International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Enhanced Rice Leaf DISEASE Detection and Categorization Using Hybrid Convolutional Vision Transformer (CVT) Model with Spatial Attention

G. Shyning Sobinsa, R. Sheeba

Submitted:05/11/2024 **Revised**: 19/12/2024 **Accepted:** 28/12/2024

Abstract: Diagnosis and classification of disease in rice leaf are essential in countering diseases in crops and enabling agricultural sustainability. Conventional methods that mostly involve manual checks are constrained by the intensity of labor, inconsistency and sensitivity to errors. The proposed paper presents a Hybrid Convolutional Vision Transformer (CVT) with Spatial Attention (SA) to improve the accuracy of detection and reliability of the classification in rice leaves. The suggested CVT architecture combines a Convolutional Neural Network (CNN), which serves as the basis of the first features extraction with a Vision Transformer (ViT) in the advanced feature representations. Convolutional Neural Network learns the most critical texture and shape, whereas Vision Transformer learns the attention over patches of an image and effectively learns the complicated spatial relationship required to recognize disease-specific features under a wide variety of field conditions. In addition, SA module further optimizes the model by giving more weight to diseased areas to eliminate interference by non-leafbackground areas. Examples with rice leafdataset show that the hybrid CVT with SA model reaches an average feature extraction error of over 98.12, a 98.56, and 98.26 feature extraction error and classification error respectively on dataset 1 and dataset 2, respectively, and utilizing multiple rice leaf classes, as opposed to baseline CNN and ViT models. In the decision making process, Spatial Attention uses heat maps to come up with a significant location that increases the interpretation of the model. This CVT hybridized system offers an extensible platform of rice leaf disease detection and classification that can be added into security in agriculture, like drone cameras to investigate remote fields. The model demonstrated is better performing than any other existing approach.

Keywords: *Agriculture, Convolutional Neural Network, Hybrid CVT, Rice leaf disease, Vision Transformer*

I Introduction

An efficient diagnosis and treatment of leaf diseases are required to achieve high and consistent yield of rice. Detection of plant diseases is also increasing faster than detection ability of current manual methods of identification, which are ineffective and yield inaccurate diagnosis. That is why it is necessary to create a clear and efficient automatic detection model of the premium rice yield [1]. Furthermore, indiscriminate use of insecticides to remove plant has an impact on ecosystem and rice quality. Thus, thorough rice leaf disease identification serves as the basis for avoiding and removing disease. At the same time, rice quality and production has to be improved while protecting the environment [2-4]. Recognizing plant is critical for agricultural production, since leaf catastrophes are recognized as the primary cause of crop output decrease. It takes time and effort to manually locate and identify disease. Agricultural technicians use hand lenses and microscopes to inspect and identify disease that are visible to the naked eye. This approach requires continuous crop monitoring. In large farms, this is an expensive, subjective, and labor-intensive task [5]. Conventional disease identification systems attempt in developing feature vectors to identify specific disease, resulting in the lack of generalizability [6]. Conversely, Deep Learning (DL)-based systems require pre-built object detection since the performance difference between generic object recognition and infectiondetection is rather significant. One could

characterize this difference as variations in object properties and detecting criteria.

Because to DL's excellent feature learning and extraction capabilities, as well as more general advances in Artificial Intelligence (AI), advanced farming settings [7] has become much easier in terms of finding disease in various plants. On the study of rice illness [8], a Deep Convolutional Neural Network (DCNN) is utilized to suggest theactualcategorization technique for nine different types of rice infection. Using the RiceBioS app [9], biotic stress is recognized in rice leaf such as rice blast and bacterial leaf rot. A DL classification system is developed in smart phone with rice sickness detection app.

Deep Learning approaches for image recognition have lately undergone major advancements. Strong performance on imagecategorization tasks by Convolutional Neural Networks (CNN) including Residual Network (ResNet) [10], Google Network (GoogLeNet) [11], Image Network (ImageNet) [12], Visual Geometry Group (VGG) [13] and Le Cun Network (LeNet) has improved the knowledge and application of Neural Networks (NN). Several researchers are focusing on developing a technique for detecting diseases in rice.

Although the traditional CNN-based rice disease detection model [14], [15] clearly extracts attributes of diseases, there are certain shortcomings in the present research. For starters, the visual backdrops are simple and the researchers use a limited number of ailments. Because of their complicated character, the presented models only go so far in generalizing rice leaf diseases. The existing models have not been sufficient in capturing the characteristics of diseases displayed in images. By extracting various meaningless features from the images, they regularly induce overfitting and poor model performance. Finally, models of disease detection in riceleaves are not particularly useful. The majority of research neither addresses any practical issues nor theoretical considerations. In another way, the current models are too big to be used on websites or mobile devices [16], [17].

Researchers have developed computer vision [18] algorithms using vision transformers [19], [20] to detect and quantify agricultural illnesses. Plants appear in digital images taken for disease identification. Backgrounds in the images of rice leaves diseases are been either simple or complex. A novel spatial Feature-Enhanced Attention (FEA) module [21] is created to increase backbone network performance for fine-grained image classification tasks.

This study uses Kaggle data to improve disease detection and classification in rice plants by combining hybrid Convolutional Vision Transformer (CVT) models with Spatial Attention (SA). While the transformer component provides global perspective by revealing connections throughout the whole image, convolutional layers of CVT model successfully capture local properties such as shapes and textures of leaves. SA also removes the unneeded background noise thereby increasing the areas that the model pays significant attention to such as disease-affected areas. With this integration, accuracy is enhanced thus enabling a high level of recognition and classification in the tough agricultural environment.

Recently the modern ML algorithms utility has made the diagnosis of rice sickness more accurate and efficient. The especially successful ones are hybrid CVT [22] models with SAmechanisms. These models are an integration of the global contextual awareness of transformers and the capacity of convolutional layers to get localized information. The model further improves its capacity to identify many plant diseases by incorporating SA which prioritizes the key image characteristics and reduces noise. This method is more accurate and higher recall than the common convolutional models and single transformers. Integrating the two technologies offers a novel level of disease management in the agricultural systems wherein they will be handled much faster and more effectively.

The study is inspired by the dire necessity to address the issues of rice leaf disease detection and classification, which is vital in ensuring food security in the world. The old methods of disease detection are time consuming, prone to errors and are not scalable. With the increasing access to high-resolution agricultural imaging data, novel, automated systems that are capable of precision and classification of disease in challenging conditions are strongly needed. This research will minimize this gap by employing hybrid CVT with SA models, which are more precise in feature extraction and deny better interpretability of the model. Besides the boundaries set to the detection

technology of diseases, this research is aimed at providing the farmers with instruments that will encourage sustainable disease management, thus, enhancing agricultural production.

Structure of the paper: The paper is divided into the following sections: The related works are discussed in section 2. Section 3 explains the presented methodologies and the section 4 discusses the results and discussions. Section 5 is the concluding part of this work.

II Related Works

Wijayanto et al. [23] was aimed at examining the successful gathering of data by drones in the image analysis of West Javan rice fields which were correlated with thermal and textual parameters to enhance the detection of Bacterial Leaf Blight (BLB). Researchers used drone data and powerful Machine Learning (ML) tools to investigate BLB impact levels. The normalized difference image proved to be quite effective because it included Haralick image features. Their findings revealed that when compared to the traditional approaches relied solely on spectral indices, incorporated image cues greatly enhanced disease diagnostic accuracy. Using vegetation and image clues, Random Forest (RF) technique achieved very high classification accuracy. The rice plant disease detection system aimed to give accurate and timely disease predictions in aiding crop disease management. Early disease identification in rice plants allowed farmers and paddy researchers to respond swiftly for safeguarding the harvest. Daniya et al. [24] presented an overview of image processing as well asML strategies for disease detection in rice plants. These authors retrieved the images of disease rice plant leaves by combining multiple segmentation algorithms.

Li et al. [25] Brown spot, rice sheath blight and stem borer were all researched utilizing the video footage. The authors used image-training techniques to detect previously unseen movies. Their laboratory findings supported the proposed video detection approaches. Their technique to identify experimental design had been ideal for developing a Deep Convolutional Neural Network (DCNN) to recognize rice videos. Using Visual Geometry Group 16 (VGG16), ResNet-50 and ResNet-101 detection of strongly blurred images was ineffective. In similar line, same detection was used to run state-of- the-art YOLOv3 on rice lesion patches. It discovered the underperformance when the objects were shown with non-standard forms and fuzzy borders.

Patil et al. [26] developed the Rice-Fusion framework, a system for autonomous rice disease diagnosis in agriculture using AI and multimodal data fusion. Their research focused on three different diseases such as rice blast, bacterial blight and brown spot. Moreover, the healthy group received one type. The collection was unique with 3600 visuals, modalities and ambient features. Earlier and delayed fusion approach was one form of fusion paradigm that incorporated both modalities. Deep LearningNN served as the system's foundation, necessitated a large number of data samples for accurate model training. With an accuracy rate of 95.31%, their study showed that the Rice-Fusion data fusion approach outperformedthe current unimodal approaches such as CNN and MLP topologies. Table 1 compares the varied methods used in rice plant detection.

Precision agriculture was the field of optimizing the farming field by utilizing advanced technology to assess and develop crop health. Earlier intervention against crop losses was essential in this domain with the task of plant disease detection. Conventional, current methods relied on manual labor or basic image tools that were slow, prone to human errors. In AI, the emergence of Vision Transformers became a transformative way of performing efficient and accurate analysis of plant diseases from complex visual datasets. Parez et al. [27] showed how visual intelligence reinvented industries in agriculture, solidifying sustainable farming through next generation, AI powered software.

Rice diseases affect the leaves have a big threat to food security all over the world hence need better methods of identifying the diseases. The usual visual techniques used in identifying diseases slow and less accurate; thus the development of Auto-ML. The idea of AI has come to the frontline when it comes to the processing of large amounts of data and this has made it easy to classify rice leaf diseases. Overall, Mukherjee et al [28] provided a detailed overview of multiple approaches in ML application that highlight both strengths and weaknesses of the techniques with the focus on their applicability to agriculture. These

authors is expected to helped to align research regarding rice disease control systems to be more sustainable and effective, so scale and reliability can be achieved in the field of agriculture.

Rice diseases controls were necessary to guarantee a stable crop production, especially in rice growing areas that were highly based on rice as food staple. Traditional ways of monitoring were laborious and commonly failed to find out the problems earlier, caused huge crop losses. Recently, Deep Learning (DL) advanced in particular CNN had created new opportunities for automated, accurate and scale disease detection. The purpose of Rahman et al. [29] study was to understand the application of CNN in identifying rice diseases, demonstrated AI driven solutions in transforming precision agriculture and diminishing manual intervention.

Russakovsky et al. [30] ImageNet Large Scale Visual Recognition Challenge (ILSVRC) greatly influenced the status quo in computer vision and DL, due to the availability of large scale dataset and competition framework on visual recognition tasks. ImageNet consisted of a richly annotated dataset of images from which algorithms for large scale image classification, object detection and localization had been developed and evaluated. The pressure to overcome the DL limitations had pushed the boundary of DL enough to highlight DCNNs as the greatest breakthrough approach helped by AlexNet architecture.

2.1 Problem Identification

Infestations in rice fields are very sensitive, affecting both quality and yield. Current rice plant disease identification methods rely on standalone DL models, which exhibit significant intra-class similarity and inter-class heterogeneity in rice leaf images. These limits reduce accuracy and efficiency, especially in real-world agricultural contexts where the data is noisy and unstructured.

III Materials And Methods

3.1 Dataset collection

To assemble the necessary information on various rice leaf diseases, the researchers has investigated and evaluated rice disease records as well as conducted interviews with Department of Agriculture—particularly those from the Regional Crop Protection Center. Images of several rice plant diseases are collected using the means available, which included digital cameras and smart phones. After gathering, all the imagesare pre-processed and included in the dataset.

Dataset 1: https://www.kaggle.com/datasets/nashehannafii/datasetleafblast

Dataset 2: https://www.kaggle.com/datasets/vbookshelf/rice-leaf-diseases

3.2 Data preprocessing

To provide high-quality input for model training, preparing a dataset for rice leaf detection involves many critical steps. Image data is initially cleaned by removing duplicates and low-quality samples such as blurry or poor lit images. The data is further labeled to represent various bug types or groups. Resizing and normalizing the images ensures that the input dimensions and pixel intensity ranges remain constant. While image improvement enhances the original dataset quality, augmentation increases the dataset's size. Increasing visual detail as well as smoothening by flattening and contracting the images. Rotation, flipping and cropping are some of the augmentation strategies used to diversify datasets and improve model robustness. These operations increase the dataset's quality, allowing for effective training of the detection model.

3.3 Hybrid Convolutional Vision Transformer with Spatial Attention

Wu et al. [31] two significant enhancements in hybrid CVT include a new convolutional token embedding in the transformer hierarchy and a convolutional transformer block with a convolutional projection. Two convolutional approaches are used in ViT architecture namely; convolutional network for projecting and embedding tokens. First, convolutional token is the embedding layer tokens constructed to fit the 2D spatial grid and performs convolution using overlapping patches. The degree of overlap is determined by stride length. Token standards are more advanced. While the number of features (feature dimension) increases with each level, number of tokens or feature resolution decreases. This produces spatial downsampling and improved representation richness, as in CNN architecture. A fully linked MLP head is used as last resort before the output classification predicts the class. The

convolutional token embedding layer has to be considered first. Next, a convolutional projection is created, which shows the successful controlling of computational costs in the Multi-Head Spatial-Attention module.

3.3.1 Convolutional Projection for Attention

To increase the efficiency, suggested Convolutional Projection layer allows undersampling of K and V matrices in order to represent the local spatial environment better. The suggested Transformer block is an adaptation of the original, which allows convolutional projection. Previous steps included more complex designs for convolution modules intended for the integration into transformer block, which increased computation costs. The model is replacing the original position-wise linear projections with depth-wise separable convolutions for constructing the Convolutional Projection layer for Multi-Head Spatial-Attention (MHSA).

Implementation Details

Figure 1 shows position-wise linear projection of ViT. Convolutional projection uses a kernel-sized convolutional layer that is depth-wise separable. Finally, a one-dimensional framework is constructed to accommodate future token storage and usage. However, another option is flattening the layer is given in equation (6):

$$x_i^{\frac{q}{k}/v} = Flatten(conv2d(Reshape2D(x_i), s) \text{ ---------- (6)}$$

Apply the following guidelines: Both depth-wise and batch-wise separable convolution is possible. $x_i$at layer i is the unaltered token before the convolutional projection, whereas *Conv2d* defines the size of the convolution kernel. The *q/k/v* matrix token input is $x_i^{\frac{q}{k}/v}$ i. By combining the previous position-wise linear projection layer with a 1 x 1 kernel size convolution layer, one can easily construct the updated Transformer Block with the Convolutional Projection layer. Convolutional projections are depicted in figure 2.

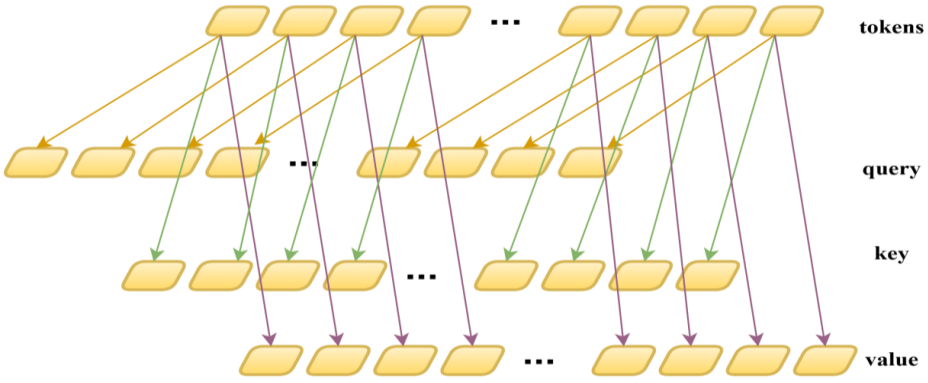

Figure 1: ViT linear projection

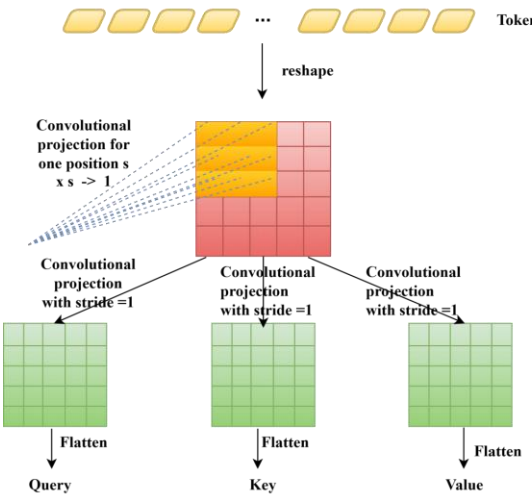

Figure 2: Convolutional projection

3.3.2 CVT with SA

Disease infections particularly in vital crop rice become the significant barriers to agricultural productivity. Reducing losses and ensuring sustainable farming systems rely on fast and accurate infection detection. Deep Learning has been recently improved forproducing hybrid methods likeCVT, which combines the capabilities of ViT with CNNs. These models are extremely useful for identifying rice leaf disease in complex agricultural contexts, since SA processes enhance these methods. The method's paradigm aims to address the shortcomings of current techniques by utilizing local feature extraction and transformers to capture long-range correlations in visual input via CNN. Convolutional layers successfully extract low-level information from leaf images such as edges and textures. Transformer layers, which usespatial-attention to encode global contextual links, assess these feature maps later. Spatial Attention systems increase the model's capability by prioritizing the image. The architecture of CVT with SA is defined by three key components: SA modules, transformer encoders and convolutional blocks. To improve generality, diseasedimages are first resized, normalized and data augmented. Convolutional layers use preprocessed images to generate feature maps. A convolutional operation is mathematically stated as *F=conv (I; W);F* is the generated feature map,I is the input image and W is the convolution kernel weight.After accepting these flattened feature maps, transformer encoder uses the following equation (7) to compute spatial-attention:

$$A(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \text{ ---------- (7)}$$

Feature maps create three matrices: value from V, query from Q and key from K. The key's dimension is denoted as d_k. Combining several images aids the model in identifying diseases on rice leaves by capturing global trends.The SA module creates an attention map $M_s$that highlights disease regions in the feature map.Thus, the operation is enhanced. The map is computed as follows in equation (8):

$$M_s(x, y) = \frac{\sum_c |F_c(x,y)|}{\sum_{x,y,c} |F_c(x,y)|} \text{ -------- (8)}$$

Where $F_c(x, y)$ represent the channel c activity at the spatial position x and y. This normalized weighting ensures that the next layers emphasize regions with high feature intensity. Finally, CVT model generates bounding box localization coordinates as well as categorization labels that represent various diseases. Using CVT with SA for rice leaf disease detection requires many processes. First, data is collected and interpreted. Rotation, flipping and cropping are some of data augmentation strategies used to simulate the real-world scenarios and reinforce the model. For the tasks likedisease location and categorization, model is trained with task-specific loss functions and Adam optimizer. Loss function L, commonly aggregates cross-entropy loss. Mean squared error is calculated in equation (9):

$$L = L_c + \lambda L_l \text{ ------ (9)}$$

Where,λ balances the importance of two tasks. Loss function for classification and bounding box linear egression are represented as $L_c$andL_l.

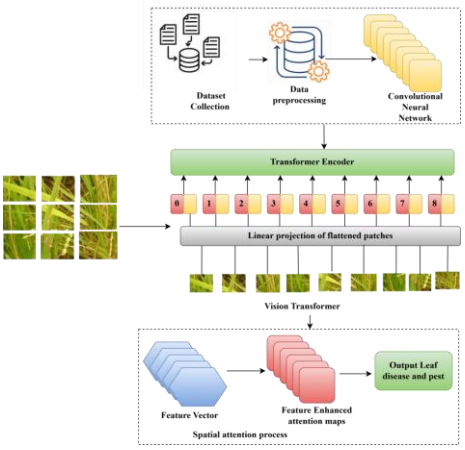

Figure 3: Overall architecture of rice leaf disease detection

The entire architecture of the proposed frame work is displayed in figure 3. The model performs well throughout testing, demonstrating high accuracy and durability in challenging scenarios like occlusion and shifting light. Convolutional Vision Transformer with SA is very useful for drone-based or mobile disease monitoring because of its ability to concentrate on the crucial portions of image, allowing for precise disease identification and localization in real time.

Algorithm 1: Convolutional Vision Transformers with Spatial Attention

Procedures:

Step 1: Preprocessing Phase

Input Image Preparation: Load the image I of size $H \times W \times C$. Where, H and W are height and width and C is the number of channels

Image Normalization**:** Normalize pixel values to the range [0, 1] for consistent feature scaling:

$I\ (norm) = I(x, y, c)\ \frac{I(x,y,c)-\mu}{\sigma}$

Data Augmentation**:** Apply transformations like rotation (θ), flipping and cropping to create diverse training examples.

Step 2: Feature Extraction using Convolutional Layers

Apply Convolutional Layers**:**

Extract feature maps F from $I(norm)$:

$F = Conv\ (I\ (norm): W)$

Where,W is the convolution kernel and F is of size $H' \times W' \times C'$

ReLU Activation**:** $F' = ReLU(F)$

Step 3: Transformer Encoding

Flatten and Tokenize Feature Maps: Convert feature maps F''into a sequence of tokens T:

$T = Flatten\ (F'')\ + Positional\ Embedding$

Multi-Head Spatial-Attention: Perform attention on tokens T:

$A\ (Q, K, V) = Softmax\ (\frac{QK^T}{\sqrt{d_k}})V$

Step 4: SA Mechanism

Compute SA Map: Aggregate channel-wise feature responses to highlight regions of interest

$$M_s(x,y) = \frac{\sum_c |F_c(x,y)|}{\sum_{x,y,c} |F_c(x,y)|}$$

Enhance features with spatial weights

$F(x, y, c) = F_c(x, y) \cdot M_s(x, y)$

Step 5: Disease Classification and Localization

Global Average Pooling:Compress feature maps to reduce dimensionality:

$G = GlobalAvgPool(F)$

Fully Connected Layers for Classification:Pass G through dense layers to predict disease class P:

$P = Softmax(W_{fc} G + b_{fc})$

Bounding Box Prediction:

Regression is used to predict the bounding box coordinates (x1, y1, x2, y2). x1, y1 are the coordinates of the top-left corner of the bounding box. X2, y2 are the coordinates of the bottom-right corner

$B=W_r G+b_r$

Step 6: Loss Function

Classification Loss:Cross-entropy loss for disease classification:

$$L_c = -\sum_i y_i \log(p_i)$$

$y_i$denotes the ground truth labelwhile the predicted probability is indicated as p_i.

Localization Loss**:** Mean Squared Error (MSE) for bounding box prediction:

$$L_l = \frac{1}{4}\sum_{j=1}^{4}(B_j - B_j^{true})^2$$

Combined Loss**:**

$$L = L_c + \lambda L_l$$

Combined loss L is the weighted combination of L_c and L_l , with the goal of maximizing bounding box accuracy and disease prediction. L_c ensures that the model properly identifies the disease by using cross-entropy, which penalizes incorrect predictions based on the probability assigned to right class.

V Results And Discussions

In this paper,Python is used for implementation. The proposed CVT with SA mechanism detects the rice leaf disease accurately than existing methods. This suggested hybrid method which extracts features from visual data using CNN, allowing for the collection of local patterns such as textures and edges while distinguishing between various rice leaf diseases.Vision Transformer is used for modeling global dependencies in images by dividing into patches and applying spatial-attention mechanisms and improving rice leaf detection accuracy. By focusing on relevant spots in images, the model improves its ability to identify and rank contaminated areas by using SA to give more weight in geographical information. This improves disease detection accuracy by reducing distractions from irrelevant background information in image. This proposed method has enhanced theleaf disease detection using these combined methods. Convolutional Vision Transformer with SA achieves 98.12% accuracy in feature extraction and 98.56% accuracy in classification using dataset1. In dataset 2, this method achieves98.26% accuracy in feature extraction and 98.67% accuracy in classification. The proposed method outperforms than the methods.

Table 1: Performance comparison of feature extraction using dataset 1

Feature Extractionusing Dataset 1				
Algorithm /Metrics	**Accuracy %**	**Precision %**	**Recall %**	**F-measure %**
CNN [33]	95.14	94.67	94.96	94.20
ViT [18]	95.87	95.12	95.45	95.02
SA [21]	96.45	96.21	96.32	96.01
CVT [53]	97.75	96.67	97.23	96.54
CVT with SA	98.12	97.85	98.02	97.67

Table 1 illustrates the comparison of feature extraction for CNN, ViT, SA, CVT and CVT with SA using dataset 1 on the performance metrics. Including SA allows the proposed CVT to outperform the existing approaches for disease detection in rice leaf. It has performed with 98.12% accuracy using dataset 1.

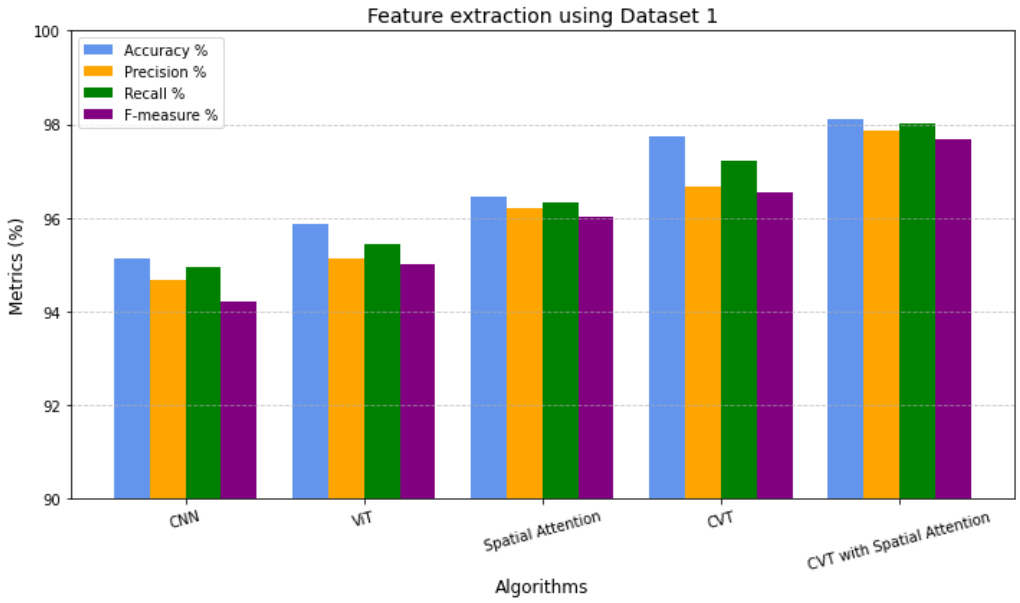

Figure 4: Comparison flow chart of feature extraction using dataset 1

Figure 4 shows CNN, ViT, SA, CVT and CVT with SA on dataset 1, together with other performance metrics such as recall, accuracy and precision. While the inscription of metrics percentage on the y-axis of the chart, the methodologies currently in use and explored in this work are represented on the x-axis.

Table 2: Comparison on classification using dataset 1

Classification using Dataset 1				
Algorithm /Metrics	**Accuracy %**	**Precision %**	**Recall %**	**F-measure %**
CNN [33]	95.80	95.01	95.76	94.98
ViT[18]	96.74	95.67	96.64	95.23
SA [21]	97.56	96.96	97.45	96.45
CVT [53]	98.12	97.54	97.89	97.04
CVT with Spatial Attention	98.56	97.78	98.34	97.64

The performance metrics of CNN, ViT, SA, CVT and CVT with SA is shown in table 2 for the classification on dataset 1. Regarding disease identification in rice leaves, the suggested CVT with SA outperforms current approaches in terms of accuracy. On dataset 1, it has achieved 98.56% accuracy.

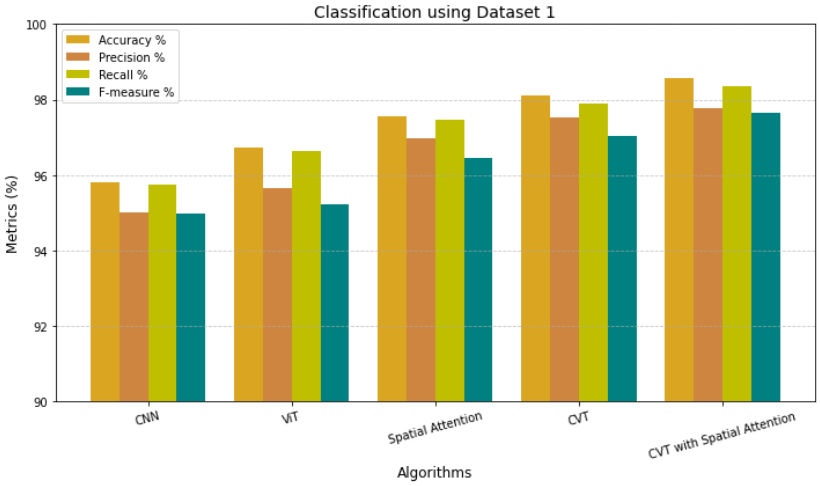

Figure 5: Comparison chart of classification using dataset 1

The algorithms such as CNN, ViT, SA, CVT and CVT are compared in figure 5 with the use of dataset 1. The metrics percentage is the y-axis of graph and the techniques that are in use and described in this study are the x-axis.

Table 3: Comparison on feature extraction using dataset 2

Feature Extraction using Dataset 2				
Algorithm /Metrics	**Accuracy %**	**Precision %**	**Recall %**	**F-measure %**
CNN [33]	95.69	94.90	95.29	94.76
ViT [18]	96.87	95.89	96.34	95.75
SA [21]	97.13	96.57	96.89	96.23
CVT [53]	97.87	97.20	97.55	97.11
CVT with SA	98.26	97.78	98.11	97.54

Table 3 compares the performance metrics of the following algorithms, CNN, ViT, SA, CVT and CVT with SA across dataset 2.

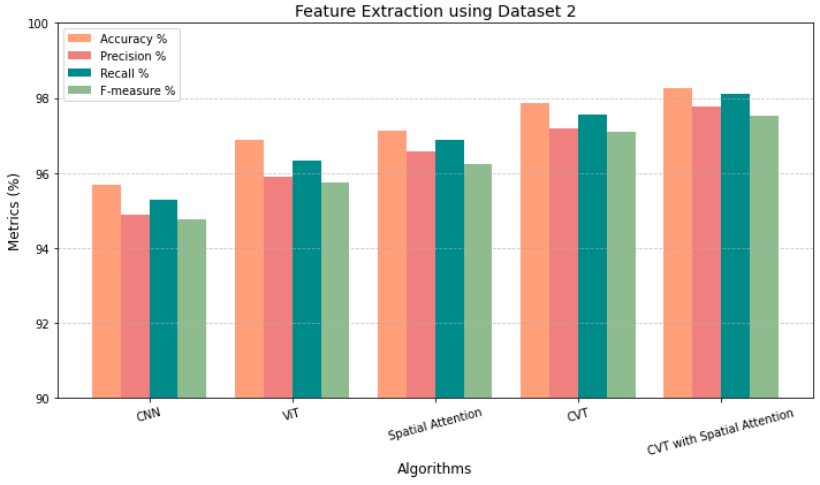

Figure 6: Comparison chart of Feature extraction using dataset 2

Figure 6 represents CNN, ViT, SA, CVT and CVT with SA alongside other evaluation parameters like recall, accuracy and precision of dataset 2. The values of metrics are presented on the y-axis of the graph and the methods, which are used at the moment and mentioned in the current research, are presented on the x-axis.

Table 4: Comparison on feature extraction using dataset 2

Classification using Dataset 2				
Algorithm /Metrics	**Accuracy %**	**Precision %**	**Recall %**	**F-measure %**
CNN [33]	95.87	94.89	95.43	94.56
ViT [18]	96.89	95.35	96.43	95.10
SA [21]	97.87	96.78	97.45	96.53
CVT [53]	98.12	97.46	98.00	97.11
CVT with SA	98.67	97.98	98.45	97.61

Table 4 uses dataset 2 and compares the accuracy, precision, recall, and f-measures of CNN, ViT, SA, CVT and CVT with SA.

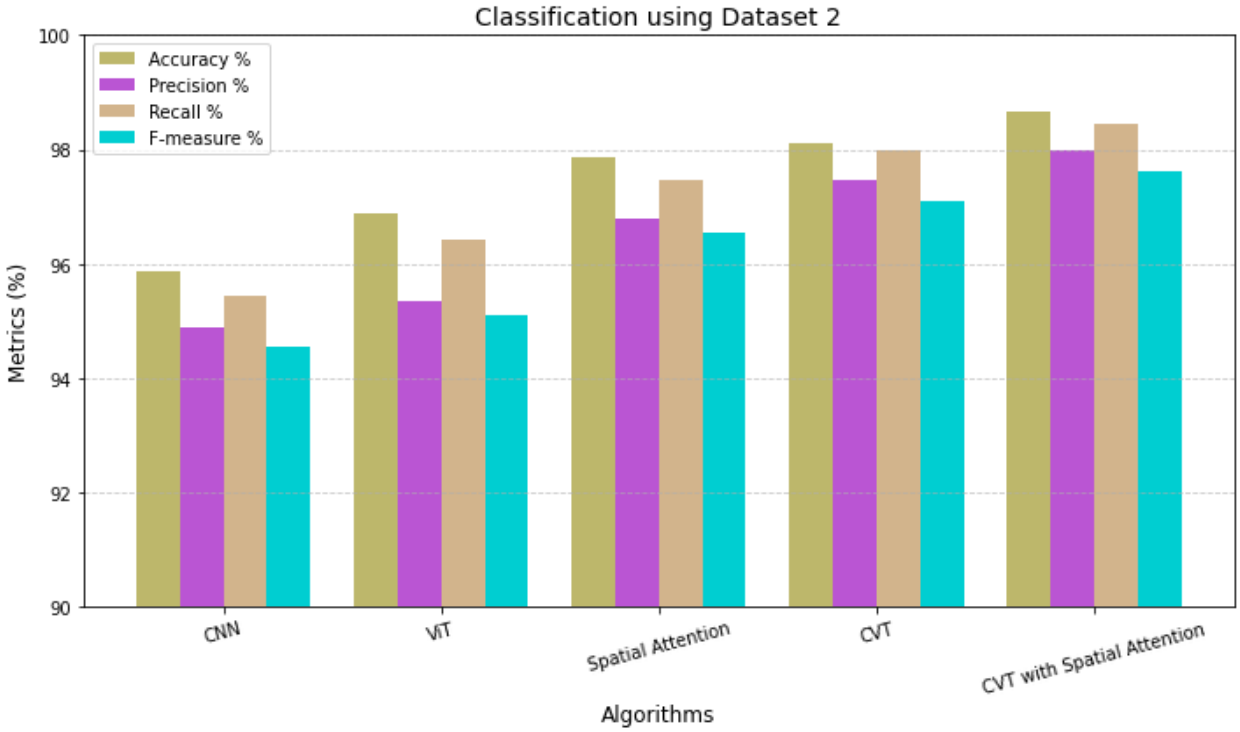

Figure 7: Comparison chart of classification using dataset 2

Figure 7 indicates the different performance indicators like accuracy, precision, recall, and f-measure comparison of CNN, ViT, SA, CVT and CVT with SA using dataset 2. In this chart, x-axis depicts the different existing and suggested that will be used in this paper and the y-axis depicts the different metrics on which the performance of algorithms should be evaluated.

VI Conclusion

Lastly, the proposed Hybrid CVT model is very accurate and reliable in the noticing of the

different rice leaf diseases. This method is better than CNN and ViT based methods because the former combines the benefit of ViT when it comes to the extraction of global dependencies and CNN when it comes to the extraction of local features. A Spatial Attention module can be used to fine-tune the given model to disease-prone areas. Therefore, accuracy in classification and interpretation of the model is enhanced. The digital results of the experiment on the rice leaf disease dataset indicate that the feature extraction accuracy (98.12) and classification accuracy (98.56) in dataset 1, feature extraction accuracy (98.26) and classification accuracy (98.67) in dataset 2, indicate that hybrid CVT is effective on the real world. The application has a high scaling capability in agricultural systems, especially in drone-based or mobile disease surveillance, using this approach. The future research is on the objectives of this work and the expansion of the data to more rice leaf diseases to develop the SA mechanism in deployment in real time in automated leaf detection. What is more, the investigation of the hybrid CVT model application to other fields of agriculture has pioneered the expansion of AI-enabled disease control technologies acceptance.

References

[1] Aggarwal, S., Suchithra, M., Chandramouli, N., Sarada, M., Verma, A., Vetrithangam, D., ...&AmbachewAdugna, B. (2022). Rice Disease Detection Using Artificial Intelligence and Machine Learning Techniques to Improvise Agro-Business. *Scientific Programming*, *2022*(1), 1757888.

[2] Bari, B. S., Islam, M. N., Rashid, M., Hasan, M. J., Razman, M. A. M., Musa, R. M., ...&Majeed, A. P. A. (2021). A real-time approach of diagnosing rice leaf disease using deep learning-based faster R-CNN framework. *PeerJ Computer Science*, *7*, e432.

[3] Bhowmik, A. C., Ahad, M. T., Emon, Y. R., Ahmed, F., Song, B., & Li, Y. (2024). A customised Vision Transformer for accurate detection and classification of Java Plum leaf disease. *Smart Agricultural Technology*, *8*, 100500.

[4] Brown, D., & De Silva, M. Plant Disease Detection on Multispectral Images using Vision Transformers. In *Proceedings of the 25th Irish Machine Vision and Image Processing Conference (IMVIP), Galway, Ireland* (Vol. 30).

[5] Deng, J., Yang, C., Huang, K., Lei, L., Ye, J., Zeng, W., ...& Zhang, Y. (2023). Deep-learning-based rice disease and insect pest detection on a mobile phone. *Agronomy*, *13*(8), 2139.

[6] Kashyap, B., & Kumar, R. (2021). Sensing methodologies in agriculture for monitoring biotic stress in plants due to pathogens and pests. *Inventions*, *6*(2), 29.

[7] Deng, R., Tao, M., Xing, H., Yang, X., Liu, C., Liao, K., & Qi, L. (2021). Automatic diagnosis of rice diseases using deep learning. *Frontiers in plant science*, *12*, 701038.

[8] Devi, R. S., Kumar, V. R., &Sivakumar, P. (2023). EfficientNetV2 Model for Plant Disease Classification and Pest Recognition. *Computer Systems Science & Engineering*, *45*(2).

[9] Fang, L., Wu, Y., Li, Y., Guo, H., Zhang, H., Wang, X., ...&Hou, J. (2021). Using channel and network layer pruning based on deep learning for real-time detection of ginger images. *Agriculture*, *11*(12), 1190.

[10] Hasan, M. M., Rahman, T., Uddin, A. S., Galib, S. M., Akhond, M. R., Uddin, M. J., & Hossain, M. A. (2023). Enhancing rice crop management: Disease classification using convolutional neural networks and mobile application integration. *Agriculture*, *13*(8), 1549.

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[12] Yang, L., Yu, X., Zhang, S., Long, H., Zhang, H., Xu, S., & Liao, Y. (2023). GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases. *Computers and Electronics in Agriculture*, *204*, 107543.

[13] Islam, M. A., Shuvo, M. N. R., Shamsojjaman, M., Hasan, S., Hossain, M. S., &Khatun, T. (2021). An automated convolutional neural network based approach for paddy leaf disease detection. *International Journal of Advanced Computer Science and Applications*, *12*(1).

[14] Thakur, P. S., Chaturvedi, S., Khanna, P., Sheorey, T., &Ojha, A. (2023). Vision transformer meets convolutional neural

network for plant disease classification. *Ecological Informatics*, *77*, 102245.

[15] Joshi, P., Das, D., Udutalapally, V., Pradhan, M. K., &Misra, S. (2022). Ricebios: Identification of biotic stress in rice crops using edge-as-a-service. *IEEE Sensors Journal*, *22*(5), 4616-4624.

[16] Kong, J., Wang, H., Yang, C., Jin, X., Zuo, M., & Zhang, X. (2022). A spatial feature-enhanced attention neural network with high-order pooling representation for application in pest and disease recognition. *Agriculture*, *12*(4), 500.

[17] Latif, G., Abdelhamid, S. E., Mallouhy, R. E., Alghazo, J., &Kazimi, Z. A. (2022). Deep learning utilization in agriculture: Detection of rice plant diseases using an improved CNN model. *Plants*, *11*(17), 2230.

[18] Lin, S., Xiu, Y., Kong, J., Yang, C., & Zhao, C. (2023). An effective pyramid neural network based on graph-related attentions structure for fine-grained disease and pest identification in intelligent agriculture. *Agriculture*, *13*(3), 567.

[19] Ullah, W., Javed, K., Khan, M. A., Alghayadh, F. Y., Bhatt, M. W., Al Naimi, I. S., &Ofori, I. (2024). Efficient identification and classification of apple leaf diseases using lightweight vision transformer (ViT). *Discover Sustainability*, *5*(1), 116.

[20] Lv, P., Xu, H., Zhang, Y., Zhang, Q., Pan, Q., Qin, Y., ...& Chen, C. (2024). An Improved Multi-Scale Feature Extraction Network for Rice Disease and Pest Recognition. *Insects*, *15*(11), 827.

[21] Tabbakh, A., &Barpanda, S. S. (2023). A deep features extraction model based on the transfer learning model and vision transformer "tlmvit" for plant disease classification. *IEEE Access*, *11*, 45377-45392.

[22] Wijayanto, A. K., Prasetyo, L. B., Hudjimartsu, S. A., Sigit, G., &Hongo, C. (2024). Textural features for BLB disease damage assessment in paddy fields using drone data and machine learning: Enhancing disease detection accuracy. *Smart Agricultural Technology*, *8*, 100498.

[23] Daniya, T., &Vigneshwari, S. (2019). A review on machine learning techniques for rice plant disease detection in agricultural research. *system*, *28*(13), 49-62.

[24] Li, D., Wang, R., Xie, C., Liu, L., Zhang, J., Li, R., ...& Liu, W. (2020). A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network. *Sensors*, *20*(3), 578.

[25] Patil, R. R., & Kumar, S. (2022). Rice transformer: A novel integrated management system for controlling rice diseases. *IEEE Access*, *10*, 87698-87714.

[26] Parez, S., Dilshad, N., Alghamdi, N. S., Alanazi, T. M., & Lee, J. W. (2023). Visual intelligence in precision agriculture: Exploring plant disease detection via efficient vision transformers. *Sensors*, *23*(15), 6949.

[27] Rahman, C. R., Arko, P. S., Ali, M. E., Khan, M. A. I., Apon, S. H., Nowrin, F., &Wasif, A. (2020). Identification and recognition of rice diseases and pests using convolutional neural networks. *Biosystems Engineering*, *194*, 112-120.

[28] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ...&Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*, 211-252.

[29] Wu, H., Xiao, B., Codella, N., Liu, M., Dai, X., Yuan, L., & Zhang, L. (2021). Cvt: Introducing convolutions to vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 22-31).

Biography

R. Sheeba

Dr R. Sheeba is working as an Assistant Professor in SRM Institute of Science and Technology,

Tiruchirapalli. She did her UG and PG with specialization in Information Technology under Anna University. She did her doctorate degree with specialization in Information and Communication Engineering. Her research areas are Network Security, Sensor Networks, Machine Learning and Data Analysis.

Mrs. G. Shyning Sobina pursuing part time Ph.D in Computer Science from SRM Institute of Science and Technology, Thiruchirapally, India. She has completed M.Sc Computer Science from Manonmaniam Sundaranar University, India. She has 3 years of teaching experience, her area of research interests include Image Processing Machine Learning. She is currently working as an Assistant Professor in Muslim Arts College, Thiruvithancode, India.