Adaptive ST-VFL: A Drift-Aware Spatio-Temporal Vertical Federated Learning Framework for Cellular Traffic Prediction

Rameen Babar
Department of Artificial Intelligence & Data Science
National University of Computer & Emerging Sciences
Islamabad, Pakistan
i221918@nu.edu.pk

Batool Rizvi
Department of Artificial Intelligence & Data Science
National University of Computer & Emerging Sciences
Islamabad, Pakistan
i221907@nu.edu.pk

Arisha Khan
Department of Artificial Intelligence & Data Science
National University of Computer & Emerging Sciences
Islamabad, Pakistan
i222020@nu.edu.pk

Nabeelah Maryam
Department of Artificial Intelligence & Data Science
National University of Computer & Emerging Sciences
Islamabad, Pakistan
nabeelah.maryam@isb.nu.edu.pk

Mohsin Khan
Department of Artificial Intelligence & Data Science
National University of Computer & Emerging Sciences
Islamabad, Pakistan
mohsin.khan@isb.nu.edu.pk

***Abstract*—Telecom traffic forecasting remains difficult due to the complex interplay of spatial relationships between neighboring cells, heterogeneous feature sources, and evolving usage patterns influenced by external factors. Existing VFL-based prediction methods treat cells independently, omit mobility-driven adjacency and inter-cell influence, and rely primarily on historical traffic values, making them insensitive to weather conditions, public events, or holiday-driven fluctuations. Furthermore, most models assume static network behavior and lack mechanisms to respond to seasonal shifts or abrupt distribution changes, resulting in reduced long-term reliability. This research introduces an Adaptive Adaptive Spatio-Temporal Vertical Federated Learning (Adaptive ST-VFL) system that integrates multimodal feature fusion, KNN graph convolution, temporal modelling with GRU and ADWIN for drift-aware optimization. The approach improves forecasting accuracy while maintaining data privacy across a distributed system. Experiments conducted on the Telecom Italia Milan dataset which was further enriched with external meteorological and event-based contextual features, confirm strong predictive performance. The model achieves an MSE of 0.02399, outperforming the baseline, LSTM (0.386) by 93%, hence, confirming its effectiveness for real-world, dynamic cellular traffic environments.**

***Index Terms*—Vertical Federated Learning (VFL), Graph Convolution Networks (GCN), Spatio-Temporal Modeling, Multimodal Feature Fusion, Concept Drift Detection**

I. Introduction

The rapid expansion of mobile data consumption driven by emerging fifth-generation (5G) and future sixth-generation (6G) networks has made cellular traffic prediction a critical function for network optimization, resource allocation, and congestion management. Accurate forecasting supports dynamic spectrum scheduling, energy-efficient base station operation, and improved quality of service. However, cellular traffic exhibits strong spatial–temporal dependencies, nonlinear variations, and fluctuations influenced by user mobility and heterogeneous service patterns [1], [2]. These characteristics make forecasting models vulnerable to instability, distributional shifts, and representation limitations, particularly when relying on centralized learning frameworks where data consolidation raises privacy concerns.

Despite significant progress, several key challenges persist. Many existing spatial–temporal forecasting methods rely on fixed adjacency matrices that cannot capture dynamic spatial relationships between nodes [3], [4]. Temporal modeling approaches based on recurrent neural networks (RNNs), including long short-term memory (LSTM) networks and gated recurrent units (GRUs) or convolutional neural networks (CNNs) often struggle with long-range temporal dependencies or multi-scale periodicity [5], [6]. Cellular traffic is further shaped by diverse external factors such as weather, mobility patterns, and special events, making it difficult for conventional models to fully represent local and global spatial–temporal interactions [7]. Additionally, most existing methods assume stationary traffic patterns, despite real-world conditions being highly dynamic and prone to sudden shifts.

To address these limitations, prior research has explored graph-based learning methods, including graph convolutional networks (GCNs), graph learning mechanisms, and hybrid graph–temporal architectures [3]–[5], [7]. These approaches improve modeling of non-Euclidean spatial structures and capture temporal behavior through modules such as temporal convolutional networks (TCNs), LSTMs, and GRUs. However, they largely operate in centralized settings and do not consider vertical federated learning (VFL), a distributed paradigm where feature partitions are held separately by different data owners. Recent work has shown that VFL and split learning can support privacy-preserving forecasting [1], but these early frameworks lack advanced spatial modeling, contextual feature integration, and mechanisms to adapt to evolving traffic patterns.

This study proposes an adaptive spatio-temporal vertical federated learning framework that incorporates graph attention networks (GATs), multimodal feature fusion, and concept-drift-aware optimization using the adaptive windowing (ADWIN) algorithm. The model is evaluated on the Telecom Italia cellular traffic dataset augmented with meteorological and event-based contextual features. The framework addresses three key gaps: limited spatial awareness, insufficient external feature utilization, and the absence of dynamic adaptation under distributional drift.

The main contributions of this study are summarized as follows:

1) Development of a VFL-based spatial–temporal prediction framework integrating graph attention for dynamic spatial modeling while preserving feature-level privacy across data owners.
2) Introduction of a multimodal fusion pipeline leveraging traffic metrics, weather variables, holidays, and event information to enhance robustness under real-world variability.
3) Design of an adaptive optimization scheme using ADWIN-based drift detection to adjust model learning behavior under changing traffic distributions.

II. Review of Literature

Early developments in cellular traffic prediction increasingly emphasized privacy-preserving learning and distributed data ownership. Li et al. introduced a vertical federated learning and split-learning scheme for core network traffic prediction, demonstrating the feasibility of distributed feature training without raw data sharing [1]. Zeng et al. designed a long short-term fusion spatial–temporal graph convolutional network that integrates multi-scale spatial information and long-range temporal patterns [5]. Hu et al. proposed a graph-learning spatial–temporal GCN that dynamically reconstructs adjacency matrices to better capture spatial correlations [3]. Wang et al. provided a comprehensive survey of deep neural models for cellular traffic, highlighting limitations in spatial coupling and non-stationarity handling [2]. Chen et al. combined GCN and LSTM modules into a spatial–temporal parallel architecture, demonstrating substantial gains in modeling cellular network dynamics [6].

Further extensions examined adaptive or attention-based spatial–temporal modeling. Gu and Deng introduced STAGCN, a spatial–temporal attention GCN capable of refining dynamic dependencies among traffic nodes [4]. Feng et al. proposed MSA-GCN, a multistage aggregation architecture that captures hierarchical spatial relations for traffic flow prediction [7]. Wang et al. developed AHSTGNN, an adaptive hybrid spatial–temporal GNN that leverages hierarchical graph learning and multi-period temporal decomposition [8]. Li et al. constructed tensor-based adjacency reconstruction to enhance spatial modeling robustness under topology uncertainty [9]. Perifanis et al. studied the energy–accuracy trade-off in federated traffic prediction, revealing optimization constraints in real-world deployments [10].

Transformers and hybrid models have also gained prominence. Habib et al. built a transformer-driven wireless traffic prediction pipeline within the O-RAN architecture, showcasing significant improvements in real-time traffic forecasting [11]. Sun et al. introduced a time-aware transformer for trajectory recovery, underscoring the importance of periodicity and temporal context in sparse mobility data [12]. Fu et al. proposed the multi-grained spatial–temporal complementary (MGSTC) learning method for online cellular prediction, demonstrating robustness under concept drift [13]. Aouedi et al. surveyed deep learning-based traffic predictors, identifying challenges in stability, scalability, and multimodal integration [14]. Liu et al. introduced a dynamic component management mechanism to handle non-stationarity in cellular traffic forecasting [15].

Several studies explored advanced hybrid deep-learning pipelines integrating statistical preprocessing or attention mechanisms. Wang et al. proposed an attention-based hybrid deep-learning method combining ARIMA, CNN-LSTM, and XGBoost for mobile traffic forecasting [16]. Jin et al. evaluated data reduction techniques to streamline machine learning-based traffic prediction while balancing accuracy and efficiency [17]. Guo et al. presented a multi-scale wavelet–transformer hybrid method for traffic forecasting in 5G environments, emphasizing multiresolution temporal learning [18]. Zhang et al. developed STP-TCN, blending temporal convolutional networks with spatial adjacency modeling for urban network prediction [19]. Wang et al. proposed TSGAN, a graph-attention-based temporal similarity GNN that achieves strong performance across varying horizons [20].

Overall, across these twenty works, several trends become clear: (1) graph neural networks remain central for capturing spatial dependencies; (2) transformer and attention-based architectures increasingly dominate temporal modeling; (3) concept drift, non-stationarity, and external contextual factors remain under-explored; and (4) federated or privacy-preserving approaches are emerging but still lack integration with advanced spatial–temporal modeling. These persistent gaps motivate the development of an adaptive spatio-temporal vertical federated learning framework capable of robust spatial modeling, multimodal fusion, and drift-aware optimization.

III. Proposed Methodology

This section describes the complete methodology developed to enable accurate, privacy-preserving, and context-aware telecom traffic forecasting under non-stationary network conditions. The workflow begins with the construction of a unified spatio-temporal dataset derived from the Telecom Italia Big Data Challenge corpus, augmented with contextual information such as weather and holiday effects. A detailed preprocessing pipeline transforms the raw activity logs into temporally aligned, cell-wise tensors suitable for deep spatial–temporal learning. The proposed Adaptive Spatio-Temporal Vertical Federated Learning (Adaptive ST-VFL) model is then introduced, combining vertical federated feature encoders, short-term temporal convolution, graph-attention-based spatial modeling, recurrent temporal learning, and drift-adaptive optimization. Mathematical formulations are provided throughout the section, and each module is motivated in the context of real-world network behavior, such as periodicity, burstiness, and cross-cell correlations.

A. Dataset Description

The Telecom Italia Big Data Challenge dataset is a large-scale, real-world cellular activity dataset containing anonymized mobile traffic information collected across hundreds of base stations deployed in Milan. Each record aggregates network activity into hourly intervals, providing information on SMS, call activity, and internet usage. These measurements capture a diverse set of urban regions including residential, commercial, and transportation hubs, offering a realistic backdrop for spatial–temporal forecasting tasks.

The dataset includes the following traffic features: (i) SMS-in, (ii) SMS-out, (iii) call-in, (iv) call-out, and (v) internet volume. These features reflect different usage behaviors and respond differently to city-wide patterns such as commuting cycles, special events, and weekends. To capture exogenous factors not directly visible in traffic logs, additional contextual features are merged from external APIs. Weather attributes such as temperature, rainfall, and wind speed are included, as these conditions influence outdoor mobility and consequently traffic demand. Holiday and event indicators are used to model abrupt activity changes during public celebrations or large gatherings. The resulting feature set forms a heterogeneous, multimodal representation of telecom demand in Milan.

To reduce noise and sparsity, only the top 500 most active cells are retained based on aggregate internet usage. This ensures that the forecasting model focuses on relevant high-demand regions while excluding underutilized cells that contain insufficient signal. The final dataset is represented as a spatio-temporal tensor:

$$X \in \mathbb{R}^{T\times C\times F}, \quad Y \in \mathbb{R}^{T\times C\times 5}, \tag{1}$$

where T is the number of time intervals, C the number of active cells, and F the total number of traffic and contextual features. This representation preserves the full spatial structure of the network, allowing graph-based modules to exploit relationships between nearby or highly correlated towers.

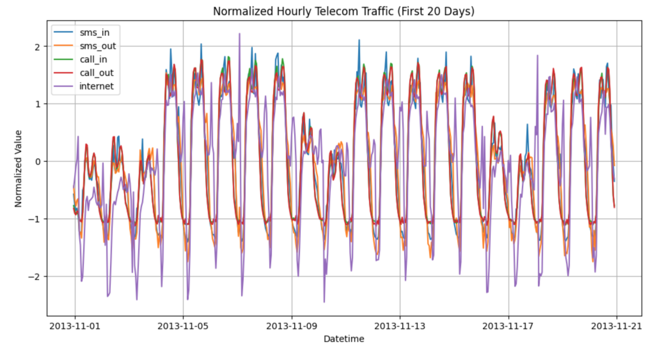

Fig. 1. Normalized hourly telecom traffic over the first 20 days, showing periodic daily structure and co-fluctuation across modalities.

B. Preprocessing and Data Pipeline

The preprocessing pipeline transforms raw logs into modeling-ready tensors by applying data cleaning, temporal aggregation, normalization, and structured indexing.

1) Data Cleaning: In real-world datasets, measurement irregularities, missing timestamps, and non-numeric fields are common. All malformed rows are discarded, traffic values are converted to floating point, and duplicates are removed. Since telecom activity is highly sensitive to precise timing, strict chronological ordering is enforced.

2) Temporal Aggregation: Because base stations may report asynchronous logs, hourly alignment is required to construct uniform sequences. Missing hourly entries are reconstructed using interpolation, ensuring that each cell contributes a continuous activity stream:

$$X_1, X_2, \dots, X_T.$$

This step is critical for convolutional and recurrent modules to capture periodicity.

3) Normalization: Traffic magnitudes vary significantly across modalities (e.g., SMS vs. internet). To prevent high-amplitude features from dominating the learning process, each modality is standardized using z-score normalization:

$$x' = \frac{x-\mu}{\sigma}, \tag{2}$$

which places all traffic signals on a comparable numerical scale and ensures stable multi-feature learning.

Figure. 1 illustrates the normalized hourly traffic for the first 20 days of the dataset.

4) Train/Test Split: A chronological split is used to preserve temporal causality:

$$\text{Train} = 87.5\%, \quad \text{Test} = 12.5\%.$$

This avoids information leakage from future intervals.

5) Spatio-Temporal Tensor Construction: After normalization and aggregation, features are grouped by timestamp:

$$X_t \in \mathbb{R}^{C\times F}, \qquad Y_t \in \mathbb{R}^{C\times 5}. \tag{3}$$

Each X_t provides a snapshot of the entire network state at time t, enabling spatial reasoning across cell towers.

C. Model Description

The Adaptive ST-VFL model integrates federated learning with deep spatial–temporal modeling to address several real-world challenges: (i) telecom data is distributed across operators and domains, making centralized training infeasible; (ii) traffic patterns vary significantly by cell type and urban region; (iii) concept drift frequently arises due to seasonality, human activity changes, and exceptional events; and (iv) different modalities respond differently to external triggers.

1) Vertical Federated Feature Encoders: Since traffic features originate from different operational domains (e.g., SMS, calls, internet usage), they are naturally suited for vertical partitioning. Let $x^{(c)}$ denote the features available to client c. Each client applies a lightweight encoder:

$$h^{(c)} = \mathrm{ReLU}(\mathrm{LN}(W_c x^{(c)})), \tag{4}$$

producing a private latent representation. These representations are securely combined:

$$Z = [h^{(1)} \| h^{(2)} \| h^{(3)}], \tag{5}$$

enabling global learning without sharing raw data.

2) Temporal Convolution Module: Short-term variations, such as morning peaks or sudden bursts, occur over small time windows. A temporal 1D convolution captures these patterns:

$$Z' = \mathrm{ReLU}(\mathrm{Conv1D}(Z)), \tag{6}$$

highlighting transitions and rapid fluctuations.

3) Graph Attention Layer (GAT): Telecom cells influence one another spatially due to mobility, handovers, and geographic proximity. To model this correlation, attention coefficients are computed:

$$\alpha_{ij} = \mathrm{softmax}_j \left(\mathrm{LeakyReLU} \left(a^\top [W z_i \| W z_j] \right) \right), \tag{7}$$

capturing directional influence from cell j to i. Spatial embeddings are aggregated:

$$h'_i = \sum_{j \in \mathcal{N}(i)} \alpha_{ij} W z_j. \tag{8}$$

This allows the model to learn both local and long-range dependencies.

4) Spatio-Temporal GRU: To capture long-term patterns such as weekly periodicity, recurring mobility cycles, or gradual trends, the GRU processes spatial embeddings over time:

$$h_t = \mathrm{GRU}(G_t, h_{t-1}), \tag{9}$$

enabling memory-based sequence learning.

5) Adaptive Federated Optimization: Telecom traffic exhibits concept drift due to evolving consumer behavior, seasonal changes, or holidays. A drift detector monitors model stability:

$$\mathrm{drift}(t) = \begin{cases} 1, & \text{if change detected,} \\ 0, & \text{otherwise,} \end{cases} \tag{10}$$

and adjusts the learning rate accordingly:

$$\eta_{t+1} = \begin{cases} 1.05\eta_t, & \text{if } \mathrm{drift}(t) = 1, \\ 0.995\eta_t, & \text{otherwise.} \end{cases} \tag{11}$$

The final prediction head produces traffic forecasts:

$$\hat{y}_t = W_o h_t + b_o. \tag{12}$$

To formalize the proposed framework, Algorithm 1 presents the complete Adaptive Spatio-Temporal Vertical Federated Learning (Adaptive ST-VFL) workflow, including vertical feature encoding, temporal convolution, graph-attention-based spatial modeling, recurrent temporal integration, and drift-aware adaptive optimization.

Algorithm 1 Adaptive ST-VFL Training Procedure

```
Input: Tensor X, labels Y, adjacency A
Initialize encoders, temporal conv, GAT, GRU, FC layer, ADWIN
for each epoch do
    for t = W to T do
        Extract window X_{t-W:t}
        Encode partitions via (4)–(5)
        Apply temporal convolution using (6)
        Compute spatial attention using (7)–(8)
        Update GRU state using (9)
        Predict ŷ_t using (12)
        Update drift status using (10)
        Modify learning rate via (11)
        Backpropagate and update weights
    end for
end for
```

D. Architecture of the Proposed Adaptive ST-VFL Framework

Figure. 2 illustrates the complete workflow of the proposed Adaptive Spatio-Temporal Vertical Federated Learning (Adaptive ST-VFL) framework. The architecture integrates five key components: vertical federated local encoders, multimodal feature fusion, KNN-based spatial graph convolution, GRU-based temporal modeling, and adaptive federated optimization with drift detection. Unlike traditional centralized forecasting models, this framework enables multiple data holders to collaboratively train a unified system without sharing raw feature data.

Each client processes its local feature subset (SMS, call, or internet traffic) using a lightweight encoder that generates a privacy-preserving latent embedding. These embeddings are fused at the server and passed through a graph convolution layer to capture spatial dependencies across neighboring cells. The resulting spatio-enhanced representations are then modeled by a GRU to learn long-term temporal patterns and periodic behaviors. During training, the ADWIN detector monitors the loss stream and adjusts the learning rate when concept drift occurs, ensuring rapid adaptation to traffic fluctuations. The final prediction head produces five traffic forecasts corresponding to SMS-in, SMS-out, call-in, call-out, and internet usage.

E. Evaluation Metrics

The performance of the proposed Adaptive ST-VFL framework is assessed using two widely adopted regression metrics: Mean Squared Error (MSE) and Mean Absolute Error (MAE). These metrics quantify the accuracy of traffic volume predictions across all five modalities (SMS-in, SMS-out, call-in, call-out, and internet usage).

MSE measures the average squared difference between predicted and true traffic values, penalizing larger deviations more heavily:

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2. \tag{13}$$

MAE computes the mean absolute difference between predictions and ground truth, providing a more interpretable measure of average prediction error:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|. \tag{14}$$

Together, MSE and MAE offer a comprehensive evaluation of forecasting performance, capturing both overall error magnitude and average deviation across temporal and spatial dimensions of the telecom network.

IV. Experimental Setup

This section describes the computational environment, software configuration, and hyperparameters used to evaluate all models, including the LSTM baseline, FedAvg, SplitNN, and the proposed Adaptive Spatio-Temporal Vertical Federated Learning (Adaptive ST-VFL) framework. All experiments—from dataset preprocessing and external feature integration to tensor construction, baseline training, and final ST-VFL training—were executed under a controlled and consistent environment to ensure reproducibility.

A. Hardware Configuration

All experiments were conducted using Google Colaboratory's hosted GPU runtime. The environment provided an NVIDIA Tesla T4 GPU with 16 GB of GDDR6 memory, an Intel Xeon-class virtual CPU, and 12.7 GB of system RAM. GPU acceleration was used for all neural network computations, including the LSTM baseline, VFL encoders, temporal convolution blocks, GAT spatial attention, and GRU-based sequence modeling. CPU resources handled preprocessing tasks such as external feature merging, cell filtering, adjacency construction, and ADWIN-based drift detection. No distributed computing infrastructure or additional accelerators were utilized.

B. Software Environment

Experiments were executed using Python 3.10 in Colab's Ubuntu-based virtualized environment. Data cleaning, merging, and alignment were performed using NumPy 1.24 and Pandas 1.5.3. Standardization relied on Scikit-learn's StandardScaler. All deep learning components including the baseline LSTM model, FedAvg client models, SplitNN client/server networks, and the full Adaptive ST-VFL architecture were implemented in PyTorch 2.1 with CUDA-enabled GPU support.

Temporal drift detection was implemented using the ADWIN algorithm from the River library (version 0.21). Visualization of normalized telecom traffic was produced using Matplotlib 3.7. External contextual signals (daily temperature, rainfall, wind speed, and public holiday flags) were fetched using the requests library and mapped into the temporal index of the telecom dataset.

C. Model Hyperparameters

To ensure consistent evaluation across all models, identical temporal windows, optimization settings, and evaluation metrics were used. The LSTM baseline was trained using a 24-hour input sequence and a hidden size of 64. The FedAvg baseline simulated five clients, each trained for 25 epochs across five global communication rounds. The SplitNN baseline used two client-side encoders paired with a server-side fusion model trained for 20 epochs.

The Adaptive ST-VFL model used three vertical encoders producing 16-dimensional embeddings, concatenated into a fused 48-dimensional representation. A temporal convolution layer with 32 output channels and kernel size 3 extracted short-term temporal features. The GAT layer produced 16-dimensional spatial embeddings using a KNN-derived adjacency matrix. A GRU with 64 hidden units captured long-term temporal dependencies, followed by a fully connected prediction layer generating five traffic-related outputs.

Training employed the Adam optimizer with an initial learning rate of 1×10^{-3}. The ADWIN drift detector dynamically adjusted the learning rate within a bounded range of $[5\times10^{-5}, 5\times10^{-3}]$ based on detected concept drift. Gradient clipping with a threshold of 1.0 ensured training stability. Each epoch iterated sequentially across all time windows.

Table I summarizes the hyperparameters used in the experimental evaluation.

V. Results and Discussion

A. Results

This section reports the performance of all evaluated models, including the LSTM baseline, FedAvg, VFL SplitNN, and the proposed Adaptive ST-VFL framework. All models were trained and tested on the same 24-hour sliding window forecasting setup. Table II summarizes the Mean Squared Error (MSE) and Mean Absolute Error (MAE) metrics.

The results indicate that traditional learning baselines, particularly FedAvg, struggle to capture the complex spatial and

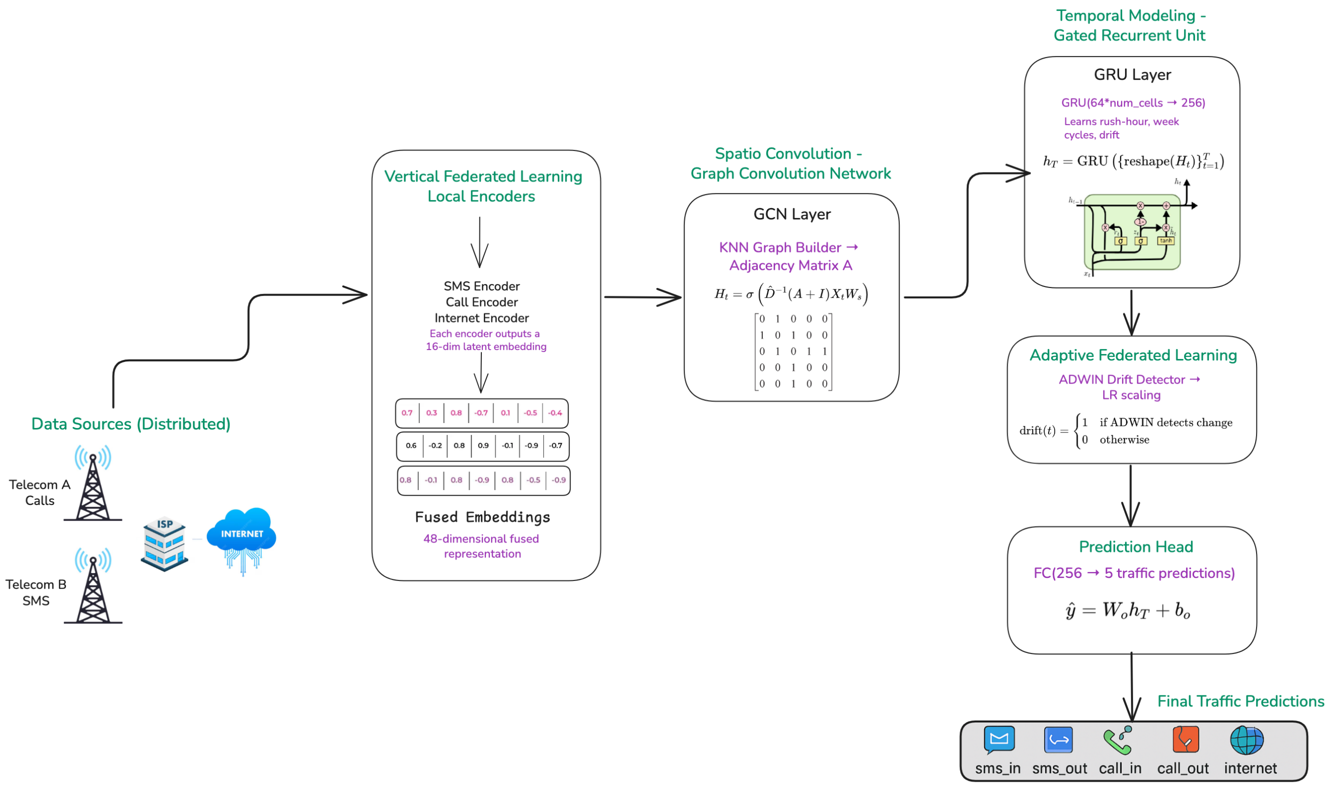

Fig. 2. Overall architecture of the proposed Adaptive Spatio-Temporal Vertical Federated Learning (Adaptive ST-VFL) framework.

TABLE I
SUMMARY OF HYPERPARAMETERS USED IN ADAPTIVE ST-VFL EXPERIMENTS

Parameter	Value
GPU Hardware	NVIDIA Tesla T4 (16 GB)
Tensor Dimensions	$6660 \times 500 \times 13$
Sequence length (window)	24 hours
Optimizer	Adam
Initial learning rate	1×10^{-3}
Learning rate range (ADWIN)	$[5 \times 10^{-5}, 5 \times 10^{-3}]$
GRU hidden units	64
Encoder output dimension	16 per partition
Temporal convolution channels	32
GAT output dimension	16
FedAvg clients	5
FedAvg rounds	5
SplitNN epochs	20
LSTM epochs	25
ST-VFL epochs	15
Batch size	32
Gradient clipping	1.0

temporal dependencies present in cellular network traffic. FedAvg exhibits the highest error due to its lack of structured feature partitioning and absence of spatial modeling capabilities. SplitNN improves performance by enabling vertical feature partitioning but still lacks explicit spatio-temporal reasoning, resulting in moderate error reduction.

The LSTM baseline performs significantly better than FedAvg and SplitNN, demonstrating the importance of temporal sequence modeling. However, the absence of spatial dependencies limits its predictive capability, especially in high-activity cells where inter-tower correlations strongly influence traffic patterns.

TABLE II
PERFORMANCE COMPARISON OF ALL MODELS ON TELECOM TRAFFIC FORECASTING

Model	MSE	MAE
LSTM	0.386262	0.524459
FedAvg	2.819680	1.112391
VFL SplitNN	1.026500	0.715511
Adaptive ST-VFL	**0.023998**	**0.142565**

The proposed Adaptive ST-VFL achieves by far the best prediction accuracy, reducing MSE by over 93% relative to the LSTM baseline and outperforming all baselines by a large margin. The combination of vertical feature encoders, temporal convolution, graph attention, and adaptive federated optimization enables the model to exploit rich cross-feature and cross-cell dependencies that simpler architectures fail to learn.

Figure. 3 shows the final comparison between actual and predicted values for the target cell across all five traffic modalities. The predicted vector closely follows the true values in all categories, demonstrating the strong generalization of the proposed architecture.

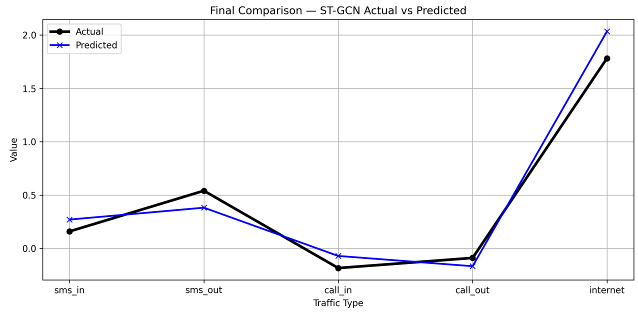

Fig. 3. Actual vs predicted traffic vector for the final timestep using the proposed Adaptive ST-VFL model. Predictions closely match the ground truth across all modalities.

B. Discussion

The experimental findings confirm that conventional federated approaches such as FedAvg and SplitNN are unable to effectively model the multi-modal and spatially interdependent characteristics of telecom traffic. Their relatively high error rates demonstrate the limitations of ignoring spatial topology or relying solely on linear feature partitions.

The LSTM baseline improves forecasting accuracy, validating the importance of sequence modeling in traffic time series. However, its failure to incorporate spatial relationships across base stations restricts its ability to capture coordinated usage patterns, especially during rush hours and event-induced peaks.

In contrast, the proposed Adaptive ST-VFL framework integrates three critical components: (i) vertical feature-learning encoders to capture fine-grained modality-specific patterns, (ii) spatio-temporal modeling through temporal convolutions and graph attention, and (iii) adaptive drift-aware optimization via ADWIN. Together, these components allow the model to dynamically adjust to evolving traffic distributions while preserving structured multi-cell dependencies.

The substantial reduction in both MSE and MAE demonstrates the effectiveness of combining federated feature partitioning with spatial graph learning and adaptive learning-rate modulation. The results confirm that hierarchical spatio-temporal modeling is essential for high-fidelity traffic prediction in modern cellular networks.

Overall, Adaptive ST-VFL achieves the lowest forecasting error, generalizes well across all five traffic modalities, and maintains stability even in the presence of temporal drift, making it highly suitable for real-world telecom resource management and network planning scenarios.

VI. Conclusion

This study addressed the challenge of accurate multi-modal telecom traffic forecasting under distributed data constraints, where heterogeneous operators cannot share raw data due to privacy and regulatory limitations. To overcome these barriers, the proposed Adaptive ST-VFL framework introduced a vertically federated architecture combining feature-specific encoders, temporal convolution, graph attention–based spatial modeling, and drift-aware optimization through ADWIN. Experimental evaluations demonstrated that Adaptive ST-VFL significantly outperforms LSTM, FedAvg, and SplitNN baselines, achieving more than a 93% MSE reduction compared to the strongest centralized baseline and delivering highly accurate predictions across all traffic modalities. These findings highlight the practical value of the framework for real-world cellular network management, enabling operators to collaboratively learn high-fidelity spatio-temporal patterns without exchanging sensitive data. Future work may integrate richer contextual signals, explore large-scale deployments with millions of cells, and incorporate differential privacy or secure aggregation mechanisms to further strengthen robustness and privacy guarantees in federated telecom forecasting environments.

References

[1] P. Li, C. Guo, Y. Xing, Y. Shi, L. Feng, and F. Zhou, “Core network traffic prediction based on vertical federated learning and split learning,” *Scientific Reports*, vol. 14, p. 4663, 2024.

[2] X. Wang, Z. Wang, K. Yang, Z. Song, C. Bian, J. Feng, and C. Deng, “A survey on deep learning for cellular traffic prediction,” *Intelligent Computing*, vol. 3, 2024.

[3] N. Hu, D. Zhang, K. Xie, W. Liang, and M.-Y. Hsieh, “Graph learning-based spatial-temporal graph convolutional neural networks for traffic forecasting,” *Connection Science*, vol. 34, 2022.

[4] Y. Gu and L. Deng, “Stagcn: Spatial–temporal attention graph convolution network for traffic forecasting,” *Mathematics*, vol. 10, 2022.

[5] H. Zeng, C. Jiang, Y. Lan, X. Huang, J. Wang, and X. Yuan, “Long short-term fusion spatial-temporal graph convolutional networks for traffic flow forecasting,” *Electronics*, vol. 12, 2023.

[6] G. Chen, Y. Guo, Q. Zeng, and Y. Zhang, “A novel cellular network traffic prediction algorithm based on graph convolution neural networks and long short-term memory through extraction of spatial-temporal characteristics,” *Processes*, vol. 11, 2023.

[7] J. Feng, J. Huang, C. Guo, and Z. Shi, “Msa-gcn: Multistage spatio-temporal aggregation graph convolutional networks for traffic flow prediction,” *Mathematics*, vol. 12, 2024.

[8] X. Wang, K. Yang, Z. Wang, J. Feng, L. Zhu, J. Zhao, and C. Deng, “Adaptive hybrid spatial-temporal graph neural network for cellular traffic prediction,” *IEEE Access*, 2023.

[9] Q. Li, X. Yang, Y. Wang, Y. Wu, and D. He, “Spatial-temporal traffic modeling with a fusion graph reconstructed by tensor decomposition,” *IEEE Access*, 2022.

[10] V. Perifanis, N. Pavlidis, S. F. Yilmaz, F. Wilhelmi, E. Guerra, M. Miozzo, P. S. Efraimidis, P. Dini, and R.-A. Koutsiamanis, “Towards energy-aware federated traffic prediction for cellular networks,” *IEEE Access*, 2023.

[11] M. A. Habib, P. E. Iturria-Rivera, Y. Ozcan, M. Elsayed, M. Bavand, R. Gaigalas, and M. Erol-Kantarci, “Transformer-based wireless traffic prediction and network optimization in o-ran,” *IEEE ICC Workshops*, 2024.

[12] T. Sun, Y. Chen, B. Zheng, and W. Sun, “Learning spatio-temporal dynamics for trajectory recovery via time-aware transformer,” *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[13] N. Fu, S. Liu, W. Xie, and Y. Huang, “Multi-grained spatial-temporal feature complementarity for accurate online cellular traffic prediction,” *ACM Transactions on Knowledge Discovery from Data*, vol. 37, no. 4, p. 111, 2025.

[14] O. Aouedi, V. A. Le, K. Piamrat, and Y. Ji, “Deep learning on network traffic prediction: Recent advances, analysis, and future directions,” *ACM Computing Surveys*, vol. 57, no. 6, p. 151, 2025.

[15] X. Liu, Y. Li, S. Zhu, Q. Su, J. Dai, C. Li, J. Zhu, and J. Zhang, “A traffic forecasting framework for cellular networks based on a dynamic component management mechanism,” *Electronics*, vol. 14, no. 4003, 2025.

[16] L. Wang, L. Che, K.-Y. Lam, W. Liu, and F. Li, “Mobile traffic prediction with attention-based hybrid deep learning,” *Physical Communication*, 2023.

[17] Y. Jin, M. Alazab, J. Wang *et al.*, "Machine learning-based cellular traffic prediction using data reduction techniques," *IEEE Access*, 2023.
[18] M. Guo, W. Zhao *et al.*, "Multi-scale wavelet transformer for 5g traffic forecasting," in *IEEE ICC*, 2023.
[19] H. Zhang, W. Liu, Q. Zhou *et al.*, "Urban cellular traffic forecasting via spatial-temporal tcn and adaptive adjacency learning," *Mathematics*, vol. 10, no. 1599, 2022.
[20] Z. Wang, J. Hu, G. Min, Z. Zhao, Z. Chang, and Z. Wang, "Spatial-temporal cellular traffic prediction for 5g and beyond: A graph neural networks-based approach," *IEEE Transactions on Industrial Informatics*, 2025.