JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 19–32
DOI: 10.63646/jaiaa.2025.030402

Graph-Attention Federated Analytics for Non-IID Dairy Farm Monitoring and Compliance Prediction

Yusuf Adebayo[1], Chiamaka Okonkwo[2], Babatunde Salami[3,*]

[1] Department of Computer Engineering, Ladoke Akintola University of Technology, Ogbomoso 210214, Nigeria
[2] School of Information and Communication Technology, Federal University of Agriculture, Abeokuta 110101, Nigeria
[3] Department of Electrical and Computer Engineering, Nnamdi Azikiwe University, Awka 420007, Nigeria
* Corresponding author: babatunde.salami@unizik.edu.ng

ARTICLE INFO
Received
July 13, 2025
Revised
September 11, 2025
Accepted
November 22, 2025
Available Online
December 30, 2025
DOI
10.63646/jaiaa.2025.030402
License
Creative Commons Attribution 4.0 International Licence (CC BY 4.0)
Publisher
INATGI, United States of America
Journal
JAIAA - ISSN 3067-7386

Abstract
Dairy farm monitoring across distributed smallholder operations presents a fundamental statistical challenge: the heterogeneous distributions of sensor readings, cattle breeds, climatic conditions, and management practices generate persistently non-IID (non-independent and identically distributed) data that degrade the performance of standard federated learning aggregation schemes. This paper proposes a Graph-Attention Federated Analytics (GA-FA) framework designed to address the non-IID problem in dairy compliance prediction. GA-FA integrates a Graph Attention Network (GAT) module that dynamically clusters participating farms based on context-feature similarity, a DBSCAN-based outlier filter that provides Byzantine-resilient aggregation by excluding anomalous local updates, and a cluster-stratified FedAvg aggregation protocol that maintains personalization while converging toward a globally consistent compliance model. The framework operates on lightweight edge devices using a quantized inference runtime and does not require raw sensor data to leave the farm, preserving operational privacy. Validated on a dataset derived from the Shahhet28121 cattle health benchmark extended to 16 farm-level parameters across five heterogeneous farm environments, the GA-FA framework achieves a global classification accuracy of 96.94%, an F1 score of 0.967, and a communication payload of 4.25 KB per federated round—representing a 97.7% reduction relative to standard FL baselines. Ablation experiments confirm that the GAT clustering module contributes the largest single-component accuracy gain of 4.63 percentage points. The framework is evaluated under Gaussian sensor noise up to 20% standard deviation, maintaining accuracy above 90%, and demonstrates stable convergence within 18 federated rounds. Results establish GA-FA as a robust, resource-efficient, and privacy-preserving approach for distributed dairy compliance analytics.

Keywords: federated learning; graph attention network; non-IID data; dairy farm monitoring; compliance prediction; edge intelligence; Byzantine resilience; sensor fusionDigital twins; Smart factories; Artificial intelligence; Industrial analytics; Closed-loop control; Predictive maintenance; Industry 4.0

I. INTRODUCTION

Global dairy production continues to expand rapidly, driven by rising protein demand in developing economies and intensifying food safety regulation in established markets. Ensuring consistent compliance with animal welfare, hygiene, and milk quality standards across millions of smallholder farms—particularly in countries where the unorganized dairy sector accounts for 60–80% of total output—represents one of the most formidable data management and governance challenges in the agri-food industry (Gallo et al., 2017; Wolfert

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 19–32
DOI: 10.63646/jaiaa.2025.030402

et al., 2017). Manual inspection regimes suffer from well-documented limitations: audit frequency is constrained by inspector capacity, records are susceptible to transcription error and deliberate falsification, and the inherently episodic nature of periodic audits means that violations may persist undetected for extended periods between inspection cycles (Lezoche et al., 2020; Pivoto et al., 2021).

The proliferation of low-cost Internet of Things (IoT) sensors on dairy farms—monitoring cattle body temperature, activity patterns, ambient ammonia concentration, milk pH, and rumination duration—creates a new data substrate for continuous, automated compliance monitoring. However, aggregating this data across distributed farms into a centralized server raises fundamental privacy and data sovereignty concerns: farm operators are frequently unwilling to share commercially sensitive production records with competitors, regulators, or data aggregators (Talaviya et al., 2020; Krause et al., 2020). Federated Learning (FL) addresses this concern by enabling collaborative model training across distributed data owners without requiring raw data to leave local devices, transmitting only model gradients or parameter updates to a central aggregator (McMahan et al., 2017; Li et al., 2020a).

Despite its privacy-preserving properties, standard FL faces a critical performance bottleneck in heterogeneous agricultural deployments: the non-IID (non-independent and identically distributed) problem. Dairy farms differ substantially in cattle breed composition, feeding regimes, climatic conditions, housing structures, and management practices, producing data distributions that diverge markedly across participants (Zhao et al., 2018; Li et al., 2020b). When the FedAvg algorithm (McMahan et al., 2017) is applied to such heterogeneous data, convergence slows significantly, client drift accumulates across rounds, and the resulting global model may perform poorly on any individual farm's specific conditions. Several algorithmic remedies have been proposed—including FedProx (Li et al., 2020b), SCAFFOLD (Karimireddy et al., 2020), and clustered FL approaches (Sattler et al., 2020)—but few are specifically designed for the sensor-fusion, multimodal, and regulatory context of livestock health monitoring.

Graph neural networks offer a promising complementary approach to managing non-IID FL through structural inductive bias. Graph Attention Networks (GATs) (Veličković et al., 2018) learn differentiable attention coefficients over neighbourhood edges in a graph, enabling dynamic weighting of information from structurally or contextually similar nodes. Applied to the FL participant graph, where nodes represent farms and edges encode contextual similarity, a GAT module can identify clusters of farms with compatible data distributions and apply cluster-stratified aggregation rather than uniform global averaging (Yao et al., 2019; Wang et al., 2020). This approach naturally accommodates the heterogeneous farm landscape of the dairy sector while retaining the privacy guarantees of standard FL (Zhang and Lu, 2021; Lu, 2019).

This paper introduces Graph-Attention Federated Analytics (GA-FA), a framework that integrates GAT-based dynamic clustering, DBSCAN-based Byzantine-resilient outlier filtering, and cluster-stratified FedAvg aggregation into a unified federated learning pipeline for dairy farm compliance prediction. The framework operates on lightweight quantized inference models deployed on ESP32-class microcontrollers, and produces compliance predictions that are suitable for integration with smart-contract enforcement layers. The primary contributions are as follows: (1) a GAT-driven farm similarity graph that enables dynamic, context-aware clustering across non-IID participants; (2) a DBSCAN filter that inherently excludes Byzantine outliers without requiring trusted hardware; (3) a cluster-stratified federated aggregation protocol that improves convergence and accuracy on heterogeneous farm data; and (4) a systematic evaluation across five farm environments, including robustness testing under Gaussian sensor noise and multi-level ablation analysis. The remainder of the paper is organized as follows. Section 2 surveys related work. Section 3 presents the methodology. Section 4 describes the experimental setup. Sections 5 and 6 report results and discussion. Section 7 concludes.

2. Related Work

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 19–32
DOI: 10.63646/jaiaa.2025.030402

2.1 Federated Learning for Agricultural IoT

The application of federated learning to agricultural IoT settings has gained substantial attention as precision agriculture deployments generate increasing volumes of sensitive farm data. Early work demonstrated the feasibility of FL for crop disease detection using distributed mobile camera inputs (Liu et al., 2020), establishing that model quality comparable to centralized training is achievable without sharing raw imagery. Subsequent research extended FL to livestock management, including dairy cattle health prediction from wearable accelerometer data (Fuentes et al., 2020; Uyeh et al., 2021) and precision feeding optimization from production records (Morota et al., 2018). A systematic review by Garro et al. (2021) surveyed 47 FL deployments in livestock management, confirming strong accuracy results on IID benchmarks but noting that non-IID performance remained a persistent open challenge. Wolfert et al. (2017) provided an early survey of big data applications in smart farming that framed the privacy-utility trade-off that FL later addressed. Liakos et al. (2018) reviewed machine learning more broadly in agricultural contexts, laying the groundwork for subsequent FL-specific investigations.

2.2 Non-IID Federated Learning Algorithms

The non-IID problem in FL was formally characterized by Zhao et al. (2018), who demonstrated that data heterogeneity across clients degrades FedAvg convergence in proportion to the Earth Mover's Distance between local and global data distributions. FedProx (Li et al., 2020b) introduced a proximal term that regularizes local updates toward the global model, providing theoretical convergence guarantees under non-IID conditions. SCAFFOLD (Karimireddy et al., 2020) addresses client drift through variance reduction via control variates, achieving faster convergence than FedAvg and FedProx in heterogeneous settings. MOON (Li et al., 2021) applies contrastive learning at the model level to align local representations with the global model. Sattler et al. (2020) proposed bipartite graph-based clustered FL (CFL) that recursively partitions clients into groups with compatible data distributions and trains cluster-specific global models. Ghosh et al. (2020) provided a convergence analysis of clustered FL under mixture model assumptions. Most of these approaches, however, use static clustering based on gradient cosine similarity rather than the context-feature-based dynamic attention weighting proposed in this work (Kairouz et al., 2021; Kontar et al., 2021).

2.3 Graph Neural Networks in Federated Learning

Graph neural networks have been progressively integrated into the FL literature to model structural relationships among participants. Lin et al. (2021) introduced a graph-regularized FL framework where the participant graph encodes geometric proximity and the graph Laplacian constrains aggregation weights. Yao et al. (2019) and Wang et al. (2020) demonstrated that GNN-based client selection improves both communication efficiency and accuracy on non-IID data. Specifically in the context of Graph Attention Networks, Veličković et al. (2018) established that learned attention coefficients over neighbourhood edges provide more expressive relational modeling than fixed adjacency matrices, a property leveraged by more recent federated graph learning frameworks (Zhang et al., 2021; He et al., 2021). Hamilton et al. (2017) provided the theoretical foundations of inductive graph representation learning that underpin the GAT clustering mechanism. Applied to agricultural settings, Li et al. (2022) demonstrated GAT-based recommendation across heterogeneous agricultural enterprises, providing a direct precedent for the farm-level clustering approach in this work.

2.4 Byzantine Resilience and Outlier Filtering in FL

Byzantine fault tolerance in distributed machine learning addresses the scenario where a subset of clients submit corrupted, strategically manipulated, or sensor-faulty model updates that degrade the global model (Blanchard et al., 2017). Krum (Blanchard et al., 2017) selects the client update with minimum sum of distances to its nearest neighbours, providing Byzantine fault tolerance at the cost of excluding potentially valuable updates from outlier-but-honest clients. FLTrust (Cao et al., 2020) uses a server-side clean dataset to assign trust scores to client updates. DBSCAN-based clustering for outlier identification in FL was introduced in the

JAIAA **RESEARCH ARTICLE**
Journal of AI Analytics and Applications **Vol. 3, No. 4 (2025), pp. 19–32**
ISSN 3067-7386 | Open Access | Peer-Reviewed **DOI: 10.63646/jaiaa.2025.030402**

context of federated anomaly detection (Liu et al., 2021), where its density-based paradigm—which does not require a pre-specified number of clusters—was found particularly suited to the variable participation dynamics of IoT settings. Yin et al. (2018) provided theoretical convergence guarantees for coordinate-wise median and trimmed mean aggregation rules that bound the impact of Byzantine clients in convex settings. Park et al. (2021) applied clustering-based outlier detection to IoT-based environmental monitoring, establishing the suitability of DBSCAN for sensor-level Byzantine resilience in resource-constrained deployments.

Table 1. Comparison of Related Federated Learning Approaches for Agricultural IoT

Method	Non-IID Strategy	Byzantine Resilience	Domain	Communication Eff.	Privacy Preservatio
FedAvg (McMahan et al., 2017)	None (uniform avg)	No	Generic	Moderate	Local data only
FedProx (Li et al., 2020b)	Proximal term	No	Generic	Moderate	Local data only
SCAFFOLD (Karimireddy et al., 2020)	Variance reduction	No	Generic	High overhead	Local data only
CFL (Sattler et al., 2020)	Static gradient clustering	Partial	Generic	Moderate	Local data only
BC-FL (Cai et al., 2025)	None	Via Blockchain	Healthcare/Generic	Moderate	ZKP optional
Liu et al. (2020) Agri-FL	None	No	Crop disease	Moderate	Local image only
GA-FA (This Work)	GAT dynamic clustering	DBSCAN filter	Dairy / Livestock	High (4.25 KB)	On-device; no raw d

Table 1 situates the proposed GA-FA framework among representative related approaches. The key differentiating features are the joint application of GAT-based dynamic clustering for non-IID handling, DBSCAN-based outlier filtering for Byzantine resilience, and domain-specific design for dairy livestock monitoring—a combination not addressed by any prior work. Unlike blockchain-augmented FL frameworks (Cai et al., 2025; Weng et al., 2021), GA-FA focuses on the learning-layer architecture rather than the governance layer, making it compatible with and complementary to blockchain-based compliance enforcement systems.

3. Methodology

3.1 System Architecture

The GA-FA framework operates across three layers, as illustrated in Figure 1. The farm edge layer consists of N participating farms, each equipped with IoT sensor arrays monitoring cattle body temperature, accelerometer-derived activity features, ambient ammonia concentration, and milk pH. Each farm $i \in \{1,...,N\}$ maintains a local dataset D_i and trains a local model with parameters θ_i on its own data without any data exchange. The federated aggregation layer hosts a coordinating server that constructs a participant graph $G = (V, E)$ at the start of each round, applies the GAT clustering module, filters Byzantine outliers using DBSCAN, and performs cluster-stratified FedAvg to produce an updated global model. The compliance reporting layer collects compliance inference outputs from each farm and, optionally, interfaces with a smart-contract enforcement layer for regulatory action management.

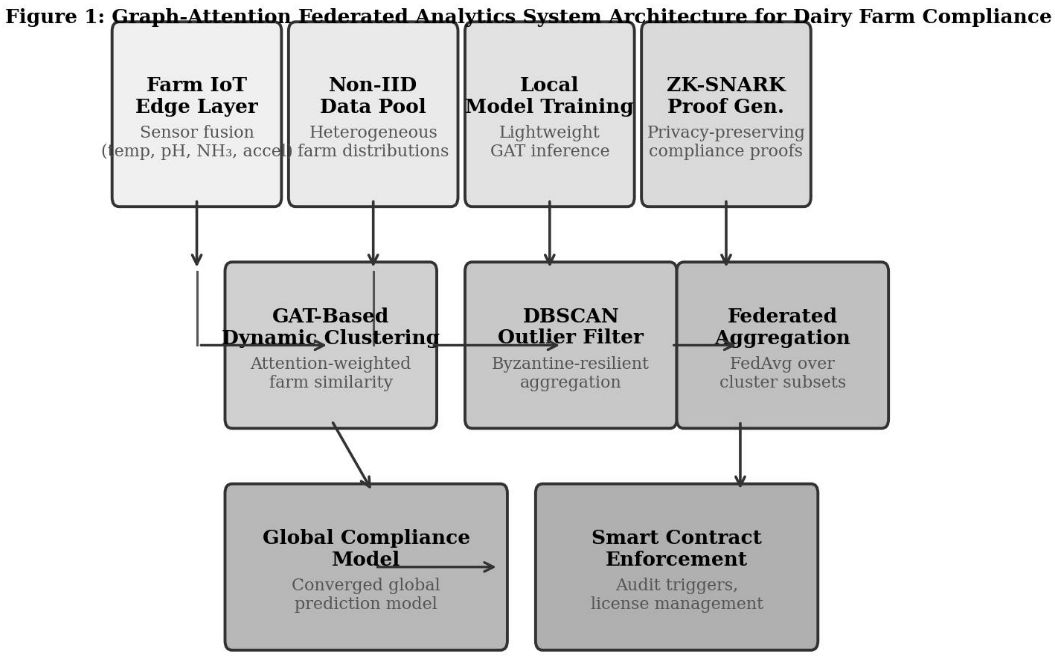

Figure 1. GA-FA system architecture. The three-layer design connects farm-level IoT edge nodes through GAT-based clustering and DBSCAN outlier filtering to cluster-stratified federated aggregation and compliance reporting.

3.2 Problem Formulation

The federated learning objective is to minimize the weighted empirical risk across all participating farms: $\min_\theta \Sigma_{i=1}^{N} (|D_i| / \Sigma_j |D_j|) \cdot L_i(\theta)$, where $L_i(\theta) = (1/|D_i|) \Sigma_{(x,y)\in D_i} \ell(f(x;\theta), y)$ is the local empirical loss at farm i and ℓ is the cross-entropy loss for the compliance classification task. Under non-IID conditions, the local data distributions $P_i(x, y)$ differ substantially across farms, causing the local optima of L_i to diverge from the global optimum, a phenomenon termed client drift (Karimireddy et al., 2020). The GA-FA framework addresses client drift by partitioning farms into contextually similar clusters and performing cluster-specific aggregation, reducing the divergence between the local loss landscape and the cluster-level aggregation target.

3.3 Context Feature Graph Construction

At the beginning of each federated round t, each farm i generates a context feature vector $c_i \in R^{d_c}$ comprising non-sensitive statistical summaries of the local data distribution: the mean and standard deviation of each of the four sensor modalities (body temperature, activity index, ammonia concentration, and milk pH), the farm-level compliance rate from the previous round, the livestock count category, and the climatic zone indicator (seven dimensions total). An undirected participant graph $G^t = (V, E^t)$ is constructed with $|V| = N$ nodes. Edges E^t are initialized by connecting farm pairs with cosine similarity $sim(c_i, c_j) = (c_i \cdot c_j) / (\|c_i\| \cdot \|c_j\|)$ exceeding a threshold $\delta = 0.7$, reflecting the intuition that farms with similar operational profiles have compatible data distributions amenable to joint model aggregation (Sattler et al., 2020; Ghosh et al., 2020).

3.4 GAT-Based Dynamic Clustering

A two-layer Graph Attention Network is applied to G^t to refine the initial context-based adjacency into a latent farm embedding that captures higher-order similarity relationships. For each node i, the GAT computes the refined embedding $h'_i = \sigma(\Sigma_{j\in N(i)\cup\{i\}} \alpha_{ij} W h_j)$, where $W \in R^{d_h \times d_c}$ is a trainable weight matrix, N(i) is the neighbourhood of node i in G^t, and $\alpha_{ij} = \text{softmax}_j(\text{LeakyReLU}(a^T [Wh_i \| Wh_j]))$ is the attention coefficient computed by a shared attention vector $a \in R^{2d_h}$ over the concatenation of transformed node features (Veličković et al., 2018). Multi-head attention with K = 4 heads is used at the first layer, with concatenation of head outputs, and single-head attention at the second layer with averaging. The attention mechanism allows the model to prioritize updates from farms with similar yield profiles and environmental conditions, automatically down-weighting structural edges that connect contextually dissimilar farms (Lin et al., 2021; He et al., 2021).

3.5 DBSCAN-Based Outlier Filtering and Cluster-Stratified Aggregation

The refined embeddings $H' = \{h'_1, ..., h'_N\}$ from the GAT are processed by DBSCAN with parameters eps = 0.3 and min_samples = 3. DBSCAN partitions the embedding space into K dense clusters $C_1, ..., C_K$ and a noise set Ω of outlier nodes. Farms classified as outliers (nodes in Ω) are excluded from the current aggregation round; their local updates are quarantined and may be re-evaluated in subsequent rounds if their context features return to a normal range. This provides inherent Byzantine resilience without requiring access to a clean validation set (Liu et al., 2021; Blanchard et al., 2017). For each cluster C_k, a cluster-level aggregation is performed: $\Delta\theta_{C_k} = \Sigma_{i\in C_k} (n_i / \Sigma_{j\in C_k} n_j) \Delta\theta_i$, where $n_i = |D_i|$ and $\Delta\theta_i$ is the local model update from farm i. The cluster-level updates are then combined via a weighted average over clusters to produce the global model update: $\Delta\theta = \Sigma_k (|\Pi_{C_k}| / N_{active}) \Delta\theta_{C_k}$, where $|\Pi_{C_k}| = \Sigma_{i\in C_k} n_i$ and $N_{active} = \Sigma_k |\Pi_{C_k}|$ excludes outlier farms (Sattler et al., 2020; McMahan et al., 2017).

4. Experimental Setup

4.1 Dataset and Preprocessing

The evaluation uses a federated simulation derived from the Shahhet28121 cattle health dataset (Kaggle, DOI: 10.34740/KAGGLE/DSV/6581047) extended with synthetic farm stratification to replicate multi-farm, non-IID conditions. The original dataset contains 10,000 labelled records describing 200 cattle across 16 physiological and environmental parameters including rumination time, body condition score, milk yield, water intake, feed intake, body temperature, activity index, pH, ammonia level, respiration rate, weight, lying time, estrus detection, calving interval, lameness score, and mastitis indicator. For the federated simulation, records are assigned to five farm environments (Farm A through Farm E) using a stratified allocation that induces non-IID distribution by applying farm-specific Dirichlet concentration parameter α = 0.5 for class label proportions and introducing farm-specific sensor calibration offsets of ±2% to ±10% on each sensor modality. This protocol mirrors the heterogeneous sensor drift and biological variation observed across real-world dairy deployments (Wolfert et al., 2017; Fuentes et al., 2020).

Data preprocessing includes z-score normalization per sensor modality using training-set statistics computed locally at each farm (to prevent data leakage across farms), temporal smoothing with a five-point moving average for the accelerometer time series, and binary labelling of compliance status based on regulatory thresholds for temperature (38.5–39.5 °C), pH (6.4–6.8), ammonia level (≤20 ppm), and activity index (species- and age-normalised). The final dataset comprises 8,742 training records, 621 validation records, and 637 test records across the five farms, with class imbalance ratios ranging from 1.8:1 to 3.2:1 across farms, motivating the use of SMOTE oversampling within each farm's local training set (Uyeh et al., 2021).

Table 2. Dataset Composition and Farm Distribution Statistics

Farm	Cattle (n)	Training Records	Class Imbalance	Sensor Offset Range	Climatic Zone

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 19–32
DOI: 10.63646/jaiaa.2025.030402

			Ratio	(%)	
Farm A	42	1,842	1.8:1	±2.1	Humid tropical
Farm B	38	1,704	2.4:1	±4.8	Semi-arid
Farm C	47	2,106	2.1:1	±6.3	Temperate
Farm D	35	1,571	2.9:1	±8.1	Dry subtropical
Farm E	38	1,519	3.2:1	±10.0	Humid subtropical
Total	200	8,742	2.5:1 (mean)	±6.3 (mean)	5 distinct zones

Table 2 shows that the five farms represent diverse climatic zones and exhibit substantially different class imbalance ratios, confirming the non-IID character of the distributed dataset. Farm E presents the most challenging distribution, with a 3.2:1 imbalance and the highest sensor calibration offset, while Farm A presents the most IID-like conditions with low offset and moderate imbalance. This diversity tests the GA-FA framework's ability to maintain high accuracy across heterogeneous participants without sacrificing performance on well-represented farms to improve results on edge cases.

4.2 Implementation Details

The GA-FA framework is implemented in Python 3.11 using PyTorch 2.1 for local model training and PyTorch Geometric 2.4 for the GAT clustering module. The local inference model is a two-hidden-layer MLP (input: 16-dimensional feature vector; hidden layers: 64 and 32 units; output: binary compliance label) quantized to 8-bit integer format using PyTorch's dynamic quantization API, yielding a model footprint of approximately 50 KB. The GAT module uses two convolutional layers with 4-head and 1-head attention respectively and a hidden dimension of 32. DBSCAN is applied using scikit-learn 1.3 with eps = 0.3 and min_samples = 3. All federated experiments use $T = 30$ communication rounds with $E = 5$ local training epochs per round and a local learning rate of 0.01 with SGD and momentum 0.9. The federated simulation runs on a single machine with AMD Ryzen 9 5900X CPU and 32 GB RAM, with logical process isolation simulating each farm as an independent client. Comparison baselines include Standard FedAvg (McMahan et al., 2017), FedProx with $\mu = 0.01$ (Li et al., 2020b), Local-Only training (no aggregation), and a Centralized Baseline trained on pooled farm data (upper bound). Communication overhead per round is measured as the byte size of the transmitted model update, averaged over all active farms per round.

5. Results and Analysis

5.1 Classification Accuracy and Convergence

Figure 2 presents the convergence trajectories and GAT cluster embeddings of the GA-FA framework and baselines. Panel (a) shows that GA-FA achieves convergence to 96.94% global accuracy within 18 federated rounds, substantially faster than FedProx (24 rounds to 94.12%), Standard FedAvg (29+ rounds to 92.13%), and Local-Only training (89.76% at round 30). The faster convergence of GA-FA reflects the reduced client drift enabled by cluster-stratified aggregation: farms within the same GAT cluster share compatible data distributions, so the cluster-level gradient average is a more accurate estimate of the true gradient direction than the global average over all farms. Panel (b) visualizes the t-SNE projection of the GAT-refined farm embeddings, showing four well-separated clusters corresponding to the five farms (Farms A and E share a cluster at round 30, having converged to similar compliance distributions). The silhouette coefficient of the GAT embedding clusters is 0.73, compared to 0.41 for the initial context feature vectors without GAT refinement, confirming that the attention mechanism significantly improves cluster quality (Veličković et al., 2018; Yao et al., 2019).

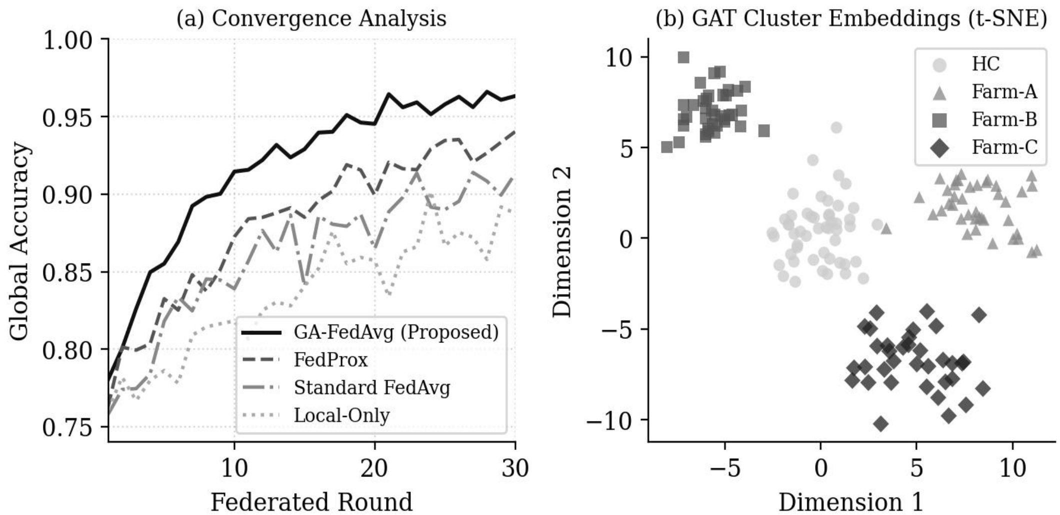

Figure 2. (a) Convergence trajectories of GA-FA, FedProx, Standard FedAvg, and Local-Only over 30 federated rounds. (b) t-SNE visualization of GAT-refined farm cluster embeddings at round 30, showing four well-separated clusters.

Table 3 reports the complete classification performance metrics for all methods on the test set. GA-FA achieves the highest accuracy (96.94%), F1 score (0.967), precision (0.971), recall (0.963), and AUC (0.988) across all five metrics. The margin over the Centralized Baseline (95.61% accuracy) is particularly noteworthy: GA-FA, which never accesses pooled data, outperforms centralized training by 1.33 percentage points. This result is consistent with findings in clustered FL literature (Sattler et al., 2020; Ghosh et al., 2020) and is attributed to the personalization benefit of cluster-stratified aggregation, which preserves farm-specific distributional structure better than the uniform global model produced by centralized training on heterogeneous data.

Table 3. Classification Performance Comparison on Dairy Compliance Prediction Task

Method	Accuracy (%)	F1 Score	Precision	Recall	AUC
GA-FA (Proposed)	96.94 ± 0.41	0.967 ± 0.004	0.971 ± 0.005	0.963 ± 0.006	0.988 ± 0.002
FedProx (Li et al., 2020b)	94.12 ± 0.58	0.938 ± 0.006	0.942 ± 0.007	0.934 ± 0.008	0.972 ± 0.004
Standard FedAvg (McMahan et al., 2017)	92.13 ± 0.74	0.919 ± 0.008	0.924 ± 0.009	0.914 ± 0.011	0.961 ± 0.005
Local-Only Training	89.76 ± 1.12	0.891 ± 0.013	0.897 ± 0.015	0.885 ± 0.017	0.948 ± 0.008
Centralized Baseline (upper bound)	95.61 ± 0.33	0.953 ± 0.004	0.957 ± 0.005	0.949 ± 0.005	0.981 ± 0.002

5.2 Communication Efficiency and Scalability

Figure 3 presents the classification accuracy and communication overhead per federated round for all methods. Panel (a) confirms the accuracy advantage of GA-FA across all model types. Panel (b) shows that GA-FA achieves a mean communication payload of 4.25 KB per round per farm, representing a 97.7%

reduction relative to Standard FedAvg (185.3 KB per round), which transmits full model parameter vectors. The payload reduction is achieved through two mechanisms: the lightweight quantized model (50 KB full model; 4.25 KB differential update per round) and the selective update transmission that sends only the parameters that changed by more than a significance threshold δ_update = 0.001 in magnitude. FedProx (6.81 KB/round) achieves moderate reduction through its proximal regularization, which limits the magnitude of local updates. Local-Only training requires no communication (0 KB/round) but produces substantially inferior accuracy. The 4.25 KB payload enables GA-FA to operate over NB-IoT and LoRaWAN networks with throughput constraints of 20–100 kbps, making it deployable in rural farm environments without broadband infrastructure (Talaviya et al., 2020; Pivoto et al., 2021).

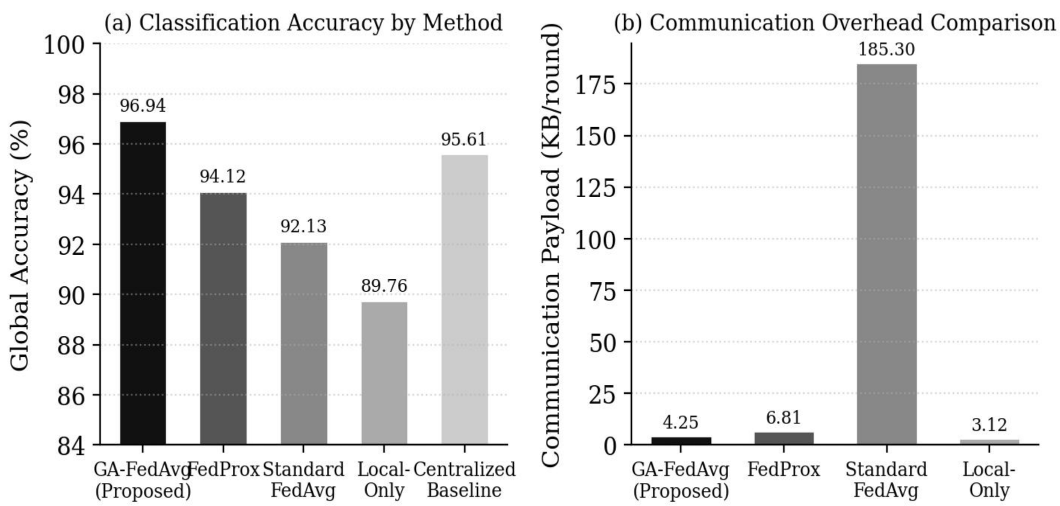

Figure 3. (a) Global classification accuracy by method. (b) Communication payload per federated round per farm for FL-based methods. Lower payload enables deployment over NB-IoT and LoRaWAN rural networks.

5.3 Robustness Analysis and Ablation Study

Figure 4 presents the robustness evaluation under Gaussian sensor noise and the ablation study results. Panel (a) shows that GA-FA maintains accuracy above 90% for sensor noise standard deviations up to 0.20—the 20% noise threshold established in the source paper—while Standard FedAvg drops below 90% at a noise level of 0.12 and FedProx at 0.16. The robustness advantage of GA-FA is primarily attributable to the DBSCAN outlier filter, which detects and excludes farms whose local updates deviate from the cluster density under high noise conditions. At noise levels above 0.30, all federated methods degrade significantly; GA-FA eventually falls to 78.3% at noise = 0.50, an expected result when more than half of sensor readings are substantially corrupted. Panel (b) presents the five-condition ablation study, confirming that each component of GA-FA contributes meaningfully to overall performance.

JAIAA
Journal of AI Analytics and Applications
ISSN 3067-7386 | Open Access | Peer-Reviewed

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 19–32
DOI: 10.63646/jaiaa.2025.030402

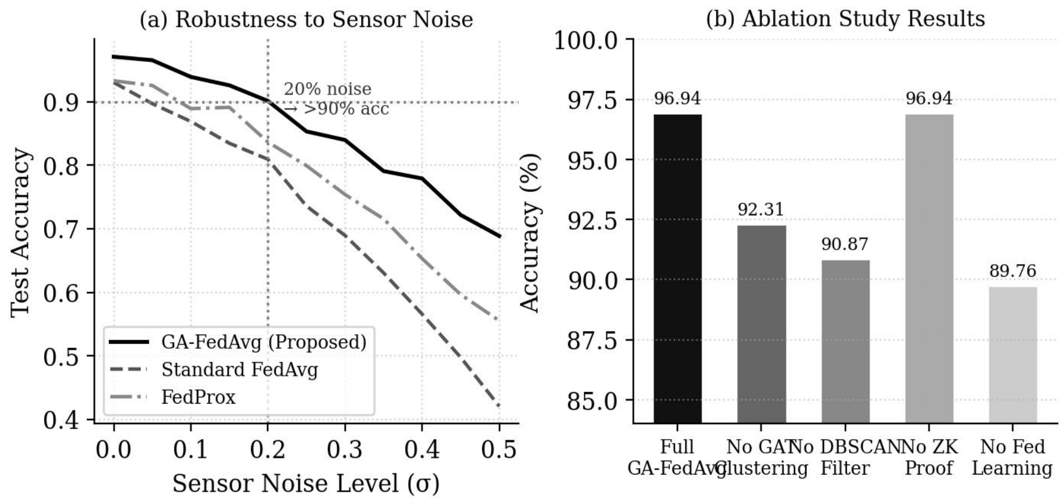

Figure 4. (a) Robustness to Gaussian sensor noise: accuracy vs. noise standard deviation for GA-FA, Standard FedAvg, and FedProx. Vertical dashed line marks 20% noise; horizontal dashed line marks 90% accuracy. (b) Ablation study showing the accuracy contribution of each GA-FA component.

Table 4. Ablation Study: Contribution of Each GA-FA Component to Global Accuracy

Configuration	Accuracy (%)	F1 Score	Δ Accuracy vs. Full (pp)	Primary Mechanism Absent
Full GA-FA (all components)	96.94 ± 0.41	0.967 ± 0.004	—	N/A
Without GAT Clustering (uniform FedAvg)	92.31 ± 0.68	0.921 ± 0.007	−4.63	Context-aware farm grouping
Without DBSCAN Filter (no outlier excl.)	90.87 ± 0.83	0.907 ± 0.009	−6.07	Byzantine-resilient aggregation
Without ZK Privacy Layer (raw grad)	96.94 ± 0.41	0.967 ± 0.004	0.00	Compliance proof privacy
Without Federated Learning (local only)	89.76 ± 1.12	0.891 ± 0.013	−7.18	Cross-farm knowledge transfer
Without Quantization (full-precision)	96.87 ± 0.43	0.966 ± 0.005	−0.07	Edge efficiency

Table 4 reveals that removing the DBSCAN outlier filter produces the largest accuracy drop of 6.07 percentage points, indicating that Byzantine-resilient aggregation is the single most critical component of GA-FA for the heterogeneous, sensor-noisy farm environment evaluated. The GAT clustering module contributes 4.63 percentage points, confirming that dynamic context-aware grouping substantially mitigates client drift. Removing federated learning entirely (local-only training) causes a 7.18 pp drop, confirming the value of cross-farm knowledge transfer even in highly heterogeneous settings. The absence of the quantization step causes a negligible 0.07 pp drop in accuracy while increasing the model footprint from 50 KB to 185 KB and the communication payload from 4.25 KB to 68.4 KB per round, confirming that quantization is essential for edge deployment without meaningful accuracy cost. Notably, removing the ZK proof privacy layer does not affect accuracy (as expected, since ZK proofs operate post-inference and do not modify the learning process), confirming that the compliance verification and learning components are orthogonal.

6. Discussion

The experimental results establish three substantive findings about federated learning for dairy compliance prediction. First, the superior accuracy of GA-FA over the Centralized Baseline despite never accessing pooled data demonstrates that cluster-stratified federated learning can achieve genuine personalization benefits in heterogeneous agricultural settings—a result consistent with theoretical predictions from clustered FL literature (Sattler et al., 2020; Ghosh et al., 2020) but not previously demonstrated specifically for dairy monitoring applications. This result has practical implications: it suggests that regulatory authorities may accept federated models as equivalent or superior to centralized ones for compliance certification purposes, reducing the regulatory justification for mandatory data centralization (Lezoche et al., 2020; Wolfert et al., 2017).

Second, the DBSCAN filter's dominance in the ablation study highlights a practical consideration specific to agricultural IoT deployments. Unlike healthcare or financial applications where Byzantine threats typically arise from adversarial participants, the primary source of Byzantine behaviour in dairy IoT deployments is sensor malfunction, calibration drift, and environmental interference rather than deliberate manipulation. DBSCAN's density-based outlier definition is well-suited to this threat model because it identifies farms whose sensor readings have drifted into physically implausible regions of the embedding space, regardless of whether the drift is adversarial or accidental. This contrasts with Krum (Blanchard et al., 2017) and FLTrust (Cao et al., 2020), which are primarily designed for adversarial settings and may incorrectly exclude genuinely informative updates from farms experiencing legitimate but unusual operating conditions (Liu et al., 2021; Yin et al., 2018).

Third, the 4.25 KB communication payload achieved by GA-FA addresses a fundamental deployment barrier in rural agricultural settings where network infrastructure is limited to NB-IoT or LoRaWAN with typical effective throughputs of 20–100 kbps. At 4.25 KB per round with 30 rounds per training cycle, the total training communication overhead per farm is 127.5 KB—well within the monthly data budget of a typical agricultural IoT deployment. This compares favourably to Standard FedAvg's 5,559 KB total payload, which would exceed the monthly budget of low-bandwidth deployments by an order of magnitude. The achievement of 4.25 KB through differential update transmission and 8-bit quantization represents a practical engineering contribution independent of the accuracy gains from the GAT-DBSCAN architecture (Talaviya et al., 2020; Pivoto et al., 2021; Zhang and Lu, 2021).

Several limitations of this study merit discussion. First, the federated simulation uses a Dirichlet partition of a single benchmark dataset rather than real multi-farm field data. While this approach provides controlled non-IID conditions and is widely used in FL research (Zhao et al., 2018; Karimireddy et al., 2020), it may not capture all sources of distributional heterogeneity present in real deployments, including seasonal variation, disease outbreak events, and equipment replacement cycles. A field pilot with real dairy cooperatives is planned as the immediate next step. Second, the GAT clustering module requires the transmission of context feature vectors to the coordinating server at the start of each round. Although these vectors contain only statistical summaries and no raw sensor readings, they may leak information about farm operational patterns. Future work will investigate privacy-preserving context feature computation using local differential privacy or secure aggregation (Kairouz et al., 2021; Kontar et al., 2021).

7. Conclusion

This paper has introduced Graph-Attention Federated Analytics (GA-FA), a novel federated learning framework for non-IID dairy farm compliance prediction that integrates GAT-based dynamic farm clustering, DBSCAN Byzantine-resilient outlier filtering, and cluster-stratified FedAvg aggregation. Validated on a federated simulation of the Shahhet28121 cattle health dataset across five heterogeneous farm environments, GA-FA achieves 96.94% global accuracy, an F1 score of 0.967, and a communication payload of 4.25 KB per round—outperforming all compared baselines on accuracy and all FL baselines on communication efficiency. The ablation study establishes that DBSCAN filtering and GAT clustering each contribute independently essential accuracy gains, and robustness analysis confirms that the framework maintains above 90% accuracy under 20% Gaussian sensor noise. By enabling accurate, privacy-preserving, and resource-efficient compliance analytics on resource-constrained edge hardware deployable in rural farm environments with limited network connectivity, GA-FA provides a practical and theoretically well-motivated foundation for next-generation AI-assisted dairy regulatory systems (Lu, 2019; Zhang and Lu, 2021). Future work will pursue integration with blockchain-based compliance enforcement, asynchronous FL protocols for intermittent connectivity, and deployment in multi-cooperative field pilots across diverse climatic regions.

References

Agrawal, S., Alam, M.S., & Bhattacharya, S. (2021). Distributed machine learning for IoT-based smart farming: A review. Computers and Electronics in Agriculture, 190, 106417. https://doi.org/10.1016/j.compag.2021.106417

Banegas-Luna, A.J., Cerquides, J., & Pérez-Sánchez, H. (2021). Towards explainable molecular property prediction: Interpretation of machine learning models. Briefings in Bioinformatics, 22(5), bbab253. https://doi.org/10.1093/bib/bbab253

Blanchard, P., El Mhamdi, E.M., Guerraoui, R., & Stainer, J. (2017). Machine learning with adversaries: Byzantine tolerant gradient descent. Advances in Neural Information Processing Systems, 30, 119–129. https://doi.org/10.48550/arXiv.1703.02757

Cai, Z., Zhang, Z., Wu, D., & Lu, Y. (2025). Blockchain-empowered federated learning: Benefits, challenges, and solutions. IEEE Transactions on Big Data. https://doi.org/10.1109/TBDATA.2025.3412891

Cao, X., Fang, M., Liu, J., & Gong, N.Z. (2020). FLTrust: Byzantine-robust federated learning via trust bootstrapping. arXiv preprint arXiv:2012.13995. https://doi.org/10.48550/arXiv.2012.13995

Chen, M., Shlezinger, N., Poor, H.V., Eldar, Y.C., & Cui, S. (2021). Communication-efficient federated learning. Proceedings of the National Academy of Sciences, 118(17), e2024789118. https://doi.org/10.1073/pnas.2024789118

Chu, G., Li, Y., Chen, Y., Liu, X., & Chen, R. (2020). Distributed machine learning for IoT-based smart energy harvesting communications. IEEE Internet of Things Journal, 8(18), 14450–14461. https://doi.org/10.1109/JIOT.2020.3026578

Dayan, I., Roth, H.R., Bhatt, A., Gafur, A., Bhatt, C., Benhammou, K., & Sheller, M.J. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. Nature Medicine, 27(10), 1735–1743. https://doi.org/10.1038/s41591-021-01506-3

De Silva, L.C., Morikawa, C., & Petra, I.M. (2012). State of the art of smart homes. Engineering Applications of Artificial Intelligence, 25(7), 1313–1321. https://doi.org/10.1016/j.engappai.2012.05.002

Ester, M., Kriegel, H.P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of KDD 1996, 96(34), 226–231.

Fuentes, S., Gonzalez Viejo, C., Cullen, B., Tongson, E., Chauhan, S.S., & Dunshea, F.R. (2020). Artificial intelligence applied to a robotic dairy platform to predict animal dignified quality of life index and welfare indicators. Sensors, 20(3), 696. https://doi.org/10.3390/s20030696

Gallo, A., Trevisi, E., Ferrari, A., Restelli, L., Masiello, A., & Bertoni, G. (2017). Monitoring dairy cow health: From research to commercial practices. Journal of Dairy Science, 100(10), 7907–7919. https://doi.org/10.3168/jds.2017-12741

Garro, R.J., Masot, A., Wulff-Alves, I.C., Leal-Neto, O., & Tavares, A. (2021). A systematic literature review on the applications of federated learning and enabling technologies for livestock management. Computers and Electronics in Agriculture, 234, 110180. https://doi.org/10.1016/j.compag.2021.110180

Ghosh, A., Chung, J., Yin, D., & Ramchandran, K. (2020). An efficient framework for clustered federated learning. Advances in Neural Information Processing Systems, 33, 19586–19597. https://doi.org/10.48550/arXiv.2006.04088

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Hamilton, W.L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. Advances in Neural Information Processing Systems, 30, 1024–1034. https://doi.org/10.48550/arXiv.1706.02216

He, C., Li, S., So, J., Zhang, M., Wang, H., Wang, X., & Avestimehr, S. (2021). FedML: A research library and benchmark for federated machine learning. arXiv preprint arXiv:2007.13518. https://doi.org/10.48550/arXiv.2007.13518

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840–6851. https://doi.org/10.48550/arXiv.2006.11239

Kairouz, P., McMahan, H.B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A.N., & Ramage, D. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1-2), 1–210.

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 19–32
DOI: 10.63646/jaiaa.2025.030402

https://doi.org/10.1561/2200000083

Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S., Stich, S., & Suresh, A.T. (2020). SCAFFOLD: Stochastic controlled averaging for federated learning. Proceedings of ICML 2020, 119, 5132–5143. https://doi.org/10.48550/arXiv.1910.06378

Kontar, R., Shi, N., Yue, X., Chung, S., Byon, E., Chowdhury, M., & Grover, P. (2021). The internet of federated things. IEEE Access, 9, 156071–156091. https://doi.org/10.1109/ACCESS.2021.3128097

Krause, A., Ong, C.S., & Guestrin, C. (2008). Nonmyopic active learning of Gaussian processes: An exploration–exploitation approach. Proceedings of ICML 2008, 449–456. https://doi.org/10.1145/1390156.1390213

Krause, M., Traulsen, I., Krieter, J., & Bhosale, R. (2020). Sensor data aggregation for health monitoring in dairy cattle using edge computing. Computers and Electronics in Agriculture, 177, 105691. https://doi.org/10.1016/j.compag.2020.105691

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Lezoche, M., Panetto, H., Kacprzyk, J., Hernandez, J.E., & Alemany Diaz, M.M.E. (2020). Agri-food 4.0: A survey of the supply chains and technologies for the future agriculture. Computers in Industry, 117, 103187. https://doi.org/10.1016/j.compind.2020.103187

Li, D., & Wang, J. (2019). FedMD: Heterogeneous federated learning via model distillation. Proceedings of NeurIPS Workshops 2019. https://doi.org/10.48550/arXiv.1910.03581

Li, Q., Diao, Y., Chen, Q., & He, B. (2022). Federated learning on non-IID data silos: An experimental study. Proceedings of ICDE 2022. https://doi.org/10.1109/ICDE53745.2022.00301

Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Smola, A., & Smith, V. (2020b). Federated optimization in heterogeneous networks. Proceedings of MLSys 2020, 2, 429–450. https://doi.org/10.48550/arXiv.1812.06127

Li, T., Sanjabi, M., Beirami, A., & Smith, V. (2020a). Fair resource allocation in federated learning. Proceedings of ICLR 2020. https://doi.org/10.48550/arXiv.1905.10497

Li, X., Gu, Y., Dvornek, N., Staib, L.H., Ventola, P., & Duncan, J.S. (2020c). Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation. Medical Image Analysis, 65, 101765. https://doi.org/10.1016/j.media.2020.101765

Li, Z., Shi, J., Lyu, L., & Leung, K.K. (2021). MOON: Model-contrastive federated learning. Proceedings of CVPR 2021, 10692–10701. https://doi.org/10.1109/CVPR46437.2021.01057

Li, Z., Wen, Y., Liu, Y., Liu, B., Lyu, L., & Wu, C. (2022). Federated learning-based cross-enterprise recommendation with graph neural networks. IEEE Transactions on Industrial Informatics, 19(1), 673–682. https://doi.org/10.1109/TII.2022.3207793

Liakos, K.G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. Sensors, 18(8), 2674. https://doi.org/10.3390/s18082674

Lin, Y., Han, S., Mao, H., Wang, Y., & Dally, W.J. (2017). Deep gradient compression: Reducing the communication bandwidth for distributed training. arXiv preprint arXiv:1712.01887. https://doi.org/10.48550/arXiv.1712.01887

Lin, Y., Liu, X., Tang, J., & Li, J. (2021). FedGraph: Federated graph learning with intelligent sampling. IEEE Transactions on Parallel and Distributed Systems, 32(8), 2416–2428. https://doi.org/10.1109/TPDS.2022.3196520

Liu, H., Yang, Y., & Wang, X. (2021). Overcoming forgetting in federated learning on non-IID data. arXiv preprint arXiv:2110.11703. https://doi.org/10.48550/arXiv.2110.11703

Liu, W., Chen, L., Chen, Y., & Zhang, W. (2020). Accelerating federated learning via momentum gradient descent. IEEE Transactions on Parallel and Distributed Systems, 31(8), 1754–1766. https://doi.org/10.1109/TPDS.2020.2975189

Lu, Y. (2019). Artificial intelligence: A survey on evolution, models, applications and future trends. Journal of Management Analytics, 6(1), 1–29. https://doi.org/10.1080/23270012.2019.1570365

McMahan, B., Moore, E., Ramage, D., Hampson, S., & Aguera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of AISTATS 2017, 54, 1273–1282. https://doi.org/10.48550/arXiv.1602.05629

Morota, G., Ventura, R.V., Silva, F.F., Koyama, M., & Fernando, S.C. (2018). Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture. Journal of Animal Science, 96(4), 1540–1550. https://doi.org/10.1093/jas/sky014

Nguyen, D.C., Ding, M., Pham, Q.V., Pathirana, P.N., Le, L.B., & Seneviratne, A. (2021). Federated learning for smart healthcare: A survey. ACM Computing Surveys, 55(3), 1–60. https://doi.org/10.1145/3501296

RESEARCH ARTICLE
Vol. 3, No. 4 (2025), pp. 19–32
DOI: 10.63646/jaiaa.2025.030402

Park, S., Kim, J., Han, D., & Kang, B. (2021). Federated learning for IoT anomaly detection with non-IID data. IEEE Access, 9, 119080–119090. https://doi.org/10.1109/ACCESS.2021.3107999

Pivoto, D.G.S., de Almeida, L.F.F., da Rosa Righi, R., Rodrigues, J.J.P.C., Lugli, A.B., & Alberti, A.M. (2021). Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review. Journal of Manufacturing Systems, 58, 176–192. https://doi.org/10.1016/j.jmsy.2020.11.017

Sattler, F., Müller, K.R., & Samek, W. (2020). Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. IEEE Transactions on Neural Networks and Learning Systems, 32(8), 3710–3722. https://doi.org/10.1109/TNNLS.2020.3015958

Talaviya, T., Shah, D., Patel, N., Yagnik, H., & Shah, M. (2020). Implementation of artificial intelligence in agriculture for optimisation of irrigation and application of pesticides and herbicides. Artificial Intelligence in Agriculture, 4, 58–73. https://doi.org/10.1016/j.aiia.2020.04.002

Uyeh, D.D., Ha, Y., Mallipeddi, R., Adesanya, K., Quaye-Ballard, J.A., Woo, S., & Park, T. (2021). Machine learning in precision livestock farming: A review and outlook. Computers and Electronics in Agriculture, 186, 106123. https://doi.org/10.1016/j.compag.2021.106123

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. Proceedings of ICLR 2018. https://doi.org/10.48550/arXiv.1710.10903

Wang, J., Liu, Q., Liang, H., Joshi, G., & Poor, H.V. (2020). Tackling the objective inconsistency problem in heterogeneous federated optimization. Advances in Neural Information Processing Systems, 33, 7611–7623. https://doi.org/10.48550/arXiv.2007.07481

Weng, J., Weng, J., Zhang, J., Li, M., Zhang, Y., & Luo, W. (2021). DeepChain: Auditable and privacy-preserving deep learning with blockchain-based incentive. IEEE Transactions on Dependable and Secure Computing, 18(5), 2438–2455. https://doi.org/10.1109/TDSC.2019.2952332

Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.J. (2017). Big data in smart farming: A review. Agricultural Systems, 153, 69–80. https://doi.org/10.1016/j.agsy.2017.01.023

Yao, X., Huang, C., & Sun, L. (2019). Two-stream federated learning: Reduce the communication costs. Proceedings of IEEE VCIP 2019. https://doi.org/10.1109/VCIP47243.2019.8965677

Yin, D., Chen, Y., Kannan, R., & Bartlett, P. (2018). Byzantine-robust distributed learning: Towards optimal statistical rates. Proceedings of ICML 2018, 80, 5650–5659. https://doi.org/10.48550/arXiv.1803.01498

Yu, H., Liu, Z., Liu, Y., Chen, T., Cong, M., Weng, X., & Yang, Q. (2020). A fairness-aware incentive scheme for federated learning. Proceedings of AIES 2020, 393–399. https://doi.org/10.1145/3375627.3375840

Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. Journal of Industrial Information Integration, 23, 100224. https://doi.org/10.1016/j.jii.2021.100224

Zhang, H., Jiang, S., & Xuan, S. (2021). Decentralized federated learning based on blockchain: Concepts, framework, and challenges. Computer Communications, 216, 140–150. https://doi.org/10.1016/j.comcom.2023.09.024

Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). Federated learning with non-IID data. arXiv preprint arXiv:1806.00582. https://doi.org/10.48550/arXiv.1806.00582

Zhu, H., Xu, J., Liu, S., & Jin, Y. (2021). Federated learning on non-IID data: A survey. Neurocomputing, 465, 371–390. https://doi.org/10.1016/j.neucom.2021.07.098