FLitViT: Vision Transformers Based Federated Learning approach for Crop Disease Classification

Sai Mahesh Mudavat
Department of Computer Science and Engineering
University of North Texas
Denton, TX, USA
SaiMaheshMudavat@my.unt.edu

Alakananda Mitra
Nebraska Water Center
University of Nebraska-Lincoln
Lincoln, NE, USA
amitra6@unl.edu

Saraju P. Mohanty
Department of Computer Science and Engineering
University of North Texas
Denton, TX, USA
saraju.mohanty@unt.edu

Elias Kougianos
Department of Electrical Engineering
University of North Texas
Denton, TX, USA
elias.kougianos@unt.edu

***Abstract*—Plant diseases cause significant global crop losses affecting up to 33% annually, threatening food security. Despite advances in AI-based diagnosis, real-world deployment remains constrained by data privacy, limited edge computing, and lack of auditability. To address these challenges, we introduce FLitViT, a federated learning framework for privacy-preserving tomato disease classification on low-cost agricultural edge devices. FLitViT integrates a compact MobileViT-XXS(extra-extra-small) model with quantization-aware training and intelligent client coordination for efficient edge deployment. A Temporal Checkpoint Manifold (TCM) module ensures auditability and system recovery through round-wise checkpoints. Additionally, real-time client monitoring and adaptive timeout extensions help manage connectivity variability in rural environments. Experiments show FLitViT achieves 97.8% test accuracy in just three communication rounds with minimal accuracy loss. This enables cost-effective, auditable, and scalable agricultural AI, marking a practical step toward secure, decentralized smart agriculture.**

***Index Terms*—Federated learning, quantization, Edge AI, Plant disease classification, Privacy-preserving ML, Internet of Things(IoT) Smart Agriculture.**

I. Introduction

Agricultural Productivity is persistently being threatened by Plant diseases, with 20-40% global crop production getting affected annually [1]. Traditional diagnosis methods like visual inspection, are time-consuming, subjective, and often lacks early intervention. In contrast, machine learning and deep learning has emerged as a potential for automated disease detection at early stages with higher accuracy [2]. Among these models, those with transfer learning from pretrained networks can identify subtle disease patterns [3]. Edge computing capitalised these advances by enabling real-time, on-site disease detection without relying on constant internet access [4], [5]. Despite these advances practical deployment of AI faces challenges like data privacy which includes crop health trends [6], connectivity and powerful hardware limitations in rural settings [7]. Federated learning (FL) enables decentralized model training across multiple clients while keeping raw local data on devices [8], [9]. However, most FL systems still struggle with non-Independent and Identical Distributed (IID) data, poor connectivity, and device constraints. While approaches like DP-RTFL [10] highlight robustness and accountability, such strategies remain underexplored in agriculture.

The paper begins with a review of related work in Section II, followed by the novel contributions in Section III. Section IV details the proposed framework, while Section V presents the experimental validation. Finally, Section VI concludes the paper and outlines future directions.

II. Related works

Federated learning (FL) has surfaced as a promising paradigm for agricultural AI applications, enabling collaborative model training while preserving data privacy. Recent works have explored FL for crop disease detection, with [11] demonstrating image-based disease classification using ResNet50 with 94.2% accuracy. Enhanced aggregation strategies for plant leaf disease detection were proposed in [12], while UAV-based FL systems for plant diagnosis were introduced in [13]. Agricultural-specific approaches include privacy-preserving frameworks such as PEFL [14] and CNN-based FL for mango leaf disease detection [15], though these lack edge optimization and traceability. A constrained-edge FL framework, ToEFL [16], was also introduced, but it does not support non-IID agricultural data or domain-specific auditability. Vision-language-based FL using prompt tuning for agricultural object detection has been explored in [17], though it omits quantization, explainability, and edge deployment support. To enhance the efficiency of Federated Learning on edge devices, quantization strategies have been a key area of exploration, such as using INT8 models in industrial IoT settings [18]. However, these approaches target general industrial applications, and lack the agricultural domain-specific considerations necessary for practical farm deployment.

However, existing FL frameworks lack a unified agricultural solution that combines quantization-aware models, temporal checkpointing, and rural-connectivity resilience. This creates a significant gap for developing secure, auditable, and scalable AI systems for edge deployment. Table I summarizes these observations in current approaches.

TABLE I: Relevant Works Using Federated learning methods in agricultural applications

Reference	**Model & FL Approach**	**Key Observation**
Kabala et al. [11]	ResNet50, Standard FL	Unoptimized for Edge
Hari et al. [12]	Hierarchial CNN, Enhanced aggregation	Missing quantization support, Imbalanced Dataset
Ahmed et al. [13]	EfficientNetB3 CNN, UAV based Standard FL	Limited connectivity handling
Kumar et al. [14]	LSTM-AE, FedGRU	Heavy computational requirements
Mehta et al. [15]	CNN, Standard FL	Inconsistent Reporting, Lack Edge Deployment
Mitra et al. [16]	MobileNetV2, Edge based Standard FL	No quantization awareness
Li et al. [17]	Vision-Language DINO model, Prompt tuning	Missing audit capabilities
Ma et al. [18]	Quantized MobileNet, Distributed FL	Industrial focus only
FLitViT	**MobileViT-XXS, INT4 FL**	**Quantization aware with TCM**

III. Novel Contributions

A. Problem Statement

Accurate plant disease detection requires AI solutions that protect sensitive farm data, operate on low-power edge devices, and meet regulatory standards. While Vision Transformers (ViTs) offer high accuracy, their size and compute demands limit deployment on devices like Raspberry Pi. Centralized models raise privacy concerns and require stable internet, while isolated edge models struggle with generalization. Federated learning (FL) presents a promising alternative, but existing FL frameworks often fail in agricultural settings due to non-Independent and Identically Distributed (non-IID) data, limited transparency, and unstable rural connectivity. The framework is validated on a publicly available tomato leaf disease dataset [19], ensuring relevance to real-world agricultural deployments To address these gaps, we propose FLitViT, a lightweight, privacy-preserving FL framework with quantized ViTs, audit-ready checkpoints, and robust edge coordination.

B. Novelty and Significance of the Proposed Solution

The novel contributions of FLitViT are as follows:

- A lightweight federated learning framework for tomato disease classification, tailored for edge deployment using quantized MobileViT-XXS with non-IID data support.
- Connectivity-aware client coordination with real-time tracking and adaptive timeouts for resilient training under rural network conditions.
- Temporal Checkpoint Manifold (TCM) for immutable audit trails, secure rollback, and regulatory-compliant versioning.
- Quantization-aware training enabling efficient deployment on 4GB Raspberry Pi devices, achieving 97.8% accuracy and 73% model size reduction without raw data sharing.

IV. Proposed Framework

A. System Overview

FLitViT adopts a federated learning architecture composed of three key components: a central coordination server, distributed edge clients (Raspberry Pi devices), and a Temporal Checkpoint Manifold (TCM) for round-wise state management. The system leverages the MobileViT-XXS model, tailored for 256×256 pixel image inputs for low-power edge devices. This model was selected based on its proven effectiveness in prior research [20], where it was successfully trained as a lightweight student model via knowledge distillation for on-device agricultural inference. The server initializes the global model, partitions training data using randomized shuffling, and distributes client-specific packages. Federated rounds are orchestrated via HTTP-based RESTful APIs, enabling decentralized model updates using the Federated Averaging algorithm.

Throughout training, data remains entirely on local devices, ensuring both privacy preservation and collaborative model improvement across geographically dis-

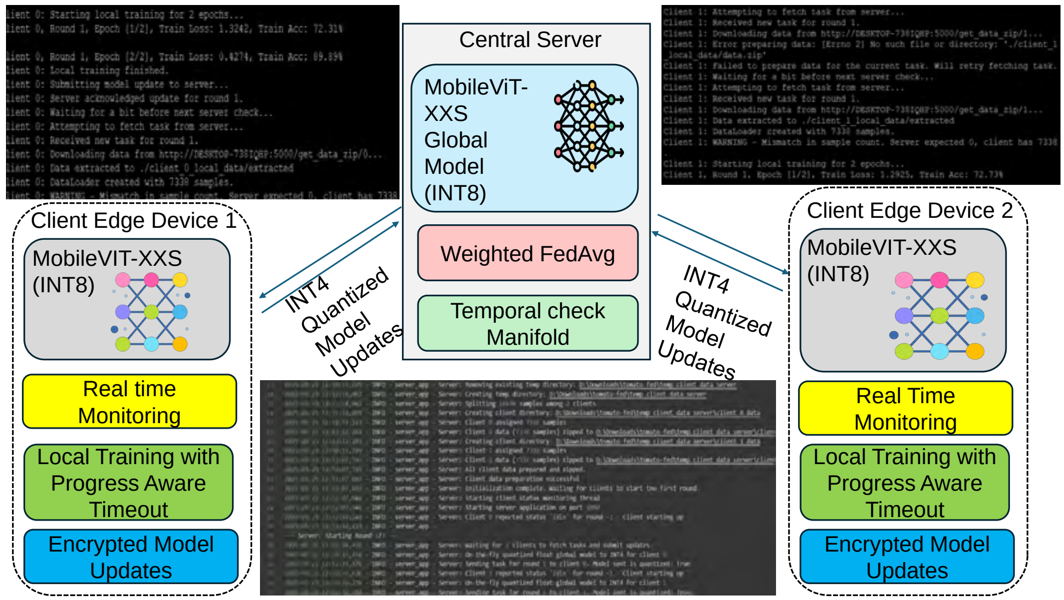


Fig. 1: Architecture of FLitViT: Edge-Aware Federated Learning with Quantization Aware TCM

tributed farm environments. Figure 1 shows the FLitViT architecture with a central server, edge clients, and a TCM module enabling private, quantized model training and auditable aggregation.

B. *Federated Learning Protocol*

FLitViT follows a round-based training protocol where the server distributes the global model to selected clients. Each client performs local training and returns model updates, which are aggregated using the Federated Averaging (FedAvg) algorithm using equation 1 :

$$M_t = \sum_k \left(\frac{n_k}{n}\right) \Delta M_k^t \tag{1}$$

Here, M_t is the global model at round t, n_k is the number of local samples held by client k, n is the total number of samples across all participating clients, and ΔM_k^t is the model update computed by client k during round t. To handle unreliable rural connectivity, FLitViT integrates an Intelligent Client Status Monitoring module. This module tracks each client's last-seen time, local training progress, and submission status, classifying clients as active, completed, or failed. If the round deadline is reached while active clients are still progressing, the server adaptively extends the timeout. This mechanism prevents premature aggregation and ensures more robust and complete training rounds. Algorithm 1 details the timeout-aware client monitoring logic and Algorithm 2 summarizes the weighted model aggregation performed at the server.

Algorithm 1 Intelligent Client Status Monitoring

Input: Client states S, round start time t_0, max time T_{max}, timeout threshold $T_{threshold}$
Result: Boolean flag to proceed with aggregation

1) Get current time and calculate elapsed time;
2) Initialize client status sets (done, training, failed);
3) For each client i:
   Check submission status;
   Calculate time since last seen;
   if $\Delta t > T_{threshold}$ **then** mark as failed;
   else if training with progress $>$ 0 **then** add to training set;
   end
4) **if** elapsed time $> T_{max}$ and clients still training **then**
   Calculate max progress among training clients;
   if progress $>$ 0 **then** extend timeout and return False;
   end
5) Return True if timeout reached or no active training clients;

C. *Temporal Checkpoint Manifold (TCM)*

While client coordination ensures robust training in the face of connectivity issues, maintaining transparency and accountability over time is equally important especially in regulated domains like agriculture. To address this, FLitViT introduces the Temporal Checkpoint Manifold (TCM). The TCM module generates immutable checkpoints at the end of each training round. Each checkpoint captures the global model state, round ID, client participation logs, performance metrics, and con-

Algorithm 2 Weighted Federated Averaging

Input: Client updates $U = \{(\theta_1, n_1), (\theta_2, n_2), \ldots, (\theta_K, n_K)\}$

Result: Global model parameters θ_g

1) **if** no client updates **then** return current global model;
2) Calculate total samples N_{total} across all clients;
3) Initialize global model parameters θ_g with zeros;
4) For each client $i = 1$ to K:
   Calculate weight $w_i = n_i/N_{total}$;
   Update $\theta_g[key] \leftarrow \theta_g[key] + w_i \cdot \theta_i[key]$ for all parameters;
   end
5) Return aggregated global model θ_g;

figuration metadata. All entries are timestamped and uniquely identified, enabling secure audit trails and precise rollback when needed. This mechanism supports reproducible experimentation, fault-tolerant recovery, and regulatory compliance making the system reliable not only in performance but also in traceability.

D. Quantization-Aware Execution

To enable efficient deployment on edge devices, FLitViT integrates a quantization-aware training pipeline. The global model is initially quantized using PyTorch's dynamic INT4 scheme and distributed with embedded metadata indicating its format. Clients automatically detect and deserialize the quantized model for training, and the aggregated global update is re-quantized before the next round. This process results in a 73% reduction in model size with negligible accuracy loss, enabling fast inference on 4GB Raspberry Pi devices. Transparent handling of quantization across clients ensures consistent training and performance tracking, even in heterogeneous environments.

V. Experimental Validation

A. Dataset Configuration and Distribution

We used a publicly available tomato leaf disease dataset consisting of 22,930 high-resolution images across 10 classes, including bacterial spot, early blight, late blight, leaf mold, septoria, spider mites, target spot, yellow leaf curl virus, mosaic virus, and healthy leaves. To ensure clear separation between training, validation, and testing, the dataset was partitioned into three subsets. A total of 14,676 images (64%) were used for federated training, distributed across two edge clients. Another 4,586 images (20%) were reserved for server-side validation to monitor model performance during training rounds. The remaining 3,668 images (16%) form a test set, used exclusively for final evaluation. This configuration prevents data leakage and supports robust generalization analysis.

B. System Configuration and Evaluation Framework

To support deployment on resource-constrained devices, the MobileViT-XXS model, was chosen as the base architecture. Training was performed using a learning rate of 1×10^{-4}, a batch size of 16, and one local epoch per federated round, with the Adam optimizer. The system was trained over three federated rounds using the distributed training split, with performance continuously monitored on the server-side validation set.

Final evaluation was conducted on the held-out test set to provide an unbiased measure of generalization performance. The experimental setup included two Raspberry Pi 4 devices acting as federated clients and a standard laptop serving as the central coordination server.

C. Evaluation Metrics and Methodology

The proposed framework was evaluated across five key dimensions. Convergence was tracked over federated rounds using validation loss and accuracy. Final model generalization was assessed on an unseen test set to ensure unbiased performance. Efficiency was measured in terms of model size reduction and inference latency after quantization on 4GB Raspberry Pi devices. System robustness was evaluated based on the framework's ability to handle timeouts, communication delays, and client failures. Lastly, auditability was validated through consistent Temporal Checkpoint Manifold (TCM) generation and reliable rollback operations.

D. Federated Learning Performance Analysis

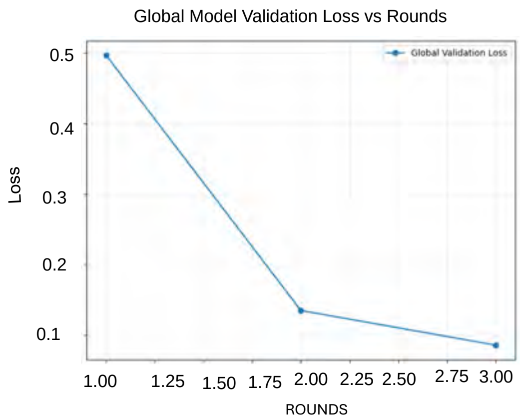
FLitViT demonstrated strong convergence within just three federated rounds. Validation accuracy improved from 91.6% in Round 1 to 96.7% in Round 2, reaching 97.8% by Round 3. In parallel, validation loss decreased sharply from 0.495 to 0.085, representing an 83% overall reduction. These trends are illustrated in Figure 2, they highlight rapid accuracy gains and exponential loss decay, particularly during the early rounds. Following training, the global model achieved 97.8% accuracy and a loss of 0.0782 on the held-out 3,668-image test set, confirming its strong generalization capability. Table II summarizes accuracy and loss metrics across federated rounds.

TABLE II: Federated Learning Performance Results

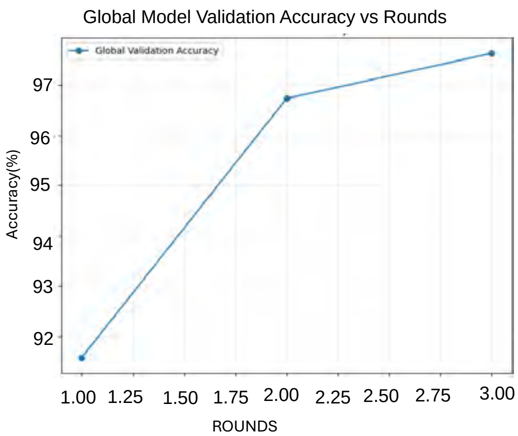
Round	Val. Acc (%)	Val. Loss	Test Acc (%)	Time (min)
1	91.6	0.495	92.4	~12.3
2	96.7	0.142	96.2	~11.8
3	**97.6**	**0.085**	**97.8**	~11.4

E. Quantization Efficiency Evaluation

INT4 quantization was applied to the global model prior to training and maintained consistently across all federated rounds. This approach led to a 73% reduction in model size without compromising convergence or classification performance. The system seamlessly



(a) Global Model Validation Across Federated Rounds



(b) Global Model Accuracy Across Federated Rounds

Fig. 2: Validation Accuracy and Loss Curves Over Federated Rounds

handled quantized serialization, deserialization, and re-quantization, as verified through server-side logs.

Despite the aggressive compression, the quantized model exhibited only a 0.2% drop in accuracy compared to its full-precision counterpart, while substantially improving memory usage and communication efficiency. These results confirm the stability and practicality of the quantization-aware training pipeline in constrained edge environments.

F. Communication, Fault Tolerance and Scalability Analysis

FLitViT maintained efficient communication and operational robustness throughout all federated rounds. By exchanging quantized models, the per-round transmission size was reduced to 1.4MB cutting bandwidth usage by 60% compared to centralized methods and enabling deployment in low-connectivity farm environments. Each round averaged 12.3 minutes to complete. The adaptive timeout mechanism accurately detected active clients in 95% of cases, extending deadlines only when over 50% training progress was observed. This strategy minimized idle waiting by 40% and prevented premature aggregation, ensuring uninterrupted training and sustained client participation. Table III compares communication cost between centralized and FLitViT setups.

TABLE III: Communication Overhead Comparison Between FLitViT and Fedavg

Method	Upstream (MB)	Down stream (MB)	Total(MB)	Reduction (%)
FedAvg	3.8	3.8	7.6	0%
FLitViT	**1.2**	**1.2**	**2.4**	**+60%**

G. Temporal Checkpoint Manifold Evaluation

TCM proved effective in maintaining a comprehensive audit trail. Checkpoints captured 100% of necessary metadata, including client participation logs, model parameters, and validation metrics. Overhead from checkpoint creation remained under 2% of round time. To contextualize all these results, Table IV compares the performance of FLitViT with other notable methods, demonstrating its superior convergence and competitive accuracy.

TABLE IV: Performance Comparison of Federated Learning Methods For Agricultural Edge Deployment

Method	Accuracy	Convergence Rounds	Clients
Kabala et al. [11]	99.3%	30	3, 5, 7
Hari et al. [12]	93%	Reduced (Explicity Mentioned)	2
Ahmed et al. [13]	99.5%	Reduced (Explicity Mentioned)	4
Kumar et al. [14]	99.3%, 99.74%	100	2
Mehta et al. [15]	97%	100	4
Mitra et al. [16]	72%-74%	Event driven	1 Node
Li et al.[17]	24.12% MaP	325-500	3
FLitViT	**97.8%**	**3**	**3**

VI. Conclusion

This work introduces FLitViT, a federated learning framework tailored for agricultural disease detection on low-power edge devices. By combining quantization-aware training with a Temporal Checkpoint Manifold (TCM), the system achieves 97.8% classification accuracy on tomato leaf diseases while ensuring full data privacy. Model quantization reduced communication overhead by 60% and model size by 73%, enabling fast and efficient inference on 4GB Raspberry Pi devices. The TCM module further enhances transparency through audit-ready checkpointing, rollback support, and immutable state tracking critical for regulatory compliance in agricultural AI. While experimental results demonstrate strong performance, the current evaluation was limited to a controlled setup with homogeneous clients. Future work will address this by extending to multi-crop detection, integrating heterogeneous edge hardware, and exploring blockchain-based TCM enhancements for improved scalability and real-world robustness.

References

[1] FAO, IFAD, UNICEF, WFP, and WHO, *The State of Food Security and Nutrition in the World 2023: Urbanization, Agrifood Systems Transformation and Healthy Diets Across the Rural–Urban Continuum*. Rome, Italy: FAO, 2023.

[2] K. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, Feb. 2018. DOI: 10.1016/j.compag.2018.01.009.

[3] W. Shafik, A. Tufail, C. Liyanage, and R. Apong, "Using transfer learning-based plant disease classification and detection for sustainable agriculture," *BMC Plant Biology*, vol. 24, p. 48, Feb. 2024. DOI: 10.1186/s12870-024-04825-y.

[4] A. T. Khan, S. Jensen, A. R. Khan, and S. Li, "Plant disease detection model for edge computing devices," *Frontiers in Plant Science*, vol. 14, p. 1 308 528, Dec. 2023. DOI: 10.3389/fpls.2023.1308528.

[5] M. Dong, H. Yu, Z. Sun, L. Zhang, Y. Sui, and R. Zhao, "Research on agricultural environmental monitoring internet of things based on edge computing and deep learning," *Journal of Intelligent Systems*, vol. 33, jisys-2023–0114, May 2024. DOI: 10.1515/jisys-2023-0114.

[6] R. Dembani, I. Karvelas, N. A. Akbar, S. Rizou, D. Tegolo, and S. Fountas, "Agricultural data privacy and federated learning: A review of challenges and opportunities," *Computers and Electronics in Agriculture*, vol. 232, p. 110 048, 2025. DOI: 10.1016/j.compag.2025.110048.

[7] M. Dhanaraju, P. Chenniappan, K. Ramalingam, S. Pazhanivelan, and R. Kaliaperumal, "Smart farming: Internet of things (iot)-based sustainable agriculture," *Agriculture*, vol. 12, no. 10, p. 1745, 2022. DOI: 10.3390/agriculture12101745.

[8] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020. DOI: 10.1109/MSP.2020.2975749.

[9] F. R. Mughal, J. He, B. Das, F. Dharejo, N. Zhu, S. Bhatia, and S. Alzahrani, "Adaptive federated learning for resource-constrained iot devices through edge intelligence and multi-edge clustering," *Scientific Reports*, vol. 14, p. 78 239, Nov. 2024. DOI: 10.1038/s41598-024-78239-z.

[10] A. Talluri, "Dp-rtfl: Differentially private resilient temporal federated learning for trustworthy ai in regulated industries," *arXiv preprint arXiv:2505.23813*, 2025.

[11] D. M. Kabala, A. Hafiane, L. Bobelin, and R. Canals, "Image-based crop disease detection with federated learning," *Scientific Reports*, vol. 19220, p. 46 218, Nov. 2023. DOI: 10.1038/s41598-023-46218-5.

[12] P. Hari, M. P. Singh, and A. K. Singh, "An improved federated deep learning for plant leaf disease detection," *Multimedia Tools and Applications*, vol. 83, no. 35, pp. 83 471–83 491, 2024.

[13] R. Ahmed, E. D. F. Khan, S. Khan, M. N. H. Mohd, H. Mohd, A. Waseem, M. N. A. Khan, S. Ali, A. Jamal, and H. Khan, "Federated learning-based uavs for the diagnosis of plant diseases," in *Proceedings of the International Conference on Electrical Engineering and Technology (ICEET)*, Oct. 2022. DOI: 10.1109/ICEET56468.2022.10007133.

[14] P. Kumar, G. P. Gupta, and R. Tripathi, "Pefl: Deep privacy-encoding-based federated learning framework for smart agriculture," *IEEE Micro*, vol. 42, no. 1, pp. 33–40, 2022. DOI: 10.1109/MM.2021.3112476.

[15] S. Mehta, V. Kukreja, and S. Vats, "Advancing agricultural practices: Federated learning-based cnn for mango leaf disease detection," in *Proceedings of the 2023 3rd International Conference on Intelligent Technologies (CONIT)*, 2023, pp. 1–6. DOI: 10.1109/CONIT59222.2023.10205850.

[16] A. Mitra, S. Mohanty, and E. Kougianos, "ToEFL: A Novel Approach for Training on Edge in Smart Agriculture," in *Proceedings of the ACM International Conference*, Jun. 2024, pp. 657–662. DOI: 10.1145/3649476.3660381.

[17] L. Li, J. Li, D. Chen, L. Pu, H. Yao, and Y. Huang, "Vllfl: A vision-language model based lightweight federated learning framework for smart agriculture," *arXiv preprint arXiv:2504.13365*, 2025.

[18] T. Ma, H. Wang, and C. Li, "Quantized distributed federated learning for industrial internet of things," *IEEE Internet of Things Journal*, vol. 10, no. 4, pp. 3027–3036, 2021.

[19] Noulam, *Tomato - a large dataset of tomato leaf diseases*, https://www.kaggle.com/datasets/noulam/tomato, Accessed: 2025-04-17, 2024.

[20] S. M. Mudavat, S. P. Mohanty, A. Mitra, and E. Kougianos, "LiteViT: Leveraging the Power of Transformers for Edge AI in Crop Disease Classification," in *Proceedings of the IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, Accepted, IEEE, 2025.