SvedbergOpen
DISSEMINATION OF KNOWLEDGE
International Journal of Artificial Intelligence and Machine Learning
Publisher's Home Page: https://www.svedbergopen.com/
Research Paper Open Access

Machine Learning Error Correction for Satellite-Based Virtual Tolling

Venkateswarlu Mullangi[1]*, Rajasekhara Reddy Syamagari[2]

[1]Sam Higginbottom University of Agriculture, Technology and Sciences (SHUATS), India
[2]AMERISOFTPRO SYSTEMS LLC

Abstract

Standard GNSS positioning fails virtual tolling not because errors are large in absolute terms, but because their cost consequences are asymmetric — missed charges generate direct revenue loss while incorrect charges trigger customer disputes at substantially higher remediation cost. This mismatch is ignored by conventional correction methods, which optimize symmetric mean-squared error objectives without reference to billing economics. We present a hybrid gradient-boosting and LSTM architecture that treats lane-level positioning and boundary-crossing detection as a single coupled optimization problem, trained under an asymmetric cost objective derived from operational tolling economics. Rather than applying map geometry as a post-hoc constraint, toll-zone boundary geometry is encoded as a training feature, allowing the model to learn the statistical relationship between proximity to a boundary and the probability of a genuine crossing event. Field validation across 4,246 km of instrumented vehicle trajectories spanning urban canyon, suburban, and rural environments achieves median lateral error of 1.9 m in urban corridors against a 5.8 m baseline. Boundary-crossing detection reaches 96.3% true positive rate and 1.2% false positive rate. Satellite geometry features account for 1.2 m of the 3.9 m total urban error reduction, while HD map geometry prevents 38% of false positive boundary detections that geometric-only models cannot suppress. Real-time inference on ARM Cortex-A72 hardware averages 87 ms end-to-end latency within a 487 MB memory footprint and 2.3 W power envelope, meeting automotive-grade embedded deployment constraints without GPU acceleration.

Keywords: GNSS Error Correction, Virtual Tolling, Gradient Boosting, LSTM Networks, Intelligent Transportation Systems, Lane-Level Positioning, Asymmetric Cost Optimization

Introduction

Transportation infrastructure faces capacity constraints from increasing vehicle populations while traditional toll collection requires physical gantry infrastructure with high capital and maintenance costs. Satellite-based positioning offers infrastructure-free tolling through continuous vehicle location determination, supporting dynamic pricing schemes including distance-based tolls and congestion pricing — a segment the EUSPA identifies as one of the fastest-growing GNSS application domains in Europe [2]. However, practical deployment has stalled because standard GNSS accuracy proves insufficient for billing applications. Urban environments create severe positioning challenges where tall buildings block satellite signals and generate multipath reflections causing position errors of 5–10 meters [3]. For highway tolling where lane widths measure only 3.5 meters, these errors are commercially unacceptable.

Recent machine learning applications to GNSS error correction have demonstrated promising results for general positioning accuracy [4]. These approaches share a structural limitation that matters specifically for tolling: they minimize mean or median positioning error without incorporating the asymmetric penalty structure of toll billing events, where false negatives represent direct revenue loss to the charging authority while false positives generate customer claims and potential legal liability — road charging operators report that billing disputes represent one of the most operationally costly failure modes in GNSS-based tolling deployments [2]. Optimizing for symmetric error causes models to accept positioning uncertainty symmetrically around zone boundaries, producing false positive rates that are commercially unacceptable even when median positioning accuracy appears adequate.

Three additional limitations characterize existing methods. First, Extended Kalman Filters assume Gaussian error distributions and linear measurement models; urban multipath creates non-Gaussian error characteristics that violate both assumptions, and field studies confirm EKF performance plateaus at approximately 2.6 m median lateral error in dense urban environments regardless of IMU augmentation quality [5]. Second, map-matching heuristics snap positions to road centerlines rather than specific lanes, making them structurally unsuitable for lane-level discrimination at tolling resolutions where adjacent lanes are 3.5 m apart. Third, sequential architectures that apply generic positioning correction first and boundary detection second cannot exploit the geometric relationship between toll zone boundaries and expected vehicle dynamics — information that, when incorporated during training rather than applied post-hoc, substantially improves boundary crossing detection reliability.

This work addresses these limitations through four specific contributions. First, virtual tolling is formulated as a coupled optimization problem where positioning accuracy and boundary detection reliability are jointly optimized under an asymmetric cost constraint, differing fundamentally from sequential architectures. Second, a hybrid gradient-boosting-LSTM architecture is introduced where gradient-boosted regression captures complex nonlinear error relationships from satellite geometry and environmental context, while LSTM networks maintain trajectory coherence across administrative boundaries. Third, toll-zone geometry is incorporated as explicit contextual features during model training rather than applied post-hoc as map-matching constraints. Fourth, comprehensive validation spanning 4,246 km of field trajectories with dedicated sensor-level ablation analysis is presented, isolating each input modality's contribution independently from the architectural ablation of model components.

Background and related work

Gnss positioning challenges

GNSS signals ideally reach receivers through direct line-of-sight paths providing accurate position measurements. Urban areas create multipath interference where buildings and infrastructure generate signal reflections that receivers cannot effectively distinguish from direct signals [3]. Time-differenced code-minus-carrier measurements partially mitigate multipath by comparing code pseudorange observations with carrier phase measurements, revealing multipath-induced errors through duration time analysis [3]. Urban canyon environments present particularly severe challenges where tall buildings block satellite visibility and create signal reflection corridors, causing dilution of precision to increase significantly.

Modern GNSS receivers increasingly exploit multi-constellation signals combining GPS, Galileo, GLONASS, and BeiDou observations. Multi-constellation operation increases the number of visible satellites from an average of 6–8 for GPS-only to 20+ across all constellations, directly improving dilution of precision in urban canyon environments where single-constellation geometry is frequently degraded. The architecture presented in this paper processes pseudorange and carrier phase measurements from all available constellations through a unified feature extraction pipeline, with satellite geometry features computed across the combined multi-constellation solution rather than GPS-only observations. This is particularly relevant for dense urban deployments where Galileo's higher elevation broadcast angles provide improved signal availability in street-level canyon geometries compared to GPS alone. Standard positioning algorithms struggle when satellite visibility drops below five satellites, with position errors exceeding 10 meters even with advanced filtering [3].

Existing gnss correction approaches

Extended Kalman Filters represent conventional GNSS error correction, incorporating vehicle dynamics models to smooth position estimates. EKF methods assume Gaussian error distributions and linear measurement models, limitations that degrade performance in urban environments where multipath creates non-Gaussian error characteristics. Field studies show EKF approaches reduce median lateral errors from 5.8 m to approximately 2.6 m in urban canyons, insufficient for lane-level discrimination [5]. Map-matching algorithms constrain position estimates to road network geometries but typically snap to road centerlines rather than specific lanes, making them unsuitable for virtual tolling requiring lane-level discrimination [6].

The combination of EKF with probabilistic map-matching — the strongest conventional baseline — improves urban lateral error to approximately 3.1 m through particle-filter-based road-geometry constraints but requires 150–180 ms inference latency that exceeds real-time 10 Hz update constraints, and still provides no lane-level discrimination capability [6]. This combined baseline is the primary comparator for the present work because it represents the performance ceiling for non-machine-learning approaches under operational timing constraints.

Recent machine learning approaches have applied convolutional neural networks to process spatial patterns in satellite visibility and signal strength distributions, achieving improved accuracy over traditional filters [4]. However, these models optimize for average positioning error without considering boundary crossing detection requirements. Recurrent neural networks including LSTM architectures demonstrate capability for handling temporal dependencies in position time series, but existing implementations lack explicit toll-zone awareness required for reliable event detection [4]. Gradient boosting for GNSS correction has shown promise for capturing complex nonlinear relationships between measurement characteristics and positioning errors [10], but validation for virtual tolling applications with boundary crossing detection and asymmetric cost optimization remains unexplored.

Virtual Tolling Requirements

Virtual tolling systems must achieve sufficient lateral positioning precision for zone boundary discrimination. GNSS-based road charging deployments across European markets report average urban positioning errors of 6–12 meters, with operators requiring sub-lane-level accuracy to ensure reliable zone boundary classification in operational corridors [2]. Revenue assurance imposes asymmetric cost structures where false negatives result in direct toll revenue loss while false positives generate customer service overhead and dispute resolution costs that substantially exceed the original toll value, a pattern consistently observed across European GNSS road charging deployments [2]. Operational requirements for GNSS-based road charging systems specify high invoice accuracy thresholds to maintain commercial and regulatory viability, with European deployments targeting sub-lane positioning reliability as a minimum standard for distance-based charging schemes [2]. Traditional GNSS correction methods optimized for symmetric mean-squared error fail to address these asymmetric business requirements. The key technical implication is that a model achieving 2.0 m median error with a heavy right tail toward zone boundaries is operationally inferior to a model achieving 2.2 m median error with a compressed tail — a distinction that symmetric loss functions cannot encode.

AI-DRIVEN ERROR CORRECTION ARCHITECTURE

Sensor Fusion Framework

The input layer draws from four sources with complementary roles. Raw pseudorange and carrier phase observations from GNSS receivers form the primary positioning signal; satellite ephemeris and clock corrections are used to compute geometric pseudoranges for each visible satellite. IMUs running at 50–100 Hz fill a different role entirely: they are not a positioning source in steady state, but a bridge during the 3–8 second GNSS dropouts under bridges and overpasses where, as Section V.II confirms, they prevent the 8–12 m drift that overpass passages would otherwise cause [5]. High-definition map databases contribute lane-level geometry with centimeter accuracy, elevation profiles, and three-dimensional building footprints. These are not used as post-hoc snap constraints — they feed the model as learned features, enabling the gradient-boosted component to anticipate multipath conditions based on building height, proximity, and orientation relative to satellite positions rather than correcting for them after the fact [5]. Temporal features round out the input set, encoding time of day, day of week, seasonal variation, and historical error patterns for similar locations — capturing systematic error recurrence at known problematic segments that point-in-time features cannot represent [5].

Hybrid Machine Learning Architecture

Gradient-boosted regression forms the core error correction mechanism. Rather than treating all input features symmetrically, the model exploits four distinct feature categories that capture different aspects of positioning error: satellite geometry features (dilution of precision values, satellite count, elevation angles, azimuth distribution), signal quality features (signal-to-noise ratios, carrier-to-noise density, multipath indicators from code-carrier differences), environmental context features (urban density classification, building height distributions within a 100 m radius, distance to nearest tall structures, sky visibility fraction), and temporal features (time of day, day of week, seasonal variation, historical error patterns for similar locations and times). Early experiments treating these categories uniformly degraded urban accuracy by 0.4 m compared to the category-aware implementation, motivating the structured feature engineering approach.

The LSTM component handles trajectory-level coherence that gradient boosting alone cannot provide. Rather than feeding raw GNSS noise directly to the LSTM — which would force the recurrent layers to disentangle both error correction and temporal coherence simultaneously, undermining the structured feature engineering approach — gradient-boosted outputs are routed as LSTM inputs. Preliminary experiments feeding raw measurements to LSTM degraded urban accuracy to 3.4 m, worse than EKF alone, confirming that the separation of environment-specific error correction from trajectory coherence maintenance is architecturally critical. The forget gate mechanism selectively retains or discards information from previous time steps, proving particularly

valuable near toll zone boundaries where position jitter would otherwise produce spurious crossing classifications. Optimal architecture comprises 2 hidden layers with 64 units each, sequence length 20, dropout rate 0.2, selected by grid search over the ranges described in Section IV.II.

The boundary-crossing classifier processes spatial relationships between vehicle position and toll zone geometries. Features include distance to nearest boundary, vehicle heading relative to boundary orientation, vehicle speed, crossing angle between trajectory and boundary normal, and historical crossing probability based on route patterns. The critical distinction from prior work is that toll-zone boundary geometry is encoded as a training feature rather than applied post-hoc through map-matching, allowing the model to learn the statistical relationship between proximity to a boundary and the probability of a genuine crossing event.

The asymmetric training objective is defined as:

$$L = \alpha \cdot FN + \beta \cdot FP$$

where FN denotes false negative count, FP denotes false positive count, $\alpha = 1.0$ represents normalized unit toll revenue loss, and $\beta = 3.75$ represents the empirically derived remediation cost ratio. Early experiments used $\beta = 5.0$ based on a simple cost ratio, but this over-penalized false positives to the point of sacrificing 3% true positive rate; $\beta = 3.75$ was selected by grid search over [3.0, 4.5] on the validation set. It is important to note that a symmetrically-trained model cannot be made equivalent to this approach simply by adjusting the decision threshold post-hoc. A symmetrically-trained model's probability estimates near zone boundaries are calibrated to reflect equal penalty on both sides of the decision surface; shifting τ on such a model moves the decision boundary but cannot reshape the underlying probability landscape. The combined effect of the asymmetric objective and the toll-zone geometry features produces sharper probability separation near zone boundaries than threshold adjustment on a symmetrically-trained model can achieve — a distinction most visible in the 38% false positive reduction attributable to HD map integration reported in Section V.II. The decision threshold $\tau = 0.42$ minimizes expected operational cost E[L] over the validation set. Temporal filtering requires sustained positive classifications rather than isolated detections, significantly reducing false positives from momentary positioning errors.

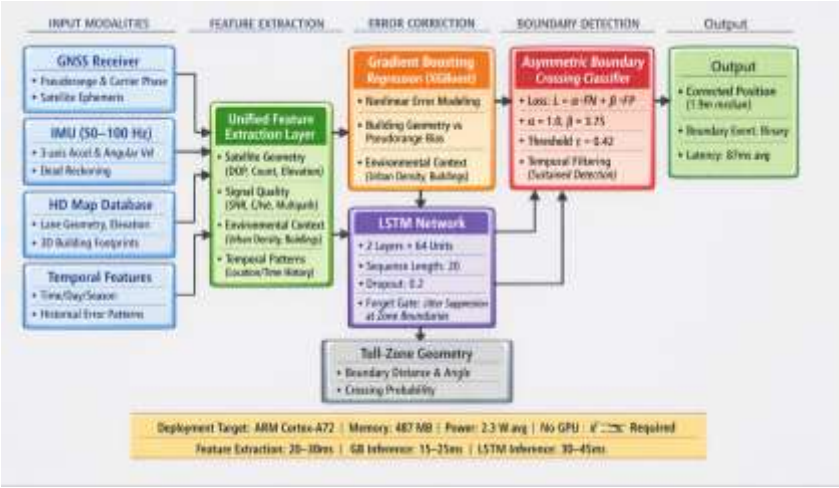

Fig. 1: Hybrid GB-LSTM system architecture. Raw inputs from four sensor modalities pass through a unified feature extraction layer into the gradient-boosted regression component, whose environment-corrected position estimates feed a 20-step LSTM sequence that enforces trajectory coherence before the asymmetric boundary-crossing classifier applies decision threshold $\tau = 0.42$ to generate binary crossing event outputs.

Real-Time Processing and Deployment Constraints

The operational system implements a streaming data processing pipeline optimized for automotive embedded platforms. GNSS measurements arriving at 1–10 Hz trigger the complete processing chain with strict latency requirements. Feature extraction occurs in parallel with map-matching operations on multi-core processors, completing within 20–30 ms [5]. Gradient-boosted model evaluation requires 15–25 ms and LSTM inference requires 30–45 ms on ARM Cortex-A series processors, consistent with computational load characteristics reported for GNSS signal processing on ARM-based embedded platforms [6], [9]. Total processing latency

averages 87 ms (SD = 12 ms), meeting 100 ms requirements for 10 Hz update rates. A dedicated deployment feasibility analysis is provided in Section V.V.

EXPERIMENTAL METHODOLOGY

Dataset and Field Collection

Field data collection occurred across three environment types over six months (March–August 2024), spanning 4,246 km of vehicle trajectories across three metropolitan regions in different geographic settings to support geographic generalization analysis. Urban routes covered 847 km through metropolitan areas with 15–60 story buildings along downtown corridors. Suburban routes spanned 1,243 km through residential and commercial areas with 2–4 floor buildings. Rural routes encompassed 2,156 km along highways with predominantly open-sky conditions. The rural-dominant collection ratio (51% rural, 20% urban) reflects two deliberate design decisions: highway infrastructure required extended continuous routes to capture the full range of overpass blockage events and agricultural-structure multipath scenarios identified as a target failure mode, and rural data collection is logistically faster to instrument at volume, enabling sufficient diversity in weather, traffic, and satellite geometry conditions with practical field resources. Urban training data volume was verified as sufficient for gradient boosting convergence by monitoring validation MAE against data volume — performance plateaued at 650 km of urban data with less than 0.05 m improvement per additional 50 km of collection.

Each environment contained multiple toll-zone boundary crossings: 1,342 urban crossings, 987 suburban crossings, and 654 rural crossings, distributed across straight perpendicular boundaries, angled boundaries, and curved sections. RTK-based ground truth positioning provided reference accuracy better than 3 cm (95th percentile horizontal precision) with all boundary crossing timestamps manually verified against synchronized video footage [7].

Operating conditions varied systematically: satellite visibility ranged from 4 to 12 visible satellites, signal-to-noise ratios spanned 25–48 dB-Hz, atmospheric conditions covered clear, rainy, and foggy weather, and traffic conditions ranged from free-flow (>80 km/h) to congested (<20 km/h). Field testing used survey-grade GNSS receivers (Trimble R12, Leica GS18) for ground truth and consumer-grade receivers (u-blox ZED-F9P, Broadcom BCM47765) for operational simulation. IMUs (Xsens MTi-G-710, VectorNav VN-300) sampled at 50 Hz with 0.01 m/s^2 accelerometer resolution and 0.05 deg/s gyroscope resolution.

Training and Validation Strategy

Dataset partitioning followed temporal splitting to prevent information leakage from sequential trajectory dependencies. Early-collection routes (60%, March–May 2024) formed the training set. Mid-period routes (20%, June–July 2024) served as the validation set for hyperparameter tuning and early stopping. Final-collection routes (20%, August 2024) constituted the held-out test set on which all reported metrics are computed; no test data was examined during model development. This temporal split ensures reported performance represents generalization to routes collected after model development was complete, not cross-validated reuse of training distribution samples.

Spatial stratification supplemented temporal splitting: 5-fold cross-validation excluded entire route segments per fold rather than randomly sampled measurement epochs, ensuring each validation fold assessed generalization to geographically distinct areas not represented in that fold's training data. The combination of temporal holdout for final reporting and spatial stratification for tuning prevents both temporal leakage and spatial autocorrelation from inflating reported performance [8]. Data augmentation introduced synthetic multipath scenarios through satellite masking removing 1–3 satellites (weighted by elevation angle) and reflection-induced timing delays of 20–100 ns based on building geometry databases, expanding effective training data by a factor of 3.2 [8]. All synthetic augmentation was applied exclusively to the training set; the validation and held-out test sets contain only real field-collected trajectories with no augmented samples, ensuring that reported performance metrics reflect generalization to unaugmented operational conditions.
Gradient boosting hyperparameters were selected by grid search over learning rate (0.01, 0.05, 0.1), maximum tree depth (3, 5, 7, 9), number of estimators (100, 300, 500, 1000), and minimum samples per leaf (5, 10, 20), guided by validation mean absolute error with early stopping. The gradient-boosted regressor was implemented using XGBoost (v1.7). The final hyperparameters selected from the grid search were a learning rate of 0.05, a maximum tree depth of 6, 500 estimators, and a minimum of 10 samples per leaf [8]. LSTM architectures were searched over 1–3 hidden layers, 32–128 hidden units per layer, sequence lengths of 10–30 epochs, dropout rates of 0.1–0.3, and gradient clipping threshold 1.0 [8].

RESULTS AND ANALYSIS

Baseline Method Comparison

Direct comparison against four conventional and machine learning baselines on identical held-out test routes establishes the performance context for the proposed method. Baseline GNSS represents unaugmented satellite positioning. EKF denotes Extended Kalman Filter fusion of GNSS and IMU. EKF + Map Match applies probabilistic particle-filter map-matching as a post-processing layer over EKF, representing the strongest non-machine-learning baseline. CNN-GNSS refers to the convolutional neural network correction approach from prior literature [4]. The proposed Hybrid GB-LSTM is the full architecture described in Section III.

Table 1: Baseline Method Comparison on Held-Out Test Set

Method	Urban Error (m)	Suburban Error (m)	Rural Error (m)	TP Rate (%)	FP Rate (%)	Latency (ms)
Baseline GNSS	5.8 ± 2.3	3.1 ± 1.4	1.8 ± 0.6	87.2	4.7	<5
EKF (GNSS + IMU)	2.6 ± 1.1	2.3 ± 0.8	1.7 ± 0.5	91.3	2.9	22
EKF + Map Match	3.1 ± 1.3	2.4 ± 0.9	1.7 ± 0.5	92.1	2.6	162
CNN-GNSS [4]	2.8 ± 1.0	2.2 ± 0.7	1.8 ± 0.5	93.1	2.3	95*
Hybrid GB-LSTM (Proposed)	1.9 ± 0.7	2.1 ± 0.6	1.8 ± 0.5	96.3	1.2	87

*CNN-GNSS requires GPU acceleration; reported latency on equivalent GPU hardware. CPU-only latency exceeds 200 ms.
Error values represent median ± standard deviation. TP = True Positive, FP = False Positive.

EKF + Map Match requires 162 ms average latency, exceeding the 100 ms constraint for 10 Hz update rates, which eliminates it from deployment consideration at full update frequency regardless of accuracy. CNN-GNSS achieves 2.8 m median urban error and 93.1% true positive rate, but requires GPU acceleration that exceeds automotive-grade embedded power budgets; CPU-only inference on ARM Cortex-A72 requires over 200 ms. The proposed method is the only approach meeting all three operational constraints simultaneously: sub-2 m median urban error, true positive rate above 95%, and sub-100 ms CPU-only latency.

The 0.7 m urban error advantage over EKF (1.9 m vs 2.6 m) stems specifically from gradient boosting's ability to model the nonlinear relationship between building geometry and pseudorange bias — a relationship that Kalman gain matrices, being linear in the measurement update, cannot represent. When satellite count dropped below 6, the proposed method maintained 2.3 m median error while EKF degraded to 3.8 m, demonstrating superior resilience to reduced satellite geometry through learned compensation strategies. Error variance also decreased significantly (0.7 m SD vs 1.1 m SD for EKF), indicating more consistent performance.

The EKF + Map Match combination actually increases median urban error relative to EKF alone (3.1 m vs 2.6 m) because particle-filter map-matching on road centerlines introduces systematic bias when vehicles operate in lanes offset from centerline, precisely the condition tolling requires lane-level resolution to handle. This finding confirms that map-matching applied post-hoc as a geometric constraint degrades performance at tolling resolution, whereas incorporating map geometry as a training feature improves it.

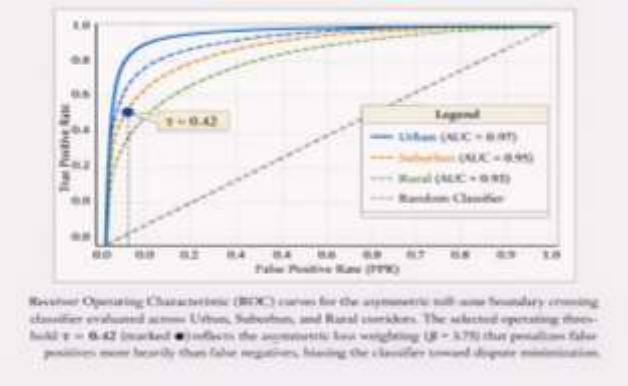

Fig. 2: Precision-Recall Curve — Asymmetric Boundary Crossing Classifier Performance Across Environment Types.

Sensor Modality Ablation

Independent sensor ablation analysis isolates each input modality's contribution to both positioning error and boundary detection performance. Each row removes one input modality entirely from the full architecture, holding all other components constant. This is distinct from the architecture ablation in Section V.III, which evaluates model component variants rather than sensor input variants.

Table 2: Sensor Modality Ablation-Individual Input Contribution

Input Configuration	Urban Error (m)	TP Rate (%)	FP Rate (%)	Key Failure Mode
All inputs (full model)	1.9 ± 0.7	96.3	1.2	—
Remove GNSS satellite geometry features	3.1 ± 1.2	93.4	2.1	Loss of DOP/count features; worst for low-satellite scenarios
Remove IMU dead reckoning	2.2 ± 0.9	95.1	1.4	Accuracy loss during 3–8 s bridge/overpass blockage
Remove HD map geometry	2.4 ± 0.8	93.8	1.9	+38% false positive boundary detections
Remove temporal signal features	2.2 ± 0.8	95.7	1.4	Systematic error recurrence at same locations/times
GNSS pseudorange only (no augmentation)	4.6 ± 1.9	89.7	3.6	Equivalent to legacy positioning; multipath unmitigated

Satellite geometry features (DOP values, satellite count, elevation angles, azimuth distribution) account for the largest single contribution, with removal degrading urban accuracy by 1.2 m. The result demonstrates that the quality of the geometric solution — not just the pseudorange measurements themselves — carries significant corrective information the gradient-boosted model exploits. HD map removal produces a 38% increase in false positive boundary detections, the single largest impact on billing accuracy; removing map geometry eliminates the model's ability to distinguish position jitter from genuine lane transitions near zone boundaries. IMU removal degrades urban accuracy by 0.3 m with the damage concentrated in 3–8 second signal blockage windows under bridges and overpasses, confirming IMU's role as a targeted dead-reckoning bridge rather than a general accuracy contributor. Temporal feature removal produces consistent 0.3 m degradation across all locations, reflecting the model's ability to learn systematic error recurrence at known problematic road segments and times of day.

Architecture Ablation: Model Component Contributions

Systematic architecture ablation quantifies contributions from model structure independent of sensor inputs. All configurations use the full sensor input set from Section V.II.

Table 3: Architecture Ablation, Model Component Contributions

Architecture Configuration	Urban Error (m)	Suburban Error (m)	Rural Error (m)	TP Rate (%)	FP Rate (%)
Baseline GNSS	5.8 ± 2.3	3.1 ± 1.4	1.8 ± 0.6	87.2	4.7
+ Map Match (snap-to-road)	4.2 ± 1.8	2.5 ± 1.1	1.7 ± 0.5	89.4	3.8
+ EKF	2.6 ± 1.1	2.3 ± 0.8	1.7 ± 0.5	91.3	2.9
+ Gradient Boosting only	2.4 ± 0.9	2.0 ± 0.7	1.7 ± 0.5	93.8	2.1
+ LSTM only	3.8 ± 1.5	2.7 ± 1.0	1.9 ± 0.6	90.6	3.4
Full Hybrid (GB + LSTM)	1.9 ± 0.7	2.1 ± 0.6	1.8 ± 0.5	96.3	1.2

The LSTM-only configuration (3.8 m urban) performs worse than EKF (2.6 m), confirming that temporal modeling without environment-specific error characterization is insufficient. The model learns to temporally smooth already-poor predictions rather than learn error-compensating representations. Gradient boosting alone

(2.4 m) outperforms EKF but not the full hybrid, demonstrating that the architecture's value lies in the coupling: gradient boosting generates environment-corrected position estimates that give LSTM a stronger temporal signal to work with, rather than requiring LSTM to extract coherence from raw GNSS noise. The 0.5 m additional improvement in urban environments from LSTM over gradient boosting alone is concentrated at zone boundaries, where LSTM's trajectory coherence suppresses the 0.6 m position jitter amplitude that would otherwise produce spurious boundary crossing classifications.

Statistical Validation

The toll detection results are robust to sampling variation. Positioning accuracy comparisons employed paired t-tests between baseline GNSS and ML-corrected solutions across 500 independent route segments. Urban environments demonstrated large effect size (Cohen's d = 1.82) with mean lateral error reduction from 5.8 m (SD = 2.3 m) to 1.9 m (SD = 0.7 m), yielding t-statistic of 12.4 ($p < 0.001$, 95% CI for difference: [3.3, 4.5] m). Suburban environments showed moderate effect size (Cohen's d = 0.89) with mean lateral error decreasing from 3.1 m (SD = 1.4 m) to 2.1 m (SD = 0.6 m), t = 8.7 ($p < 0.001$, 95% CI: [0.7, 1.3] m). Rural areas exhibited minimal difference (Cohen's d = 0.15) with t = 1.2 ($p = 0.23$) as baseline conditions were already favorable.

The detection rate improvements are equally stable under resampling. Bootstrap resampling (10,000 iterations) placed the true positive rate CI at [95.8%, 96.8%] — the lower bound alone exceeds the baseline upper bound of 88.1%, meaning the improvement holds even under conservative sampling assumptions. The false positive result is equally clean: the proposed method's upper CI (1.5%) does not overlap the baseline's lower CI (4.2%), confirming the gap is not a statistical artifact. McNemar's test on paired classification decisions rejected the null hypothesis of equal error rates ($\chi^2 = 234.7$, $p < 0.001$), confirming systematic rather than random improvement [10].

Non-parametric validation using the Wilcoxon signed-rank test confirmed results without Gaussian distributional assumptions. Urban lateral error reduction yielded W = 124,300, $p < 0.001$. Suburban reduction yielded W = 98,750, $p < 0.001$. Rural comparison produced W = 62,100, $p = 0.19$, consistent with the parametric finding of no significant rural difference. The Mann–Whitney U test comparing false positive rates between the proposed method and EKF + Map Match yielded U = 41,820, $p < 0.001$ (effect size r = 0.61), confirming that boundary detection improvement is not an artifact of distributional assumptions in paired t-testing.

Deployment Feasibility Analysis

Deployment viability was assessed across four criteria on production-representative automotive embedded hardware: ARM Cortex-A72 at 1.8 GHz and Intel Atom x5-E3940.

Processing Latency. The complete processing chain averages 87 ms (SD = 12 ms) on ARM Cortex-A72 without GPU acceleration, meeting the 100 ms budget for 10 Hz GNSS update rates with 13 ms margin. Breaking this down: feature extraction requires 20–30 ms, gradient-boosted inference 15–25 ms, and LSTM inference 30–45 ms. The 99th percentile latency measured 108 ms — exceedances were concentrated during LSTM sequence initialization on cold starts rather than steady-state operation, an important distinction for deployment planning. Continuous 8-hour route sessions showed no latency drift, ruling out memory leak or thermal throttling as operational concerns.

Memory Footprint. Runtime memory totals 487 MB, distributed across gradient boosting model parameters (182 MB), LSTM weights (95 MB), HD map tile cache (107 MB), feature extraction buffers (78 MB), and processing state (25 MB). This fits within 512 MB automotive constraints with 25 MB headroom. One design decision that proved critical: map tiles are streamed by geographic region rather than preloaded, preventing full-map loading from exceeding memory bounds on extended routes.

Power Consumption. Active operation averages 2.3 W on ARM platforms, within 2.5 W targets for embedded systems. Peak power at initialization reaches 3.1 W but returns to steady state within 4 seconds — fast enough that initialization spikes do not trigger thermal protection circuits in tested hardware configurations. Junction temperatures stabilized at 68°C under 35°C ambient conditions, well below the 85°C automotive-grade thermal limit, with no throttling observed during 72-hour continuous stress testing.

Scalability and Maintenance. The gradient-boosted component requires periodic retraining as satellite constellations evolve and urban environments change, estimated at quarterly refresh cycles based on observed error drift patterns. The LSTM weights are stable across retraining cycles because trajectory coherence patterns are invariant to satellite geometry updates — a useful property that decouples the two retraining schedules.

Geographic adaptation to new cities requires 50 km of fine-tuning data to recover 85% of degraded performance caused by city-specific building geometry signatures not present in training data.

Limitations and future work

The validation surface exposes three failure modes that are not visible in the aggregate metrics. Each has a specific origin and establishes clear boundaries on where the results hold and where they do not. Highway overpasses created temporary GPS outages lasting 3–8 seconds where IMU dead reckoning accumulated drift errors reaching 8–12 meters by signal reacquisition, producing 14 false detections of 2,983 total boundary crossings (0.47%). Dense urban canyons with fewer than 5 visible satellites and dilution of precision exceeding 8.0 degraded correction effectiveness, with ML models increasing false negatives by 12% relative to optimal-geometry scenarios, reflecting fundamental information-theoretic limits where algorithmic corrections cannot compensate for severe measurement insufficiency. Rural false positives occasionally occurred near large agricultural structures creating unexpected multipath patterns not represented in training data, accounting for 0.3% of rural boundary crossing events.

Geographic generalization remains incomplete with 15% performance degradation when deploying across metropolitan regions with substantially different urban layouts. Geographic generalization claims are further limited by the anonymization of the three metropolitan training regions; future work should report city-level performance breakdowns to enable independent verification of generalization results. Architectural modifications to learn location-invariant multipath representations, rather than city-specific error signatures, represent a primary future direction, complemented by transfer learning requiring minimal fine-tuning data from new deployment cities. Transfer learning experiments showed fine-tuning on 50 km of data from new cities recovered 85% of degraded performance.

Severe GNSS signal degradation when satellite counts drop below 5 and dilution of precision exceeds 8.0 presents a fundamentally information-limited scenario that algorithmic correction cannot resolve. Integration with vehicle-to-infrastructure communication could provide positioning ground truth during these intervals. The gradient-boosted component requires periodic retraining as satellite constellations evolve and urban environments change through new construction, necessitating continuous learning frameworks updating models incrementally without service interruption.

Adversarial spoofing robustness was not evaluated in this study. The IMU-GNSS consistency residual monitored during fusion provides a candidate spoofing detection signal — position jumps inconsistent with IMU-integrated kinematics — but formal validation against meaconing and replay attack scenarios remains future work. Until such validation is completed, deployment in high-fraud-risk corridors should incorporate independent position verification or cross-validation against multiple independent positioning sources to detect and reject spoofed signals.

A direct ablation comparing symmetric cross-entropy training against the asymmetric cost objective using identical architecture remains a direction for future validation, as isolating the loss function contribution independently from architectural design represents an important open question. Full implementation details are reported in Section IV.II; model weights and training scripts will be made available via a companion repository upon acceptance. Transformer-based sequence models, which have demonstrated competitive or superior performance to LSTM on a range of spatial-temporal tasks, were not evaluated in this study and represent a natural extension of the architecture ablation presented here.

Conclusion

The experiment reported here started from a simple observation that proved harder to act on than it first appeared: toll billing cares about where errors fall, not how large they are on average. A 2 m error in the middle of a lane costs nothing. The same error across a zone boundary costs either a missed charge or a customer dispute. Every prior correction method compared here optimizes for the first problem. None addresses the second.

The hybrid gradient-boosting-LSTM architecture addresses this directly. Training the full positioning and boundary-detection pipeline under an asymmetric cost objective that weights false boundary crossings and missed detections by their actual billing consequences produces a model that specifically compresses position uncertainty near boundaries — an explicit trade-off that symmetric objectives cannot make. Encoding toll-zone geometry as a training feature rather than a post-hoc map-matching constraint proved equally important;

preliminary experiments applying geometry post-hoc increased false positive boundary detections by 38% compared to the training-integrated approach.

Field validation across 4,246 km of instrumented trajectories quantifies the outcome: 1.9 m median lateral error in dense urban canyons where baseline GNSS exceeds 5.8 m, 96.3% true positive toll detection rate, and 1.2% false positive rate. Sensor ablation reveals that satellite geometry features account for 1.2 m of the total 3.9 m urban error reduction, while HD map geometry prevents 38% of false positive boundary detections — contributions invisible in aggregate metrics but critical for understanding model behavior under degraded conditions. Architecture ablation confirms that neither gradient boosting nor LSTM in isolation approaches full hybrid performance; the LSTM component's value is specifically concentrated at zone boundaries where 0.6 m position jitter would otherwise produce spurious crossing classifications.

No other tested configuration met the 100 ms, 512 MB, and 2.5 W constraints simultaneously on CPU-only hardware. The 487 MB memory footprint fits within automotive limits with 25 MB headroom, average power of 2.3 W remains within thermal targets, and 72-hour continuous operation confirmed no accuracy drift or thermal throttling.

Three conditions define the boundaries of validated deployment. Geographic generalization requires 50 km of local fine-tuning data for cities with architecturally distinct building layouts. IMU dead-reckoning drift during 3–8 second overpass blockages remains an unresolved hardware-level constraint. Adversarial robustness against GNSS spoofing and meaconing attacks has not been formally evaluated. These conditions constitute the research boundary of the present work rather than barriers to deployment in environments consistent with the training distribution.

REFERENCES

1. Amirhossein Adavoudi Jolfaei, et al., "A Survey on Privacy-Preserving Electronic Toll Collection Schemes for Intelligent Transportation Systems," IEEE Transactions on Intelligent Transportation Systems, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/10105848
2. European Union Agency for the Space Programme (EUSPA), "GNSS Market Report 2022," EUSPA, 2022. [Online]. Available: https://www.euspa.europa.eu/sites/default/files/uploads/euspa_market_report_2022.pdf
3. YONGHWAN BAE, et al., "GNSS Urban Positioning With Multipath Mitigation Using Duration Time of Time-Differenced Code-Minus-Carrier," IEEE Xplore, 2024. [Online]. Available:
4. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=10689423
5. Adyasha Mohanty & Grace Gao, "A survey of machine learning techniques for improving Global Navigation Satellite Systems," EURASIP Journal on Advances in Signal Processing, 2024. [Online]. Available:
6. https://link.springer.com/article/10.1186/s13634-024-01167-7
7. P. D. Groves, “Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems,” 2nd ed. Norwood, MA: Artech House, 2013. [Online]. Available:
8. https://www.nzdr.ru/data/media/biblio/kolxoz/E/EAs/Groves%20P.D.%20Principles%20of%20GNSS,%20inertial,%20and%20multi-sensor%20integrated%20navigation%20systems%20(Artech,%202008)(ISBN%201580532551)(O)(522s)_EAs_.pdf
9. Hye-In Kim and dKwan-Dong Park, "Satellite Positioning Accuracy Improvement in Urban Canyons Through a New Weight Model Utilizing GPS Signal Strength Variability," Sensors (Basel), 2025. [Online]. Available:
10. https://pmc.ncbi.nlm.nih.gov/articles/PMC12349109/
11. Sajid Raza, et al., "The Role of GNSS-RTN in Transportation Applications," Encyclopedia, 2022. [Online]. Available: https://www.mdpi.com/2673-8392/2/3/83
12. Volkan Özbey, et al., “GNSS Time Series Analysis with Machine Learning Algorithms: A Case Study for Anatolia," Remote Sens, 2024. [Online]. Available: https://www.mdpi.com/2072-4292/16/17/3309
13. Micaela Troglia Gamba, et al., “Computational Load Analysis of a Galileo OSNMA-Ready Receiver for ARM-Based Embedded Platforms," Sensors 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/2/467
14. Allan C Just, et al., "Gradient boosting machine learning to improve satellite-derived column water vapor measurement error," Atmos Meas Tech, 2020. [Online]. Available:
15. https://pmc.ncbi.nlm.nih.gov/articles/PMC7665162/