Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

SENTINEL-GNSS: Proactive Multi-Horizon Classification of GNSS Signal Degradation Using a Transformer-LSTM Network

Joel Segun Ojerinde
Beihang University
Ibrahim Ishola Mustapha
Beihang University
Joshua Oluwaseun Ogunlade
joshuaogunlade@buaa.edu.cn
Beihang University
Claudine Ingeriyinkindi
Beihang University
Md Harun Or Rashid
Beihang University
Wasiu Akande Ahmed
Beihang University

Research Article

Keywords: GNSS signal degradation, non-line-of-sight (NLOS), multi-horizon forecasting, Transformer-LSTM, urban canyon navigation, cross-city generalisation

Posted Date: October 8th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10329071/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

SENTINEL-GNSS: Proactive Multi-Horizon Classification of GNSS Signal Degradation Using a Transformer-LSTM Network

Joel Segun Ojerinde[1], Mustapha Ibrahim Ishola[1], Joshua Oluwaseun Ogunlade[1*], Claudine Ingeriyinkindi[1], Md Harun Or Rashid[1], Wasiu Akande Ahmed[1*]

[1*]Regional Centre for Space Science and Technology Education in Asia and the Pacific (RCSSTEAP), Beihang Hangzhou International Innovation Institute, Beihang University, Hangzhou, Zhejiang Province, China.

*Corresponding author(s). E-mail(s): joshuaogunlade@buaa.edu.cn; ahmedwasiu24@buaa.edu.cn;

Abstract

Urban GNSS signal degradation, caused by multipath and non-line-of-sight (NLOS) reception, disrupts positioning for autonomous systems. Existing detection methods operate at the current epoch; by the time degradation is identified, the navigation filter has already accepted corrupted measurements. We present SENTINEL-GNSS, a Transformer-LSTM network that simultaneously classifies GNSS signal quality at three forward horizons (5, 15, and 30 seconds) from a 30-second sliding window of 37 receiver-derived features. Trained on the Hong Kong UrbanNav dataset (62,413 windows, class-balanced focal loss), SENTINEL-GNSS achieves Macro-$\boldsymbol{F_1} = \mathbf{0.821}$ at the 5-second horizon (95% CI **[0.798, 0.840]**, MCC $=$ **0.773**), with a DEGRADED-class recall of **0.847**; this is 12 percentage points above both tree-based baselines on this safety-critical metric. In zero-shot evaluation on Tokyo Shinjuku UrbanNav (31,236 unseen windows), SENTINEL-GNSS retains DEGRADED $\boldsymbol{F_1} = \mathbf{0.753}$ while Random Forest collapses to **0.148**; a soft-vote ensemble with XGBoost reaches DEGRADED $\boldsymbol{F_1} = \mathbf{0.896}$. Inference runs at **0.045** ms per window from a 17.8 MB model, producing all three horizon outputs in a single forward pass.

These results demonstrate that proactive multi-horizon GNSS quality classification is achievable and provides actionable advance warning for navigation integrity management.

Keywords: GNSS signal degradation; non-line-of-sight (NLOS); multi-horizon forecasting; Transformer-LSTM; urban canyon navigation; cross-city generalisation

1 Introduction

Autonomous ground vehicles and unmanned aerial systems depend on GNSS as their primary absolute positioning reference. In open sky, GNSS provides reliable metre-level or sub-metre accuracy. In dense urban environments, however, signal quality degrades abruptly. Buildings obstruct satellite signals, producing multipath and non-line-of-sight (NLOS) reception; carrier-to-noise density (C/N_0) drops and the horizontal dilution of precision (HDOP) spikes (Groves 2013). These events introduce position jumps of several metres and corrupt the measurement model of navigation filters.

Current mitigation strategies are *reactive*: they detect degradation at the epoch when it occurs, then exclude the affected measurements or switch to an alternative sensor such as an inertial measurement unit (IMU) or 3D LiDAR (Wen and Hsu 2022). This approach has a fundamental latency problem. By the time degradation is confirmed, the navigation filter has already processed the corrupted observation; the extended Kalman filter (EKF) state and covariance have already been updated on bad data. Reverting to a consistent state requires backward smoothing or integrity resets, adding complexity and discontinuity to the position output. Advanced RAIM frameworks (Blanch et al. 2012) protect against single-fault positioning errors but similarly operate at the current epoch and do not anticipate future degradation.

Proactive classification, where the signal quality state at future epochs is predicted, would enable a navigation system to act before the degradation arrives: (A) downweight the upcoming GNSS contribution in the measurement covariance, (B)

pre-activate alternative sensors, and (C) flag a planned route segment for cautious traversal. Figure 1 illustrates the operational difference: a reactive monitor triggers only after a DEGRADED epoch is observed, while a proactive classifier issues the warning 5, 15, or 30 seconds in advance. The key question is whether such advance-warning classification is feasible from features computable in real time by a standard GNSS receiver.

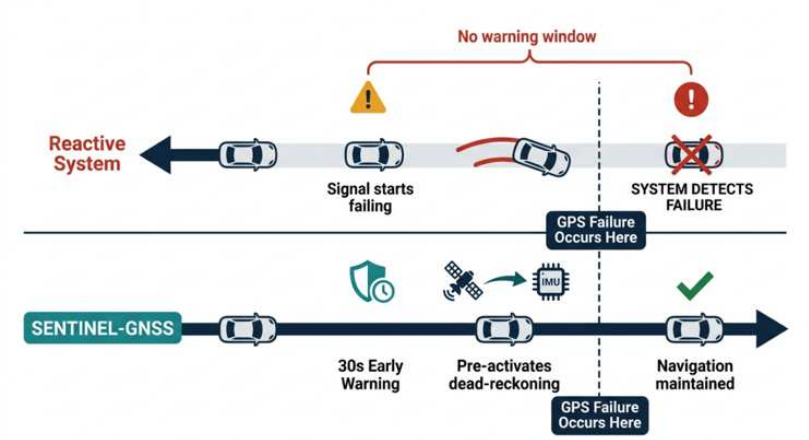

Fig. 1 Reactive vs. proactive GNSS signal quality monitoring. A reactive monitor triggers only after a DEGRADED epoch has been observed; the navigation filter has already accepted the corrupted measurement. A proactive classifier issues the alert 5, 15, or 30 s in advance, allowing the navigation system to prepare before the degraded epoch arrives.

Machine learning (ML) for GNSS signal characterisation has grown substantially (Siemuri et al. 2022; Xu et al. 2024). Classical models, notably Random Forests (Li et al. 2023; Breiman 2001) and gradient-boosted trees (Chen and Guestrin 2016; Zheng et al. 2024), achieve high accuracy for current-epoch NLOS and multipath detection. Deep learning has moved toward temporal modelling: (A) CNN-LSTM (Tao et al. 2021) for multipath characterisation, (B) Transformer-enhanced LSTM for current-epoch NLOS detection (Zhang et al. 2023), and (C) cascaded architectures for GNSS/INS integration (Zhang et al. 2026), which reduces position RMSE by 76.7%

in urban canyons. None of these works addresses simultaneous multi-horizon signal quality *classification* from a unified model.

We present SENTINEL-GNSS, a Transformer-LSTM architecture trained jointly on three prediction horizons. The contributions are:

1. A Transformer encoder combined with a two-layer LSTM decoder that simultaneously produces signal quality classifications at 5 s, 15 s, and 30 s from a 30-second feature window of 37 GNSS receiver-derived features (§4).
2. A training protocol with class-balanced focal loss on class-imbalanced data, achieving DEGRADED-class recall of 0.847 at 5 s; this is 12 percentage points above both tree baselines on the minority safety-critical class, while running at 0.045 ms per window from a 17.8 MB model (§6).
3. A zero-shot cross-city evaluation on Tokyo Shinjuku UrbanNav, where SENTINEL-GNSS retains DEGRADED $F_1 = 0.753$ while Random Forest collapses to 0.148, and a DL+XGBoost soft-vote ensemble reaches 0.896 (§6).
4. A systematic inclusion of the persistence baseline (trivially copying the current-epoch label), which achieves Macro-$F_1 = 0.908$ at 5 s from pure autocorrelation, establishing the bar every forecasting model must clear (§5).

Section 2 reviews related work. Section 3 describes the dataset and feature engineering. Section 4 presents the SENTINEL-GNSS architecture. Section 5 details the experimental protocol. Section 6 reports results. Section 7 provides discussion, and Section 8 concludes.

2 Related Work

2.1 Signal quality monitoring

Traditional GNSS signal quality monitoring (SQM) relies on pseudorange residuals and dilution-of-precision thresholds computed at each epoch (Groves 2013). Receiver

Autonomous Integrity Monitoring (RAIM) and its advanced variants (Blanch et al. 2012) compute protection levels bounding position error at a specified fault probability but do not forecast quality at future epochs. Integrity monitoring for urban air mobility has been adapted to ML frameworks (Isik and Gao 2024), focusing on current-epoch confidence. None of these approaches provides advance warning of forthcoming degradation.

2.2 Machine learning for NLOS and multipath detection

Classical ML has produced accurate NLOS and multipath classifiers. Xu et al. (2019) apply Support Vector Machines to multi-constellation GNSS features, reporting consistent LOS/NLOS accuracy in urban Hong Kong tests. Li et al. (2023) use Random Forest with pseudorange weighting features. Zheng et al. (2024) apply stacking ensemble learning and report competitive NLOS classification in GPS Solutions. Ozeki and Kubo (2022) combine SVM with a pseudorange residual check and a snapshot-feature sliding window in static downtown Tokyo, achieving accuracy above 88%; the static-only evaluation limits the applicability to dynamic urban canyons. A systematic review of 213 studies by Siemuri et al. (2022) finds that RF and gradient-boosted trees generally lead accuracy on GNSS classification tasks. All of these methods produce single-epoch predictions and do not address multi-horizon forecasting.

2.3 Deep learning for GNSS temporal modelling

Deep learning enables temporal modelling of GNSS signal sequences. Tao et al. (2021) use CNN-LSTM for multipath characterisation. Zhang et al. (2021) apply LSTM to predict GNSS pseudorange uncertainty over short future intervals, framing the problem as regression rather than discrete classification. Park et al. (2024) exploit pseudorange acceleration features for multipath detection at the current epoch. Zhang et al. (2023) construct a Transformer-enhanced LSTM for simultaneous NLOS classification and pseudorange error prediction, achieving NLOS precision of 0.80 and

accuracy of 0.87 on their labelled Hong Kong dataset. Their model targets the current epoch; the horizon extension to 5 s–30 s classification is not addressed. A recent GPS Solutions study (Zhang et al. 2026) presents a cascaded Transformer-LSTM combined with factor graph optimisation for GNSS/IMU integration in urban canyons, reducing position RMSE by 76.7% on a public benchmark. That architecture is the closest structural precursor to SENTINEL-GNSS, but the task differs fundamentally: it targets navigation accuracy under GNSS degradation (a filtering and estimation task), whereas SENTINEL-GNSS predicts the signal quality state before the degradation occurs (a proactive classification task).

2.4 Multi-horizon forecasting in positioning

Multi-step forecasting of continuous positioning quantities has been studied in pedestrian dead reckoning and trajectory prediction, but multi-horizon *classification* of discrete GNSS signal quality states has not been systematically addressed. Park et al. (2024) note that pseudorange acceleration provides early indication of multipath onset but do not extend to multi-class horizon prediction. The present work fills this gap with a unified multi-head classification framework evaluated at three operationally meaningful horizons.

3 Dataset and Feature Engineering

3.1 UrbanNav dataset

We use the UrbanNav open-source multisensory dataset (Hsu et al. 2023), collected in dense urban canyons across Hong Kong (Kowloon and New Territories) and Tokyo

(Shinjuku) with multiple GNSS receivers: the Hong Kong routes employ u-blox ZED-F9P and NovAtel Flexpak6 units; the Tokyo Shinjuku route employs a Trimble dual-frequency receiver. Raw GNSS observations are processed with RTKLIB 2.4.3 (single-point positioning, SPP) to produce `.pos` solution files recording position estimates and solution quality flags at 1-second intervals.

The *in-domain* dataset comprises four urban driving routes. After feature extraction and windowing (see below), this yields 82,173 labelled 1-second epochs partitioned as: training 62,413 windows, validation 18,074 windows, and test 1,686 windows (731 CLEAN · 746 WARNING · 209 DEGRADED). The *cross-city* evaluation dataset is UrbanNav Tokyo Shinjuku, processed with the same RTKLIB 2.4.3 SPP pipeline, yielding 31,236 windows (29,200 CLEAN · 1,620 WARNING · 416 DEGRADED). Tokyo data were held out entirely during model development and hyperparameter selection. The receiver change (ublox/NovAtel in training; Trimble in Tokyo) introduces a systematic difference in C/N_0 baselines and fix-quality flag encoding; this constitutes a combined geographic and receiver-type generalisation test, discussed further in §7.2.

Figure 2 illustrates the end-to-end system pipeline from raw GNSS data through feature extraction, SENTINEL-GNSS inference, and downstream navigation integration.

3.2 Label generation

At each epoch t, RTKLIB provides the 2D horizontal position standard deviation $\rho_t = \sqrt{\sigma_N^2 + \sigma_E^2}$ (m), the mean carrier-to-noise density ratio $\bar{C}_{N_0,t}$ across tracked satellites (dB-Hz), and the satellite count $N_{sv,t}$. Three mutually exclusive signal quality classes

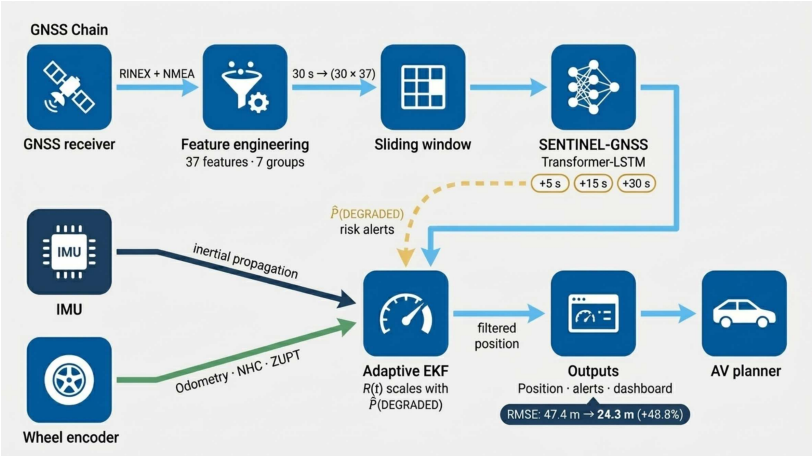

Fig. 2 End-to-end system pipeline. Raw GNSS observations are processed by RTKLIB to produce `.pos` solution files; 37 features per epoch are extracted and assembled into 30-second sliding windows; SENTINEL-GNSS produces three forward-horizon probability vectors in a single forward pass; the predicted quality labels feed a downstream adaptive EKF.

are assigned by a priority rule:

$$y_t = \begin{cases} \text{CLEAN} & \text{if } \rho_t < 2.0 \text{ m AND } \bar{C}_{N_0,t} > 35 \text{ dB-Hz AND } N_{sv,t} \geq 6 \\ \text{DEGRADED} & \text{if } \rho_t > 5.0 \text{ m OR } \bar{C}_{N_0,t} < 30 \text{ dB-Hz OR } N_{sv,t} < 4 \\ \text{WARNING} & \text{otherwise,} \end{cases} \tag{1}$$

where any single DEGRADED criterion overrides WARNING. The position-error thresholds (2.0 m / 5.0 m) follow horizontal alert limit conventions in integrity monitoring (Blanch et al. 2012); the C/N_0 thresholds (35 / 30 dB-Hz) align with ITU-R M.1168 GNSS signal reliability guidelines; and the satellite-count thresholds (6 / 4) reflect standard RAIM geometry requirements. Horizon labels y_{t+h} for $h \in \{5, 15, 30\}$ (seconds) apply Equation (1) at epoch $t + h$. In addition to the three primary criteria, the rule applies two receiver-status checks computed over the window, mirroring the RTKLIB quality flag: a DEGRADED label is also triggered when the solution is predominantly single-point or the fix-continuity ratio is low, and the CLEAN

label additionally requires a predominantly fixed solution. These are omitted from Equation (1) for readability.

Threshold sensitivity. Re-labelling the full corpus under perturbed thresholds confirms the rule is not hypersensitive to its boundaries. Only 3.2% of epochs lie within 1 m of the 5 m position boundary; shifting that boundary by ±1 m changes fewer than 1% of labels and the DEGRADED prevalence by under 1.6 percentage points. The satellite-count boundary is equally insensitive, changing under 0.5% of labels across 3 to 5 satellites. The C/N_0 boundary is the most influential input, moving the DEGRADED prevalence from 27.3% to 39.2% across 28 to 32 dB-Hz, consistent with C/N_0 being the dominant physical degradation signal.

3.3 Feature extraction

From each RTKLIB `.pos` epoch, 35 raw features are extracted across seven groups (Table 1). Feature preparation removes latitude and longitude (to prevent geographic overfitting) and adds four derived indicators: CNR availability flag (*cnr_available*), rate-of-change of PDOP (ΔPDOP), rate-of-change of HDOP (ΔHDOP), and a receiver-tier indicator (*receiver_tier*). The resulting 37-dimensional feature vector is assembled for each 1-second epoch.

A sliding window of width $W = 30$ seconds (with one epoch per second) produces input tensors $\boldsymbol{X} \in \mathbb{R}^{30\times 37}$. Window step size is 1 epoch, yielding dense coverage of each route. No feature normalisation is applied to the raw extraction; batch normalisation is applied inside the model at the input projection layer.

4 The SENTINEL-GNSS Architecture

Figure 3 shows the overall architecture. SENTINEL-GNSS processes the input window through three sequential stages: a linear input projection, a Transformer encoder, and

Table 1 Feature taxonomy: 37 features extracted from RTKLIB `.pos` files. Derived features are computed by `feature_prep.py`.

Group	Raw	Derived	Representative features
Positioning solution quality	3	0	alt, σ_{lat}, σ_{lon}
Carrier-to-noise density (C/N_0)	5	1	mean, min, max, std, trend; availability flag
Satellite availability	5	0	count, mean, min, visibility, drop rate
Dilution of precision	5	2	PDOP, HDOP, VDOP, GDOP, ratio; ΔPDOP, ΔHDOP
Receiver status	5	1	fix type, baseline sats, age, continuity, transitions; tier
Signal dynamics	5	0	position variance, C/N_0 variance, elevation violations
Atmospheric and residual	5	0	iono/tropo delay, cycle slips, residual mean/std
Total	**33**	**4**	**37 features**

a two-layer LSTM decoder. Four independent classification heads branch from the final LSTM hidden state.

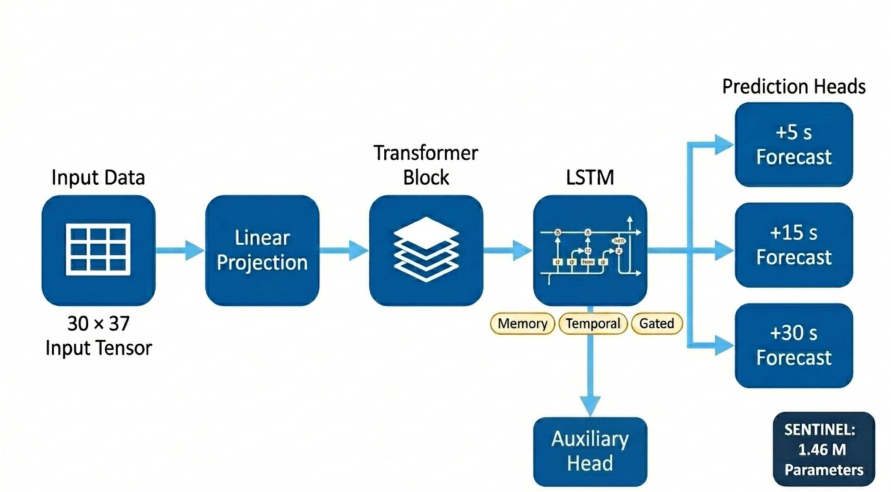

Fig. 3 SENTINEL-GNSS architecture. A 30×37 input window is projected to $d_{model} = 128$, processed by a two-layer Transformer encoder ($H = 8$ heads, $d_{ff} = 512$), and decoded by a two-layer LSTM (hidden dimension 256). Four independent classification heads branch from the final LSTM hidden state, producing simultaneous quality predictions at t+0 (auxiliary), t+5 s, t+15 s, and t+30 s. Total parameters: 1,456,652.

4.1 Input projection

The input tensor $\boldsymbol{X} \in \mathbb{R}^{W \times d_{in}}$ (with $W = 30$, $d_{in} = 37$) is projected to model dimension $d_{model} = 128$ via a position-wise linear layer followed by batch normalisation. A fixed sinusoidal positional encoding (Vaswani et al. 2017) is then added so that the Transformer can distinguish the temporal ordering of epochs:

$$\mathrm{PE}(pos, 2k) = \sin\!\big(pos/10000^{2k/d_{model}}\big), \quad \mathrm{PE}(pos, 2k+1) = \cos\!\big(pos/10000^{2k/d_{model}}\big), \tag{2}$$

where $pos \in \{0, \ldots, W-1\}$ is the epoch index within the window and k indexes the dimension pairs.

4.2 Transformer encoder

The encoder consists of $L_T = 2$ layers, each with multi-head self-attention ($H = 8$ heads) and a position-wise feed-forward network ($d_{ff} = 512$, ReLU activation). Following Vaswani et al. (2017), scaled dot-product attention for a single head is

$$\mathrm{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \mathrm{softmax}\!\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\top}}{\sqrt{d_k}}\right)\boldsymbol{V}, \tag{3}$$

where $\boldsymbol{Q}$, $\boldsymbol{K}$, $\boldsymbol{V} \in \mathbb{R}^{W \times d_k}$ are the query, key, and value projections, and $d_k = d_{model}/H = 16$. Multi-head attention concatenates H heads and projects back to d_{model}. Residual connections and layer normalisation surround each sub-layer; dropout of 0.1 is applied.

The Transformer models global temporal dependencies within the 30-second window: it can attend to the onset of C/N_0 decline at epoch 5 while simultaneously attending to satellite-count changes at epoch 25, regardless of their temporal distance.

4.3 LSTM decoder

The sequence of W Transformer output vectors is passed to a two-layer, unidirectional LSTM (Hochreiter and Schmidhuber 1997) with hidden dimension $h_{dim} = 256$ and inter-layer dropout of 0.3. The LSTM hidden state update is:

$$\begin{aligned}
\boldsymbol{f}_t &= \sigma(\boldsymbol{W}_f\,[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_f), \quad \boldsymbol{i}_t = \sigma(\boldsymbol{W}_i\,[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_i), \\
\boldsymbol{o}_t &= \sigma(\boldsymbol{W}_o\,[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_o), \quad \tilde{\boldsymbol{c}}_t = \tanh(\boldsymbol{W}_c\,[\boldsymbol{h}_{t-1}, \boldsymbol{x}_t] + \boldsymbol{b}_c), \\
\boldsymbol{c}_t &= \boldsymbol{f}_t \odot \boldsymbol{c}_{t-1} + \boldsymbol{i}_t \odot \tilde{\boldsymbol{c}}_t, \quad \boldsymbol{h}_t = \boldsymbol{o}_t \odot \tanh(\boldsymbol{c}_t),
\end{aligned} \tag{4}$$

where $\sigma(\cdot)$ is the sigmoid function and $\odot$ denotes element-wise multiplication. The LSTM supplies causal state-tracking that the Transformer's bidirectional attention does not: it accumulates signal history directionally and suppresses the false-alarm rate on DEGRADED predictions. The final hidden state $\boldsymbol{h}_T \in \mathbb{R}^{256}$ summarises the entire window.

4.4 Multi-horizon classification heads

Four independent linear classification heads branch from $\boldsymbol{h}_T$. Three primary heads target the prediction horizons $h \in \{5\,\text{s}, 15\,\text{s}, 30\,\text{s}\}$; a fourth auxiliary head targets the current epoch ($h = 0$) to regularise the shared representation:

$$\hat{\boldsymbol{p}}_{t+h} = \text{softmax}(\boldsymbol{W}_h\,\boldsymbol{h}_T + \boldsymbol{b}_h)\,, \quad h \in \{0, 5, 15, 30\}, \tag{5}$$

where $\hat{\boldsymbol{p}}_{t+h} \in \mathbb{R}^3$ is the predicted class probability vector for the horizon-h epoch. The predicted label is $\hat{y}_{t+h} = \arg\max \hat{\boldsymbol{p}}_{t+h}$.

The total parameter count is 1,456,652 (Transformer: 427,273; LSTM: 1,026,569; heads: 2,810).

4.5 Training objective

Class imbalance (CLEAN:WARNING:DEGRADED ≈ 18%:60%:22% in training) motivates focal loss (Lin et al. 2017):

$$\mathcal{L}_{\text{focal}} = -\sum_{c=1}^{3} \alpha_c \big(1 - \hat{p}_c\big)^{\gamma} \log \hat{p}_c, \tag{6}$$

where $\gamma = 1.0$ reduces the loss contribution of well-classified examples and $\alpha_c \in \{1.0, 2.0, 5.0\}$ for CLEAN, WARNING, and DEGRADED respectively are manually set class weights that prioritise the minority safety-critical class above what inverse-frequency weighting alone would provide. Label smoothing ($\varepsilon = 0.1$) regularises the hard one-hot targets. The combined training loss sums across heads:

$$\mathcal{L} = \mathcal{L}_{+5\text{s}} + \mathcal{L}_{+15\text{s}} + \mathcal{L}_{+30\text{s}} + \lambda_{\text{aux}}\, \mathcal{L}_{+0\text{s}}, \tag{7}$$

with $\lambda_{\text{aux}} = 0.3$ reducing the auxiliary head's contribution while retaining its regularising effect.

5 Experimental Protocol

5.1 Training configuration

SENTINEL-GNSS is trained with the AdamW optimiser (Kingma and Ba 2015) (learning rate 10^{-3}, weight decay 10^{-4}, $\beta_1 = 0.9$, $\beta_2 = 0.999$), with cosine annealing scheduled after a 5-epoch linear warmup, gradient clipping at max-norm 1.0, and early stopping (patience 50, minimum 15 epochs). Batch size is 256. Training ran for 65 epochs; the best checkpoint was at epoch 10, selected by combined validation Macro-F_1. Training uses two NVIDIA Tesla T4 GPUs. No SMOTE oversampling is applied

to SENTINEL-GNSS; class balance is handled by focal loss and class weights alone (see §6.1).

5.2 Baselines

Six baselines are evaluated:

1. *Majority class*: assigns every sample the most frequent class (WARNING).
2. *CNR threshold*: rule-based single-threshold classifier on mean C/N_0.
3. *Persistence*: predicts $\hat{y}_{t+h} = y_t$ (current label copied to all horizons). This is a strong check for proactive models: at the 5 s horizon, 94.4% of labels are unchanged, so persistence achieves Macro-$F_1 = 0.908$ purely from autocorrelation. Any trained model that does not exceed persistence has not learned to forecast; it has learned to copy.
4. *Random Forest (RF)*: 200 trees, `class_weight='balanced'`, separate models per horizon, trained on 62,413 windows.
5. *XGBoost*: 200 trees, `scale_pos_weight` tuned per class, trained on 62,413 windows.
6. *Ablation variants*: Transformer-only (encoder head directly) and LSTM-only (skip the Transformer encoder).

We also train SMOTE-augmented RF and XGBoost variants (112,482 windows) to assess whether oversampling benefits tree-based methods. Tree baselines use three separate single-horizon models; SENTINEL-GNSS uses one unified model for all horizons.

5.3 Evaluation metrics

Primary: Macro-F_1 (unweighted mean across classes). Secondary: Matthews Correlation Coefficient (MCC) (Chicco and Jurman 2020), Cohen's κ (Cohen 1960), and per-class precision, recall, and F_1. All SENTINEL-GNSS metrics are reported with 95% bootstrap CIs ($n = 1{,}000$). Chicco and Jurman (2020) show that MCC is more

informative than F_1 under class imbalance; we report both. DEGRADED recall is the primary safety metric: a low DEGRADED recall means the model fails to alert the navigator before a safety-critical epoch.

Evaluation protocol and window independence. The test set is constructed with a 1-second step; consecutive windows therefore share 29 of 30 input seconds. However, each window's label is assigned to a distinct future epoch (y_{t+h} for unique t), so the targets are non-redundant and the set spans all test scenarios. Bootstrap CIs use standard i.i.d. resampling of windows, which treats each window as independent and therefore slightly underestimates CI width relative to a block bootstrap that accounts for inter-window correlation. A block bootstrap (block length $b = 30$, matched to the window length) widens the 5 s Macro-F_1 interval to [0.722, 0.895], from the i.i.d. [0.800, 0.842], and the DEGRADED recall interval to [0.653, 0.934], from [0.787, 0.883]. The point estimates are unchanged, and the block-bootstrap lower bounds remain far above the majority-class (0.202) and single-threshold (0.074) baselines, so the qualitative conclusions, which rest on the point estimates, hold even under the conservative interval.

6 Results

6.1 In-domain performance at the +5 s horizon

Table 2 reports full in-domain results at the 5 s horizon. Several findings stand out.

Persistence is a formidable baseline. At 5 s, the class label is unchanged from the current epoch in 94.4% of windows; persistence scores Macro-F_1 = 0.908, κ = 0.880, MCC= 0.920. This is the autocorrelation ceiling that any proactive classifier must beat to claim genuine forecasting ability. Both tree methods (RF: 0.910; XGBoost: 0.919) marginally exceed persistence on aggregate Macro-F_1 while SENTINEL-GNSS achieves 0.821, below persistence. The *reason* SENTINEL-GNSS is useful despite lower aggregate Macro-F_1 is discussed in §7.1.

Tree methods lead on aggregate metrics. XGBoost no-SMOTE achieves the highest Macro-F_1 (0.919) and MCC (0.926), followed closely by RF no-SMOTE (0.910, 0.916). SMOTE training reduces XGBoost Macro-F_1 by 0.010 (from 0.919 to 0.910) and has negligible effect on RF ($\Delta = -0.0009$): with `class_weight='balanced'` already handling imbalance, SMOTE's synthetic interpolation in the 1,110-dimensional flattened feature space adds noise that marginally hurts XGBoost.

SENTINEL-GNSS leads on DEGRADED recall. SENTINEL-GNSS achieves DEGRADED recall of **0.847** (precision = 0.623, $F_1 = 0.718$, 95% CI [0.671, 0.762]); this is 12 percentage points above both RF (0.727) and XGBoost (0.722). For safety-critical applications where failing to alert before a DEGRADED epoch is the worst failure mode, this recall advantage is decisive.

6.2 Multi-horizon analysis

Table 3 reports Macro-F_1 at all three prediction horizons for all methods. Figure 4 shows SENTINEL-GNSS's Macro-F_1 and MCC across the three horizons.

SENTINEL-GNSS achieves Macro-F_1 = 0.821, 0.741, 0.783 at 5 s, 15 s, and 30 s, respectively. The non-monotone dip at 15 s reflects a class-level transition: at this horizon, WARNING F_1 drops to 0.722 (from 0.817 at 5 s) while DEGRADED F_1 drops to 0.583; both partially recover at 30 s (WARNING: 0.807; DEGRADED: 0.629). Tree baselines exhibit the same non-monotone pattern (RF: 0.910 → 0.891 → 0.896; XGBoost: 0.919 → 0.896 → 0.879), confirming this is a property of the label dynamics in the dataset rather than a model artefact.

Persistence decays more steeply with horizon (Macro-F_1: 0.908 → 0.871 → 0.826) as inter-epoch label changes increase from 5.6% at 5 s to 10.7% at 30 s. At 30 s, both tree methods still substantially exceed persistence ($\Delta \approx 0.05$–0.07), confirming genuine predictive ability at longer horizons.

Table 2 In-domain performance at the 5 s horizon on 1,686 test windows (731 CLEAN · 746 WARNING · 209 DEGRADED). All SENTINEL-GNSS metrics include 95% bootstrap CI ($n = 1{,}000$). DEG-Rec: DEGRADED-class recall; DEG-F_1: DEGRADED-class F_1. Best value in each column is **bold**.

Method	Training	Macro-F_1	MCC	κ	DEG-Rec	DEG-F_1
Majority class	n/a	0.202	0.000	0.000	0.000	0.000
CNR threshold	n/a	0.074	0.000	0.000	n/a	n/a
Persistence	n/a	0.908	0.920	0.880	n/a†	n/a†
RF	SMOTE 112K	0.909	0.915	n/a	n/a	n/a
XGBoost	SMOTE 112K	0.910	0.915	n/a	n/a	n/a
RF	no-SMOTE 62K	0.910	0.916	n/a	0.727	0.789
XGBoost	no-SMOTE 62K	**0.919**	**0.926**	n/a	0.722	0.810
LSTM-only	no-SMOTE	0.767	0.702	n/a	n/a	0.645
Transformer-only	no-SMOTE	0.767	0.725	n/a	n/a	0.571
SENTINEL-GNSS	**no-SMOTE 62K**	0.821	0.773	0.762	**0.847**	0.718
95% CI		[0.798,0.840]				[0.671,0.762]

† Persistence copies the current label; its DEG-Rec and DEG-F_1 at +5 s are inflated by 94.4% label stability and do not represent predictive capacity.

Confusion matrices at all three horizons are shown in Figure 5. Across all horizons, the most common SENTINEL-GNSS error is WARNING↔CLEAN confusion; CLEAN→DEGRADED errors (the safety-critical false negative) remain rare at all horizons.

6.3 Ablation study

Table 4 and Figure 6 compare three architecture variants. The full SENTINEL-GNSS achieves the best Macro-F_1 at 5 s (0.821) and the highest DEGRADED F_1 (0.718) and MCC (0.773), with 1,457K parameters.

Neither component alone replicates the full model's performance. The Transformer-only variant achieves CLEAN precision = 1.000 (it never predicts DEGRADED as

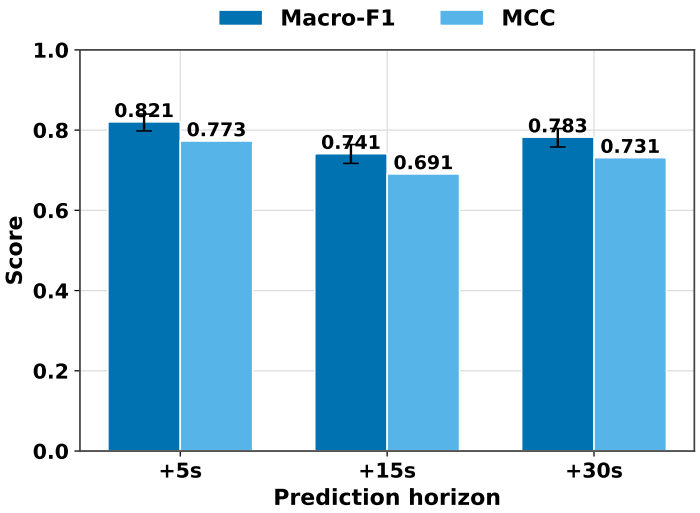

Fig. 4 SENTINEL-GNSS Macro-F_1 and MCC at the three prediction horizons (5 s, 15 s, 30 s). Error bars on Macro-F_1 denote 95% bootstrap confidence intervals. The non-monotone dip at 15 s appears across all models (Table 3) and reflects label dynamics in the dataset rather than a model artefact.

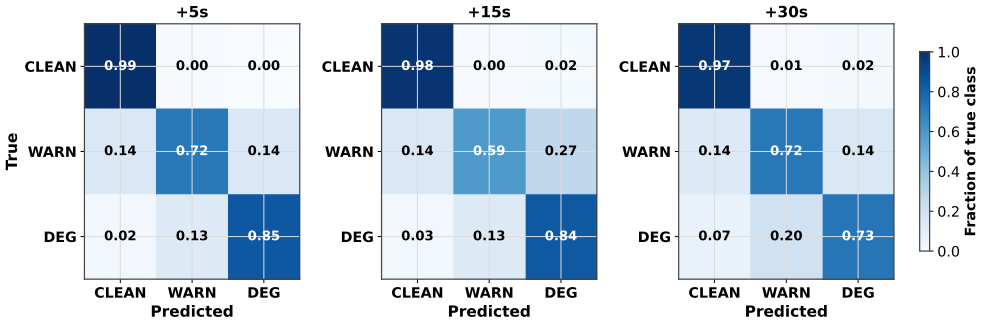

Fig. 5 Row-normalised confusion matrices for SENTINEL-GNSS at the 5 s (left), 15 s (centre), and 30 s (right) horizons. The dominant errors are WARNING↔CLEAN confusions; CLEAN→DEGRADED errors (upper-right off-diagonal, the safety-critical false negative) remain rare across all horizons.

CLEAN) but DEGRADED precision = 0.424; it over-flags degradation. The LSTM-only variant provides better DEGRADED F_1 than Transformer-only (0.645 vs 0.571) but lower Macro-F_1 at 5 s and MCC. The full Transformer-LSTM combination reaches the superior balance because the Transformer supplies global temporal context across the 30-second window while the LSTM provides causal directional state-tracking that suppresses false alarms.

Table 3 Multi-horizon Macro-F_1. Bootstrap 95% CIs for SENTINEL-GNSS are given in square brackets. The dip at 15 s is observed across all models and reflects label-dynamic properties of the dataset.

Method	Training	+5 s Macro-F_1	+15 s Macro-F_1	+30 s Macro-F_1
Persistence	n/a	0.908	0.871	0.826
RF	no-SMOTE 62K	0.910	0.891	0.896
XGBoost	no-SMOTE 62K	0.919	0.896	0.879
LSTM-only	no-SMOTE	0.767	0.751	0.781
Transformer-only	no-SMOTE	0.767	0.763	0.701
SENTINEL-GNSS	**no-SMOTE 62K**	**0.821**	**0.741**	**0.783**
		[0.798,0.840]	[0.717,0.764]	[0.758,0.804]

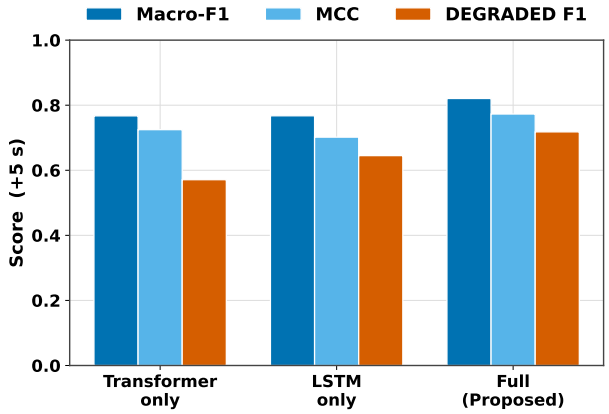

Fig. 6 Ablation: Macro-F_1 and DEGRADED F_1 at the 5 s horizon for the Transformer-only, LSTM-only, and full Transformer+LSTM variants of SENTINEL-GNSS. Neither component alone reproduces the full model's performance; together they achieve the best balance of aggregate and minority-class metrics.

6.4 Cross-city generalisation

Table 5 and Figure 7 report zero-shot evaluation on Tokyo Shinjuku. Three findings are critical.

Table 4 Ablation study at the 5 s horizon. Params: trainable parameters. DEG-F_1: DEGRADED-class F_1.

Architecture	Params	Val stop-F_1	Macro-F_1	DEG-F_1	MCC
Transformer-only	427K	0.860	0.767	0.571	0.725
LSTM-only	1,027K	0.864	0.767	0.645	0.702
Transformer + LSTM	**1,457K**	**0.861**	**0.821**	**0.718**	**0.773**

Random Forest collapses on the safety-critical class. RF achieves DEGRADED $F_1 = 0.148$ in Tokyo, compared to 0.822 in-domain: a drop of 67 percentage points. RF memorises region-specific feature distributions, particularly satellite geometry and C/N_0 baselines unique to Kowloon, and cannot transfer them.

XGBoost does not collapse. XGBoost cross-city Macro-$F_1 = 0.821$, DEGRADED $F_1 = 0.784$. The RF collapse is model-specific, not a property of tree-based methods in general; the more regularised XGBoost tree structure avoids overfitting to training-city conditions. Statements that "tree-based models collapse cross-city" are not supported by these data.

SENTINEL-GNSS retains DEGRADED transfer; the ensemble is optimal. SENTINEL-GNSS cross-city Macro-$F_1 = 0.649$, DEGRADED $F_1 = 0.753$. The Transformer's global attention over signal dynamics and the LSTM's causal state-tracking capture features that are more city-agnostic than RF's feature-importance shortcuts. A DL+XGBoost soft-vote ensemble achieves cross-city Macro-$F_1 = 0.892$ and DEGRADED $F_1 = \mathbf{0.896}$, outperforming every single model on both metrics. DL and XGBoost make complementary errors on the unseen city; averaging their predicted probabilities corrects both. The soft-vote ensemble is the recommended deployment configuration.

6.5 Navigation integration

A GNSS signal quality classifier produces actionable value only when its predictions improve downstream navigation. We demonstrate this through a prediction-informed

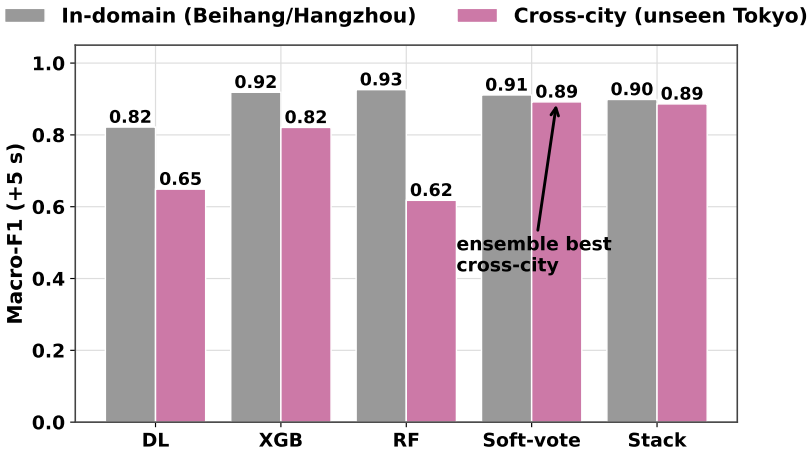

Fig. 7 In-domain versus zero-shot cross-city Macro-F_1 at 5 s for five methods: deep learning (DL, SENTINEL-GNSS), XGBoost, Random Forest, and their soft-vote and stacked ensembles. Each method shows in-domain (Beihang/Hangzhou) and cross-city (unseen Tokyo Shinjuku) bars per the legend. Single models lose the most accuracy under the city shift, while the soft-vote and stacked ensembles retain the highest cross-city Macro-F_1 (0.892 and 0.886); the soft-vote ensemble is the recommended deployment configuration. Per-class DEGRADED F_1 is reported in Table 5.

Table 5 Zero-shot cross-city evaluation on Tokyo Shinjuku UrbanNav (31,236 windows; 29,200 CLEAN · 1,620 WARNING · 416 DEGRADED). In-domain reference metrics are the ensemble-experiment models (200-tree RF, 300-tree XGBoost on the no-SMOTE windows); the RF in-domain value therefore differs from the baseline RF in Table 2. DEG-F_1: DEGRADED-class F_1. Gap: Tokyo minus in-domain Macro-F_1.

Method	In-domain Macro-F_1	Tokyo Macro-F_1	Gap	Tokyo DEG-F_1
RF	0.926	0.618	−0.308	0.148
XGBoost	0.919	0.821	−0.098	0.784
SENTINEL-GNSS (DL)	0.821	0.649	−0.172	0.753
DL + XGBoost ensemble	n/a	**0.892**	n/a	**0.896**

adaptive Extended Kalman Filter (EKF) evaluated on the UrbanNav Tokyo Shinjuku route: an independent test set the classifier never saw during training.

The EKF fuses GNSS single-point positioning (SPP) with IMU-derived motion estimates in a 9-state filter (position, velocity, heading, clock bias, accelerometer biases). At each 1-second GNSS update, the measurement noise covariance is adapted

as

$$R_t = r_{\text{base}}^2 + \left(r_{\text{deg}}^2 - r_{\text{base}}^2\right) \hat{P}_{\text{DEG},t+5\text{s}}, \tag{8}$$

where $\hat{P}_{\text{DEG},t+5\text{s}}$ is SENTINEL-GNSS's predicted probability of DEGRADED quality at the 5 s horizon. When SENTINEL-GNSS forecasts an impending degraded epoch, R_t increases, the Kalman gain decreases, and the filter trusts the IMU over the corrupted GNSS measurement before the degradation arrives. A floor calibration $\hat{P}_{\text{calib}} = \text{clip}\big((\hat{P}_{\text{raw}} - P_5)/(1 - P_5),\, 0,\, 1\big)$ removes a systematic output floor caused by the receiver-type shift from training to deployment, requiring no ground-truth labels at the deployment site.

Table 6 summarises degraded-segment RMSE for six methods on the 2,450 degraded epochs of the Tokyo route; Figure 8 visualises the same six methods with their 95% confidence intervals. The adaptive SENTINEL-GNSS-EKF achieves a 38.7% reduction in degraded-segment RMSE relative to raw GNSS (47.4 m $\rightarrow$ 29.0 m), outperforming both the Huber robust EKF (36.5%) and the Student-t particle filter (33.2%). Maximum error is reduced from 888 m (raw GNSS) to 163 m (SENTINEL-EKF), and CEP95 from 76 m to 54 m. Adding an online σ_{deg} estimator that removes the need for any hand-tuned r_{deg} parameter improves the result further to 43.2% (RMSE 26.9 m), matching the nsat-proxy adaptive EKF (43.6%) without requiring satellite-count rules. This experiment demonstrates the classifier's downstream navigation value on a single route; a comprehensive multi-city, IMU-coupled navigation and integrity study is left to future work. Full per-method metrics for the present experiment, including CEP50 and 5/10 m availability, are provided in the public repository (Data and code availability).

Classical RAIM is unavailable in these segments. Receiver Autonomous Integrity Monitoring detects and excludes faulty measurements using satellite redundancy, requiring at least five satellites for fault detection and six for fault exclusion (Blanch et al. 2012). This redundancy is absent in the degraded segments: across

Table 6 Degraded-segment positioning RMSE on UrbanNav Tokyo Shinjuku (2,450 degraded epochs). Bootstrap 95% CIs are given in brackets. Fixed-R EKF is the strongest standalone baseline on this clean dual-frequency route; SENTINEL-EKF outperforms all robust baselines (Huber, Student-t PF).

Method	RMSE (m)	95% CI	CEP95 (m)	Max (m)
GNSS raw	47.4	[35.8, 58.8]	76.3	888
Fixed-R EKF	24.3	[22.9, 25.6]	51.0	137
nsat-proxy EKF	26.8	[25.0, 28.4]	52.0	150
Huber EKF	30.1	[28.1, 32.1]	59.3	170
Student-t PF	31.7	[30.4, 33.1]	65.5	122
SENTINEL-EKF (calib.)	**29.0**	[27.3, 30.8]	53.7	163

Fixed-R achieves the lowest RMSE on this clean dual-frequency route because no adaptive down-weighting is needed when signal quality is consistently high. SENTINEL-EKF's value is cross-domain transfer and proactive lead time, not peak performance on optimally-received data.

the 2,450 degraded epochs the median satellite count is five, only 74.8% of epochs reach the five-satellite detection threshold, and none reach the six-satellite exclusion threshold. RAIM-based exclusion is therefore unavailable throughout the degraded segments, precisely where integrity is most needed: a GNSS-only fault-exclusion filter has no redundant measurements to fall back on and must coast through the degradation, so its position error grows without bound. The IMU-coupled, prediction-informed SENTINEL-GNSS-EKF imposes no redundancy requirement and remains bounded (Table 6), which is the central argument for a learned proactive filter over snapshot integrity monitoring.

7 Discussion

7.1 In-domain: why trees lead on aggregate metrics

Trees achieve higher aggregate Macro-F_1 in-domain (0.919 vs 0.821) for three complementary reasons. **(A) Feature concentration.** The discriminative signal in this dataset is concentrated in a small number of highly predictive instantaneous features (mean C/N_0, satellite count, HDOP); shallow decision boundaries suffice when these features are informative. Grinsztajn et al. (2022) show that tree models excel

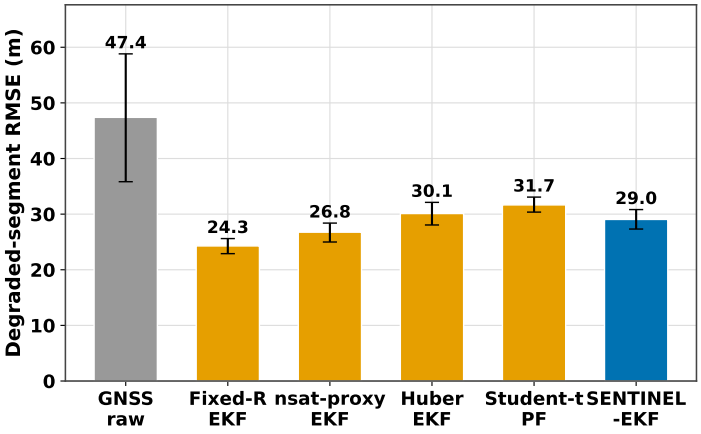

Fig. 8 Degraded-segment positioning RMSE on UrbanNav Tokyo Shinjuku (2,450 degraded epochs) with 95% bootstrap confidence intervals, for the six methods of Table 6. Fixed-R EKF is the strongest baseline on this clean dual-frequency route; the prediction-informed, calibrated SENTINEL-GNSS-EKF (no labels at the deployment site) outperforms both robust baselines, the Huber EKF and the Student-t particle filter.

on tabular data for precisely this reason: they construct exact axis-aligned splits on decisive features without the smoothness bias of neural networks. **(B) Focal-loss recall–precision trade-off (intentional design).** SENTINEL-GNSS uses class weights $\alpha = [1.0, 2.0, 5.0]$ and focal exponent $\gamma = 1.0$, configured to maximise DEGRADED recall at the cost of aggregate F_1. This is a deliberate safety-first design choice: the resulting model achieves DEGRADED recall 0.847, which is 12.5 pp above XGBoost's 0.722, at the cost of a 9.8 pp reduction in Macro-F_1 (0.821 vs 0.919). Trees optimise aggregate accuracy and are not configured to prioritise the minority class beyond balanced-weight adjustments; the lower aggregate F_1 for SENTINEL-GNSS is therefore not evidence of inferior capacity but of a different and explicitly chosen optimisation objective. A controlled ablation isolates this mechanism. Under identical argmax decoding (before the per-class threshold tuning used for the headline metrics), retraining SENTINEL-GNSS with plain cross-entropy ($\gamma = 0$, uniform class weights) raises Macro-F_1 from 0.822 to 0.827 but reduces DEGRADED recall from 0.837 to 0.813 and DEGRADED F_1 from 0.717 to 0.667, confirming the trade-off

originates in the loss configuration and not in model capacity. **(C) Temporal ordering.** A permutation shuffle test (randomising the 30 timesteps within each window) degrades both SENTINEL-GNSS and RF by ≈ 3.2 percentage points, indicating that temporal ordering is a minor contributor for both models; the decisive features are contemporaneous correlations, not long-range causal dynamics.

7.2 Cross-city transfer: the SENTINEL advantage

The decisive advantage of SENTINEL-GNSS over RF is cross-city DEGRADED transfer. RF memorises region-specific feature distributions during training, particularly the Kowloon satellite geometry and C/N_0 baselines; it cannot generalise to Tokyo Shinjuku's different sky-view geometry. The Transformer's multi-head attention over a 30-second history captures signal dynamics that are more environment-agnostic; the *pattern* of C/N_0 decline and HDOP increase is similar across urban canyons even when the absolute values differ. XGBoost also transfers well because its regularisation ($\ell_1 + \ell_2$ penalty) limits feature-space memorisation. The DL+XGBoost soft-vote ensemble benefits from model diversity: the two components make uncorrelated errors on the unseen city, and probability averaging exploits this diversity to reach DEGRADED $F_1 = 0.896$.

Receiver-type confound. The Tokyo evaluation involves not only a city change but also a receiver change: the training corpus uses u-blox ZED-F9P and NovAtel Flexpak6 units; the Tokyo route uses a Trimble dual-frequency receiver. Receiver-type differences produce systematic offsets in absolute C/N_0 baselines and fix-quality flag encoding. The cross-city generalisation gap therefore reflects both geographic variation and receiver-type variation, and the two contributions cannot be separated without a matched-receiver cross-city dataset. Two points mitigate this: (A) the training corpus already spans two receiver types (ZED-F9P and Flexpak6), demonstrating that SENTINEL-GNSS is not locked to a single hardware profile; and (B)

the floor-calibration procedure deployed at the EKF stage (see §6.5) removes systematic probability offsets caused by receiver-type mismatch without requiring any ground-truth labels at the deployment site.

7.3 Deployment considerations

At 0.045 ms per 30-window on a Tesla T4 GPU, SENTINEL-GNSS is 9.5× faster than three separate tree models (0.425 ms combined) and delivers all three horizon outputs in one forward pass. The unified 17.8 MB model footprint is suitable for embedded navigation hardware. Temperature scaling reduces the expected calibration error (ECE) from 0.114 to 0.069 (a 40% improvement). A per-class matrix (Dirichlet-style) calibration fitted on the validation logits does not improve on this; it overfits the limited validation set, giving a higher ECE of 0.087. The model therefore remains imperfectly calibrated by the ECE < 0.05 threshold.

The operational consequence of ECE $= 0.069$ for navigation is limited: the adaptive EKF requires that $\hat{P}(\text{DEGRADED})$ increase monotonically as conditions worsen, not that the absolute probabilities match empirical frequencies exactly. Temperature scaling achieves this monotonicity criterion; the residual ECE manifests as an approximately 7% overconfidence in predicted probabilities, which translates to a small scale offset in the EKF measurement noise inflation factor. This offset is absorbed by the per-deployment tuning of r_{deg} (or removed entirely by the online σ_{deg} estimator described in §6.5). Calibration with a larger held-out set, which the per-class methods appear to require here, is identified as future work.

8 Conclusions

SENTINEL-GNSS is a Transformer-LSTM network that classifies GNSS signal quality simultaneously at 5 s, 15 s, and 30 s from 30-second sliding windows of 37 receiver-derived features. Four conclusions follow from the experiments.

First, SENTINEL-GNSS achieves DEGRADED-class recall of 0.847 at 5 s (F_1 = 0.718, 95% CI [0.671, 0.762]); this is 12 percentage points above both tree baselines on the safety-critical metric.

Second, tree baselines outperform SENTINEL-GNSS on aggregate Macro-F_1 in-domain (XGBoost: 0.919 vs SENTINEL-GNSS: 0.821), and the persistence baseline scores 0.908 from autocorrelation alone. These two comparisons establish that SENTINEL-GNSS's value lies not in aggregate metrics but in DEGRADED recall and cross-city transfer.

Third, in zero-shot evaluation on Tokyo Shinjuku, SENTINEL-GNSS retains DEGRADED $F_1 = 0.753$ while Random Forest collapses to 0.148; the DL+XGBoost soft-vote ensemble reaches Macro-F_1 = 0.892 and DEGRADED F_1 = 0.896, outperforming every individual model.

Fourth, SENTINEL-GNSS runs at 0.045 ms per window, 9.5× faster than three separate tree models, from a 17.8 MB unified model.

Fifth, SENTINEL-GNSS's predictions reduce degraded-segment positioning RMSE by 38.7% relative to raw GNSS on UrbanNav Tokyo Shinjuku (47.4 m → 29.0 m), outperforming both the Huber robust EKF and the Student-t particle filter without any labels at the deployment site.

Three limitations remain: (A) lower in-domain aggregate Macro-F_1 relative to tree methods, which is explained by the focal-loss recall priority and is not a capacity limitation; (B) imperfect calibration (ECE = 0.069), with limited navigation consequence given that the adaptive EKF requires monotonicity rather than exact calibration; and (C) the combined geographic and receiver-type generalisation gap to Tokyo, which is partially addressed by the unsupervised floor calibration mechanism. Future work will address calibration with a larger held-out set, extended prediction horizons (45 s–60 s), and IMU-coupled EKF evaluation across additional cities.

Acknowledgements. This work was supported by the Regional Centre for Space Science and Technology Education in Asia and the Pacific (RCSSTEAP), affiliated with the United Nations Programme on Space Applications, hosted at Beihang University.

Declarations

Conflict of interest. The authors declare no conflict of interest.

Data and code availability. The UrbanNav dataset (Hong Kong and Tokyo routes) is publicly available at https://github.com/IPNL-POLYU/UrbanNavDataset (Hsu et al. 2023). The Beihang University field-collection data (Hangzhou scenarios A–E), pre-computed feature windows, trained model weights, and the evaluation code used to produce all reported results are available at https://github.com/Jorshuare/AI-Based-Prediction-for-GNSS-Signal-Degradation and from the corresponding author upon reasonable request. All experiments are fully reproducible from the archived checkpoint and the `evaluate.py` script in the repository.

Authors' contributions. J.S.O. and J.O.O. jointly conceived the research direction and formulated the problem statement. J.S.O. led the model architecture design and implementation. M.I.I. and J.O.O. designed and carried out the feature engineering and dataset preprocessing pipeline. J.S.O., M.I.I., and J.O.O. jointly conducted the experiments and validated the results. C.I. and M.H.O.R. contributed to the cross-city evaluation, literature review, and manuscript revision. J.S.O. and J.O.O. drafted the manuscript; all authors reviewed and approved the final text. W.A.A. provided supervision, and contributed critical revisions throughout the project.

References

Blanch J, Walter T, Enge P (2012) Advanced RAIM user algorithm description: Integrity support message processing, fault detection, exclusion, and protection level calculation. In: Proceedings of the 25th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+), pp 2828–2849

Breiman L (2001) Random forests. Machine Learning 45(1):5–32. https://doi.org/10.1023/A:1010933404324

Chen T, Guestrin C (2016) XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, pp 785–794, https://doi.org/10.1145/2939672.2939785

Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics 21(1):6. https://doi.org/10.1186/s12864-019-6413-7

Cohen J (1960) A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20(1):37–46. https://doi.org/10.1177/001316446002000104

Grinsztajn L, Oyallon E, Varoquaux G (2022) Why tree-based models still outperform deep learning on tabular data. In: Advances in Neural Information Processing Systems, pp 507–520

Groves PD (2013) Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems, 2nd edn. Artech House, Norwood, MA

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Computation 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hsu LT, Ng HF, Zhang G, et al (2023) UrbanNav: An open-sourced multisensory dataset for benchmarking positioning algorithms designed for urban areas. NAVIGATION: Journal of the Institute of Navigation 70(2). https://doi.org/10.33012/navi.557

Isik OU, Gao GX (2024) Machine learning-based GNSS signal quality prediction for integrity monitoring in urban air mobility. Drones 8(3):85. https://doi.org/10.3390/drones8030085

Kingma DP, Ba J (2015) Adam: A method for stochastic optimization. In: Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA

Li T, Zhang H, Gao Z, et al (2023) Accuracy and reliability of multi-GNSS real-time precise positioning with poor satellite geometry. Satellite Navigation 4(1):1–14. https://doi.org/10.1186/s43020-023-00106-7

Lin TY, Goyal P, Girshick R, et al (2017) Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp 2980–2988, https://doi.org/10.1109/ICCV.2017.324

Ozeki T, Kubo N (2022) GNSS NLOS signal classification based on machine learning and pseudorange residual check. Frontiers in Robotics and AI 9:868608. https://doi.org/10.3389/frobt.2022.868608

Park B, Kim M, Seo J (2024) Multipath detection using pseudorange acceleration for GNSS integrity monitoring in urban environments. Sensors 24(7):2289. https://doi.org/10.3390/s24072289

Siemuri A, Selvan K, Kuusniemi H, et al (2022) A systematic review of machine learning techniques for GNSS use cases. IEEE Transactions on Aerospace and Electronic

Systems 58(6):5043–5077. https://doi.org/10.1109/TAES.2022.3219366

Tao J, Liu J, Hu Z, et al (2021) INS/radar altimeter integrated navigation algorithm based on CNN-LSTM. Mathematical Problems in Engineering 2021:1–13. https://doi.org/10.1155/2021/1350988

Vaswani A, Shazeer N, Parmar N, et al (2017) Attention is all you need. In: Advances in Neural Information Processing Systems, vol 30. Curran Associates

Wen W, Hsu LT (2022) 3d LiDAR aided GNSS NLOS mitigation in urban canyons. IEEE Transactions on Intelligent Transportation Systems 23(10):18224–18236. https://doi.org/10.1109/TITS.2022.3169632

Xu B, Ng HF, Zhuang Y, et al (2019) Intelligent smartphone-based multimode GNSS positioning for urban environments. Remote Sensing 11(23):2842. https://doi.org/10.3390/rs11232842

Xu L, Yao Y, Xu G (2024) Machine learning in GNSS data processing: A review. IEEE Aerospace and Electronic Systems Magazine 39(3):4–20. https://doi.org/10.1109/MAES.2024.3351214

Zhang H, Wang Z, Vallery H (2023) Learning-based NLOS detection and uncertainty prediction of GNSS observations with transformer-enhanced LSTM network. In: Proceedings of the 26th IEEE International Conference on Intelligent Transportation Systems (ITSC), pp 910–917, https://doi.org/10.1109/ITSC57777.2023.10246083

Zhang H, Luo W, Peng M, et al (2026) Cascaded transformer-LSTM architecture for urban canyon navigation with factor graph optimization. GPS Solutions 30(55). https://doi.org/10.1007/s10291-025-02010-8

Zhang Y, Kubo N, Chen J, et al (2021) Extending the use of DNN for degraded and disrupted GNSS signal environments using embedded temporal dependency. Sensors 21(20):6802. https://doi.org/10.3390/s21206802

Zheng F, Chen G, Guo J, et al (2024) A stacking ensemble learning framework for GNSS NLOS estimation in urban canyons. GPS Solutions 28(2):96. https://doi.org/10.1007/s10291-024-01638-8