

Article

Not peer-reviewed version

Relabeling for Indoor Localization Using Stationary Beacons in Nursing Care Facilities

[Christina Alvarez Garcia](#)^{*} and [Sozo Inoue](#)^{*}

Posted Date: 14 December 2023

doi: 10.20944/preprints202312.0998.v1

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



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

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, Kitakyushu 808-0135, Japan; sozo@brain.kyutech.ac.jp (S.I.)

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

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 sensor data available for training. In this paper, we propose a data augmentation method to reuse the Received Signal Strength (RSS) from different sensors by relabeling to the sensors with less samples resolving data imbalance. Standard deviation and Kullback-Leibler divergence between minority and majority class are used to measure signal pattern to find matching sensors to relabel. By matching sensors between classes, two variations of relabeling are implemented specifically full and partial matching. The performance is evaluated using real world dataset we collected for five days in a nursing care facility installed with 25 BLE beacons. 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 sensor data with our proposed relabeling method for data augmentation, we achieved higher minority class F1-score as compared to augmentation with Random Sampling, Synthetic Minority Oversampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN). Our proposed method utilizes collected sensor data by leveraging majority class samples. Full matching demonstrated a 6 to 8% improvement from the original baseline overall weighted F1-score.

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

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 to a rise in patient-to-caregiver ratio [2,3]. Research efforts to better comprehend activity patterns during patient assistance is 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 [5,30]. Employing indoor positioning to monitor care routine and patient assistance is useful to support nursing record, and to optimize care response [7,8,51]. Several key positioning techniques exist but the standard approach includes trilateration, triangulation, multilateration, and fingerprinting [12]. However, these techniques are limited by environmental factors such as hardware requirements, setup complexity, signal obstruction and time synchronization.

A common challenge to indoor localization accuracy is low quality of signals. Outlier detection [20] and filtering methods are applied to resolve this issue. Moving average [21], weighted average [23], Bayesian sequential Monte Carlo [24], and Gaussian filtering [19] has been proposed for signal smoothing to achieve better positioning. Kalman filter [25,26] is a widely used method for filtering signals in IPS. In addition, machine learning [7,12,13,21,62] and deep learning [9,23,60] 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 real world setting impacts 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 employed to artificially increase the training dataset by creating modified version of the current data. To increase real data, Random Sampling (RS) [31] is commonly applied to minority class samples by duplication. On the other hand, creating synthetic samples is an alternative method. Adaptive Synthetic Sampling (ADASYN) [33] and Synthetic Minority Over-sampling Technique (SMOTE) [34] both generate synthetic data points for the minority class by interpolation based on nearest neighbors principle. Currently, existing oversampling techniques suffer from repeating data resulting to overfitting while synthetic sample is prone noise and data misrepresentation. Since augmented data is intended to reflect actual samples, oversampling should be carefully considered. At present, augmentation methods do not optimize real data from majority classes and few studies exist that investigate the use of signal pattern for augmenting training data in the context of indoor localization [2].

In this study, we propose a novel relabeling approach to address the challenge of unequally represented locations in beacon-based indoor localization. As more nursing homes are adopting IPS with IoT, this research aims to leverage sensor data collected in real-world environment to improve 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. The main objective of this research is to develop an augmentation method by increasing training samples with sensor data from different rooms to improve location detection of minority class. 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?
2. How can we use the samples from other sensors to augment location with fewer data?
3. What is the performance of oversampling based on signal pattern relabeling compared to existing augmentation methods?

To evaluate, we compared performance of different augmentation techniques applied to our collected real-world data from a nursing facility. We employed Random Sampling, ADASYN and SMOTE to the training set comprising 3.5 days of data and performed indoor localization using 1.5 days of test data. Overall and target class F1-score were measured to account for the data imbalance along with precision and recall. Our proposed method achieved an improved target class F1-score of 27% to 40% as compared to the baseline data. With full matching, the relabeling method demonstrated a 6 to 8% improvement from the original baseline overall weighted F1-score. For both full and partial matching, KL divergence resulted to better F1-score than standard deviation. This method effectively expands the training set, enhancing model accuracy, as demonstrated in a nursing care facility where beacon devices and mobile applications were 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 the proposed method, which encompasses signal pattern and relabeling approach while 4 covers the data collection process including the evaluation of SPRI in comparison to baseline and other augmentation strategies. A comprehensive analysis of the results follows in 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. IPS differ in technology, signal measurement and localization technique tailored to suit the environment and particular use-case requirements [8]. Radiofrequency-based technologies

[16] specifically Wi-Fi [9,10,13–15], Radio Frequency Identification (RFID) [31,32], and Bluetooth Low Energy (BLE) [22,24,25] are typically preferred considering flexibility to setup and integration with IoT. Signal measurements are generally based on time, angle and received signal strength (RSS) [11,18–21]. With the advent of Internet of Things (IoT), devices which easily connect to the network such as beacons, tags and mobile devices [2,6,18,27] are preferred in system design.

2.1. Indoor Localization with Beacons

Beacons are battery driven radio transmitters used in indoor positioning with proximity sensors that emits BLE signals [46,47]. Unlike the classic Bluetooth which is connection oriented, BLE has advertising functionality and don't necessarily have to pair [48,49]. 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 [50]. These devices are lightweight, and installation is straightforward, with simple mounting on walls.

BLE beacons are used to track nursing activities to get a better understanding of care routines which is crucial to optimize workload distribution considering low staff to patient ratio [52]. Existing systems combine beacons with Wireless Local Area Network (WLAN) for tracking both patients and nurses 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 routine, beacons are used for time-spatial recognition [5,51].

In some nursing facilities, multiple patients share a common room. Considering the effects of setting up indoor positioning on the privacy of elderly which can affect their social interaction, placement of beacons should be carefully considered in planning IPS [29]. Studies focused on nursing homes commonly use RSS and BLE for indoor localization [3,29,51] 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 [35–37].

Data augmentation is significant in improving indoor positioning systems. The technical and non-technical challenges of IPS in real world environments [54] have impacted the precision of indoor positioning with uncontrolled variables such as signal interference, physical obstructions, and breakdown of hardware. Data collection in on-site scenario [55] affects data quality and labeling resulting to imbalanced and unequal representation of classes in dataset [2,58]. In order to create robust and unbiased model for IPS, data augmentation shows up as a crucial and imperative approach to address imbalance. By applying several changes to the current dataset, data augmentation creates synthetic data that diversifies the training set and lessens the effects of class imbalance [38–40].

Three basic approaches dominate data-level solutions to the class imbalance issue: Random Sampling [41,42], Synthetic Minority Over-Sampling Technique (SMOTE) [40], and Adaptive Synthetic Sampling (ADASYN) [33]. By randomly adding additional copies of selected minority classes to the training data, Random Over-Sampling balances the class distribution [41]. On the other hand, instead of oversampling by replacement, ADASYN and SMOTE oversamples the minority class by producing synthetic instances [34]. 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 any new information. If the original data is already highly dimensional, this would increase 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 important data.

When training a model, applying SMOTE creates a linear mapping of the data which can lead to issues with overfitting. Also, there is a risk of overlap because SMOTE method does not account for the position of general data close to the uncommon class data [40,43–45]. Similarly with ADASYN, possible class overlap occurs in boundary areas as oversampling targets to resample between neighboring minority and majority class. In general, both synthetic approach are sensitive to noise and needs parameter tuning as both are dependent on data distribution [59].

A combination of augmentation techniques has been applied in indoor localization [56,57] including adding information to the reference dataset and deep learning-based approach [60,61], yet these involve computational complexities. At present, few studies have delved into using methods based on signal pattern 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 original class label [67]. Recently, relabeling is used to address class imbalance when using logistic regression by assigning new labels to classes with fewer instances [68]. Data augmentation leading to the loss of label information can potentially reduce the 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 signal are essential metrics providing valuable insights into the alterations of signal intensity that occur within indoor settings [63]. Trilateration and triangulation are traditional localization techniques that utilize geometric principles to assess signal behavior, measuring distance and angle from reference points.

In non-conventional approach, statistical techniques are optimized to handle the inherent variability and uncertainty in signal-based analysis. The standard deviation of the received signal reflects the fluctuations in the signal strength at different positions or distances within an indoor environment [64–66]. 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 [64]. KL divergence is a measure of difference between two probability density functions. In IPS, it is calculated to measure similarity in online phase of unlabeled data an existing database and up-to-date data [12].

The integration of multiple signal sources, sensor fusion techniques and machine learning provide more robust and reliable localization. Implementing adaptive algorithms that can learn and adjust to changes in the environment 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 care facility, we focus on leveraging the current data from stationary beacons by using the majority class to oversample locations with less sensor 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 were utilized as guide for 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 caregiving services are

delivered without interruption. To protect the privacy of patients, all device have been mounted outside the door of patient rooms.

3. Material and Methods

This section describes the processes of proposed relabeling with an overview of the indoor localization employed in the nursing facility depicted in Figure 1.

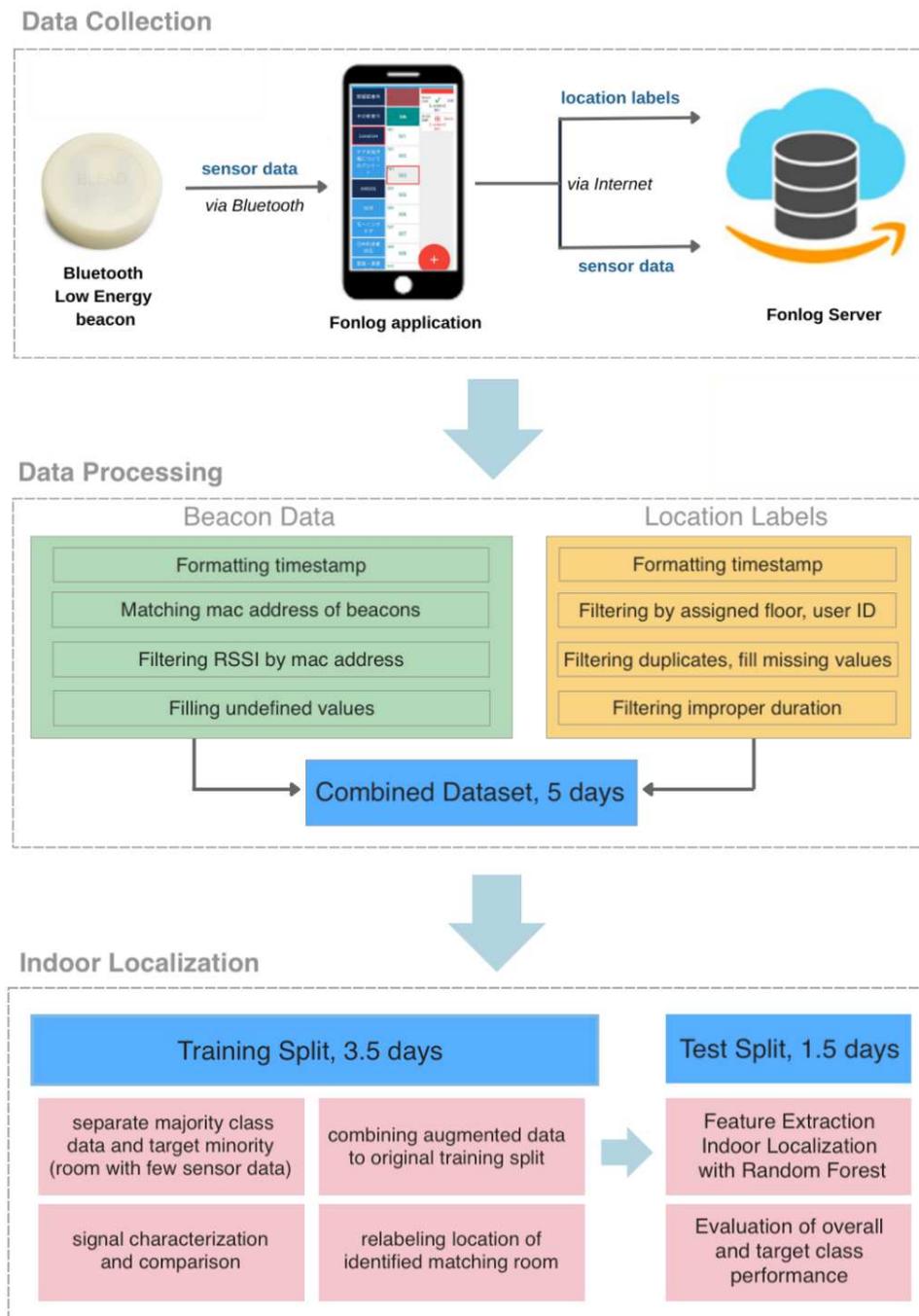


Figure 1. An overview of the proposed indoor positioning system in nursing care facility.

3.1. Signal Strength from Detected Beacons

In this section, we introduce our developed relabeling method depicted in Figure 2 for beacon data augmentation. In this approach, we utilize standard deviation and KL divergence to characterize

the signal pattern feature between locations to identify matching patterns to relabel. The subsection commences with information on the beacon data from site. To execute relabeling, understanding the signals detected from beacons to comprehend variability is necessary.

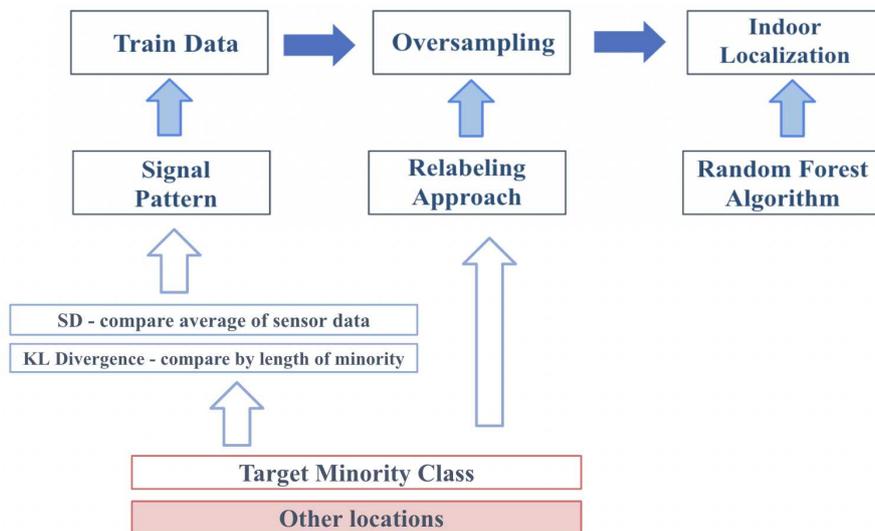


Figure 2. Proposed oversampling approach based on signal pattern relabeling for indoor localization.

Prior to deployment, the respective MAC addresses of the beacons were recorded to match with the raw sensor data from the server, as shown in Table 1. After pre-processing the raw data, the MAC addresses of detected devices were 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	2023-04-10T10:22:55.589+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 , along with their corresponding location label as in Equation 1. Overall, the signal database can be expressed as Equation 2 and the labels as Equation 3.

timestamp	mac address	RSSI	timestamp	RSSI_1	RSSI_2	RSSI_3	...
2023-04-10 13:00:00	6	-94	2023-04-10 13:00:00	0	0	0	...
2023-04-10 13:00:00	6	-96	2023-04-10 13:00:00	0	0	0	...
2023-04-10 13:00:00	4	-92	2023-04-10 13:00:00	0	0	0	...

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.

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

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

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

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

3.2. Matching

Generally, the accuracy of IPS decreases as the distance between the transmitter and receiver increases [69]. Observing this from the histogram of detected beacons and considering the setting of installed devices covering up to 5 meters 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 which are investigated for each room.

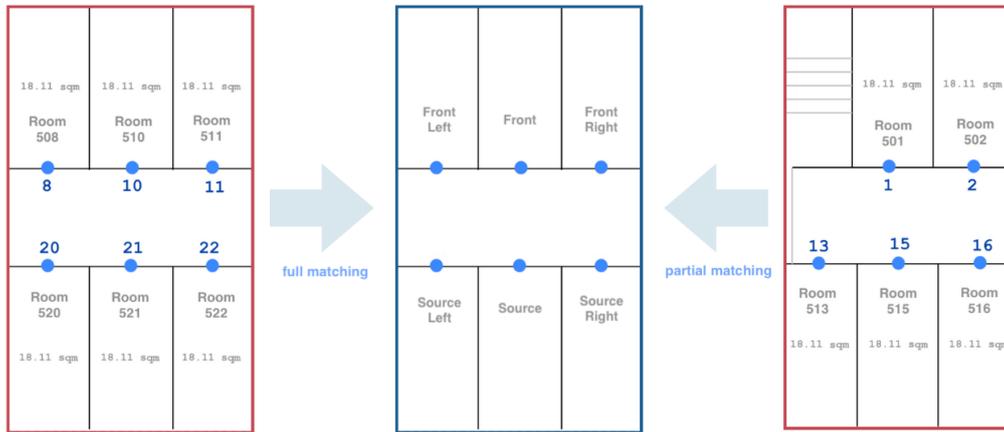


Figure 4. Full and Partial Matching based on selected sensors. (a) Full matching with complete 6 sensors. (b) 6 sensors surrounding the location. (c) Partial matching with incomplete 6 sensors.

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

$$f_m = [fl_m, sl_m, s_m, f_m, sr_m, fr_m] \quad (4)$$

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 incomplete sensors, null is filled with zero values to calculate the signal pattern feature.

3.3. Relabeling based on Signal Pattern

To define the signal variability in different rooms, a signal pattern feature is computed from the six sensors. 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. Sub-dataframe of minority class is denoted as $pattern_{min}$. From these sub-dataframes, standard deviation and KL divergence are measured to represent the signal pattern feature of each room which is then used to identify similar locations. The objective is to pinpoint the room that have matching signal pattern with 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 $pattern_{min}$.

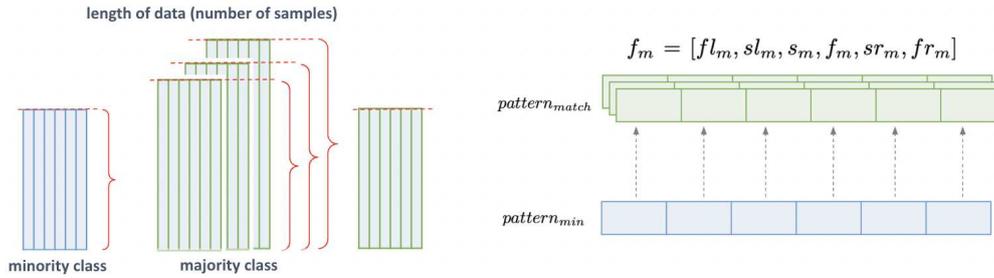


Figure 5. Comparing signal pattern feature between minority and majority class. (a) Length of samples considered. (b) Comparison between rooms per beacon.

Input:

$m \leftarrow$ room number

$d_{train} \leftarrow$ train data

Output:

$pattern_{match} \leftarrow pattern_m$

1. Identify rooms with small sample in train data, $df = d_{train}$.
2. Define $pattern_{min}$ and $pattern_m$ following Equation 4 such that for i in labels m : $d_i = df[df['m']] == "i"$
if $i == "1"$: $f_1 = [fl_1, sl_1, s_1, f_1, sr_1, fr_1]$
...
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 $pattern_{min}$ and $pattern_m$.
5. Compare the signal pattern features to identify $pattern_{match}$.

Signal pattern feature based on standard deviation calculates the absolute differences in standard deviation for each sensor between minority and majority classes. This approach considers individually the variability difference for each sensor, before summing them. The matching location from majority class is identified with the least standard deviation difference sum expressed in Equation 5.

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

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

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 $pattern_{\text{m}}$.

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 $pattern_{\text{m}}$ from all other rooms to find $pattern_{\text{match}}$. We apply both statistical measures to partial and full matching variation of relabeling forming four candidate match.

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 $d_{\text{relabelled}}$. Listed from 1 to 5 are the sequential steps followed to execute the relabeling process.

Input:

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

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. In the original floor plan depicted in Figure 7, the relabeling method is exclusively applied to patient rooms. In this study, relabeling is not applied for signals from beacons positioned in areas with diverse dimensions and layouts, such as open spaces in cafeteria. Moreover, it is crucial that the beacon sensors are installed in fixed positions.

3.4. Indoor Localization

In this paper, indoor localization is approached as a recognition problem. For room estimation considering attenuation, RSSI values from all rooms are considered in the feature extraction. Discrepancies between the timestamps of location labels and RSSI data are resolved by synchronizing data. To proceed with indoor localization, both statistical and temporal features are derived from the RSSI matrix. The Random Forest algorithm is employed for identifying locations as this classifier demonstrated effectiveness in handling imbalanced data [73] and preferred in indoor positioning [70–72].

Five statistical features were 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 effectiveness as observed in prior work. Time-based attributes specifically hour, minute, and microsecond were extracted. A non-overlapping time window of 45 seconds was implemented. The division of train and test data was date-based, adhering to train-test ratios recommended by empirical studies. 3.5 days of data comprised the training set and the remaining 1.5 days was intended for

test data. Both training and test set reflect the caregiving routine in the facility where nursing staff frequently visited the common area as compared to the patient rooms.

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

4. Data Collection and Evaluation

Data collection was taken from 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. The nursing care tasks executed by the caregiver fall under the category of patient care, medical assistance, documentation, cleaning and organization. Patient care cover 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. As majority of the caregiving activities are location-dependent, monitoring the rooms visited by nursing staff provides an estimation of the delivered workload and assistance given to elderly individuals within the floor.

4.1. Data Collection

FonLog, a smartphone application in Figure 6, is installed and used as data collection tool to log location labels and as receiver of RSSI signals [27]. In this study, FonLog was customized specifically to record activities performed in the respective floor of the nursing care facility with strategically positioned beacons.

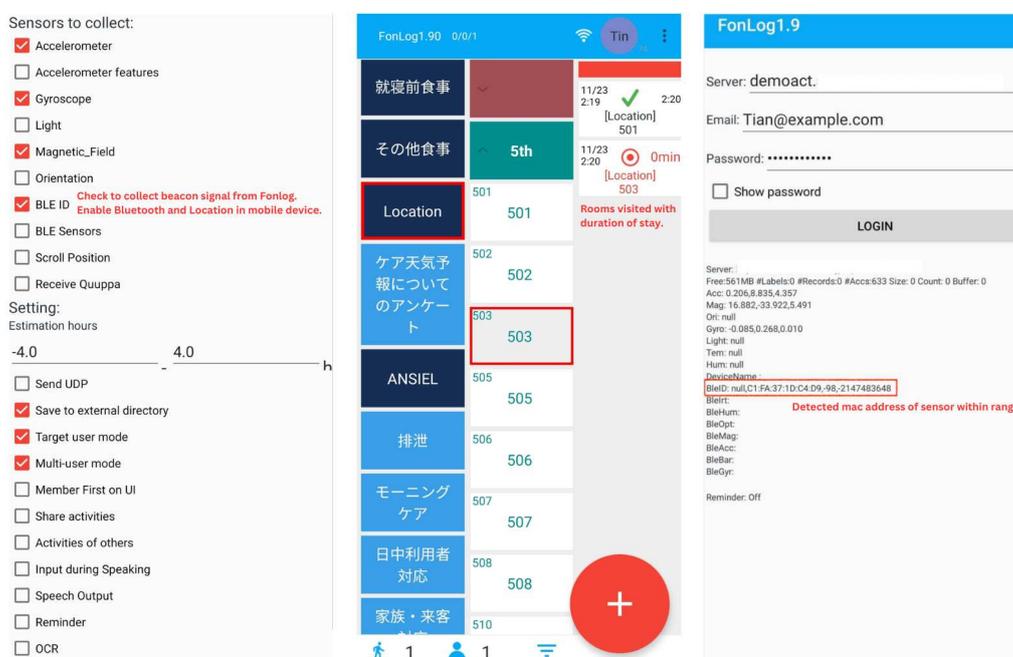


Figure 6. User interface of the FonLog application. (a) Sensor selection screen where BLE ID is enabled. (b) Activity recording screen showing respective rooms within the floor. (c) Log-in interface reflecting detected mac address of BLE device.

Two separate smartphones installed with Fonlog were setup for data gathering, one for the nursing staff with enabled BLE ID and another for the location labeler. Prior deployment, the location

and Bluetooth settings of the smartphone are set, and beacon detection is 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 perform 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 mismatch between beacon data and corresponding location label which occurs due to the delay in activity recording. Caregivers usually log the activity after completing various tasks resulting to discrepancy in timestamps during pre-processing. To mitigate this problem, the data collection process includes an observer who closely monitors the caregiver's daily routine. The labeler assigned is positioned in the hallway and records locations on the second smartphone with Fonlog choosing from the list of rooms in the app. Both carer and labeler are each assigned a unique user ID. Similarly, each patient room and area within the floor is allocated a unique customer ID. All beacon data and labels are temporarily stored in the local storage of each smartphone and uploaded to the same server once an internet connection is established. In this study, the data collection involved participating nurses of average height. Any significant height differences should be noted and impact on RSSI values should be observed in the case of multiple participants.

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 align the timezones of both the location and beacon data, and the timestamps are subsequently reformatted. Duplicates and rows with null values were removed and entries lacking in start-time and stop-time with undefined duration are filtered from the location labels.

Figure 7 depicts the map of the target floor in the nursing facility with installed beacons. Prior deployment, beacon frequency and height placement were decided based on experiments performed in the lab and hallway of the university reflecting the layout of the facility. In total, 25 stationary beacons indicated by blue points in the map were installed 2 meters 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 10hz, where the actual data showed RSSI detection ranging from 3 to 5hz.

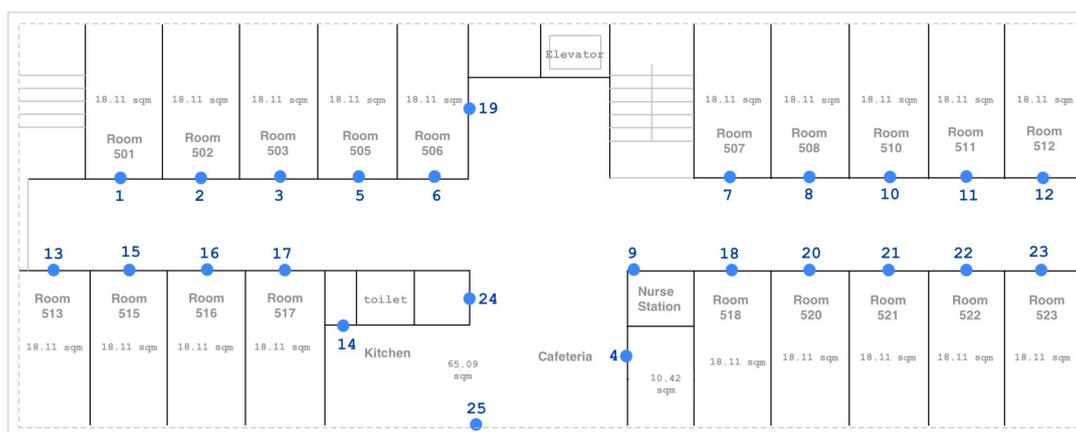


Figure 7. Layout of the facility showing the placement of installed beacons at each location.

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 [51]. This setup allows for position calculation using methods that do not rely on machine learning. For the purposes of this research, the placement of beacons was determined in agreement with the nursing home to effectively

track the movement of caregivers in 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 [51], beacons are placed in stationary positions as emitters while Fonlog installed in the smartphone of the caregiver served as the scanner.

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

4.2. Performance Evaluation

As this study aims to address data imbalance by relabeling approach, a comparative analysis of signal patterns, and assessment of the efficacy of partial and full matching of rooms is conducted. Furthermore, we evaluate the performance of relabeling against other data augmentation methods. The test data covered 1.5 days, while the training data encompassed 3.5 days, both segmented into 45-second windows without overlap.

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 both partial and full matching of rooms, we aim to refine the relabeling process and ensure the most accurate representation of location data. This comparison will also extend 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 were measured and obtained respectively using the following Equations in 7, 8, and 9 where TP is the True Positive, TN is the True Negative, FP is the False Positive, and FN the False Negative value.

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

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

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

We assessed 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, Room 516 and 507 were identified as minority classes. Filtering the sensors of all the rooms to six beacons following Equation 4 and Figure 4, we proceed with calculating signal pattern feature. Figure 8 illustrates the resulting standard deviation of each of six sensors in $pattern_{min}$ corresponding to the minority class rooms.

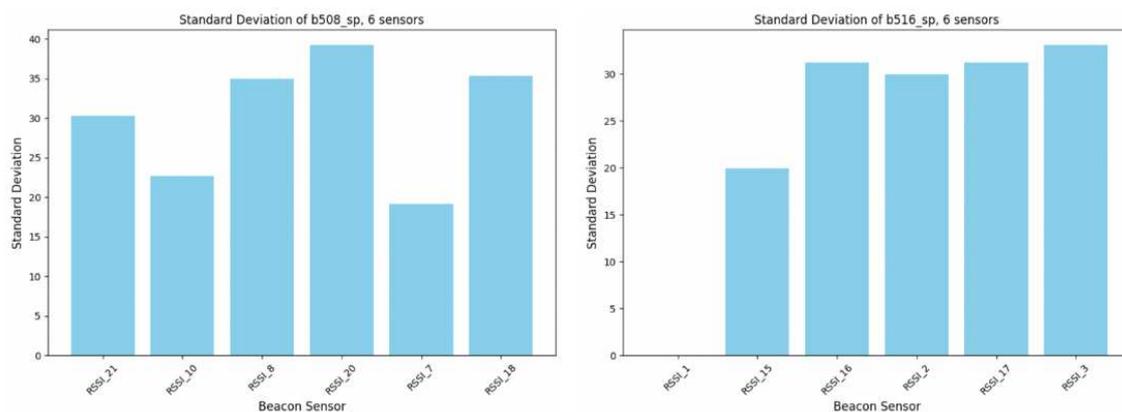


Figure 8. Standard deviation of selected sensors in target minority class and candidate math.

We then proceed with finding the match to 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 reserve as candidate match in partial matching while middle rooms with complete six beacons mounted in surrounding locations are considered in full matching.

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 implemented data augmentation to all four use-case scenario and compared with other oversampling methods. From this approach, we target to identify which statistical measure better represent signal patterns from six sensors and by comparing full and partial matching, we identify which locations to consider for relabeling.

In the case of full matching using signal patterns determined by standard deviation, as depicted in Figure 9, Room 520 emerges as the closest match to Room 508, and Room 511 is identified as the nearest match to Room 516.

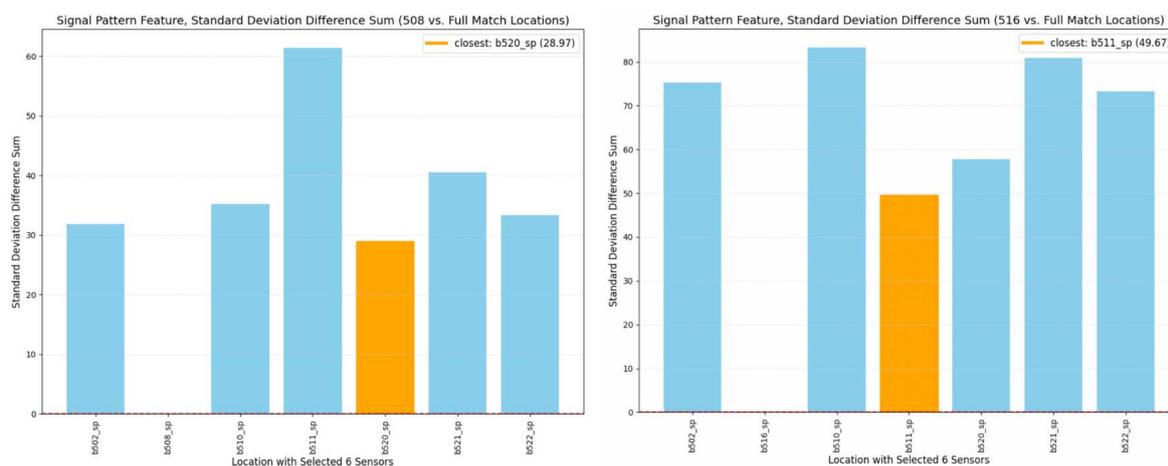


Figure 9. Full matching with signal pattern feature based on standard deviation. (a) Room 508 vs other locations. (b) Room 516 vs other locations.

Figure 10 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.

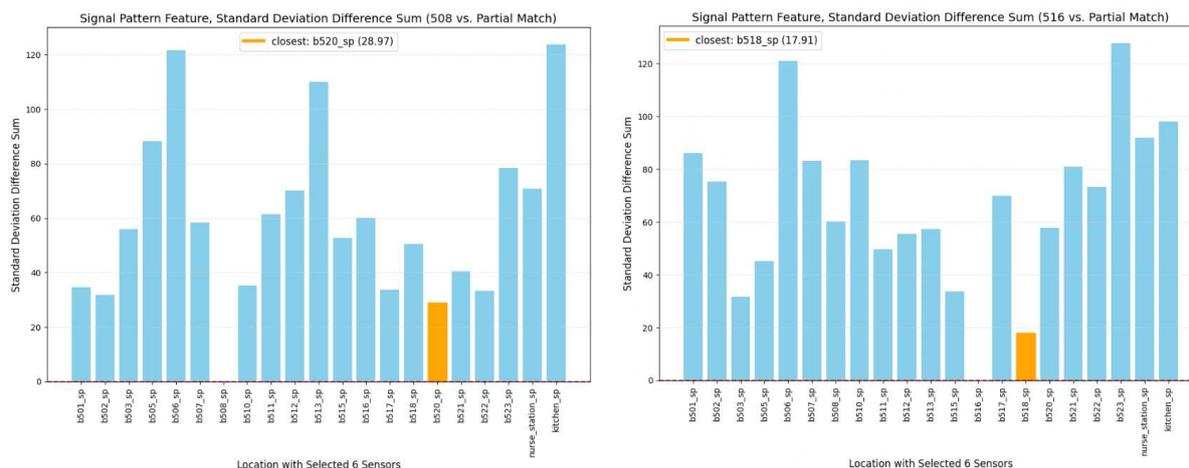


Figure 10. Partial matching with signal pattern feature based on standard deviation. (a) Room 508 vs partial matching locations. (b) Room 516 vs partial matching locations.

Continuing with assessment of signal pattern using KL divergence, Figure 11 shows the result of full matching with KL divergence where we identified 522 as the majority class nearest match to 508 while 520 is the match for 516.

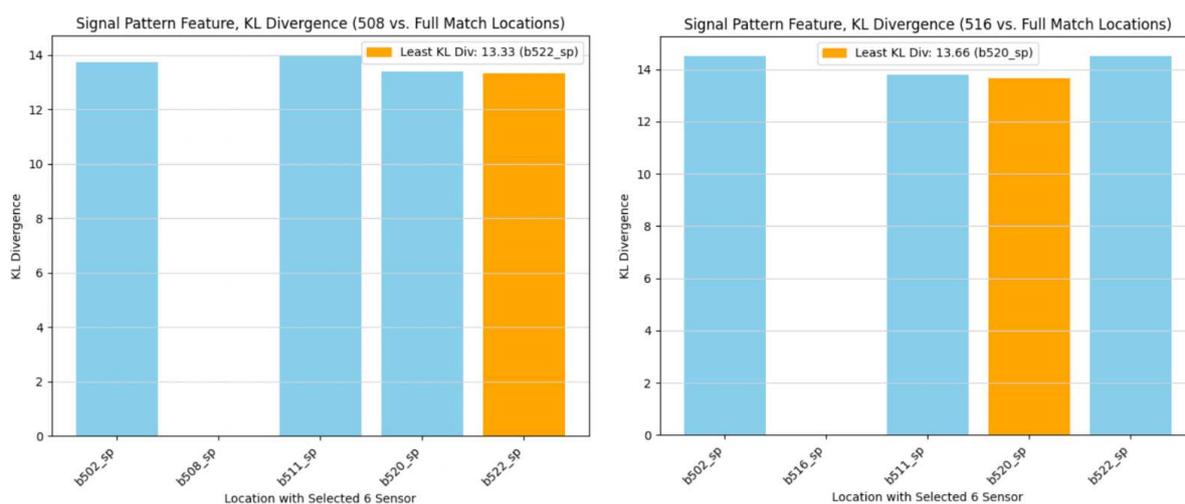


Figure 11. Full matching with signal pattern feature based on KL divergence. (a) Room 508 vs other locations. (b) Room 516 vs other locations.

Lastly, partial matching based on KL divergence resulted to 512 as nearest match of 508 while 503 is identified as top match for 516 as reflected in Figure 12.

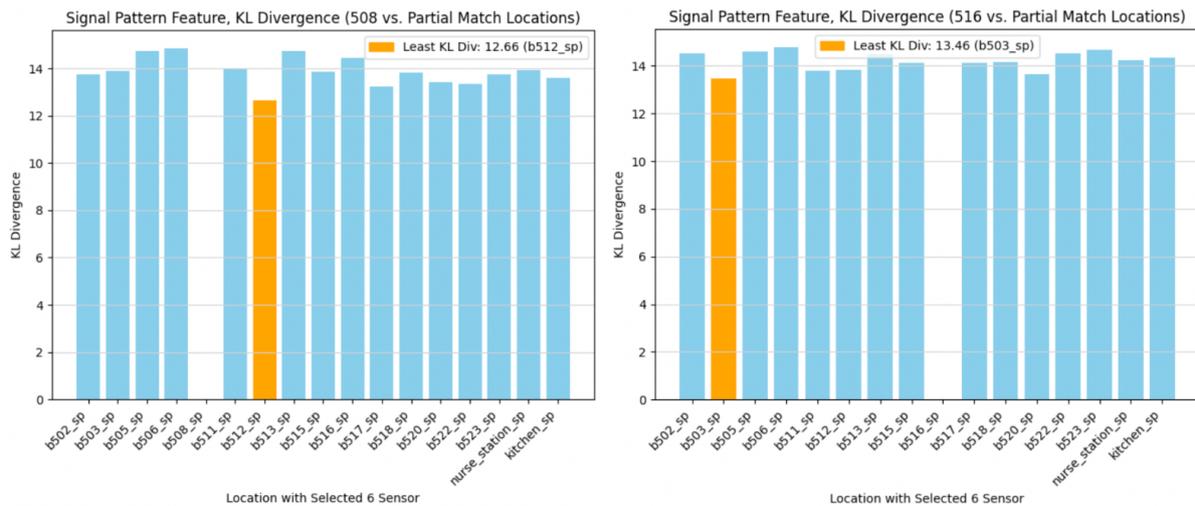


Figure 12. Signal pattern feature based on KL divergence with partial matching. (a) Room 508 vs partial matching locations. (b) Room 516 vs partial matching locations.

With the matching location identified for each minority class in all four cases, we proceed with relabeling the identified match for each minority location and concatenated the augmented data to the original training oversampling Rooms 508 and 516. Indoor localization with Random Forest model was executed using the new training data. Similarly, we applied data augmentation using Random Sampling, ADASYN and SMOTE to both Rooms 508 and 516 generating the same length of augmented data with 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 which specific combination 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 oversampling. Relabeling achieved the highest target class F1-score for both 508 and 516. In overall model F1-score, relabeling improved the model performance by 6%. ADASYN achieved the highest followed by SPRI.

Table 2. Performance of SPRI with Standard Deviation, 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	0.64
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 Method	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.

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 while all data augmentation approach has improved the performance for 508 with SMOTE as the highest followed by relabeling and ADASYN. Applying Random Sampling to the minority class resulted to lower F1-score. Regarding the overall weighted F1-score, only ADASYN and SMOTE improved the model by 2 to 3%. As for relabeling approach, the observed decrease in performance can be attributed to incomplete sensor data resulting from partial matching and the possibility of overlapping classes.

Table 3. Performance of SPRI with Standard Deviation, 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 = 0.00	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 Method	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 overall model F1-score, relabeling improved the model performance by 8% and achieved the highest result.

Table 4. Performance of SPRI 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 Method	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 the 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 other oversampling method. In overall model F1-score, only relabeling resulted to an improved model performance at 62%.

Table 5. Performance of SPRI 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 Method	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.

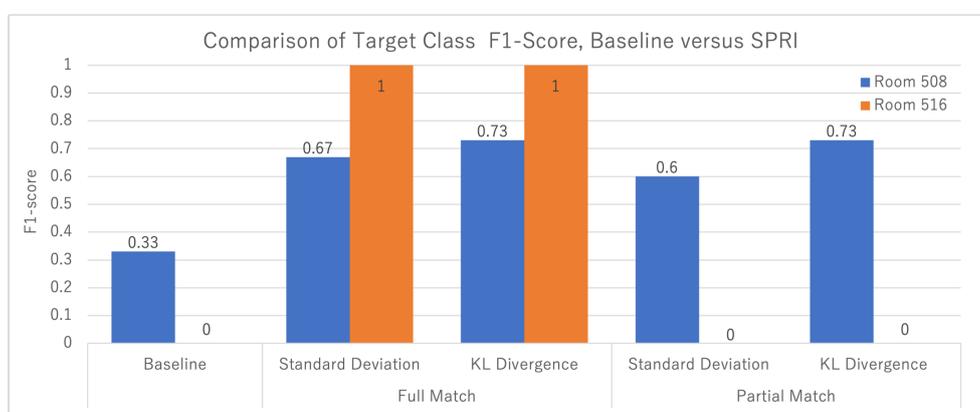
5. Discussion

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 to indoor localization in nursing facility, we illustrate the performance of the proposed method in Figure 13. In response to the first research question on identifying matching signal patterns between different locations in the facility, we calculated signal pattern features using standard deviation to compare per sensor variability and KL divergence to compare signal distributions between rooms. For better understanding of signal patterns, we filtered the analysis of sensors to six surrounding beacons in the neighboring rooms.

In response to the second research question on how we use samples from other sensors to augment location with fewer data, we applied relabeling to the resulting match from the comparison of signal patterns between rooms. We implemented 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 nursing facility in Figure 13 and Figure 14.

Between the variations of relabeling, full matching consistently achieved better performance as compared to baseline and to partial matching in terms of the target class F1-score. As reflected in Figure 13, Room 516 was only classified after employing relabeling with full matching. Matching based on signal pattern feature referenced from KL divergence resulted to better F1-score than standard deviation. Overall, the proposed method has improved the target class F1-score as compared to the baseline by 27% to 40%.

**Figure 13.** Comparison of Target Class F1-Score, Baseline versus SPRI.

By comparing the overall F1-score performance in Figure 14, we identified that full matching achieves better performance than baseline with an increase of 6% to 8%. Moreover, full matching outperformed partial matching by 6% to 7%. For both full and partial matching, KL divergence resulted to better F1-score than standard deviation. Overall, the proposed method has improved the model performance as compared to the baseline except on the case of partial matching with standard deviation. In general, increasing the training data with relabeling based on signal pattern resulted to improved localization.

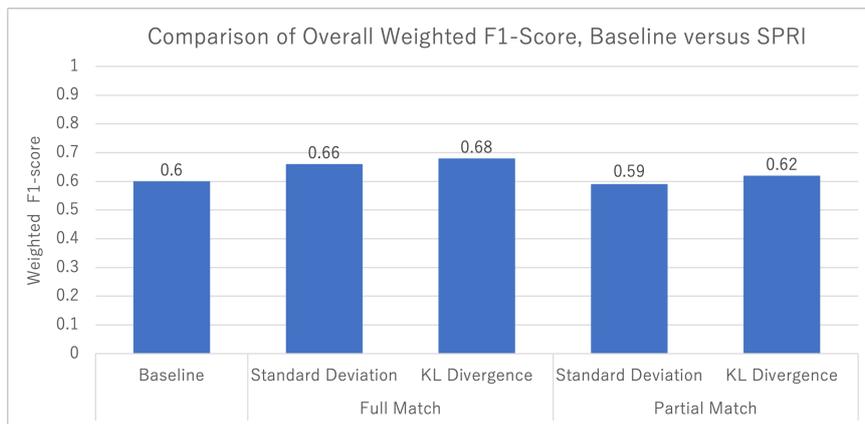


Figure 14. Comparison of Overall Weighted F1-Score, Baseline versus SPRI.

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 15. For target classes F1-score, relabeling has improved the performance of both Rooms 508 and 516. All samples of 508 and 516 are correctly localized with relabeling. In the overall model F1-score, relabeling has improved the performance by 8%. On the other hand, a trade-off of the proposed method is potential class overlap specially from majority class used in the matching which can be observed in Room 520 in the confusion matrix affecting the recognition of the room.

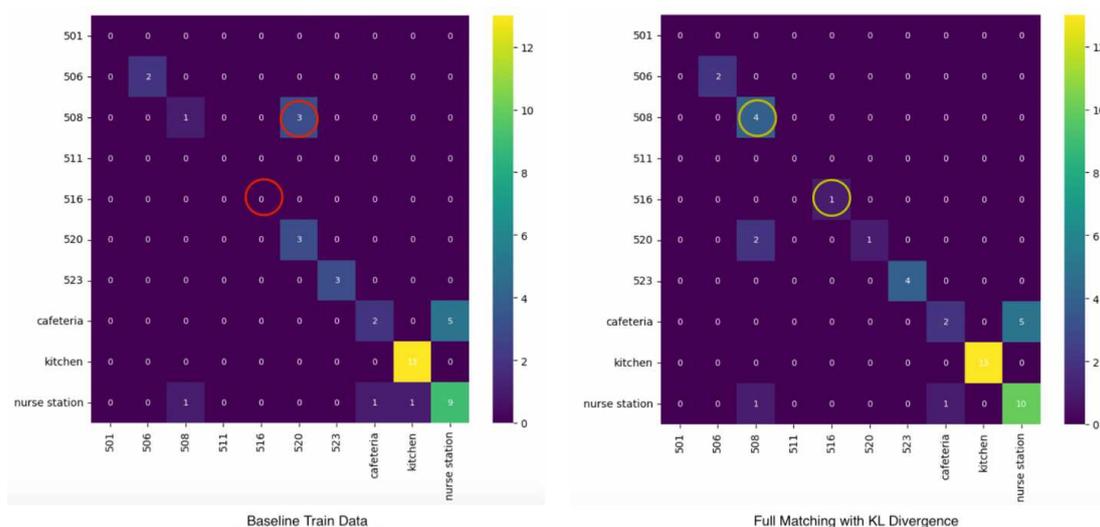


Figure 15. Comparison of indoor positioning with baseline data. (a) Confusion matrix, original data. (b) Confusion matrix, with relabeled data.

For cross-validation, we tested the proposed method using by applying data augmentation on other rooms using the same test data. In the original train data, we reduced samples for Room 520 and 523, and then re-evaluated the performance for baseline before applying augmentation. Table 6

summarizes the result after oversampling Room 520. Relabeling has improved 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 has achieved both the highest overall and target F1-score after applying data augmentation the room increasing the samples five times the original size as depicted in Table 7.

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	11174	0.00	0.53
SMOTE	11174	0.00	0.53
ADASYN	11174	0.33	0.59
Proposed method	11174	0.33	0.61

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

Between the variations of 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 sensor data from other rooms can lead to a diminished performance. Overall, as relabeling is based on leveraging samples from other sensors, the proposed data augmentation approach relies on the number of samples in 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.

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 class. By filtering beacons to surrounding sensors, we calculated the signal pattern feature of minority and majority class and identified matching rooms for relabeling. By applying a relabeling approach to identified location matching the signal patterns of minority class, indoor localization improved both in target class and overall model performance. Moreover, we presented a comparative performance of indoor localization using different augmentation techniques and confirmed our proposed method achieves better indoor localization in nursing home.

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

Through the calculation of signal pattern features, we can discern similarities between different rooms. This enables us to effectively augment locations with limited data samples by repurposing sensor data from matching rooms. By measuring both KL-divergence and standard deviation across classes, we have mapped the signal patterns among various locations. Consistently, the approach based on KL-divergence outperforms baseline data, as well as 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. Furthermore, applying the relabeling approach to other use-case scenario using other sensors aside from beacons will be analyzed to understand signal patterns for matching. Moreover, locations within common area not following specific layout will be further examined. We aim to refine our relabeling approach to encompass areas within these common spaces that feature varying geometries and dimensions, such as cafeterias, where nursing activities tend to occur over extended periods.

Lastly, expanding analysis of signal patterns by incorporating additional statistical analysis techniques is suggested to represent signal pattern feature of different sensors aside from BLE beacons. Hybrid augmentation approach by combining the proposed oversampling method with SMOTE or ADASYN is suggested to resolve limitations of respective oversampling techniques and further improve indoor localization.

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 JST-Mirai Program, Creation of Care Weather Forecasting Services in the Nursing and Medical Field, Grant Number JP21473170, 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, for allowing the staff to carry mobile devices installed with Fonlog during daily nursing routines. The authors extend their gratitude to Xia Qingxin and Ryuichiro Okuda 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
IoT	Internet of Things
IPS	Indoor Positioning System
KL	Kullback-Leibler
MAC	Media Access Control
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. *JHCLP* **2009** *12*,235.
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. *Sensors* **2018**.
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, United Kingdom, 27-29 October, 2022.
5. Calderoni, L.; Ferrara, M.; Franco, A.; et al. Indoor localization in a hospital environment using random forest classifiers. *Expert Syst Appl* **201** *42*, 125–134.
6. 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, United Kingdom, 27-29 October, 2022.
7. Kanan, R.; Elhassan, O. A combined batteryless radio and wifi indoor positioning for hospital nursing. *JCOMSS* **2016**,*12*, 33–44.
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. <https://doi.org/10.3390/s22218119>
9. Zhub, Q.; Xiongb, Q.; Wanga, K.; Lua, W.; Liua, T. Accurate WiFi-based indoor localization by using fuzzy classifier and mlps ensemble in complex environment. *J Franklin Inst* **2020**, 1420–1436. <https://doi.org/10.1016/j.jfranklin.2019.10.028>
10. 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. <https://doi.org/10.1016/j.comcom.2020.04.030>
11. Wang, J.; Ghosh, R.; Das, S. A survey on sensor localization. *J. Control Theory Appl* **2010** *8*, 2–11.
12. Yoo, J.; Park, S.; Fingerprint variation detection by unlabeled data for indoor localization. *PMC* **2020** *67*, 101219, ISSN 1574-1192. <https://doi.org/10.1016/j.pmcj.2020.101219>.
13. Oussalah, M.; Alakhras, M.; Hussein, M.I. Multivariable fuzzy inference system for fingerprinting indoor localization. *Fuzzy Sets Syst* **2015** *8*, 65–89.
14. Berrueta, B.; Baalaa, O.; Caminadab, A.; Guillet, V. An evaluation method of channel state information fingerprinting for single gateway indoor localization. *JNCA* **2020** *159*. <https://doi.org/10.1016/j.jnca.2020.102591>
15. Fei, H.; Xiao, F.; Huang, H.; Sun, L. Indoor static localization based on Fresnel zones model using COTS Wi-Fi. *JNCA* **2020** *167*, 2–pp. <https://doi.org/10.1016/j.jnca.2020.102709>
16. 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. <http://dx.doi.org/10.1016/j.adhoc.2017.06.007>
17. Huang, X.; Guo, S.; Wu, Y.; Yang, Y. A fine-grained indoor fingerprinting localization based on magnetic field strength and channel state information. *PMC* **2017** *41*, 150–165. <http://dx.doi.org/10.1016/j.pmcj.2017.08.003>
18. Sulaiman, B.; Tarapiah, S.; Natsheh, E.; Atalla, S.; Mansoor, W.; Himeur, Y. Radio map generation approaches for an RSSI-based indoor positioning system. *SASC* **2023** *5*. <https://doi.org/10.1016/j.sasc.2023.200054>
19. Wang, S. Wireless Network Indoor Positioning Method Using Nonmetric Multidimensional Scaling and RSSI in the Internet of Things Environment. *Math. Probl. Eng* **2020**. <https://doi.org/10.1155/2020/8830891>

20. 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. <https://doi.org/10.3390/app13063900>
21. Yoo, J.; Kim, H.J. Target Tracking and Classification from Labeled and Unlabeled Data in Wireless Sensor Networks. *J. Sens.* **2014** *14*, 23871-23884. doi:10.3390/s141223871
22. Cortesi, S.; Vogt, C.; Magno, M. Comparison between an RSSI- and an MCPD-Based BLE Indoor Localization System. *J. Comput* **2023** *12*, 59. <https://doi.org/10.3390/computers12030059>
23. 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. doi: 10.1109/JIOT.2019.2940368
24. 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. doi: 10.1109/ACCESS.2021.3062818.
25. Zhou, C.; Yuan, J.; Liu, H; et al. Bluetooth Indoor Positioning Based on RSSI and Kalman Filter. *Wireless Pers Commun* **2017** *96*, 4115–4130. <https://doi.org/10.1007/s11277-017-4371-4>
26. Kotanen, A.; Hannikainen, M.; Leppakoski, H.; et al. Experiments on local positioning with Bluetooth. In Proceedings of the International Conference on Information Technology: Coding and Computing, Las Vegas, NV, USA, 2003; 297-303, doi: 10.1109/ITCC.2003.1197544.
27. 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.
28. 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.
29. 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* **year2020** *15*, 4101. <https://doi.org/10.3390/s20154101>
30. Nguyen, Q. H.; Johnson, P.; Nguyen, T. T.; Randles, M. A novel architecture using iBeacons for localization and tracking of people within healthcare environment. *J. GIoT*, 1-6, doi: 10.1109/GIOTS.2019.8766368.
31. 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, 2006, doi: 10.1109/DELTA.2006.73.
32. Ni, L.; Liu, Y.; Lau, Y.C.; Patil, A. LANDMARC: Indoor Location Sensing Using Active RFID. *Wirel. Netw.* **2004** *10*, 701–710. <https://doi.org/10.1023/B:WINE.0000044029.06344.dd>
33. Haibo He, Yang Bai, E. A. Garcia and Shutao Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, 2008, pp. 1322-1328, doi: 10.1109/IJCNN.2008.4633969.
34. Chawla, N., Bowyer, K., Hall, L.O., & Kegelmeyer, W.P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *ArXiv*, abs/1106.1813.
35. Singh, A.; Purohit, A. A Survey on Methods for Solving Data Imbalance Problem for Classification. *Int. J. Comput. Appl.* **2015** *127*, 37-41.
36. Jiang, X.; Ge, Z. Data Augmentation Classifier for Imbalanced Fault Classification. *IEEE Trans. Autom. Sci. Eng* **2021** *18*,3, 1206-1217. doi: 10.1109/TASE.2020.2998467.
37. Fan, J. et al. EEG data augmentation: towards class imbalance problem in sleep staging tasks. *J. Neural Eng.* **2020** *17*, 056017. doi: 10.1088/1741-2552/abb5be
38. Maharana, K.; Mondal, S.; Nemade, B. A review: Data pre-processing and data augmentation techniques. *Glob. Transit.* **2022** *3*,1, 91-99. <https://doi.org/10.1016/j.gltpt.2022.04.020>.
39. 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; 216–220. <https://doi.org/10.1145/3136755.3136817>
40. Temraz, M.; Keane, M.T. Solving the class imbalance problem using a counterfactual method for data augmentation. *MLWA* **2022** *9*. <https://doi.org/10.1016/j.mlwa.2022.100375>.

41. 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, August 1998, 73–79.
42. He, H.; Garcia, E. A. Learning from Imbalanced Data. *IEEE Trans Knowl Data Eng* **2009** *21*, 1263-1284. doi: 10.1109/TKDE.2008.239.
43. Thabtah, F.; Hammoud, S.; Kamalov, F.; Gonsalves, A. Data imbalance in classification: Experimental evaluation. *Inf. Sci.* **2020** *513*,429-441. <https://doi.org/10.1016/j.ins.2019.11.004>.
44. Jiang, X.; Ge, Z. Data Augmentation Classifier for Imbalanced Fault Classification. *IEEE Trans. Autom. Sci. Eng.* **2021** *18*, pp. 1206-1217. doi: 10.1109/TASE.2020.2998467.
45. Lee, J., Park, K. GAN-based imbalanced data intrusion detection system. *Pers Ubiquit Comput* **2021** *25*, 121–128.
46. 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).
47. Locatify. Indoor Positioning Systems based on BLE Beacons. Available online: <https://locatify.com/blog/indoor-positioning-systems-ble-beacons> (accessed on 26 November 2023).
48. What is BLE? - Bluetooth Low Energy basics. Available online: <https://www.rfwireless-world.com/Terminology/what-is-BLE.html> (accessed on 24 November 2023).
49. Lee, W. 2021. Understanding and Using iBeacons. Available online: <https://www.codemag.com/article/1405051/Understanding-and-Using-iBeacons> (accessed on 24 November 2023).
50. Pichler, S. 2022. Indoor Positioning with Beacons. Available online: <https://www.esri.com/arcgis-blog/products/arcgis-ips/indoor-gis/indoor-positioning-with-beacons/> (accessed on 26 November 2023).
51. 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, 2018, pp. 788-793, doi: 10.1109/PERCOMW.2018.8480348.
52. Wichmann, J. Indoor positioning systems in hospitals: A scoping review. *Digit. Health* **2022** *8*. DOI 10.1177/20552076221081696
53. 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. <https://doi.org/10.3390/bdcc5030042>
54. 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. <https://doi.org/10.1016/j.cosrev.2017.03.002>
55. Ciabattini, L.; Foresi, G.; Monteriù, A. et al. 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. <https://doi.org/10.1007/s12652-017-0579-0>
56. 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, 2021, 1-7, doi: 10.1109/WCNC49053.2021.9417246.
57. 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, 2022, 1-5, doi: 10.1109/VTC2022-Spring54318.2022.9860683.
58. Yoon, J.; Oh, J.; Kim, S. Transfer Learning Approach for Indoor Localization with Small Datasets. *Remote Sens.* **2023** *15*, 2122. <https://doi.org/10.3390/rs15082122>
59. Brandt, J.; Emil Lanzén, E. A Comparative Review of SMOTE and ADASYN in Imbalanced Data Classification. Department of Statistics, Uppsala University, Sweden, 2020.
60. Liu, L.; Zhao, Q.; Miki, S.; Tokunaga, J.; Ebara, H. Indoor Fingerprinting Positioning System Using Deep Learning with Data Augmentation. *Sens. Mater* **2022** *34*, 8, 3047-3061. <https://doi.org/10.18494/SAM3912>
61. Liu, A.; Cheng, L.; Yu, C. SASMOTE: A Self-Attention Oversampling Method for Imbalanced CSI Fingerprints in Indoor Positioning Systems. *Sensors* **2022** *22*, 5677. <https://doi.org/10.3390/s22155677>

62. 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. <https://doi.org/10.3390/s23052545>
63. 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. <https://doi.org/10.3390/app9183930>
64. 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, 2007, 1-5. doi: 10.1109/ICICS.2007.4449717.
65. 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).
66. 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. DOI 10.1088/0957-0233/21/12/124007
67. Kim, M.; Jeong, C. Y. Label-preserving data augmentation for mobile sensor data. *Multidimensional Syst. Signal Process.* **2021** *32*, 115–129. <https://doi.org/10.1007/s11045-020-00731-2>
68. Li, Y.; Adams, N.; Bellotti, T. A relabeling approach to handling the class imbalance problem for logistic regression. *JCGS* **2022** *31:1*, 241–253. DOI 10.1080/10618600.2021.1978470
69. Yassin, A. et al. Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications. *Commun. Surveys Tuts.* **2017** *19,2*, 1327–1346. <https://doi.org/10.1109/COMST.2016.2632427>
70. 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*, <https://doi.org/10.1186/s13673-019-0168-7>
71. Varma, P.S.; Anand, V. Random Forest Learning Based Indoor Localization as an IoT Service for Smart Buildings. *Wireless Pers Commun* **2021** *117*, 3209–3227. <https://doi.org/10.1007/s11277-020-07977-w>
72. Ramadan, M.; Sark, V.; Gutierrez, J.; Grass, E. NLOS Identification for Indoor Localization using Random Forest Algorithm. In Proceedings of 22nd International ITG Workshop on Smart Antennas, Bochum, Germany, 2018, pp. 1-5.
73. Brown, I.; Mues, C. n experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Syst. Appl.* **2012** *39(3)*, 3446–3453. DOI <https://doi.org/10.1016/j.eswa.2011.09.033>

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.