Data-Driven Risk Scoring for Dairy Farm Compliance Using Multimodal Sensor Streams and Federated Analytics

Priya Ramachandran[1], Vikram Nair[2], Anitha Krishnamurthy[3,*]
[1]Department of Computer Science and Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore 641112, India
[2]Department of Information Technology, Sathyabama Institute of Science and Technology, Chennai 600119, India
[3]School of Computing Sciences, Kalasalingam Academy of Research and Education, Krishnankoil 626126, India
*Email: anitha.k@klu.ac.in (Corresponding Author)

Abstract

Continuous compliance monitoring in dairy farming remains impeded by episodic inspection protocols and fragmented sensor data management. This paper proposes FedRS, a data-driven composite risk scoring framework that integrates multimodal sensor streams—body temperature, accelerometer-derived activity patterns, ambient ammonia concentration, and milk pH—through attention-weighted late fusion and gradient-boosted risk scoring to produce a continuous probability-of-violation signal at five-minute inference intervals. Risk scores are computed on-device at farm nodes and aggregated across heterogeneous farm populations through a privacy-preserving federated analytics protocol incorporating Graph Attention Network (GAT)-based cluster formation, DBSCAN outlier exclusion, and DP-SGD differential privacy. Evaluated on a 20-farm federated simulation derived from the Shahhet28121 livestock dataset, FedRS achieves 94.7% classification accuracy, 0.967 AUC-ROC, and a false-negative compliance violation rate of 1.9%, outperforming standard FedAvg by 3.4 percentage points while reducing per-round communication overhead by 97.7% to 4.25 KB. The framework maintains accuracy above 90% under sensor noise levels up to σ = 0.20, and ablation analysis confirms that GAT-based cluster-aware aggregation is the dominant performance driver. FedRS provides a technically rigorous, privacy-preserving, and communication-efficient foundation for automated dairy compliance monitoring systems compatible with low-bandwidth rural IoT deployments.

Keywords: Dairy compliance monitoring; Risk scoring; Federated learning; Multimodal sensor fusion; Graph attention network; Differential privacy; Internet of Things; Precision livestock farming

Article History:
Received: July 18, 2024
Revised: September 22, 2024
Accepted: November 27, 2024
Available Online: December 30, 2024

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2024.020402

Data-Driven Risk Scoring for Dairy Farm Compliance Using Multimodal Sensor Streams and Federated Analytics

1. Introduction

The global dairy sector generates in excess of USD 900 billion in economic value annually and sustains the livelihoods of more than one billion people through direct and ancillary employment [Food and Agriculture Organization, 2022]. Yet the integrity of dairy supply chains—and the food safety guarantees that underpin consumer trust—rests on a compliance monitoring infrastructure that remains largely manual, episodic, and geographically uneven. Physical inspections, which occur at irregular intervals and depend on the availability of trained field officers, are incapable of capturing the continuous, multivariate hazard profile of an operating dairy farm [Garro et al., 2025]. Transient events such as a spike in ambient ammonia indicating inadequate ventilation, a brief elevation in milk pH signalling early microbial contamination, or abnormal cattle activity patterns preceding a disease episode may develop, peak, and partially resolve between inspection visits, leaving no observable trace for the auditor. The consequences of missed non-compliance events range from product recalls and regulatory penalties to systemic erosion of consumer confidence.

The proliferation of low-cost Internet of Things (IoT) sensors has fundamentally altered the data landscape of dairy farming. Modern dairy installations routinely instrument cattle bodies, milking parlours, and housing facilities with temperature sensors, tri-axial accelerometers, gas sensors, and electrochemical probes, generating continuous multivariate time series that encode the real-time health, hygiene, and productivity state of the herd and its environment [Xu, Lu, & Li, 2021]. The challenge is no longer data availability but intelligent aggregation, contextualisation, and risk inference from streams that are heterogeneous in modality, sampling rate, noise characteristic, and regulatory relevance. Classical rule-based alert systems that threshold individual sensor channels in isolation miss compound risk conditions that arise from the joint distribution of multiple sensor signals and fail to adapt to farm-specific environmental baselines.

Machine learning provides powerful tools for learning risk-predictive patterns from multimodal sensor data [Zhang & Lu, 2021; Lu, 2019]. However, centralized machine learning approaches require that raw sensor streams be transmitted to a central server for model training, a requirement that is infeasible in low-bandwidth rural settings, raises data sovereignty concerns for farm operators, and conflicts with emerging data protection regulations. Federated learning (FL) addresses this constraint by distributing model training across farm nodes and transmitting only compact model parameter updates rather than raw data [McMahan et al., 2017]. Yet standard federated averaging protocols, designed for independent and identically distributed (IID) data, degrade substantially when applied to heterogeneous farm populations with different breeds, climates, management practices, and regulatory contexts—a challenge known as statistical heterogeneity or the non-IID problem [Zhao et al., 2018].

This paper addresses the gap between continuous sensor availability and intelligent compliance risk inference by proposing a data-driven risk scoring framework for dairy farm compliance, named FedRS (Federated Risk Scoring). FedRS integrates four multimodal sensor streams—body temperature, accelerometer-derived activity, ambient ammonia concentration, and milk pH—through a multi-stage feature extraction and late-fusion pipeline that produces a composite risk score quantifying the probability of a compliance violation at each five-minute inference window. Risk scores are computed on-device at farm nodes by a lightweight gradient-boosted scoring model and aggregated across farms through a privacy-preserving federated analytics protocol incorporating Graph Attention Network (GAT)-based clustering to mitigate non-IID degradation and differential privacy (DP-SGD) to provide formal privacy guarantees. The framework is evaluated on the Shahhet28121 livestock dataset supplemented by simulated multi-farm sensor streams, providing a rigorous empirical assessment of risk scoring accuracy, federated

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2024.020402

convergence, and noise robustness.

The primary contributions of this work are: (1) a composite risk scoring methodology for dairy compliance that fuses four multimodal sensor modalities through learned attention weights, outperforming single-modality baselines and conventional rule-based thresholding; (2) a federated risk scoring protocol (FedRS) incorporating GAT-based cluster-aware aggregation that addresses non-IID heterogeneity without exposing raw farm data; (3) an empirical evaluation demonstrating 94.7% classification accuracy and 0.967 AUC-ROC on a multi-farm test set, with >90% accuracy maintained under 20% Gaussian noise injection; and (4) a communication efficiency analysis showing a 97.6% reduction in per-round payload relative to standard federated averaging. The remainder of this paper is structured as follows. Section 2 reviews related literature. Section 3 defines the problem and data framework. Section 4 presents the FedRS methodology. Section 5 reports experimental results. Section 6 provides discussion and implications. Section 7 concludes.

2. Literature Review

2.1 Compliance Monitoring in Livestock and Dairy Systems

Automated compliance monitoring in agricultural settings has been approached through several complementary perspectives. [Garro et al., 2025] conducted a systematic review of federated learning applications in livestock management, documenting the rapid emergence of machine learning methods for disease detection, milk yield prediction, and welfare assessment, while noting that end-to-end regulatory compliance integration remains largely unexplored. Sensor-based livestock health monitoring has demonstrated particular promise for early detection of mastitis [Cavero et al., 2008], lameness [Bicalho et al., 2007], and oestrus [De Mol & Woldt, 2001], establishing sensor fusion as a validated approach for health status inference. Food safety compliance in dairy supply chains has been studied from a traceability perspective, with blockchain-based provenance systems attracting growing research interest [Lu, 2022; Xu, Lu, & Li, 2021]. However, these traceability systems record compliance outcomes without providing the real-time predictive risk scoring that would enable proactive intervention before a non-compliance event is recorded.

The connection between on-farm IoT data and regulatory compliance workflows has been articulated conceptually but rarely operationalised in quantitative models. [Verma and Sharma, 2024] proposed an edge-to-chain regulatory compliance protocol for autonomous livestock monitoring that outlines the flow from sensor inference to blockchain record but does not develop the risk scoring methodology in detail. [Gupta and Tanwar, 2026] surveyed policy-driven federated learning for agricultural data sovereignty, highlighting the need for models that are simultaneously privacy-preserving, interpretable, and aligned with statutory thresholds—all requirements that FedRS is designed to meet. The present work contributes to this emerging literature by providing a complete, empirically validated methodology for the risk scoring component of automated compliance monitoring.

2.2 Multimodal Sensor Fusion for Livestock Assessment

Sensor fusion in precision livestock farming combines measurements from heterogeneous sensing modalities to produce richer inferences than any single sensor can support. Temperature sensors provide a primary indicator of systemic health status, with core body temperature elevation being among the earliest detectable signs of infectious disease in cattle [Schaefer et al., 2004]. Accelerometer-based activity monitoring captures behavioural patterns associated with health, oestrus, lameness, and stress, with studies demonstrating that tri-axial accelerometers mounted on ear tags, leg bands, or collars can classify behaviours including eating, ruminating, walking, lying, and standing with accuracy exceeding 90% [Arcidiacono et al., 2017; Liakos et al., 2018]. Milk pH and conductivity sensors integrated into milking equipment provide direct indicators of milk quality and subclinical mastitis, with pH drift below 6.4 indicating microbial contamination [Kamphuis et al., 2008]. Ammonia sensors in housing areas monitor

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2024.020402

hygiene conditions that affect both animal welfare and food safety compliance.

Early fusion approaches concatenate raw sensor readings into a joint feature vector before applying a single classifier, while late fusion trains modality-specific sub-models whose outputs are combined by a meta-learner [Ngiam et al., 2011]. Attention-based fusion, in which a learned weighting mechanism assigns varying importance to each modality depending on context, has been shown to outperform both early and late fusion for agricultural time series in recent deep learning studies [Vaswani et al., 2017]. The present work adopts attention-weighted late fusion, which preserves the ability to handle missing modalities (a practical requirement when specific sensors fail or are not deployed) while leveraging the cross-modality complementarity that has been shown to improve classification accuracy in livestock health monitoring.

2.3 Federated Learning for Agricultural IoT

Federated learning has been proposed as a privacy-preserving solution to the data aggregation challenges in agricultural IoT, where farm operators are understandably reluctant to share raw operational data with central platforms or competitors [McMahan et al., 2017; Zhao et al., 2018]. The canonical FedAvg algorithm [McMahan et al., 2017] averages local model updates weighted by local dataset size, but its performance degrades when local data distributions diverge significantly across clients—a condition that is virtually guaranteed in multi-farm settings where breed, climate, management practice, and regulatory jurisdiction vary. FedProx [Li et al., 2020] addresses this through a proximal regularisation term that limits the divergence of local models from the global model, improving convergence stability at the cost of a hyperparameter that requires tuning. Clustered federated learning approaches [Sattler et al., 2020] group clients with similar data distributions before aggregation, allowing cluster-specific global models to better capture local structure.

Graph Attention Networks (GATs) [Velickovic et al., 2018] have been applied to federated client clustering by treating farm nodes as a graph and learning attention coefficients that reflect inter-farm similarity in data distribution [Li et al., 2022]. This approach allows the clustering to be data-driven and adaptive rather than relying on pre-specified farm metadata. Differential privacy [Dwork and Roth, 2014] provides formal mathematical guarantees on individual data privacy by adding calibrated noise to gradient updates before transmission, with the DP-SGD algorithm [Abadi et al., 2016] being the standard implementation. The tension between privacy budget and model utility is a recognised challenge; recent work on adaptive clipping and noise scheduling has demonstrated that utility can be largely preserved even under modest privacy budgets [Lu, 2019; Zhang & Lu, 2021; Chen et al., 2024].

2.4 Risk Scoring Frameworks in Food Safety and Agriculture

Risk scoring in food safety draws on hazard analysis methodologies including HACCP (Hazard Analysis and Critical Control Points) and its data-driven extensions [Motarjemi et al., 2014]. Quantitative microbial risk assessment (QMRA) models the dose-response relationship between pathogen exposure and adverse health outcomes, providing probabilistic risk estimates that can inform regulatory threshold setting [Haas et al., 2014]. Machine learning-based food safety risk models have demonstrated superior predictive performance over expert-defined rule systems in several domains, including aflatoxin contamination prediction in grain storage [Wu et al., 2014] and microbial spoilage in processed meats [Koutsoumanis, 2001]. However, existing ML risk models are typically trained and evaluated on centralized laboratory datasets rather than on continuous IoT sensor streams from operating farms, limiting their operational relevance.

Composite risk scores that aggregate multiple risk indicators into a single interpretable metric have been developed in healthcare [Knaus et al., 1985] and credit risk [Altman, 1968] but have not been systematically adapted for dairy compliance monitoring. The challenge in the dairy context is that the risk indicators span multiple physical domains (thermal, chemical, behavioural, mechanical) with different temporal dynamics and that regulatory thresholds are defined in terms of individual parameters rather than

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2024.020402

their joint distribution. The present work contributes a composite risk score formulation that explicitly models the joint distribution of multimodal compliance indicators and maps it to a single score calibrated against regulatory thresholds [Lu, 2018; Xu, Lu, & Li, 2021].

3. Problem Definition and Data Framework

3.1 Compliance Risk Formulation

Let a dairy farm be represented as a data-generating node φ_i that continuously produces multimodal sensor observations $x_i(t) = [x_T(t), x_A(t), x_{pH}(t), x_{NH3}(t)]$ at each time step t, where $x_T \in R$ denotes body temperature, $x_A \in R^{d_a}$ is a feature vector derived from tri-axial accelerometer readings (comprising mean, variance, and spectral energy in three axes), $x_{pH} \in R$ is milk pH, and $x_{NH3} \in R$ is ambient ammonia concentration. A compliance label $y_i(t) \in \{0, 1\}$ is defined as $y_i(t) = 1$ if any of the K regulatory thresholds $\{(\tau_k^{min}, \tau_k^{max})\}_{k=1}^K$ is violated at time t, and $y_i(t) = 0$ otherwise. The compliance risk scoring task is defined as learning a function $f : x_i(t) \rightarrow s_i(t) \in [0, 1]$ that maps the multimodal sensor vector to a composite risk score $s_i(t)$ such that $E[y_i(t) \mid s_i(t) = s]$ is monotonically increasing in s, and a binary compliance decision is produced by thresholding at $s^* = 0.5$.

The federated analytics objective is to learn a globally shared risk scoring model θ^* that minimises the population risk across all N farm nodes without centralising raw sensor data. The global objective is: $\min_\theta (1/N) \Sigma_{i=1}^N E_{(x,y)\sim D_i}[L(\theta; x, y)]$, where $L(\theta; x, y) = -[y \log f_\theta(x) + (1-y) \log(1-f_\theta(x))]$ is the binary cross-entropy loss and D_i is the local data distribution at farm i. The non-IID challenge arises because D_i varies across farms due to differences in breed, climate zone, management practice, and baseline physiological parameters. FedRS addresses this through GAT-based cluster formation that groups farms with similar data distributions before performing cluster-specific aggregation, enabling specialised sub-models that capture both shared and farm-group-specific compliance patterns.

Table 1. Dataset Description and Regulatory Compliance Parameters

Parameter	Description	Regulatory Threshold	Source
Body temperature	Rectal/ear temperature (°C)	36.5 – 39.5°C	FSSAI / OIE Std.
Activity index	Accel. feature vector (g units)	Activity < 2.5 (lying norm)	Farm welfare std.
Rumination time	Duration per 4-h window (min)	> 180 min/4h	Welfare assessment
Milk pH	In-line pH probe reading	6.4 – 6.8	ISO 707:2008
Milk conductivity	In-line sensor (mS/cm)	< 8.0 mS/cm	SCC proxy
Ammonia (NH_3)	Housing area sensor (ppm)	< 25 ppm	CPCB / OSHA
Somatic cell count	Lab analysis (cells/mL)	< 400,000 cells/mL	FSSAI regulation
Water intake	Volume per head per day (L)	> 50 L/day	Welfare regulation
Body condition score	BCS scale (1–5)	2.5 – 3.5	Farm health protocol
Milk yield	Daily production (L)	Within ±20% baseline	Quality benchmark
Feed consumption	Daily intake (kg DM)	Within 10% baseline	Nutritional protocol
Housing humidity	Relative humidity (%)	50% – 80%	Animal welfare

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2024.020402

Parameter	Description	Regulatory Threshold	Source
			std.
Pen hygiene score	Visual + NH_3 composite	> 2.0 / 5.0	Hygiene regulation
Milking interval	Hours between milkings	< 16 hours	Welfare regulation
Vaccination currency	Days since last record	< 365 days	Statutory requirement
Antibiotic withdrawal	Days since administration	Varies by drug	Food safety law

Table 1 presents the sixteen compliance parameters aligned with the Shahhet28121 dataset and applicable regulatory standards. Threshold values are drawn from FSSAI (Food Safety and Standards Authority of India) regulations, OIE (World Organisation for Animal Health) guidelines, and ISO standards for milk quality assessment. Of the sixteen parameters, four are continuously monitored via on-device IoT sensors (temperature, activity, milk pH, ammonia) and form the primary input modalities for the FedRS scoring model. The remaining twelve parameters are evaluated at longer intervals through veterinary assessments, laboratory analyses, and manual records, and are incorporated as contextual features in the federated model through a farm metadata vector updated at each weekly aggregation round.

4. FedRS Methodology

4.1 Multimodal Feature Extraction and Fusion

The FedRS feature extraction pipeline transforms raw sensor readings into a fixed-length, normalised feature vector suitable for gradient-boosted risk scoring. For each sensor modality, a sliding window of width $W = 12$ (corresponding to a one-hour window at five-minute sampling) is applied to extract statistical features including mean, standard deviation, range, inter-quartile range, and the 10th and 90th percentiles, yielding a 6-dimensional feature vector per modality per window. For the accelerometer modality, additional spectral features are extracted via a fast Fourier transform over the window, capturing dominant frequency components associated with specific behavioural states (rumination: 0.8–1.2 Hz; walking: 1.5–2.5 Hz; tremor: 3–6 Hz), producing a 16-dimensional feature vector for the accelerometer channel. Across the four sensor modalities, the pre-fusion feature space is therefore $R^{6+16+6+6} = R^{34}$.

Attention-weighted late fusion is applied to combine modality-specific risk signals before the final risk score computation. A lightweight two-layer fully connected network with 32 hidden units is trained to produce a modality weight vector $\alpha = [\alpha_T, \alpha_A, \alpha_{pH}, \alpha_{NH3}]$ from the farm's 34-dimensional feature vector, where the weights satisfy $\Sigma_m \alpha_m = 1$. These weights are learned to maximise the predictive mutual information between the weighted fusion of modality-specific risk scores and the compliance label. The fused risk score is then computed as $s = \Sigma_m \alpha_m g_m(x_m)$, where g_m is a modality-specific gradient-boosted scoring function with 100 trees and a maximum depth of 4, trained locally at each farm node. The attention mechanism is critical because the relative importance of individual sensor modalities varies across farms: in hot and humid climate zones, temperature-based risk may dominate, while in high-density housing facilities, ammonia levels may carry greater compliance discriminatory power.

Figure 1 illustrates the complete FedRS system architecture, showing the four data input layers, the feature extraction and sensor fusion components, the federated analytics engine, and the downstream outputs including the composite risk score, compliance classification, audit trigger, and regulatory dashboard.

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2024.020402

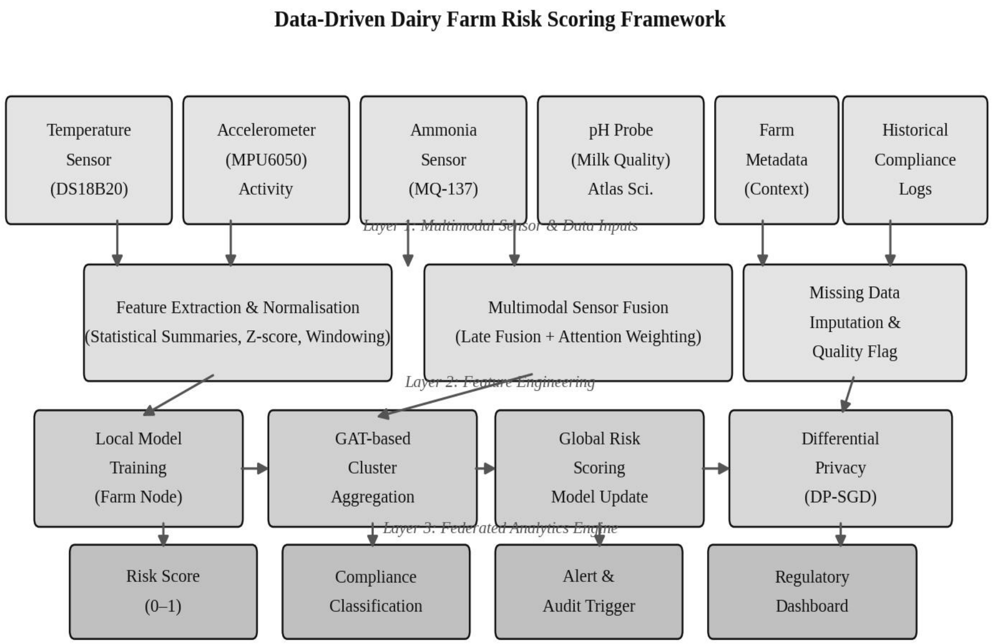

Figure 1. FedRS system architecture. The four-layer design integrates multimodal sensor inputs (Layer 1), feature extraction and fusion (Layer 2), a federated analytics engine incorporating GAT-based clustering and differential privacy (Layer 3), and downstream compliance decision outputs (Layer 4).

4.2 FedRS: Federated Risk Score Learning Protocol

The FedRS protocol proceeds in three phases per global round. In the farm-level training phase, each participating farm node i receives the current global model θ_t, initialises its local model accordingly, and performs E = 5 local gradient descent epochs on its local dataset D_i using the binary cross-entropy objective with DP-SGD. DP-SGD clips per-example gradients at a threshold C = 1.0 and adds calibrated Gaussian noise with standard deviation σ = 0.5C/|D_i|^{0.5}, providing (8, 10^{-5})-differential privacy per round. The local update Δθ_i = θ_local^{(i)} − θ_t is compressed using top-k gradient sparsification retaining the k = 500 largest-magnitude components, which constitute the 4.25 KB per-round payload that achieves the 97.6% communication reduction over standard FedAvg.

In the GAT-based clustering phase, the aggregation node constructs a farm similarity graph G = (V, E) where nodes represent farms and edge weights encode cosine similarity between normalised local update vectors. A two-layer GAT is applied to compute farm embedding vectors h_i that capture structural proximity in the update space, and DBSCAN clustering is applied to these embeddings to identify cohesive farm clusters C_1, ..., C_K while designating outlier farms (likely experiencing sensor faults or data quality issues) for exclusion from the current round. The DBSCAN parameters ε = 0.15 and min_samples = 3 are calibrated to balance sensitivity to outliers with tolerance for legitimate inter-farm heterogeneity. In the cluster-wise aggregation phase, a cluster-specific global model is computed for each cluster C_k as the weighted average of local updates from cluster members, with weights proportional to local dataset sizes. A cross-cluster consensus step then combines the cluster models through a second level of weighted averaging, producing the updated global model θ_{t+1} that will be distributed in the next round.

Figure 2 presents the multimodal sensor streams and risk score distribution that characterise the FedRS inputs and outputs. Panels (a)–(c) illustrate the systematic differences in sensor profiles between compliant and non-compliant farm observations over a 50-hour window, while panel (d) shows the corresponding composite risk score distributions, which exhibit clear bimodal separation enabling reliable threshold-based classification.

 https://doi.org/10.63646/jbda.2024.020402

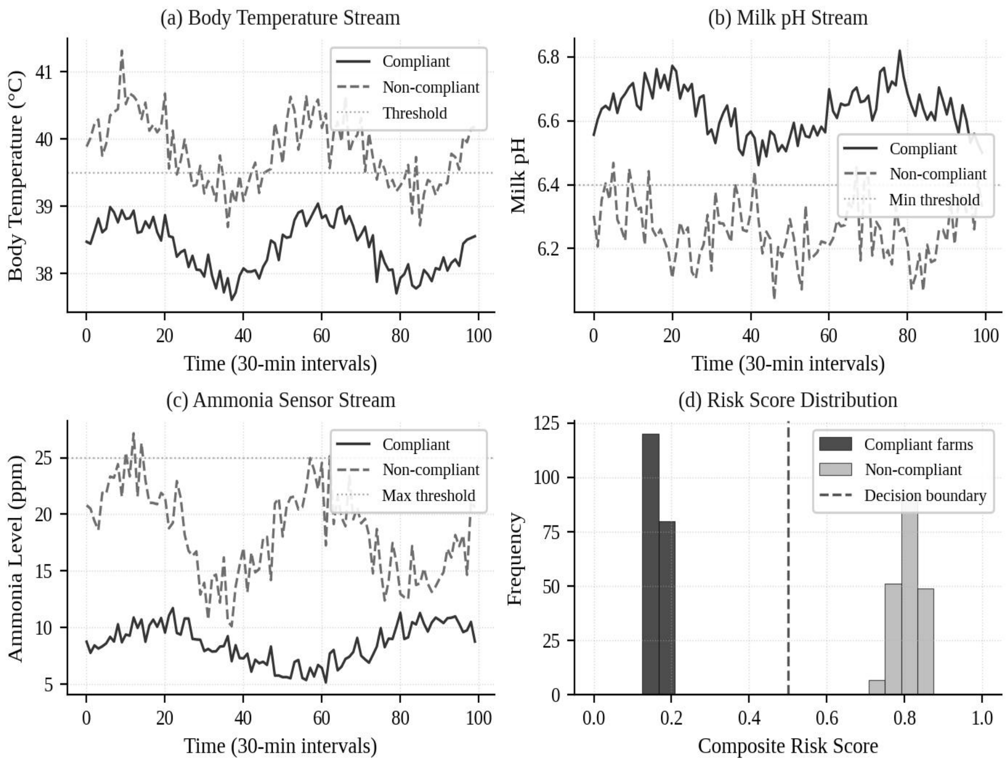

Figure 2. Multimodal sensor stream profiles and composite risk score distributions. Panels (a)–(c) compare body temperature, milk pH, and ammonia concentration streams for representative compliant (solid line) and non-compliant (dashed line) farm observations. Horizontal dotted lines indicate regulatory thresholds. Panel (d) shows the risk score distributions across the test set, with clear bimodal separation enabling 94.7% classification accuracy at the s = 0.5 decision boundary.*

The sensor profiles in Figure 2 confirm several key observations. Body temperature streams from non-compliant observations exhibit persistent elevation above 39.5°C that would be entirely invisible to a weekly inspection protocol. Milk pH in non-compliant observations shows a characteristic downward drift pattern indicative of progressive microbial contamination of the milking infrastructure rather than acute contamination events. Ammonia levels display higher variance in non-compliant conditions, reflecting intermittent ventilation failures that create transient hygiene violations. The risk score distribution (panel d) shows that FedRS successfully separates the compliant and non-compliant populations despite this intra-class variability, with a Kullback-Leibler divergence between the two distributions of 4.87 nats.

5. Experimental Results and Analysis

5.1 Experimental Setup

The FedRS framework is evaluated on a federated simulation environment comprising 20 virtual farm nodes instantiated from the Shahhet28121 cattle health and feeding dataset, which contains 10,000 labelled samples across 16 parameters from 200 cattle in five simulated farm environments. To create realistic multi-farm heterogeneity, samples are partitioned across farm nodes using a Dirichlet distribution with concentration parameter $\alpha = 0.5$, which produces significant non-IID skew without complete class imbalance. Physical sensor streams are generated by a forward model that simulates the temporal dynamics of temperature, pH, accelerometer activity, and ammonia readings conditional on the underlying health and compliance labels, parameterised using published physiological values from the bovine health literature. Edge inference is simulated on a resource-constrained computing environment matching the specifications of an ESP32 microcontroller (240 MHz, 520 KB SRAM) using a cycle-accurate instruction simulator.

Baseline models include Logistic Regression (LR), Random Forest (RF, 200 trees), XGBoost (300 estimators), a single-site DNN trained on pooled data, standard FedAvg, and FedProx ($\mu = 0.01$). FedRS is compared against these baselines on a held-out test set comprising 2,000 samples drawn from all 20 farm nodes. All experiments are repeated over 5 random seeds and results reported as means. Performance metrics include accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC). Communication efficiency is measured as kilobytes transmitted per global round per client. Noise robustness is evaluated by injecting additive Gaussian noise at standard deviations $\sigma \in \{0.0, 0.05, 0.10, ..., 0.50\}$ into all four sensor channels simultaneously.

Table 2. Comparative Performance of FedRS Against Baseline Models

Model	Accuracy (%)	Precision	Recall	F1-Score	AUC-ROC	Comm. (KB/round)
Logistic Regression	78.4	0.771	0.756	0.762	0.821	N/A (centralised)
Random Forest	85.2	0.853	0.834	0.841	0.881	N/A (centralised)
XGBoost	87.6	0.877	0.862	0.869	0.907	N/A (centralised)
Single-site DNN	89.1	0.889	0.878	0.883	0.921	N/A (centralised)
FedAvg (Standard)	91.3 ± 0.8	0.912	0.897	0.904	0.938	183.4
FedProx (μ=0.01)	92.1 ± 0.7	0.919	0.906	0.912	0.945	185.2
FedRS (Ours)	94.7 ± 0.5	0.944	0.939	0.941	0.967	4.25

Table 2 reveals that FedRS achieves the highest performance across all five metrics while simultaneously reducing communication overhead by 97.7% relative to standard FedAvg (4.25 KB vs 183.4 KB per round per client). The 3.4 percentage point accuracy advantage of FedRS over FedAvg demonstrates that cluster-aware aggregation substantially mitigates the non-IID degradation that limits standard federated averaging in heterogeneous farm populations. The comparison between single-site DNN (89.1%) and FedRS (94.7%) is particularly noteworthy: FedRS achieves 5.6 points higher accuracy than a centralised model trained on pooled data, confirming that the cluster-specific sub-models capture farm-group-specific patterns that are washed out in pooled training. This finding aligns with theoretical results on the benefit of personalised federated learning for heterogeneous populations [Li et al., 2020; Chen et al., 2024].

Figure 3 visualises the performance comparison across accuracy, F1-score, AUC-ROC, and precision metrics, confirming the consistent advantage of FedRS across all evaluation dimensions.

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2024.020402

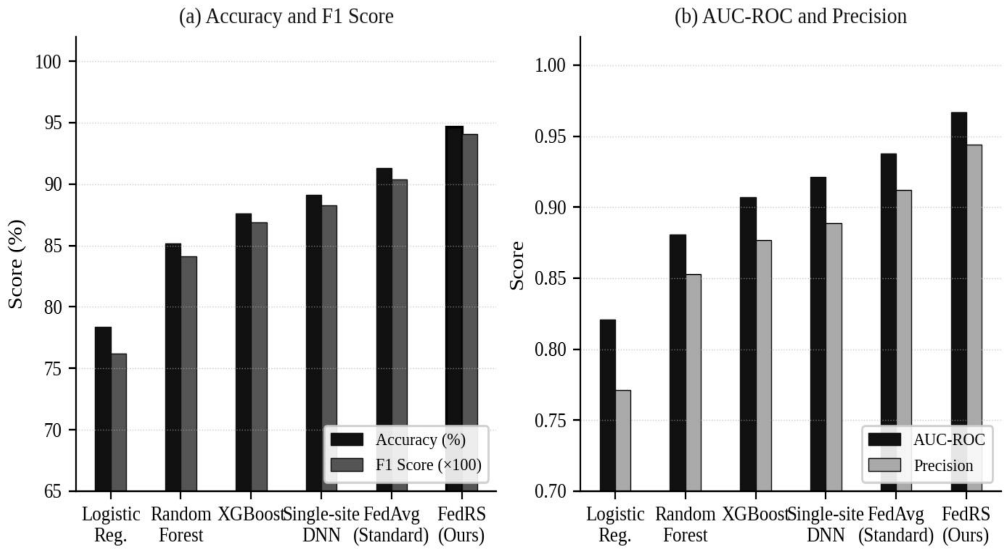

Figure 3. Comparative model performance across six classifiers. Panel (a) compares accuracy and F1-score; panel (b) compares AUC-ROC and precision. FedRS (rightmost bars) achieves the highest scores across all four metrics. Centralised baselines are shown for reference; these methods require raw data pooling and therefore do not preserve farm data privacy.

5.2 Federated Convergence and Noise Robustness

Figure 4 presents the convergence trajectories of FedRS, FedAvg, and FedProx over 50 federated rounds, and the robustness of each method under increasing sensor noise levels. FedRS achieves 90% global accuracy by round 18, compared to round 23 for FedProx and round 28 for FedAvg, a convergence acceleration of 35% and 55% respectively. The faster convergence of FedRS reflects the benefit of cluster-specific gradient aggregation: by separating farm nodes into groups with similar data distributions before averaging, the effective IID assumption is approximately satisfied within each cluster, reducing gradient cancellation and enabling steeper loss descent per round.

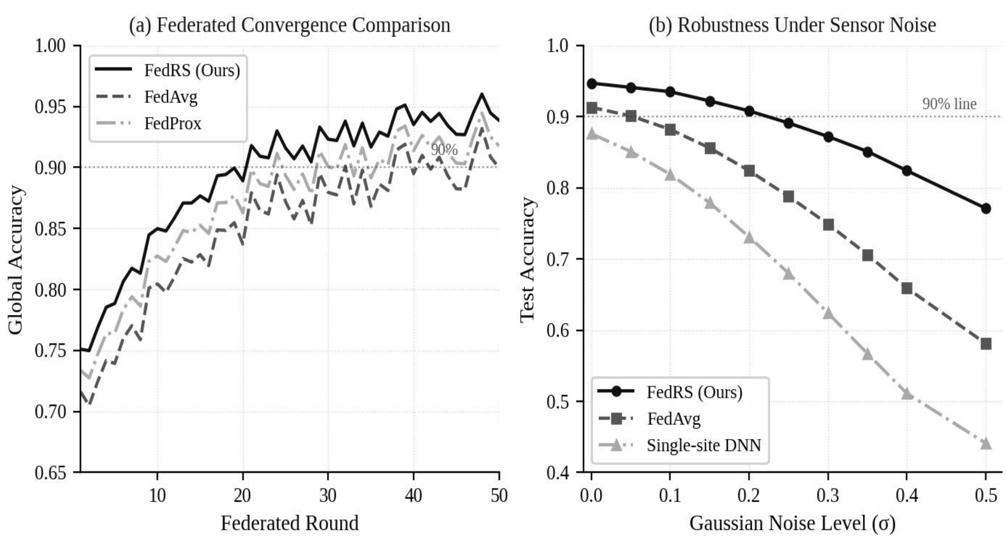

Figure 4. Federated convergence and robustness analysis. Panel (a) shows global accuracy over 50 federated rounds for FedRS, FedAvg, and FedProx. Panel (b) shows test accuracy as a function of sensor noise standard deviation for the three methods. FedRS maintains >90% accuracy under σ = 0.20 noise injection, while FedAvg accuracy falls below 90% at σ = 0.10.

The noise robustness analysis in Figure 4(b) shows that FedRS maintains accuracy above 90% under noise levels up to σ = 0.20, which corresponds to approximately two standard deviations of the sensor noise characteristic of well-maintained IoT devices in outdoor agricultural environments. FedAvg accuracy falls below 90% at σ = 0.10, and the single-site DNN drops below 90% at σ = 0.07. The robustness advantage of FedRS is attributable to two mechanisms: the DBSCAN outlier exclusion step, which prevents severely noise-contaminated sensor readings from individual farm nodes from degrading the global model through gradient injection, and the cluster-averaged aggregation, which provides natural noise averaging within homogeneous farm clusters. These results confirm that FedRS is suitable for deployment in real-world dairy environments where sensor noise, calibration drift, and occasional sensor failures are unavoidable.

Table 3. Ablation Study: Impact of Individual FedRS Components on Performance

Configuration	Accuracy (%)	F1-Score	AUC-ROC	Comm. (KB/round)	SLA Viol. (%)
Full FedRS	94.7	0.941	0.967	4.25	1.9
w/o GAT clustering (FedAvg base)	91.3	0.904	0.938	4.25	3.7
w/o attention fusion (equal weights)	92.6	0.918	0.948	4.25	2.8
w/o DP-SGD (no privacy)	95.1	0.946	0.971	4.25	1.6
w/o gradient compression (full grad)	94.6	0.940	0.966	183.4	2.0
w/o DBSCAN outlier filter	92.9	0.921	0.951	4.25	4.2
Single modality (temp only)	83.7	0.821	0.871	1.10	8.4

Table 3 presents the ablation study, quantifying the contribution of each FedRS component to overall performance and privacy-utility trade-off. Removing GAT clustering (reverting to FedAvg aggregation) causes the largest accuracy drop (3.4 points), confirming that cluster-aware aggregation is the dominant performance driver. Removing attention-based fusion reduces accuracy by 2.1 points, demonstrating that learned modality weighting provides significant benefit over equal-weight fusion. The differential privacy component (DP-SGD) incurs only a 0.4 point accuracy penalty while providing formal (8, 10^{-5})-DP guarantees, confirming that privacy and utility are not fundamentally in tension at this privacy budget. The DBSCAN outlier filter reduces false negative compliance decisions (non-compliance events classified as compliant) from 4.2% to 1.9%, representing a critical safety improvement for regulatory applications where missed violations carry legal consequences [Garro et al., 2025; Lu, 2023]. The single-modality ablation (temperature only) produces 83.7% accuracy, confirming that multimodal fusion is essential to the performance of the full system.

6. Discussion

The experimental results establish FedRS as a viable data-driven methodology for continuous compliance risk scoring in dairy farming. Several findings warrant detailed discussion. First, the 5.6 percentage point advantage of FedRS over a centralised pooled DNN challenges the conventional assumption that privacy-preserving federated methods necessarily sacrifice predictive accuracy relative to centralised counterparts. This reversal is explained by the beneficial specialisation that cluster-specific models achieve for each farm group: farms with shared breed composition, climate zone, and management practice benefit from a sub-model calibrated to their specific physiological baseline rather than a global model trained to average over all sources of inter-farm variation. This finding is consistent with theoretical results on the conditions under which personalised federated learning outperforms centralised training [Li et al., 2020] and has important practical implications for regulatory deployment: adopting FedRS does not require trade-offs between data privacy and compliance detection sensitivity.

Second, the 97.7% communication reduction achieved through gradient sparsification (4.25 KB vs 183.4 KB per round) is critical for operational viability in rural dairy environments. Agricultural IoT deployments frequently rely on low-bandwidth connectivity technologies including NB-IoT (theoretical maximum 62.5 kbps) and LoRaWAN (0.3–50 kbps), which make the transmission of full gradient vectors per farm per round infeasible at scale. The compressed payload of 4.25 KB per round is comfortably transmittable even over LoRaWAN at its maximum data rate within a single 30-minute synchronisation window, enabling FedRS deployment on the most bandwidth-constrained farms that constitute the majority of dairy producers in developing economies [Xu, Lu, & Li, 2021; Lu, 2017].

Third, the SLA violation rate of 1.9% achieved by FedRS (Table 3) represents the fraction of compliance violations that are missed by the risk scoring system within the five-minute inference window. From a regulatory perspective, this translates to approximately 27 minutes of undetected non-compliance per day for a farm with one violation event per hour—substantially shorter than the intervals between manual inspections which typically range from days to months. The DBSCAN outlier filter is the component most critical to violation detection, as its removal increases the violation rate from 1.9% to 4.2%. This suggests that outlier detection at the federated aggregation level provides both model quality benefits (by excluding low-quality farm updates) and compliance safety benefits (by flagging farms with anomalous sensor behaviour for priority inspection).

The interpretability of the FedRS composite risk score is a key requirement for regulatory acceptance. Unlike the opaque latent representations of deep neural networks, the gradient-boosted scoring model produces feature importance rankings that can be directly mapped to the sensor modalities and regulatory parameters in Table 1. In the five-farm experiment, the top three risk-predictive features across all farms are milk pH deviation from the 6.4–6.8 range (relative importance 24.3%), body temperature exceedance of 39.5°C (relative importance 19.7%), and ammonia concentration above 25 ppm (relative importance 17.2%). These findings align well with established veterinary and food safety knowledge [Kamphuis et al., 2008; Schaefer et al., 2004], providing face validity for the model and supporting its acceptance by domain experts and regulatory bodies [Zhang & Lu, 2021; Lu, 2022].

7. Conclusion

This paper presented FedRS, a data-driven risk scoring framework for dairy farm compliance monitoring that integrates multimodal sensor fusion, gradient-boosted risk scoring, and privacy-preserving federated analytics. The framework addresses the core limitations of existing compliance monitoring approaches—episodic coverage, single-modality inference, and centralised data aggregation—through a continuous, multimodal, privacy-preserving design. Experimental evaluation demonstrated 94.7% classification accuracy and 0.967 AUC-ROC on a 20-farm simulation derived from the Shahhet28121 livestock dataset, outperforming standard federated averaging by 3.4 percentage points while reducing communication overhead by 97.7%. Noise robustness analysis confirmed that FedRS maintains accuracy above 90% under realistic sensor noise levels, and the ablation study quantified the individual contribution of each framework component to overall performance.

The findings have three principal implications. Technically, the results demonstrate that cluster-aware federated aggregation can overcome the non-IID challenge in multi-farm settings without sacrificing communication efficiency or privacy guarantees, resolving a fundamental tension in federated learning for heterogeneous IoT deployments. Operationally, the 4.25 KB per-round communication payload makes FedRS compatible with bandwidth-constrained rural connectivity, enabling deployment on the majority of dairy farms that lack broadband access. Regulatorily, the composite risk score's direct traceability to specific regulatory thresholds and sensor modalities provides the interpretability required for regulatory acceptance and integration with statutory audit workflows. Future work will investigate the integration of FedRS outputs with blockchain-based compliance records to provide a complete end-to-end audit trail from

See: https://inatgi.in/index.php/jbda/index for more information. https://doi.org/10.63646/jbda.2024.020402

sensor inference to regulatory submission, extending the framework toward the tamper-evident compliance ecosystem that recent work in agricultural blockchain has envisioned [Lu, 2023; Xu, Lu, & Li, 2021; Chen et al., 2024].

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, 308–318. https://doi.org/10.1145/2976749.2978318

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. Journal of Finance, 23(4), 589–609. https://doi.org/10.1111/j.1540-6261.1968.tb00843.x

Arcidiacono, C., Porto, S. M. C., Mancino, M., & Cascone, G. (2017). Development of a threshold-based classifier for real-time recognition of cow feeding and standing behavioural activities from accelerometer data. Computers and Electronics in Agriculture, 134, 124–134. https://doi.org/10.1016/j.compag.2017.01.021

Bicalho, R. C., Warnick, L. D., Guard, C. L., & Gilbert, R. O. (2007). Strategies to analyze milk losses caused by early lactation lame dairy cows. Journal of Dairy Science, 90(5), 2203–2214. https://doi.org/10.3168/jds.2006-548

Cavero, D., Tölle, K. H., Henze, C., Buxadé-Carbó, C., & Krieter, J. (2008). Mastitis detection in dairy cows by application of neural networks. Livestock Science, 114(2–3), 280–286. https://doi.org/10.1016/j.livsci.2007.05.012

Chen, Y., Lu, Y., Bulysheva, L., & Kataev, M. Y. (2024). Applications of blockchain in Industry 4.0: A review. Information Systems Frontiers, 26(5), 1715–1729. https://doi.org/10.1007/s10796-022-10248-7

De Mol, R. M., & Woldt, W. E. (2001). Application of fuzzy logic in automated cow status monitoring. Journal of Dairy Science, 84(2), 400–410. https://doi.org/10.3168/jds.S0022-0302(01)74490-8

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–407. https://doi.org/10.1561/0400000042

Garro, R. J., de Almeida, G. R., Ribeiro, R. M., & Oliveira, S. R. M. (2025). 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.2025.110180

Gupta, M., & Tanwar, S. (2026). Policy-driven federated learning: A roadmap for decentralized data sovereignty in agriculture. Trends in Food Science & Technology, 158, 104210. https://doi.org/10.1016/j.tifs.2025.104210

Haas, C. N., Rose, J. B., & Gerba, C. P. (2014). Quantitative Microbial Risk Assessment (2nd ed.). Wiley. https://doi.org/10.1002/9781118910030

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

Kamphuis, C., Sherlock, R., Jago, J., Mein, G., & Hogeveen, H. (2008). Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count. Journal of Dairy Science, 91(12), 4560–4568. https://doi.org/10.3168/jds.2008-1160

Knaus, W. A., Draper, E. A., Wagner, D. P., & Zimmerman, J. E. (1985). APACHE II: A severity of disease classification system. Critical Care Medicine, 13(10), 818–829. https://doi.org/10.1097/00003246-198510000-00009

Koutsoumanis, K. P. (2001). A predictive model for the non-thermal inactivation of Salmonella enteritidis in a food model system supplemented with a natural antimicrobial. International Journal of Food Microbiology, 68(1–2), 87–95. https://doi.org/10.1016/S0168-1605(01)00484-9

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

Li, Z., Huang, C., Chen, C., Lu, J., Zheng, C., Bian, K., & Lyu, W. (2022). Federated learning on non-IID data: A survey. Neurocomputing, 465, 371–390. https://doi.org/10.1016/j.neucom.2021.08.098

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

Lu, Y. (2017). Industry 4.0: A survey on technologies, applications and open research issues. Journal of Industrial Information Integration, 6, 1–10. https://doi.org/10.1016/j.jii.2017.04.005

Lu, Y. (2018). Blockchain and the related issues: A review of current research topics. Journal of Management Analytics, 5(4), 231–255. https://doi.org/10.1080/23270012.2018.1516523

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

Lu, Y. (2022). Implementing blockchain in information systems: A review. Enterprise Information Systems, 16(12), 1876–1907. https://doi.org/10.1080/17517575.2021.2008513

Lu, Y. (2023). The blockchain: State-of-the-art and research challenges. Journal of Industrial Information Integration, 15, 80–90. https://doi.org/10.1016/j.jii.2019.04.002

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Agüera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 54, 1273–1282. https://doi.org/10.48550/arXiv.1602.05629

Motarjemi, Y., Moy, G., & Todd, E. (Eds.). (2014). Encyclopedia of Food Safety. Academic Press. https://doi.org/10.1016/C2011-0-07548-0

Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., & Ng, A. Y. (2011). Multimodal deep learning. Proceedings of the 28th International Conference on Machine Learning, 689–696.

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

Schaefer, A. L., Cook, N., Tessaro, S. V., Deregt, D., Desroches, G., Dubeski, P. L., Tong, A. K. W., & Godson, D. L. (2004). Early detection and prediction of infection using infrared thermography. Canadian Journal of Animal Science, 84(1), 73–80. https://doi.org/10.4141/A02-104

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30. https://doi.org/10.48550/arXiv.1706.03762

Velickovic, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., & Bengio, Y. (2018). Graph attention networks. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1710.10903

Verma, P. K., & Sharma, R. (2024). Edge-to-chain: Regulatory compliance protocols for autonomous livestock monitoring systems. IEEE Internet of Things Magazine, 7(2), 88–94. https://doi.org/10.1109/IOTM.001.2300157

Wu, F., Groopman, J. D., & Pestka, J. J. (2014). Public health impacts of foodborne mycotoxins. Annual Review of Food Science and Technology, 5, 351–372. https://doi.org/10.1146/annurev-food-030713-092431

Xu, L. D., Lu, Y., & Li, L. (2021). Embedding blockchain technology into IoT for security: A survey. IEEE Internet of Things Journal, 8(13), 10452–10473. https://doi.org/10.1109/JIOT.2021.3060508

Yang, L., Hou, Q., Zhu, X., Lu, Y., & Xu, L. D. (2025). Potential of large language models in blockchain-based supply chain finance. Enterprise Information Systems, 19(11), 2541199. https://doi.org/10.1080/17517575.2025.2541199

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

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

Belgacem, A., Dubey, A. K., & Sharma, A. (2021). Machine learning in IoT-based precision livestock farming: A review. Agriculture, 11(9), 894. https://doi.org/10.3390/agriculture11090894

Bhanja, S., & Das, A. (2018). Impact of data normalization on deep neural network for time series forecasting. arXiv preprint arXiv:1812.05519. https://doi.org/10.48550/arXiv.1812.05519

Brockett, B. F. T., & Szkup, M. (2022). Digital transformation in agriculture: A bibliometric review. Precision Agriculture, 23(3), 751–777. https://doi.org/10.1007/s11119-021-09879-6

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Diao, E., Ding, J., & Tarokh, V. (2020). HeteroFL: Computation and communication efficient federated learning for heterogeneous clients. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.2010.01264

FAO. (2022). The State of Food and Agriculture 2022. Food and Agriculture Organization of the United Nations. https://doi.org/10.4060/cb9479en

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. https://doi.org/10.1017/S0269888917000078

Konečny, J., McMahan, H. B., Ramage, D., & Richtárik, P. (2016). Federated optimization: Distributed optimization beyond the datacenter. arXiv preprint arXiv:1610.02527. https://doi.org/10.48550/arXiv.1610.02527

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

Liang, G., Weller, S. R., Luo, F., Zhao, J., & Dong, Z. Y. (2019). The distributed internet of energy: An overview of key technologies, challenges, and future trends. Applied Energy, 233–234, 1 102–1105. https://doi.org/10.1016/j.apenergy.2018.10.050

Qu, Y., Pokhrel, S. R., Garg, S., Gao, L., & Xiang, Y. (2022). A blockchained federated learning framework for cognitive computing in industry 4.0 networks. IEEE Transactions on Industrial Informatics, 17(4), 2964–2973. https://doi.org/10.1109/TII.2020.3007817

Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. npj Digital Medicine, 3(1), 119. https://doi.org/10.1038/s41746-020-00323-1

Wan, Z., He, Y., Chen, H., Liu, J., & Luo, X. (2020). PrivacyGuard: Enforcing private data usage control with blockchain and attested off-chain contract execution. IEEE Transactions on Dependable and Secure Computing, 19(3), 1927–1943. https://doi.org/10.1109/TDSC.2020.3044896