THE SCIENTIFIC TEMPER (2025) Vol. 16 (12): 5312-5321 E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2025.16.12.18 https://scientifictemper.com/
RESEARCH ARTICLE

FDBSCAN-MBKSched: A Hybrid Edge-Cloud Clustering and Energy-Aware Federated Learning Framework with Adaptive Update Scheduling for Healthcare IoT

S. Ranganathan[1*], V. Umadevi[2]

Abstract

The explosive growth of the Internet of Medical Things (IoMT) has created huge, diverse, and noisy health data streams that require processing in real time under stringent energy and latency budgets. Conventional fuzzy clustering and synchronous federated learning methodologies tend to be plagued by noise sensitivity, excessive communication overhead, and poor model convergence efficiency. To address above mentioned issues, this work introduces FDBSCAN–MBKSched, a federated learning and clustering hybrid framework combining DBSCAN-based real-time abnormal health state detection and data filtering at the edge, Mini-Batch K-Means using MapReduce in the cloud, and an adaptive update scheduling mechanism. DBSCAN removes noisy data and identifies abnormal health states in real time at the edge, while non-emergency summaries are sent to the cloud for scalable clustering. The Federated Learning (FL) module governs distributed model training without sharing raw data, with devices dynamically adapting update frequencies as a function of model freshness, battery level, and event urgency. Experimental validation on real- IoMT datasets shows that FDBSCAN–MBKSched attains 12% improved anomaly detection accuracy, 21% reduced energy usage, and 17% lower emergency latency compared to traditional fuzzy clustering–based baselines. These findings demonstrate the efficiency of the framework for latency-sensitive, privacy-preserving, and resource-constrained healthcare applications.

Keywords: Internet of Medical Things (IoMT); Edge–Cloud Collaboration; DBSCAN Clustering; Mini-Batch K-Means; Federated Learning; Adaptive Scheduling; Energy-Efficient Healthcare Analytics.

Introduction

The IoMT is a result of the rapid rise in health information produced by intelligent healthcare systems in the past few years, mainly with IoT-based devices. Although IoMT is currently used in almost every medical field, there are significant security and privacy issues because of the large and diverse amount of data it generates (Zhou *et al.*, 2025). IoMT is distributed and sensitive to information privacy, so centralized machine learning (ML) strategies are inappropriate for identifying abnormalities in IoMT data (Nguyen *et al.*, 2023).

Wearable technologies and intelligent platforms make it possible to observe health continuously because of to the growing popularity of IoMT devices, but they also present significant distributed analytics issues. Since centralized storage raises the possibility of breaches involving private medical records subject to regulations such as HIPAA as well as GDPR, security of data continues to be crucial (Rani *et al.*, 2023). When sending massive amounts of data to the cloud, scalability problems occur because edge devices have limited resources, which results in high latency and network congestion (Nayak *et al.*, 2024). Furthermore, real-time anomaly detection is made more difficult by data heterogeneity from different sensors and sampling rates. lastly, because of synchronization issues and dataset-size limitations, ineffective global aggregation using

[1]Research Scholar, PG & Research Dept of Computer Science, Nehru Memorial college, (Autonomous), Affiliated to Bharathidasan University, Puthanampatti, Trichy Dt. Tamil Nadu, India.

[2]Research Supervisor, PG & Research Dept of Computer Science, Nehru Memorial college, (Autonomous), Affiliated to Bharathidasan University, Puthanampatti, Trichy Dt.Tamil Nadu, India.

***Corresponding Author:** S. Ranganathan, Research Scholar, PG & Research Dept of Computer Science, Nehru Memorial college, (Autonomous), Affiliated to Bharathidasan University, Puthanampatti, Trichy Dt. Tamil Nadu, India, E-Mail: sranganathanmail@gmail.com

How to cite this article: Ranganathan, S., Umadevi, V. (2025). FDBSCAN-MBKSched: A Hybrid Edge-Cloud Clustering and Energy-Aware Federated Learning Framework with Adaptive Update Scheduling for Healthcare IoT. The Scientific Temper, **16**(12):5312-5321.

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.12.18

Source of support: Nil

Conflict of interest: None.

Received: 19/11/2025 **Accepted:** 20/12/2025 **Published:** 25/12/2025

conventional clustering techniques impairs efficiency on streaming IoMT information (Prasad *et al*., 2022).

To surmount these limitations, FL and Cloud–Edge cooperation have become prominent paradigms for distributed intelligence. FL facilitates decentralized training of ML models over medical devices and edge nodes in a way that maintains data privacy. Every device trains locally and exchanges model updates rather than raw data, keeping raw data on-device. FL thus becomes most suitable in IoT settings where privacy is critical (Pinto *et al*., 2025). Enhanced LSTM for heart disease prediction in IoT-enabled smart healthcare systems(Gold *et al*., 2024).

The cloud–edge end mixed computing design has become a crucial paradigm for processing and analyzing massive amounts of edge data due to the quick development of 5G and IoT technologies. (Li *et al*., 2024). In this research, edge computing improves response time by conducting initial processing as well as anomaly detection near data sources, while global model aggregation and large-scale analytics are offered by the cloud layer.

In this hybrid model, both efficient data clustering and scheduling are of crucial importance in achieving computation balance among the edge and cloud layers. Density-based clustering algorithms like DBSCAN enable local anomaly discovery by identifying outliers in medical streaming data, while cloud-side scalable aggregation is possible using Mini-Batch K-Means with MapReduce allowing high-throughput clustering of non-emergency summaries. Optimizing IoT application deployment with fog - cloud paradigm: A resource-aware approach(K Mohamed Arif Khan *et al*., 2024). In heterogeneous IoMT environment, all these hybrid methods improve overall reliability, lower transmission expense, and enable adaptive learning.

The following are the contributions of proposed work.

- To present the FDBSCAN-MBKSched framework, a novel hybrid edge–cloud architecture which includes energy-efficient federated learning, scalable global clustering, as well as density-driven local anomaly detection.
- A DBSCAN-based local clustering module is employed to remove noise during a device's level and facilitate real-time identification of unusual physiological trends.
- A Mini-Batch K-Means algorithm has been implemented under the MapReduce paradigm to carry out high-throughput global clustering in non-emergency summaries coming from distributed edges. Scalability across sizable healthcare datasets is thus guaranteed.
- To reduce redundant communication and maximize resource utilization, a novel scheduling technique is employed that dynamically regulates the frequency of model updates based on device energy, event urgency, and model freshness.

Collectively, these components form an energy-adaptive, multi-objective optimization framework that improves anomaly detection accuracy, reduces latency, and lowers energy consumption compared to existing FL-based baselines.

Related works

FL enables multi-institutional medical AI research by transmitting model parameters instead of patient data, ensuring privacy compliance. Current studies focus on imaging-based disease prediction, particularly for cancer and COVID-19 and explore strategies to manage data heterogeneity and enhance communication security (Choi *et al*., 2024).

Li *et al*., (2023) proposed the DP-Prox as well as PDP-Prox methods under various privacy budget scenarios, integrated differential privacy and customized differential privacy using FedProx, and ran simulations on several datasets. Recent studies (Sarkar *et al*.,2024), such as those extending the FedProx framework (e.g., G-Federated Proximity), focus on improving training stability, convergence speed, and model accuracy through normalization and adaptive optimization to better support real-time IIoT environments.

Machine learning models often rely on remote training, which increases resource consumption and raises privacy concerns. DRMF: Optimizing machine learning accuracy in IoT crop recommendation with domain rules and Miss Forest imputation (Sindhu *et al*.,2024). FL resolved these challenges through facilitating local model training along with aggregation. Latest methodologies, including Energy Saving Client Selection (ESCS), enhanced device participation according to battery point, learning ability, as well as network quality for attaining energy efficiency without compromising effectiveness (Maciel *et al*., 2024).

Liu *et al*., (2021) introduced an innovative method for focused poverty alleviation using the clustering technique known as DBSCAN along with a deep neural network model built on edge computing. The system allows for real-time data mining at the edge to find households living in poverty. This utilized DBSCAN to find important poverty features and help with smart household grouping. Enhancing IoT blockchain scalability through the eepos consensus algorithm (Ragul *et al*.,2025)

In 2024, Retiti Diop Emane and others came up with a new way to find anomalies through the combination of Graph Convolutional Networks (GCNs) along with DBSCAN. GCNs which is a deep learning algorithm specifically designed for graph data, uses graph topology and attribute information to get useful node and edge representations. Also, the researchers Bushra *et al*., (2024) suggested an unsupervised approach to ascertain the optimal DBSCAN parameters based on its density distribution.

Y Li *et al*., (2024) put forward a device scheduling strategy that uses adaptive batch sizes. This strategy chooses devices with data of high usefulness and changes their mini-batch sizes and gradient quantization rates based on the state of the network.

Recent research (Hicks *et al*., 2021) has emphasized the difficulty of effectively clustering millions of cells.

Conventional approaches such as k-means need to load whole datasets into memory, which makes them less scalable. The mbk means package addresses this by using a mini-batch k-means approach with on-disk data handling, enabling faster and memory-efficient clustering for large-scale single-cell datasets.

In related works, several studies have focused on improving anomaly detection in the distribution Internet of Things (IoT) to enhance reliability and automation in power networks. One method that works well for this is to use spatiotemporal correlation evaluation, an improved DBSCAN algorithm, and fuzzy logic to find and identify unusual measurement information from low-voltage tracking devices (Shao *et al*., 2022). The IoMT improved medical services via connecting devices, however it has big problems with confidentiality and safety. FL may assist with this by enabling distributed model training that preserves privacy while identifying abnormalities without disclosing raw data. (Pinto *et al*., 2025).

Most current research tackles the aforementioned challenges in distinct ways, resulting in notable gaps: federated learning analyses typically prioritize aggregation as well as convergence amidst heterogeneity, often neglecting edge-level anomaly detection; anomaly-detection literature prioritizes accuracy along with parameter optimization while overlooking downstream communication or scheduling expenses; and extensive clustering research addressed computational scalability without integrating selective, urgency-driven summary ingestion coming from edge devices. Energy-aware Security Optimized Elliptic Curve Digital Signature Algorithm for Universal IoT Networks (Jenifer *et al*.,2025). These limitations showed the requirement of single approach which deals with privacy, rapidity, energy conservation, and scalability all at once.

FDBSCAN–MBKSched bridges these gaps by performing adaptive, quartile-tuned DBSCAN at the edge for immediate emergency signalling, employing a multi-factor, energy-aware scheduler to decide Immediate/Delay/Skip actions for model updates, and aggregating non-emergency summaries in the cloud using a MapReduce Mini-Batch K-Means pipeline, forming a unified and efficient distributed framework.

Proposed Methodology

Datasets used

For evaluating the FDBSCAN-MBKSched framework, we chose two complementary recent datasets: Digital Exposome (Johnson, 2025) and PIF (Dhaliwal *et al*., 2023).

Digital Exposome – This dataset contains 40 participants with physiological (HR, HRV, EDA, BVP) and environmental signals (pollution, noise, temperature, etc.). This dataset supports edge anomaly detection, cloud clustering, and federated learning across participants. Emergency events are modeled via physiological/environmental thresholds.

PIF (Physiological & Inertial Features) includes wearable physiological and inertial measurements with overt fall events, acting as authentic emergency instances. It is beneficial for benchmarking accurate emergency detection and adaptive scheduling.

Digital Exposome serves as the main dataset (heterogeneity, scalability, FL validation) in this work and PIF as the secondary dataset (emergency validation). This pairing provides both realistic and holistic evaluation.

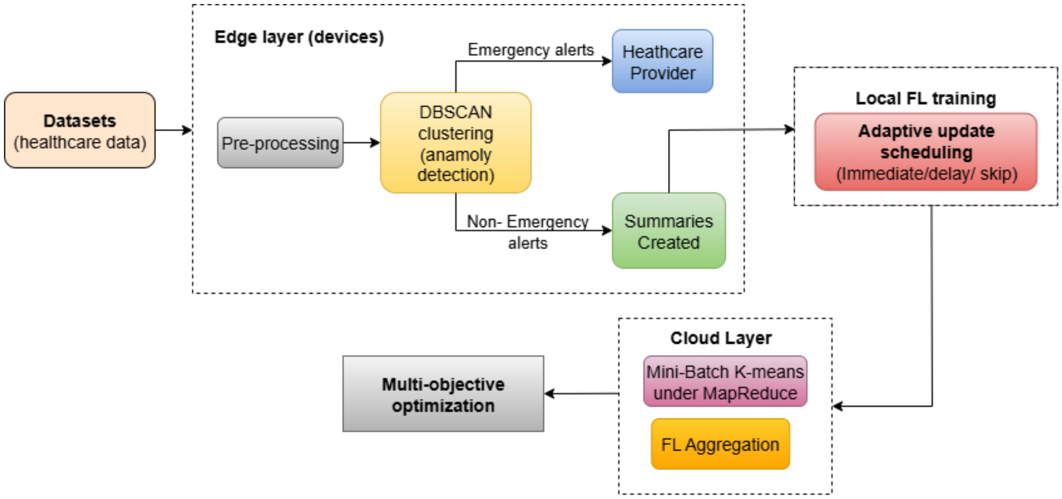
Overview of the Proposed Work

The proposed FDDBSCAN-MBKSched framework represents a hybrid edge-cloud architecture with three logical layers. At the Edge Layer, which includes devices and gateways, raw data goes through steps like denoising, normalization, as well as segmentation to get it ready for use. DBSCAN is employed on recent sliding windows to find anomalies in real time as well as flag emergency events right away. Each edge node also does local federated training on confidential time-series data to make predictions and sort data. An Adaptive Update Scheduler uses smart reasoning to decide if to send updates of the model quickly, wait, or not send them at all. It does this based on things like how urgent they are, how new the data is, and how much energy the device has. The Cloud Layer collects non-emergency data and uses Mini-Batch K-Means in a MapReduce framework to do global clustering and get insights at the population level. Additionally, it comprises a federated learning aggregator which transmits the global model back to the network after gathering model updates from edge devices and combining them using techniques like weighted averaging. In order to balance clustering accuracy, anomaly detection latency, and communication energy consumption throughout the system, the Optimization Module lastly uses a multi-objective optimizer to adjust the scheduler's hyperparameters. *Et al*ure 1 illustrates this.

The FDBSCAN-MBKSched model combines adaptive scheduling with edge anomaly detection, cloud clustering, and federated deep learning.

Pre-processing

To get the Digital Exposome and PIF datasets prepared to utilize the proposed framework, initially low-pass and median filters had been employed to eliminate high-frequency noise from ECG, EDA, as well as inertial signals. This improved the quality of the signals so that they could be analyzed accurately. Then, physiological signals like heart rate and EDA were adjusted so that their mean was zero and their variance was one. This made sure that all participants and sensors had the same input ranges (Backhus *et al*., 2025). The continuous signals were subsequently divided into overlapping time-series windows of length W seconds



*Et al***ure 1:** Flow diagram of proposed FDBSCAN-MBKSched framework

and overlap O seconds by employing the sliding window method (Bai *et al*., 2025) to identify temporal patterns. From every segment, features have been extracted, comprising statistical (mean, variance), frequency-domain (power spectral density, dominant frequencies), as well as temporal features (slopes, peaks) to facilitate clustering as well as anomaly detection (Singh and Krishnan., 2023). Lastly, the data were tagged as emergency events. For example, Digital Exposome abnormalities were simulated by setting thresholds like HR > 150 bpm above safety limits, and falls in PIF were considered as emergency instances.

DBSCAN based Anomaly Detection at Edge

DBSCAN is used by every device to identify irregularities in its local segment. Because of its exceptional capacity to handle intricate data structures as well as noisy circumstances, DBSCAN is a clustering technique. Density-Based Spatial Clustering of Applications with Noise, or DBSCAN, is a great option for this anomaly detection objective because it has a number of important benefits. Depending on the data density, DBSCAN computes the right number of clusters. The result guarantees that we can identify all significant clusters in the dataset, irrespective of their size or number, and does away with the need for manual parameter setting.

The epsilon parameter (ε), which defines the radius of the neighborhood used for calculating the density of the points, is one of the most important factors that has to be precisely changed in DBSCAN (Retiti Diop Emane *et al*., 2024). Quartiles were modified when utilizing the epsilon value for clustering [31]. These statistical measures divide a dataset as four parts, every one of which displays a quarter of the total data. We were interested in the interquartile range (I_Q), as well as the first, second, and third quartiles ($Q1$, $Q2$, and $Q3$). The first quartile ($Q1$) is the value that 25% of the data fall below. The second quartile ($Q2$) is the median and divides the dataset into two equal parts. The third quartile ($Q3$) is the value that 75% of the data fall below; and the interquartile range (IQR) is the dataset's distribution. It is calculated by deducting the third quartile ($Q3$) from the first quartile ($Q1$). Interquartile range (I_Q), the upper limit (U_b) and the lower limit (L_b), which are significant in determining the value of Epsilon.

$$\varepsilon \aleph \ddot{u}_{\ddot{u}} \tag{1}$$

Using quartiles, I_Q, U_b and L_b can be calculated. Lastly, using the epsilon value and 2-Dimensional data, Clustering is performed for separating normal data from anamoly data.

$$Clusters = \begin{cases} Anamoly\left(abnormal\right);DBSCAN\ cluster = -1 \\ Not\ anamoly\left(normal\right);DBSCAN\ cluster \neq -1 \end{cases} \tag{2}$$

Detected anomalies (emergencies) are immediately sent to healthcare providers. Non-emergency summaries (statistical features) are forwarded to the cloud.

Federated Learning with Adaptive Scheduling

The local model (W_D) is maintained by every device D. During round R, the standard FL comprises devices for calculating the local updates which will be passed to server. Adaptive scheduling, however, allows device D to determine a sending action $a_D^R \in \left\{Immediate, delay\ and\ skip\right\}$. This is dependent on three parameters as following.

- Urgency score $\left(U_D\right)$: This is calculated using DBSCAN anomaly flags along with clinical thresholds.
- Model staleness or also called freshness $\left(S_D\right)$: where final updates acknowledged time.
- Energy state $\left(E_D\right)$: This is based on battery level and calculated residual energy.

The local utility can be defined as,

$$P_D = \alpha.norm\left(U_D\right) + \beta.norm\left(1 - S_D\right) + \gamma.norm\left(E_D\right) \tag{3}$$

In above equation, weighting hyperparameters are denoted using variables α, β and γ and variables are mapped to [0,1] by norms.

Rule of decision can be considering for thresholding (example) can be taken as following: When P_D is greater than $\tau_{immediate}$, instantly send. Otherwise, if $P_D \in (\tau_{delay-low}, \tau_{immediate})$, Delay the updates from the buffer until the next available time slot or Skip Else.

Following this, using federated averaging model aggregation is carried out. A suitable FL aggregation technique, such as Federated Averaging, is used on the central server to aggregate the parameters that have been updated from every node. Despite explicitly exposing personal information, this approach creates an enhanced global model based on the knowledge of all nodes.

The model is updated by FedAvg through averaging the parameters of the model that are obtained from multiple local nodes following their training cycles. consider W_k be the model parameters of the model coming from node named k, in which k ranges from 1 to K, and n_k be the number of samples that node k utilized for training. The central server (CS) updates the global model (GM) W (Alasmari *et al*., 2025) using the equation that follows.

$$W = \frac{\sum_{k=1}^{K} n_k W_k}{\sum_{k=1}^{K} n_k} \tag{4}$$

This weighted average allows us to modify the involvement of participants based on the amount of data they have, ensuring that nodes with larger datasets have a greater effect on the construction of the global model.

Cloud Clustering: Mini-Batch K-Means under MapReduce

Once edge devices identify and locally filter out anomalies with DBSCAN, only non-emergency summary information like physiological signal statistical aggregates is sent to the cloud. Bandwidth consumption and device energy are conserved. The cloud server collects summaries from a set of edge nodes and carries out Mini-Batch K-Means (MBK) clustering based on the MapReduce model for scalable, low-latency world-wide pattern extraction. This design improves scalability and decreases computation time with high clustering accuracy for huge volumes of healthcare data streams.

Initialization

The summaries (pre-processed features) from edge devices can be denotes as:

$$Z = \{z_1, z_2, z_3, \ldots z_S\},\ z_t \epsilon F^d \tag{5}$$

where S denotes the total number of summarized instances, and d denotes the feature dimension.

Where N denotes number of clusters; mb denotes mini-batch size.

Centroids at iteration i can be denoted as,

$$C^{(i)} = \left\{c_1^{(i)}, c_1^{(i)}, \ldots c_N^{(i)}\right\} \tag{6}$$

At initialization stage (for i=0), centroids or centres are randomly selected using below equation,

$$C^{(0)} = Initialize(Z, N) \tag{7}$$

Mini-Batch Sampling

The small subset(mini-batch) of data is taken randomly at every iteration (i≥1) without any replacement.

$$M_i = sample(Z, mb) \tag{8}$$

This randomization procedure makes stochastic exploration as well as computation more efficient, which makes it possible to cluster large datasets.

Assignment Step

For available centroids $C^{(i-1)}$, each data point is allocated to nearby centroid with the help of Euclidean distance.

$$b_t = \arg\min_{u\,\epsilon\{1,2,..N\}} z_t - c_u^{(i-1)2} \tag{9}$$

The above result generates partitions based on mini-batch cluster $(M_{i,u})$.

Update Step

Following that, the centroids are modified little by little according to the average of the new samples or members in every cluster:

$$C_u^{(i)} = (1 - \eta_u)C_u^{(i-1)} + \eta_u \overline{z_{t,u}} \tag{10}$$

$$z_{t,u} = \frac{1}{|M_{i,u}|} \sum_{z_t \epsilon M_{i,u}} z_t \tag{11}$$

Also, $\eta_u - \frac{M_{i,u}}{n_u^{(i)}}$ is the learning rate multiplied through the total number of samples $n_u^{(i)}$ for cluster u.

Convergence Criterion

The iterative update procedure keeps going on as long as the centroid shift between iterations is less than an acceptable threshold T.

$$C^{(i)} - C^{(i-1)}{}_F < T \tag{12}$$

In above equation, $\cdot_F$ represents the Frobenius norm. When everything is done, all of the observations in Z are put into their closest last centroids C^*.

Integrating MapReduce

The MapReduce structure executes the algorithm simultaneously such as this:

- Map Phase: Every one mapper node creates a small batch of data on its own on a data segment.
- Reduce Phase: By employing weighted averaging, reducer nodes combine partial centroids.

$$c_u^{(r)} = \frac{\sum_{m=1}^{M} w_{m,u} c_{m,u}}{\sum_{m=1}^{M} w_{m,u}} \tag{13}$$

The centroid from mapper m is denoted as $c_{m,u}$, and the sample weight is denoted using $w_{m,u}$. This distributed formulation speeds up convergence and makes sure that real-time healthcare analytics can work on a large scale in the cloud.

Convergence and Robustness

Similar to traditional K-Means, Mini-Batch K-Means remains sensitive to its initialization. Yet it converges faster and uses less memory, which makes it perfect for aggregating high-frequency IoT data. This method converges to a local optimal value, which may be different from traditional K-Means because of random sampling.

Mutli-objective Optimization

The three objectives are balanced by system such as accuracy, latency as well as energy which is given in eqn(14).

$$\text{Objective function}(G) = \alpha(1 - accuracy) + \beta(latency) + \gamma(energy) \tag{14}$$

Here, α, β, γ denotes the weights reflecting trade-offs.

The following are the algorithm for proposed FDBSCAN-MBKSched (Algorithm 1)

Results and Discussion

Experimental setup

The proposed FDBSCAN–MBKSched were tested on two datasets mentioner earlier: DigitalExposome (40 subjects; physiological + environmental signals) and PIF (wearable physiological + inertial signals, marked fall events). Experiments are performed on 1.5 GHz CPU, 1 GB RAM devices to mimic edge and a cloud cluster of 4 nodes for aggregation and clustering. Three baselines are used for comparison:

- Centralized DBSCAN + centralized model training, no federated learning.
- Standard FedAvg, with periodic transmission each round and no adaptive scheduling.
- FedProx-style federated learning (using proximal regularization) to address heterogeneity.

The following metrics are reported: classification accuracy, precision, recall, F1-score for emergency detection; anomaly detection latency; communication cost (total data communicated); clustering quality: Silhouette Score and Davies–Bouldin Index; and energy consumption (simulated on device model, in Joules per round).

Performance evaluation on emergency detection

The comparison of performance across various methods shows the substantial gain made by the proposed FDBSCAN–MBKSched framework. The proposed method made 93.4% accurate predictions, performing better than baseline methods like Centralized training (82.3%), FedAvg without scheduling (87.5%), and FedProx (89.1%). This indicates that integrating dynamic cluster-based scheduling with mini-batch k-means in a federated learning framework impacts model generalization over disparate datasets accurately.

Precision (92%) and Recall (93%) of the proposed FDBSCAN–MBKSched approach are always better than the baselines, which show better reliability in accurate identification of positive samples and reducing false negatives. Also, F1-Score (92.5%) validates an optimal balance between precision and recall, asserting that the FDBSCAN–MBKSched approach exhibits stable performance in all metrics.

FedAvg with no scheduling achieves better performance than local centralized training and demonstrates the benefit of federated aggregation. FedProx, targeting data heterogeneity, performs somewhat better than FedAvg but still falls short of the proposed FDBSCAN–MBKSched, showing the significance of effective client scheduling according to data clusters.

The bar chart in *Et al*ure 2 clearly shows that all the performance measures of the proposed approach are best among the methods compared, validating the efficiency and superiority of FDBSCAN–MBKSched. The enhancement is especially marked in F1-Score, which is of maximum importance in applications where balanced precision and recall are important.

Latency and False Alarm Rate during Anomaly Detection

The detection latency reduces significantly when anomaly detection is done at the edge of the network in place of a centralized environment. The designed Edge + Adaptive Scheduling achieves the least latency of 0.92 s, against 2.15 s with centralized DBSCAN and 1.40 s with Edge-DBSCAN without scheduling (Table 1). This enhancement is because of local processing at the edge, which removes the latency

Table 1: Comparison of latency and false alarm rate for different anomaly detection methods.

Method	*Detection Latency (s)*	*False Alarm Rate (%)*
Centralized DBSCAN	2.15	8.7
Edge-DBSCAN (no scheduler)	1.40	5.2
Edge + Adaptive Scheduling (proposed)	0.92	2.1

Algorithm 1: Overall workflow of proposed FDBSCAN-MBKSched model
Input: DigitalExposome, PIF datasets; window length (W); overlap (O); DBSCAN (ε, MinPts); FL and scheduler parameters (α, β, γ, τ_immediate, τ_delay); MBK parameters (N, mb, T).
Output: Global model (W_{global}), cloud centroids (C^*), emergency alerts.
Begin
Preprocessing:
Apply low-pass & median filters → denoise signals.
Normalize to zero mean and unit variance.
Segment into sliding windows (W, O).
Extract statistical, frequency & temporal features.
Label emergencies using thresholds (HR > 150 bpm, falls).
Edge-level Anomaly Detection (DBSCAN):
Compute quartiles $Q1$ to $Q3$ and I_Q.
$\varepsilon \aleph \ddot{u}_{\ddot{u}}$; MinPts = 2 × dimension.
Run DBSCAN on recent window.
If cluster = –1
send emergency alert;
else
store summary features.
Adaptive Federated Scheduling:
Each device trains local model (W_D).
Compute scheduling score using Eqn (3)

$$P_D = \alpha.norm\left(U_D\right) + \beta.norm\left(1 - S_D\right) + \gamma.norm\left(E_D\right)$$

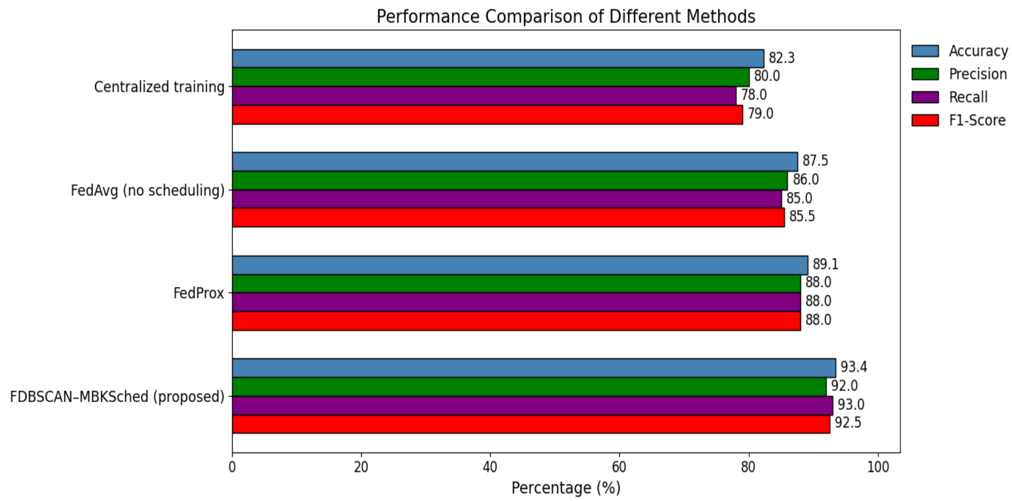
If $P_D \geq \tau_{immediate}$: send update;
else if $P_D \in (\tau_{delay-low}, \tau_{immediate})$: delay;
else: skip.
Federated Aggregation (Cloud):
Aggregate received updates using Eqn (4)

$$W = \frac{\sum_{k=1}^{K} n_k W_k}{\sum_{k=1}^{K} n_k}$$

Broadcast new global model to all devices.
Cloud Clustering (Mini-Batch K-Means + MapReduce):
Initialize centroids $C^{(0)}$ using Eqn (7)
Sample mini-batch M_i; assign each z_t to nearest centroid.
Update $C_u^{(i)}$ using eqn(10)
Map: local centroids → Reduce: weighted average combine.
Stop when $C^{(i)} - C^{(i-1)}{}_F < T$; output C^*.
Optimization:
Tune scheduler parameters (α, β, γ, τ) to balance clustering accuracy, latency, and energy.
Repeat Steps 2–6 until convergence of $w_{m,u}$ and $c_u^{(r)}$.
End

of the network, and the adaptive scheduling feature that gives precedence to urgent anomalies to indicate them right away without waiting for periodic update cycles.

The rate of false alarms is also drastically lowered, falling to 2.1% from 8.7% for Edge-DBSCAN with centralized scheduling and 5.2% for Edge-DBSCAN without any scheduling. Dynamic ε tuning, that appropriately modifies the sensitivity of clustering according to local data patterns, is responsible for this ~60–75% decrease, avoiding the labeling of noisy points as abnormalities. Edge processing as well as adaptive scheduling integrate to improve detection speed along with accuracy, making the framework ideal for real-time anomaly detection in resource-constrained environments. Overall, these results showed that edge computing and intelligent scheduling overcome the fundamental trade-off among detection speed as well as fidelity to provide a dual benefit: a faster response to anomalies and fewer false positives.



*Et al***ure 2:** Comparison of proposed method with baseline models on emergency detection

Communication and Energy Efficiency

The results (Table 2) highlight how successful the proposed FDBSCAN–MBKSched framework has been. Because many updates are delayed and omitted according to adaptive scheduling, the amount of information sent per round is reduced by at least 50% (in the range of 2.5 MB towards 1.2 MB). Compared to FedAvg and FedProx, which update continuously, energy consumption per device also drops by about 35–40% per round (from an average of 12.0–12.5 J towards 7.8 J). Such improvements demonstrated that intelligent scheduling and clustering in federated learning might effectively lower energy consumption and communication overhead without compromising the learning process, rendering the approach particularly effective for resource-constrained edge devices.

Cloud Clustering Quality

Clustering performance is assessed by two common metrics: Silhouette Score, which indicates the similarity of a point to its own cluster versus other clusters where higher range is preferable. The Davies-Bouldin Index, which measures cluster compactness and separation where lower range is desirable.

With a higher Silhouette Score (0.67) along with a lesser Davies-Bouldin Index (1.28), the proposed MBK-MapReduce method (as indicated in Table 3) works better than centralized K-Means (0.61, 1.48) and non-distributed Mini-Batch K-Means (0.63, 1.42). These results demonstrate stable clusters as well as more effectively cluster quality. Additionally, in comparison to centralized Mini-Batch K-Means, distributed MBK converges faster due to the parallel mapper and reducer setup, reducing computation time in cloud clustering about 40%. These findings prove MBK-MapReduce efficiently balances cluster quality and efficiency for large-scale, high-dimensional data set.

Table 2: Comparison of Communication and energy usage per device per round for various federated learning techniques.

Methods	*Data Sent per Round (MB)*	*Energy (J) per Device per Round*
FedAvg (every round)	2.5	12.0
FedProx	2.5	12.5
Proposed FDBSCAN–MBKSched	1.2	7.8

Table 3: Comparison of clustering quality for different K-Means approaches

Method	*Silhouette Score*	*Davies-Bouldin Index*
Centralized K-Means	0.61	1.48
Mini-Batch K-Means (non-distributed)	0.63	1.42
MBK-MapReduce (our method)	0.67	1.28

Discussion

The increased F1-score is due to two main reasons: (i) adaptive, edge-based anomaly detection using DBSCAN with dynamically calculated ε suppresses false negatives/positives, particularly in noisy physiological signals; and (ii) federated learning makes the global model leverage diverse local patterns without overfitting on any single participant.

Edge-based detection ensures emergencies are flagged immediately, bypassing cloud communication delay. Adaptive scheduling avoids needless communication under low urgency, which cuts energy and bandwidth costs. There may be a small overhead in edge computation (feature extraction, DBSCAN), but in our measurements this overhead (~0.3-0.5 s per window) is much smaller than the avoided network latency (~1-2 s) and energy cost.

Mini-Batch K-Means under MapReduce scales well, giving nearly the same Silhouette/Davies-Bouldin results as centralized clustering while reducing computation time.

Conclusion

With 93.4% accuracy along with the best precision, recall, and F1-score among every metric of assessment, the suggested FDBSCAN–MBKSched framework outperforms present federated learning techniques. By combining Mini-Batch K-Means aggregation, adaptive scheduling, and DBSCAN-based local clustering, the framework efficiently reduces data heterogeneity, speeds up convergence, and improves predictive stability. Furthermore, it is more efficient for resource-constrained edge devices than FedAvg and FedProx, reducing communication overhead by roughly 50% and device energy consumption by 35–40% per round. All things considered, FDBSCAN–MBKSched provides a scalable, energy-conscious, and reliable solution for actual IoMT-based federated healthcare systems. In order to further improve scalability and security in practical use cases, future research will concentrate on expanding the FDBSCAN–MBKSched approach to cross-silo along with multimodal IoMT environments, integrating privacy-preserving strategies like differential privacy and blockchain-driven trust systems.

Acknowledgements

We sincerely acknowledge the Head of the institution Dr. A. Venkatesan, Principal of the institution, for providing the facility to complete this paper successfully.

References

Alasmari, S., AlGhamdi, R., Tejani, G. G., Kumar Sharma, S., & Mousavirad, S. J. (2025). Federated learning-based multimodal approach for early detection and personalized care in cardiac disease. Frontiers in Physiology, 16. https://doi.org/10.3389/fphys.2025.1563185

Backhus, J.; Rao, A.R.; Venkatraman, C.; Gupta, C. Time Series Anomaly Detection Using Signal Processing and Deep Learning. Appl. Sci. 2025, 15, 6254. https://doi.org/10.3390/app15116254

Bai A, Song H, Wu Y, Dong S, Feng G, Jin H. Sliding-Window CNN + Channel-Time Attention Transformer Network Trained with Inertial Measurement Units and Surface Electromyography Data for the Prediction of Muscle Activation and Motion Dynamics Leveraging IMU-Only Wearables for Home-Based Shoulder Rehabilitation. Sensors (Basel). 2025 Feb 19;25(4):1275. doi: 10.3390/s25041275. PMID: 40006504; PMCID: PMC11861537.

Bushra AA, Kim D, Kan Y, Yi G. AutoSCAN: automatic detection of DBSCAN parameters and efficient clustering of data in overlapping density regions. PeerJ Comput Sci. 2024 Mar 14;10:e1921. doi: 10.7717/peerj-cs.1921. PMID: 38660211; PMCID: PMC11042006.

Choi, G., Cha, W. C., Lee, S. U., & Shin, S.-Y. (2024). Survey of medical applications of federated learning. Healthcare Informatics Research, 30(1), 3–15. https://doi.org/10.4258/hir.2024.30.1.3

Dhaliwal, Manpreet Kaur; Sharma, Rohini; Kaur, Rajbinder (2023), "Physiological Features and Inertial Features Based Dataset: PIFv3", Mendeley Data, V3, doi: 10.17632/phb9y6cp5c.3

Gold, O. C., & Lawrence, J. (2024). Enhanced LSTM for heart disease prediction in IoT-enabled smart healthcare systems. *The Scientific Temper*, *15*(02), 2238–2247.

Hicks, S. C., Liu, R., Ni, Y., Purdom, E., & Risso, D. (2021). mbkmeans: Fast clustering for single cell data using mini-batch k-means. PLoS Computational Biology, 17(1), e1008625. https://doi.org/10.1371/journal.pcbi.1008625

Jenifer, R. R., & Janita, V. S. (2025). Energy-aware Security Optimized Elliptic Curve Digital Signature Algorithm for Universal IoT Networks. *The Scientific Temper*, *16*(09), 4745–4761.

Johnson, T. (2025, January 30). DigitalExposome: A dataset for wellbeing classification using environmental air quality and human physiological data. Mendeley Data. https://data.mendeley.com/datasets/mbwxy48223/1

K. Mohamed Arif Khan, & A.R. Mohamed Shanavas. (2024). Optimizing IoT application deployment with fog - cloud paradigm: A resource-aware approach. *The Scientific Temper*, *15*(04), 3225–3233. https://doi.org/10.58414/SCIENTIFICTEMPER.2025.16.2.11

Li, L., Zhao, H., & Liu, N. (2023). MCD-Yolov5: Accurate, Real-Time Crop Disease and Pest Identification Approach Using UAVs. Electronics, 12(20), 4365. https://doi.org/10.3390/electronics12204365

Li, L.; Zhu, L.; Li, W. Cloud–Edge–End Collaborative Federated Learning: Enhancing Model Accuracy and Privacy in Non-IID Environments. Sensors 2024, 24, 8028. https://doi.org/10.3390/s24248028

Liu, H., Liu, Y., Qin, Z., Zhang, R., Zhang, Z., & Mu, L. (2021). A novel DBSCAN clustering algorithm via edge computing-based deep neural network model for targeted poverty alleviation big data. Wireless Communications and Mobile Computing, 2021(1). https://doi.org/10.1155/2021/5536579

Maciel, F., de Souza, A. M., Bittencourt, L. F., Villas, L. A., & Braun, T. (2024). Federated learning energy saving through client selection. Pervasive and Mobile Computing, 103(101948), 101948. https://doi.org/10.1016/j.pmcj.2024.101948

Nayak, S., Patgiri, R., Waikhom, L., & Ahmed, A. (2024). A review on edge analytics: Issues, challenges, opportunities, promises, future directions, and applications. Digital Communications and Networks, 10(3), 783-804. https://doi.org/10.1016/j.dcan.2022.10.016

Nguyen, D. C., Pham, Q.-V., Pathirana, P. N., Ding, M., Seneviratne, A., Lin, Z., Dobre, O., & Hwang, W.-J. (2023). Federated Learning for smart healthcare: A survey. ACM Computing Surveys, 55(3), 1–37. https://doi.org/10.1145/3501296

Pinto, R. P., Silva, B. M. C., & Inácio, P. R. M. (2025). Federated learning for anomaly detection on Internet of Medical Things: A survey. Internet of Things (Amsterdam, Netherlands), 101677, 101677. https://doi.org/10.1016/j.iot.2025.101677

Prasad, V. K., Bhattacharya, P., Maru, D., Tanwar, S., Verma, A., Singh, A., Tiwari, A. K., Sharma, R., Alkhayyat, A., Ţurcanu, F.-E., & Raboaca, M. S. (2022). Federated learning for the Internet-of-Medical-Things: A survey. Mathematics, 11(1), 151. https://doi.org/10.3390/math11010151

Ragul, M., Aloysius, A., & Kumar, V. A. (2025). Enhancing IoT blockchain scalability through the eepos consensus algorithm. *The Scientific Temper*, *16*(02), 3698–3709. https://doi.org/10.58414/SCIENTIFICTEMPER.2025.16.2.02

Rani, S., Kataria, A., Kumar, S., & Tiwari, P. (2023). Federated learning for secure iomt-applications in smart healthcare systems: A comprehensive review. Knowledge-Based Systems, 274, 110658. https://doi.org/10.1016/j.knosys.2023.110658

Retiti Diop Emane, C.; Song, S.; Lee, H.; Choi, D.; Lim, J.; Bok, K.;

Yoo, J. Anomaly Detection Based on GCNs and DBSCAN in a Large-Scale Graph. Electronics 2024, 13, 2625.

Sarkar, A., & Vajpayee, L. (2024). Augmenting the FedProx algorithm by minimizing convergence. In arXiv [cs.LG]. http://arxiv.org/abs/2406.00748

Shao, N., & Chen, Y. (2022). Abnormal data detection and identification method of distribution Internet of Things monitoring terminal based on spatiotemporal correlation. Energies, 15(6), 2151. https://doi.org/10.3390/en15062151

Sindhu S., & L. Arockiam. (2024). DRMF: Optimizing machine learning accuracy in IoT crop recommendation with domain rules and MissForest imputation. *The Scientific Temper*, *15*(03), 2570–2578. https://doi.org/10.58414/SCIENTIFICTEMPER.2024.15.3.24

Singh, A.K., Krishnan, S. ECG signal feature extraction trends in methods and applications. BioMed Eng OnLine 22, 22 (2023). https://doi.org/10.1186/s12938-023-01075-1

Y. Li, X. Qin, K. Han, N. Ma, X. Xu and P. Zhang, "Accelerating Wireless Federated Learning With Adaptive Scheduling Over Heterogeneous Devices," in IEEE Internet of Things Journal, vol. 11, no. 2, pp. 2286-2302, 15 Jan.15, 2024, doi: 10.1109/JIOT.2023.3292494.

Zhou, J., Sun, Y., & Tellambura, C. (2025). Revolutionizing medical data transmission with IoMT: A comprehensive survey of wireless communication solutions and future directions. In arXiv [cs.IT]. http://arxiv.org/abs/2504.02446