IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 9 July 2024, accepted 29 July 2024, date of publication 5 August 2024, date of current version 16 August 2024.
Digital Object Identifier 10.1109/ACCESS.2024.3438544

RESEARCH ARTICLE

A Comprehensive Approach Toward Wheat Leaf Disease Identification Leveraging Transformer Models and Federated Learning

MD. FAHIM-UL-ISLAM[1], AMITABHA CHAKRABARTY[1], SARDER TANVIR AHMED[1], RAFEED RAHMAN[1], HYUN HAN KWON[2], AND MD. JALIL PIRAN[3], (Senior Member, IEEE)
[1]Department of Computer Science and Engineering, Brac University, Dhaka 1212, Bangladesh
[2]Department of Civil and Environmental Engineering, Sejong University, Seoul 05006, South Korea
[3]Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea

Corresponding author: Md. Jalil Piran (piran@sejong.ac.kr)

This work was supported by Korea Environmental Industry and Technology Institute (KEITI) through the Water Management Program for Drought Project, funded by Korea Ministry of Environment (MOE) under Grant 2022003610003.

ABSTRACT Wheat is one of the most extensively cultivated crops worldwide that contributes significantly to global food caloric and protein production and is grown on millions of hectares yearly. However, diseases like brown rust, septoria, yellow rust, and other fungus diseases pose notable threats to wheat crops, impacting production and quality. Diagnosing these diseases is challenging, especially in areas with limited agricultural experts. Thus, creating computerized disease identification and decision-support technologies is crucial for safeguarding wheat leaf preservation and crop loss mitigation. The traditional approach to integrating data gathering and model training has substantial challenges in terms of data confidentiality, availability, and the costs related to data transmission. To address these challenges, federated learning (FL) is an appealing and effective option. Our study focuses on applying FL to classify agricultural diseases using image analysis. In our study, we conduct experiments on high-parameterized transfer learning (TL) models along with our proposed architecture based on the attention mechanism, introducing these models into a distributed learning strategy founded in FL. Our proposed architecture leverages the beneficial interactions of two cutting-edge vision transformer models including the advanced depthwise incorporating self-attention model referred to as CoAtNets, and the enhanced Swin Transformer V2, resulting in enhanced feature representation. Moreover, we introduce weight pruning into our model which is further classified by a reinforced linear attention mechanism (LA) to lower output dimensions. Our pruned lightweight (32M parameters) considerably decreases inference time with 624.249 ms and 644.899 on devices with low computational power, making it highly efficient in FL-based systems. The proposed model in our FL system significantly outperforms all other tested transfer learning models, including ConvNeXtBase, ConvNeXtLarge, EfficientNetV2L, InceptionResNetV2, ResNet152, and NASNetLarge, achieving accuracies up to 98% and 99%, precision up to 98%, recall up to 98%, and F-1 scores up to 95% across multiple input dimensions for wheat leaf disease classification.

INDEX TERMS Machine learning, deep learning, federated learning, transformers, attention mechanism, wheat leaf disease identification.

I. INTRODUCTION

Wheat (Triticum aestivum) is one of the most extensively consumed cereals globally and a key food source for humanity [1]. Its relevance has grown, notably due to the harmful impacts of bad agricultural practices during the epidemic

The associate editor coordinating the review of this manuscript and approving it for publication was Xinyu Du.

and the subsequent conflict between Ukraine and Russia. Both countries rank among the top 10 wheat producers worldwide. Wheat is key in reducing hunger, especially in African areas with limited access to food [2]. As a result, there are enormous incentives to produce this crop, coupled with numerous methods targeted at enhancing productivity. Wheat is one of the key staple foods and is considered the second-largest crop in the world. Significant yield advances in wheat production over the past 40 years have resulted in an ongoing equilibrium of supply versus demand [3], [4]. However, climate change and extreme weather events present substantial obstacles to agriculture and ecosystems. Fluctuations in temperature and precipitation patterns lead to an escalation in droughts, floods, and heat waves, which have a detrimental effect on the growth and productivity of wheat. Furthermore, these occurrences lead to a rise in the occurrence and intensity of diseases, which in turn impair the inherent defenses of plants and diminish their output. Wheat is a staple crop that requires control from many diseases through breeding resistance, insecticides, or other techniques. Out of the 31 pests and pathogens recorded in wheat, fungal diseases such as leaf and stripe rust, Fusarium head blight, Septoria leaf blotch, spot blotch, tan spot, and powdery mildew cause the most significant losses [5]. Yellow rust is easily detected by orange/yellow uredinial pustules on leaves, while Septoria occurs on the leaf as necrotic yellow-to-brown lesions confined by veins with little black pycnidia. Confusion can emerge at two stages: early yellow rust and Septoria both show as elongated patches of chlorosis and later yellow rust and Septoria lesions can be confused. Brown or leaf rust, another common wheat leaf disease, generates orange/brown pustules on leaves, making it difficult to identify from yellow rust and Septoria [6]. Traditional procedures for diagnosing and managing wheat crop diseases are concentrated on pathologists, making them subjective, time-consuming, and labor-intensive. Given a scarcity of experience and manpower, researchers continue to investigate computer vision algorithms for effectively detecting disease occurrences on specific plots of land used for agricultural tasks [7]. Leveraging the transformative power of deep learning models, image-based plant disease detection has emerged as a promising solution to safeguard these wheat crops, as evidenced by the works of in the literature review, Navale and Basapur [8] showcased significant enhancements in the accuracy by using the (Convolutional Neural Network) CNN models. Additionally, Ceyhan et al. [9] proposed an image-based deep learning method that uses reflection data to classify wheat varieties accurately. This method offers a more effective and affordable substitute for wheat classification in the agricultural and industrial sectors. A noteworthy issue in this regard has been the limited ability to properly address unknown diseased wheat image data linked with distinct diseases by utilizing advanced computer vision computation in a global scale approach. By using this approach, the decision-making skill of the model would be more effective and informative as known by different classes with different variants of diseases of wheat. The purpose is to develop algorithms that can fast and reliably classify previously unknown data, ensuring precise findings on a worldwide scale, thus boosting the total accuracy of classification tasks.

The present algorithms still exhibit a notable discrepancy in effectively classifying unknown photos of damaged wheat, particularly when confronted with a wide range of diseases worldwide. This difficulty emerges because these models rely significantly on pre-established disease categories and lack the adaptability to distinguish novel or rare disease variations. It is crucial to address this gap for multiple reasons. Initially, the agriculture sector is becoming more globalized, with the cultivation of wheat taking place in diverse climatic conditions and countries, each presenting distinct disease challenges. Effective disease management requires models that can generalize across multiple environments and accurately detect diseases not included in the training data. Presently, existing models cannot generally apply their knowledge to varied settings, resulting in decreasing precision and reliability in practical scenarios. Additionally, the ever-changing nature of disease development offers a substantial hurdle. Pathogens can undergo genetic alterations, leading to new disease strains. Current models may not have the capacity to recognize and interpret these new strains, underlining the need for strong models that can adapt to fresh data and continuously learn from new disease occurrences. Without this adaptability, models' decision-making abilities remain limited, leaving crops vulnerable to emerging dangers. Furthermore, climate change and extreme climatic events are projected to increase agricultural challenges and disease prevalence [10]. Federated learning tackles the difficulties presented by unpredictable weather patterns and the spread of wheat leaf diseases by facilitating decentralized, ongoing, and adaptable training of models. This method enables various stakeholders to provide localized data without disclosing sensitive information, ensuring that models stay up-to-date and strong under multiple circumstances. Therefore, federated learning gives a solution to this challenge. By allowing models to be trained on decentralized data from varied sources without sharing raw data, federated learning boosts the models' ability to generalize across different settings and adapt to new disease variants. This strategy harnesses the aggregate knowledge from different nodes, boosting the models' resilience and accuracy in recognizing and managing wheat leaf diseases globally.

The primary motivation for implementing federated learning in the context of wheat leaf disease identification lies in its ability to improve data privacy and security while utilizing distributed data sources. Conventional centralized machine learning methods necessitate aggregating all data in one place, which raises significant privacy concerns, particularly when dealing with sensitive agricultural data [11], [12], [13]. Federated learning solves this issue by facilitating model training over multiple distributed nodes, granting each participating node control over its data. This decentralized approach not only guarantees the confidentiality of data but also fosters collaboration among various parties without the need for sharing data. Furthermore, federated learning improves the ability to handle big amounts of data and increases efficiency by utilizing local computing resources. FL approach eliminates the requirement for transferring large amounts of data, resulting in reduced latency. This is especially crucial in agricultural settings where internet connectivity can be limited or unreliable.

Therefore, in our FL architecture, we employ our proposed transformer model architecture with its improved ability to capture complicated patterns in picture data and enhance the accuracy and robustness of disease classification. Therefore, FL offers a promising technique for model training, allowing the integration of input from multiple sources without compromising privacy. This decentralized strategy not only boosts the model's ability to generalize across diverse areas and situations but also ensures continual learning from new data, making the system more adaptable to emerging diseases and changing climatic conditions. Additionally, we have introduced weight pruning in federated learning systems that provide an efficient approach. Weight pruning includes lowering the number of parameters in a model by deleting less important weights and layers, hence reducing model complexity without severely sacrificing performance. This strategy is particularly advantageous in federated learning contexts where computational resources and bandwidth are potentially constrained. By performing weight pruning with the advanced method, we have reduced the overall model size, making it more viable to deploy and update models across different decentralized nodes. This strategy not only reduces the computing requirements but also boosts the efficiency and scalability of the federated learning system.

Our main contributions therefore,

- We use a four-layered preprocessing pipeline in our datasets that includes Canny Edge Detection, Gaussian Blur addition, Laplace Transformation, and High-Pass Filtering. Gaussian and Laplace noise transformation strengthen the model, and High-Pass Filtering sharpens images, making it simpler to retrieve disease-related characteristics. These methods, taken together, improve the accuracy alongside effectiveness of our wheat leaf disease categorization in an FL-based system.
- Our study introduces an advanced vision transformer architecture aimed at achieving enhanced performance with reduced parameterization in the FL distributed System. This architectural innovation combines the strengths of CoAtNet and the improved Swin Transformer version 2, leveraging a Local Attention (LA) mechanism. Compared to standard high-parameterized neural network models such as ConvNeXtBase, ConvNeXtLarge, EfficientNetV2L, InceptionResNetV2, ResNet152, and NASNetLarge, our proposed method demonstrates superior performance, particularly in FL scenarios.
- In addition, we propose an identifiable FL architecture for interactive model training across clients with different datasets which enables the storage age of local updates, allowing for dynamic interactions between local and global model instances and promoting efficient model update transmission with advanced weight pruning method.
- In our evaluation of model effectiveness, we apply multiple essential metrics and visualizations, including accuracy curves, loss curves, classification reports, confusion matrices, GPU utilization, and inference times through our training.

Following is a synopsis of the remainder of this paper. In Section II, we introduce our connected work with the previous studies comparing our proposed technique in this domain. In Section III, we address the approach we suggest for documenting our data collection, preprocessing networks, and proposed model while defining our overall system architecture and its usefulness. In Section IV, we describe our assessment metrics for the performance measure of our model. Afterward, we provide the experimental components and frameworks required to train our models in the environment. After putting our model strategies into practice, we study our outcomes and insights in Section VI to assess the efficacy of our models. In section VIII. we consider the practical considerations of our proposed system. Finally, in section IX. we conclude our work by describing the importance of our work.

II. RELATED WORK

Agriculture crops such as rice, wheat, and soybean are staple foods for many countries and because of this sustainable agricultural practices have allowed these crops to be both inexpensive and a reliable source of food security. However, plant-based ailments have put this at risk. When bacteria or fungi infect plant tissue growing above ground, lesions occur on the leaves, stems, and panicles of the plants. Image-based plant disease detection systems are developed by integrating deep learning models to process photos of sick leaves. Shoaib et al. [14], explored recent advancements in using Machine Learning (ML) and Deep Learning (DL) techniques for plant disease proof of identity, illustrating improved accuracy and efficiency, while addressing challenges and limitations, offering valuable insights for researchers, practitioners, and industry professionals. However, the authors did not apply to handling vast amounts of data with an advanced model architecture on a large scale. The applications of image preprocessing of images also exhibited the lack of their study. Moreover, transformer models are becoming increasingly important due to the development of specialized quantization methods adapted to these models.

Qin et al. in [15] proposed BiBERT, a completely binarized iteration of the BERT model that aims to substantially decrease computational and memory expenses while

TABLE 1. Overview of the existing related studies.

Research	Study Areas	Dataset	Data Preprocessing Used	Model Used
[20]	Crop Disease Recognition	PlantVillage, Apple Leaf Pathology	✗	MSCVT
[21]	Plant Disease Classification	Crawled from Baidu Baike and various agricultural websites	✓	SEViT
[22]	Plant Pathology Recognition	Plant Pathology 2020-FGVC7, Plant Pathology 2021-FGVC8	✗	DCTN
[23]	Crop Disease Prediction	Potato leaf dataset	✓	PLDPNet
[24]	Sugarcane Leaf Diseases Identification	Self-built dataset consisting of 2521 sugarcane leaf images	✗	ViT+CNN
[25]	Rice Leaf Disease Identification	Data acquisition consisting of 4523 rice leaf images	✓	ECA-ConvNeXt
[26]	Paddy Leaf Diseases	Self-built heterogeneous dataset consisting of rice-leaf disease images	✓	EfficientNetB3, MobilenetV2
[27]	Paddy Plant Disease Detection	Self-built from various paddy fields in Kerala	✓	Deep Learning-based
[28]	Tomato Disease Detection	Self-built Tomato Leaf Disease Image Dataset	✗	NanoSegmenter
[29]	Crop Classification	Smart farm datasets	✗	CNN Models
[30]	Plant Disease Classification	DiaMOS Plant	✗	EfficientNetB0+CBAM
Our Proposed Model	Plant Disease Classification	Plant Disease Classification Merged Dataset, Wheat Nitrogen Deficiency	✓	CoAtNet+Improved Swin V2+LA

preserving performance. The main advancements include the Bi-Attention mechanism, which reduces information loss during binarization by maximizing information entropy and replacing softmax with Bitwise-Affine Matrix Multiplication (BAMM), and the Direction-Matching Distillation (DMD) method, which enhances optimization accuracy by aligning the optimization directions of binarized and full-precision models through upstream distillation and similarity pattern matrices.

Chen et al. in [16] introduced DB-LLM (Dual-Binarization for Large Language Models) as a method to improve computing performance by using ultra-low bit quantization. The key innovations consist of Flexible Dual Binarization (FDB), which divides 2-bit quantized weights into two binary sets to enable efficient operations without sacrificing accuracy, and Deviation-Aware Distillation (DAD), which modifies the distillation loss to prioritize ambiguous samples. The experiments conducted on LLaMA-1 and LLaMA-2 models provide evidence that DB-LLM surpasses existing quantization approaches in terms of perplexity and accuracy, resulting in a 20% decrease in computing expenditure.

Qin et al. [17] presented IR-QLoRA, an innovative approach to enhance the precision of quantized large language models (LLMs) by LoRA finetuning. The system employs two primary methodologies: Statistics-based Information Calibration Quantization, which effectively preserves the original information in quantized parameters, and Finetuning-based Information Elastic Connection, which enables versatile representation of information. The experiments demonstrate that IR-QLoRA greatly enhances the accuracy of the LLaMA and LLaMA2 models when using 2-4 bit-widths. This improvement in accuracy is achieved with only a slight increase in time consumption, showcasing the efficiency and versatility of IR-QLoRA.

Qin et al. in [18] introduced BiBench, a benchmark designed to examine network binarization holistically. It mainly evaluates the prerequisites for viable binarization, sets evaluation metrics, and assesses milestone binarization algorithms. Key findings include the crucial importance of binarized operators on performance and deployability, large accuracy variations across jobs and architectures, and promising efficiency on edge devices despite hardware limits.

Huang et al. in [19] introduced BiLLM, a revolutionary 1-bit post-training quantization approach for LLMs. BiLLM discovered salient weights and minimized compression loss by binary residual approximation, while accurately binarizing non-salient weights using an optimal splitting search. BiLLM produced high-accuracy inference (e.g., 8.41 perplexity on LLaMA2-70B) with 1.08-bit weights, greatly surpassing state-of-the-art quantization approaches.

Zhu et al. [20] proposed MSCVT, a lightweight hybrid transformer model for crop disease detection that combines features of CNN and Transformer via multiscale self-attention (MSSA) modules, demonstrating high recognition accuracies on practical disease data. Nevertheless, the authors did not implicit any preprocessing network for measuring the robustness and flexibility of the model in the real-life data. Zeng et al. [21] developed the Squeeze-and-Excitation Vision Transformer (SEViT) model for large-scale and fine-grained plant disease classification, combining ResNet with a channel attention module for preprocessing and ViT for feature classification but yielded a lower accuracy under a field background, indicating the difficulty of field identification. However, the authors weren't interested in constructing this design in real-world environments while retaining the scale and effectiveness of the service.

Pang et al. [22] established a novel approach called Dense CNNs and Transformer Network (DCTN) for accurate field crop disease detection, utilizing a multi-head self-attention mechanism, on their dataset and a publicly available dataset, respectively, showcasing its reliability against background interference in real-field environments. Consequently, it is vital to highlight the utility of the novel computer vision algorithms that improve model efficiency in a decentralized federated network setting that was missing in these works. Traditional transfer learning methods may not be as efficient in today's world because they rely on a wide number of characteristics, which, while their abundance, may offer less precise categorization competencies. Additionally, Arshad et al. [23] constructed a novel deep learning framework for precise and efficient detection leveraging two well-established models (VGG19 and Inception-V3) under the PLDNet framework designed to automatically predict potato leaf diseases. However, this study misses the promise of retaining the standard for making that deep learning system lightweight as VGG19 comprises 143.7M parameters. Öğrekçi et al. [24] applied deep learning approaches, specifically DenseNet121, Vision Transformers (ViT), and a ViT + CNN combination, for the categorization of illnesses in sugarcane leaves, attaining high precision rates. However, the comparative study might not account for the varying nature of datasets, which can significantly impact the performance of vision transformers and CNNs.

Wang et al. [25] proposed the novel ECA-ConvNeXt model for the identification of six categories of rice leaf diseases and healthy rice leaves, combining the Efficient Channel Attention (ECA) module and utilizing transfer learning, achieving an impressive accuracy on the rice leaf disease identification dataset. Aggarwal et al. [26] proposed the federated transfer learning (F-TL) approach for rice-leaf disease categorization, using both IID and non-IID datasets, with EfficientNetB3 and MobileNetV2 showing comparable results, demonstrating the advantages of the F-TL framework for cost-effective, data-privacy-assured paddy leaf disease identification in resource-constrained edge devices. Though the use of FL showed a promising approach to handling large amounts of data is vital for fast and efficient learning throughout the system, the comparative analysis of diverse VIT models is missing in this study. Also, the noise addition process for the robust performance of the model is missing in this paper.

Haridasan et al. [27] proposed an automated computer vision-based approach utilizing image processing, machine learning, and deep learning, incorporating support vector machine classifier and convolutional neural networks, to accurately identify and categorize rice plant diseases, achieving the finest validation accuracy of 0.9145 and offering predictive remedies for disease management in Indian rice fields. However, for handling large sets of data, the traditional approach for deep learning with the assistance of a machine learning classifier would not be suitable. Liu et al. [28] proposed a NanoSegmenter model based on the Transformer structure, integrated with lightweight technologies, to achieve high-precision tomato disease detection, with a precision of 0.98, a recall of 0.97, a mIoU of 0.95, and a computational efficiency of 37 FPS, providing a viable option for this crucial agricultural application on tomato disease detection. Idoje et al. [29] investigated the application of FL in smart farming, utilizing the federated averaging model for crop classification with climatic parameters as independent variables and crop types as labels, with decentralized models converging faster and achieving higher accuracy. However, all these efforts lack the application of FL approaches that principally handle the broad dispersion of data for multiple locations or organizations for more detailed findings for one or several types of plant diseases. This is a notable constraint that necessitates the requirement of development. Table 1 has been prepared to present an overview of the strategies that have been employed for accurate diseased plant picture categorization.

Upon a detailed evaluation of existing works, it becomes obvious that these efforts generally revolve around boosting system efficiency. Yet, there are considerable gaps in addressing the essential topic of guaranteeing real-time analysis in a decentralized environment with multiple clients. Therefore, integrating FL into the domain of wheat crop disease detection could offer several key benefits. Firstly, it would enhance collaborative learning and updating by ensuring that sensitive agricultural data, such as images of infected crops, remains localized and is never shared in its raw form. Secondly, it allows for collaboration among researchers, practitioners, and industry professionals. Moreover, our advanced FL enables the development of a more robust and accurate global model by aggregating knowledge from diverse sources, potentially improving the efficiency and reliability of disease identification. By incorporating secure aggregation techniques, the model updates would be maintained efficiently during the collaborative learning process between the local and global clients.

III. METHODOLOGY

Our working methodology includes a four-layer preprocessing pipeline for data gathering, improving model correctness by using Canny Edge Detection, Gaussian Blur addition, Laplace Transformation, and High-Pass Filtering. The advanced vision transformer architecture combines CoAtNet with Swin Transformer V2, enhanced by a Local Attention mechanism, to deliver superior performance with fewer parameters in federated learning (FL) systems. Furthermore, we employ the deep ensembles technique to get higher precision and dependable uncertainty estimations. This entails the process of training these models and then calculating the average of their predictions. This method effectively captures both the uncertainty related to knowledge and the uncertainty related to chance. Nevertheless, this process needs large computer resources because of the necessity to train and keep multiple models. To mitigate this issue, we put pruning strategies into the network. Moreover, this federated learning architecture allows for dynamic interactions between local and global model instances, which enables the efficient transmission of model updates. Our pruning approaches are used to further enhance optimization across the FL system. The overall strategy of this approach is depicted in Figure 1.

A. DATA COLLECTION

We collected two datasets from Kaggle to meet the distinct needs of our clients in our federated global network for aggregate results. The first dataset, "Plant Disease Classification Merged Dataset" was captured in the uncontrolled environment of Holeta wheat farm, Ethiopia [31]. The second dataset, "Wheat Nitrogen Deficiency and Leaf Rust Image", was obtained from a controlled experiment at the IARI field during the $2019 - 20$ rabi season, using an RGB camera [32]. These datasets serve as an extensive basis for our research, meeting the needs of both real-world and controlled scenarios. The folders are organized into train, test, and validation sets, with an 80:10:10 ratio with 112×112, 128×128 and 224×224 image sizes. Table 2 gives a brief description of the total distribution of those datasets containing 4230 images.

Table 3 displays summary statistics for two datasets, with variables such as mean, standard deviation, minimum, maximum, median, and quartiles. Dataset 1 has lower values across most parameters compared to Dataset 2, which often has higher values, indicating probable differences in the distributions of the two datasets.

B. IMAGE PREPROCESSING

To increase our model's efficacy in our FL system of wheat leaf diseases, we conduct four levels of image preprocessing steps into our datasets as shown in Figure 2. These methods categories include Canny Edge Detection, Gaussian noise, Laplace noise, and High-Pass Filtering.

1) DATA AUGMENTATION

We first apply data augmentation techniques, including horizontal flipping, and slight shifts in height and width, but excluding vertical flips. These augmentations simulate variations in the positioning and orientation of the subjects within the images, further enhancing the model's ability to recognize patterns regardless of their orientation or alignment in the input data.

2) CANNY EDGE DETECTION

The Canny edge detection algorithm is a cornerstone in image analysis [33]. Its execution on the grayscale image provides an edge map that delineates the significant edges within the visual content. By picking lower and upper thresholds for 100 and 200 instances, we apply the Canny algorithm to effectively find edges, which are crucial in activities such as object identification and feature extraction for our disease images.

3) GAUSSIAN BLUR

A Gaussian filter, often referred to as Gaussian blur, serves as a smoothing filter that is vital in image processing for the objective of softening images, therefore decreasing fine features and undesirable noise [34]. This filter involves the usage of a Gaussian function, which has a close connection to the normal distribution in statistics, to determine how each pixel in the image should be adjusted. To introduce noise and fine-scale deviations in the edge map, we utilize the Gaussian blur method. This smoothing operation is carried out with a Gaussian kernel of size (5, 5), and tries to eliminate high-frequency components that may interfere with further processing steps. The blurred images preserve the most important edges while lowering noise, resulting in a cleaner and more solid basis for later research.

4) LAPLACIAN TRANSFORMATION

In our study, laplacian sharpening is done using the OpenCV library. Laplacian function, adds a level of detail enhancement to the image. The Laplacian operator increases the high-frequency components inside the image, hence accentuating edges and tiny details [35]. The sharpened image is formed by subtracting a fraction (0.5) of the Laplacian of the original image from the image itself. The outcome is a visually more effective image, keeping essential visual cues.

5) HIGH-PASS FILTERING

A high-pass filter allows high frequencies to flow through while reducing or suppressing sounds below a specific cutoff frequency [36]. In our dataset, we apply sharpening which is merely a frequency-domain high-pass operation in the

TABLE 2. Main classes with number of images in both datasets.

Dataset	Folder	Subfolder	Number of Images (Count)
First Dataset	Training	wheat_brown_rust	731
		wheat_healthy	980
		wheat_septoria	77
		wheat_yellow_rust	905
	Testing	wheat_brown_rust	93
		wheat_healthy	123
		wheat_septoria	11
		wheat_yellow_rust	114
	Validation	wheat_brown_rust	91
		wheat_healthy	122
		wheat_septoria	9
		wheat_yellow_rust	113
Second Dataset	Training	control	343
		diseased	258
	Testing	control	76
		diseased	55
	Validation	control	74
		diseased	55
		Total	4,230

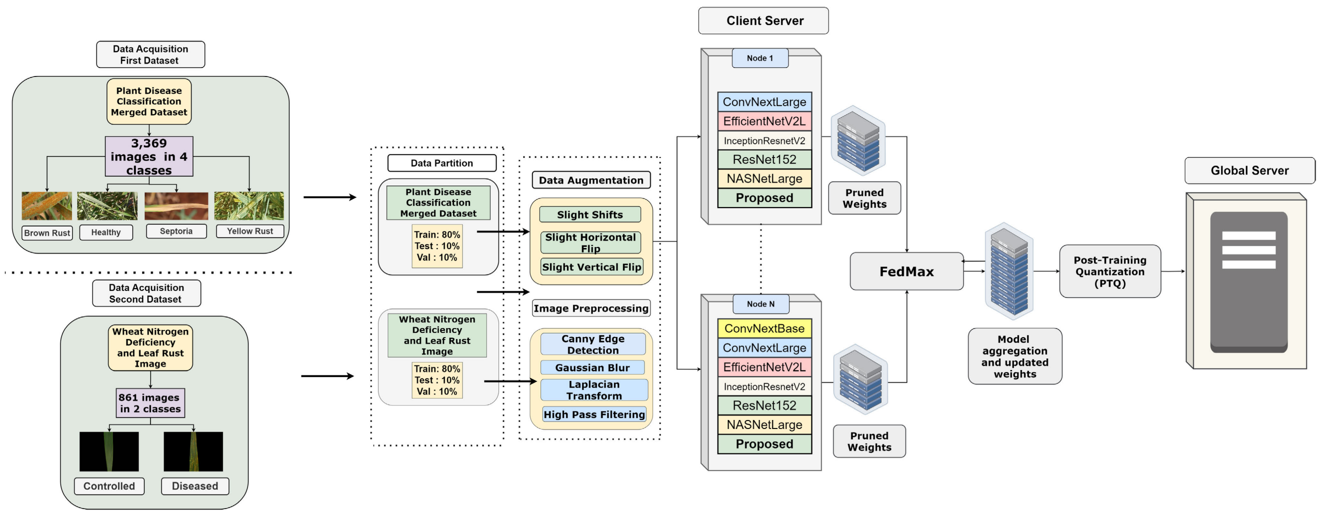


FIGURE 1. The proposed methodology involves data collection to a four-layered preprocessing pipeline incorporating canny edge detection, gaussian blur addition, laplace transformation, and high-pass filtering to enhance model accuracy. The advanced vision transformer architecture combines CoAtNet and Swin transformer V2 with a local attention mechanism, designed for high performance with reduced parameterization in federated learning (FL) systems. Additionally, the identifiable FL architecture supports dynamic interactions between local and global model instances, facilitating efficient model update transmission.

TABLE 3. Summary statistics for two datasets.

Statistic	Dataset 1	Dataset 2
Mean	842.5	2016.0
Standard Deviation	513.57	504.48
Minimum	97	2018
Maximum	1225	3024
Median	1024.0	2520.0
25th Percentile	711.25	2016.0
50th Percentile	1024.0	2520.0
75th Percentile	1155.25	3024.0

context of image processing. Here diseased images are sharpened when the contrast between adjacent areas with little fluctuation in brightness is increased.

C. THE EXISTING DEEP LEARNING MODELS FOR LEAF DISEASE RECOGNITION

To construct a precise model for the identification of wheat leaf diseases in real-world agricultural settings, our study delves into state-of-the-art models noted for their fair performance across varied datasets in recent years. Among them, we test important convolutional neural network (CNN)-based models, including ConvNeXtBase, ConvNeXt-Large, EfficientNetV2L, InceptionResNetV2, ResNet152, and NASNetLarge [25], [37], [38], [39], [40], [41]. In Table 4, we give a complete analysis of these reference models, defining their different properties and their respective applicability for diverse application situations within the domain of leaf disease identification.

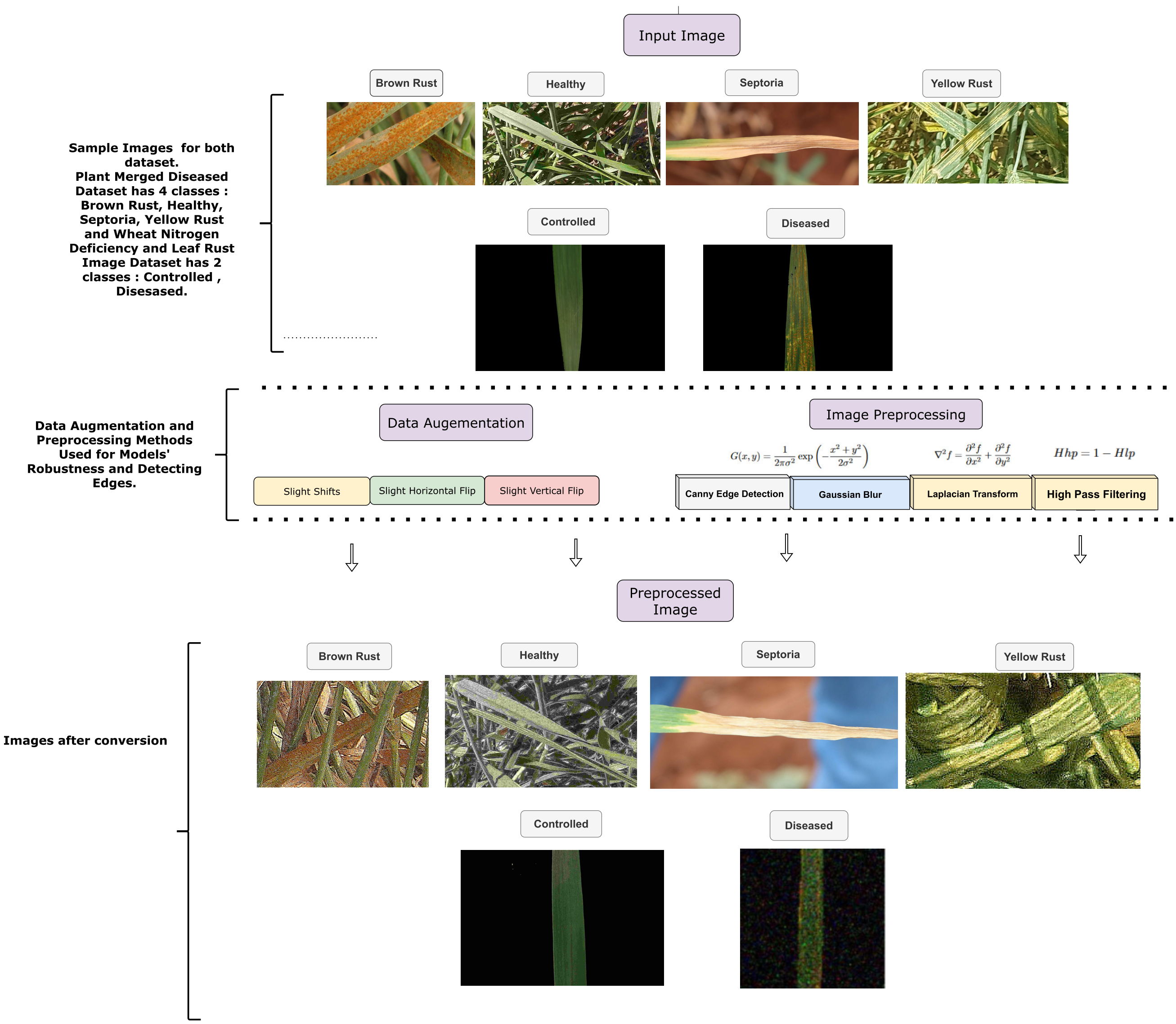


FIGURE 2. Four levels of image preprocessing methods are added to the image samples.

D. PROPOSED MODEL

In this study, we propose our model architecture for the difficult issue of wheat leaf disease classification in our FL system for handling large amounts of data, expertly combining the numerous characteristics of two foundational essential models: Swin Transformer and CoAtNet [42], [43]. We use these two models as feature extractors combining a fusion of layers in their model architecture shown in Figure 3.

The mechanism of vertically stacking convolution layers and attention layers in a logical method is effective in enhancing generalization, capacity, and efficiency in our CoAtNet model. On the other hand, The Swin Transformer V2 model features three essential strategies to boost its performance and adaptability. Firstly, it utilizes a residual-post-norm method coupled with cosine attention to improve training stability, solving major concerns faced in large-scale transformer models. Secondly, it proposes a log-spaced continuous position bias technique, enabling seamless transfer of pre-trained models from low-resolution to high-resolution tasks, hence boosting adaptability across multiple resolution settings. Lastly, the model employs a self-supervised pretraining strategy known as SimMIM, which drastically reduces the dependency on labeled data during training, allowing for more efficient exploitation of available resources. These strategies collectively contribute to the Swin Transformer V2 model's efficacy, efficiency, and adaptability across numerous applications and tasks. The whole model architecture is for the two datasets with different output shapes shown in Tables 5 and 6.

$$\mathbf{A} = \sqrt{\frac{\pi}{2e}}\left(\sum_{i=1}^{N}\mathbf{f}_{1i} + \sum_{i=1}^{N}\mathbf{f}_{2i}\right). \tag{1}$$

TABLE 4. Important description of different existing model types in leaf disease recognition.

Model	Model Type	Features	Application Scenarios in Leaf Disease Recognition
ConvNeXtBase [37]	CNN-Based	Convolutional, Multi-path	Detection of over 20 Mango leaf diseases
ConvNeXtLarge [25]	CNN-Based	Convolutional, Multi-path	Detection of 5 Rice leaf diseases
EfficientNetV2L [38]	CNN-Based	Efficient architecture	Detection of 38 types of leaf diseases in 14 different plants
InceptionResNetV2 [39]	CNN-Based	Inception and ResNet features	Detection of 3 Potato leaf diseases
ResNet152 [40]	CNN-Based	Deep residual networks	Detection of over 5 Tomato leaf diseases
NasNetLarge [41]	CNN-Based	Convolutional, Multi-path	Crop Disease Classification in 37 different categories

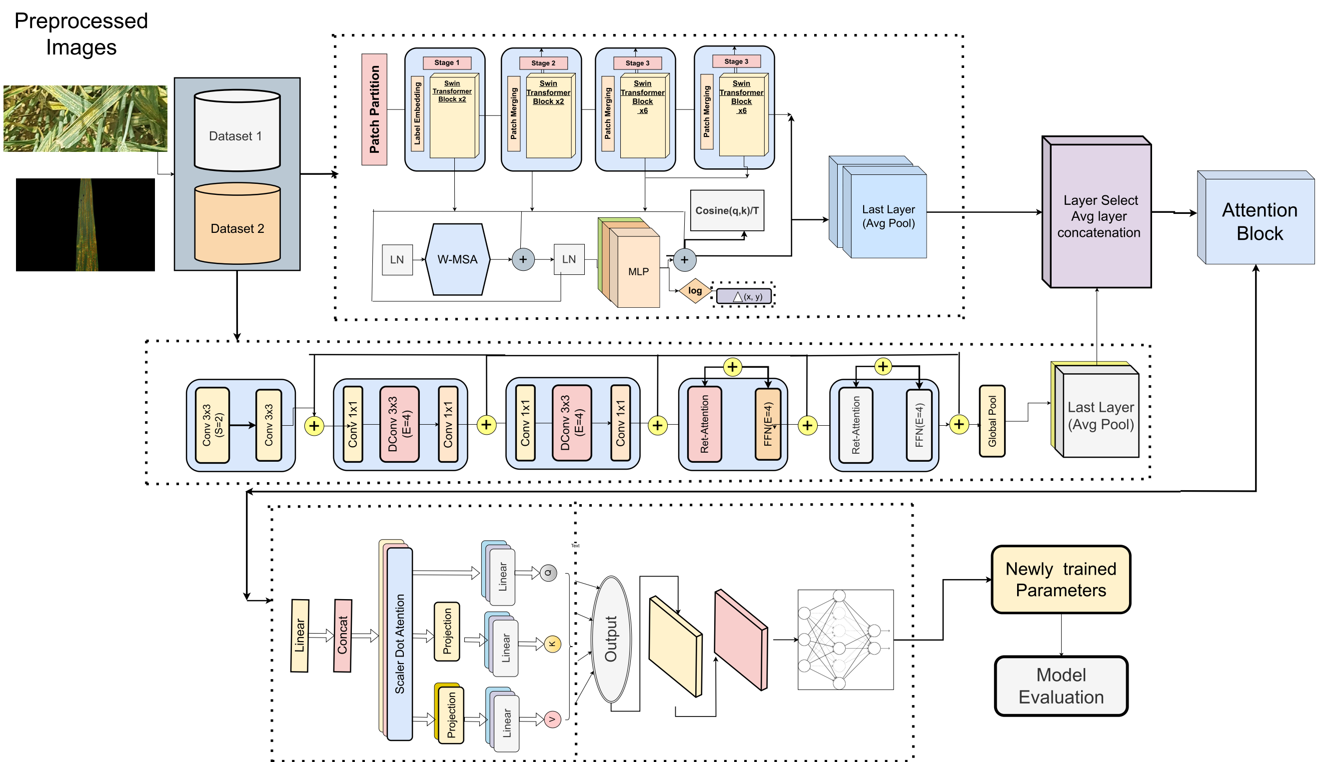


FIGURE 3. Our proposed model architecture for wheat leaf disease classification: leveraging the power of CoAtNet and Swin transformer as feature extractor applying LA mechanism in the classification phase.

In our proposed model architecture, we perform feature fusion to combine the outputs of these two network models which acted as the backbone of this proposed model as feature extractor, denoted as f_1 and f_2, representing the results of CoAtNet and Swin, respectively. Both f_1 and f_2 are $N \times 768$ matrices, where N is the number of samples. The feature fusion is achieved by computing the element-wise average, resulting in a new matrix A written in (1). Feature fusion has been mainly utilized for feature extraction which performs as the main backbone of this model.

$$D(Q, K, V) = A(QK^T W)V. \tag{2}$$

We further use the linear attention layer for classification. The LA layer makes the model more computationally efficient, reduces the danger of overfitting, and provides a more compact representation of the information by lowering the dimensionality of the feature vector from 768 to 64 aiding model interpretation. Q is the query matrix, K is the key

TABLE 5. Our proposed model architecture with feature extractor and classifier components experimenting with three input image dimensions with 4 classes and weight pruning (Freezing last 30 layers for each two models) for plant disease classification merged dataset where TP denotes trainable parameters and NTP denotes Non-Trainable parameters.

Component	Layer Name	Type	Output Shape	TP	NTP
Feature Extractor (Weight Pruning)					
input_112 (InputLayer)	Input	(None, 112, 112, 3)	0	0	0
input_128 (InputLayer)	Input	(None, 128, 128, 3)	0	0	0
input_224 (InputLayer)	Input	(None, 224, 224, 3)	0	0	0
CoAtNet [-30] (Functional)	Functional Model	(None, 768)	14,407,301 (Reduced from 22M)	14,407,301	0
Swin V2 [-30](Functional)	Functional Model	(None, 768)	17,689,560 (Reduced from 27M)	17,689,560	0
average (Average)	Average Layer	(None, 768)	0	0	0
Classification					
linear_attention_layer	Linear Layer	(None, 64)	147,456	147,456	0
output (Dense)	Dense Layer	(None, 4)	130	130	0
Total Parameters				32,244,447	0

TABLE 6. Our proposed model architecture with feature extractor and classifier components experimenting with three input image dimensions with 2 classes and weight pruning (Freezing last 30 layers for each two models) for wheat nitrogen deficiency and leaf rust image dataset where TP denotes trainable parameters and NTP denotes Non -Trainable parameters.

Component	Layer Name	Type	Output Shape	TP	NTP
Feature Extractor (Weight Pruning)					
input_112 (InputLayer)	Input	(None, 112, 112, 3)	0	0	0
input_128 (InputLayer)	Input	(None, 128, 128, 3)	0	0	0
input_224 (InputLayer)	Input	(None, 224, 224, 3)	0	0	0
CoAtNet [-30] (Functional)	Functional Model	(None, 768)	14,407,301 (Reduced from 22M)	14,407,301	0
Swin V2 [-30](Functional)	Functional Model	(None, 768)	17,689,560 (Reduced from 27M)	17,689,560	0
average (Average)	Average Layer	(None, 768)	0	0	0
Classification					
linear_attention_layer	Linear Layer	(None, 64)	147,456	147,456	0
output (Dense)	Dense Layer	(None, 2)	130	130	0
Total Parameters				32,244,447	0

matrix, V is the value matrix, W is a learned weight matrix, and D denotes the number of input dimensions shown in (2).

1) PRUNING METHOD

The weight pruning method utilized in this study is Layer-wise Weight Freezing. This method involves freezing the weights of specific layers within the neural network to reduce the number of trainable parameters. In our approach, we freeze the last 30 layers of each of the feature extractor models (CoAtNet and Swin V2). The last 30 layers of CoAtNet are frozen, reducing the number of trainable parameters from approximately 22 million to 14,407,301. Similarly, the last 30 layers of Swin V2 are frozen, decreasing the trainable parameters from approximately 27 million to 17, 689, 560. An average pooling layer combines the outputs of the feature extractors. This layer has 147, 456 trainable parameters. The output layer, is a dense layer with 4 output classes, having 130 trainable parameters. Freezing layers significantly reduces the number of parameters that need to be updated during training, lowering computational and memory requirements. With fewer parameters to optimize, the training process is accelerated, enabling faster model iterations and tuning. This method effectively utilizes pre-trained weights, which stabilizes the learning process for new tasks, particularly beneficial when the new dataset is limited in size. By limiting the model's complexity through frozen layers, overfitting is mitigated, enhancing the model's generalization capabilities on unseen data. This method leverages the robustness of pre-trained models, offering advantages in terms of computational efficiency and regularization while requiring careful selection of layers to balance efficiency and performance. Determining which layers to freeze requires careful consideration and experimentation in the model architecture of our local and global nodes in FL architecture. Freezing too many layers might hinder the learning process. Conversely, freezing too few may not yield significant computational benefits [44]. Therefore we keep the freeze of the last 30 layers. In this pruning strategy, the same deep ensemble process of that proposed model combining the model average predictions is employed to obtain high accuracy and robust uncertainty estimations and averaging their forecasts. Therefore, this approach effectively captures both evidential and algorithmic uncertainty.

2) PTQ INTEGRATION

We further employ post-training quantization into federated learning. The workflow consists of four main steps: initial model training, optimization at the central server, model distribution, and model aggregation.

- At the central server, an initial model is trained on available centralized data. This step serves as the starting point for subsequent optimization techniques.

- Following initial training of the local nodes, layer-wise weight pruning is applied to the proposed model to remove redundant or less informative parameters. After this, that is followed by post-training quantization, which reduces the precision of model weights and activations, thereby reducing memory and computational requirements.
- Our optimized model comprising pruned and quantized parameters is distributed to all federated nodes for further local training.
- After local training on each node, model updates in the form of weights and gradients are shared with the central server. The central server aggregates these updates applies necessary optimization techniques including pruning and quantization and distributes the updated model back to the nodes.

After completing full training in the global model, we reduce neural network complexity using the Post-training quantization (PTQ) pruning strategies. We utilize the post-training quantization (PTQ) [45] method that optimizes and compresses trained models to reduce memory footprint improving inference speed. We implement PTQ to convert model weights and activations from higher precision (e.g., 32-bit floating point) to lower precision (e.g., 8-bit integers) after training. This technique reduces model size and enhances computational efficiency without significantly compromising performance. The benefits of PTQ include substantial reductions in model size, making them suitable for devices with limited storage, and faster inference due to lower precision computations. Additionally, quantized models consume less power, beneficial for battery-operated devices and large-scale data center deployments. However, PTQ can lead to accuracy loss, especially in models sensitive to precision changes. To address this, we carefully calibrate and fine-tune our models.

Algorithm 1 Proposed Federated Learning for Wheat Leaf Disease Identification Using FedMax Principle

Input:
  $\boldsymbol{\theta}_g$: Initial global model parameters
  R: Number of communication rounds
  $\mathcal{D}$: Dataset containing clients
  K: Number of clients selected in each communication round
  η: Learning rate
  T: Number of local training iterations
Output:
  $\boldsymbol{\theta}_g$: Final global model parameters
Initialize $\boldsymbol{\theta}_g$
for $r \leftarrow 1$ to R **do**
  Select K clients from $\mathcal{D}$
  for each $k \in K$ **do**
    Split k into $(\mathcal{T}_k)$ and $(\mathcal{V}_k)$
    Initialize $\boldsymbol{\theta}_k$
    for $t \leftarrow 1$ to T **do**
      Train $\mathcal{T}_k(local)$:
        $P_{\text{class}_n} = \frac{e^{z_i}}{\sum_{j=1}^{N} e^{z_j}}(L * W_o + b_o)$
        $\boldsymbol{\theta}_{k_t} \leftarrow \boldsymbol{w}_{k_t}$
        $\boldsymbol{w}_{k_t+1} \leftarrow \boldsymbol{w}_{k_t} - \eta \nabla_{\boldsymbol{w}} \text{Loss}(\boldsymbol{w}_{k_t}, \mathcal{T}_k)$
        $\boldsymbol{b}_{k_t+1} \leftarrow \boldsymbol{b}_{k_t} - \eta \nabla_{b} \text{Loss}(\boldsymbol{w}_{k_t}, \mathcal{T}_k)$
    end for
  end for
  On Each Client:
    Initialize $\boldsymbol{X}_k$
    $\boldsymbol{\theta}_{r+1} \leftarrow \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{\theta}_k$
    $\mathcal{V} \leftarrow \sum_{k=1}^{K} \text{score}(\boldsymbol{X}_k) \times \boldsymbol{\theta}_k$
  On Global Server:
    $\mathcal{V} \leftarrow \sum_{k=1}^{K} \mathcal{V} \times \boldsymbol{\theta}_k$
    $\boldsymbol{\theta}_g \leftarrow \text{GlobalUpdate}(\boldsymbol{\theta}_g, \mathcal{V})$
end for

IV. PROPOSED FEDERATED LEARNING MODEL FOR WHEAT LEAF DISEASE IDENTIFICATION

Here, we use our proposed model in the distributed FL system for effective data transfer. We apply our own FL architecture for reducing dynamic allocation applying more efficient utilization of our global server device.

For the local training, the process starts with individual participants methodically training their local models which can be referred to as node devices of each region, deriving significant insights and patterns from their particular datasets. On the other hand, central servers incorporate high-performance hardware and software components suited for efficient data processing and storage, ensuring the system's usefulness and scalability in the context of our wheat leaf disease detection. The central server functions as a specialized computing device specifically tailored for the role. It operates as a center for collecting, analyzing, and aggregating the model weights and disease categorization results submitted by all participants.

Regarding the data storage, the main server efficiently manages (Hierarchical Data Format version 5) HDF5 format files, allowing for various data kinds and architectures. Common data types include integers (8, 16, 32, and 64-bit), floating-point numbers (single and double precision), complex numbers, strings, arrays, matrices, and compound kinds for structured data. These data weights go to the metadata entry creation for keeping data categorization. For each uploaded HDF5 file, a metadata item is created following the defined format. This entry is essentially an organized set of metadata that describes the file and its content. Our advanced proposed architecture of this data management pipeline with the central server at its core, is presented in Figure 4. The global server, responsible for the model aggregation of the local combined weights, successfully updates the global model with the model weights. The global server takes up a correct classification for the disease images with the secured model weights from the 1 to N number of clients with their selected classes. This comprehensive solution, coupled with the FedMax algorithm for local model weight selection, aligns with the strengths of the data management process in a more structured manner.

The incorporation of the HDF5 (Hierarchical Data Format version 5) data management pipeline with our proposed model improves the system's practical use and ability to handle huge datasets needed for deep learning tasks in wheat leaf disease identification, making it more scalable. HDF5 is a flexible file format specifically created for effective storing and organization of large amounts of data, with a hierarchical structure similar to a file system. This architecture enables the organization of complicated datasets, such as annotated photos of different variants of wheat leaf diseases, into datasets grouped that are conveniently accessible during model training and inference. The ability of HDF5 to enable quick access to specific portions of the dataset without requiring the complete file to be loaded into memory is a notable benefit. During the training phase, mini-batch processing can take advantage of this capability, enabling the model to efficiently obtain batches of images and their accompanying labels from the HDF5 file. This modification reduces the impact of input/output bottlenecks, hence improving the overall efficiency of the training process into our FL system architecture. HDF5 files are well-suited for federated learning environments because of their portability and flexibility. In our FL system, every node has the capability to store its own local data in HDF5 files, which can be utilized for local training without the need for substantial modifications to the data management process. The inclusion of local storage functionality guarantees the confidentiality and protection of data, in accordance with the principles of federated learning, which maintains data decentralization and only shares model changes. Machine learning frameworks like TensorFlow and PyTorch offer built-in support for HDF5, permitting smooth integration with data loaders that handle batching and shuffling of data during training. In our FL system, each node keeps its local HDF5 files holding the training data. The local model weights are trained on these datasets, and only the model weight updates not taking raw data are transmitted to the central server.

The objective of the FedMax algorithm is to minimize the global loss function $g(w)$, where w represents the global model parameters [46]. This global loss function is defined as follows:

$$\min_{w} g(w) = \sum_{k=1}^{m} p_k * g_k(w_k), \tag{3}$$

where, $g_k(w_k)$ mainly defines the local objective, which usually is utilized to make the predictions that are made with the local model parameters w_k. The system involves m devices or nodes selected for each communication round, where $(m = C * M)$, with C defining the ratio of selected devices. In our case this is our datasets gathering from each organization and M represents the total number of devices/nodes attending as local client-server. The notable consideration is the p_k match the condition for $(\sum_{k=1}^{M} p_k = 1)$, with $(p_k = \frac{n_k}{n})$, where (n_k) sample data(images) available on device or node k, and $(n = \sum_{k=1}^{M} n_k)$ is the total number of samples across the local clients. Recurrent development and the constant update of model weights from the central model to the local model considerably enhance interpretation while ensuring the correctness of our proposed model's effectiveness. The weighted average approach efficiently addresses this difficulty by applying a feature selection strategy for local model updates, giving weights based on their accuracy. As a result, multiple entities, such as farmers or organizations, can transmit and receive highly accurate categorization results through this strategy, which is strengthened by the centralized aggregated results.

A. INCORPORATING WEIGHTED AGGREGATION

In the FedMax algorithm, we address the issue of underperforming local models by applying a weighted aggregation strategy. Each client's contribution to the global model is weighted based on its validation performance. Specifically, the weight w_k for client k is determined by its validation accuracy a_k, normalized across all participating clients:

$$w_k = \frac{a_k}{\sum_{j=1}^{K} a_j}. \tag{4}$$

This ensures that clients with higher performance have a greater influence on the global model update, thereby reducing the impact of underperforming clients.

Our proposed Algorithm 1 for wheat leaf disease involves representing different entities and operations. The federated learning algorithm aims to collaboratively train a global model for wheat leaf disease detection across multiple clients while preserving data privacy. The process begins by initializing the global model parameters, denoted as $\boldsymbol{\theta}_{g,0}$. In each iteration, a subset of clients is selected from the overall dataset $\mathcal{D}$. For each selected client, its data is split into a training set $(\mathcal{T}_k)$ and a validation set $(\mathcal{V}_k)$. The local model parameters $(\boldsymbol{\theta}_{k,0})$ are then initialized providing an iterative training process that occurs over T iterations. During training, the local model parameters are updated along with $\nabla_{\boldsymbol{w}}$Loss, ∇_{b}Loss gradients of the loss function with respect to the weights and bias. After completion, they are aggregated across all clients to update the global model $(\boldsymbol{\theta}_{r+1})$. Each client computes local scores on its validation set using the updated global parameters using the feature selection principle of the FedMax algorithm. These local scores are then aggregated on the global server with weights assigned to each client resulting in updated global model parameters $(\boldsymbol{\theta}_g)$.

We conduct our FL approach with two datasets for wheat leaf disease classification where in the local server, we feed our pruned proposed transformer model into the system for the training in a non-independent and identically distributed (non-iid) manner.

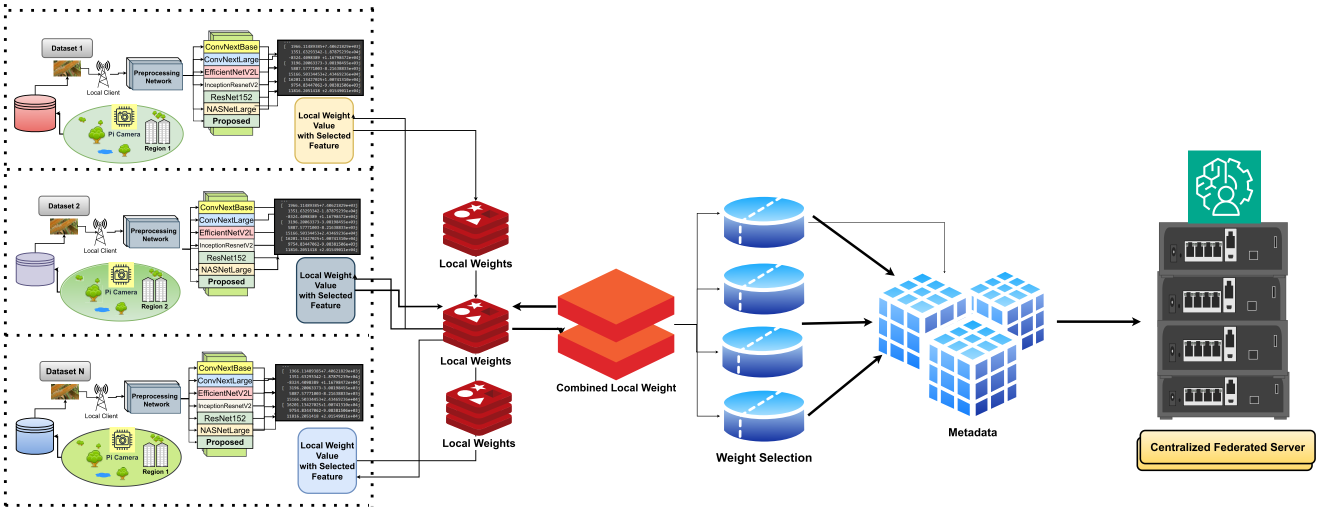


FIGURE 4. **Schematic representation of the proposed data management pipeline, emphasizing the central server's role in aggregating model weights and disease categorization results from local nodes, facilitated by the FedMax algorithm for optimal global loss minimization in wheat leaf disease detection.**

V. PERFORMANCE EVALUATION

The efficacy of each model has been scrutinized with a variety of performance evaluation metrics. To ensure that the models' proficiency is meticulously examined, evaluation metrics include accuracy curves, confusion matrices, classification reports, and Area under the ROC Curve(AUC) scores, Cohen Kappa, and inference times of all models. The accuracy curve shows the model's best attainable accuracy, and its linearity indicates its capability as a classification model. If the line is seamless, the classifier is more appropriate. By comparing the actual labels with the anticipated class identifiers, the confusion matrix makes it easy to monitor the models' accuracy and mistakes. Later the inference time is calculated for the computational efficiency.

The precision (P) is here defined as the proportion of accurately predicted outcomes compared to the number of positive instances. In other words, it measures forecast accuracy and can be expressed mathematically as follows.

$$P = \frac{T_p}{T_p + F_p}. \tag{5}$$

The recall (R) measure is calculated by dividing the number of desired results by the number of initial class evaluations.

$$R = \frac{T_p}{T_p + F_n}. \tag{6}$$

The *F1*-score is an individual metric computed by averaging precision and recall.

$$F1 = \frac{2 \times P \times R}{P + R}, \tag{7}$$

where $T_p =$ is the true positive, F_p is the false positive, F_n is the false negative, and T_n is the true negative.

The Cohen Kappa result can be explained as follows: values 0 denote no agreement, 0.01-0.20 indicate no to little agreement, 0.21-0.40 indicate reasonable agreement, 0.41- 0.60 indicate moderate agreement, 0.61-0.80 indicate substantial agreement while 0.81-1.00 denote almost perfect agreement.

$$\kappa = \frac{P_o - P_e}{1 - P_e}, \tag{8}$$

where the observed agreement, or P_0, is the percentage of instances in which both raters concur and the expected agreement, or P_e, is the percentage of occurrences in which both raters would be anticipated to agree purely by chance.

The AUC score indicates the area under the Receiver Operating Characteristic (ROC) curve, which is a graphical representation of the model's performance as the discrimination threshold is modified. It assesses the model's overall ability to discriminate between the two classes (positive and negative) across different threshold levels. AUC of 0.5 implies random performance, whereas AUC of 1.0 represents perfect discrimination.

In the above set of equations, T_p is True Positive, F_p is False Positive, F_n is False Negative, and T_n is True Negative.

Here inference time defines as T (also known as inference latency or prediction time) is the amount of time it takes a machine learning model to process input data and make predictions. It assesses the model's computational efficiency throughout the inference phase. In the context of deep learning models, such as neural networks, the inference time formula is as follows:

$$T = \frac{1}{S}, \tag{9}$$

where:

- Here S denotes as Inference Speed. The model's ability to process input data and make predictions at a rapid pace. It is the inverse of inference time and is usually measured in frames per second (FPS) or inferences per second (IPS).

GPU utilization is defined as U depicts the ratio of the workload or tasks allocated to the GPU to its full capacity.

$$U = \frac{T_b}{T_t} \times 100\%, \tag{10}$$

where:

- Time GPU is busy computing (T_b): The duration the GPU is actively processing data or performing computations.
- Total time interval (T_t): The total duration of observation.

VI. EXPERIMENTAL SETUP

All computations in our experimental setup are performed within a Windows 11 environment, harnessing the processing capabilities of two GPU accelerators: the NVIDIA® GeForce RTX 3060 Ti and the Tesla T4 with 16GB of dedicated memory. Our CPU contains an AMD RyzenTM 9 having 5900X, which ensures fast processing. TensorFlow version 2.11.0, a popular choice in the deep learning community recognized for its stability and versatility, is used for our training. TensorFlow Federated (TFF), an open-source framework, is used to build up and run our federated runtime environments. We set the total number of training epochs to 50 to handle the training process efficiently, while simultaneously implementing early stopping measures. This choice helps to prevent overfitting and directs the model toward optimal generalization. We have tried different batch size variations and finally choosing a batch size of 32 is prompted by a desire to strike a compromise between training speed and memory use. We include a global average pooling layer in the network architecture to minimize the spatial dimensions of the feature maps before connecting to the fully linked layers. A dropout rate of 0.4 is added as well to reduce overfitting and promote model robustness. To control model complexity and enhance the learning of critical characteristics, both L1 and L2 kernel regularization techniques are used for regularization. In terms of optimization, we utilize the Adaptive Gradient Algorithm (Adagrad) optimizer, which is well-known for its success in training deep neural networks [47]. through learning rate changes in each training parameter The learning rate, which was set at 0.00001, is essential to the model's convergence and overall performance. This fine-tuning of the learning rate is vital for achieving the optimal balance of convergence speed and stable training dynamics. The details are shown in Figure 7. Our study assesses the statistical significance of our method in contrast to others by analyzing performance measures using a one-sided paired t-test. Before undertaking this investigation, we assessed the normality of the data using the D'Agostino-Pearson test.

VII. RESULT ANALYSIS

Our study focuses on evaluating our performance in non-independent and identically distributed (Non-IID) scenarios. This emphasis arises from the basic characteristics of agricultural datasets, as we have two datasets with multiclasses in separate places. Agricultural data frequently demonstrates spatial heterogeneity, with different regions of a field having distinct features that influence crop health. Soil composition, solar exposure, and moisture levels can all vary dramatically throughout a field. Furthermore, temporal variability is common, influenced by factors such as weather patterns and seasonal variations.

We evaluate the performance of our model within our global FL architecture by locally running two datasets. The first dataset focuses on wheat leaf illness and has five separate classes, whereas the second dataset similarly focuses on wheat leaf disease but only includes two classes. Both datasets' performance metrics are computed, and the results are shown below.

A. PLANT MERGED DISEASED DATASET

A comparative analysis is conducted to evaluate the proposed model against various pre-trained models for the three input dimension testing 112x112,128x128,224x224, utilizing performance metrics such as accuracy, precision, recall, F-1 score, AUC, and Cohen's Kappa. Among the pre-trained models, ConvNeXtBase stands out with a 0.92 accuracy, closely followed by ConvNeXtLarge at 0.94 with an improved score, EfficientNetV2L at $0.89 - 0.92$ range, and NasNetLarge at 0.86-0.88, demonstrating their competence. In contrast, InceptionResNetV2 displays lower accuracies for all input dimensions than other models, signifying limitations in handling the dataset effectively. The proposed model emerges as the top performer, achieving a remarkable accuracy of keeping a range of $0.96 - 0.98$ for all three input dimensions, consistently high precision, recall, and F-1 score, a strong AUC of $0.94 - 0.97$, and a Cohen's Kappa score of 0.95. Table 8 shows this analysis by utilizing all class values in a combined state for this dataset.

Additionally in the confusion matrices, the proposed model outperforms its counterparts by consistently achieving true positive for all classes demonstrated in Figure 5, which other models struggled with. This highlights the robustness of the proposed model in classifying all classes effectively, making it a standout choice for our system framework. The loss curve for this data shows a trend among the six pre-trained models. All six pre-trained models show a consistent convergence phenomenon to a plateau, starting with rather high initial loss values and progressively dropping. Still, they are never above a specific threshold, which results in continuous variations during training. This pattern highlights their inability to make major progress. The proposed model, on the other hand, shows impressive properties. It begins with a much lower

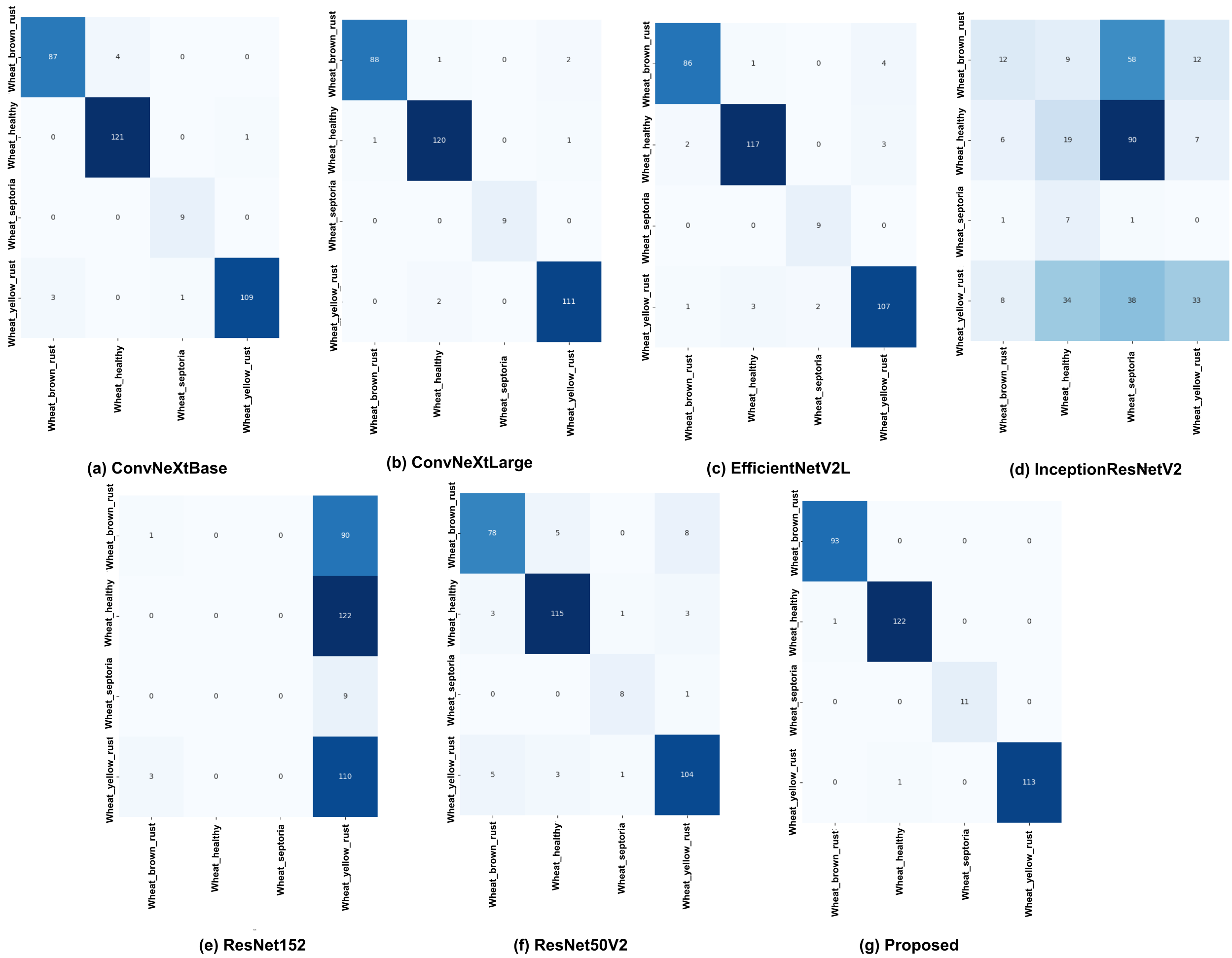


FIGURE 5. **Confusion matrix of plant disease classification merged dataset for wheat leaf disease classification for global server for (a) ConvNeXtBase (b) ConvNeXtLarge (c) EfficientNetV2L (d) InceptionResNetV2 (e) ResNet152 (f) ResNet50V2 (g) Proposed models.**

initial loss value and quickly converges to a much lower loss level as displayed in Figure 10 (a). Throughout the training phase, this behavior maintains a consistent variation around this lower value, which can be characterized as quick convergence to a lower loss. Regarding the accuracy curves for the dataset, the proposed model consistently maintains accuracy levels higher than most of the pre-trained models by the end of the training as shown in Figure 9 (a). This trend indicates that the proposed model's learning capabilities extend to accuracy as well. The proposed model, while not displaying any exceptional behavior, consistently maintains accuracy levels higher than most of the pre-trained models by the end of the training as shown in Figure 9 (a). This trend indicates that the proposed model's learning capabilities reflect its accuracy as well. Figure 7 depicts the ROC curves of all models where our proposed model 7 (g) performs better than other models in the FL system. We also perform ablation study experiments on this dataset with different dropout and batch size variations where dropout 0.2 with batch size 32 giving better results in terms of accuracy shown in Figure 17 (b).

TABLE 7. **Experimental setup summary.**

Component	Specification
Operating System	Windows 11
CPU	AMD RyzenTM 9 5900X
GPUs	NVIDIA® RTX 3060 Ti, Tesla T4
TensorFlow Version	2.11.0
Framework	TensorFlow Federated (TFF)
Total Training Epochs	50
Batch Size	32
Dropout Rate	0.4
Regularization	L1 and L2 kernel regularization
Optimizer	Adaptive Gradient Algorithm
Learning Rate	0.00001

B. WHEAT NITROGEN DEFICIENCY AND LEAF RUST IMAGE DATASET

In another dataset (Wheat Nitrogen Deficiency and Leaf Rust Image), a comprehensive comparative analysis is conducted to assess the proposed model in contrast to various pre-trained models, using the same metrics in the global FL system. Notably, ConvNeXtLarge and ConvNextLarge emerge

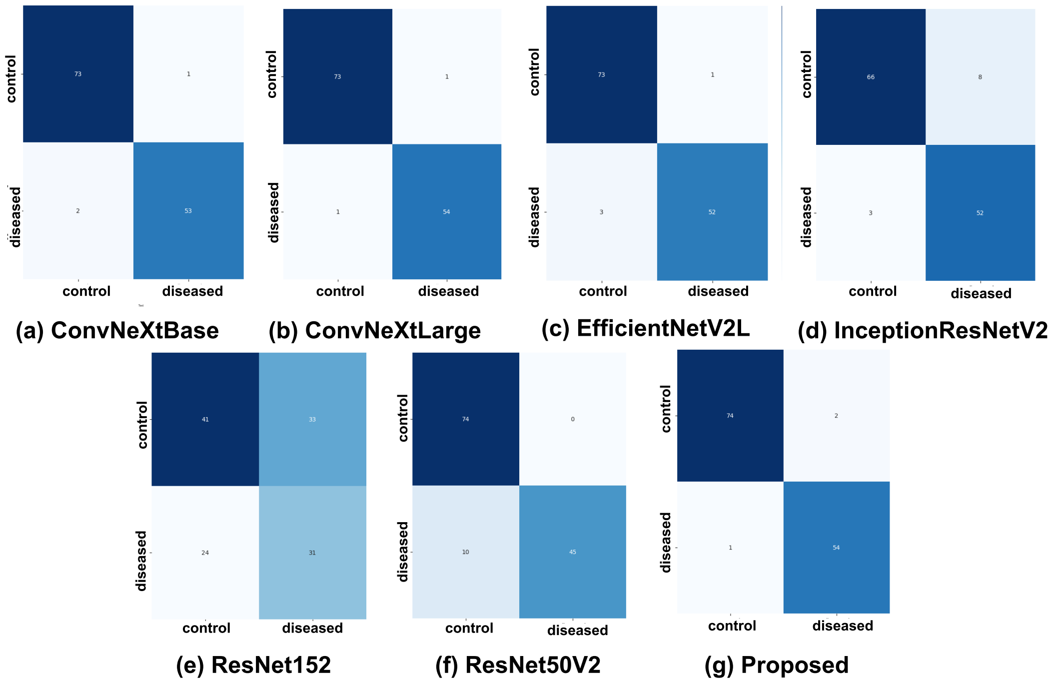


FIGURE 6. Confusion matrix of wheat nitrogen deficiency and leaf rust image dataset for wheat leaf disease classification on the global server for (a) ConvNeXtBase (b) ConvNeXtLarge (c) EfficientNetV2L (d) InceptionResNetV2 (e) ResNet152 (f) ResNet50V2 (g) Proposed models.

TABLE 8. Comparison of models on wheat nitrogen deficiency and leaf rust image for wheat leaf disease classification for plant merged diseased dataset on global server.

Model	Input Dimension	Accuracy (%)	Precision	Recall	F-1	AUC	Cohen's Kappa
			Transfer Learning Models				
ConvNeXtBase	112	0.92	0.91	0.93	0.90	0.91	0.90
	128	0.90	0.88	0.91	0.87	0.88	0.86
	224	0.92	0.91	0.91	0.90	0.91	0.93
ConvNeXtLarge	112	0.94	0.93	0.92	0.90	0.91	0.92
	128	0.91	0.93	0.95	0.91	0.92	0.91
	224	0.92	0.93	0.93	0.91	0.91	0.90
EfficientNetV2L	112	0.89	0.86	0.89	0.86	0.85	0.85
	128	0.91	0.92	0.94	0.89	0.95	0.94
	224	0.92	0.87	0.86	0.86	0.91	0.90
InceptionResnetV2	112	0.77	0.83	0.74	0.71	0.77	0.79
	128	0.78	0.83	0.74	0.71	0.84	0.79
	224	0.76	0.84	0.73	0.72	0.83	0.80
ResNet152	112	0.91	0.90	0.89	0.88	0.88	0.87
	128	0.91	0.93	0.89	0.85	0.88	0.86
	224	0.91	0.93	0.89	0.88	0.88	0.88
NASNetLarge	112	0.89	0.89	0.93	0.91	0.90	0.91
	128	0.88	0.90	0.90	0.86	0.88	0.86
	224	0.86	0.86	0.90	0.84	0.86	0.88
Proposed	112	0.97	0.96	0.97	0.95	0.92	0.96
	128	0.96	0.95	0.94	0.91	0.97	0.94
	224	0.98	0.98	0.98	0.92	0.96	0.95

good performance by achieving a range of $0.89 - 0.91$ considering all input dimension sizes. Following closely, EfficientNetV2L exhibits an accuracy span of $0.89 - 0.94$. Furthermore, other metrics show fair results for these models. Conversely, InceptionResNetV2 yields a relatively lower accuracy of 78%, suggesting certain limitations in handling this dataset too for all these dimensions. Furthermore, ResNet152 and NasNetLarge keep a good accuracy limitation

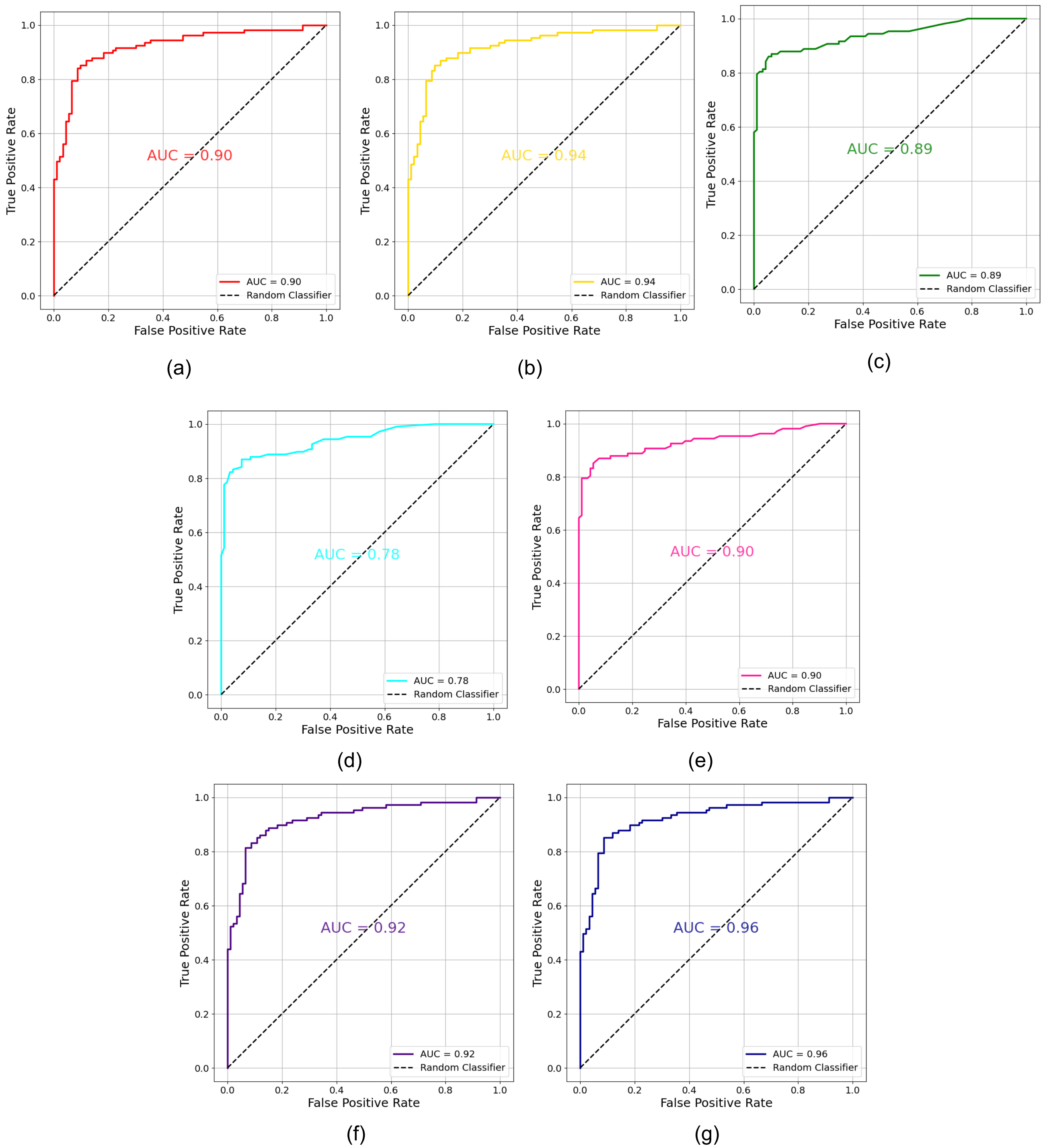


FIGURE 7. ROC curves with plant merged diseased dataset with four classes on the global FL server for (a) ConvNeXtBase (b) ConvNeXtLarge (c) EfficientNetV2L (d) InceptionResNetV2 (e) ResNet152 (f) ResNet50V2 (g) Proposed models.

of 0.90-0.95 scope. Other metric also performs better in terms of assessing the performance of each model. However, the proposed model shows the best results in its counterparts with remarkable accuracy of 0.97-0.99 -and consistently high precision, recall, F-1 score, AUC, and a Cohen's Kappa score of 0.94-0.97. The detailed information of all selective model results on various models in three input sizes of training are shown in Table 11. Additionally among the transfer learning models, both ConvNeXtBase and ConvNeXtLarge exhibit the highest true positives (TP) for the classes, reflecting their relatively strong classification performance. In contrast, the other models fail to attain the same level

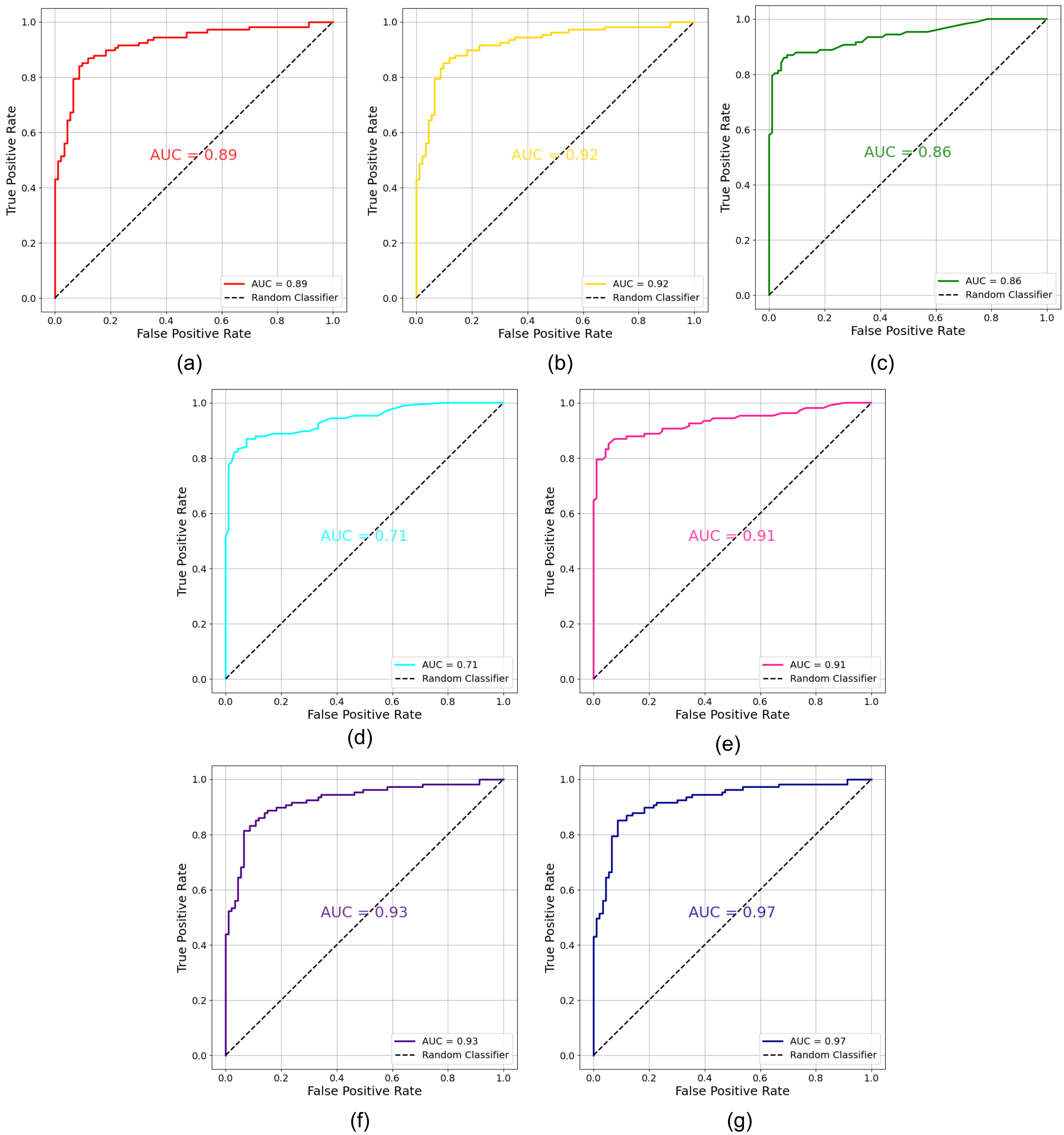


FIGURE 8. ROC curve with the wheat nitrogen deficiency and leaf rust image dataset with two classes on the global FL server assessment for (a) ConvNeXtBase (b) ConvNeXtLarge (c) EfficientNetV2L (d) InceptionResNetV2 (e) ResNet152 (f) ResNet50V2 (g) Proposed models.

of near-perfection in *TP*. However, when considering the proposed model, its confusion matrix shows near-perfect *TP* for both classes shown in Figure 6, setting it apart from the rest. When coupled with its comprehensive classification report, the proposed model demonstrates superior performance in accurately identifying both ''diseased'' and ''control'' instances. Moreover, observing the loss curves illustrated in Figure 10 (b), we notice a comparable pattern in the loss curves of all six pre-trained models. Similarly, the six pre-trained models in this training data show convergence to a plateau phenomenon, starting with somewhat large initial loss values and progressively declining. Moreover, shifting our focus to this accuracy curve is shown in Figure 9 (b), a similar pattern has also emerged. The proposed model's persistent superior performance highlights its applicability to this dataset as well. Besides, Figure 8 depicts the ROC curves

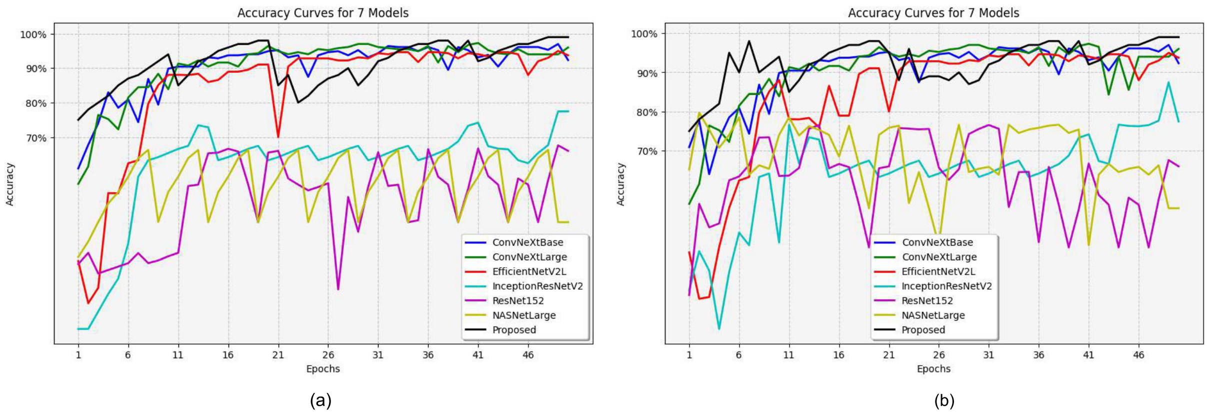


FIGURE 9. Validation accuracy curves with all the TL models along with proposed model on plant merged diseased dataset (a) and wheat nitrogen deficiency and leaf rust image dataset (b) in the global FL server.

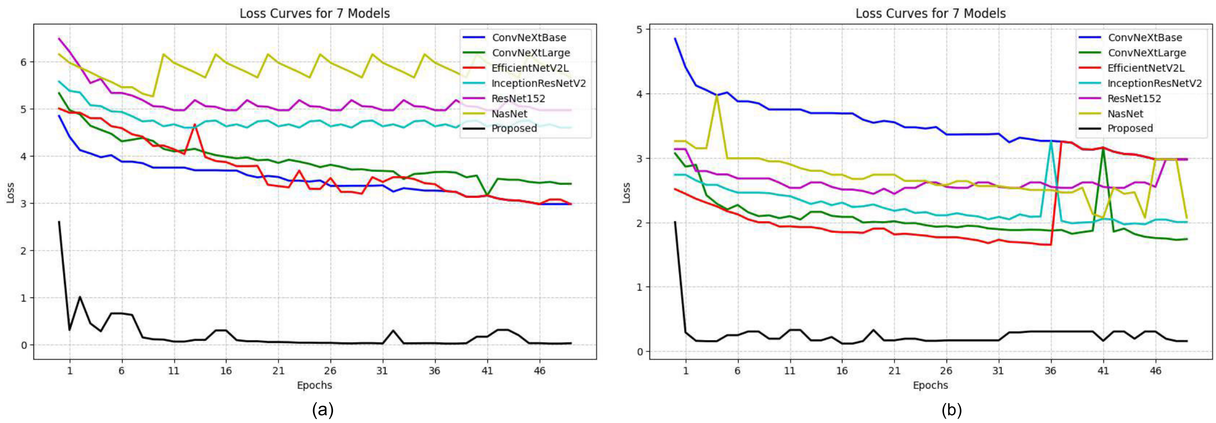


FIGURE 10. Validation loss curves with all the TL models along with proposed model on plant merged diseased dataset (a) and wheat nitrogen deficiency and leaf rust image dataset (b) in the global FL server.

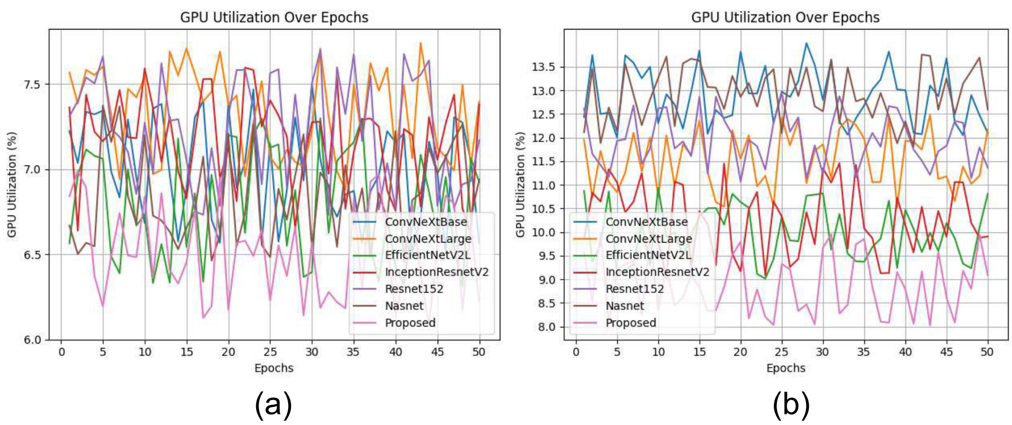


FIGURE 11. GPU utilization measurement over the training epochs of all models where proposed model performs better than other transfer learning models in this distributed federated learning system plant merged diseased dataset (a) and wheat nitrogen deficiency and leaf rust image dataset (b).

of all models where our proposed model 7 (g) performs better than other models in the FL system. Also in this dataset, we conduct ablation study tests on this dataset with different dropout and batch size changes where dropout 0.2 with batch size 32 produces better outcomes in terms of accuracy displayed in Figure 17 (b).

C. OVERALL ACCURACY AND INFERENCE TIME AND GPU UTILIZATION MEASUREMENT

In our global FL setup, we compare the overall performance of selected models across many parameters as shown in Figure 16, including accuracy, precision, recall, F1 score, AUC, and Cohen's Kappa. Notably, our proposed model

TABLE 9. Comparison of models on wheat nitrogen deficiency and leaf rust image dataset for wheat leaf disease classification for wheat nitrogen deficiency and leaf rust image dataset on global server.

Model	Input Dimension	Accuracy (%)	Precision	Recall	F-1	AUC	Cohen's Kappa
			Transfer Learning Models				
ConvNeXtBase	112	0.91	0.92	0.92	0.91	0.93	0.93
	128	0.89	0.89	0.90	0.84	0.90	0.91
	224	0.91	0.90	0.88	0.84	0.90	0.87
ConvNeXtLarge	112	0.91	0.90	0.91	0.90	0.90	0.92
	128	0.90	0.91	0.89	0.88	0.86	0.88
	224	0.91	0.93	0.91	0.88	0.90	0.93
EfficientNetV2L	112	0.92	0.90	0.91	0.90	0.92	0.92
	128	0.89	0.90	0.90	0.89	0.91	0.92
	224	0.94	0.92	0.91	0.90	0.90	0.91
IncepttionResnetV2	112	0.77	0.84	0.74	0.76	0.88	0.78
	128	0.77	0.84	0.76	0.74	0.87	0.77
	224	0.75	0.82	0.72	0.76	0.85	0.79
ResNet152	112	0.90	0.93	0.89	0.88	0.88	0.87
	128	0.91	0.90	0.89	0.86	0.88	0.87
	224	0.89	0.92	0.89	0.85	0.88	0.87
NASNetLarge	112	0.91	0.86	0.90	0.87	0.91	0.92
	128	0.92	0.93	0.91	0.86	0.92	0.91
	224	0.95	0.94	0.92	0.86	0.92	0.93
Proposed	112	0.97	0.97	0.94	0.92	0.98	0.97
	128	0.97	0.96	0.95	0.93	0.98	0.95
	224	0.99	0.97	0.97	0.91	0.97	0.97

TABLE 10. Ablation study on our proposed model of wheat nitrogen deficiency and leaf rust image dataset for wheat leaf disease classification on global server through Post-Training Quantization (PTQ).

Model	Input Dimension	Accuracy (%)	Precision	Recall	F-1	AUC	Cohen's Kappa
			Transfer Learning Models				
Proposed	112	0.82	0.82	0.80	0.80	0.83	0.82
	128	0.82	0.82	0.81	0.80	0.83	0.81
	224	0.84	0.82	0.82	0.79	0.82	0.82

TABLE 11. Ablation study on our proposed model on plant merged diseased dataset for wheat leaf disease classification on global server through Post-Training Quantization (PTQ).

Model	Input Dimension	Accuracy (%)	Precision	Recall	F-1	AUC	Cohen's Kappa
			Transfer Learning Models				
Proposed	112	0.80	0.80	0.78	0.78	0.81	0.80
	128	0.80	0.80	0.79	0.78	0.81	0.79
	224	0.82	0.80	0.80	0.77	0.80	0.80

excels in most measures, with the best accuracy, precision, F1 score, and Cohen's Kappa values, making it a strong choice for accurate centralized diseased class identification. Besides, Table 12 displays the performance metrics of various models on the final global federated model. The proposed model surpasses others in accuracy and inference time, obtaining top-5 accuracies of 99.12%, 99.04%, and 98.09% for input dimensions 112, 128, and 224, respectively. With the reduced weight, with 32 million parameters, it displays the lowest inference speeds on both GPU setups giving a fair performance in the global FL system. Besides, we calculate the GPU utilization percentage throughout the training process of our model in the FL system over 50 epochs. Our proposed model performs better than other models maintaining a lower range between 6-9 % range in both datasets shown in Figure 11.

D. PTQ QUANTIZATION ON OUR SYSTEM MODEL

The ablation evaluation of our proposed model on the Wheat Nitrogen Deficiency and Leaf Rust Image Dataset for Wheat Leaf Disease Classification is being followed by Post-Training Quantization (PTQ), on a global server. The results are summarized in Tables 8 and 11. The ablation study on our proposed model for wheat leaf disease classification

TABLE 12. Performance metrics of various models on the final global federated model calculating overall accuracy and inference time calculation.

Model	Input Dimension	Top-5 Acc	Parameters	Inference Time (ms) GPU(1)	Inference Time (ms) GPU(2)
ConvNeXtBase	112	95.01%	88.5M	800.249	735.249
	128	94.52%	88.5M	854.249	754.267
	224	94.23%	88.5M	814.239	702.787
ConvNeXtLarge	112	91.34%	197.7M	834.688	967.498
	128	93.12%	197.7M	854.249	954.249
	224	93.59%	197.7M	934.387	804.267
EfficientNetV2L	112	94.56%	119.0M	674.892	744.589
	128	95.43%	119.0M	694.249	724.249
	224	95.55%	119.0M	674.340	722.678
InceptionResnetV2	112	74.89%	55.9M	767.789	644.249
	128	74.34%	55.9M	745.249	644.249
	224	74.03%	55.9M	789.259	634.290
ResNet152	112	73.67%	60.4M	788.255	754.249
	128	72.12%	60.4M	778.256	711.239
	224	73.77%	60.4M	768.345	704.123
NASNetLarge	112	95.04%	88.9M	834.789	944.119
	128	94.10%	88.9M	854.249	954.249
	224	93.56%	88.9M	822.219	904.309
Proposed	112	99.12%	32M	644.209	622.241
	128	99.04%	32M	614.223	616.566
	224	98.09%	32M	624.249	644.899

using Post-Training Quantization (PTQ) across two datasets indicated continuous performance gains as input dimensions increased. For the Wheat Nitrogen Deficiency and Leaf Rust Image Dataset, our model has attained 82% accuracy for input dimensions 112 and 128, increasing to 84% for 224, while precision and AUC remained steady at $82 - 83\%$. For the Plant Merged Diseased Dataset, accuracy has reached 80% for input dimensions 112 and 128, increasing to 82% for 224, with precision and AUC continuously between 80 and 81%. The model's recall and F-1 scores improved slightly as input sizes developed. However after training the PTQ mechanism, the performance loss has been observed however it reduces our model complexity in the FL system.

E. 2D DIMENSIONS VISUALIZATION INSIGHTS

We evaluate the performance of our global model of two datasets of wheat leaf disease in 2D dimensions using Principal Component Analysis (PCA), t-distributed Stochastic Neighbor Embedding (t-SNE), and Isometric Mapping (Isomap), to unveil the intrinsic structure and discern any notable patterns that could elucidate underlying disease classifications. PCA performs twice under separate settings in Figure 12 (a) and (d). In the initial PCA 12 (a), a substantial separation has been observed along the first main component. However, this distinction is less obvious along the second primary component. In contrast, the second PCA application 12 (d) has exhibited a denser grouping, implying less distinction between data points based on the top two principal components. The t-SNE approach has identified clusters with high separation in both applications in Figure 12 (b)(e). The initial application Figure 12 (b) reveals unique clusters, whereas the subsequent identified application Figure 12 (e) gives well-defined, non-overlapping clusters. This indicates t-SNE's skill in finding local structures and potentially intricate groupings within the dataset. Isomap's first visualization 12 (c) shows groupings largely distributed along the first dimension, with minor distribution along the second, suggesting a fundamental differentiating feature confined within one dimension. The second Isomap depiction in Figure 12 (f) spreads the data points more broadly, reflecting a more complex manifold structure.

F. COMPARATIVE ANALYSIS OF PLANT DISEASE RECOGNITION MODELS WITH OUR PROPOSED MODEL

Table 13 compares multiple models for plant disease identification, highlighting their characteristics, Federated Learning (FL) support, and performance on two datasets. The models include CNN(general), DenseNet121(CNN) + ViT, ECA-ConvNeXt, EfficientNetB0 + CBAM, Swin Model, and a proposed Vision-Based model with LA Attention Mechanism. Our proposed model stands out, supporting FL and obtaining the maximum performance with an accuracy of 98.99% and precision of 95.98% on the first dataset,

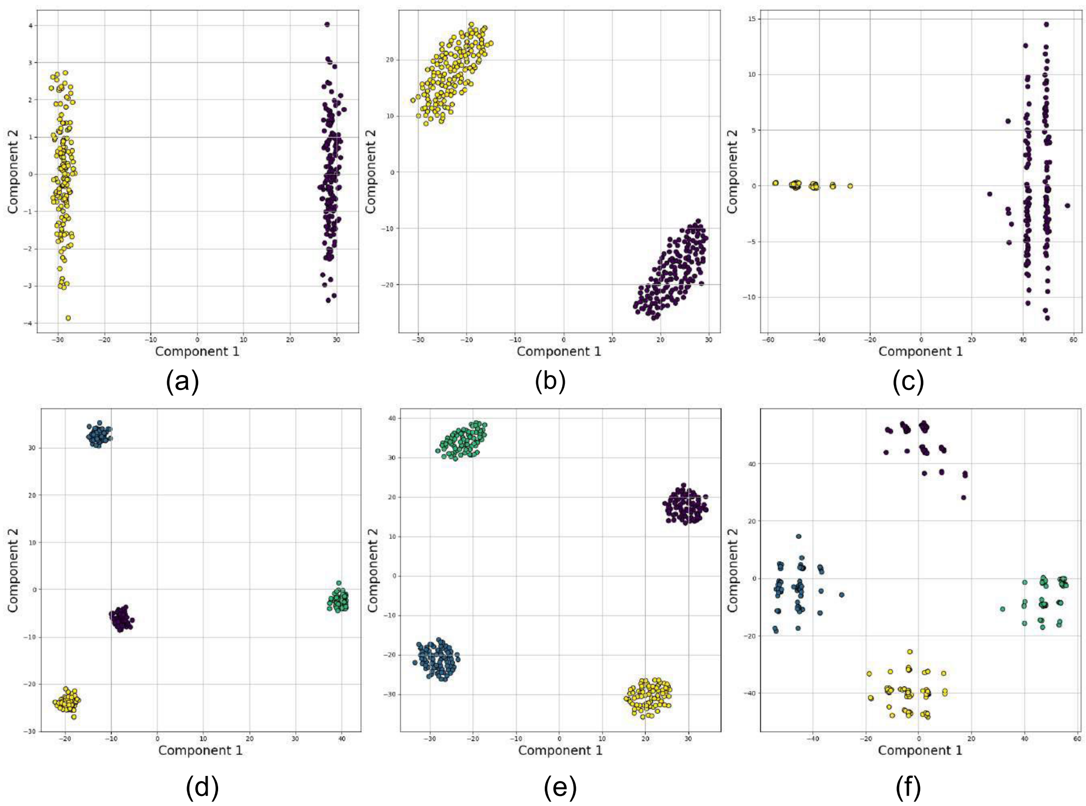


FIGURE 12. Comparative visualization showcasing the application of dimensionality reduction techniques (PCA, t-SNE, Isomap) on two wheat leaf disease datasets, depicted in 2D dimensions. Panels (a), (b), and (c) represent the techniques applied to the plant merged diseased dataset, while panels (d), (e), and (f) correspond to the wheat nitrogen deficiency and leaf rust image dataset. Each technique reveals distinct clustering patterns and structural insights, aiding in the understanding of disease distribution and classification in wheat.

and accuracy of 9.50% (possibly a typo) and precision of 96.24% on the second dataset. In contrast, the other models, such as CNN(general) and DenseNet121(CNN) + ViT, show lesser accuracy and precision, showing the proposed model's advantage in accuracy and FL support for plant disease identification.

Our proposed model provides a unique vision-based technique augmented with a Linear Attention (LA) mechanism, which considerably boosts its capacity to highlight essential aspects inside input images. This complex process, absent in standard convolutional architectures like CNNs (e.g., CNN(general)) or hybrid models such as DenseNet121(CNN) + ViT, equips the model to distinguish nuanced patterns crucial for effective plant disease identification. While existing models have shown respectable performance, they exhibit certain limitations. For instance, CNN-based techniques generally struggle with identifying intricate spatial linkages and long-range dependencies inherent in complex plant diseases. This weakness might hamper their capacity to appropriately recognize small visual clues indicative of specific diseases. Additionally, conventional models could meet difficulty in adjusting to various datasets characterized by changes in lighting conditions, plant species, and disease presentations. Furthermore, the computational needs and resource-intensive nature of existing models could provide hurdles, particularly in resource-constrained environments where access to high-performance computing equipment may be limited. Despite these challenges, our proposed model wired with the use of the Linear Attention process constitutes a considerable leap forward. By enabling more nuanced feature extraction and adaptive attention allocation, our method addresses the drawbacks of prior approaches, boosting the model's robustness and accuracy in disease identification. Moreover, the model's support for Federated Learning (FL) not only enables scalability across dispersed contexts but also facilitates collaborative learning without compromising data privacy. In summary, while existing models exhibit commendable performance, the proposed model's nuanced attention mechanism and federated learning capabilities position it as a promising solution for precise and scalable plant disease identification, essential for bolstering agricultural resilience and food security worldwide. Furthermore, In response to the current focus on model accuracy without sufficient consideration for real-time analysis in decentralized environments, our proposed model proposes a novel approach that integrates federated learning, facilitating collaborative learning across diverse organizations and locations while safeguarding data

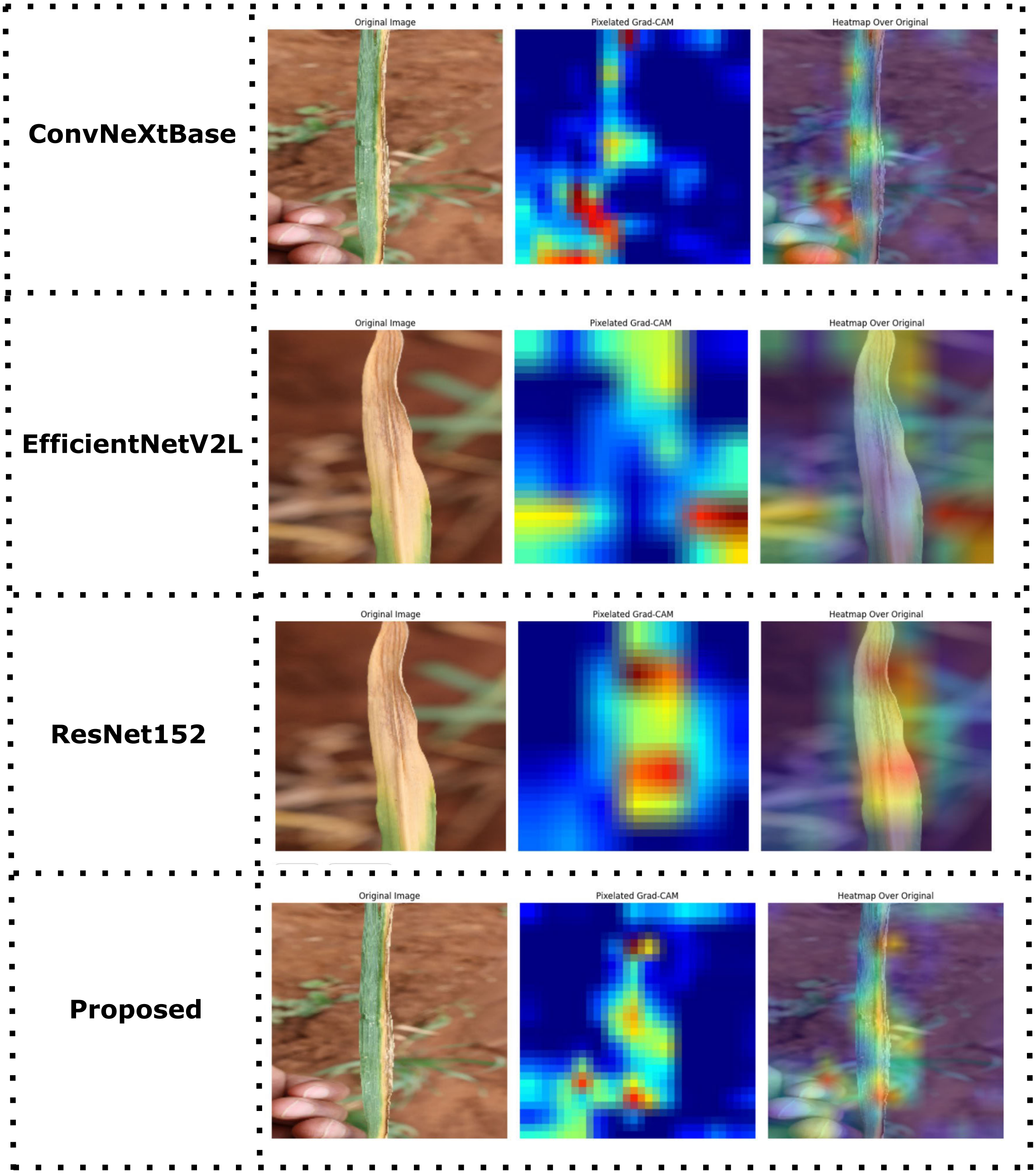


FIGURE 13. Attention maps for pre-trained TL architectures (ConvNeXtBase, EfficientNetV2L, ResNet152) and the model we propose in the FL system. Our model demonstrates a higher priority on critical diseased lesion spots with improved ROI selection in wheat.

privacy. Recognizing the limits of classic transfer learning methods in managing big, diversified datasets due to their high parameter counts, we offer for a lightweight transformer-based model architecture specifically built to perform efficiently in a federated scenario. Moreover, to address the issues given by noise and unpredictability in real-world agricultural data, we apply extensive image preprocessing techniques such as edge recognition, noise addition, and filtering, boosting the adaptability of our model.

G. ATTENTION MAPPING

Furthermore, Figure 13 presents a comparative visualization of attention maps of the pre-trained TL architectures, ConvNeXtBase, EfficientNetV2L, ResNet152, along with our proposed model in the FL system. Our proposed model appears to focus more effectively on the crucial portion of the diseased lesions with better ROI (Region of Interest) selections of the wheat than the other TL model shown in Figure 13.

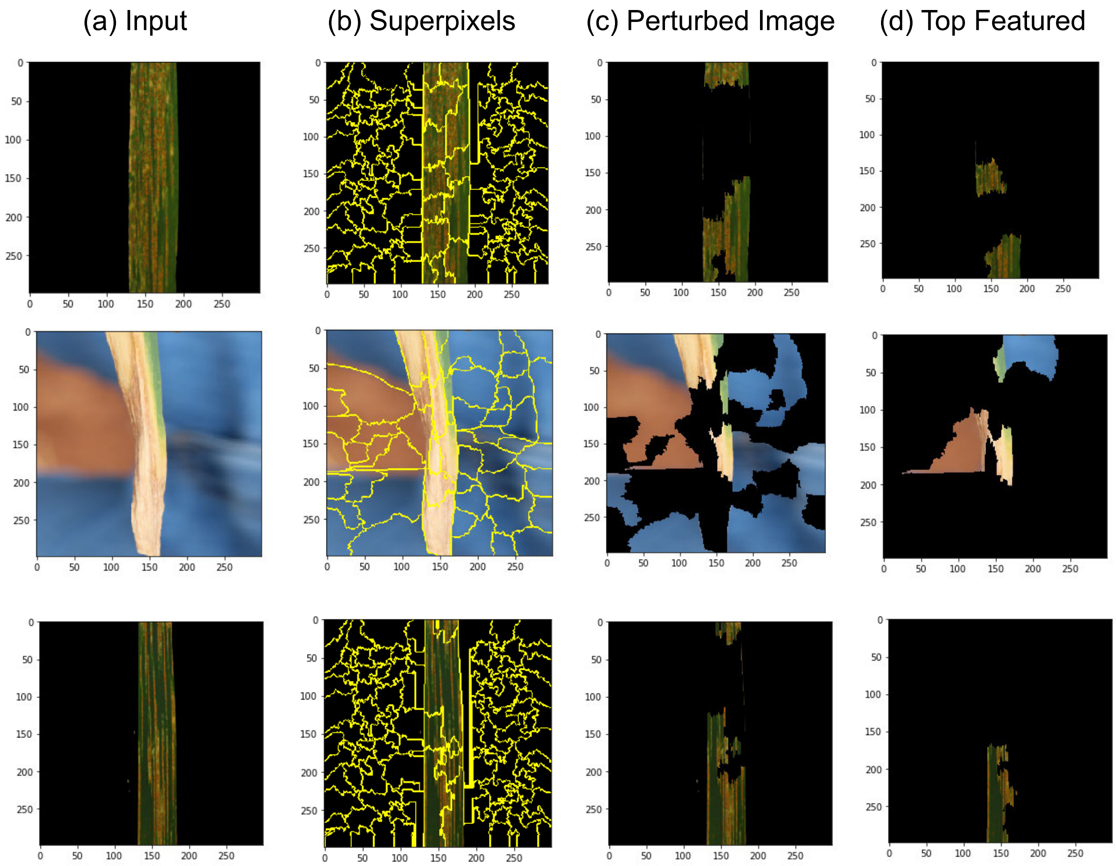


FIGURE 14. Our novel method utilizes LIME (Local Interpretable Model-agnostic Explanations) techniques to interpret our proposed model (a-d). By isolating super-pixels and employing explainable linear models, we provide transparent insights into the classification reasoning of our model.

TABLE 13. Experimentation across reference models demonstrates the efficacy of our proposed technique, particularly in terms of accuracy and support for Federated Learning (FL), positioning it as a promising solution for plant disease identification challenges.

Model Name	Model Feature	FL Support	Accuracy (1st Dataset)	Precision (1st Dataset)	Accuracy (2nd Dataset)	Precision (2nd Dataset)
CNN(general) [27]	Deep Learning-based (with ReLU and softmax functions)	✗	$91.4 \pm 1.5\alpha$	$92.1 \pm 1.1\alpha$	$90.5 \pm 1.2\alpha$	$91.3 \pm 1.4\alpha$
DenseNet121(CNN)+ViT [24]	Hybrid model	✗	$87.1 \pm 1.1\alpha$	$86.2 \pm 2.1\alpha$	$85.9 \pm 2.3\alpha$	$86.0 \pm 2.0\alpha$
ECA-ConvNeXt [25]	Based on ConvNeXt network	✗	$93.7 \pm 1.4\alpha$	$92.4 \pm 2.4\alpha$	$93.0 \pm 1.6\alpha$	$92.7 \pm 1.8\alpha$
EfficientNetB0+CBAM [23]	CNN-Based	✗	$86.2 \pm 2.1\alpha$	$84.3 \pm 1.1\alpha$	$85.0 \pm 1.4\alpha$	$84.6 \pm 1.2\alpha$
Swin Model [48]	Transformer-based	✗	$94.3 \pm 1.2\alpha$	$93.2 \pm 2.4\alpha$	$94.0 \pm 1.3\alpha$	$93.6 \pm 1.8\alpha$
Proposed	Vision Based with LA Attention Mechanism	✓	98.99	95.98	96.50	96.24

Besides, we provide a novel method for interpreting our suggested model utilizing LIME (Local Interpretable Model-agnostic Explanations) techniques. Initially, the image data is read and pre-processed to ensure compatibility with the model that we propose. Subsequently, our model is applied to estimate the class of the input image. Super-pixels are then isolated from the image to aid local interpretation. Through the application of random perturbations, new images are generated, and the classes of these perturbed images are predicted using our model. Distances between the original image and each perturbed image are calculated, followed by the generation of weights using a kernel function, indicating the relevance of each perturbed image in the interpretation process. Leveraging these perturbations, predictions, and

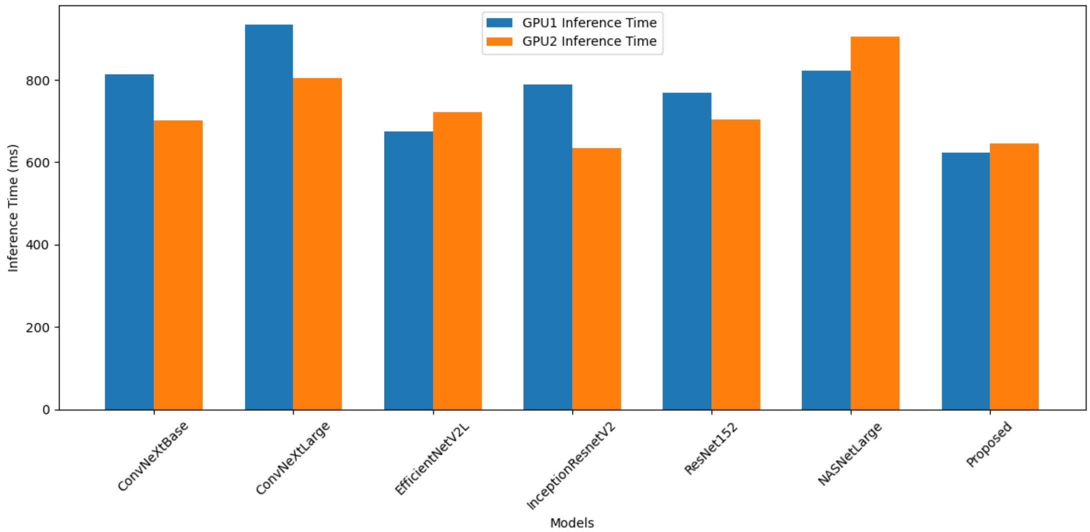


FIGURE 15. Comparision of the models of our proposed and the reference existing models on both datasets based on their aggregated inference time.

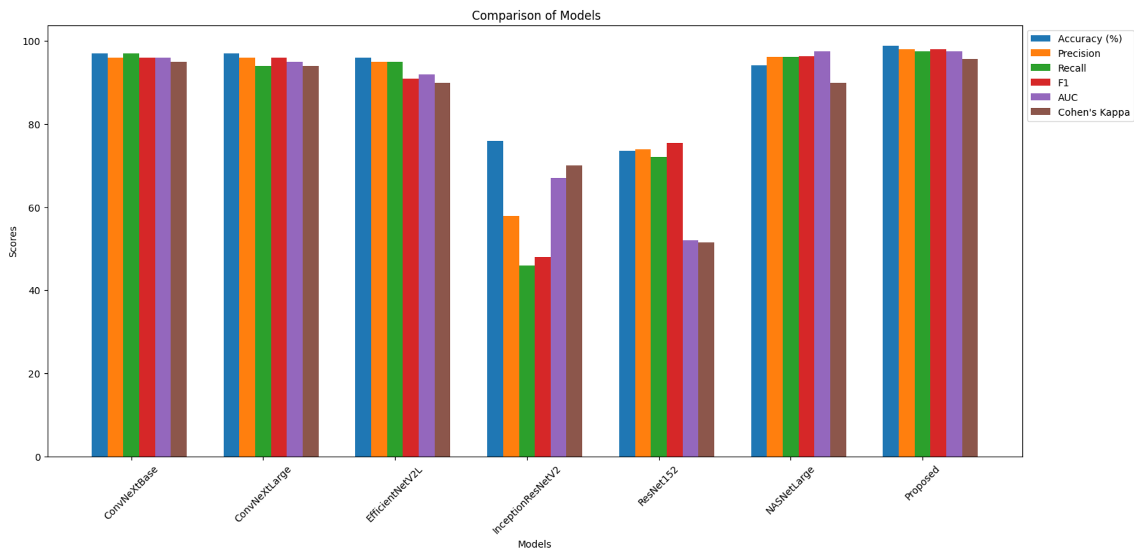


FIGURE 16. Overall performance comparison of our proposed and the existing models on both datasets in the FL system in Accuracy, Precision, Recall, F1, AUC and Cohen's kappa scores.

weights, we develop an explainable linear model to explicate the decision-making process of the underlying classifier. Finally, the top features, represented by super-pixels, are computed to provide insights into the model's classification reasoning. This approach offers a more transparent understanding of our proposed model, boosting its interpretability and dependability for stakeholders. Figure 14 shows the whole process for better visualization of our wheat-diseased images.

VIII. TECHNICAL IMPLEMENTATION OF THE PROPOSED APPROACH

The proposed approach for wheat leaf disease detection involves key technical steps: data preprocessing and augmentation, and model architecture combining CoAtNets and Swin Transformer V2 for enhanced feature extraction and image classification. The federated learning framework allows decentralized model training, preserving data privacy and enhancing scalability by aggregating model updates from multiple nodes. Weight pruning optimizes the model for resource-constrained environments by reducing redundant parameters, thus decreasing model size and computational requirements while maintaining accuracy. This method is particularly effective in federated learning settings with limited bandwidth and storage. However,implementing this approach infrastructure in real-world agricultural settings involves several practical considerations:

A. PRACTICAL CONSIDERATIONS

Here our proposed model in the FL architecture in real-world settings needs several careful considerations.

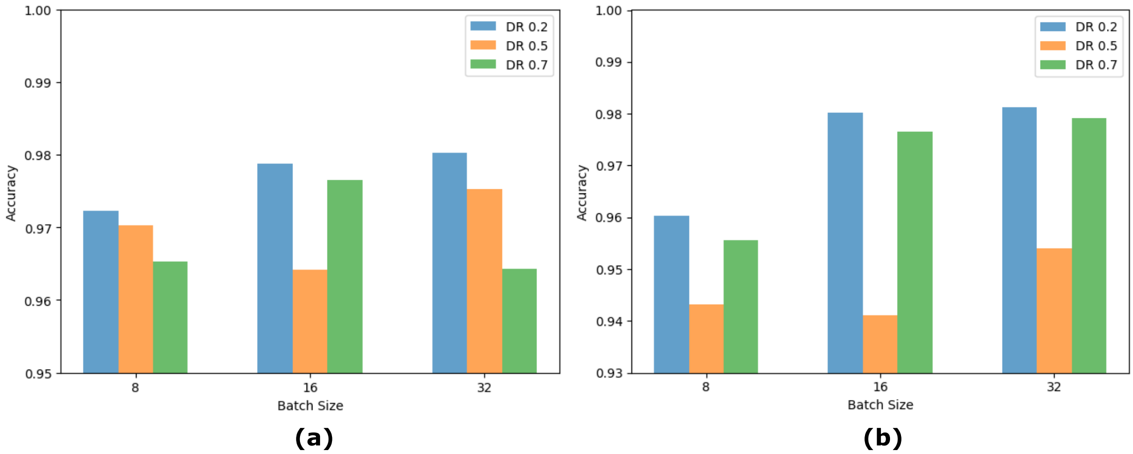


FIGURE 17. Ablations experiments assessment of accuracy with applying different dropout rates (DR) and batch sizes.

1) Network design: Establishing a robust network design is crucial for efficient communication between central servers and local nodes, especially in rural areas where connection may be unreliable. Therefore we can provide hybrid network solutions. Although wired internet connections such as DSL or fiber optics are preferred due to their reliability and speed, they are frequently unavailable in rural places. Also, cellular networks, such as 3G, 4G, and upcoming 5G technologies, can be used as a reliable backup or primary connection option in areas with sufficient signal strength. Using LTE/5G routers helps ensure connectivity for edge devices.
2) Protocols for Data Communication: Now focusing for the communication protocol with MQTT protocols. Therefore utilizing Message Queuing Telemetry Transport (MQTT) protocols for data transmission minimizes latency and handles network disruptions efficiently. MQTT is lightweight, making it suited for applications with low bandwidth and inconsistent access. It provides support for Quality of Service (QoS) levels, ensuring the transmission of messages even in the event of network outages. Also for securing with integrating data flow from edge devices to cloud servers is critical. This includes employing SSL/TLS encryption for data transport to prevent interception and unwanted access.
3) Edge Computing: For deploying system, edge devices with high processing capacity enables for local data processing, minimizing the requirement for constant data transmission to central server. Therefore the ARM Cortex-A Processors are energy-efficient and powerful enough for various computational activities required in agricultural applications. Also NVIDIA Jetson Modules are designed for AI and machine learning applications, providing the necessary GPU acceleration for real-time data processing, local training, and inference.
4) Cloud Infrastructure: Utilizing cloud-based servers from platforms like AWS, Google Cloud, or Microsoft Azure is critical for central data aggregation, model training, and storage. These platforms offer high availability and disaster recovery solutions, assuring data integrity and service continuity. Cloud services can effortlessly interface with edge devices, offering a cohesive environment for data flow, model changes, and remote management. Implementing robust security measures, such as multi-factor authentication (MFA), encryption at rest and in transit, and frequent security audits, is crucial for protecting sensitive agricultural data. For distributed storage options, such as Hadoop HDFS or cloud-based services, are recommended for handling huge datasets.
5) Power Supply and Environmental Considerations: Reliable power supply is vital for ongoing operation. Deploying solar panels with battery storage can supply sustained power in remote regions. Also ensuring that all gadgets, particularly those deployed in the field, are waterproof and can resist extreme external conditions.
6) Maintenance and Support: Remote Monitoring and Diagnostics: Implementing systems for remote monitoring and diagnostics enables preventive maintenance, decreasing downtime and boosting system reliability. Training local technicians or farmers in basic troubleshooting and maintenance can provide a timely reaction to issues and reduce dependence on distant support teams.

IX. CONCLUSION AND FUTURE WORK

Our study proposed a unique approach for diagnosing wheat leaf diseases by merging advanced transformer models (CoAtNets and Swin Transformer V2) within a federated learning framework. This combination utilized the capabilities of both architectures for robust feature extraction and classification, while preserving data privacy and security through decentralized training. Additionally, the introduction of weight pruning and linear attention processes enhanced the model for resource-constrained contexts, making it a substantial contribution to the field of agricultural disease detection. The major outcomes of our research revealed that the proposed approach outperformed classic convolutional neural networks (CNNs) and existing transformer models in terms of accuracy, precision, recall, and F1-score. The results underlined the need to combine advanced deep learning techniques with federated learning to boost model performance while ensuring data privacy.

Despite the optimistic results, our study had several limitations. The model's performance was heavily dependent on the quality and diversity of the training data, which varied across different geographical regions. Additionally, the federated learning architecture required a powerful communication infrastructure, which could be tough to construct in isolated or rural places. Furthermore, the process of weight pruning, while good for lowering model complexity, resulted in the loss of some critical features, affecting the model's overall accuracy. Future research should focus on overcoming the shortcomings found in this study. This includes strengthening the federated learning infrastructure to enable reliable communication in rural areas and boosting the diversity of the training dataset to cover a larger range of geographical differences. Additionally, exploring more complex weight pruning approaches that limit the loss of key features could further enhance the model. Future studies could potentially investigate the integration of additional data sources, such as environmental and climatic data, to increase the model's robustness and accuracy.

REFERENCES

[1] H. Khan, G. Krishnappa, S. Kumar, N. B. Devate, N. D. Rathan, S. Kumar, C. N. Mishra, S. Ram, R. Tiwari, O. Parkash, O. P. Ahlawat, H. M. Mamrutha, G. P. Singh, and G. Singh, "Genome-wide association study identifies novel loci and candidate genes for rust resistance in wheat (*Triticum aestivum* L.)," *BMC Plant Biol.*, vol. 24, no. 1, p. 411, May 2024.

[2] T. Ben Hassen and H. El Bilali, "Conflict in Ukraine and the unsettling ripples: Implications on food systems and development in North Africa," *Agricult. Food Secur.*, vol. 13, no. 1, p. 16, Apr. 2024.

[3] R. S. Khedwal, A. Chaudhary, V. K. Sindhu, D. B. Yadav, N. Kumar, R. S. Chhokar, and S. Dahiya, "Challenges and technological interventions in rice-wheat system for resilient food-ater-energy-environment Nexus in North-western Indo-Gangetic Plains: A review," *Cereal Res. Commun.*, vol. 51, no. 4, pp. 785–807, Mar. 2023.

[4] M. Nadimi, E. Hawley, J. Liu, K. Hildebrand, E. Sopiwnyk, and J. Paliwal, "Enhancing traceability of wheat quality through the supply chain," *Comprehensive Rev. Food Sci. Food Saf.*, vol. 22, no. 4, pp. 2495–2522, Apr. 2023.

[5] M. R. Simn, "Fungal wheat diseases: Etiology, breeding, and integrated management," *Frontiers Plant Sci.*, vol. 14, pp. 1–3, Jul. 2023.

[6] M. Long, M. Hartley, R. J. Morris, and J. K. M. Brown, "Classification of wheat diseases using deep learning networks with field and glasshouse images," *Plant Pathol.*, vol. 72, no. 3, pp. 536–547, Apr. 2023.

[7] A. Ahmad, D. Saraswat, and A. El Gamal, "A survey on using deep learning techniques for plant disease diagnosis and recommendations for development of appropriate tools," *Smart Agricult. Technol.*, vol. 3, Feb. 2023, Art. no. 100083.

[8] P. R. Navale and S. B. Basapur, "Deep learning based automated wheat disease diagnosis system," in *Proc. Int. Conf. Advancement Technol. (ICONAT)*, Goa, India, Jan. 2023, pp. 1–5.

[9] M. Ceyhan, Y. Kartal, K. Özkan, and E. Seke, "Classification of wheat varieties with image-based deep learning," *Multimedia Tools Appl.*, vol. 83, no. 4, pp. 9597–9619, Jun. 2023.

[10] C. Chronis, I. Varlamis, Y. Himeur, A. N. Sayed, T. M. Al-Hasan, A. Nhlabatsi, F. Bensaali, and G. Dimitrakopoulos, "A survey on the use of federated learning in privacy-preserving recommender systems," *IEEE Open J. Comput. Soc.*, vol. 5, pp. 227–247, 2024.

[11] A. Casu, M. C. Leggieri, P. Toscano, and P. Battilani, "Changing climate, shifting mycotoxins: A comprehensive review of climate change impact on mycotoxin contamination," *Comprehensive Rev. Food Sci. Food Saf.*, vol. 23, no. 2, Mar. 2024, Art. no. e13323.

[12] M. Aggarwal, V. Khullar, N. Goyal, and T. A. Prola, "Resource-efficient federated learning over IoAT for rice leaf disease classification," *Comput. Electron. Agricult.*, vol. 221, Jun. 2024, Art. no. 109001.

[13] Y. Qiu, L. Ma, and R. Priyadarshi, "Deep learning challenges and prospects in wireless sensor network deployment," *Arch. Comput. Methods Eng.*, vol. 31, pp. 1–24, Mar. 2024.

[14] M. Shoaib, B. Shah, S. Ei-Sappagh, A. Ali, A. Ullah, F. Alenezi, T. Gechev, T. Hussain, and F. Ali, "An advanced deep learning models-based plant disease detection: A review of recent research," *Frontiers Plant Sci.*, vol. 14, pp. 1–22, Mar. 2023.

[15] H. Qin, Y. Ding, M. Zhang, Q. Yan, A. Liu, Q. Dang, Z. Liu, and X. Liu, "BiBERT: Accurate fully binarized BERT," 2022, *arXiv:2203.06390*.

[16] H. Chen, C. Lv, L. Ding, H. Qin, X. Zhou, Y. Ding, X. Liu, M. Zhang, J. Guo, X. Liu, and D. Tao, "DB-LLM: Accurate dual-binarization for efficient LLMs," 2024, *arXiv:2402.11960*.

[17] H. Qin, X. Ma, X. Zheng, X. Li, Y. Zhang, S. Liu, J. Luo, X. Liu, and M. Magno, "Accurate LoRA-finetuning quantization of LLMs via information retention," 2024, *arXiv:2402.05445*.

[18] H. Qin, M. Zhang, Y. Ding, A. Li, Z. Cai, Z. Liu, F. Yu, and X. Liu, "BiBench: Benchmarking and analyzing network binarization," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2023, pp. 28351–28388.

[19] W. Huang, Y. Liu, H. Qin, Y. Li, S. Zhang, X. Liu, M. Magno, and X. Qi, "BiLLM: Pushing the limit of post-training quantization for LLMs," 2024, *arXiv:2402.04291*.

[20] D. Zhu, J. Tan, C. Wu, K. Yung, and A. W. H. Ip, "Crop disease identification by fusing multiscale convolution and vision transformer," *Sensors*, vol. 23, no. 13, p. 6015, Jun. 2023.

[21] Q. Zeng, L. Niu, S. Wang, and W. Ni, "SEViT: A large-scale and fine-grained plant disease classification model based on transformer and attention convolution," *Multimedia Syst.*, vol. 29, no. 3, pp. 1001–1010, Jun. 2023.

[22] D. Pang, H. Wang, J. Ma, and D. Liang, "DCTN: A dense parallel network combining CNN and transformer for identifying plant disease in field," *Soft Comput.*, vol. 27, no. 21, pp. 15549–15561, Sep. 2023.

[23] F. Arshad, M. Mateen, S. Hayat, M. Wardah, Z. Al-Huda, Y. H. Gu, and M. A. Al-Antari, "PLDPNet: End-to-end hybrid deep learning framework for potato leaf disease prediction," *Alexandria Eng. J.*, vol. 78, pp. 406–418, Sep. 2023.

[24] S. Ogrekçi, Y. Unal, and M. N. Dudak "A comparative study of vision transformers and convolutional neural networks: Sugarcane leaf diseases identification," *Eur. Food Res. Technol.*, vol. 249, Apr. 2023, Art. no. 18331843.

[25] X. Wang, Y. Wang, J. Zhao, and J. Niu, "ECA-ConvNeXt: A rice leaf disease identification model based on ConvNeXt," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Vancouver, BC, Canada, Jun. 2023, pp. 6234–6242.

[26] M. Aggarwal, V. Khullar, N. Goyal, R. Gautam, F. Alblehai, M. Elghatwary, and A. Singh, "Federated transfer learning for rice-leaf disease classification across multiclient cross-silo datasets," *Agronomy*, vol. 13, no. 10, p. 2483, Sep. 2023.

[27] A. Haridasan, J. Thomas, and E. D. Raj, "Deep learning system for paddy plant disease detection and classification," *Environ. Monitor. Assessment*, vol. 195, no. 1, pp. 1–28, Jan. 2023.

[28] Y. Liu, Y. Song, R. Ye, S. Zhu, Y. Huang, T. Chen, J. Zhou, J. Li, M. Li, and C. Lv, "High-precision tomato disease detection using NanoSegmenter based on transformer and lightweighting," *Plants*, vol. 12, no. 13, p. 2559, Jul. 2023.

[29] G. Idoje, T. Dagiuklas, and M. Iqbal, "Federated learning: Crop classification in a smart farm decentralised network," *Smart Agricult. Technol.*, vol. 5, Oct. 2023, Art. no. 100277.

[30] P. Alirezazadeh, M. Schirrmann, and F. Stolzenburg, "Improving deep learning-based plant disease classification with attention mechanism," *Gesunde Pflanzen*, vol. 75, no. 1, pp. 49–59, Feb. 2023.

[31] O. Getch. (2021). *Wheat Leaf Dataset, Kaggle Dataset*. [Online]. Available: https://www.kaggle.com/datasets/olyadgetch/wheat-leaf-dataset

[32] S. Arya and B. Singh, "Wheat nitrogen deficiency and leaf rust image dataset," ICAR-Indian Agricult. Res. Inst., New Delhi, India, 2020, doi: 10.17632/th422bg4yd.1.

[33] R. Song, Z. Zhang, and H. Liu, "Edge connection based Canny edge detection algorithm," *Pattern Recognit. Image Anal.*, vol. 27, no. 4, pp. 740–747, Dec. 2017.

[34] J. Flusser, S. Farokhi, C. Höschl, T. Suk, B. Zitová, and M. Pedone, "Recognition of images degraded by Gaussian blur," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 790–806, Feb. 2016.

[35] X. Ren and S. La, "Medical image enhancement based on Laplace transform, Sobel operator and histogram equalization," *Academic J. Comput. Inf. Sci.*, vol. 5, no. 6, pp. 48–54, 2022.

[36] A. Makandar and B. Halalli, "Image enhancement techniques using highpass and lowpass filters," *Int. J. Comput. Appl.*, vol. 109, no. 14, pp. 21–27, Jan. 2015.

[37] S. Gowrishankar, "ConvNeXt-based mango leaf disease detection: Differentiating pathogens and pests for improved accuracy," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 6, pp. 1–11, 2023.

[38] M. H. K. Mehedi, A. K. M. S. Hosain, S. Ahmed, S. T. Promita, R. K. Muna, M. Hasan, and M. T. Reza, "Plant leaf disease detection using transfer learning and explainable AI," in *Proc. IEEE 13th Annu. Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON)*, Vancouver, BC, Canada, Oct. 2022, pp. 166–170.

[39] M. Naveenkumar, S. Srithar, B. Rajesh Kumar, S. Alagumuthukrishnan, and P. Baskaran, "InceptionResNetV2 for plant leaf disease classification," in *Proc. 5th Int. Conf. I-SMAC (IoT Social, Mobile, Analytics Cloud) (I-SMAC)*, Palladam, India, Nov. 2021, pp. 1161–1167.

[40] L. R. Burra, J. Bonam, P. Tumuluru, and B. N. K. Rao, "Fine-tuning for transfer learning of ResNet152 for disease identification in tomato leaves," in *Intelligent Computing and Applications*. Springer, Nov. 2022, pp. 295–302.

[41] G. Yang, Y. He, Y. Yang, and B. Xu, "Fine-grained image classification for crop disease based on attention mechanism," *Frontiers Plant Sci.*, vol. 11, pp. 1–15, Dec. 2020.

[42] Z. Dai, H. Liu, Q. V. Le, and M. Tan, "CoAtNet: Marrying convolution and attention for all data sizes," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, Dec. 2022, pp. 3965–3977.

[43] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, and B. Guo, "Swin transformer v2: Scaling up capacity and resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 11999–12009.

[44] X. Geng, J. Gao, Y. Zhang, and D. Xu, "Complex hybrid weighted pruning method for accelerating convolutional neural networks," *Sci. Rep.*, vol. 14, no. 1, p. 5570, Mar. 2024.

[45] Y. Shang, Z. Yuan, B. Xie, B. Wu, and Y. Yan, "Post-training quantization on diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 1972–1981.

[46] W. Chen, K. Bhardwaj, and R. Marculescu, "FedMAX: Mitigating activation divergence for accurate and communication-efficient federated learning," in *Proc. Mach. Learn. Knowl. Discovery Databases, Eur. Conf.*, Ghent, Belgium, 2020, pp. 348–363.

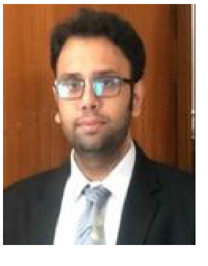
[47] R. Anil, V. Gupta, T. Koren, and Y. Singer, "Memory efficient adaptive optimization," in *Proc. 33rd Conf. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2019, pp. 1–10.

[48] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9992–10002.

MD. FAHIM-UL-ISLAM received the B.Sc. degree in computer science and engineering from Brac University, where he is currently pursuing the master's degree. His academic pursuits are complemented by his active engagement in research. He is also a Graduate Research Assistant with Brac University. His research interests include multiple fields, with a particular focus on deep learning, machine learning, artificial intelligence, natural language processing, computer vision, and software engineering.

AMITABHA CHAKRABARTY received the M.Sc. degree from the Department of Computer Science and Engineering, University of Rajshahi, in 2004, the M.Sc. degree in telecommunication engineering from Independent University, Bangladesh, and the Ph.D. degree from the Faculty of Engineering and Computing, Dublin City University, Dublin, Ireland, in 2012. He is currently a Professor with the Department of Computer Science and Engineering, Brac University, Dhaka, Bangladesh. He leads the IoT and Embedded System Research Group, Brac University. He is also involved in active research having a number of graduate and undergraduate research groups in different research projects. He has published research papers in various national and international conferences, journals, and book chapters. His research interests include the IoT, machine learning, deep learning, vision transformers, embedded systems, and switching theory. He is serving as a TCP member for various international journals and conferences. He is also serving as a Senior Judge in various national IT competition.

SARDER TANVIR AHMED received the degree in computer science and engineering from Brac University. He is currently an Engineer, deeply interested in exploring how computers can mimic human vision, focusing his efforts on understanding and advancing the field of machine learning with a touch of humility. He has contributed to projects that range from enhancing agricultural insights through disease detection in crops to aiding in the early detection of lung diseases using computer vision techniques. Through his work at Ether Technologies Ltd., and other experiences, he has demonstrated a commitment to applying his knowledge practically, aiming for incremental progress and tangible results. His journey reflects a blend of curiosity, dedication to learning, and a pragmatic approach to solving real-world problems with technology.



RAFEED RAHMAN received the B.Sc. degree in computer science and engineering from Brac University, in 2020, where he is currently pursuing the M.Sc. degree in computer science. He is also a Lecturer with the Department of Computer Science and Engineering. Brac University. His research interests include artificial intelligence, machine learning, natural language processing, and computer vision.

HYUN HAN KWON is currently a Full Professor with Sejong University. He was formerly a Professor of civil engineering with Chonbuk National University. He has twenty years of experience in stochastic hydrology, a data-driven approach to geophysical data, especially in climate-based applications. He has co-authored over 250 peer-reviewed journal articles. He was involved in integrates hydrology, stochastic systems, risk analysis, and climate variability to characterize key dynamics in water systems and to identify the means to foster sustainability in a changing world. He has been involved with projects directed at understanding and predicting at a regional scale for sustainable water resource management and development accounting for climate variability. His research interests include enhance society's capability to understand, anticipate and manage the impacts of climate fluctuations to improve human welfare, and the environment using advanced statistical approaches.

MD. JALIL PIRAN (Senior Member, IEEE) received the Ph.D. degree in electronics and information engineering from Kyung Hee University, South Korea, in 2016. He was a Postdoctoral Fellow with Kyung Hee University, until 2017. He holds a distinguished academic background and is currently an Associate Professor with the Department of Computer Science and Engineering, Sejong University, Seoul, South Korea. He has made significant contributions to the field of artificial intelligence and data science through his extensive research publications in esteemed international journals and conferences. His areas of expertise encompass machine learning, data science, big data, the Internet of Things (IoT), and cyber security. In addition to his research endeavors, he actively engages with scholarly journals as an Editor, including IEEE Transactions on Intelligent Transportation Systems, textitEngineering Applications of Artificial Intelligence (Elsevier), textitPhysical Communication (Elsevier), and textitComputer Communication (Elsevier). He represents South Korea as an Active Delegate to the Moving Picture Experts Group (MPEG). His outstanding research contributions have been recognized internationally, as evidenced by the prestigious "Scientist Medal of the Year 2017" awarded by IAAM in Stockholm, Sweden. Moreover, he received accolades from Iranian Ministry of Science, Technology, and Research, as an "Outstanding Emerging Researcher" in 2017. His exceptional Ph.D. dissertation was honored as the "Dissertation of the Year 2016" by Iranian Academic Centre for Education, Culture, and Research in the Engineering Group. He also serves as the Secretary for the IEEE Consumer Technology Society on Machine Learning, Deep Learning, and AI. Furthermore, he assumes the role of the Track Chair of Machine Learning, Deep Learning, and AI in the CE (MDA) Track for the upcoming 2024 IEEE International Conference on Consumer Electronics (ICCE). In 2022, he chaired the "5G and Beyond Communications" Session at the prestigious IEEE International Conference on Communications (ICC). His expertise as a reviewer extends to leading journals and he actively participates in various conferences.