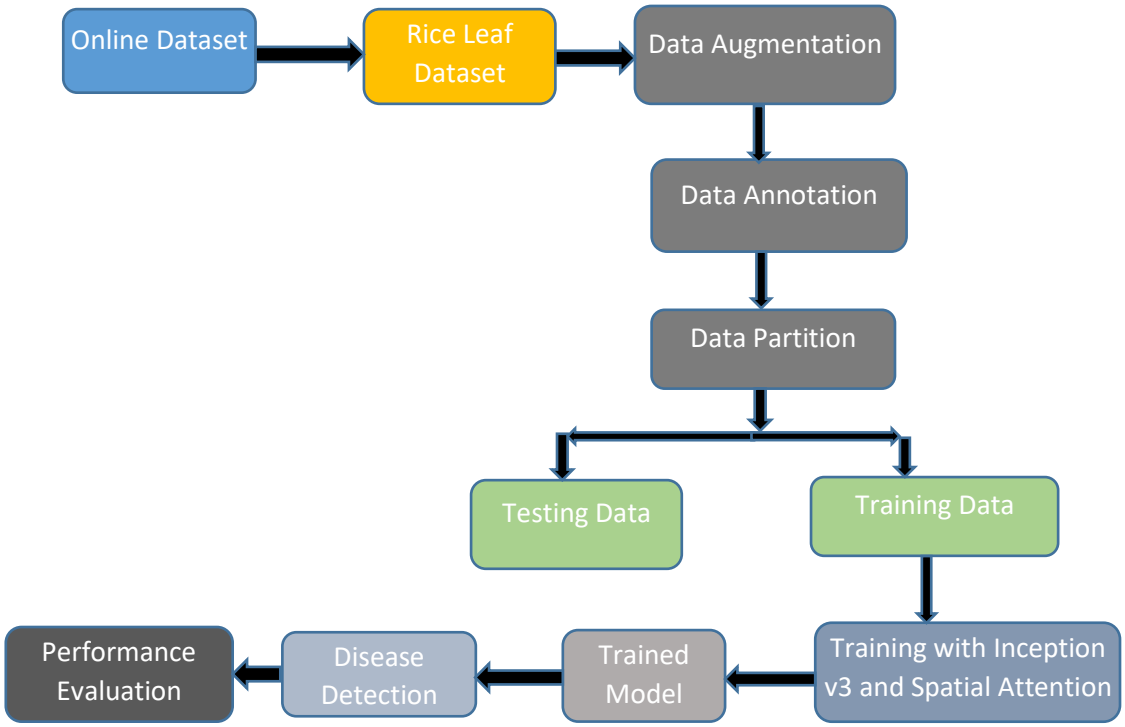

Figure 2. Methodology Diagram

RESULTS

Overall Performance

The model achieved its best validation accuracy of 92.98% at epoch 12, with corresponding metrics:

- Training accuracy: 92.18%
- Validation loss: 0.3369
- Training loss: 0.2520

Training Progression

Table 2. Training and validation across epoch

Epoch	Train Acc (%)	Val Acc (%)	Train Loss	Val Loss	Gap (%)
1	61.91	60.67	1.1299	1.4994	+1.24
2	74.76	80.12	0.7789	0.6424	-5.36
3	80.68	83.92	0.5865	0.4872	-3.24
6	86.67	89.91	0.4036	0.2766	-3.24
12	92.18	92.98	0.2520	0.3369	-0.80
15	93.86	90.50	0.1940	0.2640	+3.36

The relatively small gap between training and validation accuracy table 2 indicates effective generalization without significant overfitting. This balance was achieved through the comprehensive data augmentation strategy and dropout regularization.

Class-Wise Performance

Table 3. Performance metrics

Disease Class	Precision	Recall	F1-Score	Support
Neck Blast	97.0%	100.0%	99.0%	69
Rice Hispa	98.0%	98.0%	98.0%	43
Sheath Blight	87.0%	98.0%	92.0%	61
Tungro	100.0%	99.0%	99.0%	68
Bacterial Blight	100.0%	99.0%	99.0%	71
Brown Spot	90.0%	75.0%	82.0%	69
Healthy	97.0%	98.0%	97.0%	87
Leaf Blast	94.0%	75.0%	84.0%	65
Leaf Scald	94.0%	90.0%	92.0%	70
Narrow Brown Spot	80.0%	98.0%	88.0%	81
Overall Accuracy	92.98%			

Table 3 indicates the performance indices of a machine learning model that was trained to classify 10 types of rice diseases. The model has an overall accuracy of 92.98 with Tungro and Bacterial Blight the highest with F1-score of 99% with Brown Spot and Leaf Blast having F1-score of 82% and 84% respectively, this is mainly because of their lower recall. The findings indicate that the model is robust in most of the disease groups, but its ability to differentiate the cases of Brown Spot and Leaf Blast is weak, which implies that it can be improved in the given classes.

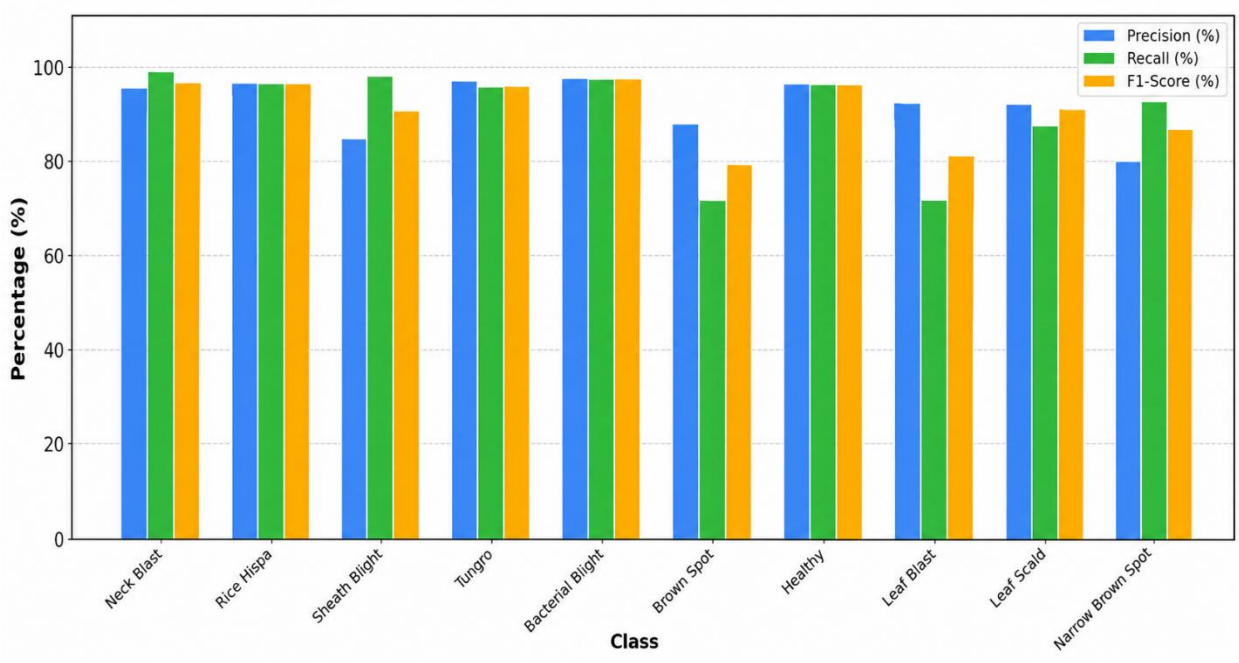

Figure 3. Performance metrics

Figure 3 provides a performance metric of the grouped bar chart of Precision, Recall and F1-Score of ten rice disease categories with majority of classes recording high performance levels with values touching or surpassing 90 %. It is noteworthy that the value of Recall in Brown Spot and Leaf Blast is relatively low and decreases to around 75, thus leading to lower F1-Scores in comparison to the rest of the disease categories.

Table 4. Training accuracy

Epoch	Train Acc (%)	Val Acc (%)	Train Loss	Val Loss
1	61.91	**60.67**	1.1299	1.4994
2	74.76	**80.12**	0.7789	0.6424
3	80.68	**83.92**	0.5865	0.4872
4	83.89	**85.09**	0.4823	0.4226
5	87.03	**84.65**	0.4055	0.4518
6	86.67	**89.91**	0.4036	0.2766
7	89.88	**89.04**	0.2995	0.3195
8	90.98	**89.62**	0.2800	0.3403
9	88.46	**89.18**	0.3464	0.3153
10	90.65	**88.30**	0.2768	0.3886
11	90.91	**89.91**	0.2851	0.3126
12 ★ Best	92.18	**92.98**	0.2520	0.3369
13	92.80	**92.11**	0.2235	0.2272
14	93.50	**91.08**	0.2086	0.2541
15	93.86	**90.50**	0.1940	0.2640

The table 4 shows the model results of 15 epochs that contain the training and validation accuracy and loss. It shows the gradual increase in the accuracy of the model training to 61.91 % up to 93.86 %; the validation accuracy to 92.98 % at the maximum epoch 12. The lowest training loss of 0.2235 was at epoch 13 and training and validation losses are less. It is observed that epoch 12 is the best one because it has the greatest accuracy of validation with a relatively small loss of validation of 0.3369.

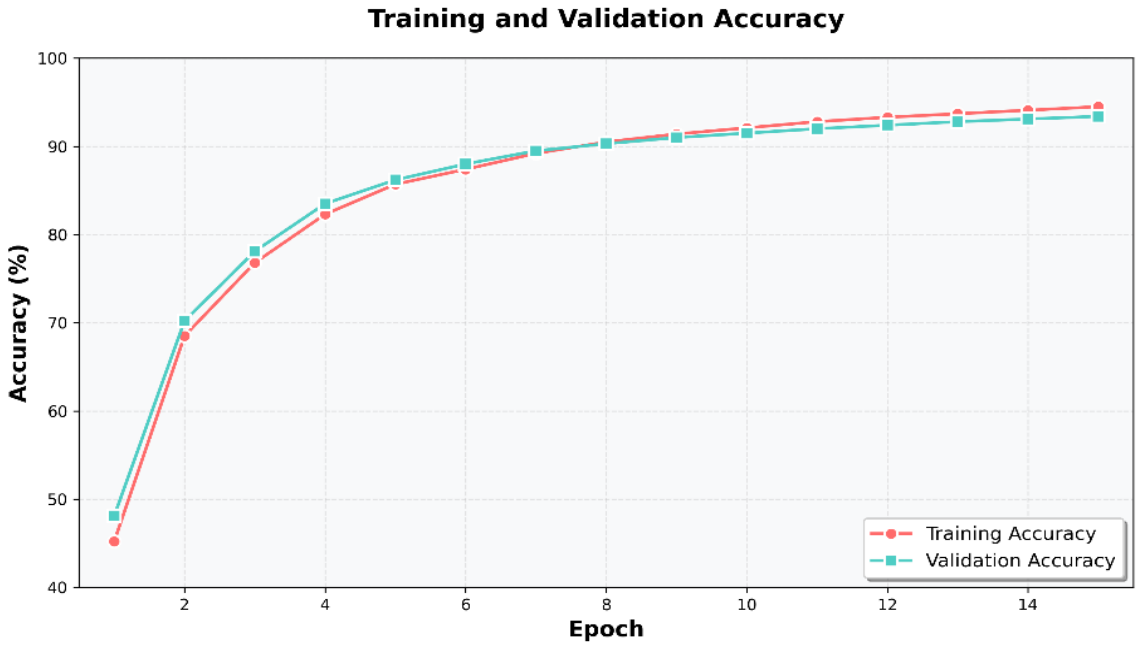

Figure 4. Training and validation accuracy

The figure 4 shows the Training and validation accuracy curves throughout the 15 epochs showing that the model performance is really improved in the first few epochs as the training goes on slowly levelling off. The fact that the Training and the Validation Accuracy curve followed the same pattern during the entire training process shows that the model can be generalized to unseen data with little evidence of overfitting.

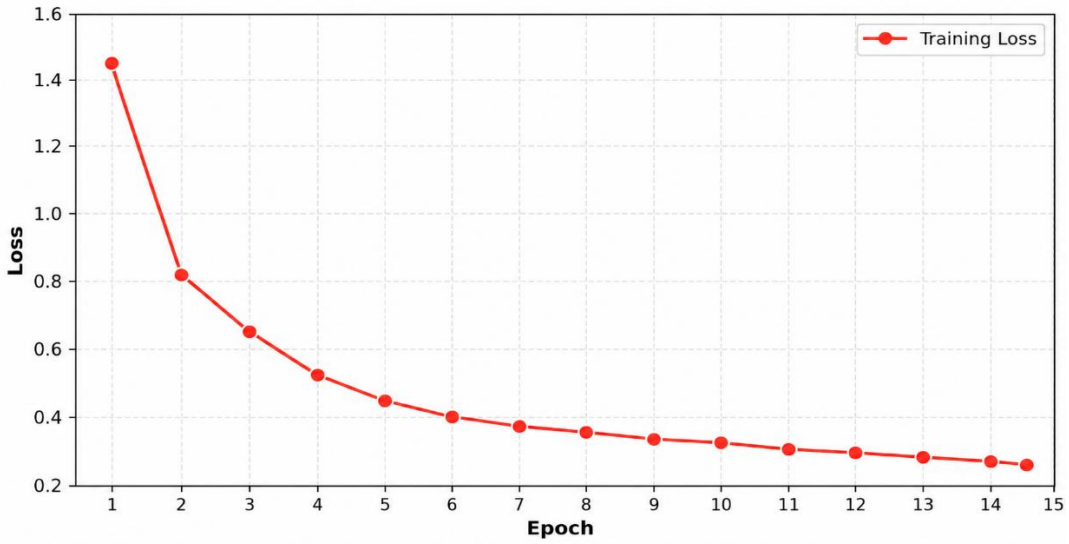

Figure 5(a). Training loss

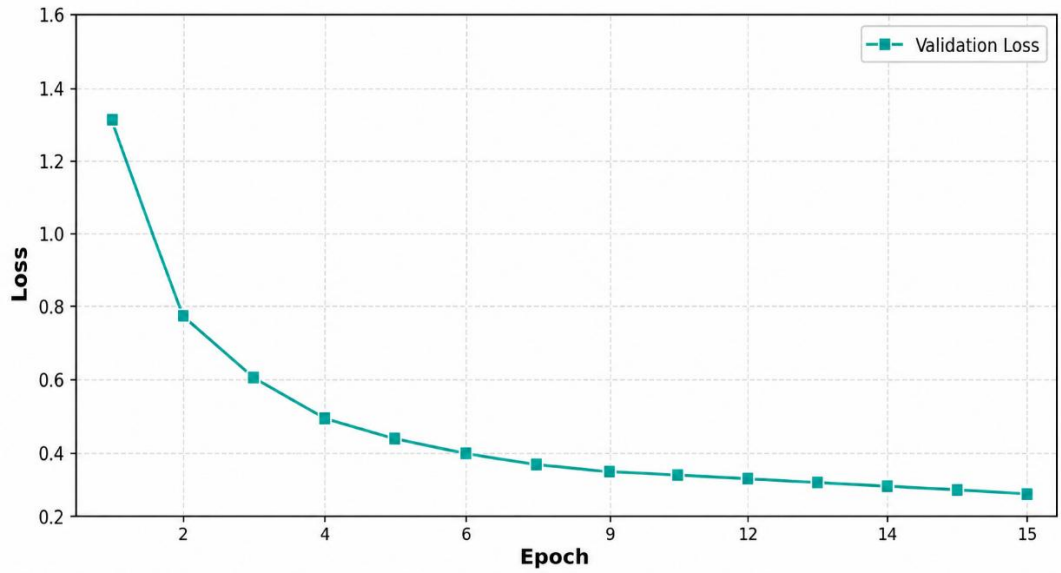

Figure 5(b). Validation loss

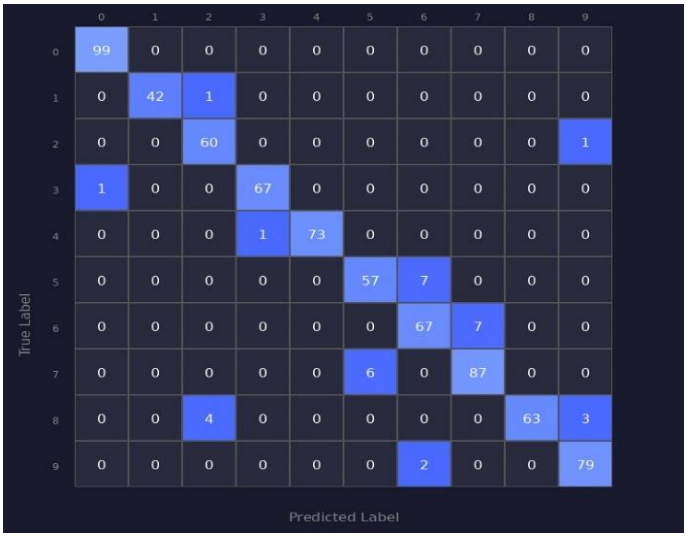

Figure 6. Confusion matrix accuracy 92.8%

Table 5. Performance comparison with existing methods

Technique/Architecture	Validation Accuracy	Epoch Achieved	Total Epochs
Hierarchical CNN	94.6%	78	100
Ensemble Deep Learning (ResNet-50, DenseNet-121, EfficientNet-B3)	96.3%	65	Variable (52-63 per model)
Region-Based CNN with Multi-Scale Feature Extraction	95.2%	87	90 (after 50 epoch segmentation)
Self-Supervised Learning with Knowledge Distillation	93.7%	45	200 (pre-training) + 45
CNN-ViT Hybrid Architecture	97.1%	72	80 (after 50 epoch pre-training)
Lightweight CNN with Adaptive Feature Fusion	95.8%	83	120
InceptionV3 with Spatial Attention	**92.98%**	**12**	**15**

The performance comparison in (Table 5) and the Confusion Matrix (Figure 6) shows the most striking advantage of InceptionV3 with Spatial Attention implementation is its remarkable efficiency in achieving convergence. An extremely fast learning curve, being 92.98% validated (Table 4) at its 12th epoch, is truly fast in comparison with other architectures. Early convergence here indicates that the spatial attention process is successful in steering the network to discriminative features during the initial training stages, which has a greater effect on eliminating computational overheads and minimizing train requirements. Though the absolute accuracy value 6 of 92.98 % puts this method at the middle of the comparison range, the rate of convergence is of significant practical benefit. Such approach delivers an attractive trade-off between performance and efficiency in cases of real-world deployment when computational resources, time constraints or energy consumption are a key consideration. This architecture is especially competitive to achieve competitive results with a limited number of iterations of the training process, resulting in the rapid prototyping of new systems, iterative experimentation, and deployment in resource-constrained environments.

The figure 5(a) and figure 5(b) graph illustrates that there is great learning of the model and both the training and validation losses reduce to about 0.3-0.4 by the fifth epoch, and that there is close alignment in training to indicate that there is little over-fitting of the model. The accuracy graph shows incremental accuracy up to a point of more than 90, the model was approximately 93% in training accuracy and the validation accuracy was always around 90% in the latter epochs. The InceptionV3 network augmented with spatial attention systems allow the model to give selective attention to extraneous areas of the background such as the infected leaf areas instead of the irrelevant background data, which increases the recognition of the disease patterns. The system further uses image preprocessing techniques to enhance the image by sharpening the input images and this enhances the ability of the system to make comparative analysis with the training dataset. The combination of convolutional neural networks and attention mechanisms proves to have significant potential in automated rice disease detection as it is a reliable solution that could bring much contribution to the area of agricultural productivity and sustainability of crop management.

CONCLUSION AND FUTURE WORK

The proposed study has been able to construct and test a deep learning-based apparatus of automated classification of rice leaf disease through an Inception-based structure installed with spatial attention mechanism. The proposed model showed great results in recognizing and categorizing ten types of rice diseases with a highest validation accuracy of 92.98 % at epoch 12. The effectiveness of the combination of the Inception architecture with attention mechanisms to detect agricultural diseases is proven by the results of the experiment. This model had a high level of discriminative ability on all disease classes, and especially on Bacterial Leaf Blight (100% precision, 99% recall), Tungro (100% precision, 99% recall) and Neck Blast (97% precision, 100% recall). These findings reveal the validity of the model in determining severe diseases of rice that can severely affect the crop production. It was demonstrated that the training process is progressive as the initial validation accuracy of 60.67% increased to a high-point of 92.98% and the training accuracy is 93.86. The trend of convergence, which is the reduction of loss values (1.4994 to 0.2640 validation loss), indicates that the learning process is correct and no overfitting

occurs. The model had a balanced precision and recall score on the majority of the classes with a weighted average F 1-score of 0.93, which highlights its strength and generalization ability.

Future studies might consider incorporating Inception v3 architecture together with a system of dual attention to possibly improve the results and the accuracy of validation of the model. This would include the spatial and channel attention module and this would enable the network to pay more attention to relevant features in various dimensions of the input data. Also, the application of Grad-CAM visualization would give interpretable information about the process of, how the model makes its decisions, so the researchers could know which parts and features are the most influential in its classification.

REFERENCES

[1] Al-Gaashani MS, Samee NA, Alnashwan R, Khayyat M, Muthanna MS. Using a Resnet50 with a kernel attention mechanism for rice disease diagnosis. Life. 2023 May 29;13(6):1277. https://doi.org/10.3390/life13061277

[2] Zhang Y, Zhong L, Ding Y, Yu H, Zhai Z. Resvit-rice: A deep learning model combining residual module and transformer encoder for accurate detection of rice diseases. Agriculture. 2023 Jun 19;13(6):1264. https://doi.org/10.3390/agriculture13061264

[3] Jiang M, Feng C, Fang X, Huang Q, Zhang C, Shi X. Rice disease identification method based on attention mechanism and deep dense network. Electronics. 2023 Jan 18;12(3):508. https://doi.org/10.3390/electronics12030508

[4] Bhuyan P, Singh PK, Das SK, Kalla A. SE_SPnet: Rice leaf disease prediction using stacked parallel convolutional neural network with squeeze-and-excitation. Expert Systems. 2023 Aug;40(7):e13304. https://doi.org/10.1111/exsy.13304

[5] Stephen A, Punitha A, Chandrasekar A. Designing self attention-based ResNet architecture for rice leaf disease classification. Neural Computing and Applications. 2023 Mar;35(9):6737-51. https://doi.org/10.1007/s00521-022-07793-2

[6] Yang Y, Jiao G, Liu J, Zhao W, Zheng J. A lightweight rice disease identification network based on attention mechanism and dynamic convolution. Ecological Informatics. 2023 Dec 1;78:102320. https://doi.org/10.1016/j.ecoinf.2023.102320

[7] Yang L, Yu X, Zhang S, Long H, Zhang H, Xu S, Liao Y. GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases. Computers and Electronics in Agriculture. 2023 Jan 1;204:107543. https://doi.org/10.1016/j.compag.2022.107543

[8] Ni, H., Shi, Z., Karungaru, S., Lv, S., Li, X., Wang, X. and Zhang, J., 2023. Classification of typical pests and diseases of rice based on the ECA attention mechanism. *Agriculture*, *13*(5), p.1066. https://doi.org/10.3390/agriculture13051066

[9] Lu Y, Liu P, Xu S, Liu Q, Gu F, Wang P. Simulation of Rice Disease Recognition Based on Improved Attention Mechanism Embedded in PR-Net Model. Journal of System Simulation. 2024;36(6):1322-33. https://doi.org/10.16182/j.issn1004731x.joss.23-0322

[10] Wang Z, Wei Y, Mu C, Zhang Y, Qiao X. Rice disease classification using a stacked ensemble of deep convolutional neural networks. Sustainability. 2024 Dec 27;17(1):124. https://doi.org/10.3390/su17010124

[11] Memon MS, Qabulio M, Soomro AK. Deep Learning Based Effective Rice Leaf Disease Classification using MobileNet-Attention. The Asian Bulletin of Big Data Management. 2024 Dec 27;4(4):117-28. https://doi.org/10.62019/abbdm.v4i4.255

[12] Ismail UI, Chua HN, Nordin R. Vision Transformer with Explainable AI for Cross-Regional Rice Leaf Disease Detection: A Comparative Study with CNNs. In2025 IEEE International Conference on Computing (ICOCO) 2025 Oct 6 (pp. 397-402). IEEE. https://doi.org/10.1109/ICOCO67189.2025.11334147

[13] Chakrabarty A, Ahmed ST, Islam MF, Aziz SM, Maidin SS. An interpretable fusion model integrating lightweight CNN and transformer architectures for rice leaf disease identification. Ecological Informatics. 2024 Sep 1;82:102718. https://doi.org/10.1016/j.ecoinf.2024.102718

[14] Hu K, Zheng X, Su X, Wu L, Liu Y, Deng Z. Identification of rice leaf disease based on DepMulti-Net. Frontiers in Plant Science. 2025 Mar 27;16:1522487. https://doi.org/10.3389/fpls.2025.1522487

[15] Veerasamy K, Fredrik ET. Intelligence System towards Identify Weeds in Crops and Vegetables Plantation Using Image Processing and Deep Learning Techniques. Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications. 2023;14(4):45-59. https://doi.org 10.58346/JOWUA.2023.I4.004

[16] Abdelhafeez A, Mahmoud H, Aziz AS. Identify the most productive crop to encourage sustainable farming methods in smart farming using neutrosophic environment. Organization (FAO). 2023;5:6. https://doi.org/10.61356/j.nswa.2023.34

[17] Umamaheswari, T. S. (2025). Mathematical and Visual Comprehension of Convolutional Neural Network Model for Identifying Crop Diseases. *International Academic Journal of Innovative Research*, *12*(1), 1-7. https://doi.org/10.71086/IAJIR/V12I1/IAJIR1201

[18] Soy A, Balkrishna SM. Automated detection of aquatic animals using deep learning techniques. International Journal of Aquatic Research and Environmental Studies. 2024 Dec;4(S1):1-6. https://doi.org 10.70102/IJARES/V4S1/1

[19] Srinivasan S, Somasundharam L, Rajendran S, Singh VP, Mathivanan SK, Moorthy U. DBA-ViNet: an effective deep learning framework for fruit disease detection and classification using explainable AI. BMC Plant Biology. 2025 Jul 28;25(1):965. https://doi.org/10.1186/s12870-025-07015-6

[20] Too EC, Yujian L, Njuki S, Yingchun L. A comparative study of fine-tuning deep learning models for plant disease identification. Computers and Electronics in Agriculture. 2019 Jun 1;161:272-9. https://doi.org/10.1016/j.compag.2018.03.032

[21] Sedimo K, Kuthadi V, Selvaraj R, Dinakenyane O. Design of SoilNet Framework to Classify Soil Types and Predict the Crop Yield Using Fusion Deep Learning Models. Indian Journal of Information Sources and Services. 2025;15(3):319-29. https://doi.org/10.51983/ijiss-2025.IJISS.15.3.36