Ad Hoc Networks 159 (2024) 103486

Contents lists available at ScienceDirect

Ad Hoc Networks

journal homepage: www.elsevier.com/locate/adhoc

RSSI-based fingerprint localization in LoRaWAN networks using CNNs with squeeze and excitation blocks

Albert Selebea Lutakamale [*,1], Herman C. Myburgh, Allan de Freitas

University of Pretoria, Department of Electrical, Electronic and Computer Engineering, Private Bag x20, Hatfield, 0028, Pretoria, South Africa

ARTICLE INFO

Keywords:
Fingerprint localization
LoRaWAN
LPWAN
Internet of Things
Wireless communications

ABSTRACT

The ability to offer long-range, high scalability, sustainability, and low-power wireless communication, are the key factors driving the rapid adoption of the LoRaWAN technology in large-scale Internet of Things applications. This situation has created high demand to incorporate location estimation capabilities into large-scale IoT applications to meaningfully interpret physical measurements collected from IoT devices. As a result, research aimed at investigating node localization in LoRaWAN networks is on the rise. The poor localization performance of classical range-based localization approaches in LoRaWAN networks is due to the long-range nature of LoRaWAN and the rich scattering nature of outdoor environments, which affects signal transmission. Because of the ability of fingerprint-based localization methods to effectively learn useful positional information even from noisy RSSI data, this work proposes a fingerprinting-based branched convolutional neural network (CNN) localization method enhanced with squeeze and excitation (SE) blocks to localize a node in LoRaWAN using RSSI data. Results from the experiments conducted to evaluate the performance of the proposed method using a publicly available LoRaWAN dataset prove its effectiveness and robustness in localizing a node with satisfactory results even with a 30% reduction in both the principal component analysis (PCA) variances on the training data and the size of the original sample. A localization accuracy of 284.57 m mean error on the test area was achieved using the Powed data representation, which represents an 8.39% increase in localization accuracy compared to the currently best-performing fingerprint method in the literature, evaluated using the same LoRaWAN dataset.

1. Introduction

In wireless networks, localization is the estimation of nodes' current location by using the knowledge of the absolute positions of a few nodes (anchor nodes) and inter-node measurements such as distance, angle of arrival or time of arrival. LoRaWAN is a communication technology which operates in unlicensed frequency bands, offering long-range and low-cost communication capabilities.

The adoption of LoRaWAN technology in IoT applications is rapidly increasing thanks to its high scalability, sustainability, and low-power nature. Its ability to offer long-range wireless communication at relatively lower costs is attributed to using inexpensive batteries that can last for many years. These characteristics of LoRaWAN technology have boosted its suitability in several long-range industrial applications ranging from luggage tracking to precision farming to smart city projects, to name a few [1].

Incorporation of location estimation capabilities into large-scale IoT applications is crucial if people are to make use of physical measurements gathered from the IoT devices deployed for different purposes. Establishing an IoT device's location with minimal localization error is crucial to its context awareness capability. It is an essential aspect of IoT applications that refers to its ability to alter behaviour based on the measurement it has conducted [2]. In Industrial Internet of Things (IIoT) settings, where many smart objects are interconnected to facilitate information management, the ability to precisely locate the positions of industrial personnel helps increase operational efficiency by giving out location-based progress of work processes [3]. It can also boost safety-related measures by enabling proper and orderly access to the machinery and management of evacuation of workers in case of emergencies [3].

In outdoor scenarios, wireless technologies based on Global Positioning System (GPS) have been preferred technologies in many localization applications [4]. However, the poor performance of GPS-based localization solutions in indoor scenarios, rich-scattering environments and urban canyons, where signal transmission is mainly in a non-line of sight (NLOS) setting, resulting in multipath propagation, capped by high cost and high-power consumption of GPS receivers, has limited

* Corresponding author.
E-mail addresses: albert.lutakamale@tuks.co.za (A.S. Lutakamale), herman.myburgh@up.ac.za (H.C. Myburgh), allan.defreitas@up.ac.za (A. de Freitas).

[1] Albert Selebea Lutakamale is with University of Dodoma currently working towards a Ph.D. degree in Electronic Engineering at the University of Pretoria.

https://doi.org/10.1016/j.adhoc.2024.103486
Received 18 August 2023; Received in revised form 20 February 2024; Accepted 24 March 2024
Available online 26 March 2024

adoption of GPS-based localization approaches in many large-scale IoT applications. Short-range wireless technologies such as WiFi, Zigbee, and Bluetooth, on the other hand, because of their short-range nature, have been predominately adopted in indoor localization scenarios [5–10]. Adopting these short-range wireless communication technologies in large-scale outdoor localization solutions would necessitate the dense deployment of access points, which is typically not economically feasible. These limitations of GPS-based and short-range based localization solutions in the context of large-scale IoT applications have paved the way for the adoption of LPWAN technologies, and in particular, the LoRaWAN communication technology, thanks to its ability to offer scalable, reliable, robust, and low power localization solutions [1].

In LoRaWAN networks, a node's current location can be estimated using either fingerprint or range-based localization methods [11,12]. Range-based localization approaches typically require establishing a path loss model that best describes signal propagation characteristics of a particular environment and adopts it to calculate the specific distance to a particular gateway using its transmitted received signal strength indicator (RSSI) value. The final step in range-based localization approaches is then to estimate the current location of a node by making use of any available geometrical (e.g., trilateration, multilateration, or min–max) or statistical (e.g., Bayesian and Maximum Likelihood) techniques [11]. However, in LoRaWAN networks, the performance of this localization approach can sometimes be challenging due to frequency hopping phenomena which can cause degradation in the localization accuracy [11]. On the other hand, fingerprinting-based localization approaches rely on a database of features, commonly known as fingerprints, at some pre-determined physical locations to compute the current location of a target node through feature matching.

As reported in [13], fingerprint-based localization approaches in large urban areas in unlicensed LPWAN, particularly LoRaWAN, can achieve lower position estimation errors than range-based counterparts. Since the implementation of fingerprint-based localization approaches is through the use of machine learning algorithms, by making use of the algorithmic advancement made in machine learning and increased computing power of computers, these localization methods can lead to satisfactory localization accuracy. The high localization performance of fingerprint-based methods over the range-based methods is due to their ability to effectively learn useful positional information even from noisy RSS data collected in NLOS environmental settings [4].

Most fingerprint-based localization solutions in LPWAN networks, particularly LoRaWAN, use classical 'shallow' machine learning techniques such as the k-nearest neighbours algorithm (and its variants such as weighted kNN), support vector machines, random forests, and decision trees, to design the localization models. Though these models can sometimes give satisfactory localization results, their complexity increases and their performance degrades with the increase in the size of fingerprint datasets, which are necessary if a large outdoor environment is to be covered for localization purposes [11]. Deep learning, particularly CNNs, has proven very efficient in dealing with large datasets without compromising performance.

The success brought by the CNNs models in solving different types of computer vision tasks has led to strong interest from research communities in adopting them in designing fingerprint-based localization methods to achieve satisfactory localization performance [14–24]. In the context of classification and regression tasks, these models have proved in several tasks to be efficient in extracting useful information from structured data [25].

Driven by the goal of improving localization performance in LoRaWAN networks, this work proposes a fingerprint-based localization approach using CNNs with squeeze and excitation blocks. The motivation to conduct this research is to develop a robust localization model that will yield better localization results compared to the currently available fingerprinting localization approaches. The following are the contributions of this work: *(i)* A localization method using CNNs enhanced with squeeze and excitation blocks to infer the location of a node through fingerprinting in LoRaWAN networks is proposed. *(ii)* The optimal data preprocessing approach to enhance the localization performance is presented. *(iii)* The proposed method achieves an 8.39% increase in localization accuracy when compared to the current fingerprinting methods in the literature [11,26–28], when evaluated on the same LoRaWAN dataset [2].

The remaining sections of the paper are as follows: Section 2 presents Related works. Preliminaries are presented in Section 3, where the LoRaWAN technology and fingerprint localization method are briefly introduced. The proposed method is presented in Section 4. Section 5 is dedicated to experimental settings and procedures. Experimental results and discussion are reported in Section 6. Conclusive remarks of the paper are presented in Section 7.

2. Related works

A location of a target node in LoRaWAN networks can be estimated by adopting either fingerprinting-based approaches or range-based approaches through analysis of RSSI, AoA, ToA, TDoA or, in some cases, a combination of more than one parameter [11]. Geometrical techniques such as triangulation and multilateration or statistical techniques such as maximum likelihood and Bayesian filtering are usually adopted in the implementation of range-based localization approaches as in [29–35]. However, the need to install dedicated hardware, like expensive antenna arrays [31,32], and the fact that some of the range-based localization techniques require accurate clock synchronization among anchor nodes [29,30,34] have made them less attractive to researchers and practitioners. Another reason for the reduced adoption of range-based localization approaches in LoRaWAN networks is the RSSI fluctuations caused by shadowing and fading phenomena in multipath propagation, which can, in most cases, lead to poor localization performance [36]. Additionally, the fact that successful implementation of some range-based localization techniques, such as those that adopt trilateration techniques as reported in [35], requires a message from a LoRaWAN node to be received by at least three gateways for its location to be determined is another major drawback. In [35], the authors reported a promising localization accuracy of 256 m mean error and 117 m median localization error on the same publicly available LoRaWAN dataset as utilized in this paper. However, these results are obtained for nodes whose messages are received by over three LoRaWAN gateways since the approach is based on triangulation. Hence, nodes which receive only one or two messages cannot be located.

Fingerprinting based localization approaches as reported in [1,3,4,11,26–28,37–40], are increasingly being applied in locating target nodes in LoRaWAN networks thanks to their robustness in challenging environments with multipath and non-line-of-sight phenomena [41] which enables them to learn useful positional information even from noisy data [4], enabling them to be relatively more accurate than range-based approaches.

In the literature, research in fingerprint-based localization in LoRaWAN generally may take three different directions. Several works focus on collecting location fingerprints at some designated locations in larger measurement campaigns and building fingerprint databases for node localization purposes, as in [2]. Others focus on finding better data preprocessing techniques that can improve the generalization capabilities of machine learning-based fingerprint localization methods, as in [26,27], and [37]. However, most research in fingerprint-based localization in LoRaWAN networks focuses on improving the localization performance of adopted machine learning algorithms, as in [1,4,11], and [38].

Researchers in [2] participated in an extensive outdoor measurement campaign to collect fingerprint datasets in LPWAN networks, which are available in three subsets, namely, Sigfox rural, Sigfox urban, and LoRaWAN urban datasets. These datasets are in version 1.1, collected in 2018, and version 1.2, collected in 2019. With these LPWAN datasets, researchers can quickly evaluate outdoor fingerprinting-based localization methods without the need to take part in measurement campaigns to build fingerprint databases.

Currently, the predominate machine learning algorithms used to develop fingerprint-based localization approaches in LoRaWAN networks are the classical machine learning models such as support vector regression, random forests, k-nearest neighbours (kNN), decision trees, and linear regression algorithms, with only a few of them using deep learning techniques. Authors in [2] implemented a basic kNN fingerprinting localization algorithm and evaluated it using version 1.1 of their LoRaWAN dataset, reporting a localization performance of 398.40 m mean error on the test area. Researchers in [39] using the same dataset, implemented and compared the localization performances of different machine learning-based fingerprinting methods, namely a multilayer perceptron neural network, kNN, and Extra Trees methods, reporting mean and median localization errors of 357 m and 206 m, respectively, for the multilayer perceptron method, which performed better compared to the other methods. A work in [4] proposed interpolation-aided fingerprinting-based localization approaches using three different machine learning architectures, achieving localization performance of 191.53 m mean error with the long short-term memory (LSTM) method using version 1.1 of the dataset reported in [2] which is slightly better compared to the performance of the other two methods that used CNN and artificial neural network (ANN).

Authors in [11,27,28], and [26] adopted version 1.2 of the publicly available LoRaWAN dataset reported in [2] to evaluate their fingerprint localization methods. A work in [11] implemented several RSSI fingerprint-based localization methods using support vector regression, kNN, random forest and several linear regression algorithms, reporting the best result of 340 m mean location error using random forest ensemble technique. In [27], the authors first created a differential fingerprint from the publicly available dataset and fused differences in RSSI with the gateway information along with the differences in arrival times (TDoA). Then two regressors, namely multilayer perceptron and random forest, were trained one after the other using the new dataset, achieving the best result of 310 m mean error using the multilayer perceptron regressor with chronological data split. Research in [28] proposed a two-layer hierarchical clustering-based fingerprint technique using RSSI measurements for urban vehicle fingerprinting localization. They used the K-means algorithm for clustering and a weighted kernel regressor to estimate the vehicle's current location, reporting a localization accuracy of 346 m mean error. A proposal to use an ensemble learning-based outdoor positioning algorithm to improve positioning accuracy using hybrid data was presented in [26]. The authors designed an ensemble algorithm incorporating kNN and Random Forest Regressor (RFR), achieving a mean localization error of 332.63 m.

Research work in [40] evaluated the performance of LoRaWAN-based RSSI fingerprinting localization approaches in a sandstorm environment using two different machine learning algorithms, namely, Gaussian process regression and support vector regression (SVR). Among the two methods, the SVR method achieved better overall localization performance. Authors in [42] deployed a weighted kNN algorithm to perform device localization using RSSI fingerprint. They evaluated their method using data from M-Bus and LoRaWAN networks. The authors also used knowledge transfer acquired from different radio maps to simplify the learning process of their proposed method. In [38], the authors propose an effective method of dealing with the issue of fluctuations experienced in RSSI values due to fading phenomena that may jeopardize the localization performance of fingerprinting-based localization methods in LoRaWAN networks. They propose using extreme RSSI (ERSS), which improves boundary autocorrelation between data points and enhances localization performance.

Authors in [1] proposed two methods to estimate node location in LoRaWAN networks based on ANN. One method used a single-layer perceptron neural network architecture, and another used a multi-layer perceptron neural network architecture. In [3], the authors applied a k-NN fingerprinting localization method to estimate the location of a node using two separate outdoor datasets collected using LoRaWAN devices. An analysis was also conducted to determine if the number of deployed gateways and the spacing between them impact the overall localization performance.

Authors in [43] proposed fingerprinting and range-based approaches to localize a target node in LoRaWAN networks using multiple features of RSSI, SF, and signal-to-noise ratio (SNR). On their experimental dataset collected in an area covering 30,000 square metres, their proposed range-based localization approach incorporating distance mapping using RF, kNN, SVM, gradient boosting (GB) and MLP with trilateration technique achieved an average localization performance of 43.97 m mean error while their fingerprinting localization approach using RF achieved a localization performance of 89.22 m mean error. With the publicly available LoRaWAN dataset similar to the one used in this paper, their range-based approach achieved a localization performance of 735.37 m mean error.

Unlike most currently available fingerprinting localization approaches in LoRaWAN that predominately adopt classical machine learning techniques, this work proposes a convolutional network-based regressor enhanced with SE blocks to localize a node in LoRaWAN networks. Two reasons drive the choice of CNN regressor. The first reason is the complexity of training classical machine learning-based localization models experienced when relatively large datasets are used, which can sometimes lead to degradation in localization performance. The second reason is CNN models' efficiency in learning useful position information in structured data as reported in [25]. The joint use of CNN and SE blocks is to improve channel-wise interdependencies.

3. Preliminaries

3.1. LoRaWAN technology

LoRaWAN technology is a communication protocol which uses LoRa as its physical layer. It is a LoRa-Alliance standardized technology which adopts Chirp Spread Spectrum (CSS) as its modulation technique. In North America, it operates at 915 MHz, in Europe at 868 MHz, and in Asia at 433 MHz bands, which are industrial, scientific and medical (ISM) bands [44]. LoRa modulation is characterized by its scalable bandwidth and frequency, relatively high immunity to fading or multipath, resistance to doppler shift, and robustness to interference.

In LoRa-based systems, the variable nature of its data rate, which can take values varying from 300 bps to 50 kbps, is strongly influenced by the channel bandwidth and the spreading factor (SF) [44]. SF, which can take values ranging from SF 7 to SF 12, is the value that indicates how fast the frequency changes in a LoRa channel. In urban areas, LoRa-based systems can offer a communication range of up to 5 km; in rural areas, the communication range can reach 15 km [45]. Key to this long-range communication is the high receiver sensitivity of LoRa systems which results in a large communication link budget [46].

The network architecture deployed in a LoRaWAN network is a star-of-stars topology consisting of gateways and end nodes/end devices (see Fig. 1). Gateways relay messages between end devices and a central server.

3.2. Fingerprinting localization method

The fingerprint (matching) localization technique [48–50] infers the current position of a node in a wireless network through feature matching by relying on a radio map built using signal features at pre-determined locations. The implementation of this localization method takes place in two steps. The first step involves the construction of a database (radio map) that relates scene features (such as received signal strength indicator (RSSI) and time of arrival (ToA)) from videos, virtual images, or electromagnetic signals with positions of nodes in pre-determined geographic locations.

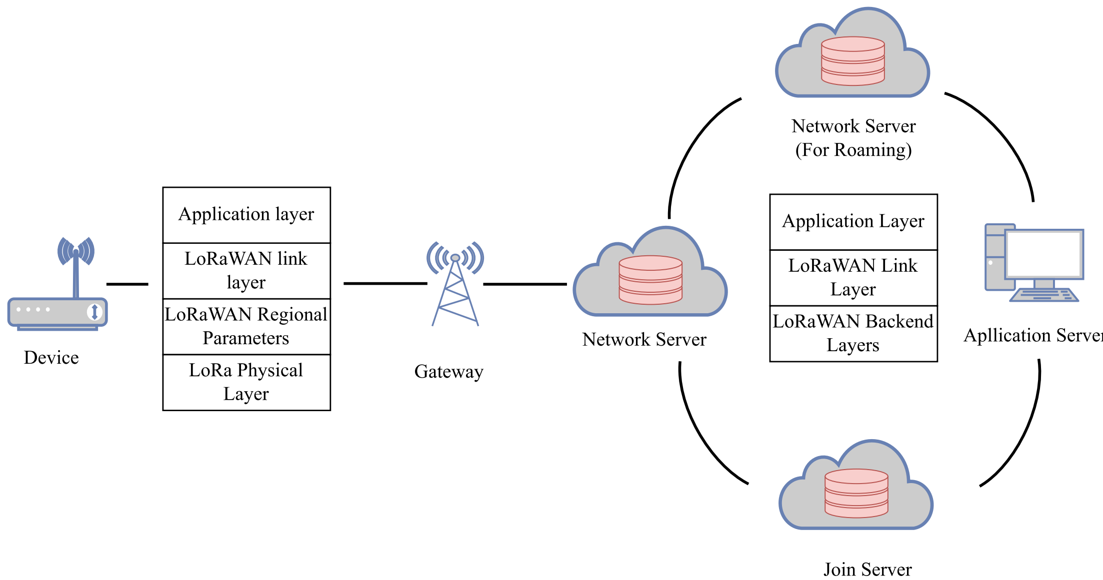


Fig. 1. LoRaWAN network architecture.
Source: Redrawn from [47].

In the second step of a fingerprint localization technique, the target node's real-time position is estimated by associating the measured data from the target node at a particular location with the nearest fingerprint location information stored in the created radio map. The localization accuracy of this localization method depends heavily on the calibration of the fingerprint database and the quality of the location-dependent measurements collected in creating the database.

The implementation of fingerprint-based localization methods is generally done by using machine learning techniques. Machine learning techniques adopted in the fingerprint localization method are the k-NN algorithm (and its variants, such as weighted k-NN), linear regression algorithms, decision tree algorithms, random forests, support vector regression, neural networks, and deep learning.

4. Proposed method

Inspired by an Inception module [51], we propose a three-branch two-dimensional convolutional neural network (CNN) regressor with six convolutional layers to perform fingerprinting localization. The useful aspect of this branched CNN architecture is that it allows processing information at various scales to extract features simultaneously at the next stage after aggregation. This model structure, refer to Fig. 2, can also prevent a blow-up in computational complexity even if the number of processing units at each stage increases significantly [51]. The first branch has a single convolutional layer, while the second and third branches have two and three convolutional layers, respectively. Each convolutional layer computes eight filters over its input with a 1×1 kernel size. The leaky ReLU activation function is used to activate the convolutional layers with a 0.3 constant gradient, allowing for a small, non-zero gradient when the unit is saturated and inactive. In order to improve channel-wise interdependencies, after each convolutional layer, a SE block with eight filters is connected. A SE block [52] is a channel-wise attention mechanism widely used to improve the overall performance of CNNs. It has been used to greater success in computer vision tasks. The three branches of the convolutional layers are then concatenated and connected to six fully connected layers. The first five fully-connected (FC) layers are activated by the ReLU activation function. The first FC layer has 512 units, the second has 256 units, the third has 128 units, the fourth has 64 units, and the fifth has 32 units. The last FC layer with two units is activated by a linear activation function for regression purposes. The decision to use five fully-connected layers before the last two-unit fully-connected layer by halving the number of units in each subsequent layer prevents information loss and thus improves the localization performance. In order to prevent overfitting and subsequently improve the model performance, several hyperparameters necessary in the model's training were tuned. The optimal hyperparameters adopted in this work are a 0.15 dropout ratio and a 0.01 L1 kernel regularizer. At the compiling stage of the model, the Adam optimizer with a dynamic learning rate is used. The learning rate is first set at 0.001 and keeps decreasing by a factor of 0.1 if the validation loss does not improve for ten consecutive epochs. The adopted loss function is the mean absolute error (MAE). The performance of the method reported in this work is evaluated using location estimation error and R^2 score. Fig. 2 is the illustration of the proposed localization approach.

5. Experimental settings and procedures

This work uses version 1.2 of the urban LoRaWAN dataset presented in [2]. The measurement campaign to collect this dataset was conducted in the city of Antwerp, Belgium, in 2019. This publicly available dataset contains 130,430 messages from 72 LoRaWAN gateways. The relevant metadata for fingerprinting localization from this dataset are the RSSI values in dBm to all gateways, spreading factor and nanosecond precise timestamps, with GPS coordinates in latitude and longitude used as ground truth references. During the creation of this database, if a gateway did not receive a message, its RSSI value was set to −200 dBm, indicating an out-of-reach RSSI value. For more details about the measurement campaign and the creation of this dataset, readers can refer to the referenced work in [2].

The proposed fingerprint localization approach uses Keras, Scikit-Learn Python libraries to implement its machine learning models with TensorFlow used as a backend. The experiments were run on Google Colaboratory Jupyter Notebooks using 32 GB RAM Core i7 LG computer. Since in the LoRaWAN dataset, the locations of nodes were given in latitude and longitude coordinates, this work uses the Haversine formula [53] to compute distances in the experiments conducted. The Haversine formula is given by [53] as:

$$hav\left(\frac{Y}{X}\right) = hav(A_2 - A_1) + cos(A_1)cos(A_2)hav(B_2 - B_1), \tag{1}$$

where 'hav' represents the Haversine function, defined as:

$$hav(C) = sin^2\left(\frac{C}{2}\right) = \frac{1 - cos(C)}{2}, \tag{2}$$

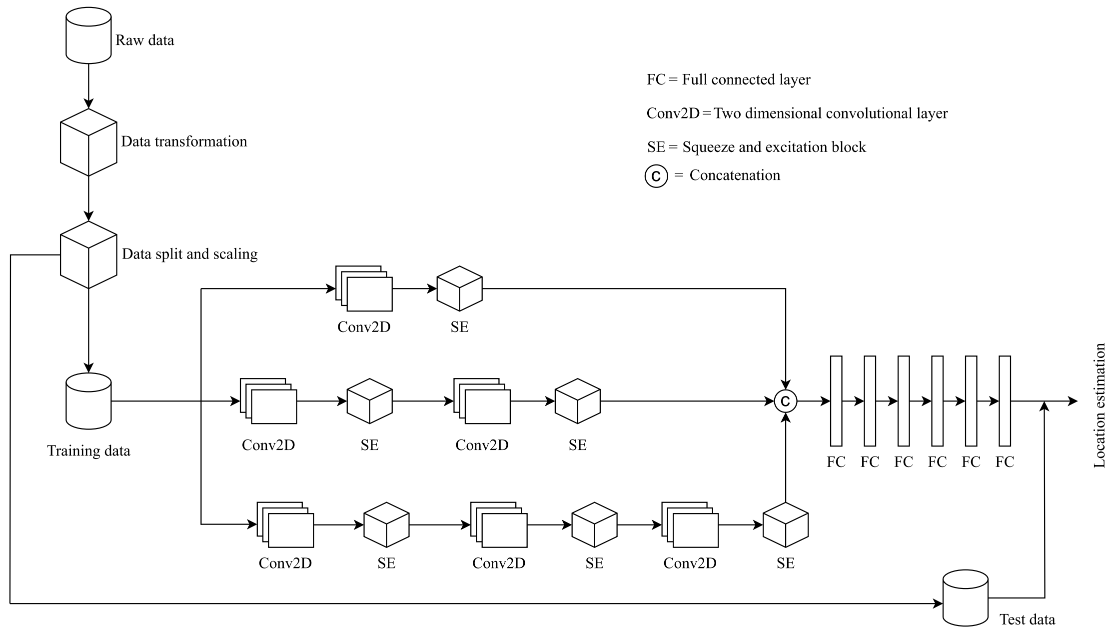


Fig. 2. Proposed fingerprinting localization method.

Y and X represent the distance from one coordinate to the other and the sphere's radius, respectively, A_1 and A_2 represent latitudes of coordinates 1 and 2, respectively, and B_1 and B_2 represent longitudes of coordinates 1 and 2, respectively, all in radians.

In order to use the provided LoRaWAN dataset as input data, it has to be preprocessed first and reshaped to tensors for it to successfully be fed into the proposed CNN-based fingerprint localization model. Each LoRaWAN gateway represents a single feature of the dataset, while every received message is considered a single sample. In this work, the RSSI values received by the base stations (gateways) are used along with LoRa spreading factor to train the models. Though in the reported LoRaWAN dataset, it is indicated that there were a total of 72 gateways; among them, there were a total of 28 gateways that did not receive at least a single transmitted message. Therefore, these gateways were removed and remained with 44 gateways. Together with a spreading factor column, the dataset becomes a 130,430 samples dataset with 45 features. The latitude and longitude columns were used as labels.

After identification of the number of samples and features of the dataset to be used, the RSSI values are then transformed into four commonly adopted RSSI data representations, namely, Positive, Normalized, Exponential and Powed RSSI data representations [54–56]. The transformation is necessary to improve the RSSI-based fingerprinting method's localization performance. According to [39] before transforming RSSI data into these four data representation forms, the minimum received RSSI value ('min') is identified first, and then all out-of-reach RSSI values are set to '$\tau = min - 1$'. The positive data representation is then given by the following equation:

$$Positive_i(x) = RSSI_i - \tau, \tag{3}$$

where, i is the base station (gateway) identifier and $RSSI_i$ is the received signal strength at gateway i. After this transformation, the out-of-reach RSSI values are set to zero, and actual received RSSI values have positive values from 1 and above. The Normalized data representation is given as:

$$Normalized_i(x) = \frac{Positive_i(x)}{-\tau}. \tag{4}$$

The Exponential and Powed data representation were first proposed in [56]. The Exponential data representation is defined as:

$$Exponential_i(x) = \frac{e^{\frac{Positive_i(x)}{\alpha}}}{e^{\frac{-\tau}{\alpha}}}. \tag{5}$$

The Powed data representation, on the other hand, is defined as:

$$Powed_i(x) = \left(\frac{Positive_i(x)}{-\tau}\right)^{\beta}. \tag{6}$$

The α and β in Eqs. (5) and (6) are parameters to be defined according to how the RSSI values are distributed. The recommended values of α and β according to [56] are 24 and e, respectively, where e is the numerical constant. But because these values were adjusted using WiFi signals in indoor experimental settings, authors in [39] re-adjusted these values for LoRaWAN outdoor RSSI values and came up with optimal values of $\alpha = 60$ and $\beta = 1.1$, which were adopted in this work.

Before feeding this newly transformed data into machine learning models, the training data is rescaled using StandardScaler of the Sklearn package in Python. The labels are rescaled using MinMaxScaler. The StandardScaler rescales the data to zero mean and unit standard deviation, creating a zero mean and unit variance distribution. The MinMaxScaler rescales the data to the range of [0, 1]. The rescaling of the training data eases the learning process of the machine learning models and subsequently enhances the performance of the machine learning models.

6. Experimental results and discussion

The following experiments were conducted to validate the proposed method's performance and robustness. In each experiment, a proposed fingerprint localization model is trained for 150 epochs in mini-batches of 512 samples. Before splitting the dataset into training data, validation data and test data, a random shuffling of the data values using numpy.random.shuffle method with a fixed seed of 42 is first conducted. Shuffling the training data helps prevent any bias during the training, prevents the model from learning the training order, and consequently improves the overall performance of the models.

6.1. Performance of the proposed method on different data representation schemes

To evaluate the performance of the proposed method on different data representations, the dataset is transformed first into the Powed data representation scheme and then split into training data, validation data and test data according to a 0.7/0.15/0.15 ratio. After the dataset

Table 1
Performance of the proposed method on different data representation schemes (optimal results are in bold).

Data split ratio	Powed	Exponential	Positive	Normalized
0.7/0.15/0.15	**292.04**	294.46	298.13	294.43
0.8/0.1/0.1	**290.2**	290.86	291.78	293.72
0.85/0.1/0.05	**284.57**	286.12	289	284.53

split, the localization model is trained using the training and validation sets, followed by the prediction step using the test set. This experiment is repeated for 0.8/0.1/0.1 and 0.85/0.1/0.05 data split ratios. The reason for trying different data split ratios was to determine if the use of different data split ratios has any impact on the overall localization performance of the proposed method.

Furthermore, the dataset is also transformed into Exponential, Positive and Normalized data representations and all the experiments conducted using the Powed data representation scheme are repeated. Table 1 presents the performance of the proposed method for each data representation scheme. From the results, it is clear that the Powed data representation scheme slightly outperforms all the other three data representation schemes over the three data split ratios, achieving mean localization errors (m) of 292.04, 290.2, and 284.57 on the 0.7/0.15/0.15, 0.8/0.1/0.1 and 0.85/0.1/0.05 data split ratios, respectively.

These results indicate that the more the test set size is reduced, the better the localization results become. This outcome is because the model is subjected to more training data. However, the ratio of the testing data relative to training and validation data used to validate the performance of machine learning models should not be too small to prevent bias towards the obtained results. The experimental results also reveal slightly better localization performance of the proposed method trained using the dataset transformed according to a Powed data representation scheme compared to the other data representation schemes. Overall, the Positive data representation scheme achieved poor localization results compared to the rest of the data representation schemes. The better localization results obtained using Powed as well as Exponential data representation schemes over the three different data split ratios could be attributed to the non-linearity nature of these two data representation schemes. Adopting one of these two data representation schemes to similar machine learning-based localization tasks using RSSI data will likely yield better results than the other data representation schemes.

Since the Powed data representation scheme yielded the best results for the proposed localization model, it is the adopted data representation scheme for the remaining experiments. Additionally, unless otherwise stated, a 0.7/0.15/0.15 ratio is adopted to split the dataset into training data, validation data, and test data for the remaining experiments. This data split ratio is the most adopted data split ratio in the literature, so it is adopted in the rest of the experiments to have a fair comparison of the localization performance with other works.

6.2. Performance of the proposed method for each retained percentages of variances after applying PCA

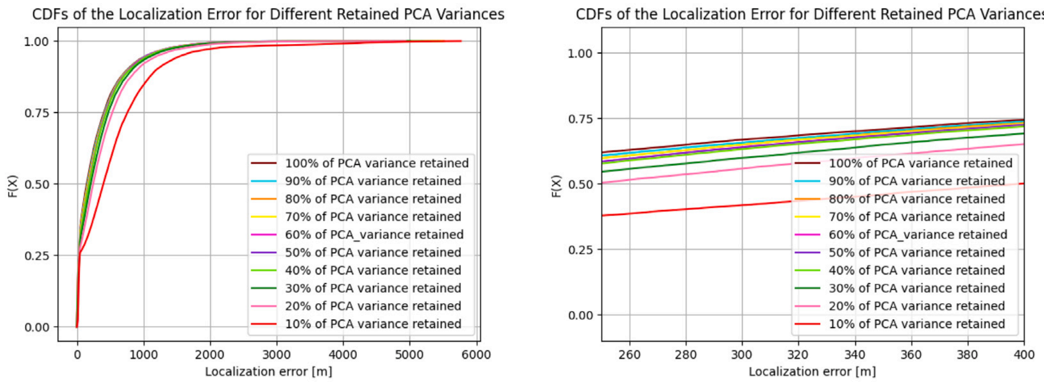
In this section, PCA is performed on the training dataset, and a series of experiments on different percentages of retained variances are performed to validate the performance of the proposed localization model. The experiments on 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100 percentages of retained PCA variances of the training dataset are performed one by one to yield localization performance in terms of mean errors as shown in Table 2 and Fig. 3. Table 2 presents the performance of the proposed method on each percentage of retained PCA variances as well as the trainable parameters as a result of the reduction in variances. Fig. 3 shows full and zoomed versions of figures presenting cumulative distribution function (CDF) curves of test errors for each of the retained percentages of the variances. A training dataset with only 10% of the retained PCA variance produced the worst localization result of 549.05 m mean localization error; this is expected given the large amount of information removed from the training dataset. As the percentage of the retained PCA variances was increased, the localization results also improved, reaching the highest localization accuracy of 292.04 m mean localization error with 100% retained PCA variance. The same trend is also seen regarding the number of trainable parameters. The use of reduced percentages of PCA variance of the training data reduces the computation burden of the method because of the reduced number of training parameters but at the expense of reduced localization accuracies. These results show that the proposed localization model can achieve satisfactory localization performance even with a 70% retained PCA variance of the data used for training the proposed method, consequently reducing the number of training parameters and, hence, the proposed method's training duration. Performing PCA is necessary because, in the LoRaWAN fingerprint databases, the number of messages received by different gateways distributed in an area for localization purposes differs for all gateways. Depending on the transmitter's location, some gateways will receive many messages, some will receive fewer, and some will fail even to receive a single message. This situation necessitates performing a PCA analysis on the dataset before using it to train the model to remove any components that have very little influence on the data to increase the model's training speed without compromising the localization performance. The proposed method's satisfactory localization performance, even with a 30% reduction in retained variances of the training data, proves the effectiveness and suitability of the proposed method to localize a node in LoRaWAN networks.

Table 2
Performance of the proposed method on different percentages of retained variances.

Retained variances (%)	Mean localization error (m)	Number of trainable parameters
10	549.05	200,230
20	378.42	224,806
30	338.36	261,670
40	317.01	310,822
50	311.27	372,262
60	309.66	421,414
70	304.58	482,854
80	303.34	544,294
90	299.29	630,310
100	292.04	728,614

6.3. Performance of the proposed method for different sample sizes of training data

This section presents experiments conducted to determine the impact different sample sizes of training data have on the performance of the proposed method. The procedure followed is that the portion of the training sample size in different percentages (10%, 25%, 30%, 50%, 70%, 85%, 100%) is extracted first and then using a ratio of 0.7/0.15/0.15, the training, validation and test sets are extracted to train and validate the method. This training procedure is carried out to determine to what extent reduced sizes of the training samples impact the overall localization performance of the proposed method. Experiments for each of the sample sizes extracted are then performed, yielding localization results as presented in Figs. 4 and 5, whereby in Fig. 4, the histograms of the localization errors in terms of mean errors (m) on the test area for each sample size are presented and in Fig. 5 full and zoomed CDF curves of test errors for each sample size are presented. As indicated in Figs. 4 and 5, it is clear that these results are in line with the general trend of machine learning algorithms of exhibiting an improved performance when the size of the training set



(a) Full CDF curves

(b) Zoomed CDF curves

Fig. 3. CDFs of localization errors (m) for different retained PCA variances.

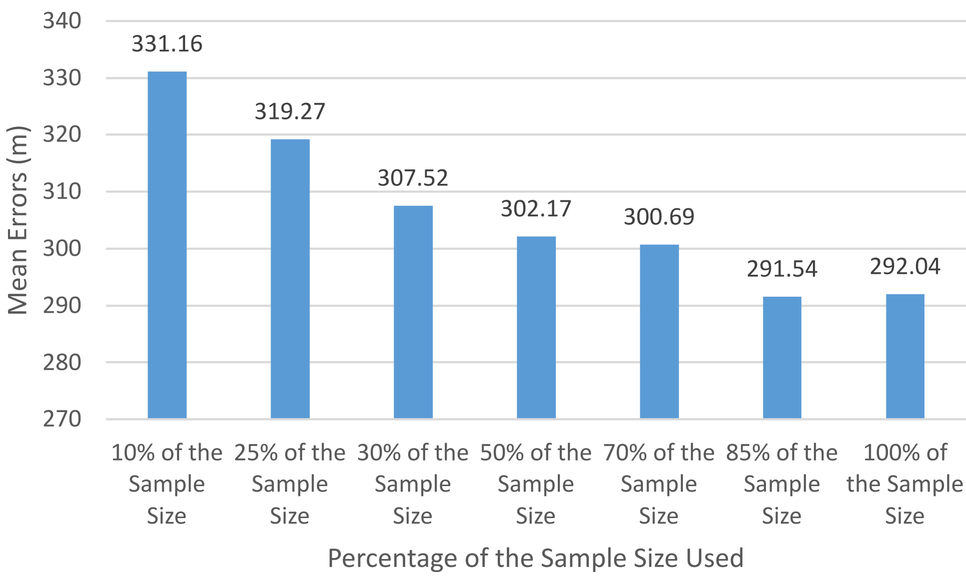


Fig. 4. Performance of the proposed method on different percentages of training dataset.

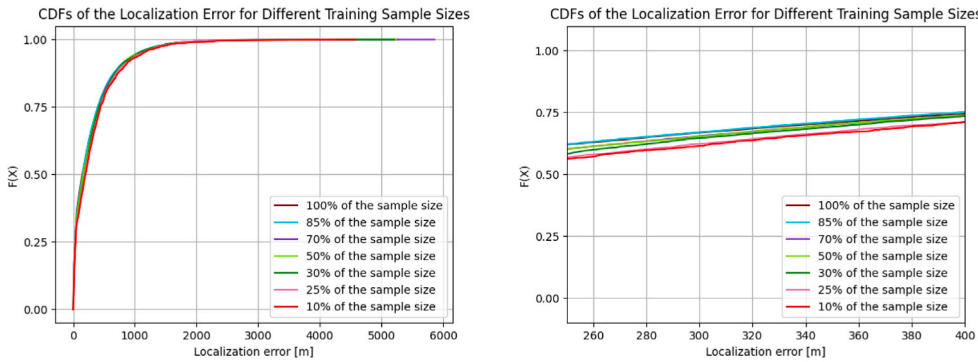
is increased. The use of only 10% of the original dataset to train and test the proposed method produced the worst localization accuracy of 331.16 m mean localization error; however, the performance improved with the increase in sample sizes of the training data to 292.04 m mean localization error when the original size of the dataset was used. From these results, it is clear that the proposed method is able to yield acceptable levels of localization performance when trained using at least 70% of the size of the original dataset. The fact that the proposed method can give satisfactory localization performance even with a 30% reduction in the original sample size of the data used to train and test it illustrates its potential to be applied in applications with even smaller datasets.

6.4. Performance of the proposed approach for different sample sizes of training data with fixed test set

In this section, unlike in Section 6.3, 15% of the original sample size of the training dataset is extracted and fixed to be used to test the proposed localization method's performance. For the remaining 85% of the sample size, a portion of the data in different percentages of 20, 40, 60, 80 and 100 are extracted to train and validate the proposed localization method. For each extracted sample, a ratio of 0.8/0.2 is adopted to split it into training and validation data. The localization model is then trained on each sample size using the training and validation data. This data splitting and localization model training strategy is adopted to determine how well the proposed method generalizes when trained on reduced sizes of training and validation sets but tested on the same test set. As a general trend, the generalization ability of a machine learning algorithm is expected to improve with an increase in the training dataset; however, a well-designed algorithm can generalize well even with reduced sizes of the training samples. Fig. 6 shows the histograms of localization performance in mean errors (m) for each percentage of the remaining sample size. Fig. 7 shows the full and zoomed CDF curves of test error lists for each percentage of the remaining sample size. It is clear from the results that the reduced sizes of the training and validation sets impact the generalization ability of the proposed method. However, the fact that the proposed method was able to obtain an accuracy of 299.79 m mean localization error when trained using a training sample size of as little as 60% of the remaining dataset and tested on the unchanged test set is a clear indication of the potential of the proposed method in localizing a node in LoRaWAN networks with acceptable levels of localization accuracies.

6.5. Performance of the proposed approach for different data shuffling seeds

In this section, different training data shuffling seeds were used to shuffle the training data. Then, experiments were conducted to evaluate



(a) Full CDF curves

(b) Zoomed CDF curves

Fig. 5. CDFs of localization errors (m) for different training sample sizes.

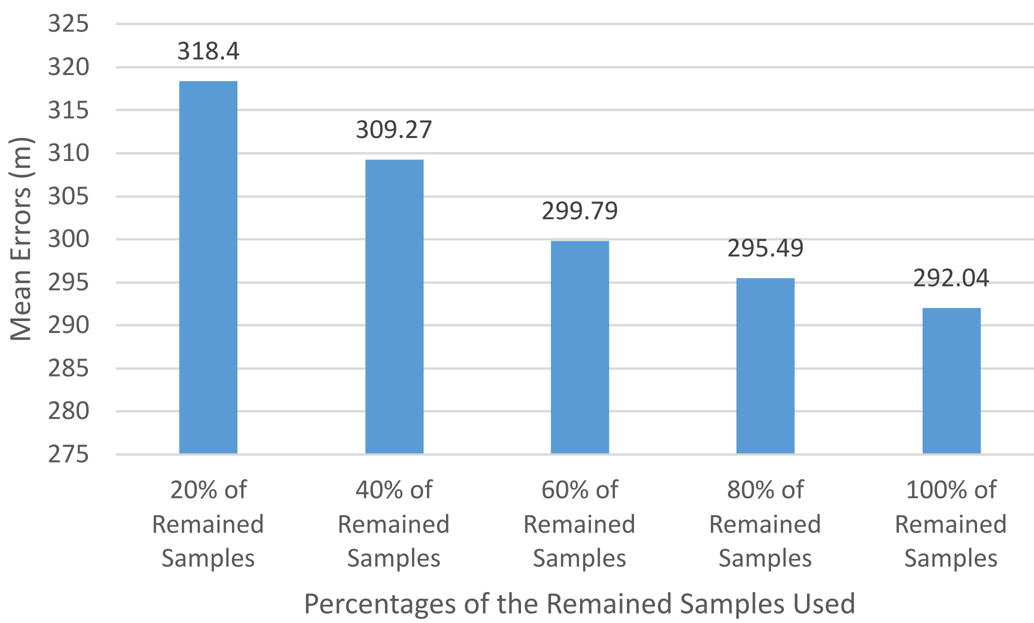
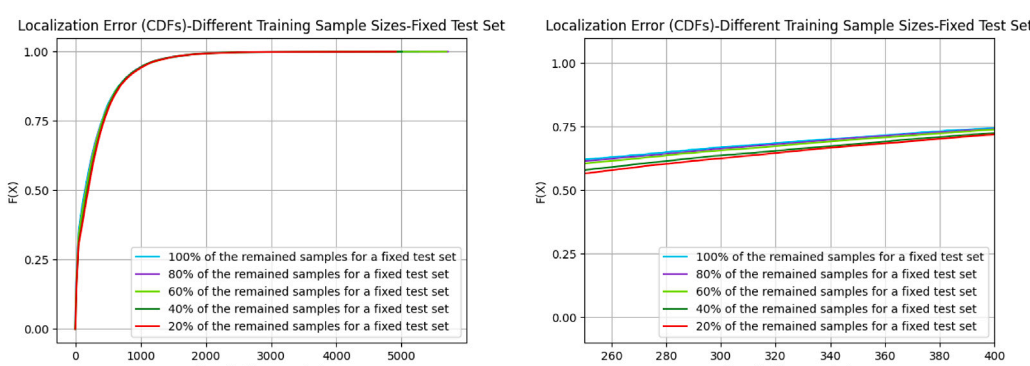


Fig. 6. Performance of the proposed method on different percentages of remaining samples with a fixed test set.



(a) Full CDF curves

(b) Zoomed CDF curves

Fig. 7. CDFs of localization errors (m) for different sample sizes with fixed training set.

how they impact the localization performance of the proposed method. In machine learning, shuffling is performed to change the order of how the individual samples appear in the dataset. For a machine learning algorithm to be robust, its performance is not expected to change significantly when it is subjected to different ordering of the sample elements in the same dataset. In order to reproduce a particular order of data elements, the same seed value is used as a base number when generating random numbers. The training data was shuffled with 1, 2, 3, 5, 10, 14, 16, 22, 42, 64, 128, and 512 random seeds before splitting into training, validation, and test sets. Because the dataset was

Table 3
Localization performance of the proposed method in terms of mean errors (m) on different training data shuffling seeds.

Data shuffling seeds	Localization performance
1	293.79
2	293.35
3	294.35
5	294.38
10	294.37
14	293.33
16	293.81
22	293.79
42	292.04
64	292.41
128	294.36
512	294.04

randomly shuffled using different seeds, the contents of the training, validation, and test sets for each of the settings will differ, impacting the final localization performance, which is evidenced in Table 3, which shows how the proposed method performs in terms of mean errors (m) for each shuffling seed. It is noted that shuffling seeds 42 and 64 give slightly better results compared to the rest of the shuffling seeds. However, despite the different shuffling seeds, the proposed method still produced consistent results, proving its robustness.

6.6. Performance comparison of the proposed approach with related works

In this section, the localization performance of the proposed method is compared with related works proposed in [11,26,27], and [28], which were evaluated using the same dataset as the proposed method. The metrics used to evaluate these related works vary from one work to another. Mean localization error, R^2 score, and validation time were adopted in [11] while mean and median errors were adopted in [27,28]. In [26], only mean localization error was reported.

In order to have a fair comparison of the localization performance of the proposed method with the related works, apart from using the same dataset, the experimental environment, experimental settings, and procedures ought to be matched with each of the works compared. However, fulfilling this condition is not always possible. For instance, none of these related works have made source codes related to their experiments publicly available; as a result, it was difficult to reproduce their results. The difficulty in reproducing the results of these works is also due to the fact that key experimental settings and procedures from these works are either missing or not clear enough. For instance, the Haversine formula is predominantly adopted in computing the equivalent distance between two points on the earth's surface in fingerprinting-based localization approaches in Sigfox and LoRaWAN networks that use GPS coordinates as ground truth references [1,37,39,54]; however, this information is missing in some of the works whose localization performances are compared with the proposed method. Because of these reasons and the fact that the end goal of all these related methods is to improve the localization performance using the same dataset, the best this work could do is just to provide performance comparison with regard to localization accuracies reported. Table 4 summarizes the experimental settings of the related works whose performances were compared with the proposed method. To allow other researchers to compare the performance of their methods with the proposed method in future, the source codes for this work will be made available in a GitHub repository accessed via https://github.com/lutakamale/CNN-SE.

The configuration of the proposed method, whose results were used for comparison, adopted a Powed data representation scheme, using training data shuffled with 42 seeds with training data, validation data, and test data extracted using a ratio of a 0.7/0.15/0.15 (the dominantly used ratio by the related works). In order to obtain a more general localization performance of the proposed method, the localization results were evaluated by repeatedly randomizing the splitting of data into three subsets. In each instance the subsets represented the training, validation and test sets. In the first randomized experiment, a localization performance of 148.06 m and 292.04 m median and mean localization errors, respectively, were observed on the test set. In the second randomized experiment a localization performance of 148.86 m and 288.82 m median and mean localization errors, respectively, was observed. Finally, in the third randomized experiment, a localization performance of 145.72 m and 293.62 m median and mean localization errors, respectively, was observed. In each experiment, the obtained R^2 score was 0.93. Therefore, on average, the proposed method achieved a localization accuracy of 147.55 m and 291.51 m median and mean localization errors, respectively, on the test set and an R^2 score of 0.93. The training procedure for each randomized experiment consisted of 728,614 trainable parameters, lasting approximately 306.34 s through 150 iterations. Table 5 shows the comparison of localization performance of the proposed method with related works in the literature. Fig. 8 illustrates the mapping of the test set's true latitude and longitude coordinate pairs and the estimated latitude and longitude coordinate pairs for the first randomized experiment.

The proposed method outperformed the best performing RF method among ten implemented methods in [11] in terms of mean error and R^2 score; outperformed the kNN-RF method reported in [26], MLP method reported in [27] and a hybrid method that used K-means and Weighted Kernel Regression proposed in [28], in terms of mean localization error. The variation in median errors, i.e., 57 m, 158.48, and 147.55 m obtained in [27,28], and the proposed method, respectively, is due to variation in the number of outliers present in the error lists of the three methods. However, since the mean and median errors of the proposed method are closer to each other than the mean and median errors from the other two methods, the distribution of errors computed by the proposed method is less skewed compared to the distribution of errors of the other methods. Research works proposed in [26,27] require more computation time and memory space due to the use of training data containing fused RSSI and TDoA values, unlike the proposed method, which used only RSSI values and the SF. The methods proposed in [11,28] are likely to require less computation time and memory space compared to the proposed method due to a PCA that was performed on the training data in [11] to remain with 95% of the variance, and the use of only 27 gateways in [28] as a result of the removal of gateways with less than 1% visibility. The satisfactory localization performance of the proposed method is attributable to the use of convolutional neural networks, which are very powerful in learning local dependencies from the structured data as well as the squeeze and excitation blocks to improve channel-wise interdependencies; as a result, the regressor was able to infer the location of a node with relatively high accuracy.

The superior localization results, particularly in terms of mean localization error of the proposed method when compared to the related works and the robustness and effectiveness demonstrated by it in sections 6.2 through 6.5 of this paper, where different strategies were taken to evaluate its performance, gives a clear proof of its potential to be deployed in real work scenarios. The 728,614 total trainable parameters and execution time of 306.34 s further indicate that the proposed method is relatively inexpensive regarding computation complexity.

7. Conclusion

This work proposed a deep learning-based fingerprinting localization method in LoRaWAN networks using CNNs with SE blocks. The system design, including CNN-SE fingerprinting-based system architecture, is presented along with four data representation schemes commonly used to improve localization performance. A series of experiments to validate the performance and robustness of the proposed method are

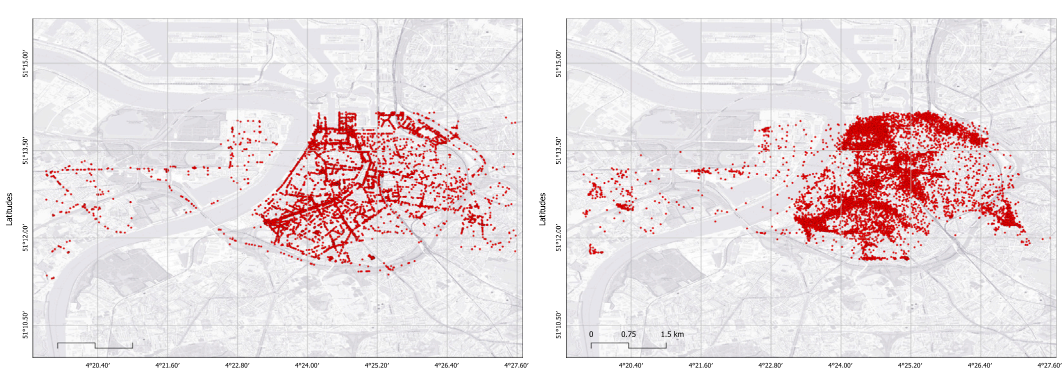
Table 4
Experimental settings of the related works whose performances were compared with the proposed method.

Research work	Python libraries used	Experimental environment	Localization parameters	Ground truth references	Distance formula
[11]	Only Scikit-Learn is mentioned	Virtual machine with 32 GB RAM memory and 10 CPU cores	RSSI	Latitudes and Longitudes	Haversine[a]
[26]	Only Scikit-Learn is mentioned	Not mentioned	RSSI and TDoA	Latitudes and Longitudes	Not mentioned
[27]	Only Scikit-Learn is mentioned	Not mentioned	Differential RSSI and TDoA	Latitudes and Longitudes	Not mentioned
[28]	Not mentioned	Not mentioned	RSSI	Latitudes and Longitudes	Not mentioned
Proposed method	Keras+Scikit-Learn+TensorFlow	Google Colab using a 32 GB RAM Core i7 LG computer	RSSI and SF	Latitudes and Longitudes	Haversine

[a] The Haversine distance is not explicitly mentioned in [11], however, it is described in the first author's PhD thesis [57].

Table 5
Localization performance of the proposed method compared to related works trained on the same dataset.

Research work	Scheme	Mean localization error (m)	Median localization error (m)	R^2 Score
[11]	Random Forest	340	Not reported	0.91
[26]	kNN-RFR	332.63	Not reported	Not reported
[27]	MLP	310	57	Not reported
[28]	K-means + Weighted Kernel Regression	346	158.48	Not Reported
Proposed method	**CNN + SE**	**291.51**	**147.55**	**0.93**



(a) True latitude and longitude coordinate pairs from the test set

(b) Estimated latitude and longitude coordinate pairs

Fig. 8. Mapping of the true latitude and longitude coordinate pairs of the test set and the estimated latitude and longitude coordinate pairs for the first randomized experiment.

conducted. Experimental results proved the proposed method's effectiveness and robustness in localizing a node in LoRaWAN networks. The proposed method achieved satisfactory localization results even with a significant reduction in PCA variances and the size of the dataset used to train the proposed method. The overall best localization performance of 284.57 m in terms of mean errors on the test area was achieved using a Powed data representation scheme with training data, validation data and test data extracted using a 0.85/0.1/0.05 ratio. Future work will jointly leverage RSSI and nanosecond precise timestamps fingerprints to localize a node in LoRaWAN networks.

Funding

The authors received no direct funding for this research.

CRediT authorship contribution statement

Albert Selebea Lutakamale: Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Herman C. Myburgh:** Writing – review & editing, Validation, Supervision, Resources, Investigation, Funding acquisition. **Allan de Freitas:** Writing – review & editing, Validation, Supervision, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The dataset used is a publicly available dataset.

Acknowledgements

Albert Selebea Lutakamale received financial support from the Ministry of Education, Science and Technology of the United Republic of Tanzania under the MoEST-HEET Project fund.

References

[1] M. Anjum, M. Khan, S. Hassan, H. Jung, K. Dev, Analysis of time-weighted LoRa-based positioning using machine learning, Comput. Commun. 193 (2022) 266–278, http://dx.doi.org/10.1016/j.comcom.2022.07.010.

[2] M. Aernouts, R. Berkvens, W. Xia, K. Vlaenderen, M. Weyn, Sigfox and LoRaWAN datasets for fingerprint localization in large urban and rural areas, Data 3 (13) (2018) 1–15, http://dx.doi.org/10.3390/data3020013.

[3] E. Svertoka, I. Marghescu, A. Rusu-Casandra, R. Burget, J. Hosek, P. Masek, A. Ometov, Evaluation of real-life LoRaWAN localization: Accuracy dependencies analysis based on outdoor measurement datasets, in: Paper Presented at the 2022 14th International Conference on Communications, COMM, 16–18 June 2022, Bucharest, Romania, 2022, http://dx.doi.org/10.1109/COMM54429.2022.9817302.

[4] J. Purohit, X. Wang, S. Mao, X. Sun, C. Yang, Fingerprinting-based indoor and outdoor localization with LoRa and deep learning, in: Paper Presented at the GLOBECOM 2020 - 2020 IEEE Global Communications Conference, 07-11 December 2020, Taipei, Taiwan, 2020, http://dx.doi.org/10.1109/GLOBECOM42002.2020.9322261.

[5] Z. Li, X. Li, G. Mou, D. Jiang, X. Bao, Y. Wang, Design of localization system based on ultra-wideband and long range wireless, in: Paper Presented at the 2019 IEEE 11th International Conference on Advanced Infocomm Technology, ICAIT, 18–20 October 2019, Jinan, China, 2019, http://dx.doi.org/10.1109/ICAIT.2019.8935892.

[6] M. Anjum, M. Khan, S. Hassan, A. Mahmood, H. Qureshi, M. Gidlund, RSSI fingerprinting-based localization using machine learning in LoRa networks, IEEE Internet Things Mag. 3 (4) (2020) 53–59, http://dx.doi.org/10.1109/IOTM.0001.2000019.

[7] A. Loganathan, N. Ahmad, P. Goh, Self-adaptive filtering approach for improved indoor localization of a mobile node with Zigbee-based RSSI and odometry, Sensors 19 (21) (2019) 4748, http://dx.doi.org/10.3390/s19214748.

[8] B. Xu, B. Hussain, Y. Wang, H. Cheng, C. Yue, Smart home control system using VLC and bluetooth enabled AC light bulb for 3D indoor localization with centimeter-level precision, Sensors 22 (21) (2022) 8181, http://dx.doi.org/10.3390/s22218181.

[9] C. Basri, A. Khadimi, Survey on indoor localization system and recent advances of WIFI fingerprinting technique, in: Paper Presented at the 2016 5th International Conference on Multimedia Computing and Systems, ICMCS, 29 September 2016 – 01 October 2016, Marrakech, Morocco, 2016, http://dx.doi.org/10.1109/ICMCS.2016.7905633.

[10] L. Chang, W. Chun, H. Qiany, A cooperative indoor localization enhancement framework on edge computing platforms for safety-critical applications, in: Paper Presented at the 2019 15th International Conference on Mobile Ad-hoc and Sensor Networks, MSN, 11–13 December 2019, Shenzhen, China, 2019, http://dx.doi.org/10.1109/MSN48538.2019.00077.

[11] T. Janssen, R. Berkvens, M. Weyn, Benchmarking RSS-based localization algorithms with LoRaWAN, Internet Things (Netherlands) 11 (2020) 100235, http://dx.doi.org/10.1016/j.iot.2020.100235.

[12] J. Torres-Sospedra, M. Aernouts, A. Moreira, R. Berkvens, LoRaWAN fingerprinting with K-means: the relevance of clusters visual inspection, in: Paper Presented at the 2022 CEUR Workshop Proceedings, Jan-2022, 2022.

[13] M. Aernouts, B. Bellekens, R. Berkvens, M. Weyn, A comparison of signal strength localization methods with sigfox, in: Paper Presented at the 2018 15th Workshop on Positioning, Navigation and Communications, WPNC, 25–26 October 2018, Bremen, Germany, 2018, http://dx.doi.org/10.1109/WPNC.2018.8555743.

[14] B. Soro, C. Lee, Joint time-frequency RSSI features for convolutional neural network-based indoor fingerprinting localization, IEEE Access 7 (2019) 104892, http://dx.doi.org/10.1109/ACCESS.2019.2932469.

[15] J. Jang, S. Hong, Indoor localization with WiFi fingerprinting using convolutional neural network, in: Paper Presented at the 2018 Tenth International Conference on Ubiquitous and Future Networks, ICUFN, 03–06 July 2018, Prague, Czech Republic, 2018, http://dx.doi.org/10.1109/ICUFN.2018.8436598.

[16] G. Zhang, P. Wang, H. Chen, L. Zhang, Wireless indoor localization using convolutional neural network and Gaussian process regression, Sensors 19 (2019) 2508, http://dx.doi.org/10.3390/s19112508.

[17] M. Stahlke, T. Feigl, M. Garcia, R. Stirling-Gallachery, J. Seitz, C. Mutschler, Transfer learning to adapt 5G AI-based fingerprint localization across environments, in: Paper Presented at the 2022 IEEE 95th Vehicular Technology Conference, VTC2022-Spring, 19–22 June 2022, Helsinki, Finland, 2022, http://dx.doi.org/10.1109/VTC2022-Spring54318.2022.9860906.

[18] S. Yang, C. Sun, Y. Kim, Indoor 3D localization scheme based on BLE signal fingerprinting and 1d convolutional neural network, Electronics 10 (2021) 1758, http://dx.doi.org/10.3390/electronics10151758.

[19] X. Song, X. Fan, C. Xiang, Q. Ye, L. Lius, Z. Wang, X. He, N. Yang, G. Fang, A novel convolutional neural network based indoor localization framework with WiFi fingerprinting, IEEE Access 7 (2019) 110698, http://dx.doi.org/10.1109/ACCESS.2019.2933921.

[20] D. Sun, E. Wei, S. Xu, Improving fingerprint indoor localization using convolutional neural networks, IEEE Access 8 (2020) 193396, http://dx.doi.org/10.1109/ACCESS.2020.3033312.

[21] Z. Liu, B. Dai, X. Wan, X. Li, Hybrid wireless fingerprint indoor localization method based on a convolutional neural network, Sensors 19 (2019) 4597, http://dx.doi.org/10.3390/s19204597.

[22] W. Njima, I. Ahriz, R. Zayani, M. Terre, R. Bouallegue, Deep CNN for indoor localization in IoT-sensor systems, Sensors 19 (2019) 3127, http://dx.doi.org/10.3390/s19143127.

[23] R. Sinha, S. Hwang, Comparison of CNN applications for RSSI-based fingerprint indoor localization, Electronics 8 (2019) 989, http://dx.doi.org/10.3390/electronics8090989.

[24] F. Qin, T. Zuo, X. Wang, CCpos: WiFi fingerprint indoor positioning system based on CDAE-CNN, Sensors 21 (2021) 1114, http://dx.doi.org/10.3390/s21041114.

[25] S. De-Bast, A. Guevara, S. Pollin, CSI-based positioning in massive MIMO systems using convolutional neural networks, in: Paper Presented at the 2020 IEEE 91st Vehicular Technology Conference, VTC2020-Spring, 25–28 May 2020, Antwerp, Belgium, 2020, http://dx.doi.org/10.1109/VTC2020-Spring48590.2020.9129126.

[26] Z. Pandangan, M. Talampas, Hybrid LoRaWAN localization using ensemble learning, in: Paper Presented at the 2020 Global Internet of Things Summit, GIoTS, 03 June 2020, Dublin, Ireland, 2020, http://dx.doi.org/10.1109/GIOTS49054.2020.9119520.

[27] G. Ferreras, M. Talampas, LoRa-based differential fingerprint localization in outdoor environments, in: Paper Presented at the 7th IEEE World Forum on Internet of Things, WF-IoT 2021, 14 June 2021 – 31 July, New Orleans, LA, USA, 2021, http://dx.doi.org/10.1109/WF-IoT51360.2021.9595880.

[28] Y. Li, J. Barthelemy, S. Sun, P. Perez, B. Moran, Urban vehicle localization in public LoRaWan network, IEEE Internet Things J. 9 (12) (2022) 10283–10294, http://dx.doi.org/10.1109/JIOT.2021.3121778.

[29] R. Muppala, A. Navnit, D. Devendra, E. Materay, N. Accetturaz, A. Hussain, Feasibility of standalone tdoa-based localization using LoRaWAN, in: Paper Presented at the 2021 International Conference on Localization and GNSS, ICL-GNSS, 01–03 June 2021, Tampere, Finland, 2021, http://dx.doi.org/10.1109/ICL-GNSS51451.2021.9452311.

[30] M. Chen, H. Zhao, C. Shi, X. Chen, D. Niu, Multi-scene LoRa positioning algorithm based on Kalman filter and its implementation on NS3, Ad Hoc Netw. 141 (2023) 103097, http://dx.doi.org/10.1016/j.adhoc.2023.103097.

[31] J. Liu, J. Gao, S. Jha, W. Hu, Seirios: Leveraging multiple channels for LoRaWAN indoor and outdoor localization, in: Paper Presented at the 27th Annual International Conference on Mobile Computing and Networking, January 31–February 4, 2022, New Orleans, LA, USA, 2022, http://dx.doi.org/10.1145/3447993.3483256.

[32] M. Aernouts, N. Bnilam, R. Berkvens, M. Weyn, TDAoA: A combination of TDoA and AoA localization with LoRaWAN, Internet of Things 11 (2020) 100236, http://dx.doi.org/10.1016/j.iot.2020.100236.

[33] A. Vazquez-Rodas, F. Astudillo-Sainas, C. Sanchez, B. Arpi, L. Minchala, Experimental evaluation of RSSI-based positioning system with low-cost LoRa devices, Ad Hoc Netw. 105 (2020) 102168, http://dx.doi.org/10.1016/j.adhoc.2020.102168.

[34] D. Guo, C. Gu, L. Jiang, W. Luo, R. Tan, ILLOC: In-hall localization with standard LoRaWAN uplink frames, ACM Interact. Mob. Wearable Ubiquitous Technol. 6 (1) (2022) 1–26, http://dx.doi.org/10.1145/3517245.

[35] A. Moradbeikie, A. Keshavarz, H. Rostami, S. Paiva, S. Lopes, A cost-effective lorawan-based IoT localization method using fixed reference nodes and dual-slope path-loss modeling, Internet Things 24 (2023) 100990, http://dx.doi.org/10.1016/j.iot.2023.100990.

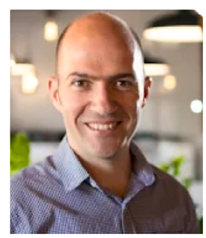
[36] E. Goldoni, L. Prando, A. Vizziello, P. Savazzi, P. Gamba, Experimental data set analysis of RSSI-based indoor and outdoor localization in LoRa networks, Internet Technol. Lett. 2 (1) (2019) http://dx.doi.org/10.1002/itl2.75.

[37] L. Marquez, M. Calle, Machine learning for localization in LoRaWAN: a case study with data augmentation, in: Paper Presented at the 2022 IEEE Future Networks World Forum, FNWF, 10–14 October 2022, Montreal, QC, Canada, 2022, http://dx.doi.org/10.1109/FNWF55208.2022.00081.

[38] H. Zhu, K. Tsang, Y. Liu, Y. Wei, H. Wang, C. Wu, H. Chi, Extreme RSS based indoor localization for LoRaWAN with boundary autocorrelation, IEEE Trans. Ind. Inform. 17 (7) (2021) 4458–4468, http://dx.doi.org/10.1109/TII.2020.2996636.

[39] G. Anagnostopoulos, A. Kalousis, A reproducible comparison of RSSI fingerprinting localization methods using LoRaWAN, in: Paper Presented at the 2019 16th Workshop on Positioning, Navigation and Communications, WPNC, 23–24 October 2019, Bremen, Germany, 2019, http://dx.doi.org/10.1109/WPNC47567.2019.8970177.
[40] I. Aqeel, M. Iorkyase, H. Zangoti, C. Tachatatzis, R. Atkinson, I. Aondonovic, Lorawan-implemented node localisation based on received signal strength indicator, IET Wirel. Sens. Syst. (2022) 1–16, http://dx.doi.org/10.1049/wss2.12039.
[41] H. Zhang, Z. Wang, W. Xia, Y. Ni, H. Zhao, Weighted adaptive KNN algorithm with historical information fusion for fingerprint positioning, IEEE Wirel. Commun. Lett. 11 (5) (2022) 1002–1006, http://dx.doi.org/10.1109/LWC.2022.3152610.
[42] A. Pimpinella, A. Redondi, M. Nikoli, M. Cesana, Machine learning based localization of LoRaWAN devices via inter-technology knowledge transfer, in: Paper Presented at the 2020 IEEE International Conference on Communications Workshops, ICC Workshops, 07–11 June 2020, Dublin, Ireland, 2020, http://dx.doi.org/10.1109/ICCWorkshops49005.2020.9145033.
[43] K. Islam, D. Murray, D. Diepeveen, M. Jones, F. Sohel, Machine learning-based LoRa localisation using multiple received signal features, IET Wirel. Sens. Syst. 13 (2023) 133–150, http://dx.doi.org/10.1049/wss2.12063.
[44] K. Mekki, E. Bajic, F. Chaxel, F. Meyer, A comparative study of LPWAN technologies for large-scale IoT deployment, ICT Express 5 (1) (2019) 1–7, http://dx.doi.org/10.1016/j.icte.2017.12.005.
[45] C. Milarokostas, D. Tsolkas, N. Passas, L. Merakos, A comprehensive study on LPWANs with a focus on the potential of LoRa/LoRaWAN systems, IEEE Commun. Surv. Tutor. 25 (1) (2022) 825–867, http://dx.doi.org/10.1109/COMST.2022.3229846.
[46] T. Gia, Z. Zou, H. Tenhunen, T. Westerlund, Comparative study of LPWAN technologies on unlicensed bands for M2M communication in the IoT: beyond LoRa and lorawan, in: Paper Presented at the 14th International Conference on Future Networks and Communications, FNC, 19–21 August 2019, 2019, http://dx.doi.org/10.1016/j.procs.2019.08.049.
[47] LoRa Alliance, What is LoRaWAN specification, 2023, https://lora-alliance.org/about-lorawan/. (Accessed 09 June 2023).
[48] X. Gong, X. Yu, X. Liu, X. Gao, Machine learning-based fingerprint positioning for massive MIMO systems, IEEE Access 10 (2022) 89320–89330, http://dx.doi.org/10.1109/ACCESS.2022.3199728.
[49] A. Khalajmehrabadi, N. Gatsis, D. Akopian, Modern WLAN fingerprinting indoor positioning methods and deployment challenges, IEEE Commun. Surv. Tutor. 19 (3) (2017) 1974–2002, http://dx.doi.org/10.1109/COMST.2017.2671454.
[50] A. Nessa, B. Adhikari, F. Hussain, X. Ferdando, A survey of machine learning for indoor positioning, IEEE Access 8 (2020) 214945–214965, http://dx.doi.org/10.1109/ACCESS.2020.3039271.
[51] C. Szegedy, Y. Liu, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Paper Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2015, pp. 1–9.
[52] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Paper Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2018, pp. 7132–7141.
[53] T. Monawar, S. Mahmud, A. Hira, Anti-theft vehicle tracking and regaining system with automatic police notifying using Haversine formula, in: Paper Presented at the 2017 4th International Conference on Advances in Electrical Engineering, ICAEE, 28–30 September 2017, Dhaka, Bangladesh, 2017, http://dx.doi.org/10.1109/ICAEE.2017.8255459.
[54] G. Anagnostopoulos, A. Kalousis, A Reproducible Analysis of RSSI Fingerprinting for Outdoor Localization Using Sigfox: Preprocessing and Hyperparameter Tuning, in: Paper Presented at the 2019 International Conference on Indoor Positioning and Indoor Navigation, IPIN, 30 September 2019 – 03 October 2019, Pisa, Italy, 2019, http://dx.doi.org/10.1109/IPIN.2019.8911792.
[55] T. Janssen, M. Aernouts, R. Berkvens, M. Weyn, Outdoor fingerprinting localization using sigfox, in: Paper Presented at the 2018 International Conference on Indoor Positioning and Indoor Navigation, IPIN, 24–27 September 2018, Nantes, France, 2018, pp. 1–6, http://dx.doi.org/10.1109/IPIN.2018.8533826.
[56] J. Torres-Sospedra, R. Montoliu, S. Trilles, O. Belmonte, J. Huerta, Comprehensive analysis of distance and similarity measures for Wi-Fi fingerprinting indoor positioning systems, Expert Syst. Appl. 42 (2015) 9263–9278, http://dx.doi.org/10.1016/j.eswa.2015.08.013.
[57] T. Janssen, Energy-efficient Positioning for the Internet of Things (Ph.D. thesis), Faculty of Applied Engineering, University of Antwerp, Belgium, 2023.

Albert Selebea Lutakamale received a B.Sc. in Electronic and Information Engineering at the Wuhan University of Technology, China, and MSc. in Information and Communication Science and Engineering, at Nelson Mandela African Institution of Science and Technology, Tanzania, in 2011 and 2017, respectively. He is currently working towards a Ph.D. degree in Electronic Engineering at the University of Pretoria, South Africa. His research interests are Wireless Communications, Fingerprinting Positioning/Localization, Massive MIMO networks, Internet of Things, LPWAN networks and Machine Learning.



Herman C. Myburgh obtained his B.Eng. in Computer Engineering, M.Eng. in Electronic Engineering and Ph.D. (wireless communication specialization) from the University of Pretoria, South Africa. He has been working at this institution in the Department of Electrical, Electronic, and Computer Engineering since 2009, where he is the head of the Advanced Sensor Networks research group. His current research interests are in wireless communication systems, sensor fusion, machine learning, and mobile health. He is the co-inventor of the hearScreen, hearZA, and hearScope smartphone-based hearing assessment solutions and is a co-founder of the hearX Group (Pty) Ltd company.

Allan De Freitas obtained the B.Eng., B.Eng. (Hons), and M.Eng. degrees in Electronic Engineering from the University of Pretoria, South Africa. He received a Ph.D. from the Automatic Control and Systems Engineering Department at the University of Sheffield, UK. He is a senior lecturer in the Department of Electrical, Electronic, and Computer Engineering at the University of Pretoria. His principal scientific interests are in the areas of signal processing and machine learning in object tracking, sensor networks, and complex systems.