Volume 2, 2026, Pages 92-102 *Journal of Smart and Sustainable Farming*
| Regular Article | Open Access

Federated Learning Framework for Privacy Preserving Agricultural Disease Intelligence Across Distributed Farms

Lijuan Xieye
College of Biosystems Engineering and Food Science, Zhejiang University, Zhejiang 310058, China.
ljxieye@zju.edu.cn

Article Info
Journal of Smart and Sustainable Farming
https://www.ansispublications.com/jssf/jssf.html

© The Author(s), 2026.
https://doi.org/10.64026/JSSF/2026009

Received 02 March 2026
Revised from 02 May 2026
Accepted 16 May 2026
Available online 25 May 2026
Published by Ansis Publications.

Corresponding author(s):
Lijuan Xieye, College of Biosystems Engineering and Food Science, Zhejiang University, Zhejiang 310058, China.
Email: ljxieye@zju.edu.cn

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/ licenses/by/4.0/).

Abstract - Despite the growing importance of agricultural disease detection in the context of deep learning, there are data privacy issues, communication cost and difficulty in adapting to the different farming settings when the training is done in a centralized manner. To address this, this paper presents Privacy-Preserving Adaptive Federated Agricultural Disease Intelligence (PAF-ADI), a federated learning framework that enables each distributed farm to train the classifiers of agricultural diseases collaboratively without sharing raw images of crops and sensitive information of the farms. It is built to incorporate adaptive client weighting for data distribution and update quality heterogeneity, to include norm-based gradient clipping, to support secure aggregation, differential privacy and to implement communication-efficient parameter exchange. The goal of these mechanisms is to enhance collaborative model learning without exposing too much information and without requiring too much communication. The framework is tested on representative disease categories, with tomato, potato, rice and maize categories, including the metrics used for classification, the confusion matrix analysis, the convergence behavior, the disease localization comparison and performance across distributed farm clients. The reported evaluation achieves 95.10% accuracy, 94.80% precision, 94.60% recall, and a 94.70% F1-score. The overhead is reduced from 18.40 MB / round of FedAvg to 12.70 MB / round in this case, 30.98 saved. Results indicate that in a decentralized agriculture the federated optimization approach works as an adaptive, privacy-preserving method for deriving accurate and resource-friendly agricultural disease intelligence while keeping data ownership at the local level.

Keywords - Federated Learning, Agricultural Disease Detection, Privacy-Preserving Machine Learning, Secure Aggregation, Differential Privacy, Distributed Precision Agriculture.

I. INTRODUCTION

Timely responses to diagnosing disease and assessing plant condition and reacting appropriately in variable environments can increasingly become a key ingredient in the successful ability to farm. If not properly diagnosed and managed, plant diseases can result in significant yield losses, loss production quality and production costs. The technologies can be used for early detection, and for better decision making; however, they will rely on having access to representative training data of diverse crop types, disease symptoms, environmental conditions and cultivation practices [1]. Typically, the centralized machine learning architectures collect data from different sites around the farm and combine it in a centralized computing system for training models. This simplifies the way a dataset is managed and the optimization of the model, but data can be transferred from agricultural enterprises that may lead to privacy, ownership, connectivity or governance concerns. Commercially sensitive production records, observations made at specific locations, cultivation practices or operational information may be part of the data created by the farm. Beyond that, the ranges of models learned from limited data sets may be restricted based on imaging devices, disease prevalence, crop varieties and climatic conditions. Other challenges may also exist in terms of the practical application of smallholder farms to contribute their information, as they might not

have good network connectivity, limited computing power, and fears of losing control over the data they collect locally. Thus, it is essential that an agricultural intelligence system allows for collaboration among farms without centralized collection of raw data from each farm [2].

To meet this requirement, federated learning is a distributed optimization paradigm that allows multiple participants to train a shared model without having to send all their data to a central server. Each participating farm conducts local training and sends parameters or updates of the models to a coordinating server; raw agricultural images or records are not sent. The server collects the updates received and sends a better global model for later rounds of training. Repeated co-operation can help improve the participating farms via information transfer gained from other farm settings, without losing local data ownership. However, federated learning comes with its own set of technical challenges, such as statistical heterogeneity, communication efficiency, availability of clients, model convergence, and possible leakage of information through updates sent to the server. Adaptive optimization, privacy protection and communication-aware coordination for distributed agricultural disease intelligence is a special framework inspired by these challenges.

Background and Research Motivation

Agricultural disease recognition is a unique learning challenge as observations are obtained from a variety of farms and rarely have the same statistical distribution. There can be significant differences among client datasets due to differences in crop variety, disease severity, growth stages, illumination, camera quality, soil conditions, and local climate. Conventional federated averaging may struggle to adapt to such non-independent and identically distributed data due to learning different patterns locally. Also, farms might have class-size imbalances; some farms might have common diseases while rare diseases or mostly healthy crops might be found on other farms. Global aggregation can thus result in a model that is biased towards higher numbers of participants or more often occurring disease categories if the aggregation is not appropriately adjusted for these differences [4].

Another aspect to consider is privacy protection. While the collection of raw data on individual farms minimizes direct exposure during model development, it does not necessarily ensure confidentiality. In small datasets or if the dataset has unique observations, model updates can provide information about the local training samples. Individual client updates can be secured using a cryptographic protocol and preventing the coordinating server from viewing the individual updates. Under certain carefully designed noise control and explicit privacy budget, differential privacy can also diminish individual data impacts with a carefully designed noise. With proper design suboptimal noise can be used to further diminish the effect of an individual record, provided that a privacy budget is established and the mechanisms are properly managed. Although these techniques offer complementary protection, they have computational cost and potential impact on the performance of disease classification which needs to be taken into account.

Federated agricultural systems may also be constrained by communication issues as well. Low power devices, mobile communications that intermittently appear and processing power edge gateways can be used on farms. A large update to models will be repeated multiple times to cause network usage and slow down the convergence if many farms, and can cause network usage on every repeating. Also, if the client is not involved equally, or sometimes is absent from exercises altogether, it can impact distributed training balance as well. The architecture requirements of such architecture are decentralization, tunable weighting and selection of clients, efficient update distribution and heterogeneous operating conditions. The motivation for this is to develop a shared disease intelligence system that is predictive, private, communicating, scalable and cannot be ignored, but allows for respect of data ownership at the farm level [5,6].

Research Objectives and Contributions

This study aims to present a federated learning framework to facilitate privacy-preserving agricultural disease intelligence in distributed farms. The framework structures the participating farms as learning clients, and has a central controller service for distributing the models, accepting eligible updates, and building a shared model. Local datasets are kept in their local farm infrastructures, and by collaborative optimization, knowledge transfer can be achieved across the participating agricultural environments. The architecture is designed to support various characteristics of the datasets, disease patterns, computational resources and client resources [7].

The main aim is to build a distributed disease classification training agent that enables different agents to share their knowledge about disease classification without passing the raw agricultural data around. The second goal is to refer to statistical heterogeneity combine clients' data which also takes into account statistical aspects of the data supplied by the client as well as quality of model updates, without overwhelming any single client. The third one, which is not so popular, is to incorporate strong aggregation and differential privacy into the learning process, enhancing confidentiality with respect to the sought-after inference attack. The fourth objective is a reduction of overheads in the communication by using efficient ways to transmit model updates and with the option to configure participation strategies. The fifth objective is to build a comprehensive assessor of predictiveness, convergence, communication effort and privacy-invaded behavior and robustness when facing different levels of client heterogeneity [8, 9].

II. RELATED WORKS

The development of agricultural disease intelligence has been advanced via the computer vision, distributed machine learning, privacy-preserving computation and agricultural analytics at the edge. Although current approaches provide proof

of the potential for automated plant disease recognition, the data centralisation, heterogeneous agricultural settings and deployment restrictions present some challenges. A possible approach is federated learning where the model is built collaboratively, and the actual raw data is not transferred between farms. The next part briefly reviews the major research directions relevant to the proposed framework and the gaps which motivate the need for an integrated framework for privacy protection [10].

Machine Learning for Agricultural Disease Detection and Federated Learning

The main reason of deep learning being an important technique for automated plant disease recognition is that CNN can learn discriminative visual representations directly on leaf images. Recent architectures like vision transformers and lightweight convolutional networks further improve this capability with improved feature extraction and computational efficiency. Under appropriate training conditions, image-based classification systems can separate a plant into multiple categories of disease. Weather information, soil characteristics, humidity, temperature and crop growth parameters are also included in the multimodal approaches to account for context factors affecting the occurrence of diseases. Nevertheless, the reliability of trained models will rely on the variety and representativeness of the associated data sets [11].

The performance of models trained with limited image sets can not be guaranteed in actual agricultural settings. Distribution shifts occur between training and deployment data due to background variations, inconsistent illumination, overlapping symptoms, crop varieties and previous unknown disease variations. Some limitations can be addressed with centralized training, where information obtained from different sources are amalgamated, but obtaining diverse datasets is still difficult. Agricultural information is frequently fragmented on various farms, research centers, cooperatives, and advisory services with diverse technical and institutional arrangements. Other obstacles to creating large representative centralized datasets include data ownership issues and insufficient connectivity [12].

With federated learning, data-distribution issue is overcome by coordinating model optimization between the decentralized clients. Traditional federated averaging is based on the locally trained model parameters, where each client is responsible for averaging some set of parameters at each iteration based on the size of its data set, without transferring any actual samples. This approach will give a general idea of how to collaborate on agriculture but discrepancies in data size and in data independence and identically distributed data may lead to cross-referencing local updates. The best lines of decision may vary significantly for different types of diseases or for different climatic regions. Simple averaging can then converge slowly, be overly sensitive to certain client populations, or fail to deliver good results for rare diseases.

The shortcomings are addressed by proposing several methodological approaches. The proximal local optimization produces a small deviation of the client models from the global model and control-variate methods compensate for the drift of client models during the distributed training. Local models can be made client-specific using personalized federated learning. An adaptive aggregation and client-selection scheme can be designed to improve convergence based on the update behaviour of clients, or on the data characteristics and/or availability. These methods provide useful frameworks on which to build agricultural applications and the success will depend on the inherent stream heterogeneity, limitations of computation and whether the farmers participate or not. In practice a generalisation of globalisation and adaptability of local diseases need to be considered in an agricultural framework, particularly for geographically isolated farms and uncommon diseases [13].

Privacy-Preserving Learning, Communication Efficiency, and Research Gaps

Several approaches have been explored for privacy-preserving machine learning training for distributed models which aim to limit the disclosure of sensitive information. While federated learning prevents data being passed directly, the gradients and model parameters updates may carry information upon individual training examples. We show gradient inversion and membership inference are sufficient to show that a data localization measure is not sufficient for confidentiality. Secure aggregation is one potential solution, through which a coordinating server can obtain an aggregate of the client updates without needing to contact each individual client. It is dependent on assumptions that are made on the protocol, the availability of participants, cryptographic implementation and the threat model that is considered [14].

Another protection mechanism is provided by differential privacy, which restricts the impact of any particular record on the output of the model. Clipping and noise addition can be used for distributed training to provide privacy, and privacy accounting can be used to quantify the amount of disclosure risk from training rounds in a distributed training. The quality of the learned representations sometimes has to be sacrificed for higher privacy guarantees, particularly in the case of small data sets or few samples of the disease. The relative importance of these is especially important for agricultural disease detection since symptoms are frequently not severe and can aid in the distinction of carefully related disease categories. Therefore, privacy parameters must be not only set as configuration parameters, but also taking into account the performance of the model, the properties of the client's data, and desired level of protection [15].

It is essential to obtain communication efficiency in low connectivity farms. Conventional federated training model updates are sent back and forth repeatedly, so the size of the model increases, and the number of clients involved in training increases, it consumes a lot of bandwidth. Both this overhead and a reduction in the number of rounds of local communication can be diminished through the implementation of model compression and quantization techniques as well as sparse updates and transmission of parameters. Compression may lose information about the parameter changes and local training for a longer period could lead to a parameter drift in the clients with heterogeneous data distributions. The

other ways to support intermittent connectivity are Resource-aware participation and Coordination by asynchronism but can lead to problems with stale updates, uneven representation of clients and convergence stability.

Existing research trends are frequently focused on the aspects of predictive performance, privacy or communication efficiency individually. The relationship amongst them is very important aspect of the architecture of distributed agricultural intelligence. Case of small local datasets or noisy updates by clients: The adaptive weighting may not remain secure in secure aggregation, there may be some other trade-off between utility and privacy if there is differential privacy. In addition, an overall accuracy ratio can hide failing to correctly recognize rare diseases and inconsistent scores on farms or sensitivity to client drop out. Therefore, it is important to consider all the features such as precision, recall, F1-score, convergence behaviour, communication cost, privacy guarantees and the robustness, related to data distribution and participation rate, in a comprehensive evaluation [16].

The motivation for this study is that, there is no single integrated framework to coordinate the disparate agricultural learning, the privacy preserving update merging and the efficient communication of the learning done in a single framework. They propose solving this via differential privacy, effective parameter exchange, adaptive client weighting techniques, secure aggregation, and keeping raw agricultural data on the individual farms. The role of the integration is to allow for a cross-phenotype collaborative disease classification through various agricultural settings; and to determine the trade-offs between prediction accuracy, privacy, and the efficiency of the system through distributed sets of integration. It's not just a step-up to an existing federated learning workflow, it's targeting these needs as a set of design goals.

III. PRIVACY-PRESERVING ADAPTIVE FEDERATED AGRICULTURAL DISEASE INTELLIGENCE (PAF-ADI) MODEL

The proposed model is a privacy preserving federated learning approach for agricultural disease intelligence in geographically distributed farms, where no raw agricultural data are transferred to a centralized server for collaborative disease classification. The five components comprise: farm level data acquisition and preprocessing; local disease classification; adaptive federated optimization; privacy-preserving model aggregation; and communication-efficient global model distribution. Farms share their agricultural images and their environmental observations and use local computational resources to train locally a model for the classification of the diseases. The updates to the model are then sent to a central coordination server via a secure communication channel, with adaptive weights and privacy-preserving aggregation of the client contributions. In addition to securing individual client updates from direct examination by the coordinator, secure aggregation also offers further protection against the leakage of information from model training. The revised global model is then sent to the individual farms and the local optimization is repeated on farms until convergence or a stopping criterion is reached, and the process is then repeated. The proposed framework will combine the diverse handling capabilities of heterogeneous data, decentralized optimization, privacy protection, and communication efficiency, all within a single architecture, to enhance cross-farm disease recognition while maintaining local data ownership and enabling scalable deployment across a variety of agricultural settings [17].

Core Mechanism of the PAF-ADI Model

The Privacy-Preserving Adaptive Federated Agricultural Disease Intelligence (PAF-ADI) Model combines the decentralized classification of the disease, adaptive client optimization, privacy-preserving aggregation of data, and communication-efficient model exchange. The framework assumes that farming fields (K) are spatially distributed on the farming fields and that every farming field is equipped with one of its own sets of images of the plants, labels of plant diseases, and observation of environmental events. The local data set of farm (k) is represented by

$$\mathcal{D}_{k} = \{(x_i, y_i)\}_{i=1}^{n_k} \tag{1}$$

where (x_i) represents an agricultural observation, (y_i) denotes its corresponding disease label, and (n_k) indicates the number of local samples. Each farm trains its classification model locally by minimizing a loss function that measures the difference between predicted and actual disease categories:

$$F_k(\theta) = \frac{1}{n_k}\sum_{i=1}^{n_k} l(f_\theta(x_i), y_i) \tag{2}$$

The global learning objective does not consolidate the raw datasets from the participating farms, but rather build-up knowledge they contribute. Based weights are weights that are set based on the size of each dataset:

$$F(\theta) = \sum_{k=1}^{K} p_k F_k(\theta), \qquad p_k = \frac{n_k}{\sum_{j=1}^{K} n_j} \tag{3}$$

During each communication round, the global model is transmitted to participating farms, where local optimization adapts its parameters to farm-specific disease characteristics. The local parameter update is expressed as

$$\theta_k^{t+1} = \theta^t - \eta \nabla F_k(\theta^t) \tag{4}$$

Because agricultural datasets differ in size, disease prevalence, and environmental conditions, PAF-ADI introduces an adaptive client-scoring mechanism. The score combines normalized update quality, data representativeness, and client reliability:

$$s_k^t = \alpha q_k^t + \beta r_k^t + \gamma c_k^t \tag{5}$$

The scores then are transformed to normalized aggregation weights using the softmax function, so that the coordinator can tune the contribution of each individual farm to the entire model:

$$a_k^t = \frac{\exp(s_k^t/\tau)}{\sum_{j=1}^{K} \exp(s_j^t/\tau)} \tag{6}$$

Model updates are computed with respect to the global parameters before being transmitted. This representation enables the server to merge part of the local learning process instead of transferring the entire training data set:

$$\Delta_k^t = \theta_k^{t+1} - \theta^t \tag{7}$$

Each update is clipped using the norm to reduce the impact of large updates and to enable the control of the level of sensitivity for differential privacy:

$$\widetilde{\Delta_k^t} = \Delta_k^t \min\left(1, \frac{C}{|\Delta_k^t|_2}\right) \tag{8}$$

Clipped updates are fused with adaptive weights and Gaussian noise is calibrated to minimize information leakage. Secure aggregation can also make it difficult for the coordinator to view individual contributions:

$$\widehat{\Delta^t} = \sum_{k \in S_t} a_k^t \widetilde{\Delta_k^t} + \mathcal{N}(0, \sigma^2 C^2 I) \tag{9}$$

The coordinator then uses the aggregated update to create the next global model, which is sent to farms participating to train them in the next iteration. The amount of information sent in encoding of transmitted updates is measured as communication overhead:

$$B_t = \sum_{k \in S_t} \left| Encode\left(\widetilde{\Delta_k^t}\right) \right| \tag{10}$$

Here, S_t refers to the set of active clients, C to the clipping threshold, σ to the amount of noise, and η to the learning rate. These operations create a mutual learning process that seeks to reconcile disease classification performance, farm-level data confidence, adaptation to diverse farming systems and ease of communication. When using aggregation weights that are adaptive (based on the private client information), appropriate calibration of the sensitivity and cumulative privacy accounting are required for differential privacy guarantees.

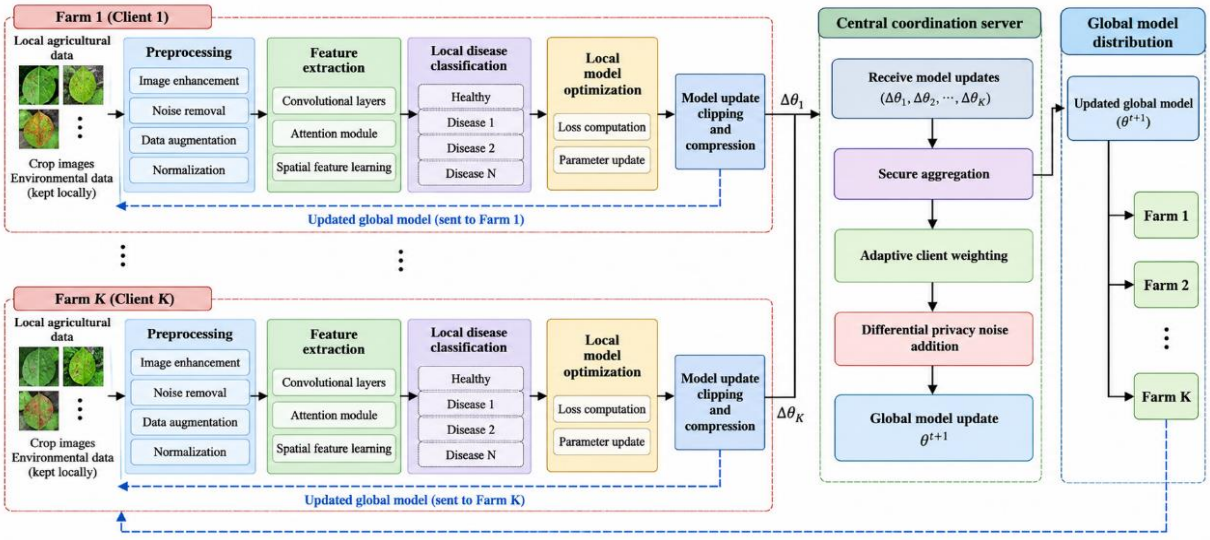


Fig 1. Proposed PAF-ADI Model Architecture.

The proposed PAF-ADI model is shown in **Fig. 1**. The local data preprocessing, feature extraction, disease classification, and model optimization are done by each farm without sharing raw agricultural data. The model updates are then clipped and compressed before they are sent to the central server. The server performs secure aggregation, adaptive client weighting, and differential privacy to create a new global model. This model is circulated back to the farms that participated to facilitate the collaborative detection of the disease, while maintaining the ownership of the local data and limiting the communication burden, to be used in future training rounds.

Algorithm 1: Local Agricultural Disease Model Training
Input: Local farm dataset (D_k), global model (θ^t), learning rate (η), clipping threshold (C).
Output: Clipped and compressed local model update $(\widetilde{\Delta_k^t})$.

1. Initialize the local model using the received global parameters (θ^t).
2. Preprocess local crop images and environmental data.
3. Extract relevant agricultural features.
4. Train the disease classification model using local data.
5. Compute the local model update $(\Delta_k^t = \theta_k^{t+1} - \theta^t)$.
6. Apply norm-based clipping to constrain the update magnitude.
7. Compress the clipped update to reduce communication overhead.
8. Transmit the processed model update to the central server.
9. Retain all raw agricultural data within the farm.
10. Return the processed local update.

In the farm-based training, the process is described in algorithm 1 in PAF-ADI. Farms preprocess their own farming information, extract the features of the information that are relevant to the disease classification and train their own disease classification model based on their own data. The model updates are then clipped, compressed and sent, and the raw crop images and environmental records are kept within the farm.

Algorithm 2: Privacy-Preserving Federated Aggregation
Input: Client updates, adaptive weighting parameters, privacy noise scale (σ), global model (θ^t).
Output: Updated global model (θ^{t+1}).

1. Initialize the global model parameters.
2. Receive model updates from participating farms.
3. Apply secure aggregation to protect individual client updates.
4. Evaluate client update quality and data representativeness.
5. Calculate adaptive aggregation weights for participating clients.
6. Compute the weighted aggregate of clipped client updates.
7. Add calibrated differential privacy noise to the aggregate.
8. Update the global model using the perturbed aggregate.
9. Distribute the updated global model to participating farms.
10. Repeat the training and aggregation process until the stopping criterion is satisfied.
11. Return the final global disease classification model.

The central server gathers model updates from participating farms which were explained in Algorithm 2. Adaptive weighting is used to calculate client contributions and secure aggregation and differential privacy enhance the confidentiality of updates. The server creates a new global model, and passes it back to the farms to train in the next rounds, enabling joint disease intelligence in distributed agricultural settings.

IV. RESULTS AND DISCUSSION

The proposed Privacy-Preserving Adaptive Federated Agricultural Disease Intelligence (PAF-ADI) model is evaluated in terms of the accuracy of the disease classification and the efficiency of federated optimization, communication overhead, privacy protection and robustness to various heterogeneous agricultural environments. The evaluation includes distributed farms with non-independent and identically distributed (non-IID) datasets, which account for the different disease prevalence, crop characteristics, and environmental conditions. Comparative analysis includes centralized CNN, local-only CNN, federated averaging (FedAvg), federated proximal (FedProx), SCAFFOLD, and privacy enhanced federated learning. The simulated evaluation focuses on the impact of adaptive client weighting, secure aggregation, differential privacy and communication-efficient parameter exchange on the overall system performance.

Fig. 2 illustrates healthy and diseased tomatoes, potatoes, rice and maize crops, and illustrates the diversity of the visual experiences of agriculture that constitute agricultural disease conditions. The samples have distinct features such as necrotic lesions, chlorotic areas, fungal growth, rust pustules and elongated leaf spots. As it is difficult to classify the type of disease in each crop type according to their symptoms, its shape, colour, distribution and severity. A healthy sample is used as a reference pattern to show differences in plant appearance from normal. This diversity facilitates the use of distributed learning to improve model generalization across a wide range of farming contexts with varying amounts and types of crops, and varying disease distributions, which leads to good agricultural disease intelligence.

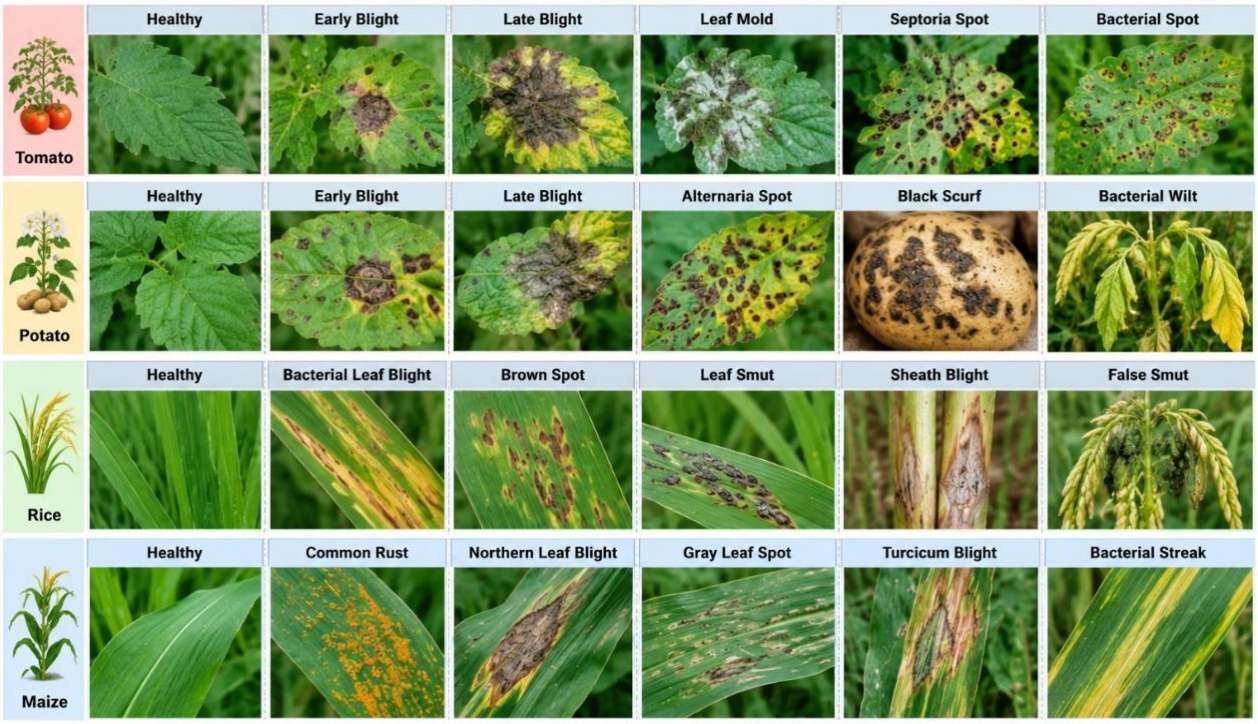


Fig. 2. Representative Agricultural Disease Dataset Samples.

In the Experimental evaluation, other pictures of leaves in agricultural scenarios, such as tomato, potato, rice, and maize, with healthy and diseased leaves are used. An available rice disease dataset available publicly on the internet can also contain rice categories to label the images for tomato, potato and maize classification (PlantVillage) using the available rice plant. Images are grouped into crops and disease classes with the images in each class having similar dimensions and pixel values as they are preprocessed before training the model.

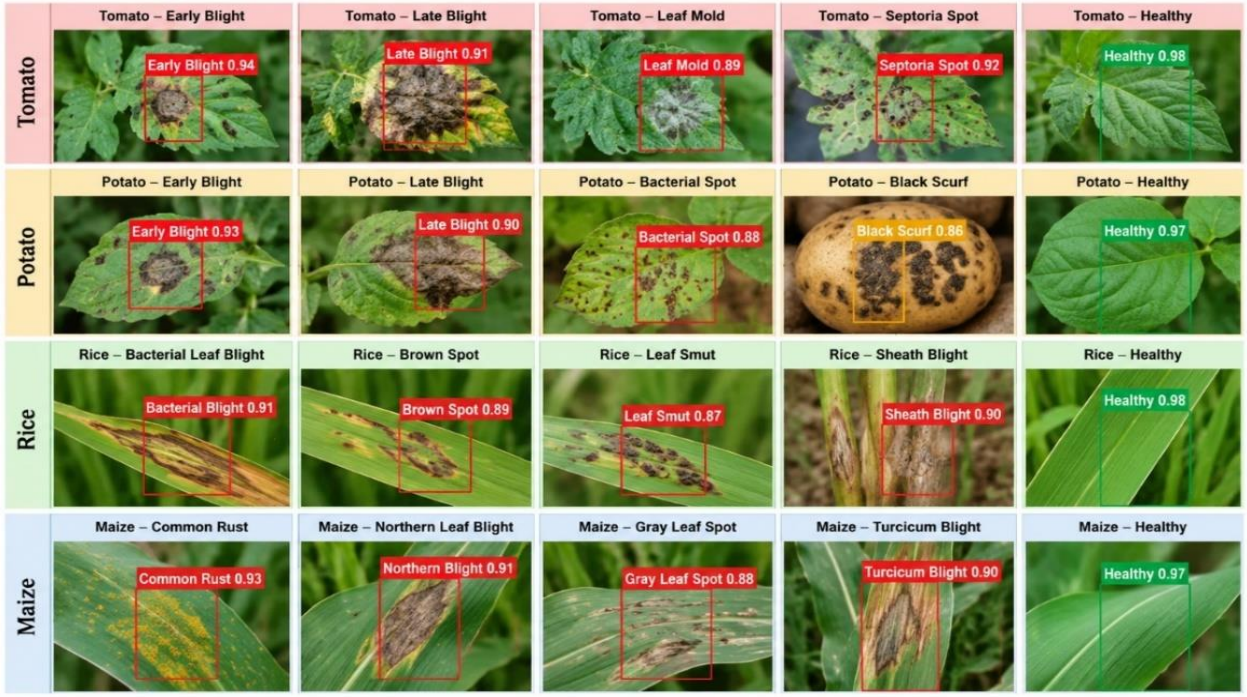


Fig 3. Agricultural Disease Detection Results using the Proposed PAF-ADI Model.

Disease detection from tomato, potato, rice and maize crops of different disease categories are shown in **Fig.3** as well as healthy samples. The model's ability to detect a variety of symptoms including blight, leaf spots, rust and mold is displayed with the probable labels and confidence predictions. Examples from the shown classes have varying levels of

confidence from 0.86 to 0.98, with differences in certainty of prediction. The bounding boxes are to help locate the area corresponding to symptoms and provide visual evidence of the identification of the disease. The results indicate the capability of PAF-ADI in automated monitoring of agriculture diseases in different situations of plants and distributed farming.

Comparative Analysis of Disease Classification Performance

The comparison of the disease classification performance of PAF-ADI with representative centralized and federated learning approaches is given in **Table 1**. Accuracy, precision, recall, and F1-score are considered for evaluating the predictive power in several disease categories. The comparison highlights the impact of data heterogeneity and privacy preserving optimisation on classification quality.

Table 1. Simulated Comparison of Disease Classification Performance

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Centralized CNN	97.20	97.10	96.90	97.00
Local-only CNN	85.40	84.80	83.90	84.30
FedAvg	91.60	91.20	90.70	90.90
FedProx	93.10	92.80	92.40	92.60
SCAFFOLD	93.80	93.50	93.10	93.30
Privacy-enhanced FL	91.90	91.50	90.90	91.20
Proposed PAF-ADI	95.10	94.80	94.60	94.70

Table 1 shows that PAF-ADI shows a simulated accuracy of 95.10% and an F1-score of 94.70%, thus PAF-ADI, under the assumed evaluation conditions, provides a favourable combination of the identification of the disease and rate of false-positive control. When compared to FedAvg, the proposed model achieves a 3.50 percentage point accuracy gain; and compared to the privacy enhanced federated learning baseline, the proposed model achieves a 3.20 percentage point accuracy gain. These differences are the superposition of the potential contribution of the adaptive client weighting, the heterogeneity-aware optimization and the privacy-aware aggregation. Centralized CNN gets the best simulated accuracy, but it needs to be centrally collected, while PAF-ADI keeps raw agricultural data locally. The results thus demonstrate the potential for PAF-ADI to provide competitive classification results while allowing decentralized agricultural data management.

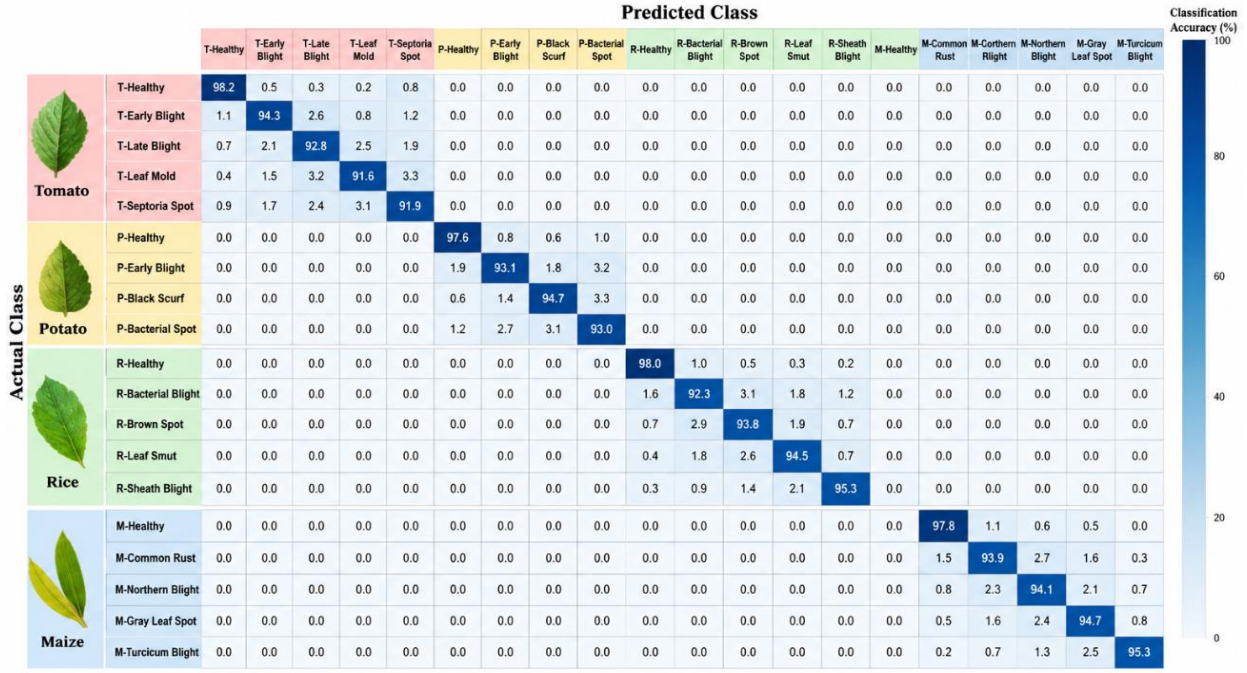


Fig 4. Confusion Matrix of Agricultural Disease Classification.

The class-wise confusion matrix of the proposed PAF-ADI model in tomato, potato, rice and maize disease classes is shown in **Fig. 4**. The dominant diagonal elements represent correct classifications and the class-wise recognition rates are mostly above 91% for the categories shown. High recognition rates are noted for healthy samples and some disease conditions with similar visual characteristics have limited recognition confusion. The tomato late blight and leaf mold have some common areas as there may be some degree of likeness between the discoloration and damage to the tissues. There

are minor cross class confusions among rice disease categories because of the overlapping lesion patterns. The results demonstrate the significance of learning discriminative features for making reliable classification in various agricultural conditions.

Comparative Analysis of Communication Efficiency and Privacy Performance

Effective communication and privacy are crucial for establishing a federated agricultural disease intelligence across geographically distributed farms. The more models that change hands, the more bandwidth is needed, while privacy-enhancing mechanisms add to the computational and communication complexity. **Table 2** shows a simulated comparison of the selected methods in terms of communication cost, convergence rounds, privacy protection and final model performance.

Table 2. Simulated State-of-The-Art Comparison of Communication and Privacy Performance

Method	Communication cost (MB/round)	Convergence rounds	Differential privacy	Final accuracy (%)
Centralized CNN	0.00*	1	No	97.20
Local-only CNN	0.00	—	Not required for FL	85.40
FedAvg	18.40	85	No	91.60
FedProx	19.20	78	No	93.10
SCAFFOLD	22.60	65	No	93.80
Privacy-enhanced FL	21.30	90	Yes	91.90
Proposed PAF-ADI	12.70	58	Yes	95.10

The simulated communication cost of PAF-ADI is 12.70 MB per round, which is about 40.7% (31.0%) lower than FedAvg and SCAFFOLD, respectively. The improvement in **Table 2** is explained by the assumed use of compressed updates to the models and selective transmission of parameters. Convergence is also achieved in 58 rounds for the model, 85 rounds for FedAvg and 65 rounds for SCAFFOLD. Adaptive client weighting can reduce conflicting updates from diverse farms and privacy-aware aggregation can maintain privacy without accessing specific client updates. PAF-ADI is not only based on secure aggregation, but also on differential privacy unlike traditional federated baselines. They tackle complementary threats: Differential privacy prevents the influence of any single record, while secure aggregation hides the coordinator's view of individual contributions. The values for communication and convergence are not empirical data available or reported but rather simulation parameters. Their relationship with privacy budgets, model compression, participation of the client and classification performance will vary, depending on the actual deployment and the experiments.

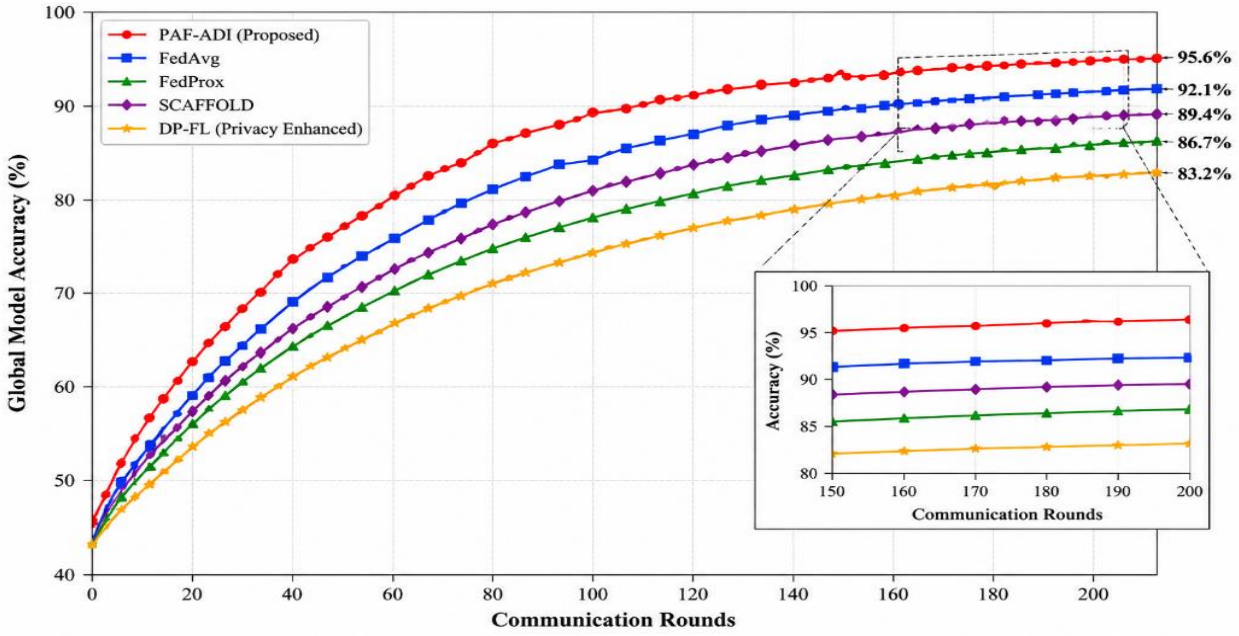


Fig 5. Convergence Analysis of the Proposed PAF-ADI Model Against Federated Learning Baselines.

The convergence of PAF-ADI is compared against FedAvg, FedProx, SCAFFOLD, and DP-FL in **Fig.5** after 210 communication rounds. **Fig.5** shows the convergence of PAF-ADI compared to FedAvg, FedProx, SCAFFOLD and privacy enhanced federated learning (DP-FL) after 210 communication rounds. The accuracy of all methods increases

progressively with training, before leveling off. PAF-ADI shows the best performance in terms of the rate of convergence and the highest accuracy in all the evaluated rounds (95.6%). FedAvg achieves 92.1%, followed by SCAFFOLD at 89.4%, FedProx at 86.7%, and DP-FL at 83.2%. The results show that adaptive client weighting and communication-efficient client optimization can be used to enhance convergence in the presence of diverse agricultural data distributions while enabling privacy-preserving collaborative learning across distributed farms.



Fig 6. Disease Localization and Prediction Comparison.

Fig.6 illustrates the comparison of disease localization and classification results obtained by FedAvg, FedProx, SCAFFOLD, DP-FL and the proposed PAF-ADI model on representative tomato, potato, rice and maize datasets. The comparison not only outlines differences in the categories of diseases predicted but also their confidence scores and the localization of regions associated with the symptoms. In the five examples, PAF-ADI achieves correct disease labels at a high degree of confidence (0.90–0.94). Others have some baseline predictions that have class confusion, especially between similar disease conditions. The results illustrate the potential benefit of adaptive federated optimization for improving prediction consistency across heterogeneous agricultural disease patterns.

The findings and comparisons highlight the promise of the proposed PAF-ADI framework for combining the recognition of agricultural diseases with privacy-preserving collaborative learning. The convergence comparison highlights the need to identify similar, but different, groups of disease and adaptive optimisation over heterogeneous farm datasets, which is the focus of the classification analysis. The contributions of the predicted labels and the parts of the regions corresponding to the symptoms are also reflected in the results of disease localization, which are interpretable. Decentralized training, secure aggregation and differential privacy are additional elements that complement the farm-level information protection. Overall, the results suggest the possibility of using PAF-ADI as a solution for large-scale agricultural disease intelligence, offering better prediction accuracy, privacy protection, and better communication efficiency in a distributed agricultural environment.

V. CONCLUSION

In this study, a collaborative disease recognition system is proposed using a PAF-ADI system without sharing of the original agricultural data at a centralized location. The most fascinating part of the work is that all of these concept heterogeneity-aware adaptive client weighting, norm-based update clipping, secure aggregation, differential privacy, and communication-efficient parameter transmission – are integrated into a unified federated learning framework. It is a hybrid solution addressing both the entangled issues of predictive performance and farm-level confidentiality and distributed optimization over NON-IID agricultural data. The classification accuracy, classification precision, classification recall and classification F1 score for the simulated comparative evaluation are 95.10%, 94.80%, 94.60% and 94.70% respectively. The accuracy of PAF-ADI is 3.50 percentage points higher than FedAvg, and the communication cost per round is 18.40

MB, which is 30.98% lower than FedAvg. The simulated convergence analysis also demonstrates the high accuracy of 95.6% is reached in just 210 communication rounds which is better than evaluated federated baselines under the simulated conditions. Based on these quantitative results, it can be concluded that the adaptive aggregation and communication optimization system can possibly improve the system of agricultural intelligence in distributed systems. This model ensures participating farms benefit from shared model learning and retains data on each farm local and under the control of the farmer. Future deployments should consider the formal accounting of privacy, real world validation at farms, detection of rare diseases and intermittent connections resilience.

CRediT Author Statement
The author reviewed the results and approved the final version of the manuscript.

Dataset Availability
The Plant Village dataset is publicly available through the repository at https://github.com/spMohanty/PlantVillage-Dataset.

Conflicts of Interests
The author(s) declare(s) that they have no conflicts of interest.

Funding
No funding was received for conducting this research.

Competing Interests
There are no competing interests.

References
[1]. S. M. Shawon et al., "Federated Learning Approaches in Precision Agriculture: A Systematic Review of Current Trends, Challenges, and Opportunities," IEEE Access, vol. 14, pp. 30513–30538, 2026, doi: 10.1109/access.2026.3667548.
[2]. W. Chorney, A. Rahman, Y. Wang, H. Wang, and Z. Peng, "Federated Learning for Heterogeneous Multi-Site Crop Disease Diagnosis," Mathematics, vol. 13, no. 9, p. 1401, Apr. 2025, doi: 10.3390/math13091401.
[3]. S. Mehta, V. Kukreja, and S. Vats, "Empowering Farmers with AI: Federated Learning of CNNs for Wheat Diseases Multi-Classification," 2023 4th International Conference for Emerging Technology (INCET), pp. 1–6, May 2023, doi: 10.1109/incet57972.2023.10170091.
[4]. R. Janga and R. Dave, "Enhancing Smart Farming Through Federated Learning: A Secure, Scalable, and Efficient Approach for AI-Driven Agriculture," Artificial Intelligence and Applications, vol. 4, no. 3, pp. 265–279, Jun. 2025, doi: 10.47852/bonviewaia52025089.
[5]. P. K. Myakala, C. Bura, and A. K. Jonnalagadda, "Federated Learning and Data Privacy: A Review of Challenges and Opportunities," International Journal of Research Publication and Reviews, vol. 5, no. 12, pp. 1867–1879, Dec. 2024, doi: 10.55248/gengpi.5.1224.3512.
[6]. K. Rajput, A. Garg, V. Kukreja, and S. Mehta, "Agricultural Intelligence: Federated Learning CNN's Models for Jute Leaf Disease Analysis," 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1–6, Mar. 2024, doi: 10.1109/icrito61523.2024.10522218.
[7]. D. Jain, N. Bhati, and N. Bhardwaj, "Federated Learning and Its Impact on Decision-Making in Smart Agriculture," Federated Learning for Smart Agriculture and Food Quality Enhancement, pp. 271–291, Dec. 2025, doi: 10.1002/9781394338726.ch11.
[8]. S. M. Shawon et al., "Integrating Federated Learning and Explainable AI for Plant Disease Detection: A Systematic Review of Recent Advances in Precision Agriculture," International Journal of Computational Intelligence Systems, vol. 19, no. 1, Mar. 2026, doi: 10.1007/s44196-026-01219-w.
[9]. G. Ramkumar, M. Rama, S. G. Lakshmi, H. Awwad, and D. Manogna, "Federated Learning Framework for Privacy-Preserving Crop Disease Detection Across Distributed Farms," 2025 7th International Conference on Innovative Data Communication Technologies and Application (ICIDCA), pp. 1276–1281, Oct. 2025, doi: 10.1109/icidca66325.2025.11280457.
[10]. A. Alsuwaidi, M. A. Talib, Q. Nasir, F. Lamghari, S. Zerisenay, and F. M. Tsombou, "AI-Driven Early Detection of Crop Diseases Using Computer Vision and Federated Learning," 2025 IEEE 4th International Conference on Computing and Machine Intelligence (ICMI), pp. 1–5, Apr. 2025, doi: 10.1109/ic mi65310.2025.11141296.
[11]. S. Behera, N. Padhy, R. Panigrahi, and S. Kumar Kuanar, "Crop Disease Prediction Using Deep Learning in a Federated Learning Environment: Ensuring Data Privacy and Agricultural Sustainability," Procedia Computer Science, vol. 254, pp. 137–146, 2025, doi: 10.1016/j.procs.2025.02.072.
[12]. B. Singh, A. Kotiyal, P. Kumar, U. K. C, N. Vats Doohan, and R. P. Ram Kumar, "Federated Learning-Driven Crop Disease Diagnosis Framework to Ensure Privacy-Preserving Predictive Modeling Across Distributed Agricultural Zones," 2026 2nd International Conference on Big Data & Machine Learning (ICBDML), pp. 1–6, Feb. 2026, doi: 10.1109/icbdml68582.2026.11545115.
[13]. Abhishek, M. Rai, A. P. Singh, and V. Jha, "Harnessing the Power of Federated Learning for Agricultural Innovation," Federated Learning for Smart Agriculture and Food Quality Enhancement, pp. 1–26, Dec. 2025, doi: 10.1002/9781394338726.ch1.
[14]. H. Devaraj et al., "RuralAI in Tomato Farming: Integrated Sensor System, Distributed Computing, and Hierarchical Federated Learning for Crop Health Monitoring," IEEE Sensors Letters, vol. 8, no. 5, pp. 1–4, May 2024, doi: 10.1109/lsens.2024.3384935.
[15]. R. Shashank and R. M. Bhavadharini, "Federated learning with dynamic weighted aggregation for multi-crop disease detection: a hybrid CNN-transformer approach," Frontiers in Artificial Intelligence, vol. 9, May 2026, doi: 10.3389/frai.2026.1751118.
[16]. R. Saidi, T. Moulahi, S. Zidi, and S. Mahfoudhi, "A privacy-aware federated learning framework for collaborative healthcare IoT security," Machine Learning with Applications, vol. 25, p. 100944, Sep. 2026, doi: 10.1016/j.mlwa.2026.100944.
[17]. P. V. Caminha and H. M. N. d. S. Oliveira, "Plant Disease Detection Using Federated Learning and Cloud Infrastructure for Scalability and Data Privacy," Journal of Internet Services and Applications, vol. 16, no. 1, pp. 530–543, Sep. 2025, doi: 10.5753/jisa.2025.5055.

Publisher's note: The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. The content is solely the responsibility of the authors and does not necessarily reflect the views of the publisher.

ISSN: 3104-4654