

Enabling Accurate Indoor Localization Using a Machine Learning Algorithm

Haidar Abdulrahman Abbas^{1*}, Kayhan Zrar Ghafoor²

¹Department of Computer, College of Science, University of Sulaimani, Sulaymaniyah, Iraq, ²Department of Software Engineering, University of Salahaddin, Erbil, Iraq



ABSTRACT

In this paper, fingerprint referencing methods based on wireless fidelity Wi-Fi received signal strength (RSS) have used for indoor positioning. More precisely, Naïve Bayes, decision tree (DT), and support vector machine (SVM) one-to-one multi-classes and error-correcting-output-codes classifier are to enable accurate indoor positioning. Then, normalization is used to reduce positioning error by reducing the fluctuation and diverse distribution of the RSS values. Different devices are used in this experiment; the training dataset is not included in the main dataset. Nonetheless, the learned model by the SVM algorithm cannot be affected by the elimination of train datasets of the test device. The efficiency of DT is lower than the other machine learning algorithms, because it performs by Boolean function, and it provides the low accuracy of prediction for dataset than the algorithms. Naïve Bayes technique based on Bayes Theorem is better than DT and close to SVM for positioning approves that 1–1.5 m positioning accuracy for indoor environments can be achieved by the proposed approach which is an excellent result than traditional protocol.

Index Terms: Received Signal Strength, Wireless Access Points, Wireless Fidelity Fingerprinting, Indoor Localization, Decision Tree, Naïve Bayes, Support Vector Machine

1. INTRODUCTION

In recent years, indoor localization became very popular due to the extensive range of applications [1]. Global navigation satellite systems and global positioning systems are used to enable accurate location, but they failed in indoor environments because of the low received signal power and satellite visibility in such places, underwater, inside building, caves, and tunnels [2]. These technologies need an open environment to work properly [3]. Signal strengths of wireless fidelity (Wi-Fi) Apps to fingerprint the location can be used in Wi-Fi fingerprint-based localization systems, these signals are

collected in the location and become the mainstream results for indoor localization. Wi-Fi fingerprint-based localization methods have two main phases offline fingerprinting phase and online localization phase Fig. 1. The offline phase is utilized in building a Wi-Fi fingerprint map by a site survey and save it at a database, and the online localization phase is used to locate the mobile devices by examining the received Wi-Fi signals with the fingerprint map [4].

Recently, users utilize mobile devices (e.g., smartphones) to access Wi-Fi networks in indoor environments (e.g., shopping malls). The investigation of indoor localization methods utilizing signals has increased widely [5]. Moreover, these methods are profitable because it does not require extra tools. One of the best advantages of location fingerprinting is capable of taking the benefits of multipath and non-line of sight problems in an indoor environment, as they truly assistant RSS to be distinct at dissimilar points of the area [3]. While there are several valuable features in fingerprint-based

Access this article online

DOI: 10.21928/uhdjst.v4n1y2020.pp96-102

E-ISSN: 2521-4217

P-ISSN: 2521-4209

Copyright © 2020 Abbas and Ghafoor. This is an open access article distributed under the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (CC BY-NC-ND 4.0)

Corresponding author's e-mail: Haidar Abdulrahman Abbas, Department of Computer, College of Science, University of Sulaimani, Sulaymaniyah, Iraq. E-mail: haidar.abbas@univsul.edu.iq

Received: 08-05-2020

Accepted: 18-06-2020

Published: 27-06-2020

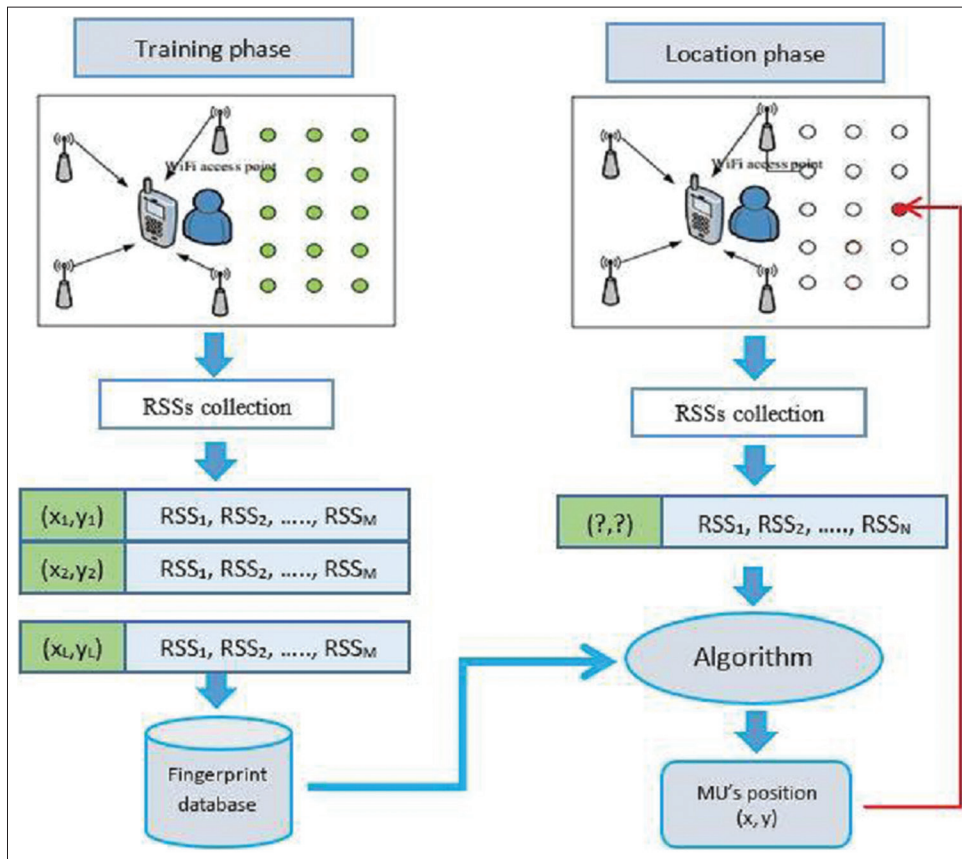


Fig. 1. Constructing the wireless fidelity fingerprinting map [7].

localization, building fingerprint landmarks for a huge area requires an important amount of time and human resources. Database fingerprint can be altered by environmental influences, time, and different devices, so it is necessary to update frequently.

Important studies have been dedicated to the online localization phase; decimeter-level localization efficiency can be obtained by utilizing advanced algorithms which are used to collect online Wi-Fi signals with the fingerprint map [6]. This field of study has been concerned by the researcher in both industry and academia. To collect RRS and fingerprint to assessment, the target location machine learning algorithms can be used such as deep learning and K-Nearest-Neighbor (K-NN) [7].

In this study, fingerprint methods utilizing the Wi-Fi strength signal is presented for indoor positioning. To decrease the positioning errors, Naive Bayes, decision tree (DT), and support vector machine (SVM) one-to-one multi-classes and error-correcting output codes (ECOC) classifiers are proposed and the contrast among these methods.

Normalization is used to reduce errors positioning in the values of RSS because of instability and diverse distributions values of RSS. Different devices are used in this experiment when the train data set is not involved in the main dataset. Nonetheless, the learned model by the SVM algorithm cannot be affected by the elimination of train datasets of the test device. The efficiency of DT is lower than the other machine learning algorithms, because it performs by Boolean function, and it provides the low accuracy of prediction for dataset than the other algorithms. Naïve Bayes techniques based on Bayes Theorem is better than DT and close to SVM for positioning accuracy. SVM error positioning approves that 1–1.5 m positioning accuracy for indoor environments can be achieved by the proposed approach which is an excellent result than traditional protocol.

2. RELATED WORK

Recently with the development of computing and the popularity of location-based services, many types of research have considered the improvement of the indoor localization

system. Some of these researches focus on the designing system for specific applications that requires a high efficiency (e.g., in the order of centimeters) [8]. Normally, developing these systems need devoted hardware with a huge application cost. Contrarily, several kinds of research have focused on general location-based services where the necessity of accuracy in the form of meters.

Wi-Fi strength signal is used by the fingerprinting method for indoor positioning. To decrease the positioning errors, the improved form of nearest neighbor algorithm is suggested which is called NK-NN, multipath, and RSS variations created the new form of NK-NN, which are utilized the basic KNN and it is variant. In the RSS testing sample, the noise can be removed by compared each testing sample to each fingerprint and based on the minimum distance, the sample is chosen for the position's calculation. After that, the process of classification is operated on the Kth-nearest training sample of diverse reference points which assistance to trim the noise of RSS training and preventing them from the localization. In the experimental outcome, the NK-NN method has better performance than other similar methods [7].

Other studies used Convolutional Neural Network (CNN)-based Wi-Fi fingerprinting for indoor localization. It can be seen that the achievement in image classifications, the suggested method can be potent to minor changes of received signals as it uses the radio map topology as well as signal strengths. In the suggested method, based on the one-dimensional Wi-Fi signals, the two-dimensional (2-D) virtual radio map is built (e.g., received signal strength indicator values) and later a CNN utilizing 2-D radio map is designed as inputs. Consequently, the proposed method is learned the signal strengths as well as the topology radio map. To enhance the efficiency of the suggested method utilizing different improving techniques as feature scaling, dropout, data balancing, and ensemble [6].

To enhance the accuracy of positioning systems, many approaches have been studies and focused on long short-term memory (LSTM) networks. A deep neural network is utilized to improve the efficiency of positioning methods which is acceptable for handling sequential datasets. Therefore, LSTM modes are used because they can recognize the dependency of long-term that is existing in the Wi-Fi data that can be seen from the deep recurrent model's performance. The architecture of RNN and LSTM can recognize the dependency of long-term and utilizing them for later prediction. It is good to examine the previous landmark position to an exact estimate of points on the radio map.

The main aim of implementing RNN is to guarantee for providing better performance of recurrent networks on the Wi-Fi dataset. Vanilla LSTM is the primary model that has a good enhancement by 47.8% over the KNN and 10.2 enhancements over RNN utilizing the complete dataset. The efficiency of Vanilla LSTM is even developed after updated to 3-Stacked LSTM. The improvement of 3-Stacked LSTM is 74.4% over the KNN and 18.1% over Vanilla LSTM [9].

There is a rich theoretical basis that is prepared by the Statistical Learning Theory for developing the model, starting a set of examples. In a specific Wi-Fi, the wireless has a signal strength measurement for standard functioning mode so that no particular hardware is desired. SVM is designed and compared to other approaches examined in the scientific literature on the equivalent data set. Experiments executed in the real-world environment illustrate that the outcomes are comparable, with the benefits of low algorithms complication in the standard functioning phase. Furthermore, the algorithm performed better than the other techniques which are mainly appropriate for classification [10].

3. METHODOLOGY

The localization algorithm is fundamentally used for making RSS related radio maps in a designated indoor environment as well as converting localization problem into an optimization problem: Obtaining RSS value measurements of an undisclosed location, the function can help in estimating the location when used in reverse order. While using a fingerprinting technique in the online phase, identical smartphones should be used in both buildings, the RSS dataset as well as testing. Using different smartphones would worsen the accuracy of the calculated position. To eliminate this problem, we propose a DT, Naive Base, and SVM model or adapt the nature of the calculated RSS values among multi-smartphones. The model is directed at various types of smartphone measurements by adopting a machine learning algorithm.

First, gathered RSS values at all the identified close positions are normalized, the normalization is achieved by subtracting the mean value of the gadgets engaged in training the model and then dividing the results by the standard deviation of the aforementioned gadgets. Before normalizing the RSS of wireless access points (AP), the error positioning high because of fluctuation and heterogeneous distribution of the RSS values and applying normalization to decrease the variation of the value and rescale the RSS value within a

uniform distribution. RSS vector will be filled with zeroes for those APs. The normalized RSS values can then be applied to train and test the algorithms.

3.1. DT

DTs are a non-parametric method that belongs to the supervised learning algorithms family. It is for classification and regression [11]. In this algorithm, a binary DT is developed from the training data set. In the beginning, basic decision rules derived from the data features are learned. It operates in three nodes; root node, internal node, and a leaf or terminal node. The terminal node has a single receiving edge and zeroes an outgoing edge. The internal has two edges, one for incoming and one for outgoing, while the root node can have zero or more outgoing edges but does not have any incoming edges. Each leaf node is given a class label. Each node is related to a decision performed on the inputs. Next, the node is split into new subsets, one for each of the node's sub-trees, in such a way that the same target location is in the same subsets [11]. The algorithm halts upon finding a pure decision meaning each node's data subset has a single target location and when uncertainty is inefficient.

3.2. Naïve Bayes

The crux of this theorem is derived from the Bayes theorem. According to the Naïve Bayes theorem, all features are independent of each other. While this assumption is usually not true in real-world applications, yet Naïve Bayes have had positive results in certain scenarios, mostly when there is a small number of training samples [12]. With an end goal to accomplish high accuracy while decreasing pre-deployment trials, we select this strategy for processing the probabilities of the locations' given measurements. The event with the most elevated likelihood is considered as the candidate. Naïve Bayes classifier depends on two fundamental assumptions: (1) The features do not affect each other and (2) the prominence of all the features is equal [13].

3.3. SVM

SVMs [14], [15] are non-parametric supervised learning models with related learning algorithms that analyze data used for pattern recognition problems. SVMs are applied in the localization system by training the support vectors on a radio map that consists of grid points. SVMs study the association between the trained fingerprints and their grid points by taking into account each grid point as a class. This method can be expanded to multiple class classification rather than just two classes. In our training dataset, we have 105 classes, so we used the ECOC one-to-one SVM which is used to classification when classes are more than two after representing the training data by mapping the data to the feature space. The SVM

algorithms identify hyperplane, which separates the support vector trained with a distance [14].

4. PERFORMANCE EVALUATION

In this section, designing radio maps of RSS in the studied indoor environment and positioning the difficulty as an optimization problem is presented which is the main idea behind the localization algorithm: providing RSS value measurements of a new position, the function is reversed to determine the evaluated position. Localization fingerprinting methods utilized two main phases (online positioning phase and offline training phase). Afterward, the fingerprint is collected and builds our dataset. The dataset consists of the true location of pre-selected positions and equivalent RSS of nearby AP [16]. The approach that we proposed illustrates the decision tree, Naïve Bayes, and SVN model and compares those models, normalization was applied to fingerprint landmarks.

4.1. Dataset and Simulation Setup

Our RSS data are collected from KIOS Research Center which is a 560 m² office environment. This center has many open cubicle-style and special offices, labs, and conference rooms. Wireless LAN standard has been used to install nine local Apps and offer full coverage all over the floor. We utilize five diverse mobiles to collect our data including, HP iPAQ hw6915 personal digital assistant with Windows Mobile, an Asus EeePC T101MT laptop running Windows 7, and HTC Flyer Android tablet and two other Android smartphones (HTC Desire, Samsung Nexus S). We use fingerprinting for our training data, documenting these fingerprints have RSS measurement from entire existing APs, at 105 separate reference positions by carrying all five devices concurrently. We utilize each device to collect fingerprints, 2100 training fingerprints are available, equivalent to 20 fingerprints per reference position. For building device-specific radio maps, these data are utilized by computing the measure of mean values RSS that analogous to each reference position. We indicate that the device-specific radio maps are only required to estimate purposes. After 2 weeks, we utilized a predefined router to gather more test data by walking forward to the router. The router contains two segments and 96 positions; most of them do not concur with the mention position. Each router is tested 10 times using all devices concurrently, while one fingerprint was documented at all test positions [17].

MATLAB toolbox is used to estimate the performance of models. In the first scenario, we tested the DT, Naïve Bayes,

and SVM as matching algorithms without applying the RSS normalization (mean and standard deviation) parameters while the device is tested, the training dataset of the device was excluding. The second scenario was using the machine algorithms by applying the RSS normalization, like the previous scenario, the testing device, the training dataset was excluding.

4.2. Simulation Results

We used root mean square error between the estimated and the true locations to evaluate the localization accuracy DT, Naïve Bayes, and SVM algorithms, while there are many methods to evaluate the accuracy.

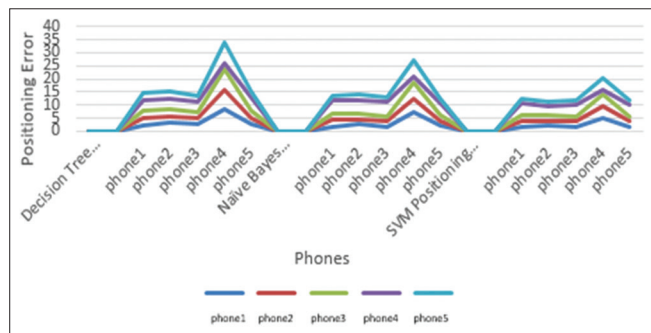


Fig. 2. Decision tree, Naive Bayes, support vector machine positioning accuracy when normalization not applied.

TABLE 1: Positioning accuracy of DT, Naïve Bayes, and SVM algorithms when normalization not applied

Decision tree positioning accuracy when normalization not applied					
	Phone 1	Phone 2	Phone 3	Phone 4	Phone 5
Phone 1	2.203	2.8795	2.5406	4.4574	2.4796
Phone 2	3.4706	2.1509	2.7549	3.7918	2.9944
Phone 3	2.8008	2.5236	1.9669	3.9633	2.3818
Phone 4	8.4164	7.2392	7.9662	2.4191	8.0881
Phone 5	2.5555	2.6016	2.479	5.1455	2.6042

Naïve Bayes positioning accuracy when normalization not applied					
	Phone 1	Phone 2	Phone 3	Phone 4	Phone 5
Phone 1	1.8598	2.6695	2.0898	4.9985	1.8697
Phone 2	2.7418	1.8852	2.1934	5.0275	2.4532
Phone 3	1.7627	2.2526	1.5297	5.6689	1.8885
Phone 4	7.4514	5.1601	6.2018	1.8415	6.4267
Phone 5	2.0015	2.1527	1.9874	4.3407	1.7289

SVM positioning accuracy when normalization not applied					
	Phone 1	Phone 2	Phone 3	Phone 4	Phone 5
Phone	1.748	2.3896	1.9668	4.4945	2.0451
Phone	2.3941	1.7739	1.9912	3.2393	2.083
Phone	1.7156	2.3555	1.4941	4.344	1.9079
Phone	5.0985	4.2347	4.6483	1.8507	4.7268
Phone	1.791	2.0599	1.9353	4.1604	1.6278

DT: Decision tree, SVM: Support vector machine

The average positioning accuracy of the first scenario shown in Table 1 which contains all the devices tested. The DT is generally represented by Boolean function and gives a dataset low prediction accuracy compared to other machine learning algorithms. The SVM has better positioning of the accuracy than other algorithms Fig. 2.

The positioning accuracy for all algorithms is higher in the second scenario after normalization exercised on the dataset than before normalization, caused by fluctuation and heterogeneous distribution of the RSS values. The findings are listed in Table 2. Compared to the SVM with DT and Naïve Bayes, we can see that the SVM exhibits more accuracy in positioning. SVM has the most elegant maths behind them and uses the Kernel trick in the dual problem. The results of Naïve Bayes have a second degree in positioning accuracy at each scenario Fig. 3.

Phone 4’s positioning accuracy is weaker than most phones, its return to phone 4’s RSS values, Phone 4 read the signal strength from -11 to -90 dB, and most are between -11 and -40 dB, unlike other phones. There is a big gap in RSS values where we have categorized phone4 and others, so the positioning accuracy is worse than some, so the error is large.

TABLE 2: Positioning accuracy of DT, Naïve Bayes, and SVM algorithms when normalization applied

Decision tree positioning accuracy when normalization applied					
	Phone 1	Phone 2	Phone 3	Phone 4	Phone 5
Phone 1	2.2143	2.6063	2.3258	2.5884	2.2332
Phone 2	2.0923	2.1043	2.1062	2.3519	2.2207
Phone 3	2.1431	2.2638	1.9752	2.4954	2.213
Phone 4	2.2618	2.5557	2.4014	2.5941	2.4816
Phone 5	2.2688	2.5763	2.3769	2.6545	2.5061

Naïve Bayes positioning accuracy when normalization applied					
	Phone 1	Phone 2	Phone 3	Phone 4	Phone 5
Phone 1	1.9256	2.1813	1.787	2.3594	1.8646
Phone 2	1.9088	1.8985	1.834	2.2155	1.9758
Phone 3	1.7655	1.9025	1.5323	2.0569	1.7883
Phone 4	1.8462	2.009	2.0197	1.8725	1.7213
Phone 5	1.7523	1.9018	1.9706	2.0426	1.7213

SVM positioning accuracy when normalization applied					
	Phone 1	Phone 2	Phone 3	Phone 4	Phone 5
Phone	1.7554	1.982	1.8261	2.2321	1.9163
Phone	1.7485	1.6998	1.6029	1.8001	1.7596
Phone	1.4364	1.853	1.4621	2.2092	1.7774
Phone	1.8266	1.9311	1.8866	1.8432	1.6975
Phone	1.6622	1.963	2.0058	2.0025	1.6215

DT: Decision tree, SVM: Support vector machine

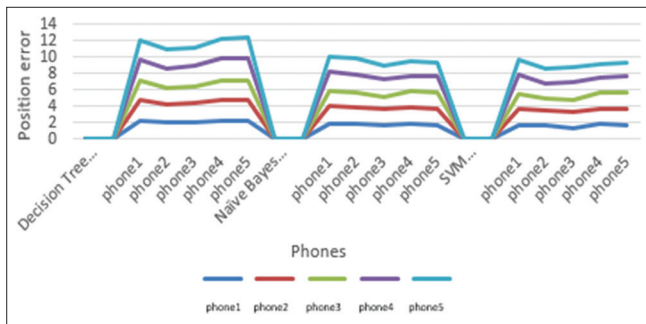


Fig. 3. Decision tree, Naïve Bayes, support vector machine positioning accuracy when normalization applied.

Each of the algorithms has strengths and weaknesses, if we compare DT to other algorithms, it has less requirement for data pre-processing and not affected by missing values in data set, but any changes in data set impact the structure of it which is lead to instability. NB needs a smaller amount of training data to evaluate the test data, and implementation is easy. However, the main problem of NB is the assumption of independence. The SVM algorithm has more effective when the number of dimensions is greater than the sample number, and it comparatively memory efficient. SVM has a disadvantage like it is not the best option for the large data set and does not execute well with a data set that has more noises. In general, our proposed approach has many advantages; it does not need extra hardware to be installed, and high performance was achieved. Our proposed concept's disadvantages, require more computational work (especially SVM) compared to others by the system. To compare our results with other works, this system has more positioning accuracy. This is a very different outcome than conventional protocol.

5. CONCLUSION

In this research article, RSS fingerprint-based Wi-Fi localization was assessed in regards to the in-operation infrastructure of an indoor environment. We review the modern resolutions for very accurate localization in indoor schemes. Next, we outline the rise in positioning error when dissimilar platform-devices are used in the fingerprinting technique for training and testing the dataset. In addition, RSS measurements produce different values for the same position and time when dissimilar platform-devices are used. We implement the most popular and reliable machine learning algorithms, namely, DTs, Naïve Bayes, and SVM learning algorithms. Examine ensemble estimators that apply multiple algorithms to

estimate the position and then we choose a combination the leads to the most efficient performance. SVM error positioning shows that 1–1.5 m positioning accuracy for indoor environments can be accomplished by the presented technique which is an obvious improvement compared to existing approaches. Thus, fingerprinting localizations can utilize RSS data to minimize the notable amount of time and energy.

REFERENCES

- [1] J. Xiao, Z. Zhou, Y. Yi and L. M. Ni. "A survey on wireless indoor localization from the device perspective," *ACM Computing Surveys*, vol. 49, no. 2, p. 2933232, 2016.
- [2] A. S. Paul and E. A. Wan. "RSSI-Based indoor localization and tracking using sigma-point kalman smoothers," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 5, pp. 860-873, 2009.
- [3] N. Alikhani, S. Amiranloo, V. Moghtadaiee and S. A. Ghorashi. "Fast Fingerprinting Based Indoor Localization by Wi-Fi Signals," *2017 7th International Conference on Computer and Knowledge Engineering*, vol. 2017 Janua, pp. 241-246, 2017.
- [4] S. Dai, L. He and X. Zhang. "Autonomous WiFi Fingerprinting for Indoor Localization". In: *2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPs)*, pp. 141-150, 2020.
- [5] F. Li, M. Liu, Y. Zhang and W. Shen. "A two-level wifi fingerprint-based indoor localization method for dangerous area monitoring," *Sensors (Basel)*, vol. 19, no. 19, p. 4243, 2019.
- [6] J. W. Jang and S. N. Hong. "Indoor Localization with WiFi Fingerprinting Using Convolutional Neural Network," *International Conference on Ubiquitous and Future Networks*, vol. 2018, pp. 753-758, 2018.
- [7] M. Alfakih and M. Keche. "An enhanced indoor positioning method based on Wi-fi RSS fingerprinting," *The Journal of Communications Software and Systems*, vol. 15, no. 1, pp. 18-25, 2019.
- [8] C. Chen, Y. Chen, Y. Han, H. Q. Lai and K. J. R. Liu. "Achieving centimeter-accuracy indoor localization on wifi platforms: A frequency hopping approach," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 111-121, 2017.
- [9] A. Sahar and D. Han. "An LSTM-based Indoor Positioning Method Using Wi-Fi Signals," *ACM's International Conference Proceedings*, 2018.
- [10] M. Brunato and R. Battiti. "Statistical learning theory for location fingerprinting in wireless LANs," *Computer Networks*, vol. 47, no. 6, pp. 825-845, 2005.
- [11] Y. Li. "Predicting materials properties and behavior using classification and regression trees," *Materials Science and Engineering A*, vol. 433, no. 1-2, pp. 261-268, 2006.
- [12] N. Gutierrez, C. Belmonte, J. Hanvey, R. Espejo and Z. Dong. "Indoor Localization for Mobile Devices," *Proceeding. 11th IEEE International Conference on Sensing Control*, pp. 173-178, 2014.
- [13] Z. Wu, Q. Xu, J. Li, C. Fu, Q. Xuan and Y. Xiang. "Passive indoor localization based on CSI and naive bayes classification," *IEEE Transactions on Systems, Man, and Cybernetics Systems*, vol. 48, no. 9, pp. 1566-1577, 2018.
- [14] B. Schölkopf. "Slides learning with kernels," *Journal of the Electrochemical Society*, vol. 129, p. 2865, 2002.

- [15] T. Joachims. "Transductive Inference for Text Classification Using Support Vector Machines," *Proceeding 20th International Conference on Machine Learning*, 2000.
- [16] Z. Zhong, Z. Tang, X. Li, T. Yuan, Y. Yang, M. Wei, Y. Zhang, R. Sheng and N. Grant. "XJTLUIndoorLoc: A New Fingerprinting Database for Indoor Localization and Trajectory Estimation Based on Wi-Fi RSS and Geomagnetic Field," *Proceeding 2018 6th International Symposium Computer Network*, pp. 228-234, 2018.
- [17] A. H. Salamah, M. Tamazin, M. A. Sharkas and M. Khedr. "An Enhanced WiFi Indoor Localization System Based on Machine Learning," *2016 International Conference Indoor Position Indoor Navigation*, pp. 4-7, 2016.