Lightweight LoRa Signal Quality Classification Using RSSI and SNR

Boubaker Abdallah[1,3,*], Rym Chéour[1,2], Jalel Ktari[1] and Olfa Kanoun[3]

[1]*CES Lab, National Engineering School of Sfax (ENIS), University of Sfax, Tunisia*
[2]*Higher Institute of Applied Science and Technology of Kasserine, University of Kairouan, Tunisia*
[3]*Measurement and Sensor Technology, Technische Universität Chemnitz, Reichenhainer Straße 70, 09126 Chemnitz, Germany*

Abstract

LoRa/LoRaWAN, based on Chirp Spread Spectrum, enables long-range and low-power connectivity but exhibits sensitivity to fading, interference, and deployment heterogeneity. This paper addresses supervised classification of LoRa signal quality using non-PHY measurements drawn from a public IEEE Dataport dataset. We operationalize a four-level quality taxonomy (fear, poor, good, excellent) and propose a reproducible pipeline that includes outlier curation, normalization, class-imbalance mitigation by class weighting, and leakage-safe preprocessing within model pipelines. We benchmark representative classifiers, namely Logistic Regression, support vector machines with radial basis kernel, Random Forest, and a compact multilayer perceptron, under stratified five-fold cross-validation. Evaluation relies on macro-F1 as the primary endpoint, complemented by per-class precision, recall, and F1 as well as normalized confusion matrices. Error analyses highlight that residual confusions occur mainly between adjacent regimes, consistent with overlapping ReceSSI and SNR bands. The results indicate that lightweight models trained on RSSI and SNR can deliver robust, deployment-ready quality classification without access to physical-layer waveforms, supporting quality-of-service monitoring and adaptive configuration at LoRa gateways. The study provides a transparent and reproducible baseline for future work on dataset shift, probability calibration, and semi-supervised adaptation on IEEE Dataport resources.

Keywords

LoRa, LoRaWAN, IEEE Dataport, Signal quality classification, RSSI, SNR, Stratified cross-validation, Macro-F1, Random Forest

1. Introduction

Wireless networks are driving a rapidly expanding IoT ecosystem that covers transportation, agriculture, metering, security, healthcare, and smart cities, up to hundreds of billions of connected devices by 2030. To support such scale and heterogeneity, LPWANs offer long-range connectivity with low energy consumption [1]. For effective localization and communication, signal-based ranging remains central: trilateration/triangulation exploits features such as RSSI, SNR, path loss, angle of arrival, and propagation time, while Time of arrival (ToA)/Time Difference of Arrival (TDOA) avoid fingerprinting but demand tight synchronization and are sensitive to Line of Sight (LoS) and bandwidth constraints [2]. Within this landscape, LoRa/LoRaWAN built on chirp spread spectrum (CSS), has emerged as a leading LPWAN thanks to its long range, interference robustness, and ultra–low power operation [3, 4]. Within this context, LoRa signal classification which includes transmitter identification (RF fingerprinting), inference of physical layer parameters (e.g. spreading factor, SF), recognition of the quality state and detection of interference has become essential for network safety, radio orchestration, and performance optimization. Recent approaches used both time-frequency descriptors and deep learning to extract, from raw IQ streams or spectral representations, discriminative signatures in nonstationary environments and low-SNR regimes [5, 6]. These studies highlight the dual role of hardware impairments (CFO, IQ imbalance, PA nonlinearities). They offer a specific feature that can be exploited for classification, but

Cooperative Information Systems – Early Research Achievement and Demos, 2025
*Corresponding author.
boubaker.abdallah@enis.tn (B. Abdallah); rym.cheour@enis.tn (R. Chéour); Olfa.Kanoun@etit.tu-chemnitz.de (J. Ktari); jalel.ktari@enis.tn (O. Kanoun)
0009-0006-9006-472X (B. Abdallah); 0000-0002-9970-3915 (R. Chéour); 0000-0002-0520-9415 (J. Ktari); 0000-0002-7166-1266 (O. Kanoun)

they also induce biases that compromise the generalization of learned models [3].
This methodology involves systematically comparing linear (logistic regression), nonlinear (SVM-RBF), and ensemble-based classifiers (random forest) on the LoRa dataset. This comparison provides a better understanding of the relative strengths of these approaches for signal quality classification. By relying on macro-F1 rather than raw accuracy, the methodology ensures fair evaluation under class imbalance. The resulting pipeline constitutes a reproducible framework for LoRa signal classification and contributes to advancing adaptive strategies in IoT networks.
The remainder of this paper is organized as follows. Section 2 surveys related work on RSSI/SNR-driven LoRa signal classification and QoS assessment. Section 3 details the dataset from IEEE Dataport, preprocessing and label construction, the baseline and proposed models, and the stratified five-fold cross-validation protocol. Section 4 presents the classification framework and training setup. Section 4 reports the empirical findings, including cross-validated macro-F_1, per-class precision/recall/F_1, normalized confusion matrices, and signal characterization figures. Section 6 concludes and discusses limitations and future directions, with emphasis on class-imbalance mitigation, probability calibration, and robustness to distribution shift.

2. Related Work

Recent studies increasingly exploit RSSI and SNR as low-cost yet discriminative descriptors for LoRa/LoRaWAN classification across several tasks, including device fingerprinting, spreading-factor prediction, environmental sensing, link reliability estimation, and localization. For device identification, a gated graph neural network that fuses RSSI with phase and temporal encoding improves cross-environment robustness over flat feature vectors [7]. For SF prediction, systematic study on public data shows that the joint use of RSSI and SNR forms the most effective feature subset across classical models [8]. Multi-gateway RSSI enables city-scale traffic sensing and campus-scale area classification with standard classifiers [9, 10], while temporal aggregation of RSSI sequences enhances packet success prediction in long-term deployments [11]. Semi-supervised transfer learning that combines RSSI, SNR, and timestamps further improves outdoor localization under label scarcity [12]. Heuristic methods, such as floor-level estimation and RSSI-based SF assignment, demonstrate the scale of the problem, but are not generalizable and lack the robustness of a supervised learning framework [13, 14]. The literature suggests that RSSI/SNR individually, jointly, or temporally aggregated support reliable classification, though open challenges remain in feature normalization across SFs, inter-gateway calibration, and standardized evaluation protocols. These gaps motivate further research into reproducible, generalizable classification pipelines for LoRa/LoRaWAN systems.

3. Methodology

3.1. Dataset and Ingestion

Four indoor LoRa measurement files were consolidated into a single corpus: *Lab 10 (LoRa Receiver 2)*, *Lab 6 (LoRa Receiver 3)*, *Lab 7 (LoRa Receiver 1)*, and *Lab 8 (LoRa Receiver 4)*. The merged dataset comprises per-packet `RSSI` and `SNR` measurements collected across multiple rooms, floors, and receiver locations [15]. After cleaning and concatenation, a total of **5385 packets** were retained.

3.2. Preprocessing and Feature Engineering

Raw `RSSI` and `SNR` values were coerced to numeric (`NaN` on parse failure), and missing values were imputed column-wise using the mean. In addition to the raw features, three lightweight descriptors were engineered to capture basic signal geometry and robustness:

$$\texttt{rssi\_abs} = |\texttt{rssi}|, \tag{1}$$

$$\texttt{snr\_rssi\_ratio} = \frac{\texttt{snr}}{|\texttt{rssi}| + \epsilon}, \tag{2}$$

$$\texttt{signal\_quality} = \texttt{snr} - |\texttt{rssi}|, \tag{3}$$

with a small ϵ added to avoid division by zero. These features are intentionally simple to remain compatible with resource-constrained gateways. Before model training, features were standardized to zero mean and unit variance using statistics from the training split only.

3.3. Labeling Strategy

To enable supervised learning of link quality, continuous `RSSI` and `SNR` were discretized into four ordinal classes: {*fear*, *poor*, *good*, *excellent*}. The mapping follows a 2-D rule derived from link-budget considerations and BER/PER analysis:

$$\text{label} = \begin{cases} \text{excellent}, & \text{if SNR} \geq 10 \text{ and RSSI} \geq -80, \\ \text{good}, & \text{if SNR} \geq 6 \text{ and RSSI} \geq -90, \\ \text{poor}, & \text{if SNR} \geq 2 \text{ and RSSI} \geq -100, \\ \text{fear}, & \text{otherwise.} \end{cases}$$

After applying this mapping, the class distribution was strongly imbalanced: **fear:** 39 samples (0.7%), **poor:** 1774 samples (33.0%), **good:** 1193 samples (22.2%), and**excellent:** 2379 samples (44.1%)

This imbalance motivated the use of class weights in both ML and DL training (Section 4). Labels were encoded in the fixed order $\{\text{fear}=0,\ \text{poor}=1,\ \text{good}=2,\ \text{excellent}=3\}$ using a `LabelEncoder`, ensuring consistent mapping across all models.

3.4. Experimental Protocol

We evaluate under stratified five-fold cross-validation (`StratifiedKFold`, $n=5$, shuffle, random_state $=42$) to preserve empirical class proportions in each fold. The primary metric is macro-F1 to emphasize minority classes; we also report accuracy and per-class F1. Where a single held-out test split is required (for DL), the training set is further split (80/20) for validation, keeping stratification.

3.5. Baseline Classification Models (Machine Learning)

Three supervised baselines were implemented with `scikit-learn`:

- **Logistic Regression (LogReg):** linear decision boundary, `max_iter`=400, `class_weight=balanced`.
- **Support Vector Machine, RBF kernel (SVM-RBF):** non-linear classifier with `probability`=True and `class_weight=balanced`.
- **Random Forest (RF):** `n_estimators`=300, fixed `random_state` for reproducibility.

Macro-F1 was computed by cross-validation via `cross_val_score` on the standardized feature set [`rssi`, `snr`, `rssi_abs`, `snr_rssi_ratio`, `signal_quality`]. These baselines, trained on the identical labels defined above, serve as a coherent reference for our deep model.

4. Proposed Classification Method

4.1. Lightweight Deep Learning Enhancement (MLP)

To improve robustness on minority classes while keeping inference cost low, we adopt a compact Multi-Layer Perceptron (MLP) implemented in TensorFlow/Keras: Dense(64, ReLU) → BatchNorm → Dropout(0.3), Dense(32, ReLU) → BatchNorm → Dropout(0.2), and Dense(4, Softmax).

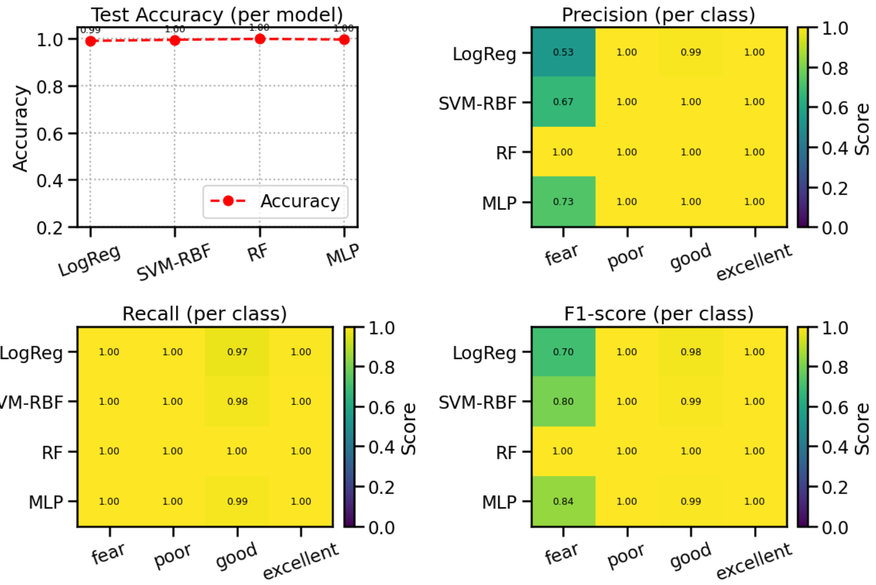

Figure 1: Model performance metrics of SVM-RBF, Random Forest, Logistic Regression classifiers and MLP

The model is trained with Adam and `sparse_categorical_crossentropy`. To counter class imbalance we use class weights (`compute_class_weight`) applied at training time. Regularization and generalization are promoted via early stopping (patience $= 10$, restore-best-weights) and model checkpointing on validation accuracy. Input dimensionality matches the standardized feature vector used by the baselines.

4.2. Unified Evaluation

The MLP is trained on the same training folds, validated on a stratified split of the training data, and evaluated on the held-out test portion; we report accuracy, precision, recall, and F1 (macro/weighted and per-class), together with a normalized confusion matrix. For direct benchmarking, the baseline models are evaluated under the identical preprocessing, feature set, label mapping, and splits. This design enforces a *baseline* $\rightarrow$ *proposed* progression rather than parallel pipelines and allows fair, apples-to-apples comparisons.

5. Experiments and Results

The four-panel visualization provides a comprehensive overview of the performance of three classical machine learning models, Logistic Regression (LogReg), SVM with RBF kernel (SVM-RBF), Random Forest (RF) together with the proposed deep learning model (MLP) in the classification of LoRa signal quality classes (*fear*, *poor*, *good*, *excellent*). Each subplot highlights complementary aspects of predictive ability, offering insights into both global model accuracy and class-wise discriminative capacity.

5.1. Quantitative Evaluation of Classifier Metrics

The comparative results in Figure 1 demonstrate that Random Forest, SVM-RBF, and the proposed MLP all achieve perfect accuracy (1.00), while Logistic Regression slightly underperforms at 0.99. Precision and F1-scores reveal that the minority *fear* class remains the most challenging: Logistic Regression records the weakest F1 (0.70), SVM-RBF improves to 0.80, and the MLP further enhances performance to 0.84, whereas Random Forest sustains flawless scores across all categories. These results confirm Random Forest as the most robust baseline, but also highlight that the proposed MLP provides a lightweight deep learning alternative that significantly improves the detection of weak-signal instances without compromising overall accuracy.

5.2. Characterization of LoRa Signals

The temporal behavior of RSSI, SNR, and simulated CSS chirps reveals characteristic channel dynamics: RSSI spans roughly -120 to $-70\,\mathrm{dBm}$ with pronounced fluctuations (samples 40–90) consistent with fading and time variation; SNR is mostly $> 8\,\mathrm{dB}$ with brief dips below $0\,\mathrm{dB}$ indicating transient interference; and the chirp trace reflects amplitude and instantaneous-frequency variability induced by these conditions as shown in figure 2.

Figure 2: Temporal evolution of RSSI, SNR, and simulated CSS chirps.

Reliability metrics exhibit clear dependencies on channel indicators: BER decreases rapidly with increasing SNR, with a knee near $\approx 7\,\mathrm{dB}$; higher RSSI aligns with lower BER under AWGN assumptions; and PER shows a steep transition around a 20–$30\,\mathrm{dB}$ link margin, marking an operational threshold. The link-margin distribution concentrates within 30–$50\,\mathrm{dB}$, while high BER is chiefly confined to the joint low-RSSI/low-SNR regime. These patterns indicate that hybrid indicators (e.g., effective SNR or margin) are more informative than single-channel metrics as shown in figure 3.

The class distribution is highly imbalanced, with the *excellent* category exceeding 95% of samples. SNR box plots follow the expected ordering *excellent* shows the highest median and tightest interquartile range, whereas *poor* exhibits a wider spread while a modest overlap between *good* and *fear* accounts for the residual confusion observed in evaluation. These patterns motivate complementing raw SNR with additional descriptors (e.g., temporal statistics and link-margin dynamics) to improve separability as shown in figure 4.

6. Conclusion

This work presented a leakage-safe, reproducible pipeline for supervised LoRa signal classification using non-PHY measurements (RSSI and SNR) obtained from an IEEE Dataport dataset. We formalized a four-level quality taxonomy and benchmarked representative classifiers under stratified five-fold cross-validation, reporting macro-F1 together with per-class precision, recall, and normalized confusion matrices. The results indicate that lightweight models can reliably separate quality regimes using simple,

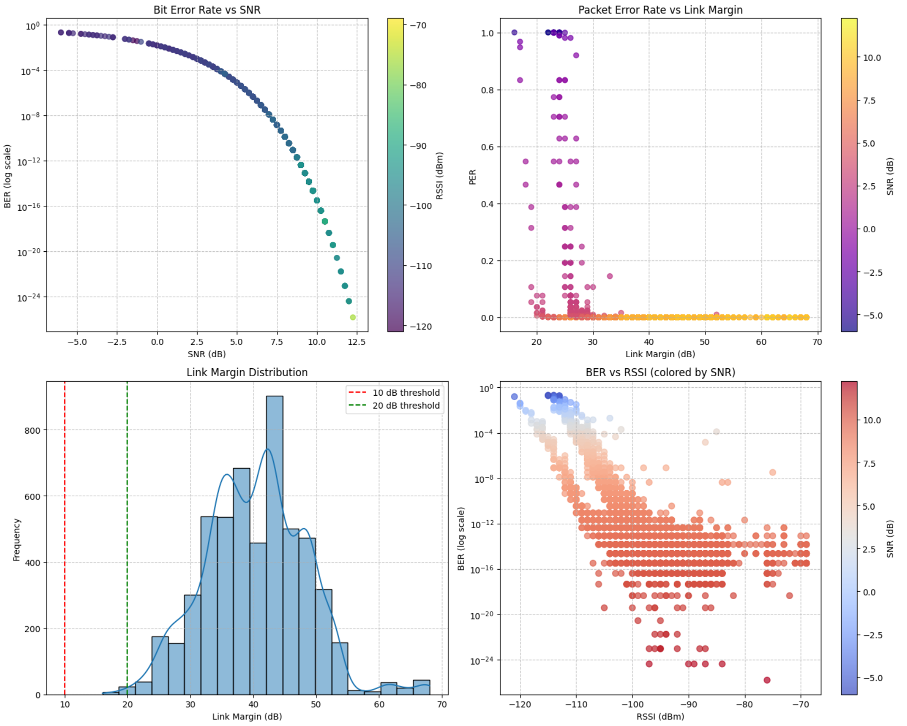

Figure 3: BER vs. SNR, PER vs. link margin, margin distribution, and BER vs. RSSI (SNR-colored).

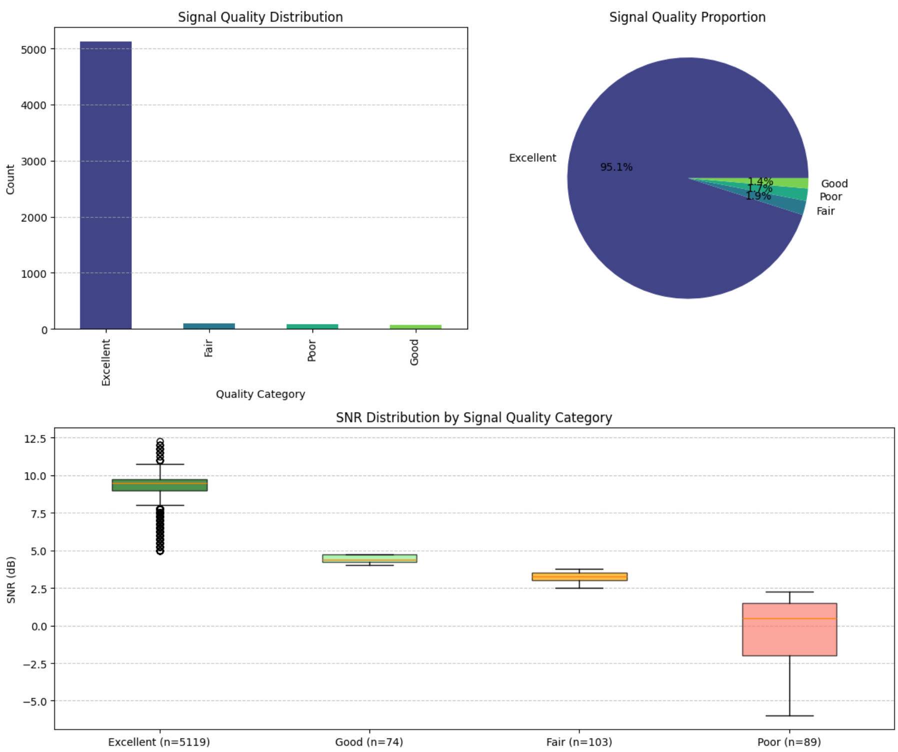

Figure 4: Quality-class distribution and per-class SNR statistics.

physically meaningful descriptors, supporting deployment at LoRa gateways for quality-of-service monitoring and adaptive configuration.

A central limitation of the present study is the markedly imbalanced class distribution in the public dataset, with a dominant majority class and comparatively scarce minority classes. Although we employed class weighting and macro-F1 to mitigate and properly quantify imbalance effects, residual bias may persist in decision thresholds, probability calibration, and generalization under distributional shift. In future work, we will curate balanced training splits and collect or subsample additional minority-class instances, investigate cost-sensitive learning and focal losses alongside calibrated decision thresholds, explore physics-consistent augmentation (e.g., small RSSI/SNR perturbations) and semi-supervised enrichment, and adopt group and time-aware validation to stress-test robustness. These steps aim to deliver more equitable class performance and improve external validity across deployments and operating conditions.

Declaration on Generative AI

During the preparation of this manuscript, the authors utilized ChatGPT to assist in grammar and spelling correction, as well as paraphrasing of the text. Following the use of this tool, the authors carefully reviewed and revised the content to ensure its accuracy, coherence, and integrity, and assume full responsibility for the final version of the manuscript.

References

[1] P. L. Ngwenyama, R. C. Webber-Youngman, Recent advances, challenges and future trends for the applications of low power wide area networks (lpwans) technologies in underground mines, International Journal of Mining, Reclamation and Environment (2025) 1–53.

[2] S. M. Asaad, H. S. Maghdid, A comprehensive review of indoor/outdoor localization solutions in iot era: research challenges and future perspectives, Computer Networks 212 (2022) 109041. doi:10.1016/j.comnet.2022.109041.

[3] Z. Xu, S. Tong, P. Xie, J. Wang, From demodulation to decoding: Toward complete lora PHY understanding and implementation, ACM Transactions on Sensor Networks 18 (2023) 1–27. doi:10.1145/3546869.

[4] B. Abdallah, S. Khriji, R. Chéour, C. Lahoud, K. Moessner, O. Kanoun, Improving the reliability of long-range communication against interference for non-line-of-sight conditions in industrial internet of things applications, Applied Sciences 14 (2024) 868.

[5] G. Shen, J. Zhang, A. Marshall, L. Peng, X. Wang, Radio frequency fingerprint identification for lora using deep learning, IEEE Journal on Selected Areas in Communications 39 (2021) 2604–2616. doi:10.1109/JSAC.2021.3087250.

[6] A. Al-Shawabka, P. Pietraski, S. B. Pattar, F. Restuccia, T. Melodia, Deeplora: Fingerprinting lora devices at scale through deep learning and data augmentation, in: Proceedings of the 22nd ACM International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing (MobiHoc '21), Association for Computing Machinery, Shanghai, China, 2021, pp. 251–260. doi:10.1145/3466772.3467054.

[7] B. Zhang, Gnn for lora device fingerprint identification, IEEE Open Journal of the Communications Society (2025). doi:10.1109/OJCOMS.2024.3492922.

[8] A. Prakash, A. Gorrela, Effective feature selection for predicting spreading factor with ml in large lorawan-based mobile iot networks, in: Proceedings of the National Conference on Communications (NCC), 2025. doi:10.1109/NCC63735.2025.10983488.

[9] H. B. Pasandi, S. I. Lopes, Low-cost traffic sensing system based on lorawan for urban areas, in: Proceedings of the 1st International Workshop on Emerging Topics in Wireless, 2022. doi:10.1145/3565474.3569069.

[10] C.-C. Wei, C.-C. Chang, The campus localization system based on lora, in: Proceedings of the 5th International Conference on Computer Communication and the Internet (ICCCI), 2023. doi:10.1109/ICCCI59363.2023.10210152.

[11] A. Strzoda, A. Madej, Link failure prediction in lora networks, in: Proceedings of the International

Wireless Communications and Mobile Computing Conference (IWCMC), 2023. doi:10.1109/IWCMC58020.2023.10183185.

[12] Y.-S. Chen, C.-Y. Huang, A semi-supervised transfer learning with grid segmentation for outdoor localization over lorawans, Sensors (2021). doi:10.3390/s21082640.

[13] K. Nakano, N. Ikeda, Experimental evaluation of floor height estimation using unlicensed-band lpwa signals toward 3d nlos indoor positioning, in: IEEE PIMRC, 2022. doi:10.1109/PIMRC54779.2022.9978130.

[14] P. Tempiem, R. Silapunt, Spreading factor allocation using the standard deviation classification method, in: International Symposium on Antennas and Propagation (ISAP), 2020. doi:10.23919/ISAP47053.2021.9391249.

[15] M. A. Kamal, M. M. Alam, A. Abu Bakar Sajak, M. M. Su'ud, Snr and rssi based an optimized machine learning based indoor localization approach: Multistory round building scenario over lora network, Computers Materials and Continua (2024).