A Study on Estimating the Scale of Terrestrial Smartphone Density Using Low Earth Orbit Satellites and Machine Learning

Futo Noda, Gia Khanh Tran
Institute of Science Tokyo, Tokyo, Japan
Email: noda.f.4847@m.isct.ac.jp, tran.k.527f@m.isct.ac.jp

***Abstract*—Japan frequently experiences natural disasters such as earthquakes and torrential rains, and the risk of future disaster occurrences remains high. In such circumstances, the terrestrial communication infrastructure may become inoperable. Estimating the number of mobile terminals within a specific area under these conditions can serve as a proxy for population estimation, thus playing an essential role in supporting emergency response planning. Even under normal conditions, the ability to quickly assess the distribution of people in areas where communication infrastructure is temporarily unavailable offers significant advantages for various applications such as urban planning and traffic management. To address these challenges, the authors have previously proposed a system for estimating the density of smartphones on the ground by analyzing the Received Signal Strength Indicator (RSSI) obtained through sensing from Low Earth Orbit (LEO) satellites. In this paper, we demonstrate through estimation simulations using multiple machine learning algorithms that the proposed system can accurately estimate the scale of terminal density within a target area. The results confirm both the feasibility and the effectiveness of the proposed approach.**

***Index Terms*—LEO, Machine Learning, Smartphone, Wi-Fi, Density Estimation, RSSI**

I. Introduction

A. Background

In recent years, satellite communication has attracted growing attention as a means to enhance communication resilience in the event of terrestrial infrastructure failure during natural disasters, and as a complementary technology to overcome the increasing difficulty of deploying dense base-station networks in the 5G/6G era. Earth-orbiting satellites are classified by altitude into Geostationary Earth Orbit (GEO, 36,000 km), Medium Earth Orbit (MEO, 2,000–36,000 km), and Low Earth Orbit (LEO, up to 2,000 km). GEO and MEO satellites offer wide-area coverage and are widely utilized in systems such as the Global Navigation Satellite System (GNSS) and satellite broadcasting. In contrast, LEO satellites, owing to their lower altitude, provide shorter propagation distances and reduced latency, making them particularly suitable for applications requiring high spatial resolution.

Building on these characteristics, numerous studies and commercial deployments of wireless communication systems using LEO satellites have emerged in recent years. Shayea *et al.* [1] organized the challenges and opportunities associated with the integration of LEO satellites with 5G/6G and IoT systems, emphasizing the societal value of LEO-based networks capable of low-latency and high-resolution observation. Similarly, Castro-Carrera *et al.* [2] conducted a systematic review of LEO satellite services for IoT applications, showing that their usage is expanding beyond communication to include observation and monitoring purposes.

These studies demonstrate that LEO satellites possess strong potential not only as communication relays but also as independent, wide-area observational platforms. Particularly, their ability to operate unaffected by ground-level damage during disasters makes them highly effective for obtaining information—such as population or terminal distribution—without relying on terrestrial infrastructure.

B. Related Works

A variety of studies have investigated the use of LEO satellites for wireless communication and sensing applications. Dwivedi *et al.* [3] analyzed the performance of LEO satellite-based IoT networks under interference and quantitatively evaluated how terminal density and satellite altitude affect communication quality. Zhang *et al.* [4] proposed a resource optimization method for LoRa-based LEO IoT systems and demonstrated the improved efficiency of satellite IoT communications. These studies suggest that LEO satellites can be effectively applied to large-scale sensing and ground terminal observation.

In the terrestrial wireless sensing domain, numerous studies have explored the estimation of crowd size or device count using the RSSI. Janssens *et al.* [5] proposed a non-contact crowd estimation method based on wireless sensor networks (WSN) and demonstrated that the number of individuals can be accurately classified from RSSI distributions. Similarly, Zaidan *et al.* [6] estimated the number of people in indoor environments using Wi-Fi RSSI measurements and achieved reasonable accuracy even with a single receiver. Depatla *et al.* [7] further investigated crowd counting using Wi-Fi signals through walls, showing that spatial variations in received power can be exploited as indicators of human presence.

While these studies demonstrate that spatial variations in signal strength can serve as reliable indicators of population or terminal density, they are all limited to terrestrial or indoor environments. To the best of our knowledge, no prior work has attempted to estimate the number of ground terminals from a macroscopic, top-down perspective using LEO satellite observations. This study aims to bridge this research gap by proposing a novel estimation framework that leverages RSSI data observed from LEO satellites and applies machine learning techniques to estimate terminal density, even under

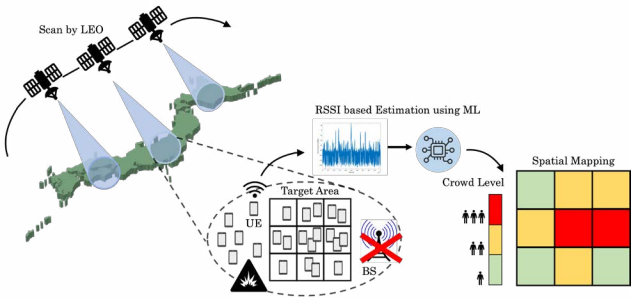

Fig. 1: Concept of the proposed estimation system.

disaster conditions where terrestrial infrastructure is unavailable.

In previous work, the authors proposed a system to estimate the number of smartphones by observing radio signals received by LEO satellites and applying a regression function to RSSI data [8]. Furthermore, we have explored the enhancement of estimation accuracy through the integration of machine learning [9]. The proposed system relies solely on RSSI measurements, offering the following advantages:

- It avoids privacy-related concerns, as it does not require access to personally identifiable information.
- It operates independently of smartphone models or the mobile carriers with which users are subscribed.

In this paper, we propose a method for estimating the scale of mobile terminal density in a target area using machine learning. The effectiveness of the proposed method was evaluated through simulations conducted in MATLAB. Additionally, we performed verification using multiple machine learning algorithms to investigate whether the estimation results vary depending on the algorithm employed and what effects arise from combining multiple algorithms. The results suggest the potential of the proposed system to estimate the scale of UE density and indicate that the introduction of machine learning algorithms is effective for estimation, enabling highly accurate scale estimation. Furthermore, to evaluate whether the proposed system should include smartphones located indoors as part of its target, we also conducted an investigation into the signal attenuation characteristics of radio wave propagation from indoor smartphones to aerial platforms such as Unmanned Aerial Vehicles (UAV).

The remainder of this paper is organized as follows. Section II describes the architecture of the proposed system, the system model used for simulation, and the assumed use cases. Section III presents the details of the MATLAB based simulation conducted to validate the proposed estimation method, including the simulation scenarios, the introduction of a 3D city model, the implementation of an ideal antenna model, parameters and the definition of evaluation metrics. Additionally, this section includes an investigation into the propagation characteristics of radio waves emitted from smartphones located indoors. Section IV presents the results obtained from the simulations and provides a comprehensive discussion. Finally, Section V summarizes the paper and discusses future directions.

II. System Overview and Use Cases

This section presents an overview of the proposed system architecture, the system model used for simulation, and the assumed use cases for the proposed terminal density estimation system.

A. Architecture

Figure 1 illustrates the overview of the proposed terminal density estimation system. In this system, a LEO satellite continuously performs radio scanning while orbiting the Earth. During this process, it observes radio signals emitted by smartphones (hereinafter referred to as user equipment, or UE) located on the ground and collects RSSI data.

When terrestrial communication infrastructure such as base stations is operational, the number of UEs communicating with base stations, along with GNSS and other positional data, are used to associate the observed RSSI values with the actual number of UEs present in a given target area. These data are used as training samples for machine learning models. Once trained, the system can estimate the number of UEs in a target area using only RSSI data from LEO satellites, even when the ground infrastructure is inoperative, such as during large-scale natural disasters.

B. System Model

The system model used in the simulation consists of the following components: a receiver simulating the LEO satellite, transmitters simulating UEs, a 3D urban environment model, and an ideal antenna model. UEs are randomly distributed in outdoor locations within the 3D city model and emit radio signals at random time intervals. These signals are observed by a highly directional antenna mounted on a LEO satellite orbiting at an altitude of 400 km.

Further details regarding the parameters, assumptions, and configurations of each component are provided in Section III.

C. Use Cases

A primary use case of the proposed system is estimating the number of disaster victims in affected areas during emergency situations. When communication infrastructure becomes inoperable, it becomes difficult to assess how many people are in need of rescue, which is critical for planning effective relief operations. In such cases, the proposed system can offer a valuable solution, as it leverages LEO satellites that remain unaffected by ground-level damage caused by disasters such as earthquakes.

Moreover, the smartphone penetration rate in Japan exceeds 90% [10], making it feasible to use the number of smartphones in an area as a proxy for the number of individuals. The proposed system can help identify regions requiring further detailed investigation and enable coordination with other systems such as UAV-based surveys.

In addition to disaster response scenarios, the proposed system can be utilized in peacetime situations. For instance, when localized outages in communication infrastructure occur, estimating the number of people in affected areas can assist in

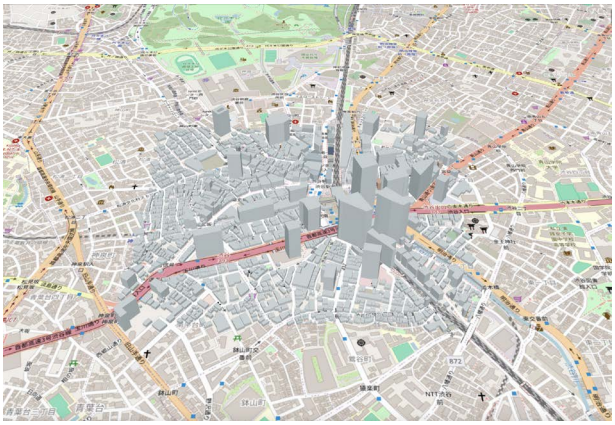

Fig. 2: 3D urban environment model(around Shibuya Station) in MATLAB.

planning the deployment of mobile base stations. This highlights the versatility and broad applicability of the proposed approach.

III. System Model Components and Simulation Overview

This section describes the simulation overview based on the system model outlined in Section II, detailing the model components, simulation scenario, parameter settings, and an evaluation related to indoor terminals.

A. System Model Components

1) 3D Urban Environment Model: To reproduce a realistic outdoor environment, a 3D building model was constructed in MATLAB by importing OpenStreetMap (OSM) data. The simulation targeted a 1 km^2 urban area around Shibuya Station in Tokyo, assuming it as a disaster-affected region. The coordinates of the target area range from 35.65387° N to 35.66250° N in latitude and from 139.69364° E to 139.70608° E in longitude. Figure 2 shows the corresponding 3D urban model implemented in MATLAB. The constructed model accurately reproduces the actual building environment and includes elevation information, enabling precise calculation of the propagation distance from UE to the LEO satellite and evaluation of the corresponding path loss.

2) Antenna Model: Due to the large propagation distance between the LEO satellite and the user equipment UE, significant path loss occurs. Therefore, the antenna mounted on the LEO satellite must exhibit high gain to compensate for this loss. Additionally, to achieve high spatial resolution within the target area, the antenna is required to have high directivity.

In this study, we adopt an ideal directional antenna model whose gain is defined by the following equation [11]:

$$G = \alpha \left| \frac{\sin\left(k\frac{\pi}{2}\phi_{\mathrm{rad}}\right)}{k\frac{\pi}{2}\phi_{\mathrm{rad}}} \right| - \beta \quad [\mathrm{dBi}] \tag{1}$$

$$\phi_{\mathrm{rad}} = \frac{180(\phi_{\mathrm{deg}} - 90)}{\pi} \quad [\mathrm{rad}] \tag{2}$$

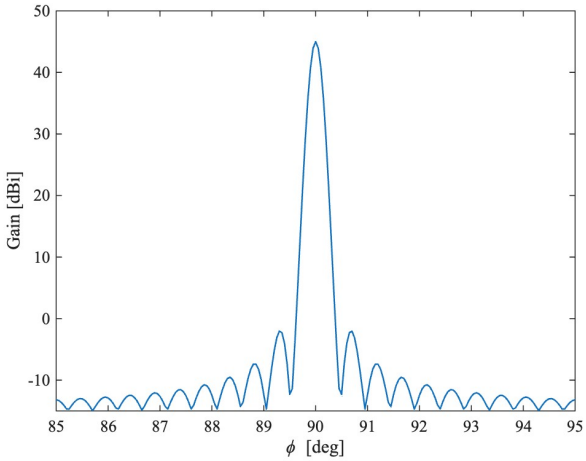

Fig. 3: Antenna pattern for simulation.

In this model, α determines the maximum antenna gain, k controls the sharpness of the peak (i.e., beamwidth), and β represents a gain offset. Figure 3 illustrates the antenna pattern defined by this model.

For the simulation, we set the constants to $\alpha = 60$, $k = 240$, and $\beta = 15$ such that the antenna's half-power beamwidth corresponds approximately to the coverage of a 1 km^2 target area [12].

The formulation in (2) shows that the antenna gain is maximized when the UE is located directly beneath the LEO satellite.

3) LEO and UE Configuration: The LEO satellite is modeled as a receiver equipped with the directional antenna described in Section III-A.2 and is positioned at an altitude of 400 km directly above the area defined in Section III-A.1.

UE is modeled as a transmitter equipped with an omnidirectional antenna, assuming an antenna gain of 0 dBi. UEs are placed exclusively in outdoor locations within the target area. The rationale for restricting UEs to outdoor settings is discussed later in Section III-C.

The transmission frequency used by the UEs is set to the 5.6 GHz band, which is legally permitted Wi-Fi band for outdoor use in Japan [13].

B. Simulation Overview

The simulation conducted in this study is based on the following scenario:

1) Between 1 and 500 UEs are randomly distributed in the target area, each emitting a signal with a power of 10 dBm at 5.6 GHz, at randomly assigned time instances.
2) A LEO satellite orbits at an altitude of 400 km and observes the radio signals as it passes over the target area.
3) Each UE transmits at least once during the observation window of the LEO satellite.
4) The LEO satellite is capable of performing beamforming toward the target area.

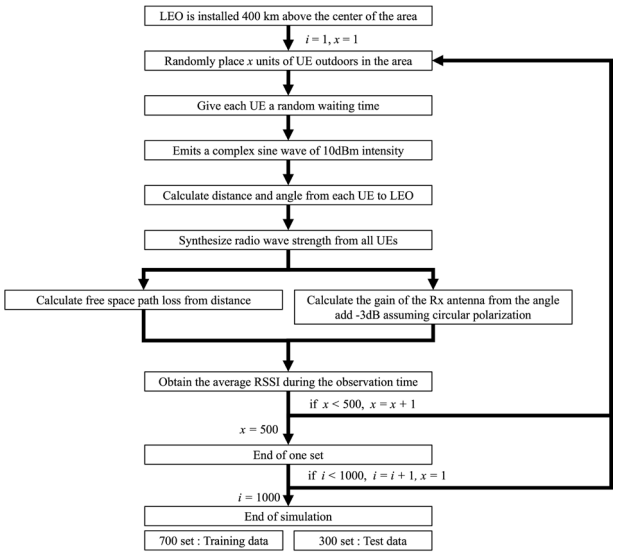

Fig. 4: Flowchart of dataset generation.

5) The LEO satellite conducts repeated observations regardless of whether the ground communication infrastructure is functional. During normal operation, the system collects both RSSI values and the number of connected UEs (via base stations and GNSS), and stores them as training data. In failure scenarios (e.g., post-disaster), only RSSI is available, and the stored training data is used to estimate the number of terminals in the area.

To simplify the simulation, the following assumptions are made:

- The receiving antenna is assumed to be circularly polarized, allowing reception of both horizontally and vertically polarized signals.
- Only free-space path loss calculated by Friis propagation formula [14] and a 3 dB attenuation due to polarization mismatch are considered as losses.
- The observation time is assumed to be the instant when the LEO satellite is directly above the center of the target area.

The estimation accuracy for terminal density is defined as follows:

- The number of UEs is categorized into three groups: small-scale (1–150 devices), medium-scale (151–350 devices), and large-scale (351–500 devices).
- If the predicted group matches the actual group, the estimation is considered successful. The ratio of successful estimations over the total number of trials is used as the performance metric.

Figure 4 shows the flowchart representing the process for generating the simulation dataset.

C. Indoor Terminal Evaluation

To examine whether the proposed system can also cover indoor terminals, we investigated the attenuation characteristics of radio wave propagation from indoor to outdoor

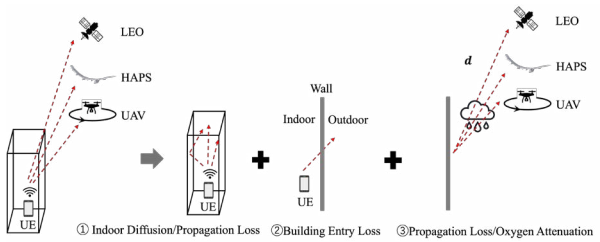

Fig. 5: Indoor to outdoor propagation loss model.

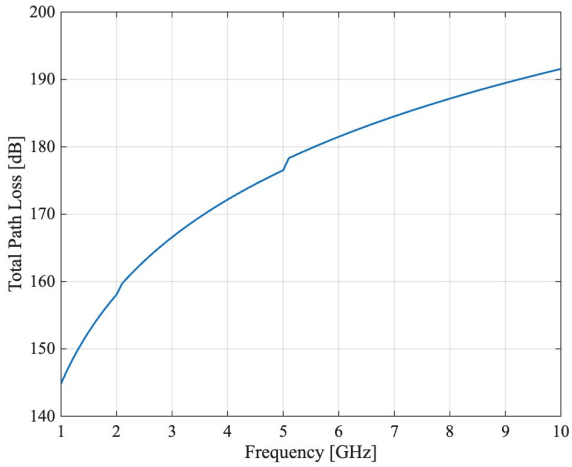

Fig. 6: Indoor-to-Outdoor Total Path Loss vs Frequency (1–10 GHz).

environments. The loss model constructed for this study is illustrated in Fig. 5.

The total loss model consists of the following three components:

1) Indoor Diffusion and Distance Loss: [15]

According to ITU-R Recommendation P.1238, signal attenuation in indoor environments across frequencies from 900 MHz to 100 GHz is primarily caused by signal diffusion and distance-dependent loss. For transmissions from non-top floors, additional floor penetration loss is also considered. The indoor propagation loss L is given by:

$$L = 20\log_{10} f + N\log_{10} d + L_f(n) - 28 \quad \text{[dB]} \qquad (3)$$

where f is the frequency [MHz], N is the distance power loss coefficient, d is the propagation distance [m], $L_f(n)$ is the floor penetration loss [dB], n is the number of floors penetrated, and 28 is an offset such that $L = 0$ when $f = 1$ MHz and $d = 1$ m.

2) Wall Penetration Loss: [16]

When radio waves exit a building, additional attenuation occurs as the signal penetrates the wall. ITU-R Recommendation P.2109-2 provides a method to compute this building entry loss (BEL), which is expressed as:

$$L_{BEL}(P) = 10\log_{10}\left(10^{0.1A(P)} + 10^{0.1B(P)} + 10^{0.1C}\right) \text{ [dB]} \qquad (4)$$

Here, f is the frequency [GHz], P is the probability that the actual loss does not exceed the calculated value, and C is a correction constant. $A(P)$ and $B(P)$ depend on the building material and require complex intermediate calculations, which are omitted in this paper. For full details, refer to ITU-R P.2109-2 [16].

3) Free-Space and Atmospheric Loss: [14]

Outdoor propagation loss includes free-space path loss and atmospheric attenuation. In this study, free-space path loss is computed using Friis' transmission formula. Atmospheric attenuation is negligible below 10 GHz and thus is not considered in our evaluation [17].

4) Numerical Evaluation of Indoor-to-Outdoor Loss: Based on the models in Sections III-C.1 to C.3, we performed numerical evaluations to estimate the total loss incurred when signals propagate from indoor terminals to the outdoor environment. The evaluation results are presented in Fig. 6.

The simulation conditions are as follows:

- Frequency range: 1–10 GHz
- Building is single-story or the transmitter is located on the top floor (no floor penetration loss)
- Signal exits through a typical glass window
- Incidence angle on the window is 60 degrees
- Probability P that actual loss does not exceed the calculated value is set to 0.5
- Indoor propagation distance: 5 m; Outdoor distance: 50 m

The total path loss curve shown in Fig. 6 is generally monotonically increasing; however, a slight inflection is observed around the 5 GHz. This anomaly arises from the discrete switching of the coefficient N used in the indoor distance loss model, defined in equation (3), for each frequency band, as adopted in Section III.C.1.

As shown in Fig. 6, signal propagation from indoor to outdoor environments in the 1–10 GHz frequency range incurs significant attenuation, exceeding 140 dB. Since the proposed system utilizes LEO satellites, the additional free-space path loss further increases the overall attenuation when compared to terrestrial receivers.

Given these findings, it is reasonable to limit the scope of the current study to outdoor terminals only. In practical use cases, it would be effective to utilize the proposed system to obtain a coarse estimate of terminal distribution, and then perform more detailed assessments using systems capable of closer-range observations, such as UAV-based platforms.

IV. SIMULATION RESULTS AND DISCUSSION

Using average RSSI data, we performed terminal density group classification through multiple machine learning methods. The classification results are shown in Figs. 7 and 8.

All machine learning models employed in this study were implemented using built-in MATLAB commands. The parameters for each model are listed below. Some parameter tuning was supported by OpenAI's GPT-5.1 model.

1) Decision Tree Regression (Tree)
   - Uses all training data
   - Maximum of 100 splits (nodes)
2) Random Forest (RF)
   - Uses all training data
   - 100 trees with random sampling
   - Maximum of 8 splits per tree
3) Gradient Boosting (GB)
   - Uses all training data
   - 250 boosting stages
   - Learning rate: 0.05
   - Maximum of 8 splits per tree
4) Support Vector Machine (SVM)
   - Uses 5000 training data
   - Radial Basis Function (Gaussian) kernel
5) XGBoost (XGB)
   - Uses all training data, with target variable p transformed to $\log(1+p)$
   - 350 boosting stages
   - Learning rate: 0.03
   - Maximum of 12 splits per tree
6) Multilayer Perceptron (MLP)
   - Uses all training data
   - One hidden layer with 50 nodes
   - 50 training iterations

The mixed model(Mix) applies a weighted stacking of the six models mentioned above (Tree, RF, GB, SVM, MLP, XGB) with weights $w = [0.06, 0.06, 0.25, 0.18, 0.15, 0.30]$. The weights were optimized using nonlinear least squares on the training dataset.

Figure 7 shows the accuracy of group classification for small-scale (1–150 UEs), mid-scale (151–350 UEs), and large-scale (351–500 UEs) when each model is applied. All six machine learning algorithms and the mixed model(Mix) achieved classification accuracy of 85% or higher, with most reaching approximately 90%. This result suggests that the proposed system has strong potential to estimate terminal density within target areas from LEO satellite observations with high accuracy. However, two main challenges were also identified. First is the method for selecting machine learning models. As shown in Fig. 7, while RF demonstrates superior results compared to other models in the mid-scale range (151–350 UEs), it shows the lowest accuracy in the large-scale range (351–500 UEs). Although there are differences among models, the largest difference is approximately 5%, indicating that the six machine learning algorithms and the Mix model yield nearly identical estimation results. In other words, while high-accuracy results can be expected regardless of which model is used, further validation is necessary to select the best model for all cases. The second challenge is improving estimation accuracy near group boundaries. Figure 8 shows the relationship between the true number of UEs and group classification success rate, revealing a significant drop in success rate near 150 and 350 units where groups transition. This is clearly the cause of degraded classification accuracy and requires countermeasures in the future.

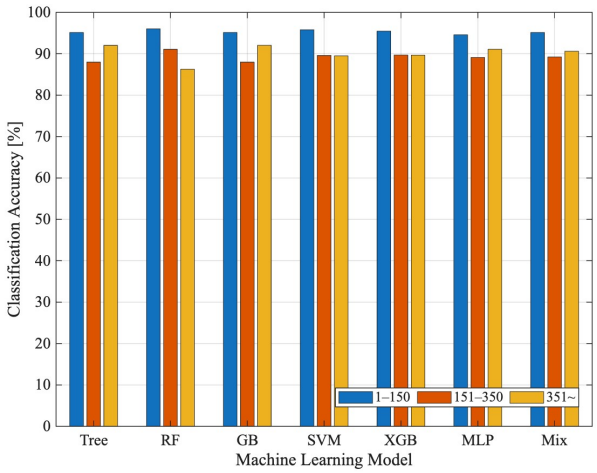

Fig. 7: Classification accuracy.

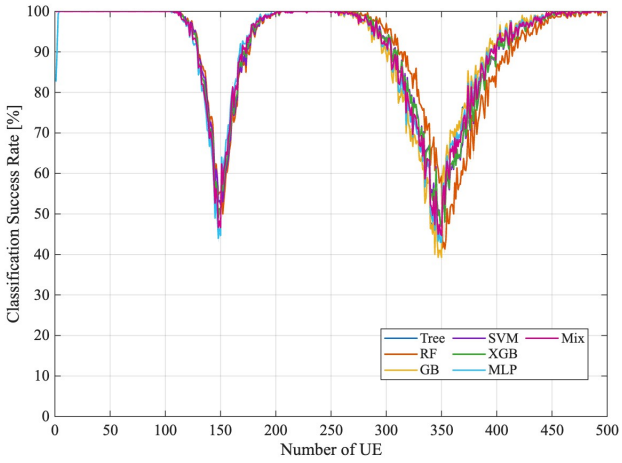

Fig. 8: Classification success rate vs Number of UE.

V. Conclusion and Future Works

In this paper, we proposed a system for estimating the density of ground-based mobile terminals by leveraging LEO satellite observations and machine learning. Although the simulation included several simplifying assumptions, the six machine learning models and the mixed model evaluated in this study achieved scale classification accuracy ranging from a minimum of 85% to a maximum of 95%, suggesting the feasibility and effectiveness of the proposed system.

However, the current study is limited in that the number of UEs was capped at 500 due to computational constraints. Further evaluation is necessary for scenarios involving larger numbers of terminals. Additionally, future work will address several remaining challenges, including the consideration of interference from terminals located outside the target area, development of an estimation model capable of handling any scenario, improvement of estimation accuracy near group boundaries, and the integration of complementary systems—such as UAV-based short-range observation platforms—for real-world deployments.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 24K00940. The authors would also like to express their sincere gratitude to the Satomi Scholarship Foundation.

References

[1] I. Shayea, A. A. El-Saleh, M. Ergen, B. Saoud, R. Hartani, D. Turan, and A. Kabbani, “Integration of 5G, 6G and IoT with Low Earth Orbit (LEO) networks: Opportunity, challenges and future trends,” Results in Engineering, vol. 23, p. 102409, 2024, doi: 10.1016/j.rineng.2024.102409.

[2] A. Castro-Carrera, “A systematic review of LEO satellite services for IoT,” Multidisciplinary Reviews, vol. 8, no. 3, p. 2025083, 2024, doi: 10.31893/multirev.2025083.

[3] A. K. Dwivedi, S. Chaudhari, N. Varshney, and P. K. Varshney, “Performance analysis of LEO satellite-based IoT networks in the presence of interference,” IEEE Internet of Things Journal, vol. 11, no. 5, pp. 8783–8799, Mar. 2024, doi: 10.1109/JIOT.2023.3321574.

[4] C. Zhang, H. Peng, Y. Ji, T. Hong, and G. Zhang, “Adaptive resource optimization for LoRa-enabled LEO satellite IoT system in high-dynamic environments,” Sensors, vol. 25, no. 11, p. 3318, 2025, doi: 10.3390/s25113318.

[5] R. Janssens, E. Mannens, R. Berkvens, and S. Denis, “Device-free crowd size estimation using wireless sensing on subway platforms,” Applied Sciences, vol. 14, no. 20, p. 9386, 2024, doi: 10.3390/app14209386.

[6] H. M. Zaidan, E. A. Mohammed, and D. H. Alhelal, “Estimation of the number of people in an indoor environment based on WiFi received signal strength indicator,” Bulletin of Electrical Engineering and Informatics, vol. 10, no. 3, pp. 1473–1481, Jun. 2021, doi: 10.11591/eei.v10i3.3038.

[7] S. Depatla and Y. Mostofi, “Crowd counting through walls using WiFi,” Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom), Athens, Greece, pp. 1–10, 2018, doi: 10.1109/PERCOM.2018.8444589.

[8] F. Noda and G. K. Tran, ”Proposal of LEO Based Population Estimation System Using Smartphone Emitted WLAN signals,” 2024 IEEE 21st Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2024, pp. 1080-1081, doi: 10.1109/CCNC51664.2024.10454793.

[9] F. Noda and G. K. Tran, ”Study on Estimation Method of the Number of Smartphones on the Ground Assuming Observation from Low Earth Orbit Satellites,” IEICE Technical Report, SR2025-7, vol. 125, no. 52, pp.21-22, Jun. 2025.

[10] Ministry of Internal Affairs and Communications (in Japan), ”Information and Communication White Paper, 2023 Edition,” pp.137, accessed 30.Jul.2025. [Online]. Available: https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r05/pdf/00zentai.pdf

[11] C.Balanis, ”Antenna Theory: Analysis and Design,” John Wiley & Sons, pp.575-602, 2016.

[12] E. Kang, Y. Park, J. Kim, and H. Choo, ”Downlink Analysis of a Low-Earth Orbit Satellite Considering an Airborne Interference Source Moving on Various Trajectory,” Remote Sensing, vol. 16, no. 2, p. 321, Jan. 2024, doi: 10.3390/rs16020321.

[13] Ministry of Internal Affairs and Communications (in Japan), Radio Use Portal, ”Outdoor/aerial use of wireless LAN,” accessed 30.Jul.2025. [Online]. Available: https://www.tele.soumu.go.jp/j/sys/others/wlan_outdoor/

[14] H. T. Friis, ”A Note on a Simple Transmission Formula,” in Proceedings of the IRE, vol. 34, no. 5, pp. 254-256, May 1946, doi: 10.1109/JRPROC.1946.234568.

[15] International Telecommunication Union, ”Propagation data and prediction methods for the planning of indoor radiocommunication systems and radio local area networks in the frequency range 900 MHz to 100 GHz,” Recommendation ITU-R P.1238-7, Feb. 2012.

[16] International Telecommunication Union, ”Prediction of building entry loss,” Recommendation ITU-R P.2109-2, Aug. 2023.

[17] L. Luini and C. G. Riva, ”A Simplified Model to Predict Oxygen Attenuation on Earth-Space Links,” in IEEE Transactions on Antennas and Propagation, vol. 65, no. 12, pp. 7217-7223, Dec. 2017, doi: 10.1109/TAP.2017.2765541.