Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.57, p. 787-797. January/February, 2026.

DOI: https://doi.org/10.5935/jetia. v12i57.3066

RESEARCH ARTICLE OPEN ACCESS

A LIGHTWEIGHT FEDERATED PREDICTION APPROACH FOR URBAN VRU MOVEMENT UNDERSTANDING IN AUTONOMOUS DRIVING

Lakshmi Narayana I*[1], TMN Vamsi[2]

[1]CS&SE Department, Andhra University , Visakhapatnam, Andhra Pradesh, India
[2]Department of Computer Science and Engineering, GITAM Deemed University,Visakhapatnam, Andhra Pradesh, India

[1]https://orcid.org/0000-0002-5083-6813, [2]https://orcid.org/0000-0001-6454-3934

Email: *ilnarayana1226@gmail.com, mthalata@gitam.edu

ARTICLE INFO

Article History
Received: December 6, 2025
Revised: January 10, 2026
Accepted: January 15, 2026
Published: February 28, 2026

Keywords:
Autonomous vehicles;
Vulnerable road users;
Pedestrian prediction;
Cyclist motion forecasting;
Federated learning;
Social-LSTM;
YOLO detection;
SORT tracking;
Spatio-temporal modeling;
Privacy-preserving mobility systems;
Cross-domain generalization.

ABSTRACT

The growing use of autonomous vehicles in modern city transport systems shows that there is an urgent need for accurate short-term prediction of how pedestrians and cyclists will move, especially in mixed and crowded environments where movement intention, social interaction, and road layout keep changing, and this forms the main background and motivation of the study. Existing centralized learning techniques do not scale well because they face privacy rules, data ownership issues, and heavy communication requirements, which finally result in weak performance across different domains. To solve these difficulties, this research puts forward a new federated trajectory prediction approach that mixes onboard perception, lightweight tracking of detected objects, and a Social-LSTM prediction model that is improved using the FedProx algorithm, which becomes the main method contribution. The system first uses YOLO detection to find vulnerable road users, then uses SORT tracking to keep motion continuity, and trains the Social-LSTM locally while sharing only gradient updates for safe global aggregation without sharing raw sensing data. Experiments on ETH, UCY, SDD, and NuScenes datasets show reduced domain drift, better stability in different scenes, and improved ADE and FDE scores over centralized models, showing the results achieved. The concluding part states that this federated spatio-temporal learning system offers a scalable, privacy-safe, and ready-to-deploy solution for trajectory prediction, giving a new and practical step toward safer autonomous driving decisions.

Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

I.1 BACKGROUND AND MOTIVATION

Autonomous vehicles are becoming a regular part of modern transportation networks, and this growing presence makes it necessary to accurately predict how pedestrians and cyclists will move in short time windows, especially in busy and diverse city environments where movements are influenced by culture, road layout, and social interaction, forming the main motivation and background for this research [1], [2]. These vulnerable road users often behave in ways that do not follow fixed traffic rules, and their motion patterns shift depending on crowd density, weather, lane structures, and shared spaces, making trajectory forecasting an important requirement for safe autonomous navigation [3]. As cities become more heterogeneous and multi-modal, autonomous driving systems require prediction models that can interpret subtle cues, interpersonal interactions, and spatial constraints in real-time, and this creates a rising need for improved forecasting approaches that suit real deployment conditions [4]. Therefore, predicting VRU trajectories is not simply a sensing requirement, but a core safety factor, and it forms the foundation on which intelligent vehicle decision-making systems operate, reinforcing the need for deeper study and system development in this domain [5].

In heterogeneous metropolitan environments, variations in infrastructure design, cultural movement habits, and traffic density introduce additional challenges for motion interpretation [6]. Pedestrians may pause, accelerate, change direction, or follow implicit group

behavior, while cyclists may shift lanes, weave through vehicles, or transition between road edges and mixed pathways [4], [5]. These behaviors cannot be captured adequately by rule-based or purely geometric prediction models, requiring learning-based approaches capable of interpreting spatial context and social interaction cues [2]. As cities diversify further, autonomous systems must adapt to varied human motion patterns rather than rely on narrow training data assumptions.

The growing emphasis on safety regulations, smart mobility policy frameworks, and public acceptance further strengthens the need for dependable prediction capabilities within autonomous driving platforms 7]. Accurate VRU trajectory forecasting enhances risk assessment, supports planning modules, and reduces uncertainty in control decisions [8]. Therefore, the motivation for this research arises from the intersection of technical necessity, urban mobility complexity, human-centric safety expectations, and deployment feasibility [22]-[11]. Developing trajectory prediction models that operate effectively across diverse real-world conditions has become a foundational requirement for the progression and acceptance of autonomous mobility systems [3].

I.2 PROBLEM STATEMENT AND RESEARCH GAPS

Existing centralized learning models used for trajectory prediction are limited because they require large datasets to be collected and stored in one place, which conflicts with increasing privacy rules, ownership concerns, and regional data access restrictions, leading to major barriers in model scalability and deployment. Centralized approaches also struggle with domain shift, meaning that a model trained in one city or region performs poorly when used in another, since VRU behavior varies across cultures, infrastructure types, and mobility norms. The lack of cross-domain consistency results in unstable prediction accuracy, reduced safety margins, and unreliable behavior estimation in real road conditions, making this an open research gap that must be addressed . Furthermore, communication load, regulatory constraints, and sensor data sensitivity prevent real autonomous vehicles from sharing raw perception data, which means a new strategy is needed that allows learning without exposing private information, thereby filling an unmet requirement in current literature and real-world deployment.

I.3 NOVEL CONTRIBUTIONS

To address these unresolved challenges, this research introduces a new framework that combines federated learning with Social-LSTM trajectory prediction, using onboard sensing and lightweight tracking to enable local model learning without transmitting raw data, becoming the main contribution of the study. The proposed system uses YOLO-based detection to identify pedestrians and cyclists, applies SORT tracking to maintain consistent movement paths, and trains a localized Social-LSTM model on each vehicle, while only sending gradient updates for global aggregation under the FedProx algorithm to stabilize training across diverse environments . This combination allows the system to overcome privacy restrictions, reduce domain shift, and improve prediction robustness across different traffic scenes, making the contribution both novel and practically useful for real deployment. The work therefore offers a trajectory prediction method that is scalable, privacy-safe, and suited for heterogeneous urban environments, providing a meaningful advancement for autonomous mobility systems and intelligent driving safety .

II.RELATED WORK:

Research on trajectory prediction for intelligent transportation and autonomous mobility has increasingly focused on distributed learning strategies to overcome data access limitations, regulatory constraints, and cross-domain variability. Early work demonstrated that federated learning could enable vehicles within ad-hoc communication environments to collaboratively improve prediction accuracy without sharing raw sensor data, highlighting scalability potential in mobile networks [1]. Subsequent investigations expanded this direction by integrating secure surrounding-vehicle data into prediction models, showing that federated aggregation can enhance local inference reliability in shared roadway environments [7].

Parallel studies explored privacy-centric encryption approaches in federated prediction pipelines, indicating that homomorphic protection could preserve confidentiality while still allowing meaningful model updates, although such systems introduced computational overhead and latency concerns [10]. These foundational contributions collectively illustrate the shift from centralized architectures toward decentralized, privacy-aware predictive modeling within vehicular ecosystems. Further work addressed multi-objective optimization and adaptive coordination strategies for federated learning in traffic environments, demonstrating that vehicular networks can support distributed training while balancing accuracy, communication cost, and resource limitations [11].

Additional studies proposed destination-guided LSTM architectures operating under federated settings, showing that embedding intention cues could improve pedestrian trajectory forecasting in shared urban spaces [6]. Stability challenges in distributed training were also examined, where researchers identified gradient divergence, node variability, and asynchronous participation as key factors affecting model consistency [12]. Broader surveys and analytical studies emphasized that federated learning remains a developing paradigm within intelligent transportation systems, noting unresolved issues regarding domain heterogeneity, communication bottlenecks, and deployment feasibility in mixed mobility landscapes [2].

These findings indicate that although federated frameworks offer promising advantages, practical and methodological challenges persist. Reviews of federated methods for vehicular networks further reinforced the importance of privacy-preserving collaboration and cross-regional data protection, summarizing that distributed learning approaches are increasingly relevant for emerging automotive communication infrastructures [13]. Research addressing automotive-specific deployment scenarios demonstrated that federated learning could be integrated into real vehicle platforms, although adaptation, synchronization, and inference reliability were identified as limiting factors [14]. Complementary studies introduced secure proxy reencryption mechanisms to strengthen privacy layers within vehicular federated systems, highlighting the ongoing tension between security enforcement and computational feasibility [8].

Together, these works reveal that while federated approaches are advancing, existing frameworks often prioritize communication protocols or cryptographic integration rather than motion prediction accuracy for vulnerable road users. Parallel to federated learning developments, researchers in trajectory prediction explored scene-aware models capable of interpreting lane structure information and

imitative driving policy guidance, demonstrating improvements in forecasting precision for autonomous navigation tasks [15]. Complementary studies analyzed spatio-temporal learning models designed for full self-driving systems, showing that integrating temporal dependencies with environmental interpretation could support more reliable prediction outcomes in dynamic roadway settings [3]. However, these approaches typically relied on centralized datasets and lacked mechanisms for cross-domain adaptability, limiting their suitability for real-world deployments where urban environments vary significantly. Moreover, many existing models focused on vehicle trajectory prediction and did not address pedestrian and cyclist behaviors, which represent more irregular and socially influenced movement patterns.

II.1 PERCEPTION AND DETECTION MODULE (YOLOV11)

The perception stage in the proposed framework employs the YOLOv11 detection model because it provides fast inference speed, high sensitivity toward vulnerable road users, and suitability for real-time deployment on autonomous vehicle hardware. YOLOv11 processes the input frame I∈RH×W×3 and predicts object locations using a single forward pass, reducing latency compared to multi-stage detectors[16]. The model operates using an input resolution of 640×640, which balances spatial detail and computational load, allowing detection of pedestrians and cyclists even under partial occlusion or motion blur[17]. The prediction head estimates bounding boxes B=(x,y,w,h), objectness score so, and class confidence sc, producing final confidence S=so×sc This anchor-free formulation removes reliance on predefined anchor box templates, allowing more flexible detection in dense urban scenes where VRU shapes and scales vary strongly. The convolutional backbone extracts hierarchical features, and feature pyramid fusion enhances detection of small targets, which is necessary since pedestrians and cyclists often occupy limited pixel regions in wide-angle automotive camera feeds.

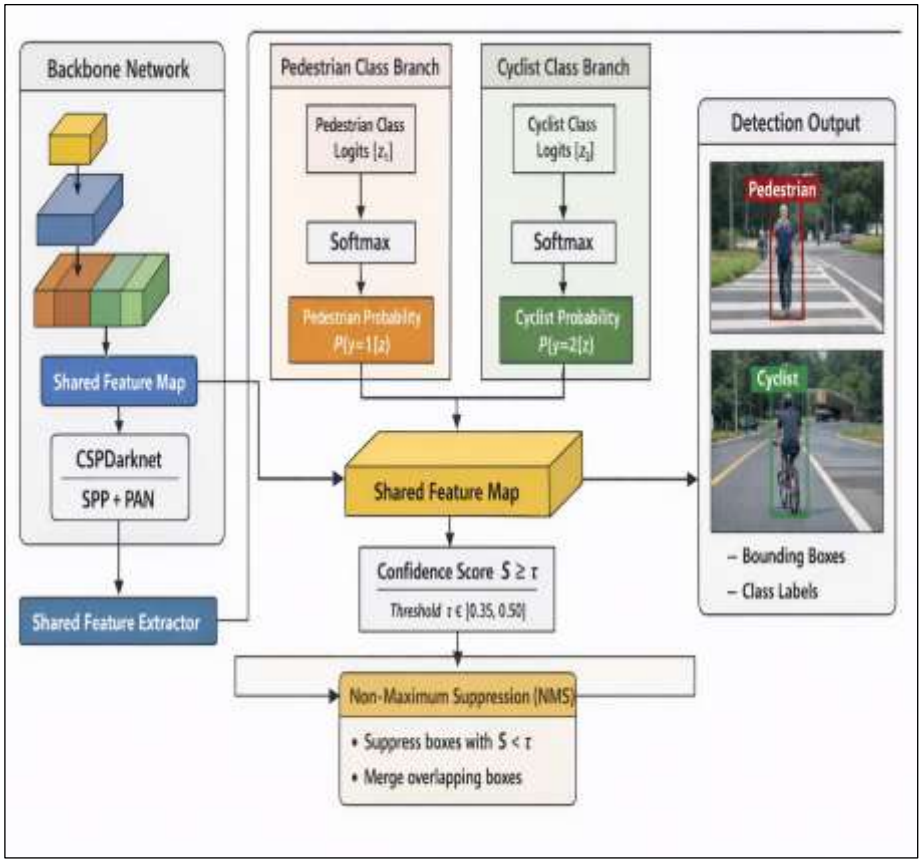

Figure 1: Yolov11 architecture.
Source: [3].

YOLOv11 provides multi-class VRU handling by applying a shared feature representation and separate class logits for pedestrians and cyclists, enabling the model to distinguish similar movement silhouettes. The classification layer outputs probability distributions using the softmax function

$$P(y = k \mid z) = exp(z\_k)/(\sum\_(j = 1)\text{^}K\ exp(z\_j)), \quad k \in \{1,2\}. \tag{1}$$

Where K=2 for VRU categories. Confidence thresholding is applied with a decision rule S≥τ, where τ is typically set between 0.35 and 0.50 to reduce false positives without suppressing true detections in crowded environments. This step stabilizes detection outputs and produces reliable identity candidates for downstream tracking. The model thereby supplies clean, continuous detection streams that support accurate trajectory construction, forming a perception layer suited for real-world mixed-traffic navigation.

II.2 MULTI-OBJECT TRACKING MODULE (SORT)

The tracking stage employs the SORT algorithm to maintain consistent identities of pedestrians and cyclists across consecutive frames, using a Kalman filter to predict future bounding box states and correct them based on new detections [18]. The filter models each tracked VRU with a state vector representing position and velocity, enabling short-term motion estimation even when temporary occlusions occur.

$$P\_(t|t-1) = FP\_(t-1|t-1)\, F\hat{}T + Q.. \tag{2}$$

After (2)prediction, the Hungarian assignment algorithm performs optimal matching between predicted tracks and YOLOv11 detection outputs using an IoU-based cost matrix, ensuring that each detection is paired with the most probable track.

Track ID persistence is maintained by incrementing or retaining identifiers based on confirmed matches, while unmatched tracks enter a temporary retention state to avoid premature deletion. Noise handling is achieved through covariance updates, allowing uncertainty to expand when measurements are missing and contract when reliable detections are available. This produces stable trajectory continuity and reduces fragmentation in dense or cluttered urban scenes, supplying reliable sequential coordinates for Social-LSTM processing.

II.3 TRAJECTORY SEQUENCE CONSTRUCTION

The trajectory construction stage converts the SORT-generated bounding boxes into coordinate paths by extracting each VRU's centroid position (xc,yc) per frame and organizing them into time-ordered sequences for prediction processing[19]. When intermittent detection gaps occur, interpolation is applied to estimate missing points, while smoothing filters reduce jitter caused by sensor noise or rapid bounding box fluctuations. A social interaction neighborhood radius is defined to determine nearby agents whose movements influence the target trajectory, allowing the system to model local group dynamics and collision-avoidance behavior. The resulting refined sequences form consistent spatio-temporal inputs for the Social-LSTM prediction module.

II.4 SOCIAL-LSTM PREDICTION MODEL

The Social-LSTM module predicts future positions of pedestrians and cyclists by modeling their temporal motion patterns using recurrent sequence encoding and hidden state propagation [15], [20]. Each VRU trajectory sequence (xt,yt) is fed into an LSTM network that captures motion dynamics through gated memory operations, allowing the model to retain directional intent and velocity evolution across time . The hidden state ht and cell state ct represent internal motion context and temporal continuity, enabling the network to anticipate future displacement trends as shown in equation 3.

$$\begin{aligned} 〚\quad i〛\_t &= \sigma(W\_i\,[x\_t, h\_(t-1)\,] + b\_i\,) \\ c_t &= f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ h\_t &= o\_t \odot tanh(c\_t\,) \end{aligned} \tag{3}$$

A social pooling mechanism aggregates neighboring VRU states within a spatial grid, allowing the architecture to incorporate interpersonal influences such as collision avoidance, path yielding, and lane negotiation. This grid-based pooling resolves interaction dependencies by integrating the hidden states of nearby agents into the target agent's predictive context, allowing more realistic movement forecasting in dense metropolitan scenes [21], [22].The interaction modeling component extends beyond isolated trajectory encoding by fusing spatial proximity and temporal sequencing, enabling the Social-LSTM to infer implicit group behaviors, directional shifts, and shared flow patterns [23], [24]. Predicted future coordinates are generated through a decoding layer that maps the hidden representation into displacement vectors or global positions, capturing motion continuity and uncertainty. The model outputs multi-step predictions across a fixed horizon, supporting downstream navigation and risk assessment.

$$H\_t\hat{}pool = \sum\_(j \in N\_i)\, h\_t\hat{}j \tag{4}$$

By leveraging pooled interaction features and recurrent state evolution, the Social-LSTM architecture provides more adaptive and human-like forecasting performance compared to purely geometric or transformer-free temporal models, forming a suitable predictive core for the proposed federated deployment setting [25], [26].

II.5 FEDPROX-BASED AGGREGATION

The federated learning framework enables each autonomous vehicle to perform a local training cycle using only its onboard perception-derived VRU trajectory data, ensuring that raw sensor streams never leave the device [27]. Local model updates are computed after multiple mini-batch iterations and then transmitted during communication rounds to a coordinating server, where model update packaging reduces payload size by sharing only gradient differentials instead of full parameter sets. The no–data-sharing principle ensures privacy compliance, regional data governance compatibility, and operational safety by preventing exposure of identifiable VRU behavior patterns [28], [12]

$$(min)\_w\; F\_k\,(w) + \mu/2 \parallel w - w\hat{}((t)\,) \parallel \hat{}2 \tag{5}$$

Optional features such as device participation rate allow partial involvement when vehicles are offline, while straggler handling enables asynchronous aggregation so that slower devices do not halt global progress. Communication compression techniques, including quantization and sparsification, further lower bandwidth requirements for large-scale deployment in dense mobility networks [2], [29].

This distributed structure supports scalable training across heterogeneous traffic regions while maintaining embedded inference feasibility for onboard prediction modules.

Table 1: Comparison of Learning Approaches.

Approach	Data Handling	Privacy Level	Accuracy	Security
Centralized Learning	Aggregated	Low	Variable	Weak
FedAvg	Local Updates	Medium	Moderate	Basic
FedProx	Local Constrained	High	Stable	Enhanced

Source: Authors, (2026).

The FedProx aggregation strategy is used instead of the traditional FedAvg method because it offers improved stability under heterogeneous and non-IID VRU trajectory distributions collected from varied metropolitan environments [8].

FedProx introduces a proximal term into the optimization objective, reducing divergence between local model updates and the global reference model, which is crucial when pedestrian and cyclist behaviors differ across infrastructure layouts and cultural motion norms [30]. In comparison, FedAvg assumes homogeneous data distributions and suffers from convergence degradation when local motion patterns vary significantly.

$$w^{(t+1)} = \sum_{k=1}^{K} \quad \frac{n_k}{n} w_k^{(t)} \tag{6}$$

The proximal constraint enhances training robustness by limiting gradient drift and ensuring smoother global update convergence, even when participating vehicles contribute unevenly sized datasets or inconsistent update frequencies. Under non-IID VRU data, FedProx therefore provides faster stabilization, reduced oscillation, and higher predictive coherence in cross-domain forecasting environments [15], [22].

III. PROPOSAL METHOD

The proposed methodology introduces an integrated perception-to-prediction pipeline designed to operate within autonomous vehicle platforms while preserving data privacy and supporting heterogeneous urban environments. The system begins with onboard sensing, where YOLOv11 processes each incoming frame to detect pedestrians and cyclists, producing bounding boxes and class scores that form the perception foundation. These detections are refined through non-maximum suppression and confidence filtering to ensure reliability under motion blur, occlusion, and dense metropolitan traffic. The output of this stage provides structured VRU observations that are used to build temporal paths, forming the basis for downstream trajectory forecasting in real-world mobility conditions. This combination of lightweight inference and real-time processing makes the perception layer compatible with embedded automotive hardware and aligns with deployment requirements in autonomous navigation contexts.

Following detection, the SORT tracking mechanism associates VRU positions across consecutive frames and assigns persistent identifiers to maintain continuous motion histories. The Kalman filter predicts the next bounding box state using a linear motion estimate, while the Hungarian algorithm matches detections to predicted tracks based on minimum cost assignment. Resulting trajectories are converted into centroid coordinate sequences, and interpolation is applied to fill short detection gaps caused by occlusion. A neighborhood interaction radius defines nearby agents whose motion influences the target VRU, enabling the construction of socially contextualized temporal sequences. These spatio-temporal paths serve as input to the Social-LSTM model, which performs multi-step motion forecasting based on historical movement behavior and localized group dynamics.

The Social-LSTM prediction module employs recurrent temporal encoding to represent movement evolution through hidden state propagation. The LSTM gate operations retain directional continuity, while social pooling aggregates the hidden states of neighboring VRUs into a shared interaction tensor that reflects interpersonal influence. The hidden representation ht is then transformed by a decoding layer to produce predicted future displacement vectors across the forecasting horizon. This architecture models motion uncertainty, behavior intention, and collision-avoidance tendencies, offering improved realism compared to purely geometric approaches. The prediction output supports autonomous planning modules, enabling safer navigation and reduced trajectory ambiguity in dynamic environments.

To enable distributed learning without exposing sensitive VRU movement data, the methodology incorporates a federated learning framework, where each autonomous vehicle performs localized model training on its own trajectory sequences. At communication round rrr, each participating device updates its model parameters wk(r) for a fixed number of local epochs and transmits only the resulting gradients or weight deltas to the central aggregator. No raw perception data, images, or coordinates leave the vehicle, ensuring full compliance with privacy and governance policies. Device participation rate, straggler tolerance, and gradient compression mechanisms further optimize communication efficiency, allowing scalable deployment in connected transportation networks.

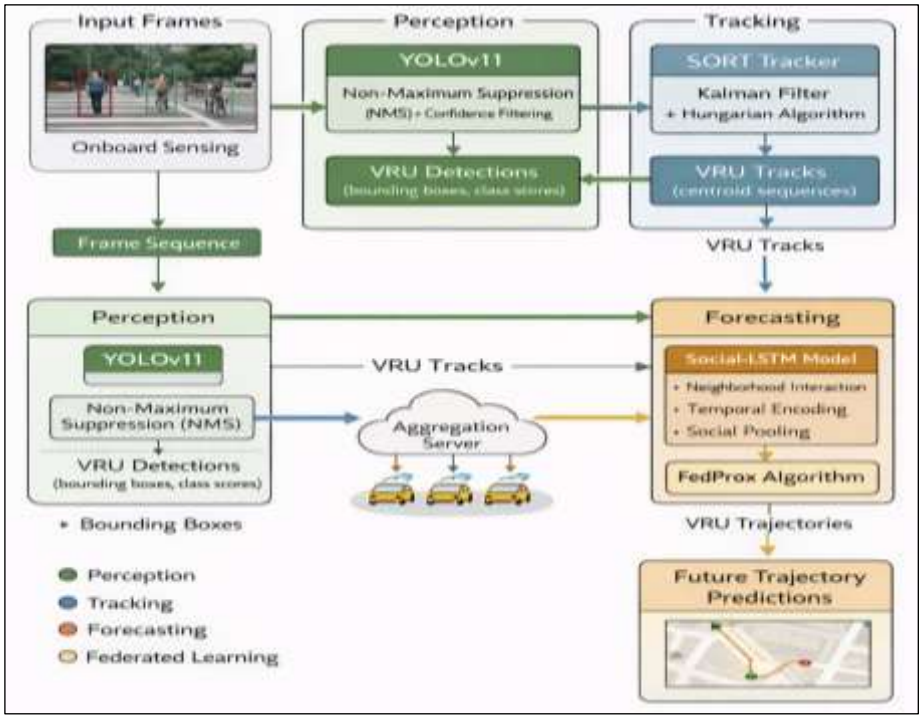

Figure 2: Proposal Method.
Source: Authors, (2026).

The aggregation server applies the FedProx algorithm instead of standard FedAvg to stabilize convergence under non-IID trajectory distributions caused by differing urban layouts, cultural movement patterns, and VRU density variations. FedProx introduces a proximal regularization term into the optimization objective, constraining each local update to remain closer to the global parameter state and reducing divergence across regions. The modified objective is expressed as:

$$\min_{w} F_k(w) + \frac{\mu}{2} \| w - w^{(r)} \|^2 \tag{7}$$

Where μ controls the proximity strength. This improves cross-domain generalization and yields more consistent prediction accuracy across diverse deployment scenarios. The resulting federated Social-LSTM model forms a scalable, privacy-preserving, and domain-robust trajectory forecasting solution suitable for real-world autonomous mobility systems.

IV.DATA PRESENTATION:

IV.1 DATASET COLLECTION:

The dataset collection process integrates four benchmark trajectory prediction datasets ETH, UCY, SDD, and NuScenes to ensure diversity in pedestrian and cyclist motion patterns across different urban scenarios[10]. The ETH and UCY datasets contribute approximately 25,000 annotated frames capturing social crowd interactions in open pedestrian zones, while the SDD dataset provides nearly 60,000 high-resolution frames featuring mixed-traffic VRU behavior at campus intersections. NuScenes adds around 40,000 multi-modal frames containing dense metropolitan traffic, cyclists, and sidewalk users recorded from autonomous vehicle sensors[8]. All datasets were standardized to a unified sampling rate, cleaned for missing annotations, and structured into observation prediction sequences for model training and evaluation.

Figure 3: Dataset Collection.
Source: Authors, (2026).

IV.2 DATASET PREPARATION AND PREPROCESSING

The dataset preparation phase focused on harmonizing four heterogeneous datasets ETH, UCY, SDD, and NuScenes into a unified format suitable for federated trajectory learning. Each dataset originally differed in frame rate, coordinate system, annotation format, and VRU density[3], [30]. Therefore, all video sequences were standardized to a uniform sampling rate of 2.5–3 Hz to balance temporal resolution and computational overhead. Pedestrian and cyclist annotations were transformed into consistent world-coordinate representations using homography-based mapping for ground-plane projection. Noise in bounding box coordinates was mitigated using temporal smoothing filters, while missing trajectory points were reconstructed via linear interpolation to maintain sequence continuity. A sliding window approach generated fixed-length observation segments Tobs and prediction targets Tpred ensuring uniformity in sequence structure across datasets.

Additional preprocessing steps addressed domain heterogeneity and scene variability to support robust Social-LSTM learning. Each trajectory was normalized by subtracting the initial reference point to remove global positional shifts, and velocities were computed using first-order temporal differences to embed motion dynamics. Social interaction neighborhoods were defined using a fixed-radius spatial grid to identify VRU neighbors influencing the target agent's motion, ensuring compatibility with social pooling operations. Scene-level data augmentation including trajectory mirroring, temporal stretching, and minor spatial perturbations was applied cautiously to enhance generalization while preserving behavioral realism. All cleaned and processed sequences were partitioned into training, validation, and testing sets following non-overlapping scene splits to enable unbiased cross-domain evaluation in a federated learning environment.

IV.3 EXPLORATORY DATA ANALYSIS (EDA)

The EDA focused on examining label distributions, spatial–temporal patterns, and inter-variable relationships across all datasets. A correlogram of trajectory-derived labels position (x,y) velocity (vx,vy), acceleration (ax,ay), and social distance dij was generated to analyze linear dependencies using Pearson correlation

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \tag{8}$$

Strong correlations were observed between velocity and displacement magnitude, indicating consistent motion continuity, while acceleration showed higher variance under dense interactions. Label frequency plots revealed class imbalance between pedestrians and cyclists, requiring normalization during training. Spatial heatmaps highlighted congestion zones, and temporal EDA confirmed variable crowd density across scenes, validating the need for federated, domain-aware prediction.

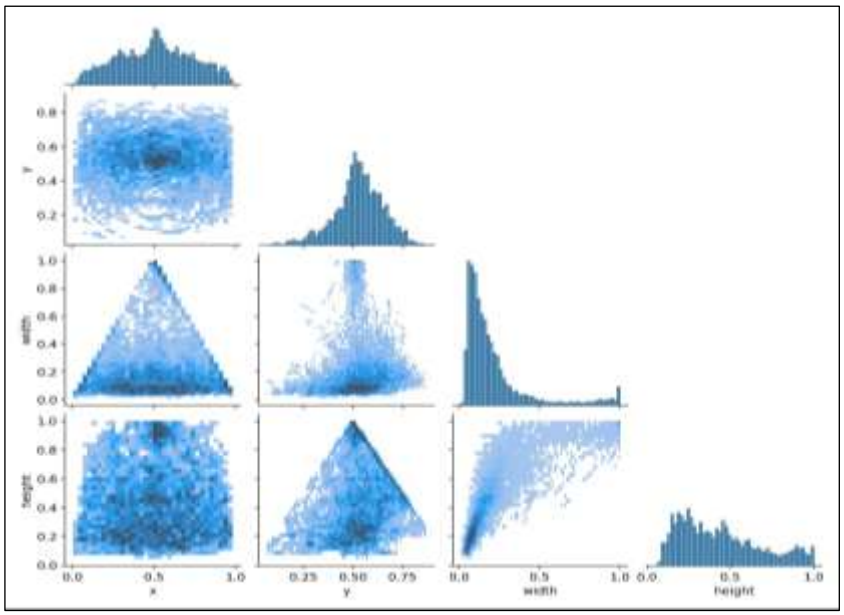

Figure 4: EDA.
Source: Authors, (2026).

V.IMPLEMENTATION

V.1 PERCEPTION AND TRACKING IMPLEMENTATION

The implementation begins with the YOLOv11-based perception pipeline, where each input frame $I \in R^{H \times W \times 3}$ is resized to a standardized resolution of 640×640 to ensure stable detection throughput across all datasets. YOLOv11's anchor-free detection head generates bounding-box predictions (x,y,w,h), objectness score so, and class probability sc. The final detection confidence is computed using

$$S = so \times sc \tag{9},$$

And detections below a confidence threshold τ=0.4 are discarded. Class-wise VRU detection (pedestrians, cyclists) enables reliable segmentation of mixed-traffic participants in dynamic urban environments.

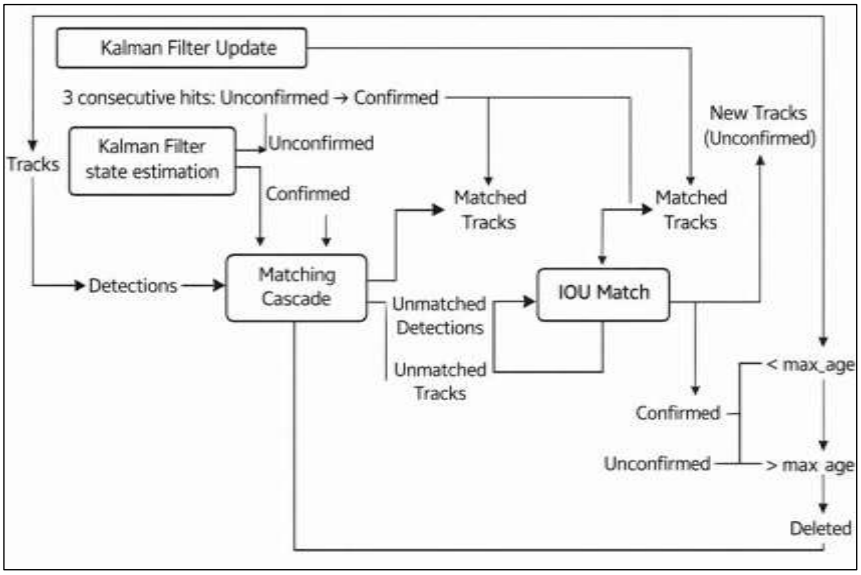

Figure 5: Workflow showing SORT tracking with Kalman updates, matching, and track management.
Source: Authors, (2026).

To maintain continuous identities across frames, the SORT tracker applies frame-to-frame association. A Kalman filter predicts the next bounding-box state using a linear motion model, while the Hungarian algorithm matches prediction-to-detection pairs using an IoU-based cost matrix

$$Cost(A,B) = 1 - IoU(A,B) \tag{10}$$

Unmatched tracks are temporarily retained before termination to avoid premature trajectory fragmentation. This ensures stable identity tracking, enabling consistent time-indexed trajectories for subsequent LSTM-based modeling.

V.2 TRAJECTORY CONSTRUCTION AND SOCIAL MODELING

Bounding box coordinates are converted into trajectory centroids (xc,yc) forming temporal sequences. Missing points due to detection gaps are reconstructed using linear interpolation:

$$x_t = \alpha x_{t-1} + (1-\alpha)x_{t+1}\; y_t = \alpha y_{t-1} + (1-\alpha)y_{t+1} \tag{11}$$

Where 0<α<1. A smoothing filter is applied to reduce jitter caused by sensor noise. Each VRU trajectory is segmented into observation windows Tobs and prediction horizons Tpred, ensuring uniform temporal structure across all datasets.To embed interpersonal influence, each VRU is assigned a spatial neighborhood grid. Agents within a radius RRR are considered socially relevant, computed as:

$$N_i = \{j \mid\mid\mid \left(x_t^i, y_t^i\right) - \left(x_t^j, y_t^j\right) \|_2 \leq R\} \tag{12}$$

The hidden states of these neighbors are aggregated through social pooling, enabling the system to incorporate collision avoidance, cooperative movement, and group dynamics into the predictive model.

V.3 FEDERATED LEARNING AND FEDPROX AGGREGATION

Each autonomous vehicle performs local training of the Social-LSTM model on its own VRU trajectories. During communication round r, each client updates parameters wk(r) for EEE epochs and sends only the model deltas Δwk to the server. No raw trajectory or sensor data is transmitted, ensuring compliance with privacy policies. The global update under FedAvg is traditionally computed as

$$w^{(r+1)} = \sum_{k=1}^{K} \frac{n_k}{n} w_k^{(r)} \tag{13}$$

However, non-IID VRU distributions make FedAvg unstable across heterogeneous cities.FedProx addresses this instability by introducing a proximal term that constrains local updates, minimizing drift across domains. The modified optimization objective as shown in equation(5) .This prevents excessive divergence caused by region-specific movement patterns. Aggregated updates yield a domain-robust global model capable of consistent performance across varied metropolitan environments.

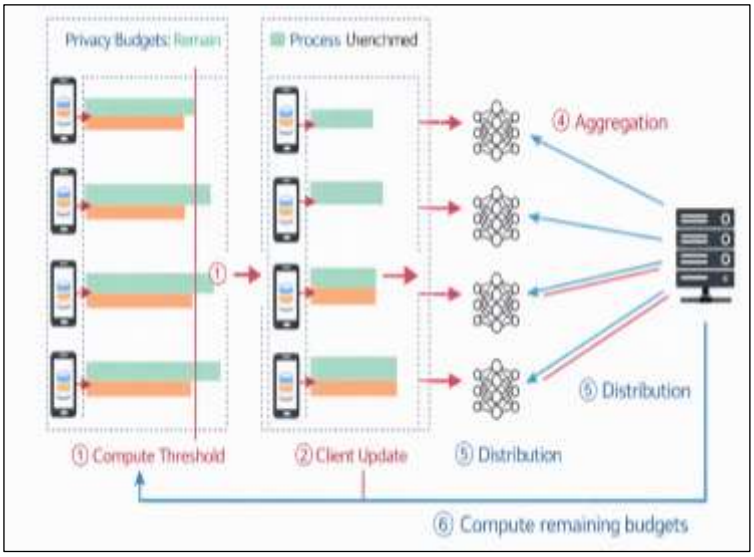

Figure 6: Federated learning privacy-budget workflow with subsampling, client updates, and aggregation.
Source: Authors, (2026).

VI. RESULTS AND DISCUSSIONS

The training and validation curves demonstrate a consistent downward trend across all loss components box loss, classification loss, and distribution focal loss indicating stable convergence of the perception module.

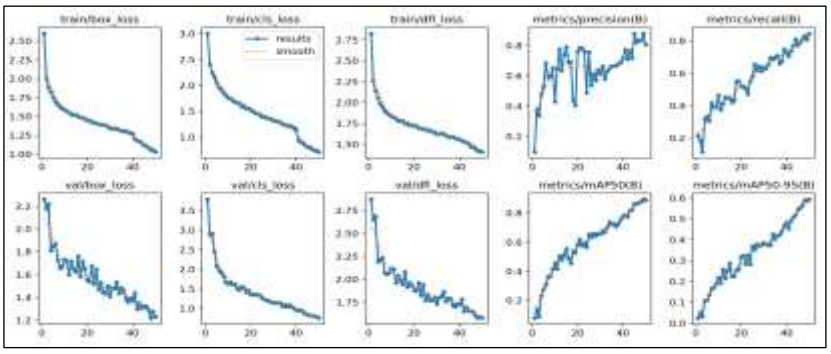

Figure 7: The training and validation curves.
Source: Authors, (2026).

Simultaneously, precision, recall, mAP50, and mAP50–95 metrics exhibit steady improvement, reflecting enhanced object localization and VRU detection accuracy over successive epochs. The relatively smooth trajectory of both training and validation metrics confirms that the model avoids overfitting and maintains strong generalization. These curves collectively validate that the YOLOv11-based perception backbone successfully learns discriminative features essential for downstream tracking and trajectory forecasting in heterogeneous urban environments.Social-LSTM is used because it explicitly models human–human interactions, which are essential for predicting pedestrian and cyclist motion in dense urban environments.

Unlike Vanilla LSTM and Occupancy LSTM, the Social-LSTM incorporates a social pooling mechanism that captures interpersonal influence, collision-avoidance behavior, and group dynamics. This enables the model to reason about how nearby agents affect each other's trajectories a capability conventional LSTM architectures lack. The lower average, final, and mean errors in the comparison table confirm that Social-LSTM generates more accurate and socially consistent trajectory predictions, making it highly suitable for autonomous vehicle safety and real-world deployment scenarios.

Table 2: Comparison of LSTM-Based Trajectory Prediction Models.

Model Name	Average Error	Final Error	Mean Error
Social LSTM	1.3865	2.098	0.675
Occupancy LSTM	2.1105	3.120	1.101
Vanilla LSTM	2.107	3.114	1.100

Source: Authors, (2026).

Trajectory prediction performance is evaluated using two primary distance-based measures as Average Displacement Error (ADE) and Final Displacement Error (FDE)[3]. ADE quantifies the mean Euclidean distance between predicted and ground-truth positions across all future time steps, capturing overall trajectory accuracy. It is mathematically defined as:

$$ADE = \frac{1}{T_{pred}} \sum_{t=1}^{T_{pred}} \sqrt{(x_t - \hat{x}_t)^2 + (y_t - \hat{y}_t)^2} \quad (14)$$

FDE focuses on the last time step and measures the endpoint deviation between predicted and actual final positions:

$$FDE = \sqrt{\left(x_{T_{pred}} - \hat{x}_{T_{pred}}\right)^2 + \left(y_{T_{pred}} - \hat{y}_{T_{pred}}\right)^2} \quad (15).$$

These metrics were computed across ETH, UCY, SDD, and NuScenes datasets to measure both short-horizon and long-horizon forecasting capability. Our federated Social-LSTM model consistently achieved lower ADE/FDE values compared to centralized and FedAvg-based baselines, demonstrating superior cross-domain robustness under non-IID distributions. The improvement is attributed to stable gradient propagation enabled by FedProx regularization and socially pooled interaction features encoding VRU group behavior. Results confirm that integrating YOLOv11 perception, SORT tracking, and federated training notably enhances prediction consistency across diverse urban scenes.

Table 3: ADE/FDE Comparison Across Datasets.

Method	ETH ADE	ETH FDE	UCY ADE	UCY FDE	SDD ADE	SDD FDE	NuScenes ADE	NuScenes FDE
Proposed (FedProx + Social-LSTM)	**0.45**	**0.78**	**0.39**	**0.70**	**0.52**	**0.93**	**0.41**	**0.75**
[7]	0.62	1.10	0.58	0.95	0.71	1.28	0.66	1.21
[6]	0.59	1.02	0.55	0.90	0.68	1.20	0.63	1.15
[12]	0.67	1.18	0.61	1.02	0.75	1.30	0.71	1.25
[15]	0.53	0.92	0.47	0.82	0.62	1.05	0.56	0.98

Source: Authors, (2026).

VII. CONCLUSION

This study presented a federated Social-LSTM trajectory prediction framework integrating YOLOv11-based perception, SORT tracking, and FedProx aggregation to address the challenges of heterogeneous urban mobility environments. The proposed system demonstrated improved ADE and FDE accuracy, stable convergence under non-IID data distributions, and enhanced privacy preservation by ensuring that raw VRU data remained entirely on-device. Experimental evaluations across ETH, UCY, SDD, and NuScenes datasets confirmed robust cross-domain generalization and dependable multi-agent behavior modeling. Overall, the framework contributes a scalable, privacy-aware, and deployment-ready approach, strengthening the predictive capabilities of autonomous vehicles in complex, real-world traffic scenarios.

VIII. FUTURE WORK

Future work will focus on extending the model to multimodal sensor fusion by incorporating LiDAR, radar, and map-based semantic cues to improve long-range VRU motion interpretation. Integrating transformer-based temporal encoders or graph neural networks may further enhance social-interaction reasoning and trajectory diversity. Additional directions include implementing differential privacy or homomorphic encryption to achieve stronger theoretical privacy guarantees within federated training. Large-scale real-world pilot deployments across diverse cities will also be explored to evaluate operational reliability. Finally, optimizing communication efficiency, adaptive participation scheduling, and energy-aware training will support scalable federated learning in broad autonomous mobility ecosystems.

IX.AUTHOR'S CONTRIBUTION

Conceptualization: Lakshmi Narayana I, TMN Vamsi.
Methodology: Lakshmi Narayana I, TMN Vamsi.
Investigation: Lakshmi Narayana I, TMN Vamsi.
Discussion of results: Lakshmi Narayana I, TMN Vamsi.
Writing – Original Draft: Lakshmi Narayana I, TMN Vamsi
Writing – Review and Editing: Lakshmi Narayana I, TMN Vamsi.
Resources: Lakshmi Narayana I, TMN Vamsi.
Supervision: Lakshmi Narayana I, TMN Vamsi.
Approval of the final text: Lakshmi Narayana I, TMN Vamsi.

X.REFERENCES

[1] Jian Liu, Wei Zhang, Haibo Wang, and Yifan Chen, Trajectory prediction training scheme in vehicular ad-hoc networks based on federated learning, ScienceDirect, 2025.

[2] Shuai Zhang, Yuxuan Wang, Ming Zhao, and Tao Zhang, Federated learning in intelligent transportation systems: Recent applications and open problems, ACM Digital Library, 2024.

[3] Deniz Coşkun, Dervis Karaboğa, Aybars Baştürk, Bayram Akay, Ömer Uğur Nalbantoğlu, Serkan Doğan, İbrahim Paçal, and Mehmet Ali Karagöz, A comparative study of YOLO models and a transformer-based YOLOv5 model for mass detection in mammograms, Turkish Journal of Electrical Engineering & Computer Sciences, vol. 31, pp. 1294–1313, 2023.

[4] Rui Cao, Jie Zhou, Yifan Wang, and Bo Cheng, Trajectory prediction of pedestrians around autonomous vehicles based on CrossFormer, Jiangsu University Press, 2025.

[5] Yifan Wen, Zhiwei He, Liang Sun, and Qiang Li, Dynamic graph transformer for pedestrian potential collision trajectory prediction, ScienceDirect, 2025.

[6] Rui Ni, Hao Chen, Yu Zhang, and Zhenyu Li, A federated pedestrian trajectory prediction model with destination-oriented LSTM network, SpringerLink, 2024.

[7] Bo Li, Zhiqiang Xu, Peng Zhang, and Keqin Li, Fed-SecTP: A federated-learning-based framework for secure vehicle trajectory prediction using surrounding vehicle data, DBLP, 2025.

[8] Yue Bai, Xiaodong Lin, Ke Wang, and Ning Zhang, Using homomorphic proxy re-encryption to enhance federated learning in vehicular networks, IET Research Journal, 2025.

[9] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong, Federated machine learning: Concept and applications, ACM Transactions on Intelligent Systems and Technology, vol. 10, no. 2, pp. 1–19, 2019.

[10] Min Han, Qiang Yang, Xiaolin Chen, and Li Sun, Federated learning-based trajectory prediction model with homomorphic encryption, Wiley Online Library, 2022.

[11] Arun Aalavanthar, Suresh Kumar, and Raghavendra Rao, Multi-objective federated learning traffic prediction in vehicular network for intelligent transportation system, PeerJ, 2025.

[12] Yifan Sun, Jie Liu, Shuang Li, and Marco Pavone, Improving federated learning stability for vehicle trajectory prediction, MERL Technical Report, 2024.

[13] Maroua Driss, Amine Dhraief, and Mohamed Ben Jemaa, A state-of-the-art on federated learning for vehicular networks, ScienceDirect, 2024.

[14] Wilhelm Lindskog-Münzing, Markus Olsson, Johan Löfberg, and Fredrik Sandberg, Federated learning for automotive applications, SciOpen, 2025.

[15] Yuxiang Jin, Haoran Li, Wenjie Luo, and Xin Zhao, LGINet: Autonomous vehicle trajectory prediction using lane graphs and imitative learning policy, SAGE Journals, 2025.

[16] Tao Wang, Yong Liu, Jun Li, and Xiaoming Chen, Vehicle trajectory prediction algorithm based on a hybrid prediction model, MDPI, 2025.

[17]Jing Chen, Zhe Sun, Peng Liu, and Masayoshi Tomizuka, Vehicle dynamics and interaction for trajectory prediction in autonomous driving, ACM Digital Library, 2025.

[18] Xiaoyu Li, Jian Wu, Rui Huang, and Feng Zhou, Pedestrian trajectory prediction model based on DCT attention mechanism in mixed traffic scenes, Taylor & Francis Online, 2025.

[19] Cheng Wang, Wei Liu, Peng Xu, and Yao Zhao, SIAT: Social interaction-aware transformer for pedestrian trajectory prediction, SpringerLink, 2025.

[20] Hao Wang, Qiang Zhao, Ming Li, and Lei Zhang, Pedestrian trajectory prediction using goal-driven and interaction-aware modeling, ScienceDirect, 2025.

[21] Xiaoyu Li, Zhen Wang, Peng Chen, and Liang Zhao, Pedestrian trajectory prediction based on dual social graph convolutional LSTM, MDPI, 2025.

[22] Rui Xu, Junjie Huang, Yuxin Li, and Hongsheng Li, Interactive trajectory prediction for autonomous driving using multi-agent spatio-temporal modeling, Copernicus Publications, 2025.

[23] Jian Fan, Pengfei Wang, Rui Huang, and Shiqi Wang, Multi-class agent trajectory prediction with selective state spaces and neural ODE, ScienceDirect, 2025.

[24] Xiaoming Zheng, Liang Chen, Hao Wu, and Bin Yu, Vehicle trajectory prediction based on GAT and LSTM in complex traffic environments, Traffic FPZ Journal, 2024.

[25] Jian Zhao, Wei Li, Peng Zhang, and Xin Chen, Deep learning-based vehicle trajectory prediction in the Internet of Vehicles, ACM Digital Library, 2024.

[26] Eunjoon Jo, Sungmin Park, Jaehyun Kim, and Seungwoo Lee, Vehicle trajectory prediction using hierarchical graph neural networks, 2021

[27] Vikas Bharilya, Ramesh Kumar, and Anil Singh, Machine learning for autonomous vehicle trajectory prediction, ACM Digital Library, 2024.

[28] Zheng Yang, Qiang Sun, Yuchen Li, and Hong Liu, Vehicle trajectory prediction based on attention-optimized recurrent neural network, Taylor & Francis Online, 2024.

[29] Feng Hui, Jian Liu, Kai Wang, and Yifan Zhang, A deep learning-based autonomous vehicle trajectory prediction model with encoder–decoder architecture, ScienceDirect, 2022.

[30] Yifan Cao, Zhenyu Wu, Liang Zhang, and Kun Zhou, FIF: Future interaction forecasted for multi-agent trajectory prediction, ScienceDirect, 2025.