Machine Learning Based Indoor Localization using Wi-Fi and Smartphone

Zulqarnain Khokhar ¹, Murtaza Ahmed Siddiqi ²

Abstract – Wi-Fi based indoor positioning with the help of access points and smart devices have become an integral part in finding a device or a person's location. Wi-Fi based indoor localization technology has been among the most attractive field for researchers for a number of years. In this paper, we have presented Wi-Fi based in-door localization using three different machine-learning techniques. The three machine learning algorithms implemented and compared are Decision Tree, Random Forest and Gradient Boosting classifier. After making a fingerprint of the floor based on Wi-Fi signals, mentioned algorithms were used to identify device location at thirty different positions on the floor. Random Forest and Gradient Boosting classifier were able to identify the location of the device with accuracy higher than 90%. While Decision Tree was able to identify the location with accuracy a bit higher than 80%.

Keywords – Machine Learning, Random Forest, Decision Tree, Gradient boosting, Wi-Fi localization, indoor positioning.

I. INTRODUCTION

Indoor Location-Based Services (ILBs) have been quite a popular and attractive subject for quite some time. As ILBs became an essential component for numerous indoor-based applications. Such applications assist in information repossession, device location, person location, pedestrian direction-finding etc. [1]. With the demand of high precision and cost effective localization for industrial applications, a number of researchers find indoor localization quite attractive. Able to identify the exact position of a mobile device in an indoor area i.e. building can be quite useful for a number of industrial, domestic and rescue based applications [2].

A number of researchers have proposed a number of different approaches to facilitate ILBs with precision. Clearly, every proposed approach comes with its advantages and disadvantages.

Some of the most common methods include localization with the help of Wi-Fi, acoustic, blue-tooth, cellular network, visible light, global positioning system etc. [3]. Fingerprint localization with the help of Wi-Fi is one the most common method used these days [4]. Wi-Fi based fingerprinting method involves mapping based on physical location with the help of Received Signal Strength (RSS) received from surrounding Wi-Fi access points (AP). Samples of RSS and related values are collected at different locations to make a database of values identifying each location associated with an AP. Such database creates a fingerprint or a mapping of each location based on values received from close-by APs. As per literature [5], fingerprint based ILBs suffer from two major issues. One dataset construction and accuracy of localization due to hindrances or crowded area. There are two main phases of Fingerprinting [6]. An offline phase to collect data and an online phase for processing and identifying locations of collected data. Both of the mentioned phases are illustrated in Figure 1 [7].

Figure 1 Fingerprint-based Localization model

To make the dataset for fingerprinting, signal strength readings are taken at each point of targeted area's Wireless Local Area Network (WLAN). All the readings are then stored in a database that represents radio map (RM) of the targeted area. In the online phase, the smartphone collects the RSS from the APs and forward them to ML based algorithm and on

¹ Zulqarnain Khokhar works at UBL as Assistant Manager IT service delivery. Email: khokharzulqarnain@gmail.com

² Murtaza Ahmed Siddiqi is with the department of computer science, Sukkur IBA University, Pakistan. Email: Murtaza.siddiqi@iba-suk.edu.pk

server to compare the predefined fingerprint of the offline phase with the RSS in the online phase. This process creates an estimation of location on the grid map. Such grid map represents a fingerprint of each location based on signals received from nearby APs. Due to high uncertainty of Wi-Fi signals, fingerprinting based on Wi-Fi is not an ideal solution for ILBs. A number of studies are conducted to overcome the mentioned issue, but environmental setting cannot be controlled. That brings us to the main challenge in ILBs, which is high accuracy [8].

On the other hand, a due preprocess and appropriate feature selection can make a huge difference when using Wi-Fi based finger printing. With appropriate pre-processing Wi-Fi based Indoor localization can achieve high accuracy in indoor tracking [9]. Another advantage of Wi-Fi based ILBs is that every device now a day is compatible and connected with Wi-Fi. That means without installing any specific software Wi-Fi based ILBs can be performed. Wi-Fi based ILBs is cost effective and it does not require any extra hardware other than Wi-Fi device. Existing WLAN infrastructure can be used for measuring signal strength with the help of wireless interface of hand help devices [10]. Due to the mentioned advantages Wi-Fi based ILBs is on the rise [11]. ILBs based on Wi-Fi generally includes based on fingerprinting and multi-lateration [12]. Multi-lateration approach based on Wi-Fi suffer from distance estimation [13]. While, fingerprinting based approach have the concern of creating a fingerprinting database for the location. In our experiment, we have used Random Forest (RF), Decision Tree (DT) and Gradient Boosting (GB). RF is an ensemble learning technique for regression and classification. DT is a supervised learning technique; it works by splitting the data based on certain parameters. GB is considered an algorithm based on greedy approach. Generally, GB is an ensemble algorithm in ML. GB is use for regression and classification based problems [14]. In our experiment, the ML models try to identify locations that are provided to them in the form of testing dataset. Unlike other ILBs paper, our model identifies ILBs as a classification problem. Each location is labeled based on closest AP and the ML model predicts which location is provided to it in the test dataset. Due to the fingerprinting method adapted in this research, the result is represented as a classification system. Unlike customary research, that highlight the distance between device detected and original device location. In this research, we simply treated locations as a label and used ML based methods to train a model to classify and detect the location of device presented as a testing dataset.

II. RELATED WORK

A number of researchers and organizations have been working on ILBs and positioning methods. Approaches used by recent researchers are not limited to ML. Authors F. Zafari et al. [15] published an up-to-date survey that includes some of the recently proposed ILBs solutions. The paper cover methods that acquired high accuracy with minimum false rate. The authors highlighted the importance of ML in the field of ILBs and provided a detailed survey of different ML techniques that are appropriate for ILBs. Paper covers a brief description of prominent methods such as Time of Flight (ToF), Angle of Arrival (AoA), Received Signal Strength (RSS) and Return Time of Flight (RTOF). Mentioned methods rely on Wi-Fi, Ultra Wideband (UWB), Radio Frequency Identification Device (RFID) and Bluetooth systems. Furthermore, author discussed how different ML techniques could further improve location detection with efficiency. ML techniques discussed by authors include Neural Networks (NN), Support Vector Machine (SVM) and k-Nearest Neighbors (K-NN). Authors highlighted that most of the ML based techniques use RSSI fingerprints to obtain device or user location. The paper mainly discussed localization and positioning of users and their devices. In discussed literature, author highlighted the strengths of the existing systems, while comparing different approaches. Author also assessed different proposed methods from the perspective of availability, energy efficiency cost, tracking accuracy, reception range, scalability and latency. Paper also cover a number of use case samples of ILBs to show their importance.

In another paper [16] author proposed a recurrent neural networks (RNNs) based approach for Wi-Fi fingerprinting and then using the fingerprint for ILBs. Authors proposed solution identifies objects at different paths and takes into account the relationship between the received RSSI values in a path. To increase the accuracy between the sequential variations of RSSI, a biased regular filter is used for both input RSSI data and sequential output positions. Authors used different types of RNN to get a comparative result and to identify the RNN configuration for highest precision. RNN types used by the authors include vanilla RNN, bidirectional GRU (BiGRU), bidirectional LSTM (BiLSTM), bidirectional RNN (BiRNN), gated recurrent unit (GRU) and long short-term memory (LSTM). Experimental results in the paper have frequently shown that the proposed LSTM structure succeeds an average ILBs error of 0.75m with 80% of the prediction errors under 1m. The proposed method performs better then feed forward neural network, K-nearest neighbors (K-NN), Kalman filter and probabilistic methods. As per authors, some of the

core challenges for ILBs include spatial uncertainty, RSSI variability and the RSSI small collecting time. The proposed method in the paper provides a much effective and efficient method to overcome mentioned challenges.

Another paper by H. Salamah et al. [17] Identify a tricky concern with Global Navigation Satellite Systems (GNSS). The GNSS suffer from correctness deterioration and outages in areas surrounded by concrete and GNSS is nearly inaccessible for indoor locations. To address this problem, authors proposed a ML based approach. In the proposed method flawless match between the fingerprint and pre-defined set of AP on the wireless map are utilized. The authors employed Principle Component Analysis (PCA) to decrease the computational cost and increase the performance of the ILBs with the help of ML method. The authors conducted experiments in a real world indoor environment with both static and dynamic settings. The proposed approaches were developed using Android-based smart phone with IEEE 802.11 WLANs. The evaluation of the proposed approaches was compared with SVM, DT, RT and K-NN classifiers. The results highlighted that the proposed

References	Method	Results/ limitation
[18]	Decision trees, Random forest	Proposed method gave reasonable results but results can be improve by using Deep learning models.
[19]	Supervised Learning	Achieved the required results but the proposed method is highly dependent on the wireless device interface used for finger printing and localization.
[20]	Deep Learning	Proposed approach gave good results for limited area localization but the proposed scheme is not suitable for complex environments.
[21]	Supervised Learning	Results show good accuracy but further improvements can be achieved in complex environment by utilizing advance ML approaches.

approach reduced the computational strain by 70% with the help of RT classifier in the static environment and by 33% when K-NN was used for classification. It was noted that the location correctness was improved in event of using K-NN and RT classifiers. The Comparison of some recent related work are shown in Table 1.

Table 1 Recent papers on Wi-Fi localization

III. METHODOLOGY

In our proposed method, we used Wi-Fi based grid mapping for localization. In our proposed method, first step was to collect data from the 30 AP located on the

building floor. The data collected will show the Wi-Fi signal behavior at each location. The AP taken in to consideration for the experiment are distributed on the floor. The floor have 24 rooms, which include research rooms, faculty office, lecture rooms and conference room. Wi-Fi AP are available in each room and in the hallway. Fingerprint data was collected while standing in the hallway. Each observation was collected while standing in front of each room. Such observation method is useful to label the observations at each point. Table 2 shows a sample of how observations were collected.

Room Numbers	Signal strength parameters/ features
301	Time, RSSI, Channel, Type, SubType, SourceMAC, BSSID, DestinationMac, Size, Description, Building ID, Floor, Device Model, Longitude, Latitude
302	
303	

Table 2 Fingerprint dataset observation sample

The data was collected using HUAWEI mate 10 cell phone, and the application used was Packet Capture for android-based systems. The basic flow of proposed method is shown in Figure 2:

Figure 2 Experiment flow from data collection to classification

The dataset made after collecting data contained 30 APs signal strength at different locations on the floor. Initially the collected data was in PCAP format and was converted to CSV format. Originally, the CSV files contained features such as Time, RSSI, Channel, Type, SubType, SourceMAC, BSSID, DestinationMac, Size, Description, Building ID, Floor, Device Model, Longitude, Latitude, Altitude, SSID and RSSI. Table 3 shows the extracted features that were used for the experiments. The four feature we selected were based on geographical identity (i.e latitude and longitude), basic service identifier and received signal strength. Reason behind selecting only mentioned features was due to the application used to collect fingerprint dataset i.e. Packer capture. The application is not feature rich and was designed for sniffing purpose.

Feature	Longitude	Latitude	BSSID	RSSI
Format	127.3676	36.3800	88:36:6c:...	-68
	127.3678	36.3800	2c:36:f8:...	-78

Table 3 Features selected for experiments

Due to mentioned reason, we selected only the four most appropriate features for our experimentation. Most of the Wi-Fi based localization methods indent to use mentioned features as well. Common features selected for Wi-Fi localization can be seen in Table 4 [28].

References	Technology	Features	Algorithm
[22]	Wi-Fi	RSSI	K-NN, RF
[23]	Wi-Fi	RSSI, SSD, HLF	K-NN
[24]	Wi-Fi	RSSI	K-NN
[25]	Wi-Fi	RSSI, RSSIF, CMF, FoCF, SSF, FLOMF	Multiple Classifiers Multiple Samplers fusion algorithm
[26]	Wi-Fi	RSSI	K-NN, SVM, logical regression
[27]	Wi-Fi	BSSID, RSSI	RF

Table 4 Common features for ML based Wi-Fi localization

Keeping in mind that irrelevant feature could degrade ML based algorithms. After feature selection, labels were added based on AP location. The labels were added as unique numbers to identify each AP location. The labels are shown in Figure 3:

Longitude	Latitude	BSSID	RSSI	Label
127.367820	36.38007	d0:17:c2:9e:12:a0	-62	1013
127.367932	36.38007	2c:36:f8:42:fe:51	-47	1003
127.367932	36.38007	2c:36:f8:e8:a1:e2	-64	1003
127.367910	36.38007	2c:36:f8:0e:8f:32	-78	1005

Figure 3 Dataset features with labels

The BSSID was encoded using python library Label Encoding. Label encoding is one of the most popular encoding method for handling categorical variables in a dataset. In this encoding method, each value is assigned a unique integer value based on alphabetical order. Figure 4 shows the data after applying label encoding on BSSI feature.

Longitude	Latitude	BSSID	RSSI	Label
127.367820	36.38007	74	-62	1013
127.367932	36.38007	11	-47	1003
127.367932	36.38007	36	-64	1003
127.367910	36.38007	2	-78	1005

Figure 4 Dataset after applying Label Encoding

The dataset was then divided in to 80% training set and 20% testing dataset. A dataset can be divided in to a number of different ratios, but in this experiment, the

fairest splitting recommended for a small dataset [29] is performed. Min-Max scaling was applied to normalize the datasets before classification. In Min-Max scaling, the data is scaled to a static range for better performance and understanding of ML algorithms. Mathematically, Min-Max scaling can be represented as Equation 1 [30].

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

In Equation 1, 'x' is the original value from the data and x' is the new normalized value. After data scaling, ML algorithms were applied. Based on training dataset, we made three different models for each ML algorithm. The flow of the process can be seen in Figure 5.

Figure 5 Experiment flow diagram

IV. EXPERIMENT AND RESULTS

The computer used for the experiment was as Intel i5-9400 with 16GB RAM and GTX 1050 card. We trained three models based on DT, RF and GBC. After training the three ML models, we provided them with testing data set for prediction. The detection rate of each model based on testing dataset can be seen in Table 5:

	Decision Tree (DT)	Random Forest (RF)	Gradient Boosting Classifier (GBC)
Detection Rate	75.96%	98.50%	89.81%

Table 5 ML algorithms test predictions

As seen in Table 5, RF was able to detect the device location with 98.50% detection rate. GBC was the second with location detection rate of 89.81% and DT model was able to detect device location with 75.96%. Table 6 displays the confusion matrix for the test results of each algorithm.

Matrix	DT	RF	GBC
Accuracy	81.04%	98.80%	94.31%
Precision	79.02%	99.01%	89.67%
Recall	79.08%	98.20%	91.88%
F1-Score	76.86%	98.05%	90.94%

Table 6 Confusion matrix for each algorithm

The accuracy, precision, recall and F1-score were calculated by using Equations 2, 3, 4 and 5 respectively. Accuracy represents the ratio of accurate prediction to the total observations. Precision represent the ratio of correctly predicted observations out of total true and false positive predictions. Recall indicate the ratio of correct prediction out of all predictions of an individual label. F1-score represent the weighted average of recall and precision.

$$Accuracy = \frac{True\ Negative + True\ Positive}{Total} \quad (2)$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (3)$$

$$Recall = \frac{Ture\ Positive}{True\ Positive + False\ Negative} \quad (4)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Where, true positive represent the correct location of the device while true negative represent correct predictions of the device not present at the location. False positive indicate incorrect predictions of the device being present at a location while it is not and false negative represent incorrect predictions of a device not present at a location while it is at that location. Based on the detection rate and classification matrix in Table 4 & 5, RF was able to predict the location of the device with highest precision. The superior performance by RF can be credited to a number of factors. However, the most significant reason behind RF's enhanced performance is it does not depend vastly on any specific set of features [31]. As mentioned earlier, that the tool (i.e. Packet Capture) used for data collection was not very rich in feature representation. Rudimentary dataset resulted in complications for the algorithms to perform classification. Despite dataset issues, RF was able to perform with high accuracy as compared to the other algorithms.

Table 7 presents a result comparison between this study and paper with similar methodologies and ML algorithms. Based on Tables 5, 6 and 7 the proposed method in this paper performed with competitive accuracy. But then again, comparing accuracy or detection rate might not be a very good scale to conclude superior performance method.

References	Algorithm	Accuracy
[18]	DT	87.96%
	RF	85.58%
[32]	RF	93.12%
[27]	RF	97.5%
[33]	RF	90.13%
[34]	DT	95.3%
	RF	95.6%

Table 7 Result Comparison

As each paper had experimented using a different fingerprint dataset, dissimilar indoor space and diverse Wi-Fi equipment. Mentioned reason make it difficult to present a concrete comparison ground between proposed method and existing methods.

V. CONCLUSION

Based on the experiments conducted and results, we conclude that the three algorithms were able to identify most of the locations with good accuracy. However, due to highly unpredictability of Wi-Fi indoor signal, we find it effective to selected very limited features for our experiment. RF based model gave us the most accurate positions based on our experiments. RF was also able to identify positions of small sampled locations. With due data preprocessing models were trained very effectively. Nevertheless, considering the high computational requirement of ML algorithms, mentioned approaches will require further improvements to minimize the resources required. Understandability this study has its limitations and the goal of treating localization as a classification problem may need more research. As majority of the localization based research focus on identifying locality in term of geographical location. Such work present the findings in term of cm or similar scale to categorize the difference from original and predicted location. In future work, we plan to expend experiment area i.e. multiple floors. Our plan also include working on method to identify AP, which could increase accuracy for indoor localization.

REFERENCES

- [1] A. Basiria, E. Lohan, T. Moore, A. Winstanley, P. Peltola, C. Hill, P. Amirian and P. Silvab, "Indoor location based services challenges, requirements and usability of current solutions," *Computer Science Review*, vol. 24, pp. 1-12, 2017 - <https://doi.org/10.1016/j.cosrev.2017.03.002>.
- [2] C. Laoudias, A. Moreira, S. Kim, S. Lee, L. Wirola and C. Fischione, "A Survey of Enabling Technologies for Network Localization, Tracking, and Navigation," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3607-3644, 2018 - DOI: 10.1109/COMST.2018.2855063.
- [3] J. Li, X. Gao, Z. Hu, H. Wang, T. Cao and L. Yu, "Indoor Localization Method Based on Regional Division with IFCM," *MDPI, Electronics*, vol. 8, no. 559, 2019 - DOI: 10.3390/electronics8050559.
- [4] J. Jun, L. He, Y. Gu, W. Jiang, G. Kushwaha, V. A. L. Cheng, C. Liu and T. Zhu, "Low-Overhead WiFi Fingerprinting," *IEEE Transactions on Mobile Computing*, vol. 17, no. 3, pp. 590-603, 2018 - DOI: 10.1109/TMC.2017.2737426.
- [5] E. Lohan, J. Torres-Sospedra, H. Leppäkoski, P. Richter, Z. Peng and J. Huerta, "Wi-Fi Crowdsourced Fingerprinting Dataset for Indoor Positioning," *MDPI, Data*, vol. 2, no. 32, 2017 - <https://doi.org/10.3390/data2040032>.
- [6] O. A. Abdullah and I. Abdel-Qader, "Machine Learning Algorithm for Wireless Indoor Localization," in *Machine Learning - Advanced Techniques and Emerging Applications*, London, IntechOpen-DOI: 10.5772/intechopen.74754, 2018.
- [7] T. Tan, L. Zhang and Q. Li, "An Efficient Fingerprint Database Construction Approach Based on Matrix Completion for Indoor Localization," *IEEE Access*, vol. 8, pp. 130708 - 130718, 2020 - DOI: 10.1109/ACCESS.2020.3009441.
- [8] X. Song, X. Fan, X. He, C. Xiang, Q. Ye, X. Huang, G. Fang, L. L. Chen, J. Qin and Z. Wang, "CNNLoc: Deep-Learning Based Indoor Localization with WiFi Fingerprinting," in *SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI*, Leicester, United Kingdom, 2019 - DOI: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00139.
- [9] S. Srivastava, "iotforall," 2019. [Online]. Available: <https://www.iotforall.com/wifi-indoor-positioning>.
- [10] Z. Farid, R. Nordin and M. Ismail, "Recent Advances in Wireless Indoor Localization Techniques and system," *Hindawi Publishing Corporation Journal of Computer Networks and Communications*, vol. 2013, p. 12, 2013 - <https://doi.org/10.1155/2013/185138>.
- [11] Y. Zhuang, Z. Syed, Y. Li and N. El-Sheimy, "Evaluation of Two WiFi Positioning Systems Based on Autonomous Crowdsourcing of Handheld Devices for Indoor Navigation," *IEEE Transactions on Mobile Computing*, vol. 15, no. 8, pp. 1982 - 1995, 2016 - DOI: 10.1109/TMC.2015.2451641.
- [12] H. Cao, Y. Wang, J. Bi, S. Xu, M. Si and H. Qi, "Indoor Positioning Method Using WiFi RTT Based on LOS Identification and Range Calibration," *MDPI*, vol. 9, 2020 - <https://doi.org/10.3390/ijgi9110627>.
- [13] Y. Zhang, D. Li and Y. Wang, "An Indoor Passive Positioning Method Using CSI Fingerprint Based on Adaboost," *IEEE Sensors Journal*, vol. 19, no. 14, pp. 5792 - 5800, 2019 - DOI: 10.1109/JSEN.2019.2907109.
- [14] J. Brownlee, "How to Develop a Gradient Boosting Machine Ensemble in Python," *Machine Learning Mastery*, 4 May 2020. [Online]. Available: <https://machinelearningmastery.com/gradient-boosting-machine-ensemble-in-python/>. [Accessed 22 Sep 2020].
- [15] F. Zafari, A. Gkelias and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," *IEEE Communications Surveys*

- & *Tutorials*, vol. 21, no. 3, pp. 2568 - 2599, 2019 - DOI: 10.1109/COMST.2019.2911558.
- [16] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp and K. Reddy, "Recurrent Neural Networks For Accurate RSSI Indoor Localization," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10639 - 10651, 2019 - DOI: 10.1109/JIOT.2019.2940368.
- [17] A. H. Salamah, M. Tamazin, M. A. Sharkas and M. Khedr, "An Enhanced WiFi Indoor Localization System Based on Machine Learning," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Madrid, Spain, 2016 - DOI: 10.1109/IPIN.2016.7743586.
- [18] M. Anjum, M. A. Khan, S. A. Hassan, A. Mahmood, H. K. Qureshi and M. Gidlund, "RSSI Fingerprinting-based Localization Using Machine Learning in LoRa Networks," *IEEE Internet of Things Magazine*, vol. 3, no. 4, pp. 53 - 59, 2020 - DOI: 10.1109/IOTM.0001.2000019.
- [19] A. Vashist, D. R. Bhanushali, R. Relyea, C. Hochgraf and A. Ganguly, "Indoor Wireless Localization Using Consumer-Grade 60 GHz Equipment with Machine Learning for Intelligent Material Handling," in *IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas, 2020 - DOI: 10.1109/ICCE46568.2020.9043072.
- [20] M. Kim, D. Han, June-Koo and K. Rhee, "Machine Learning for Practical Localization System Using Multiview CSI," *IEEE Access*, vol. 8, pp. 184575 - 184584, 2020 - DOI: 10.1109/ACCESS.2020.3029598.
- [21] C. L. Nguyen, O. Georgiou and G. Gradoni, "Reconfigurable Intelligent Surfaces and Machine Learning for Wireless Fingerprinting Localization," *ArXiv*, vol. 1, 2020 - DOI: arXiv:2010.03251.
- [22] X. Guo, L. Li, N. Ansari and B. Liao, "Knowledge Aided Adaptive Localization via Global Fusion Profile," *IEEE INTERNET OF THINGS JOURNAL*, vol. 5, no. 2, pp. 1081 - 1089, 2018 - DOI: 10.1109/JIOT.2017.2787594.
- [23] X. Guo, L. Li, N. Ansari and B. Liao, "Accurate WiFi Localization by Fusing a Group of Fingerprints via a Global Fusion Profile," *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, vol. 67, no. 8, pp. 7314 - 7325, 2018 - DOI: 10.1109/TVT.2018.2833029.
- [24] J. Niu, B. Wang, L. Shu, T. Q. Duong and Y. Chen, "ZIL: An Energy-Efficient Indoor Localization System Using ZigBee Radio to Detect WiFi Fingerprints," *IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS*, vol. 33, no. 7, pp. 1431 - 1442, 2015 - DOI: 10.1109/JSAC.2015.2430171.
- [25] X. Guo and N. Ansari, "Localization by Fusing a Group of Fingerprints via Multiple Antennas in Indoor Environment," *IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY*, vol. 66, no. 11, pp. 9904 - 9915, 2017 - DOI: 10.1109/TVT.2017.2731874.
- [26] X. Guo, S. Zhu, L. Li, F. Hu and N. Ansari, "Accurate WiFi Localization by Unsupervised Fusion of Extended Candidate Location Set," *IEEE INTERNET OF THINGS JOURNAL*, vol. 6, no. 2, pp. 2476 - 2485, 2019 - DOI: 10.1109/JIOT.2018.2870659.
- [27] S. Lee, J. Kim and N. Moon, "Random forest and WiFi fingerprint-based indoor location recognition system using smart watch," *Lee et al. Hum. Cent. Comput. Inf. Sci.*, vol. 9, no. 6, 2019 - DOI: 10.1186/s13673-019-0168-7.
- [28] A. Nessa, B. Adhikari, F. Hussain and X. N. Fernando, "A Survey of Machine Learning for Indoor Positioning," *IEEE Access*, vol. 8, pp. 214945 - 214965, 2020 - DOI: 10.1109/ACCESS.2020.3039271.
- [29] A. A. Tokuç, "Splitting a Dataset into Train and Test Sets," Baeldung, 14 January 2021. [Online]. Available: <https://www.baeldung.com/cs/train-test-datasets-ratio>. [Accessed 1 March 2021].

- [30] S. Raschka, "MinMax Scaling," [Online]. Available: http://rasbt.github.io/mlxtend/user_guide/preprocessing/minmax_scaling/. [Accessed 21 8 2020].
- [31] A. Sharma, "Decision Tree vs. Random Forest – Which Algorithm Should you Use?," Analytics Vidhya, 12 May 2020. [Online]. Available: <https://www.analyticsvidhya.com/blog/2020/05/decision-tree-vs-random-forest-algorithm/>. [Accessed 2 March 2021].
- [32] Y. Wang, C. Xiu, X. Zhang and D. Yang, "WiFi Indoor Localization with CSI Fingerprinting-Based Random Forest," *Sensors*, vol. 18, no. 9, 2018 - <https://doi.org/10.3390/s18092869>.
- [33] E. Jedari, Z. Wu, R. Rashidzadeh and M. Saif, "Wi-Fi Based Indoor Location Positioning Employing Random Forest Classifier," in *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Alberta, Canada, 2015 - DOI: 10.1109/IPIN.2015.7346754.
- [34] A. Ç. Seçkin and A. Coskun, "Hierarchical Fusion of Machine Learning Algorithms in Indoor Positioning and Localization," *Applied Sciences*, vol. 9, no. 18, 2019 - <https://doi.org/10.3390/app9183665>.