Fingerprinting Based Positioning Techniques Using Machine Learning Algorithms Principles, Approaches and Challenges

Chapter · April 2022

DOI: 10.1201/9781003194538-6

5 authors, including:


Safar Maghdid Asaad
Koya University


Kayhan Zrar Ghafoor
Knowledge University


Halgurd Sarhang Maghdid
Koya University


Aos Mulahuwaish
Saginaw Valley State University

Fingerprinting Based Positioning Techniques Using Machine Learning Algorithms: Principles, Approaches and Challenges

Safar Maghdid Asaad[1, 2]

[1] Department of Information System Engineering Techniques, Erbil Technical, Engineering College, Erbil Polytechnic University, Erbil, Kurdistan Region-F.R., Iraq. Email: safar.dei20@epu.edu.iq

[2] Department of Software Engineering, Faculty of Engineering, Koya University, Koya KOY45, Kurdistan Region-F.R., Iraq. Email: safar.maghdid@koyauniversity.org

Kayhan Zrar Ghafoor

Department of Computer Science, Knowledge University, University Park, Kirkuk Road, 44001 Erbil, Iraq. Email: kayhan.zrar@knu.edu.iq

Halgurd Sarhang

Department of Software Engineering, Faculty of Engineering, Koya University, Koya KOY45, Kurdistan Region-F.R. Iraq. Email: halgurd.maghdid@koyauniversity.org

Aos Mulahuwaish

Department of Computer Science and Information Systems, Saginaw Valley State University MI, USA. Email:amulahuw@svsu.edu

Abbas M. Ali

Department of Software Engineering, Salahaddin University-Erbil, Iraq.
E-mail: abbas.mohamad@su.edu.krd

Abstract

The demand for using location-based services (LBSs) is rapidly increased, specifically in the last decade. Most people's daily activities are related to LBS services, including navigation, billing address, tracking stuff, transportation, and other point-of-interest (POI). In the same manner, many solutions are widely available to process the positioning from outdoors to indoors. One of the most utilized positioning solutions is using fingerprinting-based techniques via different technologies, including WiFi, Bluetooth, 3G/4G, and UWB. Many attempts have been made to enhance fingerprinting-based positioning and then to provide an accurate solution. The recent attempts are referred to use modern machine learning algorithms as fingerprinting matching process. However, there is no single solution to provide an accurate, low-cost, on-the-go, and seamless positioning solution. Therefore, this article aims to address the issues of using fingerprinting-based positioning. A new taxonomy for the recent solutions, which are related to fingerprinting-based techniques, is also designed. Accordingly, machine learning algorithms which have been used in fingerprinting-based technique and their challenges are investigated.

Keywords: *Indoor Localization, Fingerprint, Machine Learning, Indoor Positioning Systems, Inertial Navigation.*

I. Introduction

Nowadays, one of the world's most prominent innovations is LBSs. Global navigation satellite systems (GNSS) [1]–[4] including (1) Global positioning systems (GPS), (2) BeiDou navigation satellite systems (BDS), (3) GLONASS, and (4) Galileo are applicably utilized when users outdoors. Due to suffering GNSS from obstruction of signals by objects like trees, roofs, walls, buildings, their performances deteriorate, and they cannot be used in the indoor environment. To fill this gap in urban or indoor environments, numerous researches have been conducted recently. Among these researches, many technologies have attracted the researchers to propose new indoor localization solutions based on some available technologies including, ZigBee, cellular, Bluetooth, infrared, Ultra-Wide Band (UWB), radio frequency identification (RFID), Micro-electromechanical systems (MEMS), geomagnetic field, pseudo files (PL) and Wireless Fidelity (WiFi).

Because of the availability of WiFi signals in almost all indoor environments for internet accessing and the increasingly growing up smartphone users, several measurements, including (1) angle-of-arrival (AoA) [5], (2) received signal strength (RSS) [6], and (2) time-of-arrival (ToA) [7] have been derived from establishing fingerprint databases. Besides, based on these fingerprint databases, many algorithms have been designed to estimate indoor locations accurately. Typical fingerprinting positioning techniques consist of two phases, which are Offline and online phases, as illustrated in Figure 1. In the first phase, the fingerprint database is constructed, containing the measurements derived from the power of WiFi signal values in preferred points via the applicable device. After that, to represent the mapping relationship between signal information and positions, a position model can be trained based on the collected data. After that, the positioning model can be used to predict the desired point's location during the online process.

Most of the current indoor WiFi localization techniques concentrated on fingerprints based on the received signal strength (RSS) [8]–[11]. This is because; the RSS can be extracted from some common WiFi devices. However, the RSS signal is sensitive to the environment changing and heterogeneous hardware. To overcome the drawbacks of the technologies and fingerprinting data, machine learning algorithms have been applied to improve the mapping accuracy. However, to find the drawbacks and suggest new models, and improve the performance of indoor positioning solutions, such research needs to be surveyed. Therefore, this article aims to compare existing approaches and solutions in indoor positioning using fingerprinting techniques via machine learning algorithms.

Therefore, the main contribution of this chapter is to present recent fingerprinting-based positioning techniques via using Machine Learning Algorithms. The chapter also investigates the issues of WiFi indoors positioning in terms of time and space complexity, Big data, privacy, and security. Finally, the main purpose of this chapter is to make a guideline for the reader to know the recent progress of indoors positioning challenges and approaches.

The remainder of the chapter is structured like the current fingerprinting-based positioning technique using machine learning algorithms is discussed in section II. This is accompanied by a contrast between all the solutions grouped into classical techniques and Modern Techniques categories. When advanced machine learning methods are used, section III addresses the recent research challenges and potential solutions. Finally, this research is concluded in section IV.

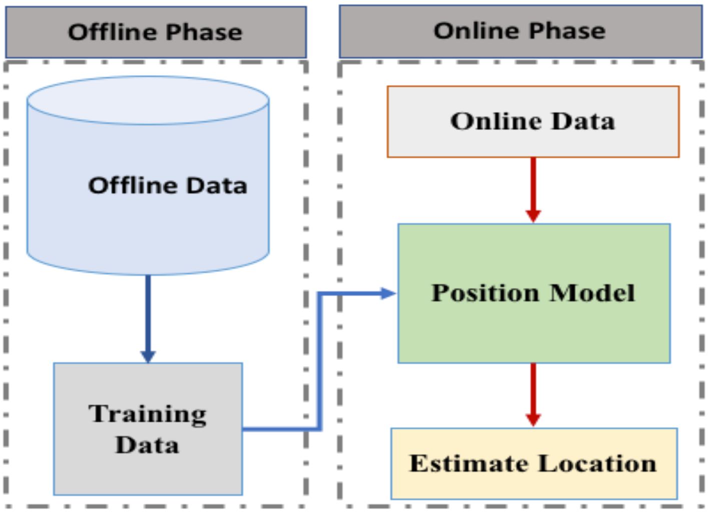

Figure 1: Sample Structure of Fingerprint-based Indoor Positioning System

II. Literature Review

The increasing expectations for applications based on LBSs have encouraged investigators to introduce a pool of opinions to ensure these systems' effectiveness. Besides, these concepts are not without limitations. Thus, so many machine learning techniques are applied to improve them. The existing solutions that propose using machine learning algorithms to provide more significant relativity are examined in this research to highlight the available challenges and provide recommendations to be explored in future works. As shown in Figure 2, the current Machine learning algorithms for the purpose of localization are categorized into two categories, which are (1) classical algorithms and (2) modern algorithms.

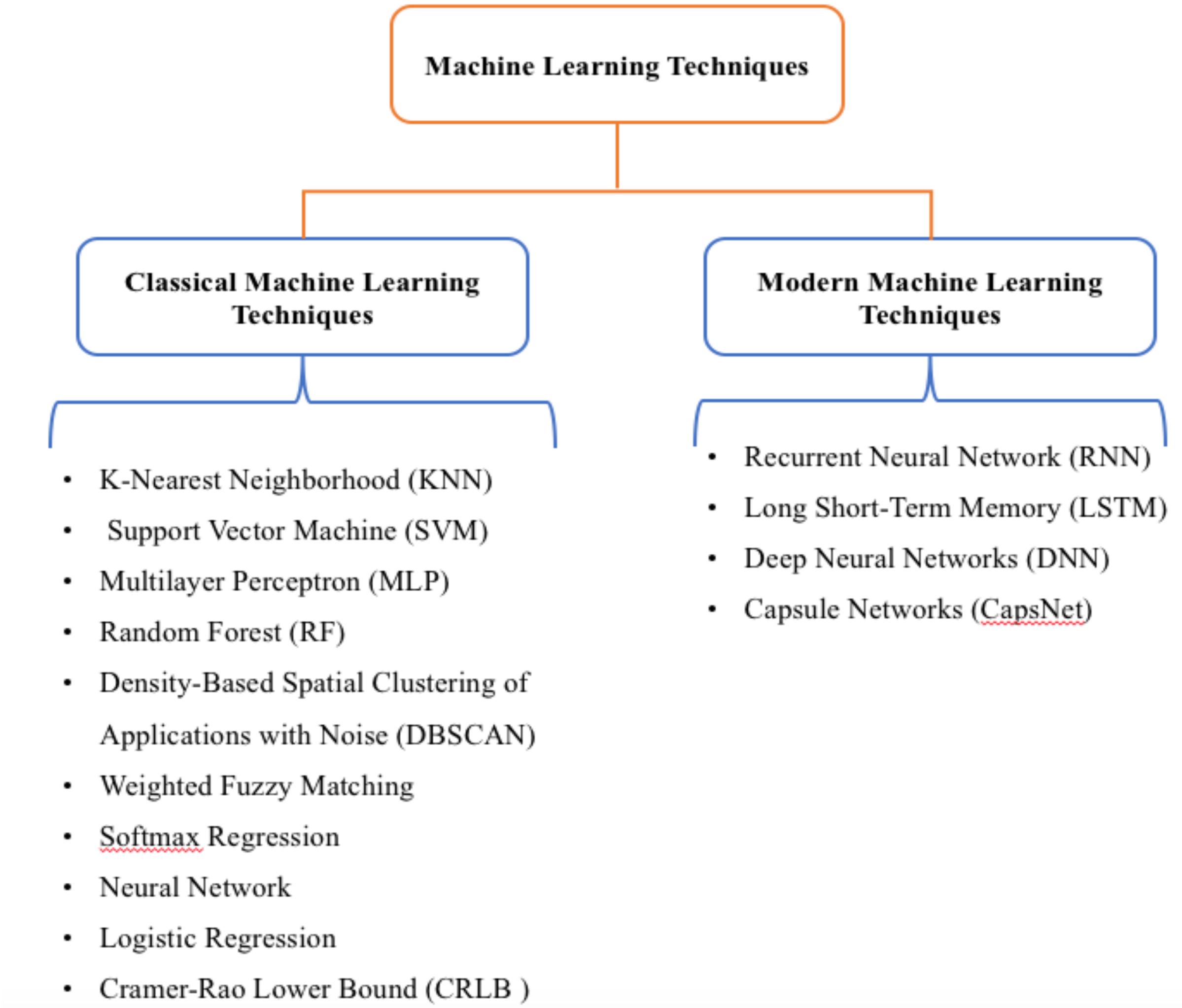

Figure 2: Taxonomy of Machine Learning Algorithms for the purpose of Localization

1. Classical Machine Learning based Solutions

Machine learning algorithms are still used in various fields, including tracking systems, image recognition, speech recognition, recommendation systems, and so on. Further, because of their effectiveness in solving available problems. As presented in Table 1, the latest localization solutions based on classical machine learning techniques are explained.

Walls and many obstacles in the indoor environment are some factors that affect the accuracy of indoor navigation. Therefore, many applications have been used to tackle these types of issues. These kinds of applications are based on communication technologies, including the Internet of Things (IoT), Artificial Intelligence (AI), Bluetooth, UWB, and WiFi. The authors in [16] proposed an algorithm based on WiFi technology to improve indoor location tracking accuracy. Essentially, the fingerprinting database is established via measuring and collecting the RSSI values for each RP. Afterward, a weighted fuzzy matching algorithm is used to track the user's location. The weighted fuzzy matching algorithm matches the user's RSSI value with the pre-stored RSSI values in the fingerprinting database. Finally, Particle Swarm Optimization (PSO) algorithm is applied to improve indoor localization accuracy further. The study performs a simulation with assuming (8m x 8m) empty room with four WiFi Access Points (APs) at each vertex. The Simulation results verified that the accuracy of the location tracking could be improved when weighted fuzzy matching and PSO algorithms are used together. Approximately obtained results determine that the average position error is about 2m when the PSO algorithm is not applied, while the average position error is about 1.2m when the PSO algorithm is applied. However, only 4 APs are used, which are not enough for real environments. The multi-floor, environmental movement, and different object availability concerns have remained as future challenges.

In a similar study, authors in [17] investigated a new indoor localization prototype based on the WiFi (RSS) fingerprinting, but with employing Machine learning techniques. In the offline phase, the authors measured the RSSI values of MikroTik APs signals via Smartphones to establish the fingerprint database. Later on, in the online phase, the proposed prototype uses various classifiers including, (1) Support Vector Machine (SVM). (2) K-Nearest Neighborhood (KNN), and (3) Multilayer Perceptron neural network (MLP). Figure 3 presents the block diagram of the proposed prototype. The classifiers are trained and tested in the same training dataset. A part of the dataset is used as a validating dataset to assess the achievement of the proposed prototype. The experimental area is the third floor of a building with placing five APs. The experiments' obtained results express that the performance of the KNN is better than the other classifiers. However, the KNN requires an absolutely large database fingerprint, and its performance deteriorates when the RSS values fluctuate.

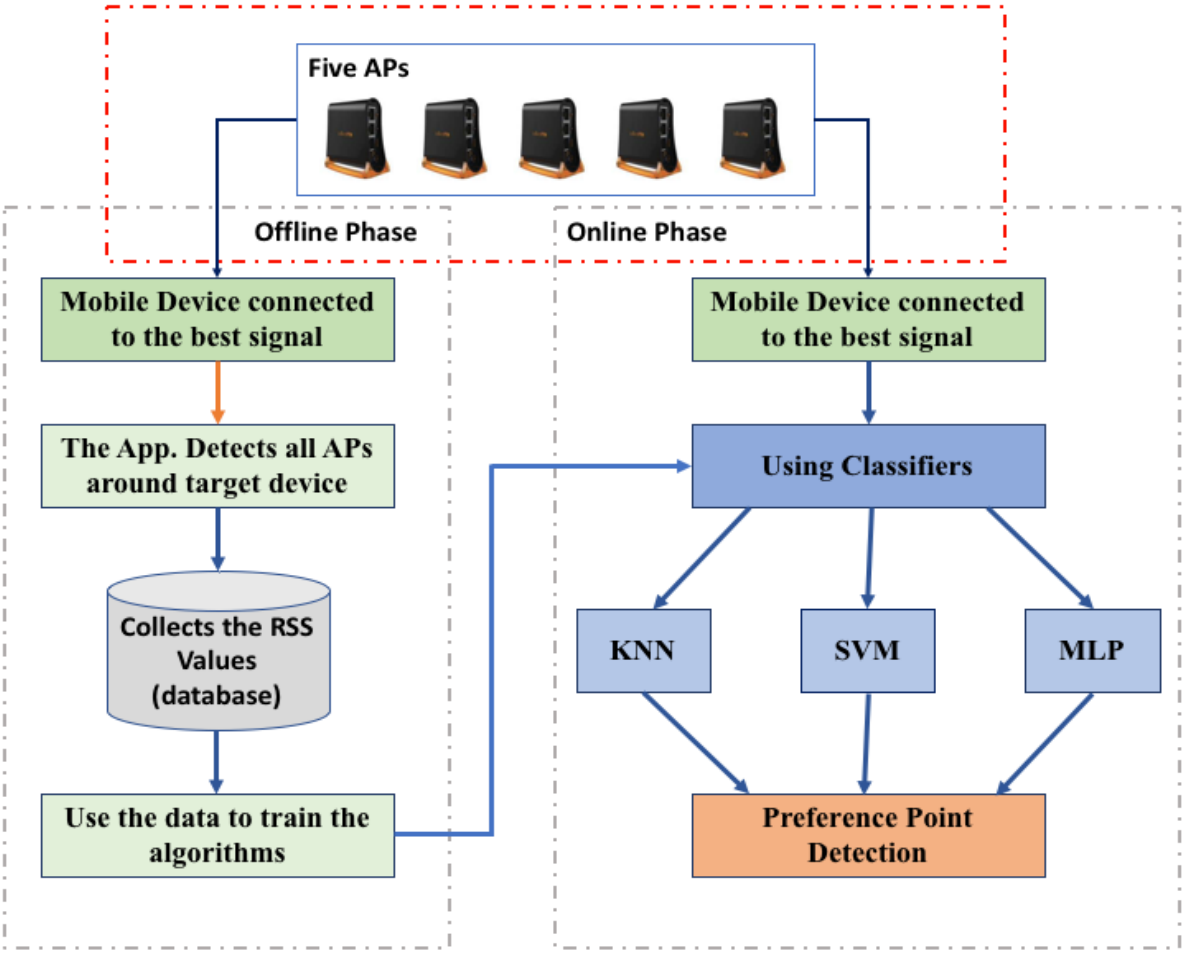

Figure 3: Block diagram of the proposed prototype

Since the instability of Wi-Fi signals, the conventional distance-based calculation techniques fail to evaluate the RSS fingerprints of some adjacent reference positions accurately. KNN algorithm cannot obtain accurate positioning when RSS fingerprints are clustered into only one region. To this end, the authors in [18] propose a new algorithm named DBSCAN-KRF and a new concept called "the insensitive region of the RSS fingerprint." The DBSCAN-KRF is an integration of three models. (1) **"Density-Based Spatial Clustering of Applications with Noise (DBSCAN)"** model. It is used to detect the RSS fingerprint signal's insensitive region from the fingerprint Library and delete the noise sample. (2) The **KNN** model is used to obtain the positioning when the region is sensitive. (3) **"Random Forest (RF)"** model is selected to obtain the positioning when the region is insensitive. The study conducted many experiments considering two important factors that affect the accuracy of indoor positioning models. These factors are the number of APs and the number or density of RSS records. The experimental findings present that the DBSCAN-KRF algorithm's accuracy is notably higher than the baseline algorithms when the number of APs and the quantity of RSS samples in each reference point are 8 and 40, respectively. However, with owning the simplicity characteristic, the performance of WKNN declines with RSS fluctuations, and the WKNN requires a massive dataset. Additionally, the change of environment, the positioning target movement, multi-floor localization, and the positioning environment factors' area size are not considered in the proposed algorithm.

Another study [19] focused on received signal strength (RSS) from cellular signals to build an indoor localization model. The study employs Weighted K-Nearest Neighbor (WKNN) to split the large cluster into tinier ones. Then, WKNN is integrated with a multi-layer neural network to benefit from the robust clustering ability of WKNN. The flowchart of the suggested model is shown in figure 4. Another benefit of the integrated model is that the implemented multi-layer neural network can predict the indoor location within each cluster. Three experiments are conducted, including two urban environments and an individual rural environment. The obtained results are 5.9 m, 5.1 m, and 8.7 m in the urban environments and the rural environment, respectively, in mean distance localization error. These results present that the accuracy of the integrated approach is better than the WKNN-only algorithm. However, the cellular base transceiver station (BTS) and RSS mathematical model are not considered.

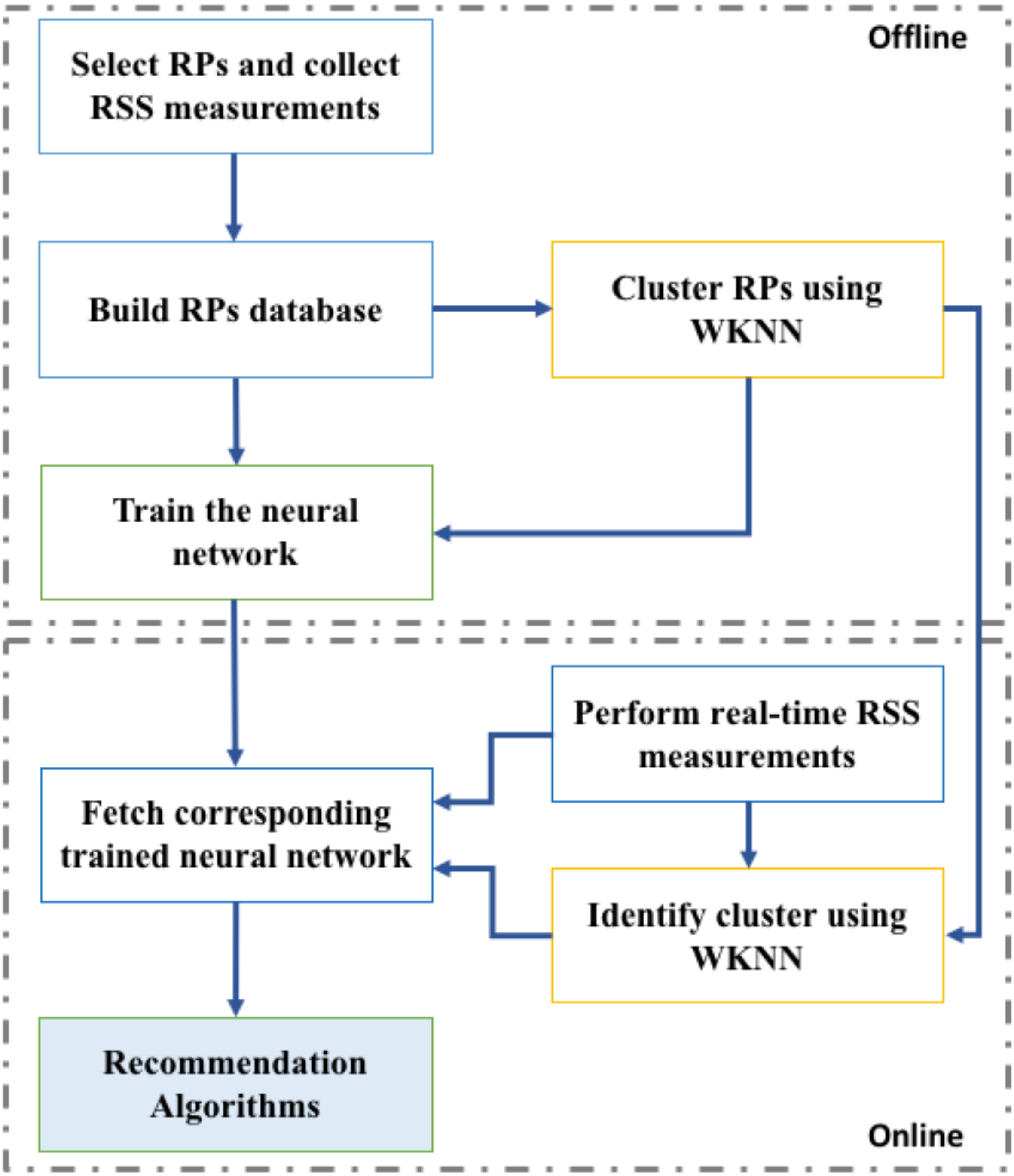

Figure 4: Flowchart of Proposed Model with WKNN and Neural Network

Cost and installation are other issues in indoor positioning systems. The studies in [20] used Bluetooth Low Energy (BLE) for fingerprinting-based indoor positioning. That is because of the easiness of use and the low-cost characteristics of Bluetooth. The authors prepared the fingerprint database by deploying 14 Bluetooth beacons in a corridor. The RSSI values are collected from the k nearest BLE devices to a mobile device. The architecture of the proposal is demonstrated in Figure 5. Accordingly, two popular machine learning algorithms are implemented to evaluate the accuracy of indoor positioning estimation, including (1) SVM and (2) logistic regression. The achieved results from a set of trial experiments show that the SVM technique is more accurate than Logistic Regression. Approximately, the average error of SVM is 50cm. In contrast, the average error of logistic regression is 90cm. However, the performance of the SVM is limited to such big and complex data. Moreover, other concerns are overlooked, including the environment movement objects and multi-floor localization.

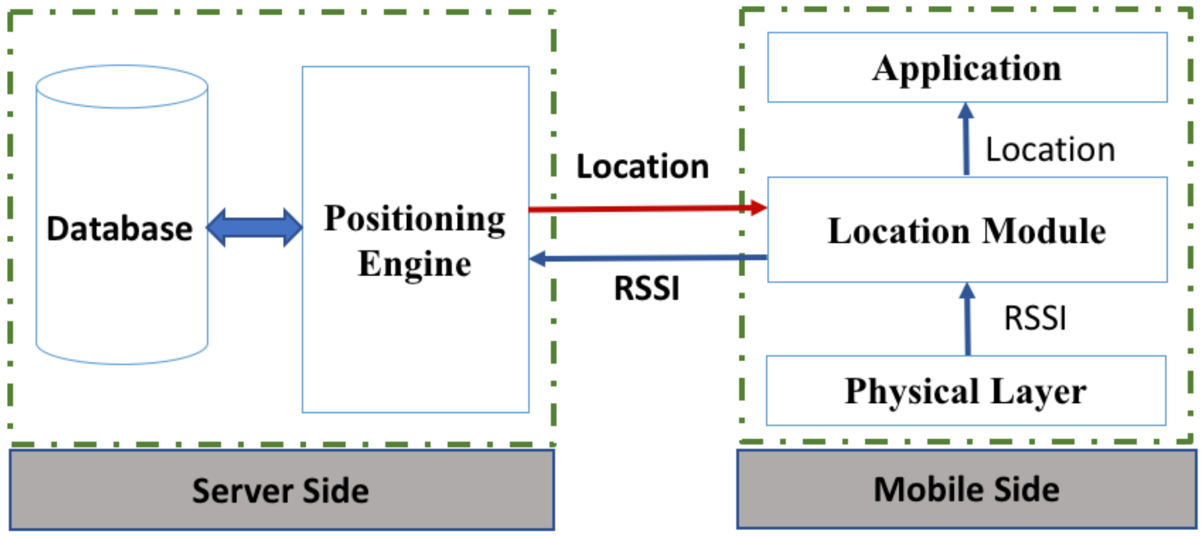

Figure 5: Architecture of the Proposed Model

The millimeter-wave (mmWave) has some characteristics that can help reduce the positioning error in an indoor environment. These characteristics are narrow beam, rapid signal attenuation, broad bandwidth, and so forth. A new location fingerprint positioning solution based on the narrow beam and rapid signal attenuation characteristics of mmWave is proposed in [21]. The solution is called (DoA-LF). The fingerprint dataset of DoA-LF contains RSSI information and direction of arrival (DoA) information of APs operating on the mmWave spectrum, which is achieved via the multiple signal classification (MUSIC) method. Figure 6 depicts the proposed DoA-LF solution.

Subsequently, the weighted K nearest neighbor (WKNN) technique is engaged to calculate the weighted mean of the chosen K candidate RPs with the most similar features. Afterward, Cramer-Rao Lower Bound (CRLB) tool is applied to analyze the influence of the quantity of APs, the interval of RPs, the channel type of mmWave, and the failure of the DoA prediction model on positioning error. The simulation results from a 100m x 100m indoor environment are verified that (1) positioning error is minimized notably when an optimal K is selected (2) the DoA information of APs working on mmWave spectrum band can significantly reduce the positioning error in comparison to fingerprinting based positioning with signal below 6 GHz (3) since mmWave has a more extensive path loss exponent and more petite variance of shadow fading contrasted with low-frequency signals, the mmWave can be utilized to decrease the positioning error.

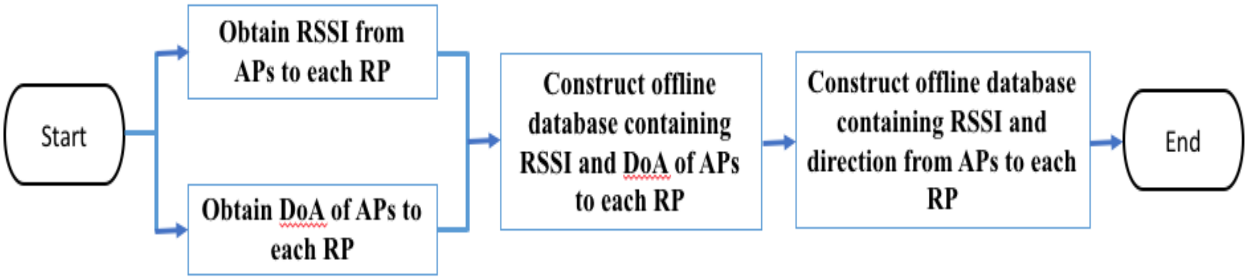

Figure 6: Proposed DoA-LF Solution

Table 1 Comparisons of Existing Indoor Localization Solutions Based on the Classical Machine Learning Algorithms

Year	Solutions	Fingerprint	Methods	Environments	Accuracy	Characteristics
2020	[16]	WiFi (RSS)	- Weighted fuzzy matching - PSO	(8m x 8m) Simulation empty Room	1.2m	Combined weighted fuzzy matching and PSO algorithms to achieve higher position accuracy.
2020	[17]	WiFi (RSS)	- KNN - SVM - MLP	(3170 M) floor	- 1.2m - 5.88m - 3.06m	The performance of the localization is further improved by training multiple classifiers.
2019	DBSCAN-KRF [18]	WiFi (RSS)	- DBSCAN - KNN - RF	- Two (8.8m × 5.6m) offices - Two corridors	N/A	- DBSCAN detect insensitive region of the RSS fingerprint signal from the fingerprint Library and delete the noise sample - KNN and RF obtain the positioning when the region is sensitive and insensitive respectively.
2018	[19]	Cellular (RSS)	- WKNN - Neural Network	- Two urban environments - One rural environment	- 5.9 m - 5.1 m - 8.7 m	- division the environment into small clusters - estimating the position within each cluster
2018	[20]	Bluetooth (RSS)	- SVM - Logistic regression	Corridor	50 cm	Used Bluetooth as low cost and ease of use technology.
2017	DoA-LF [21]	WiFi (RSS + DoA Info of APs)	- WKNN - CRLB	(100m × 100m) Simulation indoor environment.	N/A	- DoA information of APs working on mmWave spectrum reduce the positioning failer in comparison to fingerprinting based positioning with signal below 6 GHz. - The mmWave can reduce the positioning error.

2. Modern Machine Learning based Solutions

Due to existing challenges associated with classical machine learning techniques, a range of solutions are proposed to improve the accuracy of indoor positioning based on modern machine learning algorithms. The comparisons between existing solutions using advanced machine learning algorithms are provided in Table 2.

Classical matching algorithms, utilized in Fingerprinting technique, regularly suffer from restricted performance in analyzing complex and noisy data. Equally, deep learning algorithms are more powerful to analyze very complicated and noisy values. Therefore, a local feature-based deep long short-term memory (LF-DLSTM) is proposed for this purpose [22]. Classical localization systems collect the RSSI data via smartphones. The smartphones are used to scan for WiFi routers signals and collect RSSI data. However, within the initial step, the proposed algorithm faced the problems of low sampling rate and high battery-power drain of smartphones. To address this issue, the authors are focused on the advantages of a passive scanning system. A huge amount of RSSI data is collected in a passive scanning system by employing WiFi routers for scanning smartphones with a high sampling rate. The proposed technique reduces the noise effect in RSSI and obtains robust local features from the row RSSI data via the local feature extractor. Accordingly, for final accurate localization, the DLSTM network is applied to encode temporal dependencies and generate a more discriminatory representation. In the proposed model, two practical experiments have been carried out. The first one has been in a research lab, and the second one has been in an office. The results present that the proposed model outperforms the state-of-art methods for indoor localization. The achieved localization performance of mean localization errors is 1.48 and 1.75 meters within the research laboratory and office environments. However, multi-floor localization and change of environmental concerns have not been addressed.

Authors in [14] used WiFi-RSS in indoor environments to suggest a modern WiFi-based indoor localization method known as WiFiNet. The proposed method aims to get advantages of Convolutional Neural Networks' superior capability to solve classification problems. The authors compare three strategies, a new architecture called WiFiNet that developed and built specifically to address this challenge, and the most common pre-trained networks that use both transfer learning and feature extraction. Furthermore, the top-performing traditional architectures for WiFi-based indoor localization were tested as a benchmark. The experiments were conducted in a realistic environment, 3600 m2, at a university campus. WiFiNet achieves the greatest generalization and adaptation to practical environments, with a Root Mean Square Error (RMSE) of 3.3 m when measured during walking around the real environment. The processing time in a medium area (30 locations and 113 APs) notably reduced as compared with baseline WiFi indoor positioning algorithms, including SVM. However, with the increasing number of APs, the processing time is increasing as well.

Another proposed work [23] is known as the CapsLoc system, as presented in Figure 7. The CapsLoc is based on the RSS fingerprinting data to estimate indoor locations. Authors employ Capsule Networks (CapsNet) by considering the fingerprinting wireless positioning scheme to predict indoor positions accurately. The CapsNet model extracts a convolutional layer, a primary capsule layer, plus a feature capsule layer from WiFi fingerprint to establish its hierarchical structures. Real experimental tests are conducted on a corridor with three IoT labs on the third floor using 6 APs (two APs in each lab) to deal with an area of 460m2. The achievements indicate that applying the CapsLoc model can provide better indoor positioning accuracy with an averaged error of 0.68 m. The results also express that the model outperforms KNN, SVM, and CNN.

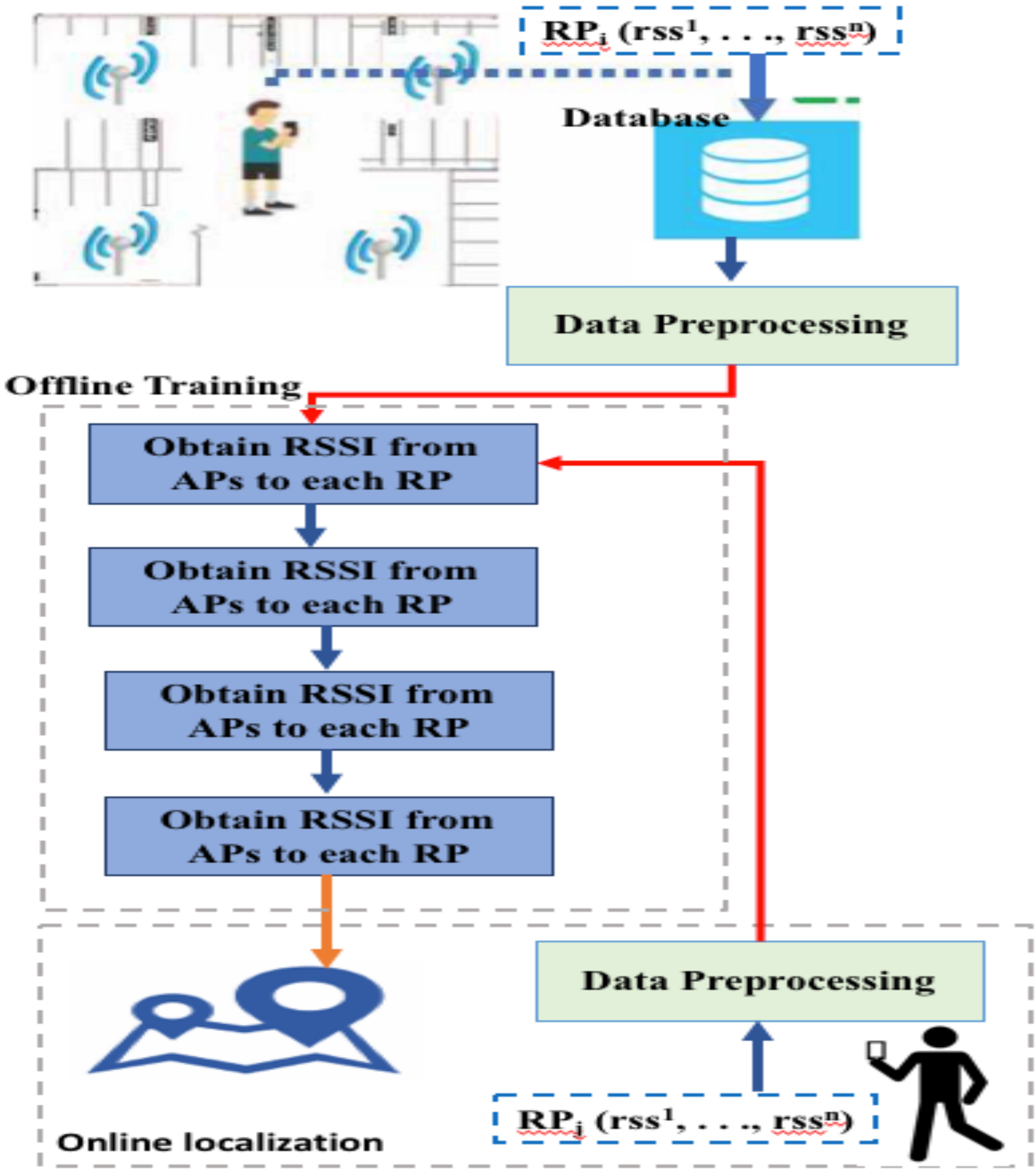

Figure 7: The Framework of the CapsLoc System

Another paper [24] presented a deep learning-based Wi-Fi and magnetic field fingerprint-based localization framework. Since magnetic field strength (MFS) is difficult to discriminate over wide regions, the unsupervised learning density peak clustering algorithm is initially utilized to choose multiple MFS center points as geotagged features to aid indoor positioning. Where the unsupervised learning density peak clustering algorithm is based on the comparison distance (CDPC) technique. Authors create a position fingerprint picture for localization using Wi-Fi and magnetic field fingerprints, considering the state-of-the-art use of deep learning in image classification. The suggested deep residual network (Resnet), which is smart enough to learn key features from a large fingerprint image dataset, is used to perform the localization. An MLP-based transferred learning perfect localizer is applied to refine the pre-trained Resnet coarse localizer by using the prior knowledge of the pre-trained Resnet coarse localizer. Many data enhancement techniques have used and dynamically modified the learning rate (LR) to improve the robustness of our localization scheme. The proposed method achieves adequate localization efficiency including both indoor and outdoor settings, according to experimental findings.

One of the main issues of fingerprinting-based localization systems is to extract valid features from the utilized technologies. This is to establish the fingerprint database that contains the mapping between the Reference Points and their corresponding signal features. Authors in [25] proposed a hierarchical localization approach based on multipath “multiple-input multiple-output (MIMO) channel state information (CSI)” fingerprints. The approach is based on time domain-based multipath MIMO CSI instead of the frequency domain-based multipath MIMO CSI. Two deep neural networks (DNN) are trained for coarse and fine positioning in the offline phase. The coarse positioning is used to reduce the training time when the targeted region is wide. A softmax regression classifier follows this to generate soft information on the classified output data. Two machine learning algorithms, softmax regression classifier, and K, weighted the nearest neighbor (KWNN), enhanced localization accuracy in the online phase. The simulation outcomes demonstrate that the localization accuracy can be increased when multipath MIMO CSI in the time domain is considered instead of the frequency domain. As well as, the coarse-to-fine process can obtain a more reliable performance compared to the present approaches.

Several factors influence the localization accuracy of the existing Radio Frequency (RF) fingerprinting-based techniques such as WiFi, Bluetooth Low Energy (BLE). These factors include (1) the number of available AP, (2) physical obstructions and complexity of the indoor space, (3) moving objects in the environment, and (4) physical scale. Specifically, the value of the RSS fluctuates when the indoor space becomes complicated, and the physical scale is enlarged. To overcome the issue of RSS fluctuation, the authors in [26] proposed a solution based on the geomagnetic field signal for indoor positioning instead of the RF signal. Since the geomagnetic sensor signal is available everywhere in the environment, the objects have a unique sequence of the geomagnetic field signals as long as the objects are moving. The proposed work establishes the fingerprinting database via collecting geomagnetic field signals in many locations from two large testbed environments. The first environment is the first floor of Incheon International Airport, with a selected (608m x 50m) area known as a large-scale testbed. The Second environment is the first floor of the Korea University Science and Engineering campus. The testbed size is (94m x 26m) area, and it is called a medium-scale testbed. Collecting the geomagnetic field stable signal is more manageable than collecting the unstable radio signals from multiple beacons or APs. Two DNN models, including basic recurrent neural network (basic RNN) and Long Short-Term Memory (LSTM), are applied to map between a precise position and the previously collected geomagnetic field signals. After training the DNNs, the obtained results present that achieved average localization accuracy via LSTM is notably higher than the RNN in both environments. The achieved results are 1.2m, 4.1m, 0.51m, and 1.04m with RNN and LSTM, respectively, in both environments medium-scale and large-scale testbeds. However, the proposed technique might suffer from measuring the same geomagnetic field signal value that might arise in multiple locations. That is due to installing differently characterized magnetometers by manufacturers in the devices, including smartphones. Furthermore, the magnetometers are disturbed by several factors including, electronic devices and ferromagnetic architectural features.

Table 2 Comparisons of Existing Indoor Localization Solutions Based on the Modern Machine Learning Algorithms

Year	Solutions	Fingerprint	Methods	Environments	Accuracy	Characteristics
2021	WiFiNet [14]	WiFi (RSS)	CNN	($3600m^2$) real world environment (university area)	3..3 m	- Achieves the greatest generalization and adaptation to practical environments, with an RMSE of 3.3 m when measured during walking around the real environment. - The processing time in a medium area (30 locations and 113 APs) as opposed to baseline WiFi indoor positioning algorithms including SVM.
2020	LF-DLSTM [22]	WiFi (RSS)	DLSTM	- Lab - Office	- 1.48 m - 1.75 m	Reduces the noise effect in RSSI and obtains robust local features from the row RSSI data via the local feature extractor.
2020	CapsLoc [23]	WiFi (RSS)	CapsNet	($460m^2$) a corridor with three IoT labs	0.68m	- extracting higher level features from the WiFi RSS fingerprint dataset via CapsNet.
2020	[24]	Wi-Fi (RSS) Magnetic field	- Resnet	- Indoor and outdoor test sites (divided into dozens of grids	- 97.1 %	- It is smart enough to learn key features from a large fingerprint image dataset.
2019	[25]	CSI	- DNN - Softmax regression - KWNN	(40m X 40m) urban microcell environment	0.8349m	- Depending on "multipath MIMO CSI in the time domain" instead of the frequency domain, and using coarse-to-fine process increases the localization accuracy.
2019	[26]	Geomagnetic sensor signal	- Basic RNN - LSTM	- (94m x 26m) indoor testbed - (608m x 50m) indoor testbed	- 1.2m, 4.1m - 0.51m, 1.04m	- Geomagnetic signal is available everywhere in the environment. - Objects have a unique sequence of the geomagnetic field signals as long as the objects are moving. - Data collection process based Geomagnetic signal on is simpler than to be based on the RF signal.

III. Challenges

In this section, the most common challenges that face the researchers who are have been working on localization, specifically fingerprint-based localization systems, are addressed.

A. Time Consuming

Constructing the fingerprint database and training the model takes a long time. Consequently, the number of measurements and the size of the reallocation environment, and the time to predict the intended locations are increasing linearly. We consider that by monitoring user trajectories, data can be obtained dynamically, but other procedures are further required to minimize the training period. Using crowd-sourcing technology as the method for solving this issue would be superior.

B. Environment Complexity

One of the necessities for positioning based on radio frequency signals is line-of-sight communication connecting transmitter and transceiver. Sometimes signals attenuate through occlude obstacles. Besides, the signals might be prevented by obstacles and cannot be accessed by the receiver. For example, the WiFi Access Point signal cannot penetrate an obstacle to access the receiver. To tackle these issues, high-precision receivers are required to be utilized, which maximizes the indoor positioning technology cost but adds to the accuracy of positioning within tens of meters. Other points affect the signal's strength, including the spatiotemporal changes of signals and the geometric changes of the environment.

C. Storage Capacity and Computing Power

Other limitations belong to the receiver devices, including storage size and computing power. The recording of RSS from many nearby APs and storing this data in a database along with the client device's identified locations in an offline phase is the basic process for wireless fingerprinting-based localization. Big data would be collected in a big enough space for an appropriate amount of APs, which necessitates a very large storage capacity, which linearly raises the expense and complicates the computing method. Furthermore, the capacity and processing ability of the clients' devices have a significant impact on the success of the proposed solutions. For example, many algorithms are implemented on smartphones, and their computing power is varied according to their model and manufacture. Therefore, the positioning accuracy is linearly increased when the more powerful smartphone is used to estimate the local position based on the machine learning techniques. In terms of energy consumption, clients' electronic devices cannot continue to execute position algorithms at high speeds, so the battery will easily drain. Remote terminal processing activities can be offloaded to local cloud data centers owing to IoT technology. Smartphones are only utilized to gather signals and obtain positioning results in order to prolong their life cycle. However, the other big data issue could be handled via implementing a distributed database system. This is to solve the issue of reducing the time delay of retrieving location information and avoiding space complexity. Since the data is distributed in many servers/sources, the users will easily contact the servers/sources to export the required location information.

D. Security and privacy

Real-life individuals gain great benefits from Location-based services, but still, some issues have faced the individuals. When the background applications collect the location information from the individuals, this information might be used to reveal their privacy for attackers' benefit. Additionally, some proposed indoor positioning algorithms utilize cloud computing to store the fingerprint data and store the implemented model. In that case, the related positioning data must be transmitted between the user's phone and the cloud. During this transmission process, there may be some vulnerable points to attack, specifically MITM. To overcome these vulnerabilities, robust cryptographic approaches have to be implemented to encrypt the data before being transmitted. So, the service provider should implement protecting algorithms the prevent the man-in-the-middle attack (MITM). Usually, MITM is defended by authentication techniques such as Public Key Infrastructures (PKI) mutual authentication, stronger mutual authentication like secret keys or passwords. If the attacker knows secret key information, although the cryptography mechanism is done, it is useless. So secret key management also a very important issue [15].

To propose a robust solution, many machine learning algorithms have been suggested to establish a model to predict a real-time indoor location with acceptable accuracy. Some of them require a huge amount of data to be trained. To this end, the time and memory complexity are increased sine more times are taken to train them.

IV. Conclusion

Recently, LBSs attracted the attention of a huge number of researchers. This is because the services can be used in a variety of applications, both indoors and outdoors. Correspondingly, many solutions have been conducted to improve the positioning accuracy, specifically for indoor environments. Using fingerprinting-based techniques through various technologies, including WiFi, Bluetooth, Geomagnetic, 3G/4G, and UWB, is one of the most used positioning solutions. Several attempts have been made to improve positioning based on fingerprinting and then to provide reliable solutions. A detailed overview of the recent fingerprinting-based position techniques using machine learning algorithms is presented in this work. The early achievements are referred to as the fingerprinting matching method using advanced machine learning algorithms. However, to provide reliable, low-cost, on-the-go, and seamless positioning solutions, there is no single solution. This article, therefore, attempted to address the issues of using positioning based on fingerprinting. The issues involve the storage size of the receiver device, computing power, and security aspects issues. A new taxonomy is also being developed for the latest solutions relating to fingerprinting-based techniques. Machine learning algorithms that have been used in fingerprinting-based techniques are then investigated, and their limitations are discussed.

References

[1] C. Tiberius and E. Verbree, "GNSS positioning accuracy and availability within Location Based Services: The advantages of combined GPS-Galileo positioning," *NaviTec*, no. 1, 2004.

[2] C. Stallo *et al.*, "GNSS-based location determination system architecture for railway performance assessment in presence of local effects," in *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, Apr. 2018, pp. 374–381, doi: 10.1109/PLANS.2018.8373403.

[3] K. Zhang, M. Spanghero, and P. Papadimitratos, "Protecting GNSS-based Services using Time Offset Validation," in *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, Apr. 2020, pp. 575–583, doi: 10.1109/PLANS46316.2020.9110224.

[4] J. Paziewski, "Recent advances and perspectives for positioning and applications with smartphone GNSS observations," *Meas. Sci. Technol.*, vol. 31, no. 9, 2020, doi: 10.1088/1361-6501/ab8a7d.

[5] S. Tomic, M. Beko, R. Dinis, and L. Bernardo, "On target localization using combined RSS and AoA measurements," *Sensors (Switzerland)*, vol. 18, no. 4, pp. 1–25, 2018, doi: 10.3390/s18041266.

[6] L. Kanaris, A. Kokkinis, A. Liotta, and S. Stavrou, "Fusing bluetooth beacon data with Wi-Fi radiomaps for improved indoor localization," *Sensors (Switzerland)*, vol. 17, no. 4, pp. 1–15, 2017, doi: 10.3390/s17040812.

[7] A. G. Ferreira, D. Fernandes, A. P. Catarino, and J. L. Monteiro, "Performance analysis of ToA-based positioning algorithms for static and dynamic targets with low ranging measurements," *Sensors (Switzerland)*, vol. 17, no. 8, pp. 9–11, 2017, doi: 10.3390/s17081915.

[8] N. A. M. Maung and W. Zaw, "Comparative Study of RSS-based Indoor Positioning Techniques on Two Different Wi-Fi Frequency Bands," *17th Int. Conf. Electr. Eng. Comput. Telecommun. Inf. Technol. ECTI-CON 2020*, pp. 185–188, 2020, doi: 10.1109/ECTI-CON49241.2020.9158211.

[9] J. Golenbiewski and G. Tewolde, "Wi-Fi based indoor positioning and navigation system (IPS/INS)," *IEMTRONICS 2020 - Int. IOT, Electron. Mechatronics Conf. Proc.*, 2020, doi: 10.1109/IEMTRONICS51293.2020.9216376.

[10] L. Chen, I. Ahriz, and D. Le Ruyet, "CSI-Based Probabilistic Indoor Position Determination: An Entropy Solution," *IEEE Access*, vol. 7, pp. 170048–170061, 2019, doi: 10.1109/ACCESS.2019.2955747.

[11] M. Alfakih, M. Keche, H. Benoudnine, and A. Meche, "Improved Gaussian mixture modeling for accurate Wi-Fi based indoor localization systems," *Phys. Commun.*, vol. 43, p. 101218, 2020, doi: 10.1016/j.phycom.2020.101218.

[12] L. Ma, A. Y. Teymorian, X. Cheng, T. George, and W. Dc, "A Hybrid Rogue Access Point Protection Framework for Commodity Wi-Fi Networks," *Proc. - IEEE INFOCOM*, vol. 2018-April, pp. 1894–1902, 2008.

[13] S. H. Fang and T. Lin, "Principal component localization in indoor wlan environments," *IEEE Trans. Mob. Comput.*, vol. 11, no. 1, pp. 100–110, 2012, doi: 10.1109/TMC.2011.30.

[14] N. Hernández *et al.*, “WiFiNet: WiFi-based indoor localisation using CNNs,” *Expert Syst. Appl.*, vol. 177, p. 114906, 2021, doi: 10.1016/j.eswa.2021.114906.

[15] R. Daş, A. Karabade, and G. Tuna, “Common network attack types and defense mechanisms,” *2015 23rd Signal Process. Commun. Appl. Conf. SIU 2015 - Proc.*, pp. 2658–2661, 2015, doi: 10.1109/SIU.2015.7130435.

[16] H. K. Yu, S. H. Oh, and J. G. Kim, “AI based Location Tracking in WiFi Indoor Positioning Application,” in *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Feb. 2020, pp. 199–202, doi: 10.1109/ICAIIC48513.2020.9065227.

[17] A. A. Careem, W. H. Ali, and M. H. Jasim, “Wirelessly Indoor Positioning System based on RSS Signal,” *Proc. 2020 Int. Conf. Comput. Sci. Softw. Eng. CSASE 2020*, pp. 238–243, 2020, doi: 10.1109/CSASE48920.2020.9142111.

[18] K. Wang *et al.*, “Learning to improve WLAN indoor positioning accuracy based on DBSCAN-KRF Algorithm from RSS Fingerprint Data,” *IEEE Access*, vol. 7, pp. 72308–72315, 2019, doi: 10.1109/ACCESS.2019.2919329.

[19] A. A. Abdallah, S. S. Saab, and Z. M. Kassas, “A machine learning approach for localization in cellular environments,” *2018 IEEE/ION Position, Locat. Navig. Symp. PLANS 2018 - Proc.*, pp. 1223–1227, 2018, doi: 10.1109/PLANS.2018.8373508.

[20] P. Sthapit, H. S. Gang, and J. Y. Pyurr, “Bluetooth Based Indoor Positioning Using Machine Learning Algorithms,” *2018 IEEE Int. Conf. Consum. Electron. - Asia, ICCE-Asia 2018*, pp. 3–6, 2018, doi: 10.1109/ICCE-ASIA.2018.8552138.

[21] Z. Wei, Y. Zhao, X. Liu, and Z. Feng, “DoA-LF: A Location Fingerprint Positioning Algorithm with Millimeter-Wave,” *IEEE Access*, vol. 5, no. c, pp. 22678–22688, 2017, doi: 10.1109/ACCESS.2017.2753781.

[22] Z. Chen, H. Zou, J. F. Yang, H. Jiang, and L. Xie, “WiFi Fingerprinting Indoor Localization Using Local Feature-Based Deep LSTM,” *IEEE Syst. J.*, vol. 14, no. 2, pp. 3001–3010, 2020, doi: 10.1109/JSYST.2019.2918678.

[23] Q. Ye, X. Fan, G. Fang, H. Bie, X. Song, and R. Shankaran, “CapsLoc: A Robust Indoor Localization System with WiFi Fingerprinting Using Capsule Networks,” in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, Jun. 2020, pp. 1–6, doi: 10.1109/ICC40277.2020.9148933.

[24] D. Li, Y. Lei, X. Li, and H. Zhang, “Deep learning for fingerprint localization in indoor and outdoor environments,” *ISPRS Int. J. Geo-Information*, vol. 9, no. 4, 2020, doi: 10.3390/ijgi9040267.

[25] J. Fan, S. Chen, X. Luo, Y. Zhang, and G. Y. Li, “A Machine Learning Approach for Hierarchical Localization based on Multipath MIMO Fingerprints,” *IEEE Commun. Lett.*, vol. PP, no. 1, p. 1, 2019, doi: 10.1109/LCOMM.2019.2929148.

[26] H. J. Bae and L. Choi, “Large-Scale Indoor Positioning using Geomagnetic Field with Deep Neural Networks,” *IEEE Int. Conf. Commun.*, vol. 2019-May, pp. 1–6, 2019, doi: 10.1109/ICC.2019.8761118.