

Article

Relabeling for Indoor Localization Using Stationary Beacons in Nursing Care Facilities

Christina Garcia * and Sozo Inoue 

Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu Ward, Kitakyushu 808-0135, Japan; sozo@brain.kyutech.ac.jp

* Correspondence: garcia.christina-alvarez442@mail.kyutech.jp or alvarez7.christina@gmail.com

Abstract: In this study, we propose an augmentation method for machine learning based on relabeling data in caregiving and nursing staff indoor localization with Bluetooth Low Energy (BLE) technology. Indoor localization is used to monitor staff-to-patient assistance in caregiving and to gain insights into workload management. However, improving accuracy is challenging when there is a limited amount of data available for training. In this paper, we propose a data augmentation method to reuse the Received Signal Strength (RSS) from different beacons by relabeling to the locations with less samples, resolving data imbalance. Standard deviation and Kullback–Leibler divergence between minority and majority classes are used to measure signal pattern to find matching beacons to relabel. By matching beacons between classes, two variations of relabeling are implemented, specifically full and partial matching. The performance is evaluated using the real-world dataset we collected for five days in a nursing care facility installed with 25 BLE beacons. A Random Forest model is utilized for location recognition, and performance is compared using the weighted F1-score to account for class imbalance. By increasing the beacon data with our proposed relabeling method for data augmentation, we achieve a higher minority class F1-score compared to augmentation with Random Sampling, Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN). Our proposed method utilizes collected beacon data by leveraging majority class samples. Full matching demonstrated a 6 to 8% improvement from the original baseline overall weighted F1-score.

Keywords: oversampling; data augmentation; machine learning; signal measurement; signal pattern; relabeling; indoor localization; beacon; nursing care



Citation: Garcia, C.; Inoue, S. Relabeling for Indoor Localization Using Stationary Beacons in Nursing Care Facilities. *Sensors* **2024**, *24*, 319. <https://doi.org/10.3390/s24020319>

Academic Editor: Boon Giin Lee

Received: 9 December 2023

Revised: 31 December 2023

Accepted: 2 January 2024

Published: 5 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The increasing population of older adults is impacting the nursing workforce, leading to a shortage of skilled staff [1]. As the demand for services grows, the use of nursing homes is also escalating, resulting in a rise in the patient-to-caregiver ratio [2,3]. Research efforts to better comprehend activity patterns during patient assistance are significant to leverage staff and improve elderly care delivery [4].

Indoor Positioning System (IPS) allows localization within enclosed spaces, facilitating navigation and the tracking of individuals or objects through a network of transmitters and receivers [5,6]. Employing indoor positioning to monitor care routine and patient assistance is helpful to support nursing records and to optimize care response [7–9]. Several key positioning techniques exist, but the standard approach includes trilateration, triangulation, multilateration, and fingerprinting [10]. However, these techniques are limited by environmental factors such as hardware requirements, setup complexity, signal obstruction, and time synchronization.

IPS differ in technology, signal measurement, and localization techniques tailored to suit the environment and particular use-case requirements [8]. Radiofrequency-based technologies [11], specifically Wi-Fi [12–16], Radio Frequency Identification (RFID) [17,18],

and Bluetooth Low Energy (BLE) [19–21], are typically preferred considering flexibility to setup and integration with IoT. Recently, studies on Wi-Fi have been optimizing line of sight (LOS) and non-line of sight (NLOS) [22,23] approaches, including combination with vision [24] for localization. Furthermore, sensors such as IMU [25] and Geomagnetism [26] are integrated in movement and positioning. However, considering restrictive data collection in caregiving activities, beacons are preferred in nursing homes with straightforward installation and lesser complexity imposed on participants [3,27]. The presence of unavoidable equipment in the facility can also hamper the collection of geomagnetic data. Given that Indoor Positioning Systems (IPS) are contingent upon the specific use case and environment, Bluetooth Low Energy (BLE) has been selected based on target availability, straightforward deployment, and privacy considerations pertinent to elderly individuals. Additionally, the energy-efficient power consumption of BLE allows longer data collection with minimal disturbance in the facility monitoring the device remotely.

A common challenge to indoor localization accuracy is low quality of signals. Outlier detection [28] and filtering methods are applied to resolve this issue. Moving average [29], weighted average [30], Bayesian sequential Monte Carlo [20], and Gaussian filtering [31] has been proposed for signal smoothing to achieve better positioning. Kalman filter [21,32] is a widely used method for filtering signals in IPS. In addition, machine learning [7,10,14,29,33] and deep learning [12,30,34] methods are being employed to optimize signal features which are not covered by standard localization techniques. Network issues, environmental factors affecting signal quality, and hardware malfunction can impact collected data. Moreover, data imbalance due to unequal representation of different areas or activities within the facility in a real-world setting affects indoor positioning accuracy [2].

Data augmentation methods are utilized to address the issue of class imbalance, including in IPS applications. Data augmentation is a method to artificially increase the training dataset by creating a modified version of the current data. To increase real data, Random Sampling (RS) [17] is commonly applied to minority class samples by duplication. On the other hand, creating synthetic samples is an alternative method. Adaptive Synthetic Sampling (ADASYN) [35] and Synthetic Minority Over-sampling Technique (SMOTE) [36] both generate synthetic data points for the minority class by interpolation based on the nearest neighbors principle. Currently, existing oversampling techniques suffer from repeating data, resulting in overfitting, while synthetic samples are prone to noise and data misrepresentation. Since augmented data are intended to reflect actual samples, oversampling should be carefully considered. At present, augmentation methods do not optimize real data from the majority classes, and few studies exist that investigate the use of signal patterns for augmenting training data in the context of indoor localization [2].

This paper introduces a data augmentation approach to address the challenge of unequally represented locations in beacon-based indoor localization. By analyzing signal patterns in different rooms, we successfully implement a relabeling strategy, utilizing Received Signal Strength Indicator (RSSI) values from one location as a proxy in another. As more nursing homes adopt IPS with IoT, this research aims to leverage beacon data collected in real-world environments to improve indoor localization. The main objective of this research is to develop an augmentation method by increasing training samples with beacon data from different rooms to improve location detection of minority classes. We focus on identifying semantic locations where rooms are preferred over exact geographical positions to support caregiving records. Specifically, this study aims to address the following research questions:

1. How do we identify and match signal patterns between different locations in the facility? We utilize standard deviation and Kullback–Leibler divergence to analyze signal patterns, facilitating the identification of similar beacons across minority and majority classes.
2. How can we use the samples from other beacons to augment location with less data? We propose a novel relabeling approach that reassigns signals from matching beacons to areas with fewer samples, thereby addressing data imbalance issues.

3. What is the performance of oversampling based on signal pattern relabeling compared to existing augmentation methods? To evaluate, we compare the performance of relabeling with Random Sampling, ADASYN, and SMOTE applied to our collected real-world data from a nursing facility. To further assess, we test relabeling to different rooms to validate its effectiveness.

We employ different augmentation techniques in the training set comprising 3.5 days of data and perform indoor localization using 1.5 days of test data. Overall and target class F1-scores are measured to account for the data imbalance, including precision, and recall. Our proposed method achieves an improved target class F1-score of 27% to 40% compared to the baseline data. With full matching, the relabeling method demonstrates a 6 to 8% improvement from the original baseline overall weighted F1-score. KL divergence results in a better F1-score than the standard deviation for both full and partial matching. This method effectively expands the training set, enhancing model accuracy, as demonstrated in a nursing care facility where beacon devices and mobile applications are employed for data collection.

The subsequent sections of this paper are structured as follows: Section 2 covers the relevant research on indoor localization with BLE technology, including the distinction of our work over other localization and oversampling techniques. In Section 3, we elaborate on the proposed method, which encompasses signal pattern and relabeling approach, while Section 4 covers the data collection process, including the evaluation of proposed relabeling in comparison to baseline and other augmentation strategies. A comprehensive analysis of the results follows in Section 5 emphasizing the significant insights from our research. Section 6 concludes the paper, highlighting the main contributions and outlining potential directions of future work.

2. Related Literature

In this chapter, we cover relevant works on the design of indoor positioning implemented in the nursing facility. With the advent of the Internet of Things (IoT), devices that easily connect to the network such as beacons, tags, and mobile devices [2,37–40] are preferred in system design. Signal measurements are generally based on time, angle, and received signal strength (RSS) [28,29,31,38,41].

2.1. Indoor Localization with Beacons

Beacons are battery-driven radio transmitters used in indoor positioning with proximity sensors that emit BLE signals [42,43]. Unlike the classic Bluetooth, which is connection-oriented, BLE has advertising functionality and does not necessarily have to pair [44,45]. Beacons are preferred in IPS since they are flexible, easy to deploy, and cost-effective, with low power consumption that can last up to a year [46].

BLE beacons are used to track nursing activities to better understand care routines, which is crucial to optimize workload distribution considering the low staff-to-patient ratio [47]. Existing systems combine beacons with Wireless Local Area Networks (WLAN) for tracking both patients and nurses by analyzing RSSI, Time of Arrival (ToA), and the Angle of Arrival (AoA) [7]. This involves the setup of more hardware as receiver and transmitter modules. BLE beacons are often paired with smartphones used by caregivers for detection. To automatically record daily caregiving routines, beacons are used for time-spatial recognition [5,9].

In some nursing facilities, multiple patients share a common room. Considering the effects of setting up indoor positioning on the privacy of the elderly, which can affect their social interaction, the placement of beacons should be carefully considered in planning IPS [27]. Studies on nursing homes commonly use RSS and BLE for indoor localization [3,9,27] with its energy-efficient power consumption. However, noise and multipath can affect the signal quality and positioning performance.

2.2. Data Augmentation

The uneven distribution of classes within a dataset is referred to as data imbalance, which is a prevalent problem in machine learning. This indicates that the classifier performs well for the majority class but poorly for the minority class. Since minority classes are crucial to solving many real-world issues, correctly categorizing samples from this class is similarly crucial [48–50].

Data augmentation is significant in improving indoor positioning systems. The technical and non-technical challenges of IPS in real-world environments [51] have impacted the precision of indoor positioning with uncontrolled variables such as signal interference, physical obstructions, and hardware breakdown. Data collection in on-site scenario [52] affects data quality and labeling, resulting in an imbalanced and unequal representation of classes in dataset [2,53]. In order to create a robust and unbiased model for IPS, data augmentation shows up as a crucial and imperative approach to address data imbalance. By applying several changes to the current dataset, data augmentation creates synthetic data that diversify the training set and lessens the effects of class imbalance [54–56].

Three basic approaches dominate data-level solutions to the class imbalance issue: Random Sampling [57,58], Synthetic Minority Over-Sampling Technique (SMOTE) [56], and Adaptive Synthetic Sampling (ADASYN) [35]. By randomly adding additional copies of selected minority classes to the training data, Random Over-Sampling balances class distribution [57]. On the other hand, instead of oversampling by replacement, ADASYN and SMOTE oversample the minority class by producing synthetic instances [36]. However, these techniques have several limitations.

Overfitting can result from Random Over-Sampling since duplicates of minority-class samples are added to the dataset without adding new information. If the original data are already highly dimensional, this increases the calculation cost and lengthen the classifier's training period. Conversely, Random Under-Sampling randomly removes instances from the majority class, potentially leading to the neglect of essential data.

When training a model, applying SMOTE creates a linear mapping of the data, which can lead to overfitting issues. Also, there is a risk of overlap because the SMOTE method does not account for the position of general data close to the uncommon class data [49,56,59,60]. Similarly with ADASYN, possible class overlap occurs in boundary areas as oversampling targets to resample between neighboring minority and majority classes. In general, both synthetic approaches are sensitive to noise and need parameter tuning as both are dependent on data distribution [61].

A combination of augmentation techniques has been applied in indoor localization [62,63] including adding information to the reference dataset and deep learning-based approach [34,64], yet these involve computational complexities. Few studies have focused on using methods based on signal patterns between sensors to augment training data specifically for indoor localization purposes in nursing homes.

Accurate labels ensure that models learn the correct patterns and relationships in the data. In machine learning, relabeling is employed in the data augmentation process to handle proper class labeling, which is crucial for the accuracy of supervised learning models. In existing studies, relabeling is commonly applied to preserve the original class labels [65]. Recently, relabeling has been used to address class imbalance when using logistic regression by assigning new labels to classes with fewer instances [66]. Data augmentation leading to the loss of label information can reduce model performance.

2.3. Signal Pattern

Signal patterns refer to the specific way that radio frequencies, like Bluetooth, behave in an indoor setting. Fluctuations in the signal are essential metrics providing valuable insights into the alterations of signal intensity that occur within indoor settings [67]. Trilateration and triangulation are traditional localization techniques that utilize geometric principles to assess signal behavior, measuring distance and angle from reference points.

In the non-conventional approach, statistical techniques are optimized to handle inherent variability and uncertainty in the signal-based analysis. The received signal's standard deviation reflects the signal strength fluctuations at different positions or distances within an indoor environment [68–70]. On the other hand, variance measures the spread of the RSSI values around the mean, offering perspectives on the consistency of the signal strength measurements at different locations [68]. KL divergence is a measure of the difference between two probability density functions. In IPS, it is calculated to measure similarity in the online phase of unlabeled data in existing database and up-to-date data [10].

Integrating multiple signal sources, sensor fusion techniques, and machine learning provides more robust and reliable localization. Implementing adaptive algorithms that can learn and adjust to environmental changes further improves positioning performance. In this study, we delve into the signal patterns between locations by employing statistical techniques, specifically standard deviation and KL divergence, to measure features to find matching patterns for oversampling.

In this work, we propose to compare the signal pattern features of labeled beacon data from different locations and determine the divergence between minority and majority classes as a foundation for data augmentation. By considering the layout of the care facility, we focus on leveraging the current data from stationary beacons by using the majority class to oversample locations with less BLE data. We utilize relabeling to update labels of the augmented data derived from other locations, aligning them with the labels of the minority class. Specifically, signal patterns observed from selected positioned beacons from various rooms are utilized to guide the relabeling process. Full and partial matching represent two distinct relabeling variations that take into account the comprehensive arrangement of installed beacons.

This research strategically deploys BLE beacons to enable indoor localization of nursing staff, ensuring that the architecture of the nursing home remains unaltered and that caregiving services are delivered without interruption. This work prioritizes identifying semantic room locations over exact coordinates, considering the relevance to caregiving records. To protect the privacy of patients, all devices are mounted outside the doors of patient rooms.

3. Material and Methods

This section introduces our developed relabeling method depicted in Figure 1 for beacon data augmentation.

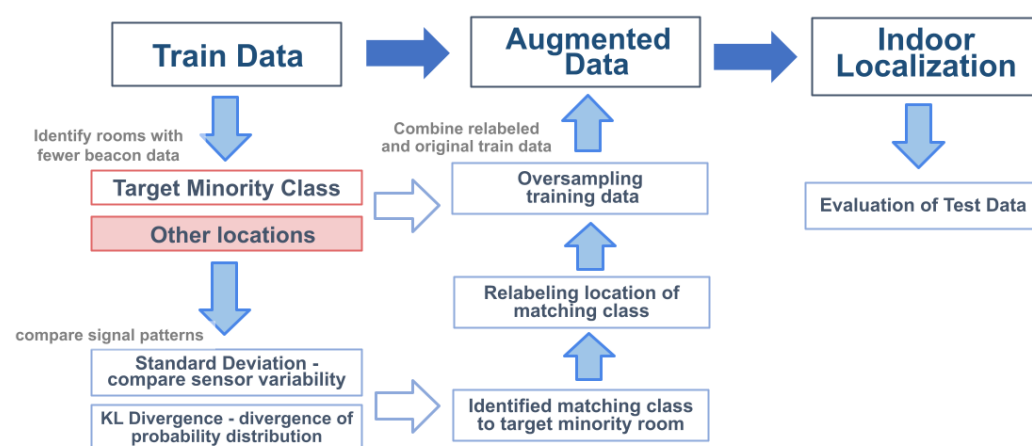


Figure 1. Proposed oversampling approach based on signal pattern relabeling for indoor localization.

In this approach, we aim to increase the training data sample, particularly locations with less beacon data. Initially, training data are analyzed to identify rooms with fewer beacon signals, termed as the "Target Minority Class". To oversample, we identify matching locations to minority classes from other locations by comparing signal patterns using

standard deviation and KL divergence. To proceed with data augmentation, we apply relabeling to the identified matching class by updating the location label similar to the target minority. The relabeled set is then added to the original training data, increasing the set. The augmented data are then used to enhance the indoor localization employing machine learning, with the model's efficacy evaluated using test data.

The subsection commences with information on the beacon data from the site. Understanding the signals detected from beacons to comprehend variability is necessary to execute relabeling. Figure 2 describes the overview of the indoor localization employed in the nursing facility from data collection to the use of the relabeling approach.

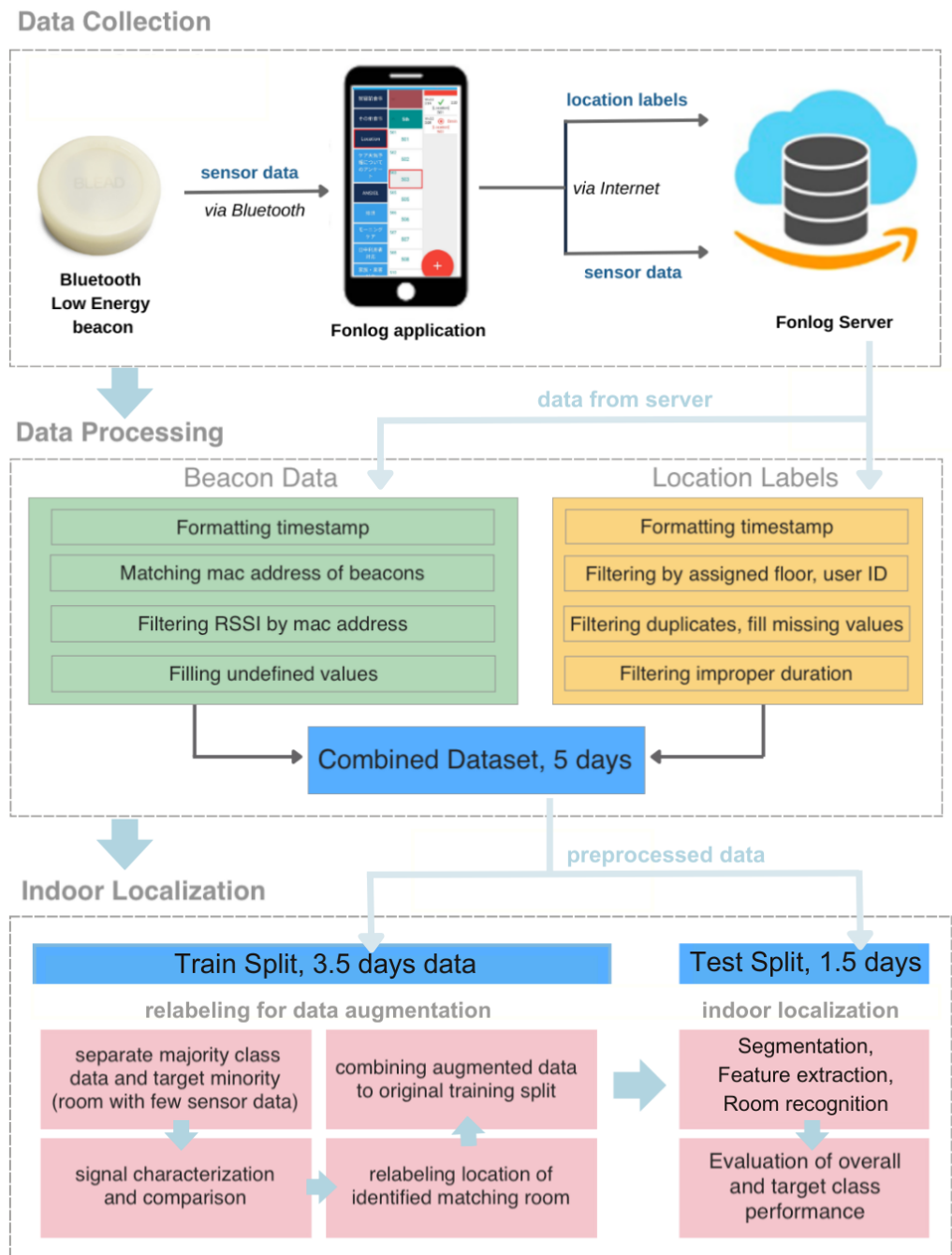


Figure 2. An overview of the proposed indoor positioning system in nursing care facility.

3.1. Signal from Detected Beacons

Prior to deployment, the respective MAC addresses of the beacons are recorded to match the raw data from the server, as shown in Table 1. After pre-processing the raw data, the MAC addresses of detected devices are filtered to match the beacons. The final dataframe of RSSI values is shown in Figure 3.

Table 1. Collected raw beacon data stored in the server.

User_id	Timestamp	Mac Address	RSSI
90	2023-04-10 10:22:55.589+0900	FD:07:0E:D5:28:AE	−75
90	2023-04-10 10:22:55.599+0900	D2:1C:25:72:FB:E3	−62

Based on Figure 3, we can suppose $RSSI_t$ represents the detected RSSI at timestamp t at location m . We then define the RSSI measurement for all beacons n , and the corresponding location label as in Equation (1). Overall, the signal database can be expressed as Equation (2) and the labels as Equation (3).

$$RSSI_t = [RSSI_{t1}, \dots, RSSI_{tn}], label_t = m_t, \quad (1)$$

$$RSSI = \{[RSSI_{11}, \dots, RSSI_{1n}], \dots, [RSSI_{t1}, \dots, RSSI_{tn}]\}, \quad (2)$$

$$label = \{m_1, m_2, m_3, \dots, m_t\}, \quad (3)$$

where t represents the timestamp, n denotes the total number of beacons installed on site, and m refers to the associated elderly room. With location label m , training data are observed to identify rooms with fewer detected signals.

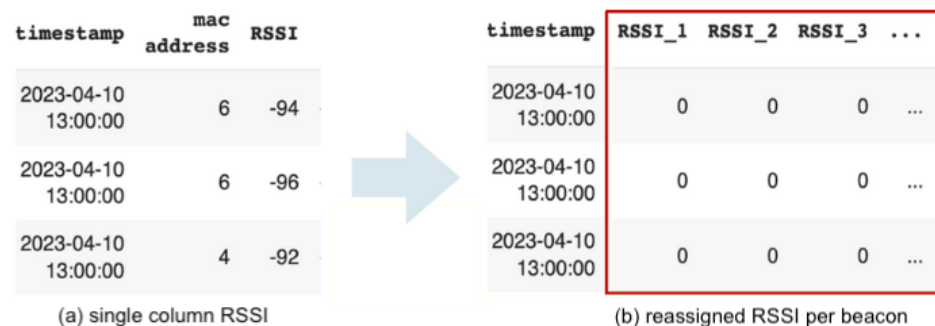


Figure 3. Matrix of the received signal strength from beacons. (a) Single column RSSI. (b) The corresponding RSSI for each beacon, repositioned into separate columns.

3.2. Matching

Generally, the accuracy of IPS decreases as the distance between the transmitter and receiver increases [71]. Observing this from the histogram of detected beacons and considering the setting of installed devices covering up to a 5 m range, we limit the analysis to six beacons to better comprehend signal patterns between rooms. Specifically, we only focus on elderly rooms installed with stationary beacons following a similar layout. Figure 4 shows the targeted surrounding beacons investigated for each room.

Prior to calculating the signal pattern feature, we classify rooms based on the completeness of the selected surrounding beacons. For each location, six BLE transmitters are filtered such that

$$f_m = [fl_m, sl_m, sm, fm, sr_m, fr_m]. \quad (4)$$

In Equation (4), s is the beacon on the location (source) room, f is the beacon on the room in front, fl and sl are the beacons on the left side rooms while sr and fr are beacons on the right side, respectively. In full matching, we only consider elderly rooms forming complete six beacons as candidate match to the minority class. On the other hand,

with partial matching, all elderly rooms are considered as possible match to the minority class regardless of incomplete sensors forming the targeted six. As an example, in Figure 4, Room 515 is considered only in partial matching where $f_{515} = [null, 13, 15, 1, 16, 2]$. Room 521, on the other hand, is also a candidate in full matching, where $f_{521} = [8, 20, 21, 10, 22, 11]$. For locations with incomplete beacons resulting in null values representing no signal detected, we replace null with zeroes for the model to better differentiate from distant beacons with higher negative values and to calculate the signal pattern feature.

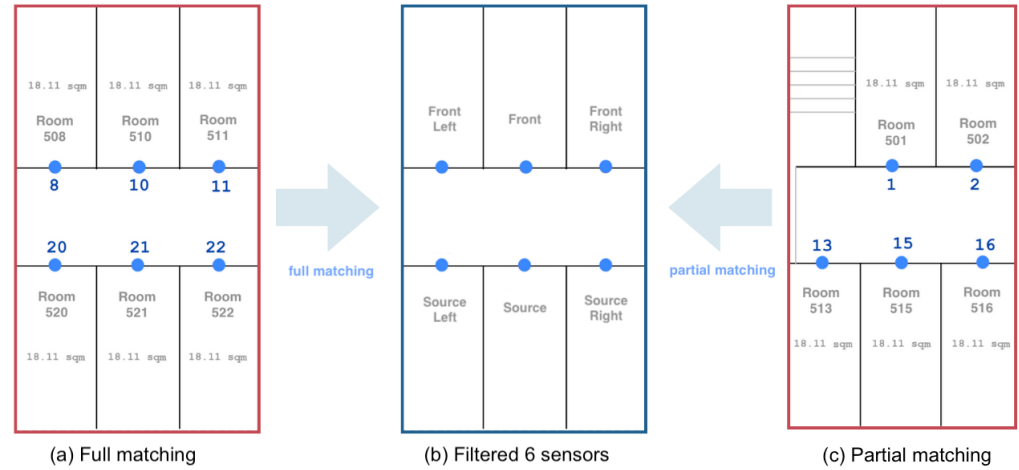


Figure 4. Full and Partial Matching based on selected sensors. (a) Full matching with complete 6 sensors. (b) A total of 6 sensors surrounding the location. (c) Partial matching with incomplete 6 sensors.

3.3. Relabeling Based on Signal Pattern

A signal pattern feature is computed from the six beacons to define the signal variability in different rooms. A sub-dataframe $pattern_m$ containing the RSSI values of the beacons in the list f_m is created for each room in the majority class. The sub-dataframe of the minority class is denoted as $pattern_{min}$. From these sub-dataframes, standard deviation and KL divergence are measured to represent each room's signal pattern feature, which is then used to identify similar locations. The objective is to pinpoint the room that has a matching signal pattern with the minority class in order to move forward with the relabeling process. Initially, the majority class is downsampled to match the number of samples in the minority class as depicted in Figure 5. Outlined in Steps 1 to 4 are the procedures to calculate the signal pattern feature of $pattern_m$ and $pattern_{min}$. See Algorithm 1.

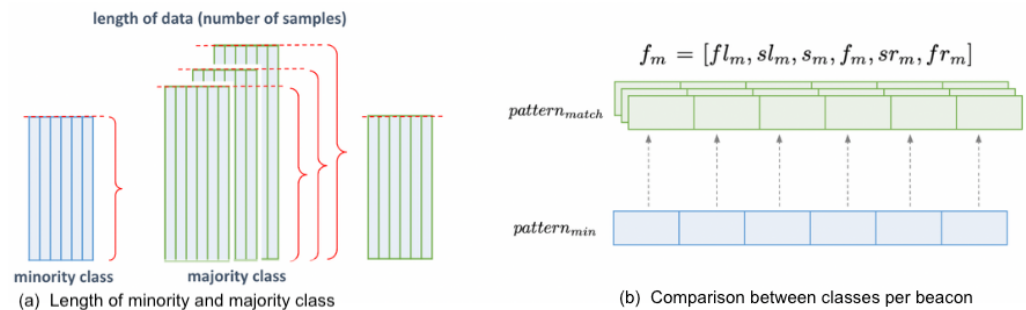


Figure 5. Comparing signal pattern feature between minority and majority class. (a) Length of samples considered. (b) Comparison between rooms per beacon.

The signal pattern feature based on standard deviation calculates the absolute differences in standard deviation for each beacon between minority and majority classes. This approach considers individually the variability difference for each beacon before summing

them. The matching location from the majority class is identified with the least standard deviation difference sum expressed in Equation (5),

$$\Delta\sigma_{\text{total}} = \sum |\sigma_{\text{min}} - \sigma_{\text{match}}|, \quad (5)$$

where $\Delta\sigma_{\text{total}}$ represents the total difference in standard deviation values, σ_{min} is the standard deviation for a specific beacon in the minority subdataframe, σ_{match} is the standard deviation of each beacon in the candidate sub-dataframes $pattern_m$, and $\sum$ represents the summation symbol, indicating the sum of absolute differences across all beacons.

Algorithm 1: Algorithm for comparing the signal pattern feature between rooms

Input:

$m \leftarrow$ room number

$d_{\text{train}} \leftarrow$ train data

Output:

$pattern_{\text{match}} \leftarrow pattern_m$

1. Identify rooms with small sample in train data, $df = d_{\text{train}}$.
 2. Define $pattern_{\text{min}}$ and $pattern_m$ following Equation (4) such that for i in labels, m
 $d_i = df[df['m']] == 'i']$
 if $i == "1"$: $f_1 = [fl_1, sl_1, s_1, f_1, sr_1, fr_1]$
 $\dots$
 elif $i == "25"$: $f_{25} = [fl_{25}, sl_{25}, s_{25}, f_{25}, sr_{25}, fr_{25}]$.
 3. Group $pattern_m$ accordingly as candidates for full and partial match.
 4. Calculate the signal pattern feature from $pattern_{\text{min}}$ and $pattern_m$.
 5. Compare the signal pattern features to identify $pattern_{\text{match}}$.
-

On the other hand, Kullback–Leibler (KL) Divergence is calculated from minority to majority classes. With this approach, we measure the divergence between the probability distribution represented by minority and majority classes, normalizing the data to represent probability distributions. The KL Divergence applied is expressed in Equation (6),

$$D_{\text{KL}}(P \parallel Q) = \sum P(i) \log \left(\frac{Q(i)}{P(i)} \right), \quad (6)$$

where $P(i)$ represents the probability of observation i in $pattern_{\text{min}}$ and $Q(i)$ represents the probability of observation i in candidate match $pattern_m$. After calculating and plotting all the statistical values, we observe the trend of the signals in $pattern_{\text{min}}$ and locate similarity in the rest of $pattern_m$ from all other rooms to find $pattern_{\text{match}}$. We apply both statistical measures to partial and full matching variations of relabeling forming four candidate matches.

To proceed with relabeling, train data d_{train} , $pattern_{\text{min}}$ for the location with low sample and identified $pattern_{\text{match}}$ are required. The expected output is the augmented train data $d_{\text{train,new}}$ with concatenated original train data d_{train} and relabeled data d_{relabel} . Listed from one to five are the sequential steps followed to execute the relabeling process. See Algorithm 2.

Nursing homes frequently feature an evenly distributed space across rooms. Given the environment's influence on signal behavior, the relabeling method is specifically designed for floors where the rooms share the same geometric layout. The relabeling method is exclusively applied to patient rooms in the original floor plan based on matching as depicted in Figure 4. In this study, relabeling is not applied for signals from beacons positioned in areas with diverse dimensions and layouts, such as open spaces in the cafeteria. It is crucial that the beacon devices are installed in fixed positions. Moreover, as the proposed approach specifically filters data to the six surrounding beacons, Standard Deviation and KL Divergence were chosen to measure signal patterns as they represent the variability of the targeted signals. To apply relabeling to other sensors, the representation of signal patterns

can be updated using other statistical measures that better suit the behavior of the data being analyzed.

Algorithm 2: Algorithm for relabeling the location of matching class

Input:

$d_{\text{train}} \leftarrow \text{train data}$
 $pattern_{\text{match}}, pattern_{\text{min}}$

Output:

$d_{\text{train, new}} \leftarrow d_{\text{train}}, d_{\text{relabelled}}$

1. Create a dataframe for the relabeled data $d_{\text{relabelled}}$.
 2. Populate $d_{\text{relabelled}}$ with values from $pattern_{\text{match}}$ following the same columns representing RSSI values of the six beacons.
 3. Add remaining columns from d_{train} in $d_{\text{relabelled}}$ for beacons not included in $pattern_{\text{match}}$. Fill with zero values.
 4. For relabeling, assign the location of the minority class to the labels for $d_{\text{relabelled}}$.
 5. To oversample, combine the augmented data $d_{\text{relabelled}}$ to the original train split d_{train} in a new dataframe $d_{\text{train, new}}$.
-

3.4. Indoor Localization

In this paper, indoor localization is approached as a recognition problem using machine learning [72] to identify semantic room locations. This approach is designed to enhance caregiving records by emphasizing the functional nature of spaces within the caregiving environment rather than their physical coordinates. For room estimation considering attenuation, RSSI values from all rooms are considered in feature extraction. Discrepancies between the timestamps of location labels and RSSI data are resolved by synchronizing data. To proceed with indoor localization, statistical and temporal features are derived from the RSSI matrix. The Random Forest algorithm is employed for identifying locations as this classifier demonstrates effectiveness in handling imbalanced data [73] and preferred in indoor positioning [74–76]. As we deal with a similar scenario of data imbalance between locations, we opt to adopt this model to focus on relabeling.

Five statistical features are extracted, specifically mean, standard deviation, minimum, maximum, and RSSI count. Integrating the quantity of detected signals per device into the feature set improves the model's effectiveness, as observed in prior work. Time-based attributes, specifically hour, minute, and microsecond, are extracted. A non-overlapping time window of 45 s is implemented. The train and test data division is date-based, adhering to train-test ratios recommended by empirical studies. The training set comprises 3.5 days of data, and the remaining 1.5 days are intended for test data. Both training and test sets reflect the caregiving routine in the facility, where nursing staff frequently visited the common area compared to the patient rooms.

To assess the effectiveness of relabeling, the indoor localization performance of baseline train data is compared to the performance of data augmentation. In addition, various data augmentation methods must be applied to the minority class. For cross-validation, the relabeling approach is applied to different rooms and compared with other methods. The next chapter covers the evaluation of real-world data collected in the nursing home.

4. Data Collection and Evaluation

The majority of the caregiving activities are location dependent; hence, monitoring the rooms visited by nursing staff is necessary for an estimation of the delivered workload and assistance given to elderly individuals on the floor. This information is relevant in automatic care recording, necessitating inputted FonLog data. The nursing care tasks executed by the caregiver fall under the category of patient care, medical assistance, documentation, cleaning, and organization. As the movement of the caregiver is dependent on nursing activities, we elaborate the details on caregiving to record the workload of staff performed

in respective locations. Patient care covers bathing, excretion assistance, daytime support, meal preparation, and recreation. Medical assistance encompasses checking patient vitals, administering treatment, and overseeing rehabilitation. Cleaning and organization tasks include changing linens and handling laundry. Care record falls under documentation.

4.1. Data Collection

The data collection approach implemented in the nursing facility is depicted in Figure 6. Pre-deployment testing was conducted with the caregiver to ensure the detection of beacons, connection to the server, and correct reading of both location labels and BLE signals. Furthermore, initial data gathering with the placement of beacons was conducted to make adjustments accordingly following the nursing home's protocol. Initially, user feedback was collected via discussion with the participating staff and user study during the pre-deployment setup. A location labeler was employed to address the staff's challenges in recording locations while carrying out their tasks.

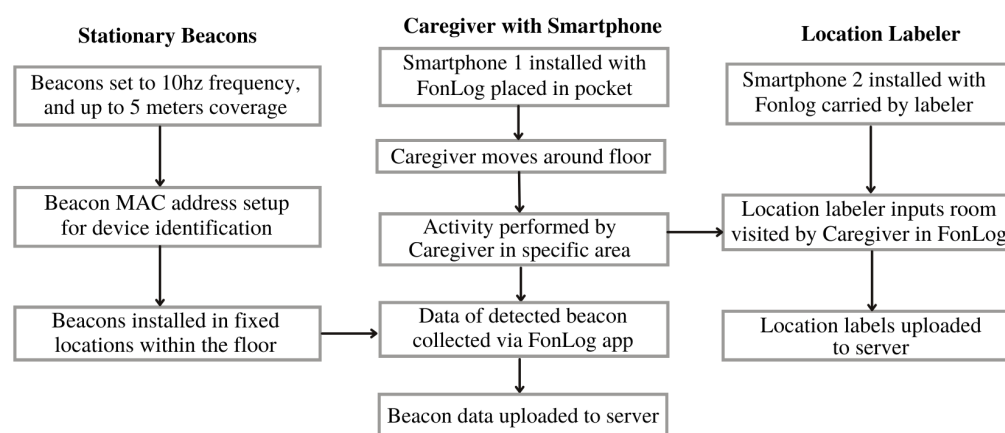


Figure 6. Overview of data collection approach in nursing facility.

Final data collection was performed on the fifth floor of a nursing care facility over a period of five days, from morning to afternoon of April 10 to 14, capturing the locations a caregiver visited during their daily routine. Daily, about seven hours of data collection was performed: three hours in the morning and four hours in the afternoon. The schedule followed the caregiving routine in the facility. In adherence, the placement of beacons outside patient rooms was decided to protect privacy, and data collection was executed following the facility's protocol and proposed schedule.

FonLog, a smartphone application in Figure 7, was installed and used as a data collection tool to log location labels and as a receiver of RSSI signals [39]. In this study, FonLog was explicitly customized to record activities performed on the respective floor of the nursing care facility with strategically positioned beacons.

Two separate smartphones installed with FonLog were set up for data gathering, one for the nursing staff with enabled BLE ID and another for the location labeler. Prior to deployment, the location and Bluetooth settings of the smartphone were set, and beacon detection was verified on the login interface, where the MAC (Media Access Control) addresses are displayed. The nursing staff was requested to bring one smartphone as they performed caregiving across various rooms on the floor. To avoid disturbance during tasks and maintain the quality of nurse care, the device was placed in the front pocket of the uniform.

A common challenge of on-site data collection is a mismatch between beacon data and the corresponding location label, which occurs due to the delay in activity recording. Caregivers usually log the activity after completing various tasks, resulting in discrepancies in timestamps during pre-processing. To mitigate this problem, the data collection process included an observer closely monitoring the caregiver's daily routine. The labeler assigned was positioned in the hallway and recorded locations on the second smartphone,

with Fonlog choosing from the list of rooms in the app. Both the carer and labeler were each assigned a unique user ID. Similarly, each patient room and area within the floor was allocated a unique customer ID. All beacon data and labels were temporarily stored in the local storage of each smartphone and uploaded to the same server once internet connection was established. In this study, the data collection involved participating nurses of average height. Any significant height differences should be noted, and the impact on RSSI values should be observed in the case of multiple participants.

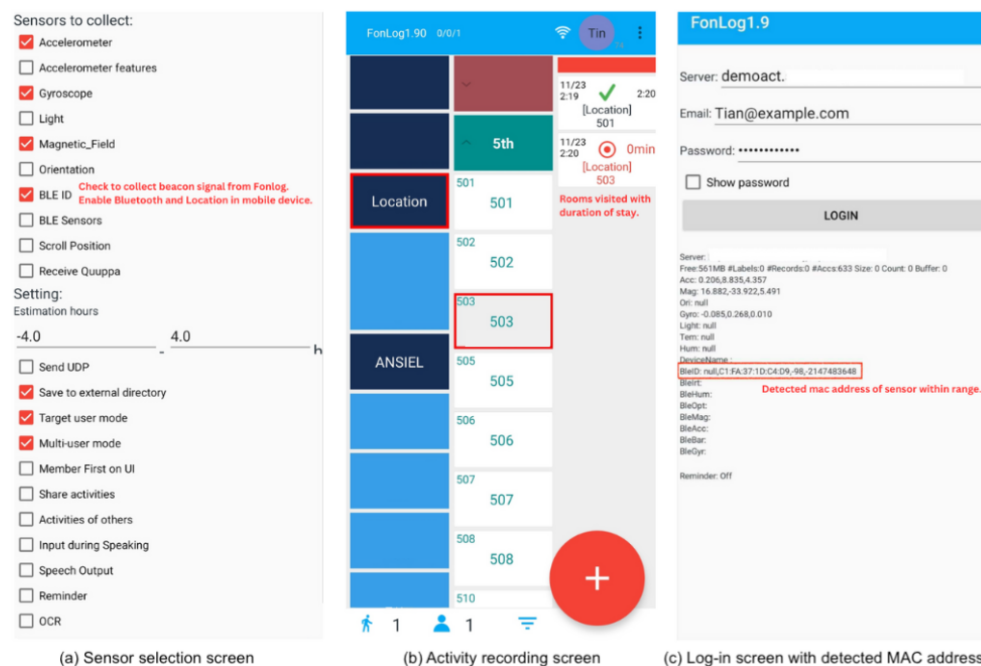


Figure 7. User interface of the FonLog application. (a) Sensor selection screen where BLE ID is enabled. (b) Activity recording screen showing respective rooms within the floor. (c) Log-in interface reflecting detected mac address of BLE device.

Beacon data recorded from Fonlog contain user_id, timestamp, MAC address, and RSSI, while location data by labeler consists of user_id, activity performed, start and stop time, customer_id, and specified location. In our initial approach, we aligned the timezones of both the location and beacon data, and the timestamps were subsequently reformatted. Duplicates and rows with null values were removed, and entries lacking in start-time and stop-time with undefined duration were filtered from location labels. Figure 8 depicts the map of the target floor in the nursing facility with installed beacons.

Before deployment, beacon frequency and height placement were decided based on experiments performed in the lab and hallway of the university, reflecting the facility's layout. In total, 25 stationary beacons indicated by blue points on the map were installed 2 m from the ground, strategically positioned outside 19 elderly rooms, and in common areas to cover the usual route of the nursing staff. All beacons were configured to a frequency of 10 Hz, where the actual data showed RSSI detection ranging from 3 to 5 Hz.

In medical and elderly care facilities, there are substantial limitations on where BLE beacons can be placed. In scenarios where users carry the beacon tag, scanners are strategically positioned in corners and specific areas, taking into account the geometry of the location [9]. This setup allows for position calculation using methods that do not rely on machine learning. For this research, the placement of beacons was determined in agreement with the nursing home to effectively track the movement of caregivers on the respective floor. Specifically, beacons were mounted at an appropriate height on the doors outside each patient room. This approach was adopted to respect privacy and to minimize disturbance to the elderly during the installation and maintenance of the devices. In contrast

to other studies [9], beacons were placed in stationary positions as emitters while Fonlog installed in the caregiver's smartphone served as the scanner.

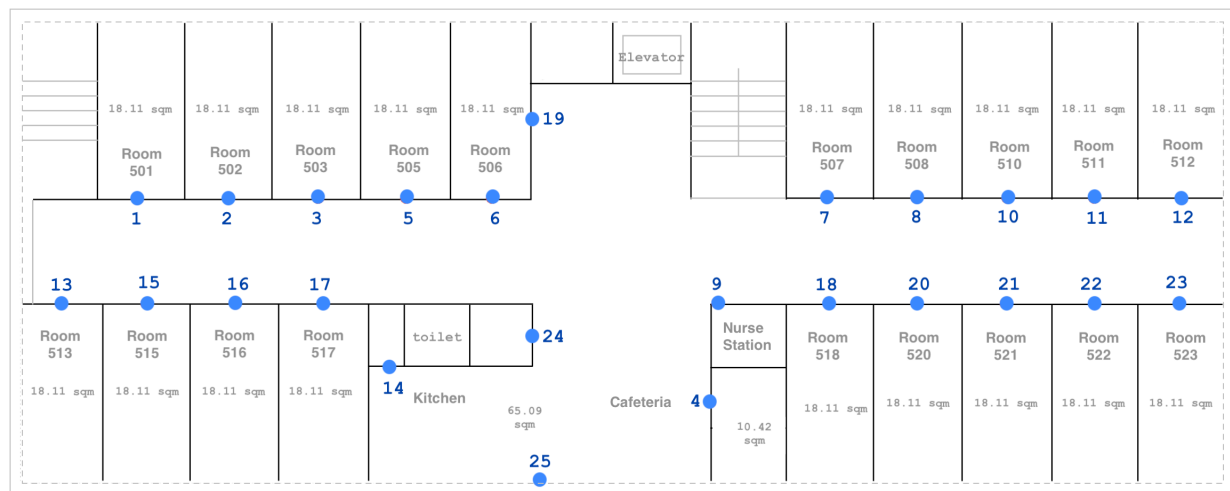


Figure 8. Layout of the facility showing the placement of installed beacons at each location.

The transmission power of beacons varies depending on the calibration of the RSSI at a distance of one meter. Beacons configured for broader coverage have higher transmission power. According to the specification sheet of the BLE device used, an RSSI detection range of 100 m corresponds to a transmission power of +4 dBm. All beacons were set to a detection range of zero to five meters coverage with a transmission power of up to −30 dBm (decibel-milliwatts).

Figure 9 displays the preprocessed beacon data with location labels used in the baseline of the indoor positioning model.

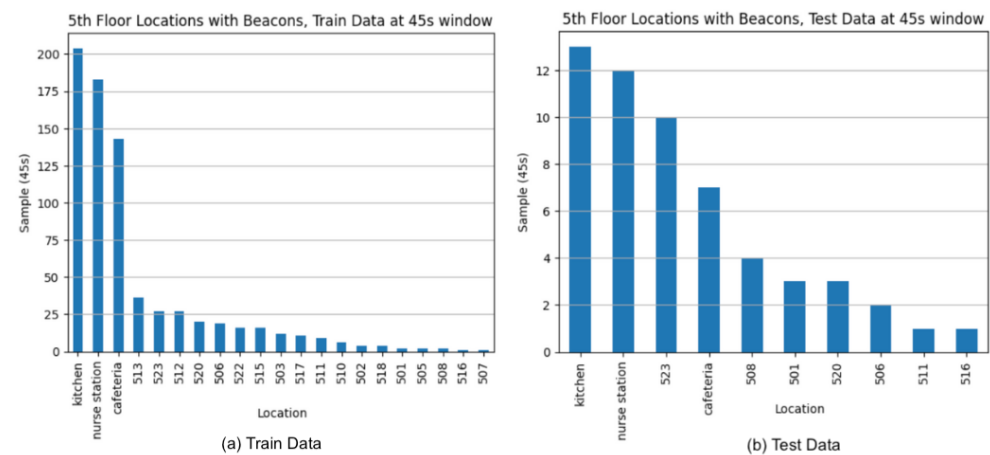


Figure 9. Baseline beacon data with location labels segmented for feature extraction. (a) Training split, 45 s window (b) Test split, 45 s window.

From the data, we can see an overview of the caregiving routine, which involves more extended activities in the kitchen, nurse station, and cafeteria, where meal assistance and nurse care recording are performed. We can also identify rooms with more extended data, suggesting longer patient assistance is delivered in the respective elderly room. The same test data from the figure are retained to evaluate and compare the performance of different data augmentation approaches in Section 4.3.

4.2. Performance Evaluation

As this study aims to address data imbalance by relabeling the approach, a comparative analysis of signal patterns and an assessment of the efficacy of partial and full room matching are conducted. Furthermore, we evaluate the performance of relabeling against other data augmentation methods. The test data cover 1.5 days, while the training data encompass 3.5 days, both segmented into 45 s windows without overlapping.

The target use case for evaluation in this study is centered on enhancing the performance of indoor localization through the meticulous comparison of signal patterns by specific features. By implementing partial and full matching of rooms, we aim to refine the relabeling process and ensure the most accurate representation of location data. This comparison also extends to assessing the efficacy of our proposed methods against other data augmentation approaches, thus providing a comprehensive analysis of performance improvements in indoor localization systems.

To evaluate the impact of data augmentation to indoor positioning performance, *F1-score*, *Precision* and *Recall* are measured and obtained, respectively, using the following Equations in (7)–(9), where *TP* is the True Positive, *TN* is the True Negative, *FP* is the False Positive, and *FN* the False Negative values.

$$F1\text{-score} = \frac{TP}{TP + 0.5(FP + FN)}, \quad (7)$$

$$Precision = \frac{TP}{(TP + FP)}, \quad (8)$$

$$Recall = \frac{TP}{(TP + FN)}. \quad (9)$$

We assess the efficacy of applying relabeling to underrepresented locations within the training dataset from the collected data. For this evaluation, reviewing the beacon data for each location in the training set, Rooms 516 and 507 are identified as minority classes. Filtering the device of all the rooms to six beacons following Equation (4) and Figure 4, we proceed with calculating the signal pattern feature. Figure 10 illustrates the resulting standard deviation of each of six beacons in $pattern_{min}$ corresponding to the minority class rooms.

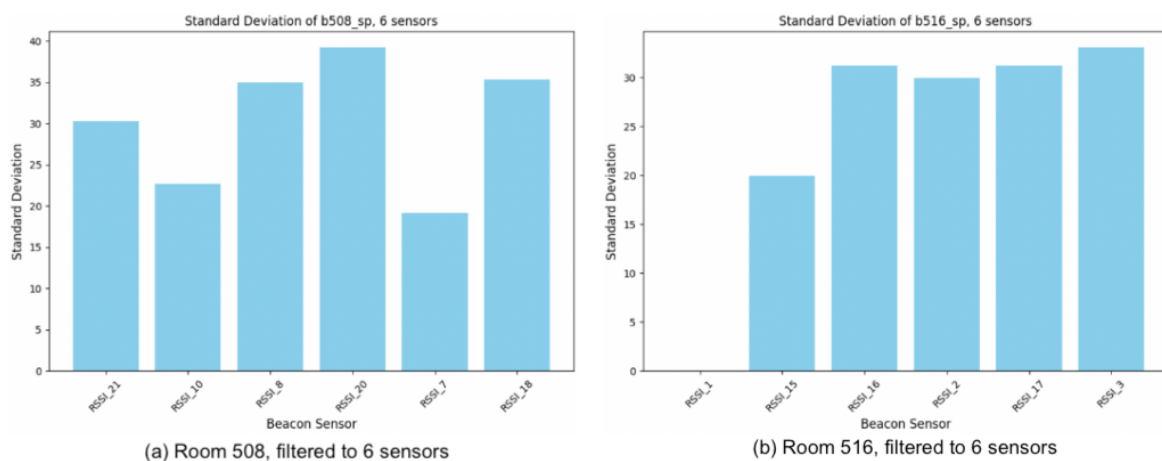


Figure 10. Standard deviation of selected sensors in target minority class rooms. (a) Room 508, filtered to 6 sensors (b) Room 516, filtered to 6 sensors.

We then find the match to the minority class by comparing the signal pattern feature of other rooms using standard deviation and KL divergence. Based on the floor layout, corner rooms with only less than six beacons are reserved as candidate matches in partial matching. In comparison, middle rooms with complete six beacons mounted in surrounding locations

are considered in full matching. This process allows for two types of relabeling: full and partial matching, by correlating beacons across different classes.

The integration of two variations of matching and two statistical measures for signal pattern features results in four potential scenarios for candidate matching in the relabeling process. To identify the best combination for relabeling, we implement data augmentation to all four use-case scenarios and compare it with other oversampling methods. From this approach, we aim to identify the statistical measure that better represents signal patterns from six beacons, and by comparing full and partial matching, we determine the locations to consider for relabeling.

In the case of full matching using signal patterns determined by standard deviation, as depicted in Figure 11, Room 520 emerges as the closest match to Room 508, and Room 511 is identified as the nearest match to Room 516.

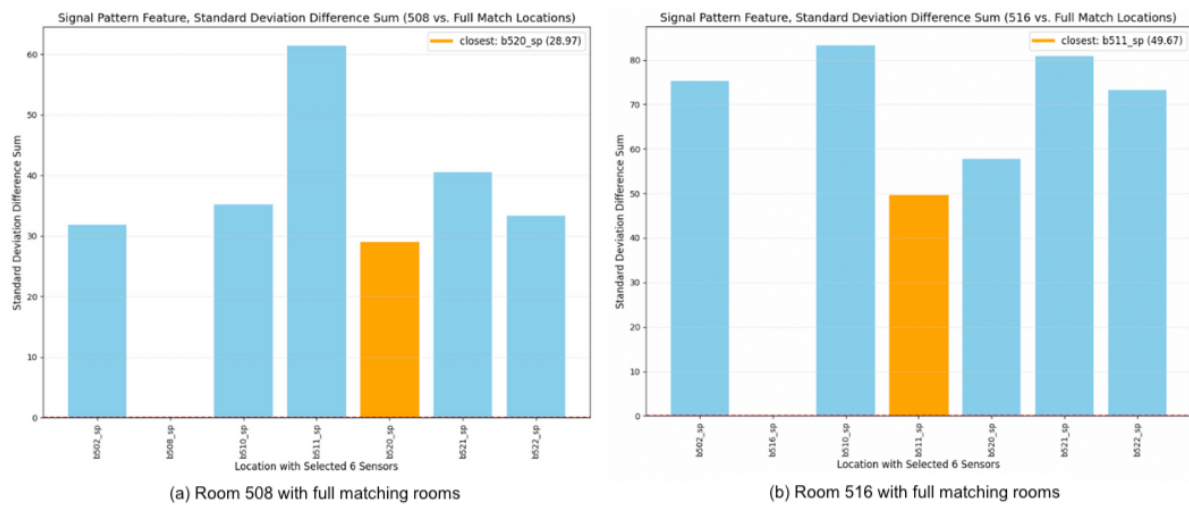


Figure 11. Full matching with signal pattern feature based on standard deviation. (a) Room 508 vs. other locations. (b) Room 516 vs. other locations.

Figure 12 reveals that, under partial matching with signal pattern analysis based on standard deviation, Room 520 remains the closest match to Room 508, while Room 518 is identified as the nearest match to Room 516.

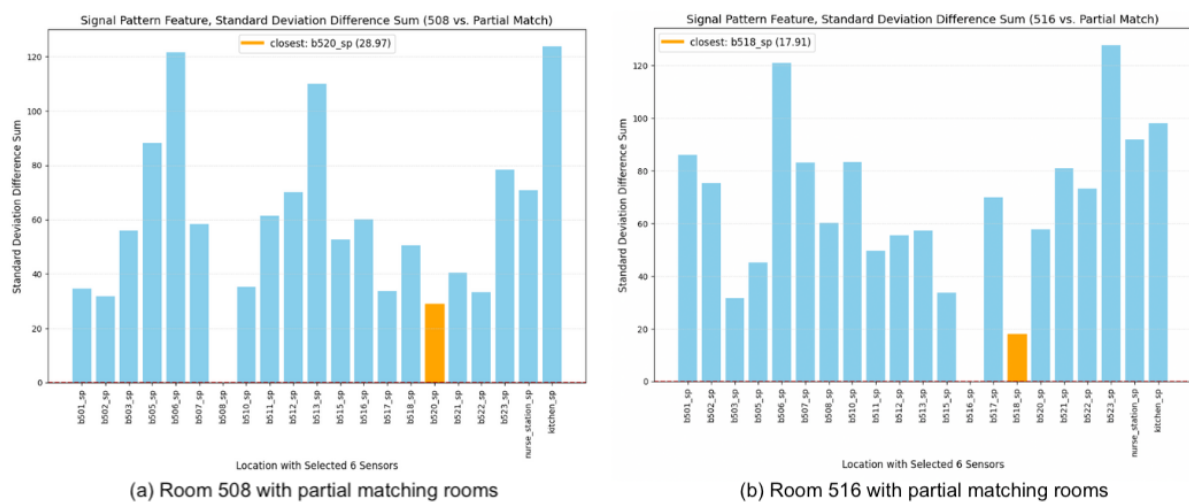


Figure 12. Partial matching with signal pattern feature based on standard deviation. (a) Room 508 vs. partial matching locations. (b) Room 516 vs. partial matching locations.

Continuing with the assessment of the signal pattern using KL divergence, Figure 13 shows the result of full matching with KL divergence where we identify 522 as the majority class nearest match to 508. At the same time, 520 is the match for 516.

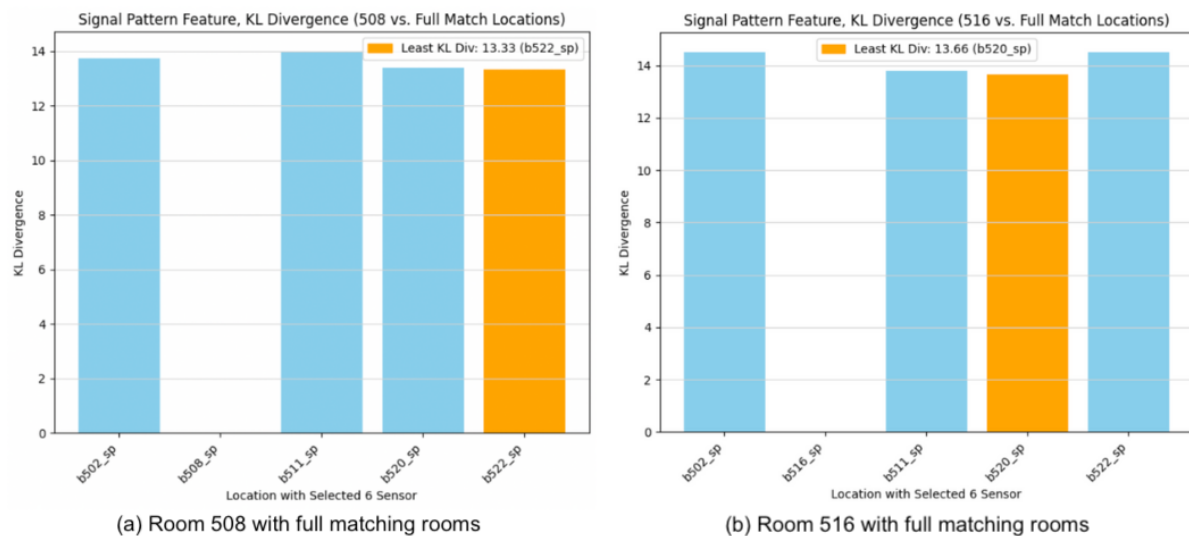


Figure 13. Full matching with signal pattern feature based on KL divergence. (a) Room 508 vs. other locations. (b) Room 516 vs. other locations.

Lastly, partial matching based on KL divergence results in 512 as the nearest match of 508, while 503 is identified as the top match for 516 as reflected in Figure 14.

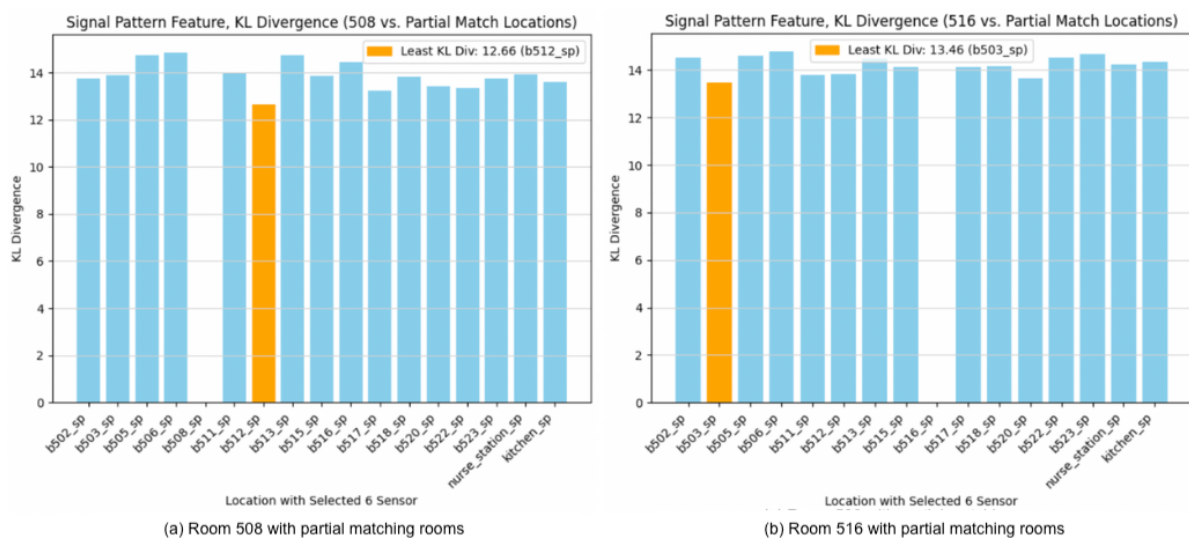


Figure 14. Signal pattern feature based on KL divergence with partial matching. (a) Room 508 vs. partial matching locations. (b) Room 516 vs. partial matching locations.

With the matching location identified for each minority class in all four cases, we proceed with relabeling the identified match for each minority location. We concatenate the augmented data to the original training oversampling Rooms 508 and 516. Indoor localization with the Random Forest model is executed using the new training data. Similarly, we apply data augmentation using Random Sampling, ADASYN, and SMOTE to both Room 508 and 516, generating the same length of augmented data as that of the relabeled matches for comparison.

4.3. Results

In our evaluation of indoor localization, we assessed the effectiveness of various combinations of relabeling approaches and statistical measures applied to signal patterns. This assessment aimed to determine the specific combination that yields the most optimal performance in terms of minority class localization.

Comparing the performance of relabeling using full matching based on standard deviation in Table 2, only the proposed relabeling method classified Room 516 after over-sampling. Relabeling achieved the highest target class F1-score for both 508 and 516. In the overall model F1-score, relabeling improved model performance by 6%. ADASYN achieved the highest value, followed by the proposed method.

Table 2. Performance of Relabeling with Standard Deviation, Full Matching versus other Sampling.

Oversampling Approach	Target Class Precision	Target Class Recall	Target Class F1-Score	Overall Model Weighted F1-Score
Baseline *	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.25 Room 516 = 0.00	Room 508 = 0.33 Room 516 = 0.00	0.60
Random Sampling	Room 508 = 0.67 Room 516 = 0.00	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.57 Room 516 = 0.00	0.64
SMOTE	Room 508 = 1.00 Room 516 = 0.00	Room 508 = 0.25 Room 516 = 0.00	Room 508 = 0.40 Room 516 = 0.00	0.63
ADASYN	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.50 Room 516 = 0.00	0.69
Proposed Relabeling	Room 508 = 0.57 Room 516 = 1.00	Room 508 = 1.00 Room 516 = 1.00	Room 508 = 0.73 Room 516 = 1.00	0.66

* Original train data with no augmentation applied.

The performance of relabeling with partial matching based on standard deviation is summarized in Table 3. In target class F1-score, no oversampling method was able to classify Room 516. At the same time, all data augmentation approaches improved the performance for 508, with SMOTE as the highest, followed by relabeling and ADASYN. Applying Random Sampling to the minority class resulted in a lower F1-score. Regarding the overall weighted F1-score, only ADASYN and SMOTE improved the model by 2 to 3%. As for the relabeling approach, the observed decrease in performance can be attributed to incomplete beacon data resulting from partial matching and the possibility of overlapping classes.

Table 3. Performance of Relabeling with Standard Deviation, Partial Matching versus other Sampling.

Oversampling Approach	Target Class Precision	Target Class Recall	Target Class F1-Score	Overall Model Weighted F1-Score
Baseline *	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.25 Room 516 = 0.00	Room 508 = 0.33 Room 516 = 0.00	0.60
Random Sampling	Room 508 = 0.67 Room 516 = 0.00	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.57 Room 516 = 0.00	0.60
SMOTE	Room 508 = 1.00 Room 516 = 0.00	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.67 Room 516 = 0.00	0.63
ADASYN	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.75 Room 516 = 0.00	Room 508 = 0.60 Room 516 = 0.00	0.62
Proposed Relabeling	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.75 Room 516 = 0.00	Room 508 = 0.60 Room 516 = 0.00	0.59

* Original train data with no augmentation applied.

On the other hand, performance comparison with full matching based on the KL divergence is depicted in Table 4. With this relabeling approach, only the proposed method was able to classify Room 516 after performing data augmentation. Relabeling achieved the highest target class F1-score for both Room 508 and 516. In the overall model F1-score, relabeling improved the model performance by 8% and achieved the highest result.

Table 4. Performance of Relabeling with KL Divergence, Full Matching versus other Oversampling.

Oversampling Approach	Target Class Precision	Target Class Recall	Target Class F1-Score	Overall Model Weighted F1-Score
Baseline *	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.25 Room 516 = 0.00	Room 508 = 0.33 Room 516 = 0.00	0.60
Random Sampling	Room 508 = 0.67 Room 516 = 0.00	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.57 Room 516 = 0.00	0.66
SMOTE	Room 508 = 0.67 Room 516 = 0.00	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.57 Room 516 = 0.00	0.66
ADASYN	Room 508 = 0.60 Room 516 = 0.00	Room 508 = 0.75 Room 516 = 0.00	Room 508 = 0.67 Room 516 = 0.00	0.66
Proposed Relabeling	Room 508 = 0.57 Room 516 = 1.00	Room 508 = 1.00 Room 516 = 1.00	Room 508 = 0.73 Room 516 = 1.00	0.68

* Original train data with no augmentation applied.

Lastly, Table 5 summarizes indoor positioning comparison with partial matching based on KL divergence. Similarly, in target class F1-score, no oversampling method was able to classify Room 516, while relabeling achieved the highest performance for Room 508 at 73%, which is 16% to 29% higher than that of other oversampling methods. In the overall model F1-score, only relabeling resulted in an improved model performance at 62%.

Table 5. Performance of Relabeling with KL Divergence, Partial Matching versus other Oversampling.

Oversampling Approach	Target Class Precision	Target Class Recall	Target Class F1-Score	Overall Model Weighted F1-Score
Baseline *	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.25 Room 516 = 0.00	Room 508 = 0.33 Room 516 = 0.00	0.60
Random Sampling	Room 508 = 0.67 Room 516 = 0.00	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.57 Room 516 = 0.07	0.60
SMOTE	Room 508 = 0.67 Room 516 = 0.00	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.57 Room 516 = 0.00	0.58
ADASYN	Room 508 = 0.40 Room 516 = 0.00	Room 508 = 0.50 Room 516 = 0.00	Room 508 = 0.44 Room 516 = 0.00	0.57
Proposed Relabeling	Room 508 = 0.57 Room 516 = 0.00	Room 508 = 1.00 Room 516 = 0.00	Room 508 = 0.73 Room 516 = 0.00	0.62

* Original train data with no augmentation applied.

5. Discussion

In this section, we outline the findings of our study, building upon the methodologies and approaches discussed previously. We highlight the contribution to the field of indoor localization, particularly in the context of nursing care facilities.

To investigate the impact of relabeling on indoor localization in nursing facilities, we illustrate the performance of the proposed method in Figure 15. In response to the first research question on identifying matching signal patterns between different locations in the facility, we calculate signal pattern features using standard deviation to compare per beacon variability and KL divergence to compare signal distributions between rooms. To better

understand signal patterns, we filter the analysis of beacons to six surrounding devices in the neighboring rooms.

In response to the second research question on how we use samples from other beacons to augment location with fewer data, we apply relabeling to the resulting match from the comparison of signal patterns between rooms. We implement two variations to relabeling based on the completeness of six beacons filtered around the surrounding location. We summarize the resulting performance of applying relabeling to the stationary beacons installed in the nursing facility in Figures 15 and 16.

Between the variations of relabeling, full matching consistently achieves better performance compared to baseline and partial matching in terms of the target class F1-score. As reflected in Figure 15, Room 516 is only classified after employing relabeling with full matching. Matching based on signal pattern features referenced from KL divergence results in a better F1-score than standard deviation. Overall, the proposed method improves the target class F1-score compared to the baseline by 27% to 40%.

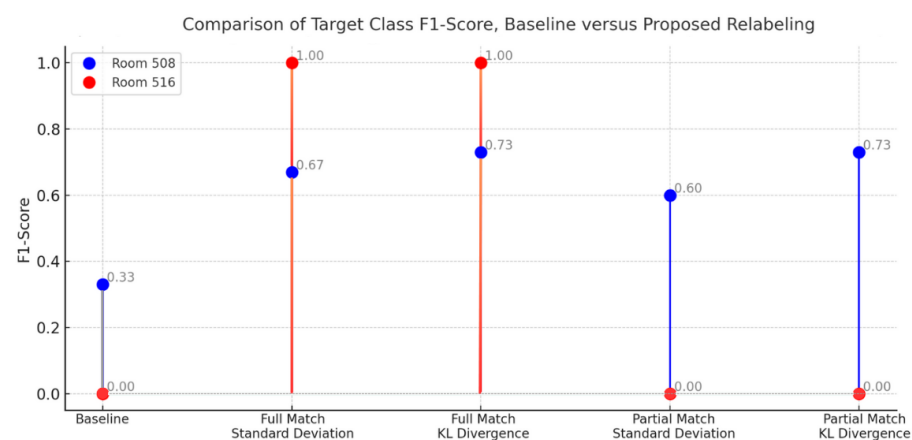


Figure 15. Comparison of Target Class F1-Score, Baseline versus Proposed Relabeling.

By comparing the overall F1-score performance in Figure 16, we determined that full matching achieves better performance than baseline with an increase of 6% to 8%. Moreover, full matching outperforms partial matching by 6% to 7%. For both full and partial matching, KL divergence results in a better F1-score than standard deviation. Overall, the proposed method improves model performance compared to the baseline, except in the case of partial matching with standard deviation. In general, increasing the training data with relabeling based on signal pattern results in improved localization.

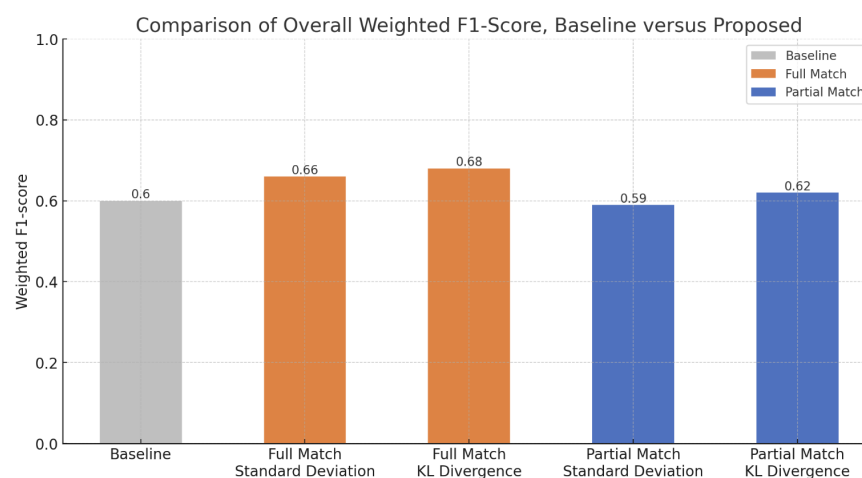


Figure 16. Comparison of Overall Weighted F1-Score, Baseline versus Proposed Relabeling.

Choosing the best combination of relabeling from the matching variation and signal pattern feature, we compare full matching based on KL divergence with the baseline confusion matrix as depicted in Figure 17. For target class F1-scores, relabeling improves the performance of both Room 508 and 516. All samples of 508 and 516 are correctly localized with relabeling. In the overall model F1-score, relabeling improves performance by 8%. On the other hand, a trade-off of the proposed method is potential class overlap, especially from the majority class used in the matching, which can be observed in Room 520 in the confusion matrix affecting the recognition of the room.

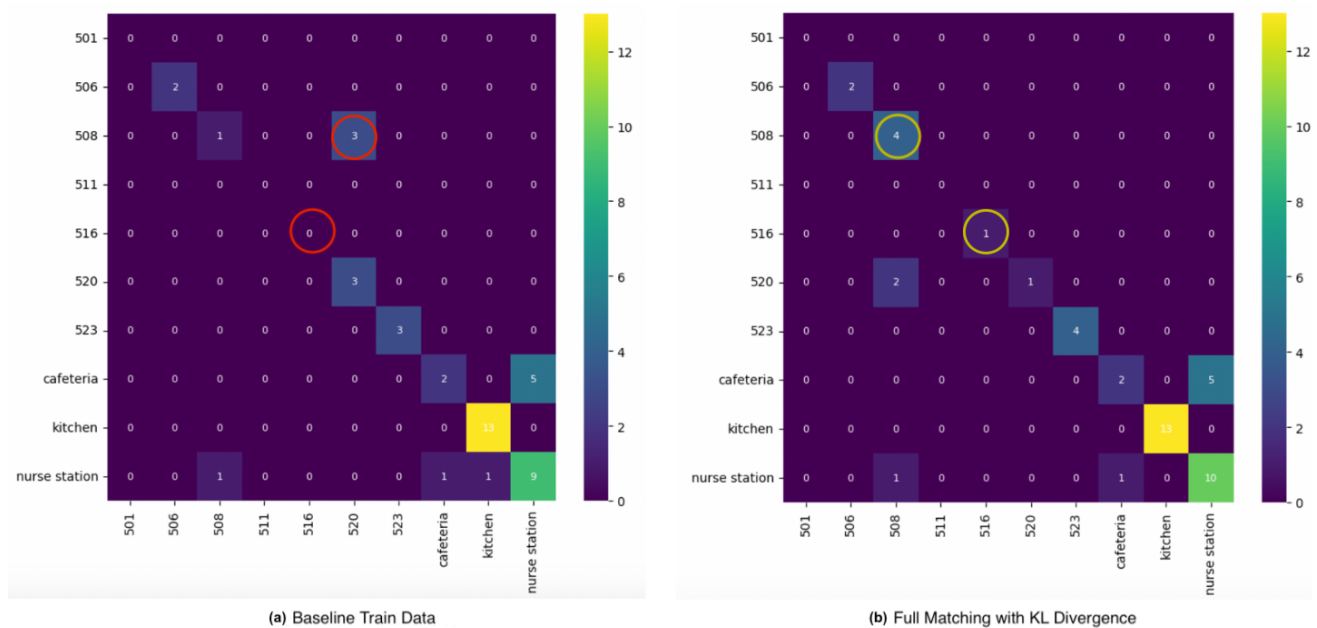


Figure 17. Comparison of indoor positioning with baseline data. (a) Confusion matrix, original data. (b) Confusion matrix, with relabeled data.

For cross-validation, we test the proposed method by applying data augmentation on other rooms using the same test data. In the original train data, we reduce Rooms 520 and 523 samples and then re-evaluate the performance for baseline before applying augmentation. Table 6 summarizes the result after oversampling Room 520. Relabeling improves both overall and target F1-score of baseline after applying data augmentation and increasing the samples by 96.7%.

Table 6. Comparison of Indoor Localization Performance, Oversampling Room 520.

Oversampling Approach	Train Data Room 520	Room 520 F1-Score	Overall Weighted F1-Score
Baseline	1000	0.00	0.56
Random Sampling	1969	0.50	0.67
SMOTE	1969	0.40	0.67
ADASYN	1969	0.86	0.67
Proposed method	1969	0.40	0.63

For Room 523, the proposed relabeling achieves both the highest overall and target F1-score after applying data augmentation to the room, increasing the samples five-fold the original size as depicted in Table 7.

Given the varying quantities of samples added through oversampling, evaluating and contrasting the augmented data produced by each technique is crucial, as they each function distinctly.

Presently, we remain positive that the application of the proposed method to other rooms relevant insights into the current effectiveness of relabeling. Between the variations in relabeling, full matching effectively utilizes all sensor data from other beacons. Conversely, while partial matching demonstrates some improvement in localization accuracy, its reduced retention of beacon data from other rooms can lead to diminished performance. Overall, as relabeling is based on leveraging samples from other beacons, the proposed data augmentation approach relies on the number of samples in the majority class. In line with other oversampling techniques, the volume of data gathered within the facility for training significantly influences the outcome of data augmentation and the performance of indoor localization.

Table 7. Comparison of Indoor Localization Performance, Oversampling Room 523.

Oversampling Approach	Train Data Room 523	Room 523 F1-Score	Overall Weighted F1-Score
Baseline	2178	0.00	0.53
Random Sampling	11,174	0.00	0.53
SMOTE	11,174	0.00	0.53
ADASYN	11,174	0.33	0.59
Proposed method	11,174	0.33	0.61

6. Conclusions

In this paper, we proposed a relabeling method for data augmentation towards indoor localization in nursing care facilities. Our proposed method addresses the challenge of data imbalance due to unequal representation of locations with low samples in minority classes. By filtering BLE devices to surrounding beacons, we calculated the signal pattern feature of minority and majority classes and identified matching rooms for relabeling. By applying a relabeling approach to identified locations matching the signal patterns of the minority class, indoor localization improved both in the target class and overall model performance. Moreover, we presented a comparative performance of indoor localization using different augmentation techniques and confirmed that our proposed method achieves better indoor localization in nursing homes.

The main contribution of this work is utilizing real data from other beacons referenced from the majority class to augment locations with fewer samples. With our proposed full and partial matching based on calculated signal pattern features, we showed the advantage of having a flexible method where the statistical measure can be varied in accordance with the targeted analyzed data. Moreover, our method displayed model generalization leveraging majority class data. On the other hand, dependence on the number of samples of majority classes should be taken into account when applying relabeling.

We could discern similarities between different rooms by calculating signal pattern features. This enabled us to effectively augment locations with limited data samples by repurposing data from matching rooms. By measuring both KL divergence and standard deviation across classes, we mapped the signal patterns among various locations. The approach based on KL divergence consistently outperforms baseline data and other methods, specifically Random Sampling, SMOTE, and ADASYN. However, it is important to acknowledge the potential for class overlap with the matching class. The augmented data should be carefully examined.

The reliance of the proposed method on the volume of majority class samples and the potential for class overlap necessitates careful application and consideration. To this end, our future work will focus on expanding data collection to enable more robust evaluations. Broadening the scope of data collection to various nursing homes with different layouts is a key component to validate the effectiveness of relabeling in diverse settings. Furthermore, locations within common areas not following specific layouts will be further examined. We aim to refine our relabeling approach to encompass areas within these common spaces with varying geometries and dimensions, such as cafeterias, where nursing activities

tend to occur over extended periods. Applying the relabeling approach to other use-case scenarios using different sensors, such as an accelerometer and gyroscope, will be analyzed to understand signal patterns for matching. The aim is to apply the proposed relabeling to other sensor data to assess the performance in improving models after data augmentation.

Currently, the analysis is performed in the offline phase using real-world data. Expanding the system to a more dynamic real-time relabeling and indoor positioning for nursing homes is suggested to adapt to immediate changes in the caregiving environment. Furthermore, elaborating analysis of signal patterns by incorporating additional statistical analysis techniques is suggested to represent signal pattern features of different devices aside from BLE beacons. A hybrid augmentation approach by combining the proposed oversampling method with SMOTE or ADASYN is suggested to resolve the limitations of respective oversampling techniques and further improve indoor localization. Algorithm optimization should be explored to enhance the accuracy and robustness of indoor localization. We plan to investigate the use of additional machine learning algorithms and the implementation of filtering techniques to improve signal quality, potentially increasing the precision of our localization system. On the other hand, the proposed relabeling maintains its effectiveness even with filtered or high-quality signal inputs.

Lastly, as we continue to improve the FonLog application and indoor positioning system, we will integrate user feedback collection via an in-app survey to incorporate suggestions in the system design cycle. This is to ensure adherence to the protocol of the partner facility and adapt to user experience with practical challenges faced in real-world scenarios.

Author Contributions: Conceptualization, C.G. and S.I.; methodology, C.G.; software, C.G. and S.I.; validation, C.G. and S.I.; formal analysis, C.G.; investigation, C.G.; resources, C.G.; data curation, C.G.; writing—original draft preparation, C.G.; writing—review and editing, C.G. and S.I.; visualization, C.G.; supervision, S.I.; project administration, C.G.; funding acquisition, S.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the JST-Mirai Program, Creation of Care Weather Forecasting Services in the Nursing and Medical Field, Grant Number JP21473170, and Health Labour Sciences Research, Grant Number 19AC1003 in Japan.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We extend our heartfelt gratitude to the managers of Kaikyokan Facility for their invaluable collaboration in securing approval for data collection on site and for assisting with the installation of beacons. This research was made possible with the participating caregiver department, allowing for the staff to carry mobile devices installed with Fonlog during daily nursing routines. The authors extend their gratitude to the participating graduate students from Maekawa Laboratory of Osaka University for their assistance during data collection. Lastly, we thank the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ADASYN	Adaptive Synthetic Sampling
AoA	Angle of Arrival
BLE	Bluetooth Low Energy
HAR	Human Activity Recognition
LOS	Line of Sight

IMU	Inertial Measurement Unit
IoT	Internet of Things
IPS	Indoor Positioning System
KL	Kullback–Leibler
MAC	Media Access Control
NLOS	Non-line of Sight
RF	Random Forest
RFID	Radio Frequency Identification
RS	Random Sampling
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
SMOTE	Synthetic Minority Oversampling Technique
ToA	Time of Arrival
Wi-Fi	Wireless Fidelity
WLAN	Wireless Local Area Network

References

1. López, M.P.; Tsitouras, D.J. Flatlining: How the Reluctance to Embrace Immigrant Nurses is Mortally Wounding the U.S. Healthcare System. *J. Health Care Law Policy* **2009**, *12*, 235. [\[CrossRef\]](#)
2. Garcia, C.; Quynh, V.N.P.; Kaneko, H.; Inoue, S. A Relabeling Approach to Signal Patterns for Beacon-based Indoor Localization in Nursing Care Facility. In Proceedings of the 5th International Conference on Activity and Behavior Computing, Kaiserslautern, Germany, 7–9 September 2023.
3. Morita, T.; Taki, K.; Fujimoto, M.; Suwa, H.; Arakawa, Y.; Yasumoto, K. Beacon-Based Time-Spatial Recognition toward Automatic Daily Care Reporting for Nursing Homes. *J. Sens.* **2018**, *2018*, 2625195. [\[CrossRef\]](#)
4. Garcia, C.; Inoue, S. Challenges and Opportunities of Activity Recognition in Clinical Pathways. In Proceedings of the 4th International Conference on Activity and Behavior Computing, London, UK, 27–29 October 2022.
5. Calderoni, L.; Ferrara, M.; Franco, A.; Maio, D. Indoor localization in a hospital environment using random forest classifiers. *Expert. Syst. Appl.* **2015**, *42*, 125–134. [\[CrossRef\]](#)
6. Nguyen, Q.H.; Johnson, P.; Nguyen, T.T.; Randles, M. A novel architecture using iBeacons for localization and tracking of people within healthcare environment. In Proceedings of the 2019 Global IoT Summit (GIoTS), Aarhus, Denmark, 17–21 June 2019; pp. 1–6. [\[CrossRef\]](#)
7. Kanan, R.; Elhassan, O. A combined batteryless radio and wifi indoor positioning for hospital nursing. *J. Commun. Softw. Syst.* **2016**, *12*, 33–44. [\[CrossRef\]](#)
8. Bibbò, L.; Carotenuto, R.; Della Corte, F. An Overview of Indoor Localization System for Human Activity Recognition (HAR) in Healthcare. *Sensors* **2022**, *22*, 8119. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Morita, T.; Taki, K.; Fujimoto, M.; Suwa, H.; Arakawa, Y.; Yasumoto, K. BLE Beacon-based Activity Monitoring System toward Automatic Generation of Daily Report. In Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Athens, Greece, 19–23 March 2018; pp. 788–793. [\[CrossRef\]](#)
10. Yoo, J.; Park, S. Fingerprint variation detection by unlabeled data for indoor localization. *Pervasive Mob. Comput.* **2020**, *67*, 101219. [\[CrossRef\]](#)
11. Palipana, S.; Pietropaoli, B.; Pesch, D. Recent advances in RF-based passive device-free localisation for indoor applications. *Ad. Hoc. Netw.* **2017**, *64*, 80–98. [\[CrossRef\]](#)
12. Zhu, Q.; Xiong, Q.; Wang, K.; Lu, W.; Liu, T. Accurate WiFi-based indoor localization by using fuzzy classifier and mlps ensemble in complex environment. *J. Franklin Inst.* **2020**, *57*, 1420–1436. [\[CrossRef\]](#)
13. Wang, J.; Tian, Z.; Yang, X.; Zhou, M. TWPalo: Through-the-wall passive localization of moving human with Wi-Fi. *Comput. Commun.* **2020**, *157*, 284–297. [\[CrossRef\]](#)
14. Oussalah, M.; Alakhras, M.; Hussein, M.I. Multivariable fuzzy inference system for fingerprinting indoor localization. *Fuzzy Sets Syst.* **2015**, *8*, 65–89. [\[CrossRef\]](#)
15. Berrueta, B.; Baalaa, O.; Caminadab, A.; Guillet, V. An evaluation method of channel state information fingerprinting for single gateway indoor localization. *J. Netw. Comput. Appl.* **2020**, *159*, 102591. [\[CrossRef\]](#)
16. Fei, H.; Xiao, F.; Huang, H.; Sun, L. Indoor static localization based on Fresnel zones model using COTS Wi-Fi. *J. Netw. Comput. Appl.* **2020**, *167*, 102709. [\[CrossRef\]](#)
17. Xu, B.; Gang, W. Random sampling algorithm in RFID indoor location system. In Proceedings of the Third IEEE International Workshop on Electronic Design, Test and Applications (DELTA'06), Kuala Lumpur, Malaysia, 17–19 January 2006. [\[CrossRef\]](#)
18. Ni, L.; Liu, Y.; Lau, Y.C.; Patil, A. LANDMARC: Indoor Location Sensing Using Active RFID. *Wirel. Netw.* **2004**, *10*, 701–710. [\[CrossRef\]](#)

19. Cortesi, S.; Vogt, C.; Magno, M. Comparison between an RSSI- and an MCPD-Based BLE Indoor Localization System. *J. Comput.* **2023**, *12*, 59. [\[CrossRef\]](#)
20. Danış, F.S.; Cemgil, A.T.; Ersoy, C. Adaptive Sequential Monte Carlo Filter for Indoor Positioning and Tracking with Bluetooth Low Energy Beacons. *IEEE Access* **2021**, *9*, 37022–37038. [\[CrossRef\]](#)
21. Zhou, C.; Yuan, J.; Liu, H.; Qiu, J. Bluetooth Indoor Positioning Based on RSSI and Kalman Filter. *Wirel. Pers. Commun.* **2017**, *96*, 4115–4130. [\[CrossRef\]](#)
22. Cao, H.; Wang, Y.; Bi, J.; Zhang, Y.; Yao, G.; Feng, Y.; Si, M. LOS compensation and trusted NLOS recognition assisted WiFi RTT indoor positioning algorithm. *Expert Syst. Appl.* **2024**, *243*, 122867. [\[CrossRef\]](#)
23. Huang, X.; Guo, S.; Wu, Y.; Yang, Y. A fine-grained indoor fingerprinting localization based on magnetic field strength and channel state information. *Pervasive Mob. Comput.* **2017**, *41*, 150–165. [\[CrossRef\]](#)
24. Tang, C.; Sun, W.; Zhang, X.; Zheng, J.; Sun, J.; Liu, C. A Sequential-Multi-Decision Scheme for WiFi Localization Using Vision-Based Refinement. *IEEE Trans Comput.* **2023**, 1–16. [\[CrossRef\]](#)
25. Blandeau, M.; Guichard, R.; Hubaut, R.; Leteneur, S. IMU positioning affects range of motion measurement during squat motion analysis. *J. Biomech.* **2023**, *153*, 111598. [\[CrossRef\]](#)
26. Kuo, Y.-H.; Wu, E.H.-K. Intelligent Geomagnetic Indoor Positioning System. *Electronics* **2023**, *12*, 2227. [\[CrossRef\]](#)
27. Yang, A.C.H.; Lau, N.; Ho, J.C.F. The Role of Bedroom Privacy in Social Interaction among Elderly Residents in Nursing Homes: An Exploratory Case Study of Hong Kong. *J. Sens.* **2020**, *15*, 4101. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Yaro, A.S.; Maly, F.; Prazak, P. Outlier Detection in Time-Series Receive Signal Strength Observation Using Z-Score Method with Sn Scale Estimator for Indoor Localization. *Appl. Sci.* **2023**, *13*, 3900. [\[CrossRef\]](#)
29. Yoo, J.; Kim, H.J. Target Tracking and Classification from Labeled and Unlabeled Data in Wireless Sensor Networks. *J. Sens.* **2014**, *14*, 23871–23884. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Hoang, M.T.; Yuen, B.; Dong, X.; Lu, T.; Westendorp, R.; Reddy, K. Recurrent Neural Networks for Accurate RSSI Indoor Localization. *IEEE Internet Things J.* **2019**, *6*, 10639–10651. [\[CrossRef\]](#)
31. Wang, S. Wireless Network Indoor Positioning Method Using Nonmetric Multidimensional Scaling and RSSI in the Internet of Things Environment. *Math. Probl. Eng.* **2020**, *2020*, 8830891. [\[CrossRef\]](#)
32. Kotanen, A.; Hannikainen, M.; Leppakoski, H.; Hamalainen, T.D. Experiments on local positioning with Bluetooth. In Proceedings of the International Conference on Information Technology: Coding and Computing, Las Vegas, NV, USA, 28–30 April 2003; pp. 297–303. [\[CrossRef\]](#)
33. Yaro, A.S.; Maly, F.; Prazak, P. A Survey of the Performance-Limiting Factors of a 2-Dimensional RSS Fingerprinting-Based Indoor Wireless Localization System. *Sensors* **2023**, *23*, 2545. [\[CrossRef\]](#)
34. Liu, L.; Zhao, Q.; Miki, S.; Tokunaga, J.; Ebara, H. Indoor Fingerprinting Positioning System Using Deep Learning with Data Augmentation. *Sens. Mater.* **2022**, *34*, 3047–3061. [\[CrossRef\]](#)
35. He, H.; Bai, Y.; Garcia, E.A.; Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–8 June 2008; pp. 1322–1328. [\[CrossRef\]](#)
36. Chawla, N.; Bowyer, K.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *arXiv* **2002**, arXiv:1106.1813.
37. Fikry, M.; Garcia, C.; Quynh, V.N.P.; Inoue, S. Improving Complex Nurse Care Activity Recognition Using Barometric Pressure Sensor. In Proceedings of the 4th International Conference on Activity and Behavior Computing, London, UK, 27–29 October 2022.
38. Sulaiman, B.; Tarapiah, S.; Natsheh, E.; Atalla, S.; Mansoor, W.; Himeur, Y. Radio map generation approaches for an RSSI-based indoor positioning system. *Syst. Soft Comput.* **2023**, *5*, 200054. [\[CrossRef\]](#)
39. Mairittha, N.; Mairittha, T.; Inoue, S. A mobile app for nursing activity recognition. In Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, Singapore, 8–12 October 2018; pp. 400–403.
40. Luschi, A.; Villa, E.A.B.; Gherardelli, M.; Iadanza, E. Designing and Developing a Mobile Application for Indoor Real-time Positioning and Navigation in Healthcare Facilities. *J. Health Care Technol.* **2022**, *30*, 1371–1395. [\[CrossRef\]](#) [\[PubMed\]](#)
41. Wang, J.; Ghosh, R.; Das, S. A survey on sensor localization. *J. Control Theory Appl.* **2010**, *8*, 2–11. [\[CrossRef\]](#)
42. Using Beacons for Indoor Positioning, Tracking and Indoor Navigation. Available online: www.infsoft.com/basics/positioning-technologies/bluetooth-low-energy-beacons (accessed on 20 November 2023).
43. Locatify. Indoor Positioning Systems based on BLE Beacons. Available online: <https://locatify.com/blog/indoor-positioning-systems-ble-beacons> (accessed on 26 November 2023).
44. What Is BLE?—Bluetooth Low Energy Basics. Available online: <https://www.rfwireless-world.com/Terminology/what-is-BLE.html> (accessed on 24 November 2023).
45. Lee, W. Understanding and Using iBeacons. 2021. Available online: <https://www.codemag.com/article/1405051/Understanding-and-Using-iBeacons> (accessed on 24 November 2023).
46. Pichler, S. Indoor Positioning with Beacons. 2022. Available online: <https://www.esri.com/arcgis-blog/products/arcgis-ips/indoor-gis/indoor-positioning-with-beacons/> (accessed on 26 November 2023).
47. Wichmann, J. Indoor positioning systems in hospitals: A scoping review. *Digit. Health* **2022**, *8*, 20552076221081696. [\[CrossRef\]](#)

48. Singh, A.; Purohit, A. A Survey on Methods for Solving Data Imbalance Problem for Classification. *Int. J. Comput. Appl.* **2015**, *127*, 37–41. [\[CrossRef\]](#)
49. Jiang, X.; Ge, Z. Data Augmentation Classifier for Imbalanced Fault Classification. *IEEE Trans. Autom. Sci. Eng.* **2021**, *18*, 1206–1217. [\[CrossRef\]](#)
50. Fan, J.; Sun, C.; Chen, C.; Jiang, X.; Liu, X.; Zhao, X.; Meng, L.; Dai, C.; Chen, W. EEG data augmentation: Towards class imbalance problem in sleep staging tasks. *J. Neural Eng.* **2020**, *17*, 056017. [\[CrossRef\]](#)
51. Basiri, A.; Lohan, E.S.; Moore, T.; Winstanley, A.; Peltola, P.; Hill, C.; Amirian, P.; Silva, P.F. Indoor location based services challenges, requirements and usability of current solutions. *Comput. Sci. Rev.* **2017**, *24*, 1–12.2 [\[CrossRef\]](#)
52. Ciabattini, L.; Foresi, G.; Monteriù, A.; Pepa, L.; Pagnotta, D.P.; Spalazzi, L.; Verdini, F. Real time indoor localization integrating a model based pedestrian dead reckoning on smartphone and BLE beacons. *J. Ambient Intell. Human. Comput.* **2019**, *10*, 1–12. [\[CrossRef\]](#)
53. Yoon, J.; Oh, J.; Kim, S. Transfer Learning Approach for Indoor Localization with Small Datasets. *Remote Sens.* **2023**, *15*, 2122. [\[CrossRef\]](#)
54. Maharana, K.; Mondal, S.; Nemade, B. A review: Data pre-processing and data augmentation techniques. *Glob. Transit.* **2022**, *3*, 91–99. [\[CrossRef\]](#)
55. Um, T.T.; Pfister, F.M.J.; Pichler, D.; Endo, S.; Lang, M.; Hirche, S.; Fietzek, U.; Kulić, D. Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, UK, 13–17 November 2017; pp. 216–220. [\[CrossRef\]](#)
56. Temraz, M.; Keane, M.T. Solving the class imbalance problem using a counterfactual method for data augmentation. *Mach. Learn. Appl.* **2022**, *9*, 100375. [\[CrossRef\]](#)
57. Ling, C.X.; Li, C. Data mining for direct marketing: Problems and solutions. In Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 27–31 August 1998; pp. 73–79.
58. He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284. [\[CrossRef\]](#)
59. Thabtah, F.; Hammoud, S.; Kamalov, F.; Gonsalves, A. Data imbalance in classification: Experimental evaluation. *Inf. Sci.* **2020**, *513*, 429–441. [\[CrossRef\]](#)
60. Lee, J.; Park, K. GAN-based imbalanced data intrusion detection system. *Pers. Ubiquit Comput.* **2021**, *25*, 121–128. [\[CrossRef\]](#)
61. Brandt, J.; Emil Lanzén, E. *A Comparative Review of SMOTE and ADASYN in Imbalanced Data Classification*; Department of Statistics, Uppsala University: Uppsala, Sweden, 2020.
62. Hilal, A.; Arai, I.; El-Tawab, S. DataLoc+: A Data Augmentation Technique for Machine Learning in Room-Level Indoor Localization. In Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC), Nanjing, China, 29 March–1 April 2021; pp. 1–7. [\[CrossRef\]](#)
63. Yu, S.M.; Park, J.; Ko, S.-W. Combinatorial Data Augmentation for Real-Time Indoor Positioning: Concepts and Experiments. In Proceedings of the IEEE 95th Vehicular Technology Conference: (VTC2022-Spring), Helsinki, Finland, 19–22 June 2022; pp. 1–5. [\[CrossRef\]](#)
64. Liu, A.; Cheng, L.; Yu, C. SASMOTE: A Self-Attention Oversampling Method for Imbalanced CSI Fingerprints in Indoor Positioning Systems. *Sensors* **2022**, *22*, 5677. [\[CrossRef\]](#)
65. Kim, M.; Jeong, C.Y. Label-preserving data augmentation for mobile sensor data. *Multidimens. Syst. Signal Process.* **2021**, *32*, 115–129. [\[CrossRef\]](#)
66. Li, Y.; Adams, N.; Bellotti, T. A relabeling approach to handling the class imbalance problem for logistic regression. *J. Comput. Graph. Stat.* **2022**, *31*, 241–253. [\[CrossRef\]](#)
67. Yoo, J.; Park, J. Indoor Localization Based on Wi-Fi Received Signal Strength Indicators: Feature Extraction, Mobile Fingerprinting, and Trajectory Learning. *Appl. Sci.* **2019**, *9*, 3930. [\[CrossRef\]](#)
68. Bose, A.; Foh, C.H. A practical path loss model for indoor WiFi positioning enhancement. In Proceedings of the 6th International Conference on Information, Communications and Signal Processing, Singapore, 10–13 December 2007; pp. 1–5. [\[CrossRef\]](#)
69. Sahoo, S. Comparing RSSI between Thinly Enclosed and Exposed Devices. Available online: <https://ysjournal.com/computer-science/comparing-rssi-between-thinly-enclosed-and-exposed-devices/> (accessed on 26 November 2023).
70. Lee, P.Q.; Seah, W.; Tan, H.; Yao, Z. Wireless sensing without sensors—An experimental study of motion/intrusion detection using RF irregularity. *Meas. Sci. Technol.* **2010**, *21*, 124007. [\[CrossRef\]](#)
71. Yassin, A.; Nasser, Y.; Awad, M.; Al-Dubai, A.; Liu, R.; Yuen, C.; Raulefs, R.; Aboutanios, E. Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications. *IEEE Commun. Surv. Tutorials* **2017**, *19*, 1327–1346. [\[CrossRef\]](#)
72. Thakur, N.; Han, C.Y. Indoor Localization for Personalized Ambient Assisted Living of Multiple Users in Multi-Floor Smart Environments. *Big Data Cogn. Comput.* **2021**, *5*, 42. [\[CrossRef\]](#)
73. Brown, I.; Mues, C. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Syst. Appl.* **2012**, *39*, 3446–3453. [\[CrossRef\]](#)
74. Lee, S.; Kim, J.; Moon, N. Random forest and WiFi fingerprint-based indoor location recognition system using smart watch. *Hum. Cent. Comput. Inf. Sci.* **2019**, *9*, 6. [\[CrossRef\]](#)

75. Varma, P.S.; Anand, V. Random Forest Learning Based Indoor Localization as an IoT Service for Smart Buildings. *Wirel. Pers. Commun.* **2021**, *117*, 3209–3227. [[CrossRef](#)]
76. Ramadan, M.; Sark, V.; Gutierrez, J.; Grass, E. NLOS Identification for Indoor Localization using Random Forest Algorithm. In Proceedings of the 22nd International ITG Workshop on Smart Antennas, Bochum, Germany, 14–16 March 2018; pp. 1–5.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.