Advances in Artificial Intelligence and Machine Learning; Research (Ahead of Print) Received 14-06-2026; Accepted 12-09-2026; Published online 19-09-2026

Privacy-Preserving Cross-Domain Federated Learning for Real-Time Traffic Anomaly Detection

Ahmed A. A. Gad-Elrab aaahmad4@kau.edu.sa
Department of Computer Science,
Faculty of Computing and Information Technology
King Abdulaziz University, Jeddah, Saudi Arabia

Department of Mathematics, Computer Science Branch, Faculty of Science
Al-Azhar University, Nasr, Cairo 11651, Egypt

Mohammed Ibrahim Altwijri maltwijri@kau.edu.sa
Department of Computer Science,
Faculty of Computing and Information Technology
King Abdulaziz University, Jeddah, Saudi Arabia

Corresponding Author: Ahmed A. A. Gad-Elrab.

Abstract

Smart city traffic operations require learning from data generated by multiple cities agencies and sensing platforms under strict privacy and latency constraints. Centralized training is impractical due to regulatory limits bandwidth cost and operational risk while existing federated learning methods struggle with cross domain shift real time data streams and static privacy control. This paper proposes a privacy preserving cross domain federated learning framework for real time traffic anomaly detection that integrates federated transfer learning adaptive differential privacy and streaming model updates within a unified architecture. Regional clients train locally on UAV and UGV sensor or video data and share privacy controlled model updates for global aggregation without exposing raw data. A domain adaptation mechanism enables zero shot generalization to new cities with unseen traffic patterns while an adaptive privacy budget dynamically balances data sensitivity trust level and update frequency against model utility. Online inference and incremental learning support early detection of congestion incidents and abnormal traffic events under live traffic streams. The cross-city traffic deployment scenario drives the architectural design, and experimental assessment under controlled non-IID federated conditions shows increased learning performance and resource efficiency compared with the studied federated baselines.

Keywords: Privacy preserving, Machine learning, Real time anomaly detection, Intelligent transportation systems.

Citation: Ahmed A. A. Gad-Elrab and Mohammed Ibrahim Altwijri. Privacy-Preserving Cross-Domain Federated Learning for Real-Time Traffic Anomaly Detection. **Advances in Artificial Intelligence and Machine Learning**. 2026. (Ahead of Print) https://dx.doi.org/10.54364/AAIML.2026.65346

1. INTRODUCTION

Urban traffic systems generate continuous data streams from heterogeneous sources including road-side sensors UAVs and UGVs. These data are critical for congestion monitoring incident detection and real time operational decision making. Centralized learning approaches require large scale data transfer and expose sensitive mobility patterns which conflicts with privacy regulations and operational constraints. As a result centralized training remains difficult to deploy in real city scale environments.

Federated learning enables collaborative model training without raw data sharing by keeping data local and exchanging model updates only. It has been thoroughly researched in a variety of application fields and has become a viable, privacy-conscious substitute for centralized and distributed learning [1]. Federated learning facilitates collaboration across several agencies or regions while limiting data exposure and bandwidth consumption in smart city scenarios. However, the usefulness of traditional federated learning for real-time traffic monitoring is limited since it relies on reasonably stable data distributions and frequent batch updates.

Cross silo federated learning extends the federated paradigm to settings with a small number of reliable participants such as cities institutions or organizations. This setting is well aligned with inter city collaboration but still suffers from slow adaptation under non IID data and limited support for streaming or real time updates [1]. In addition most existing approaches rely on static privacy budgets which either degrade model accuracy unnecessarily or fail to protect sensitive data when data importance and trust levels vary over time.

To reduce communication overhead several communication efficient federated optimization methods have been proposed. Federated Block Coordinate Descent introduces multiple local updates over vertically partitioned features before synchronization and achieves significant communication savings [2]. Despite these advantages such methods are designed for batch learning on static datasets. They do not address cross domain shift across cities nor the requirements of real time anomaly detection.

Recent studies have applied federated learning to intelligent traffic monitoring systems demonstrating improvements in privacy protection communication efficiency and energy management [3]. However, most methods do not offer continuous learning from real-time data streams and usually concentrate on offline training cycles with fixed objectives or incentive mechanisms. Dynamic privacy control and zero shot adaptability to new cities are yet mostly unexplored.

In order to fill these shortcomings, this paper suggests a cross-domain federated learning system for real-time traffic anomaly detection that preserves privacy. To enable continuous learning across cities without exchanging raw data, the architecture incorporates streaming model updates, federated transfer learning, and adaptive differential privacy. The suggested method allows zero shot deployment in new areas while preserving real-time detection performance under stringent privacy restrictions by fusing domain adaptation with adaptive privacy management.

The main contributions of this work are summarized as follows:

- we provide a cross-domain federated learning system that protects privacy for real-time traffic anomaly detection across several smart cities without exchanging raw traffic data.
- we combine streaming federated updates, adaptive differential privacy, and federated transfer learning into a single architecture to enable continuous learning in diverse traffic environments
- we present a common encoder and city-specific anomaly detection heads to facilitate efficient cross-city knowledge transfer while maintaining local traffic features.
- We design an adaptive privacy technique that improves the privacy-utility trade-off by dynamically modifying the privacy budget based on data sensitivity, trust level, and update frequency.
- we design a federated learning approach that is communication-efficient and shares only protected transferable representations to minimize communication overhead without sacrificing model performance.
- we show that the suggested framework outperforms representative federated learning baselines in terms of accuracy, scalability, and resource efficiency through comprehensive simulations.

It is important to distinguish this work's contribution from suggesting a new stand-alone federated optimizer or differential privacy technique. The system-level articulation of cross-domain transfer, adaptive privacy management, and streaming federated updates as a coordinated learning process are what make it new. In particular, the system breaks down each regional model into a private city-specific detection head and a globally transferable encoder. Local heads maintain domain-specific behavior, whereas only clipped and privacy-perturbed encoder updates take part in global aggregation. This division restricts the information disclosed during federation and combines local expertise with cross-domain knowledge sharing. Additionally, rather of repeatedly processing the entire historical dataset, the framework works with recent streaming windows. The joint operational formulation of transfer, privacy adaptation, local specialization, and streaming updates for heterogeneous cross-domain federated systems is thus the contribution.

Technically, the suggested system is different from traditional full-model federated aggregation in that it keeps domain-specific detection heads locally while limiting global exchange to the transferable encoder. Therefore, rather than affecting the entire client model, privacy perturbation is applied to the shared representation update. This parameter separation is coupled with streaming-window optimization and a privacy controller whose degree of protection is determined by update frequency, sensitivity, and trust. Instead of replacing FedAvg, differential privacy, or transfer learning separately, the resultant contribution is a coordinated cross-domain federated formulation.

2. RELATED WORK

Motivated by privacy preservation and regulatory compliance, federated learning has arisen as a decentralized learning paradigm that allows collaborative model training without sharing raw data. According to early research, statistical heterogeneity, communication overhead, and privacy leakage are the main obstacles to federated learning [1, 4]. By permitting many local updates

prior to aggregation, Federated Averaging (FedAvg) improved communication efficiency in sample-partitioned environments and established a common baseline [5]. FedAvg experiences reduced convergence for non-IID data and heterogeneous systems, according to subsequent analyses [6, 7].

Situations with a small number of reliable and consistent participants, like hospitals, businesses, or local authorities, are addressed by cross-silo federated learning. Although this configuration is consistent with inter-city smart transportation systems, it still uses static objectives and synchronous, round-based optimization [1, 8]. Because of this, cross-silo FL is not very flexible when it comes to changing data distributions and domain changes that are typical of real-world traffic situations.

In federated optimization, communication efficiency continues to be a key area of study. In addition to FedAvg, methods like variance-aware local updates and proximal federated optimization seek to stabilize training under heterogeneity [6, 9]. Distributed coordinate descent and block coordinate descent techniques decouple optimization across feature subsets in vertically partitioned or feature-partitioned contexts [10, 11]. These concepts are extended to privacy-aware collaborative learning by Federated Block Coordinate Descent (FedBCD), which allows for numerous local block updates with lower synchronization frequency and theoretical convergence guarantees [2]. FedBCD assumes batch learning on static datasets and ignores real-time inference and streaming data, despite its communication efficiency and support for federated transfer learning.

Using secure aggregation and cryptographic methods, privacy protection in federated learning has been thoroughly investigated. Although they prohibit the publication of individual updates, secure aggregation techniques have significant computational and communication cost [12]. Although secure multi-party computation and homomorphic encryption offer robust privacy guarantees, they are frequently unsuitable for latency-sensitive applications [13, 14]. Although previous research demonstrates that fixed privacy budgets cause needless accuracy erosion in long-running or continuously updated systems [15, 16], differential privacy provides formal privacy guarantees by disturbing updates or gradients.

Federated transfer learning has been suggested as a solution to participant cross-domain variability and limited data overlap. Early research shows that it works well in situations where there are few labels or mismatched feature spaces [2, 17]. However, the majority of current approaches rely on offline training and episodic knowledge transfer, which restricts their use in dynamic settings.

Recently, intelligent transportation and traffic monitoring systems have used federated learning. Using car, roadside, UAV, and UGV data, previous research investigates FL-based traffic flow prediction and anomaly detection, showing better privacy protection and lower communication costs as compared to centralized methods [18, 19]. In order to increase convergence speed and energy efficiency in air-ground traffic monitoring systems, more recent research incorporates Bayesian prediction, client selection, and incentive mechanisms [3]. However, these methods do not offer zero-shot deployment across cities, adaptive privacy control, or continuous learning; instead, they mainly concentrate on offline or episodic training.

Strong foundations in federated optimization, cross-silo collaboration, communication efficiency, privacy preservation, and traffic monitoring are provided by existing research. Nevertheless, cross-domain shift, real-time streaming data, and dynamic privacy constraints are still difficult to jointly handle with present approaches. By combining federated transfer learning, adaptive differential

Table 1: Qualitative comparison between representative existing studies and the proposed framework

Method	Cross-domain adaptation	Privacy mechanism	Streaming / real-time FL	Traffic anomaly detection	Communication efficiency	Zero-shot transfer	Main limitation
FedAvg [5]	No	Basic federated training	No	No	Moderate	No	Sensitive to non-IID data and not designed for cross-domain real-time traffic settings
Local SGD [6]	No	Basic federated training	No	No	High	No	Improves communication efficiency but does not address privacy adaptation or cross-city deployment
FedProx [9]	No	Basic federated training	No	No	Moderate	No	Improves robustness under heterogeneity but still assumes static batch learning
Cross-silo FL frameworks [1, 8]	Limited	General FL privacy support	No	No	Moderate	No	Supports stable institutional collaboration but lacks explicit real-time and anomaly-aware mechanisms
FedBCD [2]	Partial	No raw data or model parameter sharing	No	No	High	Limited	Designed for feature-partitioned batch optimization rather than live traffic streams
Federated transfer learning [2, 17]	Yes	Depends on implementation	No	No	Moderate	Partial	Handles domain mismatch but is typically studied in offline settings
Secure aggregation based FL [12]	No	Secure aggregation	No	No	Moderate	No	Provides stronger privacy with added communication and computation overhead
Differentially private FL [15, 16]	No	Differential privacy	No	No	Moderate	No	Fixed privacy budgets can degrade utility in long-running learning systems
Traffic flow FL [18]	Limited	FL-based privacy preservation	Limited	No	Moderate	No	Focuses on traffic prediction rather than anomaly detection and cross-city adaptation
Privacy-preserving traffic monitoring FL [19]	Limited	Privacy-preserving FL	Limited	Partial	Moderate	No	Does not incorporate adaptive privacy control or streaming federated updates
Energy-efficient traffic monitoring FL [3]	No	FL-based privacy preservation	Partial	Partial	High	No	Focuses on client selection and incentives rather than transfer learning and adaptive privacy
Proposed method	**Yes**	**Adaptive differential privacy**	**Yes**	**Yes**	**High**	**Yes**	**Unifies cross-domain transfer, adaptive privacy control, and real-time anomaly detection for operational smart city traffic systems**

privacy, and streaming federated updates into a single framework for real-time traffic anomaly detection across cities, this work expands upon these baselines.

The key distinctions between representative previous research and the suggested framework are outlined in TABLE 1. Cross-domain transfer, real-time streaming updates, and adaptive privacy control are not addressed by early federated techniques like FedAvg, Local SGD, and FedProx,

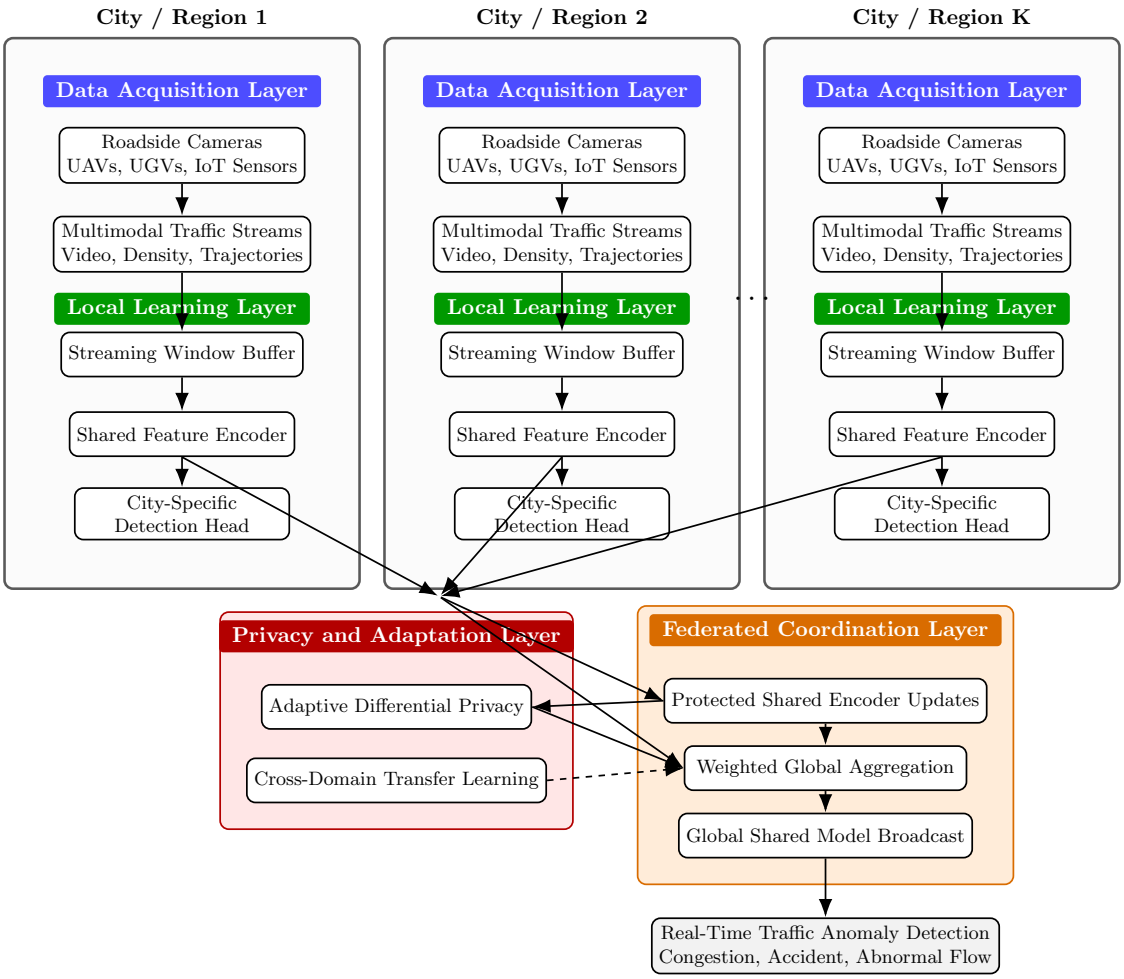

Figure 1: A cross-domain federated learning architecture that protects privacy is suggested for real-time traffic anomaly detection. Every city uses multimodal traffic data for streaming local training, applies adaptive differential privacy prior to upload, and communicates only protected updates of the shared representation. For continuous real-time anomaly detection, the coordination server transmits the updated representation after compiling shared information among cities.

which primarily focus on communication-efficient decentralized optimization [5, 6, 9]. Cross-silo federated learning provides a suitable collaboration setting for smart city agencies, yet existing frameworks remain largely round-based and static [1, 8]. FedBCD improves communication efficiency in feature-partitioned scenarios and supports collaborative learning without raw data exchange, but it is still designed for batch learning rather than live anomaly detection [2]. In intelligent transportation applications, prior works on traffic-focused federated learning improve privacy, energy efficiency, and client participation, but they do not jointly support cross-city adaptation, adaptive privacy budgeting, and continuous learning from live streams [3, 18, 19]. In contrast, the proposed method integrates cross-domain federated transfer learning, adaptive differential privacy, and streaming federated updates in a unified framework, enabling zero-shot generalization across cities and real-time traffic anomaly detection under operational constraints.

The comparison shows that the suggested framework's uniqueness is not predicated on any one element alone. Federated optimization is addressed by FedAvg and FedProx, communication-efficient collaborative optimization is emphasized by FedBCD, domain mismatch is addressed by federated transfer learning, and information is protected by differentially private FL. Additionally, application-specific cooperation is addressed in traffic-oriented FL studies. The suggested approach combines streaming-window learning, adaptive protection of shared updates, private domain-specific heads, and transferable shared representations to address the confluence of these criteria. Therefore, rather than implying that the various constituent processes are themselves newly introduced, the comparison should be understood at the architectural and operational levels.

3. PROPOSED ARCHITECTURE AND SYSTEM MODEL

This section describes the system architecture and the operational model of the proposed framework for privacy-preserving cross-domain federated learning in real-time traffic anomaly detection. The framework integrates cross-silo federated learning, federated transfer learning, adaptive differential privacy, and streaming data processing to enable collaborative traffic monitoring across multiple cities without exposing raw data.

3.1 System Architecture

The proposed architecture consists of multiple regional traffic monitoring nodes connected through a federated coordination server. Each region represents a traffic authority or smart city control center that collects traffic observations from heterogeneous sensing platforms including roadside cameras, UAVs, UGVs, and IoT traffic sensors. Instead of sharing raw traffic data, each region performs local training and periodically exchanges privacy-preserving model updates with the global server. FIGURE 1 illustrates the architecture of the proposed system. The architecture consists of four main layers:

1) Data Acquisition Layer. Road cameras, car sensors, UAV surveillance feeds, and ground-based mobile sensors are just a few of the diverse sensing sources from which traffic data is regularly gathered. These sources produce multimodal streams that include ambient signals, vehicle trajectories, traffic density measures, and video frames. Every participating city or organization keeps their data locally.

2) Local Learning Layer. A local anomaly detection algorithm that learns from locally observed traffic patterns is kept up to date by each regional node. Deep neural networks or hybrid architectures that can handle spatiotemporal traffic data may be used in the model. Batches of incoming streaming data are used for local training. In order to identify anomalous traffic behavior, such as spikes in traffic, collisions, or strange vehicle movements, the local model extracts traffic representations.

3) Federated Coordination Layer. Using a federated aggregation method akin to federated averaging, a global coordination server compiles model changes from participating areas. A global model that depicts common traffic patterns among cities is created during the aggregation stage. To enhance local model performance while maintaining data locality, the global model is subsequently redistributed to each participant.

4) Privacy and Adaptation Layer. Adaptive differential privacy is used to safeguard sensitive mobility data in model changes. The sensitivity of the local dataset, participant trust, and update frequency all influence the privacy budget's dynamic adaptation. This layer also includes a cross-domain adaptation module that enables knowledge transfer between cities with different traffic distributions.

3.2 System Model

Consider a federation of K participating regions or cities. Each region k maintains a local dataset D_k consisting of traffic observations collected from local sensing infrastructure. The data may include heterogeneous modalities such as sensor measurements, video features, and mobility statistics. Let $D_k = \{(x_i^k, y_i^k)\}_{i=1}^{n_k}$ denote the local dataset at region k, where x_i^k represents the traffic observation and y_i^k represents the anomaly label or target signal. Due to privacy and regulatory constraints, the raw datasets $\{D_k\}$ cannot be shared among regions. The goal of federated learning is to collaboratively train a global model with parameters w that minimizes the overall loss across all regions

$$\min_w F(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w) \tag{1}$$

where $n = \sum_{k=1}^{K} n_k$ and $F_k(w)$ denotes the local objective function for region k

$$F_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} \ell(w; x_i^k, y_i^k) \tag{2}$$

Here $\ell(\cdot)$ represents the anomaly detection loss function such as cross-entropy or reconstruction error depending on the detection model. During training, each regional client performs local updates using its private dataset and periodically sends model updates to the central aggregator. Let w_t denote the global model at communication round t. Each client updates its local parameters through several local iterations

$$w_k^{t+1} = w_t - \eta \nabla F_k(w_t) \tag{3}$$

where η is the learning rate. The server aggregates the local updates to obtain the new global model

$$w_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} w_k^{t+1} \tag{4}$$

3.3 Cross-Domain Transfer Learning

Due to variations in infrastructure, vehicle behavior, sensing conditions, and environmental factors, traffic patterns fluctuate significantly between cities. The suggested system uses federated transfer learning to overcome this difficulty. A domain-specific adaption layer and a shared representation layer are featured in every local model. While the domain-specific layer learns geographical variances, the shared representation retains common traffic features. The model may transfer knowledge from previously trained cities to newly participating locations thanks to this approach. Consequently, zero-shot or few-shot deployment in new cities is supported by the system without requiring a significant amount of retraining.

3.4 Adaptive Privacy Mechanism

Differential privacy is used to disturb model updates in order to protect privacy during collaborative training. The suggested architecture replaces a fixed privacy budget with an adaptive privacy mechanism that dynamically modifies the privacy parameter ϵ based on participant trust levels and data sensitivity. Let Δ be the model updates' sensitivity. The privacy-preserving update is generated by adding calibrated noise

$$\tilde{w}_k = w_k + \mathcal{N}(0, \sigma^2) \tag{5}$$

where the adaptive privacy budget determines σ. Particularly in long-running real-time learning systems, this approach strikes a balance between model accuracy and privacy protection.

3.5 Real-Time Streaming Learning

Because traffic situations are so dynamic, constant observation is necessary. The suggested architecture analyzes streaming traffic data and gradually changes models to enable real-time anomaly detection. The method uses newly observed data windows to execute periodic federated updates rather than retraining models from start. The system can adjust to changing traffic circumstances and quickly identify anomalies like accidents, congestion bursts, and unusual mobility patterns thanks to its streaming federated learning process. Overall, the suggested architecture enables scalable and privacy-preserving traffic anomaly detection across several smart cities by combining streaming analytics, adaptive privacy management, and cross-domain federated learning.

4. PROPOSED PRIVACY-PRESERVING CROSS-DOMAIN FEDERATED LEARNING FRAMEWORK

The suggested system uses a hierarchical cross-silo federated design, as seen in FIGURE 1, where each city shares only privacy-protected updates of the transferable representation module while performing local streaming learning on multimodal traffic data. By separating city-specific anomaly detection behavior from globally shared information, this architecture enables the model to gen-

eralize across varied traffic domains without disclosing raw observations. Compared to fixed-budget techniques, the adaptive differential privacy component offers a more flexible privacy-utility trade-off by further adjusting protection strength based on data sensitivity, trust level, and update frequency.

4.1 Problem Formulation

Imagine a federation of regions with K smart cities. Each region $k \in \{1, \ldots, K\}$ owns a local streaming traffic dataset

$$\mathcal{D}_k = \{(x_{k,i}, y_{k,i})\}_{i=1}^{N_k},$$

where $x_{k,i}$ denotes a multimodal traffic observation collected from local sensing infrastructure such as UAVs, UGVs, roadside cameras, and IoT sensors, and $y_{k,i}$ denotes the corresponding anomaly label. Due to privacy, regulatory, and operational constraints, raw data cannot be exchanged across regions. The proposed model consists of a shared encoder $h_\phi(\cdot)$ and a client-specific anomaly detector $g_{\theta_k}(\cdot)$. For a given input $x_{k,i}$, the prediction is defined as

$$\hat{y}_{k,i} = g_{\theta_k}(h_\phi(x_{k,i})).$$

The shared encoder captures domain-invariant traffic representations across cities, while the local head preserves city-specific traffic characteristics. At communication round t, client k receives a recent streaming window

$$\mathcal{W}_k^{(t)} = \{(x_{k,i}^{(t)}, y_{k,i}^{(t)})\}_{i=1}^{m_k^{(t)}},$$

where $m_k^{(t)}$ is the number of samples observed during the current time interval. Instead of retraining on the full historical dataset, the local objective is computed on the current streaming window:

$$\mathcal{L}_k^{(t)}(\phi, \theta_k) = \frac{1}{m_k^{(t)}} \sum_{i=1}^{m_k^{(t)}} \ell\Big(g_{\theta_k}(h_\phi(x_{k,i}^{(t)})), y_{k,i}^{(t)}\Big). \tag{6}$$

4.2 Cross-Domain Representation Learning

To improve generalization across heterogeneous traffic domains, the local objective is augmented with a domain regularization term. The overall local objective becomes

$$\mathcal{J}_k^{(t)}(\phi, \theta_k) = \mathcal{L}_k^{(t)}(\phi, \theta_k) + \lambda_d \mathcal{L}_{\mathrm{dom},k}^{(t)}(\phi) + \lambda_r \|\theta_k\|_2^2, \tag{7}$$

where λ_d and λ_r are balancing coefficients. The term $\mathcal{L}_{\mathrm{dom},k}^{(t)}(\phi)$ encourages the shared encoder to learn transferable representations across cities. Depending on implementation, this term can be instantiated as a feature-alignment loss, discrepancy minimization loss, or adversarial domain confusion objective. The collaborative objective across all clients is thus

$$\min_{\phi, \{\theta_k\}_{k=1}^K} \mathcal{F}^{(t)}(\phi, \{\theta_k\}) = \sum_{k=1}^{K} \omega_k^{(t)} \mathcal{J}_k^{(t)}(\phi, \theta_k), \tag{8}$$

where

$$\omega_k^{(t)} = \frac{m_k^{(t)}}{\sum_{j=1}^{K} m_j^{(t)}}$$

denotes the aggregation weight at round t.

Instead of directly exchanging domain-specific classifiers, the suggested approach uses a shared encoder to facilitate cross-domain transfer. While each local detection head is kept private during federated training, encoder parameters are tuned using data from participating domains and then aggregated. As a result, the shared parameter set ϕ represents transferable information, whereas θ_k maintains client-specific decision characteristics. To lessen domain-specific dependence in the encoder space, the domain regularization term $L_{\mathrm{dom},k}$ is only connected to the shared representation. Deploying the learnt shared representation to an unknown target domain without utilizing target-domain samples to update the shared encoder prior to beginning inference is referred to in this formulation as zero-shot transfer. Few-shot adaptation is the process of employing a small amount of target-domain data to update the target-specific detection component later on. The shared-encoder/local-head architecture's supported deployment options are referred to as few-shot and zero-shot. The non-IID client experiment described in Section 5 should not be confused with these. The ability of federated training to function with distributional disparities is evaluated in the current evaluation using heterogeneous client partitions. A dedicated held-out-city protocol is different from the controlled benchmark under consideration here since it would need a target traffic domain that is not included in federated training.

4.3 Local Update Rule

At round t, each client initializes the shared encoder using the global parameters $\phi^{(t)}$ and updates the shared and local parameters through stochastic gradient descent:

$$\phi_k^{(t,e+1)} = \phi_k^{(t,e)} - \eta \nabla_{\phi} \mathcal{J}_k^{(t)}(\phi_k^{(t,e)}, \theta_k^{(t,e)}), \tag{9}$$

$$\theta_k^{(t,e+1)} = \theta_k^{(t,e)} - \eta \nabla_{\theta_k} \mathcal{J}_k^{(t)}(\phi_k^{(t,e)}, \theta_k^{(t,e)}), \tag{10}$$

for $e = 0, \ldots, E-1$, where E is the number of local epochs. After local training, only the shared encoder update is uploaded:

$$\Delta\phi_k^{(t)} = \phi_k^{(t,E)} - \phi^{(t)}. \tag{11}$$

The local detector head θ_k remains private and is not shared.

4.4 Adaptive Differential Privacy

To protect mobility-sensitive information, each shared update is first clipped:

$$\bar{\Delta\phi}_k^{(t)} = \Delta\phi_k^{(t)} \cdot \min\left(1, \frac{C}{\|\Delta\phi_k^{(t)}\|_2}\right), \tag{12}$$

where C is the clipping threshold. Next, client k applies Gaussian perturbation:

$$\widetilde{\Delta\phi}_k^{(t)} = \bar{\Delta\phi}_k^{(t)} + \mathcal{N}(0, \sigma_k^{2(t)} I). \tag{13}$$

Unlike fixed-budget mechanisms, the proposed framework uses an adaptive privacy controller:

$$\epsilon_k^{(t)} = \Psi\left(s_k^{(t)}, \tau_k^{(t)}, \rho_k^{(t)}\right), \tag{14}$$

where $s_k^{(t)}$ denotes data sensitivity, $\tau_k^{(t)}$ denotes trust level, and $\rho_k^{(t)}$ denotes update frequency. The noise scale $\sigma_k^{(t)}$ is then determined from $\epsilon_k^{(t)}$ and the clipping threshold. This adaptive design allows stricter privacy when data are highly sensitive and lower noise when higher utility is required. Three operational elements are mapped to the privacy parameter utilized for the current client update by the controller Ψ, which should be seen as a policy. The update-frequency term reflects recurrent exposure brought on by frequent communication, the sensitivity term represents the privacy relevance of the current information, and the trust term indicates the predetermined confidence connected with the participating client. Stronger protection is favored by higher sensitivity or more frequent updates, while the trust component allows the policy to distinguish between participating entities. The contribution of a single update is limited prior to noise injection because the encoder update is confined by the clipping threshold C. The Gaussian noise scale used to the clipped shared update is then determined using the resulting privacy parameter. As a result, adaptation maintains the same clip-then-perturb processing pipeline while varying the protection level across clients and communication cycles. Also, the operational privacy budget consumption utilized by the simulator is reflected in the privacy cost provided in this evaluation. It is not a full cumulative (ϵ,δ)-DP accounting guaranty throughout all communication cycles. Establishing such a guaranty is outside the purview of the current simulation-based evaluation and necessitates precise composition accounting for recurring customer engagement.

4.5 Global Aggregation

The coordination server aggregates the protected shared updates as

$$\phi^{(t+1)} = \phi^{(t)} + \sum_{k=1}^{K} \omega_k^{(t)} \widetilde{\Delta\phi}_k^{(t)}. \tag{15}$$

Because only the shared encoder is aggregated, the global model captures cross-city transferable knowledge while preserving local traffic specialization in the private heads $\{\theta_k\}$.

4.6 Real-Time Inference and Streaming Adaptation

During operation, each regional node uses its local detection head and the current shared encoder to forecast abnormalities from incoming observations:

$$\hat{y}_k^{(t)} = g_{\theta_k^{(t)}}(h_{\phi^{(t)}}(x_k^{(t)})). \tag{16}$$

New streaming windows are used to refresh the model on a regular basis, allowing for continual adaptation to changing traffic situations such incidents, road closures, atypical mobility behavior, and congestion build-up. In operational smart city scenarios, this formulation results in a single framework that concurrently solves low-latency anomaly detection, privacy-preserving collaboration, and cross-domain transfer.

Algorithm 1 Adaptive Privacy-Preserving Cross-Domain Federated Learning

Require: Number of clients K, communication rounds T, local epochs E, local batch size B, learning rate η

Require: Initial shared encoder parameters $\phi^{(0)}$, local detector heads $\{\theta_k^{(0)}\}_{k=1}^{K}$

Require: Privacy controller $\Psi(\cdot)$, clipping threshold C

for $t = 0$ to $T-1$ **do**

  Server broadcasts current shared encoder $\phi^{(t)}$ to all participating clients

  for all clients $k \in \{1, \ldots, K\}$ in parallel **do**

    Receive streaming traffic window $W_k^{(t)}$

    Initialize local shared parameters $\phi_k^{(t,0)} \leftarrow \phi^{(t)}$

    Keep local head $\theta_k^{(t,0)}$

    for $e = 1$ to E **do**

      for all mini-batches $\mathcal{B} \subset W_k^{(t)}$ of size B **do**

        Compute local loss

$$\mathcal{L}_k(\phi, \theta_k) = \mathcal{L}_{\text{det}}(\phi, \theta_k) + \lambda_d \mathcal{L}_{\text{dom}}(\phi) + \lambda_r \|\theta_k\|_2^2$$

        Update local shared encoder and local head

$$\phi_k^{(t,e)} \leftarrow \phi_k^{(t,e-1)} - \eta \nabla_{\phi} \mathcal{L}_k$$

$$\theta_k^{(t,e)} \leftarrow \theta_k^{(t,e-1)} - \eta \nabla_{\theta_k} \mathcal{L}_k$$

      end for

    end for

    Compute shared encoder update

$$\Delta\phi_k^{(t)} \leftarrow \phi_k^{(t,E)} - \phi^{(t)}$$

    Clip update

$$\Delta\phi_k^{(t)} \leftarrow \Delta\phi_k^{(t)} \cdot \min\left(1, \frac{C}{\|\Delta\phi_k^{(t)}\|_2}\right)$$

    Determine adaptive privacy budget

$$\epsilon_k^{(t)} \leftarrow \Psi\left(s_k^{(t)}, \tau_k^{(t)}, \rho_k^{(t)}\right)$$

    Sample Gaussian noise with variance $\sigma_k^{2(t)}$ derived from $\epsilon_k^{(t)}$

    Form protected update

$$\widetilde{\Delta\phi}_k^{(t)} \leftarrow \Delta\phi_k^{(t)} + \mathcal{N}(0, \sigma_k^{2(t)} I)$$

    Send $\widetilde{\Delta\phi}_k^{(t)}$ to the server

    Keep $\theta_k^{(t+1)} \leftarrow \theta_k^{(t,E)}$ locally

  end for

  Server aggregates protected shared updates

$$\phi^{(t+1)} \leftarrow \phi^{(t)} + \sum_{k=1}^{K} \frac{n_k^{(t)}}{\sum_{j=1}^{K} n_j^{(t)}} \widetilde{\Delta\phi}_k^{(t)}$$

end for

return Final shared encoder $\phi^{(T)}$ and local heads $\{\theta_k^{(T)}\}_{k=1}^{K}$

Algorithm 1 describes the proposed adaptive privacy-preserving cross-domain federated learning procedure. The algorithm operates in iterative communication rounds coordinated by a central server. At the beginning of each round, the server broadcasts the current shared representation parameters to all participating clients. Each client then performs local training using its recent streaming data window while maintaining a private anomaly detection head that captures city-specific traffic patterns. During local optimization, the model jointly minimizes the anomaly detection loss and a cross-domain regularization objective that encourages the shared encoder to learn transferable traffic representations across regions. After completing the local updates, each client computes the update of the shared encoder parameters. The update is first trimmed to limit its sensitivity and then disrupted using Gaussian noise produced by an adaptive differential privacy controller in order to safeguard critical mobility data. The privacy controller dynamically modifies the privacy budget based on update frequency, trust level, and data sensitivity. The coordination server receives the protected updates and uses weighted averaging to combine them to create the updated global representation. The global parameters are subsequently redistributed to all clients for the next training round. By separating globally shared representations from local anomaly detection heads, the proposed algorithm enables collaborative learning across heterogeneous cities while preserving data privacy and supporting real-time traffic anomaly detection.

Operationally, a streaming round starts when client k uses observations from the appropriate interval to create the current window $Wk(t)$. The client runs E local epochs over mini-batches from the current window, keeps its local head, and initializes the shared encoder from $\varphi(t)$. The encoder difference $\Delta\varphi k(t)$ is then calculated, this update is clipped, privacy perturbation is applied, and only the protected encoder update is uploaded. The server broadcasts $\phi(t+1)$ for the next round after aggregating the received protected updates. The update rule specified in Algorithm 1 does not require historical observations outside of the active window.

4.7 Computational Complexity Analysis

The computational complexity of the proposed framework can be analyzed at both the client side and the server side.

4.7.1 Local computation complexity

Let $m_k^{(t)}$ denote the number of samples in the streaming window at client k during round t, let E denote the number of local epochs, and let B denote the mini-batch size. Denote by C_h and C_g the forward-backward computational costs of the shared encoder h_ϕ and local detection head g_{θ_k}, respectively. The local training complexity at client k for one communication round is

$$O\left(E \cdot \frac{m_k^{(t)}}{B} \cdot (C_h + C_g)\right). \tag{17}$$

The clipping and Gaussian perturbation steps introduce an additional linear overhead in the size of the shared parameter vector $|\phi|$, yielding

$$O(|\phi|). \tag{18}$$

Hence, the total client-side complexity per round is

$$O\left(E \cdot \frac{m_k^{(t)}}{B} \cdot (C_h + C_g) + |\phi|\right). \tag{19}$$

4.7.2 Server aggregation complexity

At the coordination server, only the protected shared encoder updates are aggregated. If K_t clients participate in round t, the aggregation complexity is linear in both the number of participants and the size of the shared model:

$$O(K_t|\phi|). \tag{20}$$

This is more scalable than transmitting full client models when the client-specific heads remain local.

4.7.3 Communication complexity

Since only the shared encoder update is uploaded, the uplink communication cost per participating client is

$$O(|\phi|), \tag{21}$$

and the total uplink cost per round is

$$O(K_t|\phi|). \tag{22}$$

Similarly, the server broadcasts only the shared encoder, resulting in a downlink cost of

$$O(K_t|\phi|). \tag{23}$$

Because the local detector heads are not transmitted, the proposed design reduces communication overhead compared with methods that synchronize the full model.

4.7.4 Scalability discussion

The proposed framework scales favorably in cross-silo smart city settings for three reasons. First, communication is reduced by sharing only the transferable representation parameters. Second, regional specialization is possible without additional server overhead since local heads are lightweight and remain private. Third, to limit local processing and enable real-time deployment, training is done on streaming windows instead of the entire historical dataset. As a result, the architecture is appropriate for operational traffic systems where privacy and latency are top priorities. The resulting platform allows real-time anomaly detection through streaming model updates, facilitates privacy-preserving collaborative learning across cities, and adapts to heterogeneous traffic conditions using federated transfer learning.

5. SIMULATION AND EVALUATION

The simulation environment, experimental setup, and evaluation criteria used to evaluate the suggested cross-domain privacy-preserving federated learning framework are explained in this section. A discrete-event federated learning simulator built with `SimPy` and `PyTorch` was used for all tests. The federated-learning and system-level characteristics of the suggested architecture are the main topics of review in this section. The full sensing and feature distributions of an implemented intelligent transportation system are not replicated. Specifically, under controlled conditions, the benchmark separates non-IID client behavior, federated convergence, privacy-budget behavior, communication cost, delay, and client scalability.

5.1 Simulation Environment

A communication-aware environment that replicates actual network conditions between clients and the central server was used to simulate the federated learning system. The simulator mimics real-world network behavior by incorporating wireless channel fading, signal-to-noise ratio (SNR), and bandwidth limitations. The following parameters are used to model the communication channel between clients and the server:

- Channel bandwidth: 5 Mbps
- Signal-to-noise ratio: 15 dB
- Fading: enabled
- Packet size range: 64 KB – 512 KB

With this configuration, the simulator may record network variability, transmission delay, and communication cost during federated learning rounds.

5.2 Dataset and Data Distribution

The MNIST dataset was used for the experiments. In order to replicate varied data environments across cities, the dataset was divided across clients using a non-IID data distribution. The following is a summary of the data dissemination configuration:

- Dataset: MNIST
- Total clients: 90 and 100
- Non-IID partitioning: 2 shards per client
- Local batch size: 64

A portion of the dataset reflecting local traffic observations in a smart city setting is held by each client. Realistic cross-domain traffic distributions across regions are simulated by the non-IID partitioning. Instead of using MNIST as a direct depiction of traffic data or anomalies, this work uses it as a controlled federated-learning benchmark. Its function is to offer a replicable environment that allows for the systematic generation of heterogeneous client distributions and the examination of the effects of federated aggregation, adaptive privacy, communication limitations, and client population without regard to traffic-specific feature extraction. Statistical heterogeneity across participating domains is abstractly represented by the non-IID shard split. Therefore, rather than providing direct empirical validation on actual traffic streams, UAV/UGV observations, or traffic video, the experiments presented here validate the federated learning behavior of the suggested architecture under controlled heterogeneous conditions.

5.3 Model Configuration

For every experiment, a multi-layer perceptron (MLP) neural network served as the fundamental learning model. There are two hidden layers in the architecture: Hidden layer 1: 256 neurons and Hidden layer 1: 256 neurons. To guarantee a fair comparison, all baseline methods used the same architecture.

5.4 Federated Learning Setup

For a predetermined number of global communication cycles, the federated learning process was carried out. Local training is followed by server aggregation in each round, where the primary training parameters are (1) Number of communication rounds: 100, (2) Client participation: selected using Bayesian client selection. (3) Random seed: 42, and (4) Device: CUDA if available. Each client was given a distinct communication budget and packet size in order to replicate resource variability with configurations: Minimum client budget was 1, Maximum client budget was 5, and Total system budget was 20,000. These budgets show the computational or communication resources that each client has access to during training.

5.5 Baseline Methods

A number of sample federated learning techniques were compared to the suggested approach:

- **FLI**: Federated learning with Bayesian client selection.
- **FedBCD**: Block coordinate descent for communication-efficient federated learning.
- **Cross-Silo FL**: Weighted aggregation in standard cross-silo federated learning. Federated learning with adaptive differential privacy (AdaDP).
- **Proposed Method**: Federated learning with adaptive differential privacy and anomaly detection while maintaining cross-domain privacy.

Every technique was assessed using the same simulated parameters. The purpose of the baseline comparison is to evaluate the integrated framework under the same simulation settings against representative federated configurations. Performance variations should be understood as the result of the integrated architecture rather than as separate causal effects of transfer learning, adaptive differential privacy, streaming updates, or client selection because the current experiments assess the entire suggested configuration.

5.6 Evaluation Metrics

Several metrics measuring system performance, resource efficiency, and model correctness were used to assess the effectiveness of the suggested federated learning framework. These measurements enable a thorough comparison of the suggested method with baseline techniques such as Cross-Silo federated learning, FedBCD, and FLI.

- **Final Accuracy**

  The global model's classification performance at the conclusion of training is measured by final accuracy. It shows how well the learning algorithm uses the aggregated federated model to identify abnormalities.

$$FinalAccuracy = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(y_i = \hat{y}_i)$$

  where y_i represents the true label and $\hat{y}_i$ denotes the predicted label.

- **Mean Accuracy**

  The average model accuracy over all communication rounds is represented by mean accuracy. This indicator assesses the consistency of model performance during training as well as the general stability of the learning process.

- **Maximum Accuracy**

  The highest accuracy attained throughout training is indicated by maximum accuracy. This metric highlights the best performance level reached by the learning algorithm.

- **Budget Utilization**

  Budget utilization measures the proportion of the available system budget consumed during client participation and model training. This metric reflects how efficiently each algorithm uses its allocated computational or communication resources.

$$BudgetUtilization = \frac{UsedBudget}{TotalBudget} \times 100$$

- **Cost Efficiency**

  Cost efficiency evaluates the trade-off between learning performance and computational cost. It is defined as the ratio between achieved accuracy and the total training cost.

$$CostEfficiency = \frac{Accuracy}{Cost}$$

Higher values indicate better cost-effective learning performance.

- **Energy Efficiency**

  Energy efficiency measures the accuracy achieved per unit of energy consumption. This metric evaluates how effectively the learning system converts energy resources into model performance.

$$EnergyEfficiency = \frac{Accuracy}{Energy}$$

- **Total Cost**

  Total cost represents the cumulative resource cost incurred during the federated learning process. It includes the cost associated with client participation and communication overhead.

- **Total Energy Consumption**

  Total energy measures the energy consumed by participating clients during training. This metric reflects the computational effort required by the learning process.

- **Total Delay**

  Total delay represents the total communication and processing latency during the training process. This indicator measures how responsive the federated system is to various client demographics.

- **Privacy Cost**

  The privacy budget used for training is measured by the privacy cost. It shows how much differential privacy techniques are used to safeguard model updates.

- **Composite Performance Score**

  A composite performance score that takes into account accuracy, cost, energy usage, and delay was calculated in order to identify the ideal client population.

$$CompositeScore = Accuracy - Cost - Energy - Delay$$

  By striking a balance between resource efficiency and model accuracy, this statistic offers a comprehensive picture of system performance.

The range of the controlled federated benchmark is reflected in the chosen metrics. While the system-level behavior of federation is characterized by budget use, cost, energy consumption, delay, and privacy cost, accuracy evaluates predictive behavior. A benchmark with clear normal/anomalous traffic events and temporal event annotations is necessary for traffic-specific anomaly measures like precision, recall, F1-score, AUC, false-alarm rate, and event-detection delay. Reporting these metrics as proof of traffic anomaly identification would be deceptive because MNIST does not have such traffic-event semantics. Therefore, rather than being a comprehensive operational traffic-anomaly evaluation, the study's results are understood as an assessment of federated learning and resource behavior.

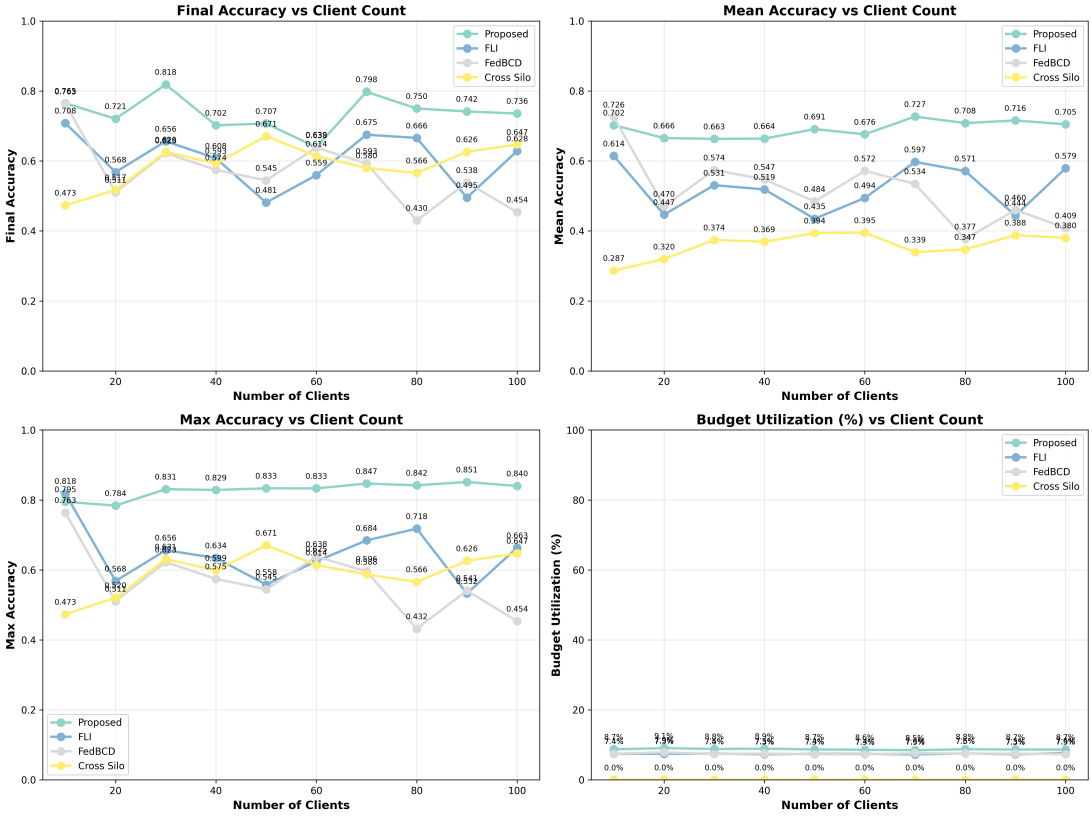

Figure 2: Accuracy performance is compared with different client numbers. The figure shows final accuracy, mean accuracy, maximum accuracy, and budget usage for the proposed framework and baseline methods such as FLI, FedBCD, and Cross-Silo federated learning.

6. RESULTS AND DISCUSSION

The experimental findings from the simulation study are examined in this section. Three baseline methods—FLI, FedBCD, and Cross-Silo federated learning—are contrasted with the suggested architecture. Model correctness, resource efficiency, and overall system performance under varying client participation numbers are the main evaluation criteria.

6.1 Accuracy Performance

The final accuracy, mean accuracy, and maximum accuracy attained by the assessed methods over various client population sizes ranging from 10 to 100 customers are shown in FIGURE 2.

Out of all the compared methodologies, the suggested framework consistently gets the best result. In most settings, the ultimate accuracy of the suggested method falls between around 0.70 and 0.82. On the other hand, FedBCD and Cross-Silo techniques exhibit more fluctuations and usually poorer performance, whereas FLI achieves final accuracy values between 0.49 and 0.71. The suggested method's stability during training is shown by the mean accuracy results. Across various client setups, the suggested approach keeps mean accuracy values between 0.66 and 0.73. This suggests

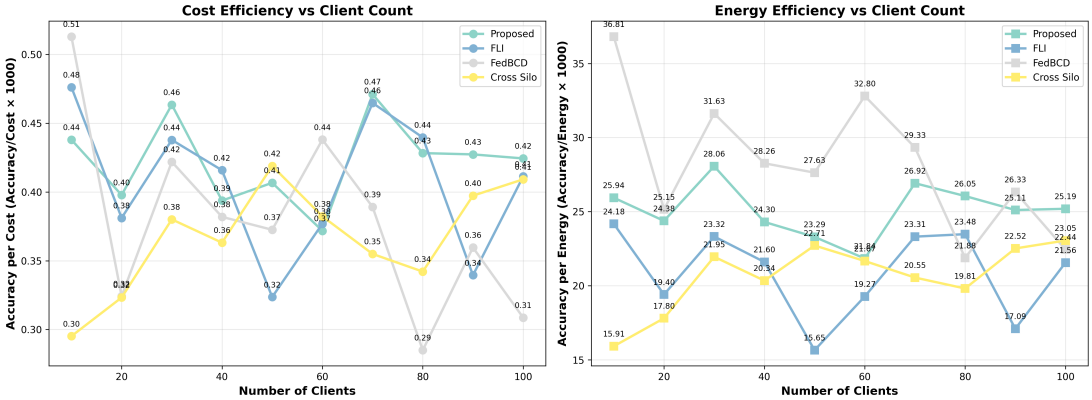

Figure 3: Energy and cost efficiency are compared for different client population sizes. The results show how resource consumption and model accuracy are traded off for the evaluated federated learning methods.

that compared to the baseline approaches, the model converges more reliably. With mean accuracy values ranging from 0.45 to 0.61, FLI performs moderately, while Cross-Silo has the lowest stability. The suggested method's efficacy is further demonstrated by the comparison of maximum accuracy. The suggested framework outperforms the baseline algorithms, achieving peak accuracy values above 0.84 for a number of client settings. This enhancement suggests that improved generalization across diverse clients is made possible by the cross-domain learning process.

6.2 Budget Utilization

Additionally, the budget consumption across various client populations is reported in FIGURE 2. The suggested approach shows effective resource management during training by maintaining a comparatively steady budget utilization of 7–9%. The alternative approaches, on the other hand, either use very few resources or are unable to make the most of the given funds. This finding implies that resource allocation and model improvement are successfully balanced by the suggested adaptive client selection technique.

6.3 Cost and Energy Efficiency

The cost and energy efficiency attained by the assessed techniques are shown in FIGURE 3. Across all client population sizes, the suggested approach consistently delivers high cost efficiency values. The suggested framework maintains good accuracy in relation to the training costs, as evidenced by the cost efficiency, which runs roughly between 0.40 and 0.47. In contrast, under some setups, FLI exhibits greater oscillations because to ineffective client selection. Results for energy efficiency show a similar pattern. The suggested method keeps energy efficiency numbers per energy unit between roughly 24 and 28 accuracy units. In certain situations, FedBCD shows greater energy efficiency, but at the cost of noticeably reduced accuracy. Consequently, a more balanced trade-off between performance and energy consumption is achieved by the suggested framework.

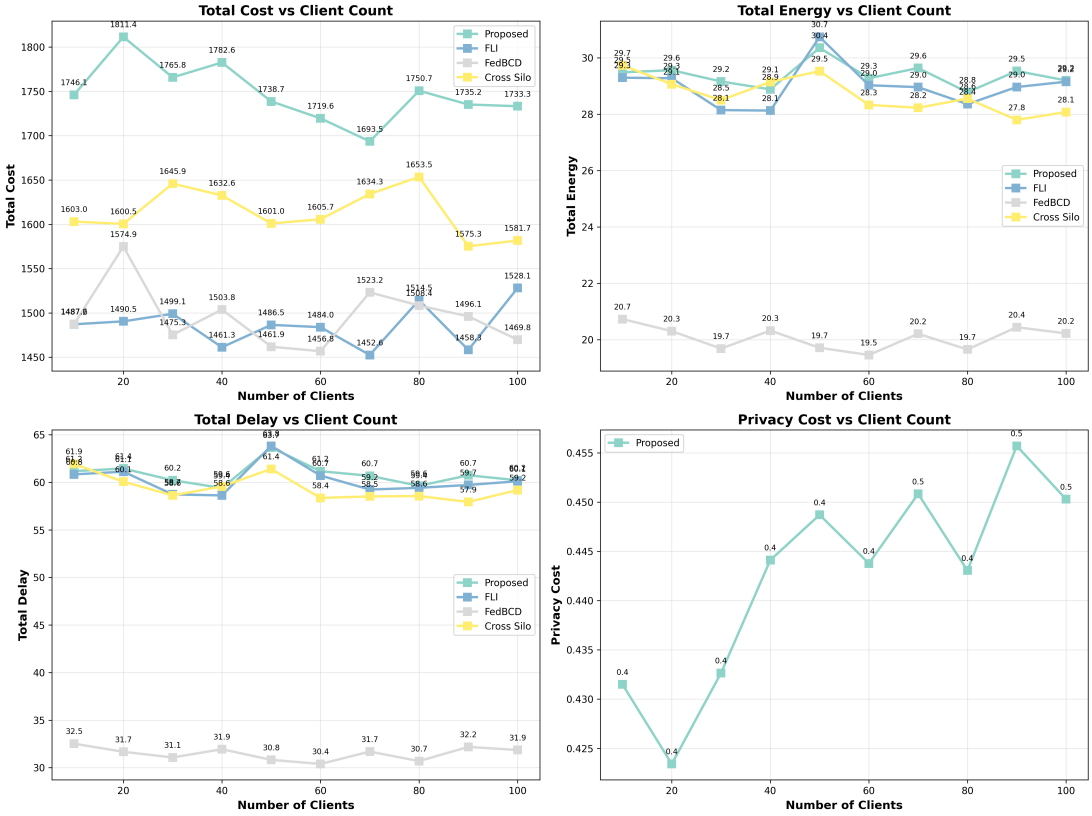

Figure 4: system resource usage for varying client counts. The total cost, total energy usage, total delay, and privacy cost of the assessed federated learning techniques are contrasted in the figure.

6.4 System Resource Consumption

The total cost, total energy usage, total delay, and privacy cost for the assessed approaches are compared in FIGURE 4. Because of its improved client selection process and adaptive privacy mechanism, the suggested approach has a little higher overall cost than various baseline solutions. With delay levels ranging from 59 to 61 units across various client counts, the overall delay is still comparable to the other methods. This suggests that system responsiveness is not greatly impacted by the extra computation needed for privacy protection. The suggested framework maintains competitive energy usage despite obtaining improved accuracy, as evidenced by the fact that energy consumption stays within a comparable range to existing approaches.

Here, privacy-related resource behavior across configurations is compared using the privacy cost metric, which captures the simulated consumption connected to the adaptive privacy mechanism. This metric should not be construed as a formal end-to-end privacy guaranty because the current simulation does not explicitly account for cumulative (ϵ,δ)-DP composition across successive communication rounds. The observed privacy cost values, which show regulated privacy spending while preserving model performance, fall roughly between 0.42 and 0.46.

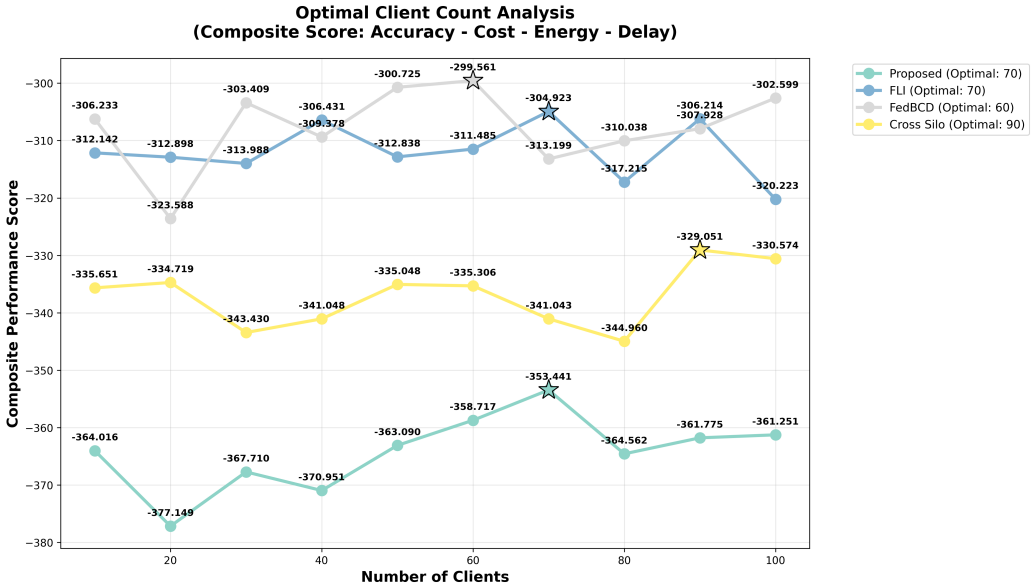

Figure 5: The ideal number of participating clients is determined using a composite performance score. The score determines the optimal client population size for each method by combining accuracy, cost, energy consumption, and latency.

6.5 Optimal Client Population Analysis

The composite performance score used to calculate the ideal number of participating clients is shown in FIGURE 5. To assess overall system performance, the composite score incorporates accuracy, cost, energy usage, and delay. The findings show that about 70 clients are the ideal amount to participate in the suggested framework. The suggested approach strikes the ideal balance between resource usage and model accuracy for this setup. The ideal configurations for the basic approaches, on the other hand, vary: FLI works best with about 70 clients, FedBCD with about 60 customers, and Cross-Silo with about 90 clients. These results demonstrate that the suggested system maintains balanced resource consumption and scales well with growing client numbers.

Because the measure is defined as a weighted mixture of several system characteristics, such as accuracy, cost, energy consumption, and latency, the composite performance score utilized in FIGURE 5 may result in negative results. Cost, energy, and delay are penalties that lower the total system performance score in the suggested formulation. Even when the model performs well, the resultant composite score may turn negative because these resource measures usually have bigger numerical magnitudes than the normalized accuracy values. As a result, the score's absolute value is not the main measure of performance. Rather, the key to identifying the best configuration is the relative comparison of scores across various client numbers. A better trade-off between model accuracy and resource usage is indicated by a higher (less negative) composite score. As shown in FIGURE 5, the optimal number of participating clients corresponds to the configuration that achieves the highest composite score among the evaluated options.

As shown in FIGURE 5, the raw composite score is negative for all methods because the penalty terms associated with cost, energy, and delay dominate the accuracy term in magnitude. Therefore, the score is used as a relative ranking criterion rather than an absolute utility value. To improve

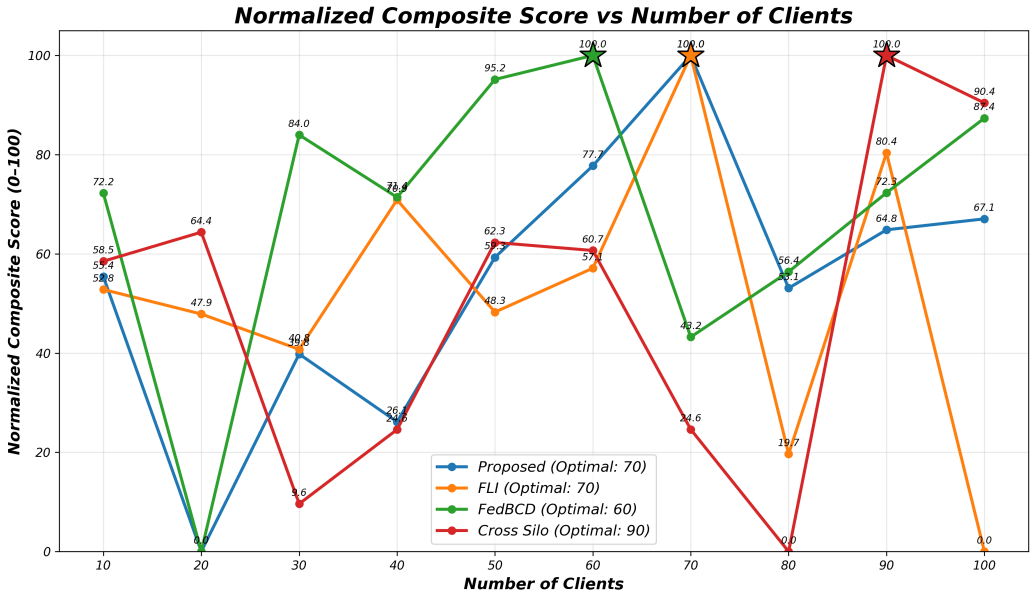

Figure 6: Normalized composite performance score for different client population sizes. The scores are obtained using min–max normalization of the raw composite scores to map values into the range 0–100 for each method. Higher values indicate better trade-offs between accuracy, cost, energy consumption, and delay.

readability, a normalized positive version of the same metric was also generated using min–max scaling, where higher values indicate better overall trade-offs. Based on this analysis, the optimal client count is 70 for the proposed method, 70 for FLI, 60 for FedBCD, and 90 for Cross-Silo federated learning.

FIGURE 6 presents the normalized composite performance score used to determine the optimal number of participating clients. Model accuracy, computational cost, energy consumption, and system latency are just a few of the performance factors that are included into the composite score. Because the penalty factors related to cost, energy, and delay outweigh the accuracy term in magnitude, the initial composite scores have negative values. The scores were standardized using min–max scaling so that the numbers for each approach lie between 0 and 100 in order to improve interpretability. Higher values under this transformation indicate improved overall performance. The findings show that when about 70 clients take part in the federated learning process, the suggested architecture performs at its best overall. Out of all the studied client population sizes, the approach earns the greatest normalized composite score at this configuration. Different ideal setups are provided by the baseline approaches. Additionally, the FLI approach performs best with about 70 consumers. The Cross-Silo approach achieves its maximum score at roughly 90 clients, whereas FedBCD performs best at 60 clients. These findings show that the suggested architecture maintains a fair trade-off between model accuracy and system resource consumption while scaling effectively as the number of participating clients rises.

The composite performance ratings can be converted into a normalized positive scale to enhance interpretability. The original formulation computes the composite score by combining multiple penalty variables, such as cost, energy, and time, with one positive item, accuracy. Even when the overall system performance is adequate, the resultant score may turn negative because the penalty

terms have greater numerical magnitudes than the normalized accuracy values. On its own, this does not suggest bad model behavior. Instead, it shows how the contributing terms scale relative to one another. As a result, the initial composite score should be interpreted in a comparison manner, with a higher score—that is, a less negative value—indicating a better trade-off between accuracy, cost, energy consumption, and time. The raw composite scores were also transformed into a normalized positive score using min–max normalization for each method across all assessed client numbers in order to facilitate display. In particular, the normalized score $\hat{S}(c)$ for a method with raw score $S(c)$ at client count c is calculated as

$$\hat{S}(c) = 100 \times \frac{S(c) - S_{\min}}{S_{\max} - S_{\min}}$$

where $S_{\min}$ and $S_{\max}$ represent the method's lowest and highest raw scores across all tested client counts. Each method's best client count earns a score near 100 under this transformation, while the lowest receives a score near 0. This normalized view makes the figure easier to read and comprehend while maintaining the configuration ranking.

6.6 DISCUSSION

The experimental results demonstrate that the proposed cross-domain privacy-preserving federated learning framework provides several advantages over existing approaches. First, throughout the assessed client population sizes, the suggested framework consistently attains more accuracy. These outcomes show how well the integrated setup performs under the simulated conditions that were taken into consideration. The observed improvement is ascribed to the entire configuration rather than any one process because the individual framework components are not isolated experimentally. Second, the framework maintains strong cost and energy efficiency, indicating that improved model performance does not come at the expense of excessive resource consumption. Third, the adaptive client participation strategy allows the system to scale efficiently as the number of clients increases. Overall, the results confirm that the proposed method provides a robust and scalable solution for privacy-preserving real-time anomaly detection in distributed smart city environments.

Overall, the simulation findings show that, in the assessed non-IID federated situations, the suggested integrated framework can improve predictive performance while maintaining competitive resource consumption. These results validate the viability of the underlying federated architecture, which includes communication-constrained learning, shared-representation aggregation, local specialization, and privacy-aware updates. However, the results should not be construed as direct validation of cross-city zero-shot performance or traffic-event detection because the current experiments use MNIST as a controlled heterogeneous benchmark. Therefore, the traffic-monitoring scenario reflects the architecture's planned application environment, whereas the current trials determine its behavior at the federated-system level.

7. CONCLUSION

in this paper, a privacy-preserving cross-domain federated learning system for real-time traffic anomaly detection in smart city settings was introduced in this research. The suggested framework enables safe and effective collaborative learning across several cities without disclosing raw traffic data by combining streaming federated updates, adaptive differential privacy, and federated transfer learning. When compared to traditional federated learning methods, experimental results showed increased detection accuracy, cross-domain flexibility, and resource efficiency. The suggested approach enables efficient anomaly detection in heterogeneous and non-IID traffic scenarios and offers a scalable and privacy-conscious solution for distributed intelligent transportation systems. Future research will concentrate on extending the architecture with cutting-edge deep learning and safe aggregation techniques, as well as assessing the framework using real-world traffic datasets.

ACKNOWLEDGMENT

The project was funded by KAU Endowment (WAQF) at king Abdulaziz University, Jeddah, Saudi Arabia. The authors, therefore, acknowledge with thanks WAQF and the Deanship of Scientific Research (DSR) for technical and financial support.

References

[1] Kairouz P, McMahan HB, Avent B, Bellet A. Advances and Open Problems in Federated Learning. Found Trends Mach Learn. 2021;14:1-210.

[2] Liu Y, Kang Y, Zhang X, Li L, Cheng Y, Chen T et al. A Communication-Efficient Collaborative Learning Framework for Distributed Features. In: Neurips Workshop on Federated Learning for Data Privacy and Confidentiality. 2019. ArXiv preprint: https://arxiv.org/pdf/1912.11187

[3] Wang Y, Sui M, Xia T, Liu M, Yang J, Zhao H. Energy-Efficient Federated Learning-Driven Intelligent Traffic Monitoring: Bayesian Prediction and Incentive Mechanism Design. Electronics. 2025;14:1891.

[4] Konečný J, McMahan HB, Ramage D, Richtárik P. Federated Learning: Strategies for Improving Communication Efficiency. In: NIPS Workshop on Private Multi-Party Machine Learning; 2016. Arxiv Preprint: https://arxiv.org/pdf/1610.05492

[5] McMahan B, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-Efficient Learning of Deep Networks From Decentralized Data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Proc Mach Learn Res. 2017;54:1273-1282.

[6] Stich SU. Local SGD Converges Fast and Communicates Little. In: International Conference on Learning Representations (ICLR). 2019.

[7] Li T, Sahu AK, Zaheer M, Sanjabi M, Talwalkar A, Smith V. Federated Optimization in Heterogeneous Networks. In: Proceedings of the machine learning and systems. 2020;2:429-450.

[8] Bonawitz K, Eichner H, Grieskamp W, Huba D, Ingerman A, et al. Towards Federated Learning at Scale: System Design. In: Proceedings of the machine learning and systems. 2019;1:374-388.

[9] Li T, Sahu AK, Talwalkar A, Smith V. FedProx: A Robust Federated Learning Framework. In: Proceedings of the machine learning and systems. 2020.

[10] Wright SJ.Coordinate Descent Algorithms. Math Program. 2015;151:3-34.

[11] Richtárik P, Takáč M. Distributed Coordinate Descent Method for Learning With Big Data. J Mach Learn Res. 2016;17:1-25.

[12] Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S et al. Practical Secure Aggregation for Privacy-Preserving Machine Learning. In: Proceedings of the ACM conference on computer and communications security. ACM. 2017:1175-1191.

[13] Shokri R, Shmatikov V. Privacy-Preserving Deep Learning. In: Proceedings of the ACM conference on computer and communications security; 2015.

[14] Gentry C. Fully Homomorphic Encryption Using Ideal Lattices. In: Proceedings of the ACM symposium on theory of computing. ACM. 2009:169-178.

[15] Dwork C, Roth A. The Algorithmic Foundations of Differential Privacy. Found Trends Theor Comput Sci. 2014;9:211-487.

[16] Abadi M, Chu A, Goodfellow I, McMahan HB, Mironov I, et al. Deep Learning With Differential Privacy. In: Proceedings of the ACM conference on computer and communications security.ACM; 2016:308-318.

[17] Yang Q, Liu Y, Chen T, Tong Y. Federated Machine Learning: Concept and Applications. ACM Trans Intell Syst Technol. 2019;10:1-19.

[18] Zhang J, Chen X, Li Y. Federated Learning for Traffic Flow Prediction. IEEE Trans Intell Transp Syst. 2021.

[19] Chen Y, Wang L, Zhang H. Privacy-Preserving Traffic Monitoring Using Federated Learning. IEEE Internet Things J. 2022.