JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Machine Learning-Based Indoor Localization Using Wi-Fi Fingerprints

Muhammad Harris Hashmi[1]
Department of Computer Science,
University of Alabama at Birmingham, 1402 10th Avenue S., Birmingham, AL 35294, USA

Bhanu Priya Mahant[2]
Director of Research
Voxbourne Services Pvt Ltd 300 Prospect Ave 4H Hackensack NJ 07601

Muhammad Salik Salam
Department of Computer Science,
University of Alabama at Birmingham, 1402 10th Avenue S., Birmingham, AL 35294, USA

Muhammad Ahsan
Department of Computer Science
University of Alabama at Birmingham,1402 10th Avenue S.
Birmingham, AL 35294, USA

Abstract— IPS (Indoor Positioning System) aims to locate individuals and objects within buildings using radio waves, magnetic fields, audio signals, images, and other sensor data. Specific positioning systems, like the Received Signal Strength Indicator (RSSI), rely on the strength of radio signals from emitting devices (beacons) in the area. Wi-Fi, which can be accessed on almost any consumer device, has recently become the most popular method for indoor localization. This research looked at how to enhance collaborations between an indoor Back Propagation-based Neural Network (BPNN) and a Random Forest (RF) by integrating open-source tool & public data repository UJIIndoorLoc databases into Jupyter Notebook. This proposed structure is offered in two stages: offline & online. BPNN trains the network with data from each reference location before turning to RF to acquire an estimated position for the end user. This study presents RF-HAIL (Random Forest-based WiFi fingerprint-based indoor location method), an indoor location technique that uses WiFi fingerprints and has a high degree of versatility so that it may be used in various contexts and with a wide range of devices. In the offline phase, a radio map database is built & BPNN is trained. An unknown point is located in the online setting using the deterministic position method. The RP coordinates of an unidentified point with the highest fingerprint similarity are ones used by the RF method. According to the results of experiments, RF has a lower average localization error and a higher localization precision than the present WKNN-based HAIL approach.

Keywords—Indoor Localization, WIFI Fingerprint, Machine Learning, RSSI, Random Forest, Decision Tree.

I. INTRODUCTION

Positioning technology is crucial for keeping track of resources and people in today's environment. The necessity for automated positioning technologies in a real-time setting is on the rise in modern industries. Positioning technology like Global Navigation Satellite System is accessible for outdoor or open situations. It has been widely used since it is so precise. However, widespread & reliable indoor positioning technology is still lacking [1][2][3]. By using positioning technology, location intelligence enables the development of real-time mapping and tracking apps. Data analysis from location intelligence systems allows decision-makers to see corporate expansion, employee safety, and efficiency possibilities. So, there is a need for reliable indoor location tracking solutions [4][5][6].

There is already a variety of indoor localization systems available. They provide indoor localization also may be roughly categorized into three groups: vision-based, wireless-signal-based, and other methods [4]. Several parameters, including RSSI, time of flight, time of arrival, channel state information, time difference of arrival, as well as time difference of flight, are used by the system in wireless signal-based techniques to forecast the location of mobile devices linked to the wireless system [7][8][9]. To forecast the position of mobile devices linked to the system, the system employs computer vision algorithms with the help of different cameras. RSSI-based fingerprinting is widely used in literature as a wireless-based approach because of its ease and lack of additional hardware requirements [10].

There has been a substantial increase in the usage of WiFi technology in residential, industrial, & commercial settings in recent times, which has led to the acceptance of RSSI-based fingerprinting approaches for indoor localization. Because of this, some different efforts have been attempted to perform indoor localization using Wi-Fi, fingerprinting, Machine Learning (ML), & Received Signal Strength (RSS) metrics [4][5]. There are three kinds of fingerprint techniques: probabilistic, deterministic, & ML approaches, with the first two incurring a substantial computing cost because an ML technique has the potential to be more computationally effective and also is becoming more prevalent [11], that is the approach that we will be using in this work.

The remaining paper is organized as follows. Section 2 discusses most related work of the research topic. Section 3 provides a complete proposed work, including a problem statement, methodology, proposed approaches, proposed algorithm, and flowchart. Section 4 evaluates experimental results and performance metrics. Section 5 summarizes the study with a discussion of future work.

II. LITERATURE REVIEW

This section analyzes research about indoor location. For indoor monitoring & Wi-Fi fingerprinting, many ML techniques, with an emphasis on RFs (Random Forests) & DTs (Decision Trees), are investigated. Wi-Fi indoor localization discusses wifi fingerprinting & RSSI position estimation signals. This section shows the prior work of many researchers on the subject of indoor localization systems.

Tewes et al. (2019) centered indoor location depending on inter-channel measurement. Due to its importance in various applications, indoor location is an active study field. We present a hybrid strategy that combines two valuable signaling and machine learning techniques. Our technique depends upon actual data from 2 possible configurations. “IEEE CTW 2019 - Competition positioning algorithm,” whereby the approach achieves precision with RMSE values of below ten centimeters, demonstrates a highly optimistic outcome. Unceasingly constructing installation in another indoor setting, the algorithm's performance is still superior to that of the most recent indoor location technologies [12].

Li et al. (2019) obtained comparable results in terms of location utilizing the suggested approach. This study presented a multi-client system prototype with a mobile surveyor relying on the crowdsourced fingerprint. A linear regression model is utilized to validate all training devices. During the training phase, geometric distribution generates a non-visible access point as conditional ineligibility for the client. Field test outcomes demonstrate the efficacy & benefit of adapting previous radio maps to identify new – an anti-device by average accuracy of 94 percent in real-world wireless devices [13].

Abbas et al. (2019) introduce WiDeep, an indoor localization solution powered by DL (Deep Learning) that provides precise positioning even in noisy environments. WiDeep consists of a DL model of stacked autoencoders and probabilistic architecture that manages the WiFi signal's acute noise. WiDeep provides a variety of unique modules, for example, to avoid overtraining & manage heterogeneous devices. WiDeep is evaluated in two testing sites of varying sizes and densities. Findings indicate that the average positional accuracy for smaller and bigger testbeds is 2.64m and 1.21m, respectively. This precision provides cutting-edge technology and is effective under all testing conditions for heterogeneous devices [14].

Alitaleshi, Jazayeriy & Kazemitabar (2020) In this study, the Hierarchical Structure of Extreme Learning Machine (H-ELM) is used to generate a floor based on the Wi-Fi fingerprinting approach. This ELM's intricate architecture comprises two components: supervised multiclass classification (original ELM) and multilayer feature encoding with unsupervised learning (ELM-sparse-autoencoder). Identifying H-ELM floors may be more exact than standard ELM floor identification. To evaluate the proposed technique, they accessed data on IT construction from the publicly accessible UJIIndoorLoc dataset. According to the results of our simulations, the suggested Wi-Fi fingerprint floor detection system may obtain a more precise hit rate than existing advanced systems [15].

Chen et al. (2020) presented a novel technique for enhancing real-time location accuracy. At first, the D-CNN model using images of received signal intensity was created. Next, D-CNN outcome prediction errors were used to train the SVR model. Experimentations were conducted using the public database from Spain's Jaume I library. Research results were superior to CNN's findings. In addition, the results demonstrate that the average time suggested by the SVR + D-CNN technique was only 0.612 seconds also that localization precision in an indoor environment was reduced by 86.27% compared to P-CNN + GPR (Gaussian Process Reversal) [16].

Khokhar et al. (2020) The indoor Wi-Fi location is determined using three different ML techniques. Three ML methods were constructed and compared: the DT, RF, & Gradient Booster classifiers. Following creating a floor fingerprint depending upon Wi-Fi signals, the disclosed approaches were used to determine the device's position at 30 separate floor locations. Classification of RF and GB (Gradient Boosting) pinpoints the location of above 90 percent of devices. In contrast, DT could recognize the place with an accuracy above 80% [17].

Koovimol and Pattaramalai (2021) This study is an experiment in ML using LSTM, BiLSTM, & GRU simulations in MATLAB. Adam, RMSProp, & SDGM are the three training options utilized by all ML to evaluate Wi-Fi RSSI fingerprint data. Since three wireless routers were set up in a similar room under three distinct conditions, six pairs of fingerprint data were obtained. First, all MLS & option combinations are simulated and compared to determine the optimal ML. Then, specific parameters are modified to enhance performance. Maximum VA (Validation Accuracy) for the ML GRU with RMSProp is 62.86 percent, minimum LA (Loss Accuracy) is 0.0020 & VL (Validation Loss) is 1.8002, respectively. In conclusion, adapting parameters increases VA by approximately 33 percent and decreases LA & VL by approximately 1.8605 & 2.0947, respectively [19].

III. RESEARCH METHODOLOGY

This section describes a comprehensive research project. The indoor localization method explained in this part also has issues that must be resolved. The general approach that has been suggested, including the gathering of RPs, the selection of APs, the development of a radio map, and the training using a technique of ML, has been explained in depth. Then, the suggested flowchart & algorithm for HAIL-RF are described.

A. Problem Statement

RSS variation over time, device heterogeneity, & considerable fingerprint similarities are three prevalent factors that may significantly reduce the accuracy of the RSS-fingerprint approach for localization in practise. RSS values change over time because of factors such as multidirectional impact, non-line-of-sight (NLOS), and environmental changes. Typically, the RSS variance problem is prevalent during fingerprint generation. Some parts of the total WiFi scanning time may not be able to record many APs, and APs with a low RSS can have a poor response rate. Device heterogeneity causes fingerprint variance, making it more difficult for the customer to discover similar fingerprints in radio map databases supplied by other devices and also reduces location accuracy. Radio map generation is labor-intensive and time-consuming, inhibiting the broad implementation of RSS fingerprinting localization. When walls or mobilizers block WiFi signals, the intensity difference between the transmitting and receiving ends increases dramatically. It's hard to tell the neighbours' RPs since their fingerprints seem identical.

B. Proposed Methodology

Numerous indoor location services have emerged in recent years, thanks to the meteoric rise in the use of smartphones as well as other wireless devices. The term "indoor location" describes the process through which information about a user's physical location within a building is gathered. Within the last decade, a wide variety of sensors & mobile phones for wireless communications have allowed location & monitoring in connection with monitoring & tracing of analog consumers, in addition to enabling numerous applications and associated services. Due to its flexibility and ability to be adapted to different environments and heterogeneous networks, the Wi-Fi fingerprint locating approach established in this research yields accurate findings. The era of unrestrained obstructive context is over. There are two parts to this phase: offline and online. Reference Point (RP) data are collected, and APs are chosen during the offline stage. Because of this, not all scanned APs should be used for localization when selecting APs. In practical indoor settings, many APs are scannable and unnecessary. Reducing overlap is vital for two motives.

On the one hand, specific APs may only be detected in a relatively small region, which adds nothing to fingerprint measurement. APs may exacerbate the fingerprint chaos by low RSS values, which often exhibit significant temporal volatility and tiny appearance. After training the Random Forest, fingerprints were mapped using a radio map database. It includes rough positioning for AP rankings at Test Points RSS and precise positioning by RSS values, with weights of each higher RP determined by RF depending upon RSS values, both of which take place throughout the online phase. Additionally, the location of the consumer is derived using weighted average coordinates. Finally, the device's location has been pinpointed at each of these places.

a) RP Collection

To acquire accurate fingerprint data from all RPs in the area, automated collectors are used in RP collection. Initially, numerous equidistant locations are identified as RPs & their associated coordinates are recorded throughout the appropriate area. While at each RP, the collector pauses to record the RSS values for Nsam samples for all scanned APs. If the RSS variation is reduced over time, Nsam should be more than 50. Finally, the averaged AP RSS values serve as our definition of the sample modes. [20].

b) AP Selection

Global average RSS levels determine the selection of the AP. APs with global RSS averages over a predetermined Amin threshold are chosen to form a new S database for other radio map construction. Mainly chosen APs have requirements they must meet.

$$\frac{1}{N_{RP}}\sum_{i=1}^{N_{RP}} RSS_I \geq A_{min}, \qquad (1)$$

NRP is the total number of RPs, while RSSi is the value measured at APi. It is crucial to remember that the RSS value calculated in this RP is set to C dBm if RP is not accessible to scan an AP. In addition, the actual atmosphere is used to establish the Amin threshold, and C is less than the least RSS scanned value.

c) Radio Map Construction

It is impossible to locate a fingerprint without a radio map. Measurements must be taken at several reference sites to collect representative RSS samples in the study region. Radio mapping usually entails activities like fingerprint collection and location labelling to represent the area accurately. Below are the directions for constructing the radio map, which relies on S [21].

Radio map Φ is shown by

$$\Phi = \{(fp_i, p_i, \mathrm{i} = 1, \dots \dots, N_{RP})\}, \qquad (2)$$

where fp_i is a fingerprint of RPi, Pi is the coordinates of RPi, & NRP is the total number of RPs.

To characterize fingerprint fpi, a NAP ×3 matrix is used, with NAP standing for "number of APs selected."

$$fp_i = \begin{bmatrix} x_1^i, & rank_1^i, & pair_1^i \\ \dots, & \dots, & \dots, \\ x_{N_{RP}}^i, & rank_{N_{RP}}^i, & pair_{N_{RP}}^i \end{bmatrix} \qquad (3)$$

Matrix in (3.3) is defined below:

- $x^i = \left(x_1^i, \dots \dots, x_{N_{RP}}^i\right)^T$ is a vector, which includes in descending order measured RSS values for all APs in S.
- $rank^i = \left(rank_1^i, \dots \dots, rank_{N_{RP}}^i\right)^T$ is a vector, which includes in similar order corresponding BSSIDs as xi. $rank_1^i$ is BSSID of x_1^i in particular.
- $pair^i = \left(pair_1^i, \dots \dots, pair_{N_{RP}}^i\right)^T$ is set that comprises N_{RP} Subsets. When $j = 1,2 \dots, N_{RP}, pair_1^i$ is built as per the following rules: if $x_g^i (g = j+1, j+2, j+N_{key2})$ is smaller than $(x_j^i - \beta)$ dBm, i.e., $x_g^i \leq (x_j^i - \beta), rank_g^i$ will become an element in a set $pair_j^i$. In addition, the measured RSS value of APj is β dB greater than APg if $j = N_{key1}+1, \dots, N_{AP}, pair_j^i = \Phi$ is the agreeable pairing relationship among APj & APg. To enhance the reliability of agreeable pair association & decrease the amount of time consumed on RSS variance, parameters N_{key1}, N_{key2} and β are adjusted.

d) Machine Learning Techniques for Training

In recent years, image recognition using Machine Learning (ML) has attracted lots of attention. ML has advanced in several areas, like image recognition, language translation, wireless transmission scheduling, and self-driving, providing a robust framework for dealing with a wide range of challenging practical situations. In theory, ML uses real-world data to train ML solutions to monitor intricate relationships among input data (features) & output values (labels). ML-based location may be used with any of the localization above strategies. Direct techniques, in which the location of the observations is explicitly provided, are the most apparent option for a full-stack ML solution. However, ML may be coupled with other location analyses [22]. ML has also been discovered to be an effective method for combining multidimensional data collected from various sensors, technologies, and locations. These challenges have recently been widely studied, and ML approaches have shown decent performance. There are some issues with conventional indoor localization techniques that may be resolved by using ML algorithms. Large-scale indoor localization systems in places like shopping malls, multi-story buildings, & airports with massive training sets often cannot function successfully using conventional methodologies [23]. ML strategies may be implemented in several various ways. There are two broad categories of ML methods known supervised and unsupervised. In the following sections, we provide a concise overview of supervised methods.

In this study, the following machine learning methods were employed:

- **Decision Tree (DT)**

A decision tree, also known as a classification tree, is a powerful supervised learning technique. It creates a branching chart or tree to display all possible outcomes of a choice. When describing a decision tree, each node assesses a feature, every branch is compared to the outcome of the parent node, and each leaf provides the class label. The sample is categorized using a top-down approach, first at the broadest level. For a given feature or node, the path leading to the data point value for that characteristic is investigated until the leaf is located or the label is selected [24].

DT is a kind of tree in which the internal nodes (trends in the data input) can be interpreted as tests as well as leaf nodes (the outcomes) can be interpreted as groups (of these trends). In general, it is a template matching model that checks the given input x against the prototypes in μ_k.

$$\varphi(x) = [\kappa(x, \mu_1), \ldots, \kappa(x, \mu_N)] \tag{4}$$

Instead of using a predetermined kernel function κ, basis function φ is learned from input data by selecting related attributes (that is adaptive base function model of the type method belongs officially).

$$f(x) = \omega_0 + \sum_{m=1}^{M} \omega_m \varphi_m(x)) \tag{5}$$

Each area's response value is represented by its weight, defined by the base function as a portion of the input feature space partitioned at nodes [25].

- **Random Forest (Decision Tree Forests)**

As one of the more established ML techniques, RF has seen usage across various domains. RF is a novel technique that can be employed not only for classification but also for regression analysis. It is presumed that several types of trees populate the forest using the RF technique. An abundance of trees makes a forest more robust & secure. The same holds for many trees in the forest; this is when the random approach is most effective. This improves the accuracy of the RF approach. In the RF method, several different types of DTs are used. All DTs in a random forest may be broken down into individual nodes [26].

The results from the training of several individual DTs are then averaged. The uncertainty of a forecast is reduced when many estimates are averaged. Following is a detailed explanation of how the ensemble is determined:

$$f(x) = \sum_{m=1}^{M} f_m(x) \tag{6}$$

Where m'th tree is f_m. Because the same learning approach is being applied to different subsets of data, the resulting predictors may be strongly correlated with one another. To remove correlation from the fundamental learners, RFs analyze trees based on a randomly selected subset of input variables [25].

C. Proposed Algorithm

Following is a description of algorithm one, which contains the pseudocode for the HAIL-RF localization algorithm:

Algorithm 1: Pseudocode of HAIL-RF Localization Algorithm

Input: Fingerprinting at n-test points ($Test_{n\text{-}pt}$), radio map.
Output: Predicted Position of n-test Points ($Test_{n\text{-}pt}$).

Method:

1. Set required input parameters and reference points (Rfr_{pt})
2. For every Rfr_{pt} in the radio map comprises k^{th} Rfr_{pt} fingerprint and k^{th} Rfr_{pt} co-ordinates where k is ranged from 1 to overall Rfr_{pt}
3. Firstly, for the execution of complete $Test_{n\text{-}pt}$ parameter's ranking for correspondent pairings & matching with appropriate BSS_ID ranking of k^{th} Rfr_{pt}.
4. Compute the number of CEs (Common Elements) in a pair of $T_{n\text{-}point}$ & i^{th} Ref_{point}
5. Compute summation of all calculated CEs number
6. Compute common APs (access points) correspondent pairs (that is, score) amid $Test_{n\text{-}pt}$ & k^{th} Rfr_{pt}
7. End of For Loop
8. End of For Loop
9. Constructed collection of top Rfr_{pt} and all Rfr_{pt} are organized in descendant sequence as per their calculated score values
10. For every top $Rfer_{pt}$ do
11. Train the random forest model
12. Compute the weights value of the top Rfr_{pt}
13. End of For Loop
14. Compute $Test_{n\text{-}pt}$ position co-ordinates
15. A user's finally predicted location coordinates are stored.
16. End

D. Proposed Flowchart

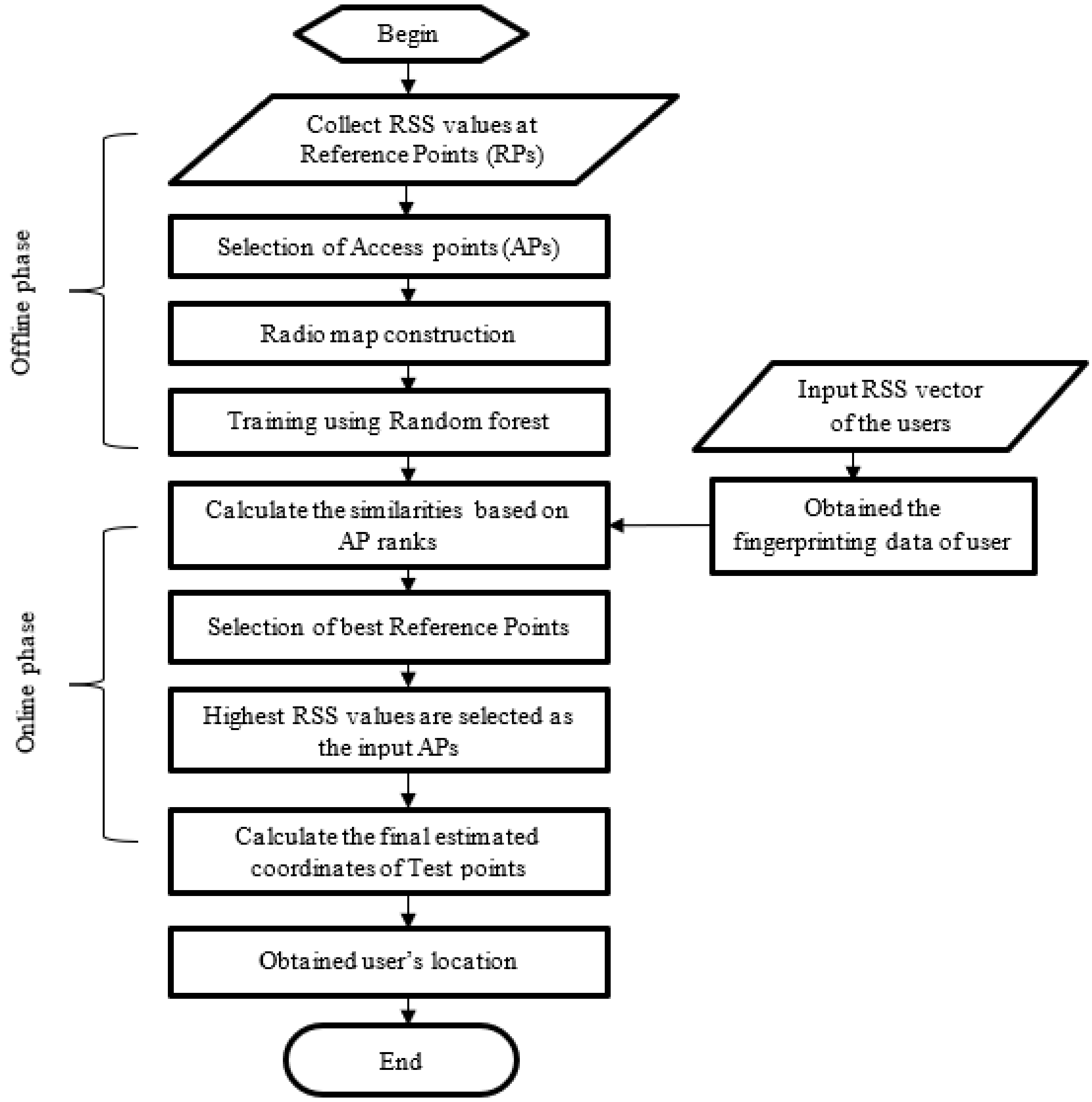


Fig. 1: Flowchart of Proposed HAIL-RF Model

The recommended HAIL-RF model is shown in Figure 1 below, and it is a flowchart built on the random forest ML method. The whole process is now broken up into online & offline phases. First, RSS values from reference points are gathered, and then APs (Access Points) are chosen. It is time now to build a radio map. After the radio map has been built, it can be trained with ease thanks to ML methods like RF. The next step is to rank the available access points and select the best reference locations depending upon calculated similarities. Here, the RSS vector value of the user is input, and the person's fingerprint information is acquired, both of which are used to find patterns of similarity. Select RSS feeds with the highest value as input to the access points. At last, estimate the coordinates of test locations to identify the users' precise location.

IV. RESULTS AND DISCUSSION

This section describes simulated results or consequences of the applied RF-HAIL strategy. The effectiveness of the executed strategy is shown by a visual depiction of the obtained outcomes. Here, we use the accuracy and error values determined by the train & test sets as our output parameters. Python's simulation tool was utilized for this study.

A. Performance Metrics

To evaluate the performance of the deployed RF-HAIL strategy, I conducted a battery of experiments using wireless fingerprints indoor location. The goal of the test suite was to determine which set of our parameters yields the best performance of the proposed system. For this reason, we have modified the system configuration settings [27].

True positive (TP) indicates that a classifier accurately predicted the number of positive occurrences, whereas false negative (FN) indicates that a classifier incorrectly assessed the number of negative examples. The number of positive occurrences.

The F1-score, precision, & recall are often employed as Performance Metrics in imbalanced domains and are derived from findings in the confusion matrix.

- **TP (True Positives):** This is the sum of all good outcomes that were accurately anticipated.
- **TN (True Negatives):** It was correct to expect negative values.
- **FP (False Positives):** False positive projected value when the actual class is no. However, the predicted class is yes.
- **FN (False Negatives):** False negative prediction real class, but predictions are made in class.

1) **Accuracy:** As a performance metric, classification model accuracy is the ratio of correctly predicted values to total values. A model is considered to be of more excellent quality when its accuracy is increased. It is a correct solution when the dataset is almost symmetrical, and FP & FN are nearly identical.

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \quad (7)$$

2) **Precision:** Positive values accurately predicted the overall positive forecasts. The greater the accuracy, the lower the false positive rate.

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

3) **Average Localization Error:** This (measured as the square root of MSE) is a function of RSS measurement standard deviation for a given technique and environment. The discrepancy between the true position as well as estimated position.

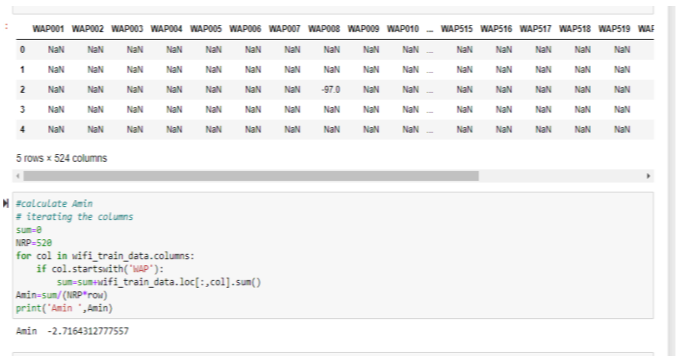
4) **Complexity:** It is possible to divide an IPS's complexity into software & hardware components. Hardware complexity is responsible for selected technology & IPS signal measurement method. Challenge stems from the computational burden required to find the software. In a server-based intrusion prevention system, localization algorithms are run on a centralized server, which permits quick positioning calculations due to its powerful processing capability & ample power generation.

5) **Scalability & Robustness:** The ability of the localization system to continue working correctly despite changes to the target area's location and/or signal source is what scalability in IPS means. The adjustments may include an enlargement of the target area or a broader range for the sent signals.

6) **Cost:** Implementing and maintaining an IPS system represents the resources invested in this endeavour. Cost is very conditional on factors like the site's size, the accuracy required, the technology used, the amount of power consumed, and so on.

B. *Results of the Implemented Approach*

The findings of the tests carried out in this part are shown, and these results include a description of the model's performance and accuracy. The field test results demonstrate the efficacy of the prevalent RF Classifier in determining the indoor location of a stimulus room. The classification has a high granularity level and has been used in the semantic localization of indoor environments. Comparative analyses of the proposed Random forest-based adaptive indoor localization system and other classifiers have been conducted. Wi-Fi fingerprinting data set developed by UJIIndoorLoc. All 21049 entries from the recommended database were shown. Each of the 529 records has a direct connection to a specific capture and provides the value of a number element. These range from 001 to 520 RSSI levels, 521 to 523 real sample point world coordinates, 524 for BuildingID, 525 for SpaceID, 526 for Relative position respecting spaceID, 527 for UserID, 528 for PhoneID, and 529 for Timestamp.



	WAP001	WAP002	WAP003	WAP004	WAP005	WAP006	WAP007	WAP008	WAP009	WAP010	...	WAP515	WAP516	WAP517	WAP518	WAP519
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	-97.0	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN

5 rows × 524 columns

```
#calculate Amin
# iterating the columns
sum=0
NRP=520
for col in wifi_train_data.columns:
    if col.startswith('WAP'):
        sum=sum+wifi_train_data.loc[:,col].sum()
Amin=sum/(NRP*row)
print('Amin ',Amin)
```

Amin -2.7164312777557

Fig. 2: Dataset and A_{min} calculated

As demonstrated in Fig. 2, the proposed database comprises a 520-element vector of integer values that reflects 98% of the data in each record (520 vector locations out of 529). While numbers represent RSSI levels, Vector locations map to unique WAP IDs (MAC addresses). For each WAP that has been located, its MAC address and relative strength are shown here. MAC addresses are encoded as strings, and RSSI values correspond to negative integer values used to measure dBm; for example, a value of -100dBm indicates a very weak signal, while a value of 0dBm indicates a robust signal to the determined WAP. Since NaN is used in place of -100dBm. Figure 2 also displays Amin's threshold value, which is -2.716.

Fig. 3: Selected WAPs

As many WAPs as possible have been chosen using the Amin threshold. Each record in the database has a single Wifi scan with the raw intensities of discovered WAPs, and the total number of WAPs shown in the database is 520. Even when WAPs are present in a scan, not all of them will be immediately apparent. For WAPs whose identities could not be determined during the scan shown in Figure 3, RSSI levels remain constant, by default making use of the +100dBm artificial value of WAPs which were not detected by the device.

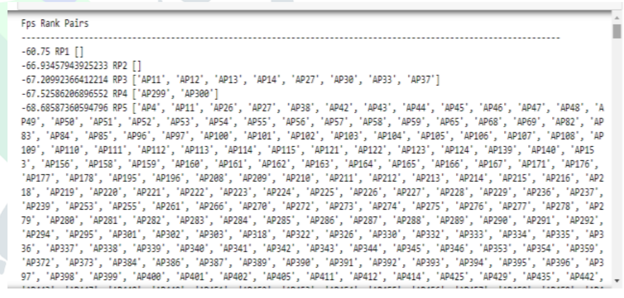


Fig. 4: Rank and Pairs

Fig. 4 illustrates the screenshot that displays the ranking of the fingerprints & their pairing with other APs. This design is derived from the database, including the picked APs. To begin, the selected APs are arranged in descending order according to the measured RSS values. Since specific APs cannot be scanned, C dBm has been assigned as a value for RSS measured at the RP location for these APs. As a result, the corresponding rankings are determined by the BSSIDs of the selected APs. Last but not least, the proposed rule may be used to produce pairings in which Nkey1, Nkey2, & are each equal to 6, 4, & 3, respectively.

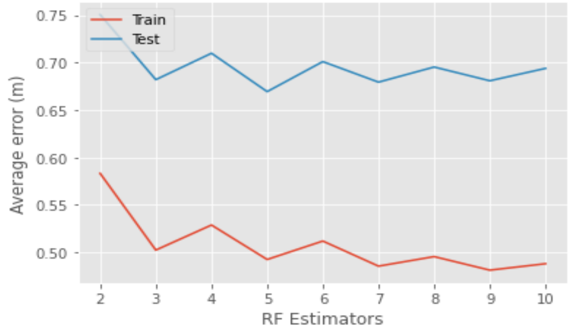


Fig. 5: Average Error Performance of RF Classifier at Corridor

Fig. 5 shows their average errors as a function of several RF estimators ranging from 2 to 10. It signifies train & test error. According to the graph, it can be seen that the training error has decreased together with the increasing number of RF estimators. Similarly, test error has been depicted, which is greater than train error.

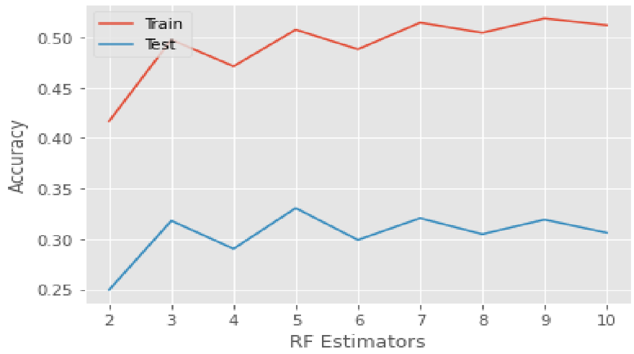


Fig. 6: Accuracy Performance of RF Classifier at Corridor

Fig. 6 illustrates the accuracy of their training and testing procedures as the number of RF estimators increases from 2 to 10. It also suggests that the RF estimator, abbreviated as RFE, is connected to the accuracy of the localization, and in the subsequent experiments, we set the RFE value to 6. The accuracy of the test is represented by the blue colour, while the red colour represents the precision of the train. The findings indicate that accuracy of signal localization is worse than that of RF-HAIL, particularly concerning the RSS value calculated using BPNN.

V. CONCLUSION AND FUTURE WORK

In recent years, there has been a noticeable increase in demand for precise localization inside indoor environments. Based on the WiFi signals used by the WiFi infrastructure deployment and the availability of mobile devices authorized over WiFi, localization solutions may be relatively cost-effective. The fluctuations in WiFi signals caused by many obstacles, like walls, furniture, & people moving through the space, are referred to as interference in multipath. This is one of the most challenging aspects of an indoor setting. These problems were identified, and a solution was suggested as a high-adaptability indoor location method. This method is known as RF-HAIL, and it is a precise and resilient approach to an indoor location that can be applied to a variety of contexts on a wide variety of devices and adapted to various environments. To improve the speed at which localization accuracy may be increased and the number of mistakes that can be made, RF-HAIL offers a revolutionary strategy that uses the advantages offered by both AP rank & RSS. With support for RSS-based Random Forest Value, this approach may be applicable in a wide range of situations. The location system with automated collectors may make the installation process more manageable.

Based on RF recommendation, experiments have been conducted and compared with WiFi fingerprint data from the UJIIndoorLoc database. The proposed system has been tested and shown to work admirably in terms of accuracy. This is crucial since the system's performance remains consistent regardless of the depth of the data utilized to place the customer. In-depth experiments have demonstrated that RF-HAIL can offer high localization accuracy, mitigating effects of RSS variations, environmental complexity, & device heterogeneity.

Based on performed experiments & analysis of findings, it can be determined that implemented RF-HAIL technique accurately identified the majority of the sites. Because of the significant unpredictability of Wi-Fi indoor signals, we deem it most beneficial to choose just a small number of features for this experiment. RF-based models provided the most precise placements depending on the results of our studies. Additionally, RF was able to detect the locations of tiny sampling places. With proper data preparation, models were efficiently trained. Nevertheless, given the high computing requirements of ML algorithms, the methods above will need more refinement to reduce the number of resources needed.

For future work, we want to boost system accuracy by developing more complicated localization algorithms that use the temporal distribution of potential consumer locations to reduce false positive results. In addition, we will investigate the analytical method of using indoor signal propagation models for multilateration.

REFERENCES

[1] J. Huang, H. Luo, W. Shao, F. Zhao, and S. Yan, "Accurate and robust floor positioning in complex indoor environments," *Sensors (Switzerland)*, 2020, doi: 10.3390/s20092698.

[2] R. Chen and L. Chen, "Smartphone-Based Indoor Positioning Technologies," in *Urban Book Series*, 2021.

[3] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Modern WLAN Fingerprinting Indoor Positioning Methods and Deployment Challenges," *IEEE Communications Surveys and Tutorials*. 2017, doi: 10.1109/COMST.2017.2671454.

[4] N. Singh, S. Choe, and R. Punmiya, "Machine Learning Based Indoor Localization Using Wi-Fi RSSI Fingerprints: An Overview," *IEEE Access*. 2021, doi: 10.1109/ACCESS.2021.3111083.

[5] F. Zafari, A. Gkelias, and K. K. Leung, "A Survey of Indoor Localization Systems and Technologies," *IEEE Commun. Surv. Tutorials*, 2019, doi: 10.1109/COMST.2019.2911558.

[6] S. M. Asaad and H. S. Maghdid, "A Comprehensive Review of Indoor/Outdoor Localization Solutions in IoT era: Research Challenges and Future Perspectives," *Comput. Networks*, vol. 212, p. 109041, 2022, doi: https://doi.org/10.1016/j.comnet.2022.109041.

[7] J. Wang and J. Park, "An Enhanced Indoor Positioning Algorithm Based on Fingerprint Using Fine-Grained CSI and RSSI Measurements of IEEE 802.11n WLAN," *Sensors*, vol. 21, p. 2769, 2021, doi: 10.3390/s21082769.

[8] H. Rizk, A. Elmogy, and H. Yamaguchi, "A Robust and Accurate Indoor Localization Using Learning-Based Fusion of Wi-Fi RTT and RSSI," *Sensors*, vol. 22, p. 2700, 2022, doi: 10.3390/s22072700.

[9] T. Yang, A. Cabani, and H. Chafouk, "A survey of recent indoor localization scenarios and methodologies," *Sensors*. 2021, doi: 10.3390/s21238086.

[10] J. M. Rocamora, I. W. H. Ho, W. M. Mak, and A. P. T. Lau, "Survey of CSI fingerprinting-based indoor positioning and mobility tracking systems," *IET Signal Processing*. 2020, doi: 10.1049/iet-spr.2020.0028.

[11] A. H. Salamah, M. Tamazin, M. A. Sharkas, and M. Khedr, "An enhanced WiFi indoor localization System based on machine learning," 2016, doi: 10.1109/IPIN.2016.7743586.

[12] S. Tewes, A. A. Ahmad, J. Kakar, U. M. Thanthrige, S. Roth, and A. Sezgin, "Ensemble-based learning in indoor localization: A hybrid approach," 2019, doi: 10.1109/VTCFall.2019.8891416.

[13] Y. Li, S. Williams, B. Moran, and A. Kealy, "A Probabilistic Indoor Localization System for Heterogeneous Devices," *IEEE Sens. J.*, 2019, doi: 10.1109/JSEN.2019.2911707.

[14] M. Abbas, M. Elhamshary, H. Rizk, M. Torki, and M. Youssef, "WiDeep: WiFi-based accurate and robust indoor localization system using deep learning," 2019, doi: 10.1109/PERCOM.2019.8767421.

[15] A. Alitaleshi, H. Jazayeriy, and S. J. Kazemitabar, "WiFi Fingerprinting based Floor Detection with Hierarchical Extreme

Learning Machine," 2020, doi: 10.1109/ICCKE50421.2020.9303624.

[16] H. Chen, B. Wang, Y. Pei, and L. Zhang, "A WiFi Indoor Localization Method Based on Dilated CNN and Support Vector Regression," 2020, doi: 10.1109/CAC51589.2020.9327326.

[17] Z. Khokhar and M. A. Siddiqi, "Machine Learning Based Indoor Localisation Using Wi-Fi And Smartphone," *J. Indep. Stud. Res. Comput.*, 2020, doi: 10.31645/06.

[18] G. Junlin, Z. Xin, W. Huadeng, and Y. Lan, "WiFi fingerprint positioning method based on fusion of autoencoder and stacking mode," 2020, doi: 10.1109/ICCST50977.2020.00075.

[19] P. Koovimol and S. Pattaramalai, "Experimental Machine Learning for RSSI fingerprint in Indoor WiFi Localization," 2021, doi: 10.1109/ecti-con51831.2021.9454865.

[20] D. Hou *et al.*, "Robust indoor localization in ultra dense networks: A fingerprint similarity approach," 2017, doi: 10.1109/WCSP.2017.8170985.

[21] J. Xue, J. Liu, M. Sheng, Y. Shi, and J. Li, "A WiFi fingerprint based high-adaptability indoor localization via machine learning," *China Commun.*, 2020, doi: 10.23919/J.CC.2020.07.018.

[22] R. Zhang, X. Jing, S. Wu, C. Jiang, J. Mu, and F. Richard Yu, "Device-Free Wireless Sensing for Human Detection: The Deep Learning Perspective," *IEEE Internet Things J.*, 2021, doi: 10.1109/JIOT.2020.3024234.

[23] A. Nessa, B. Adhikari, F. Hussain, and X. N. Fernando, "A Survey of Machine Learning for Indoor Positioning," *IEEE Access*, 2020, doi: 10.1109/ACCESS.2020.3039271.

[24] K. Das and R. N. Behera, "International Journal of Innovative Research in Computer and Communication Engineering A Survey on Machine Learning: Concept, Algorithms and Applications," *Int. J. Innov. Res. Comput. Commun. Eng.*, 2017.

[25] R. Clark, M. Hong, S. Sasaki, and S. Huang, "Indoor Localization based on WiFi Fingerprinting."

[26] T. S. Kumar, S. S. Vennam, and S. S. Chamarti, "EasyChair Preprint A Survey on Machine Learning Algorithms," 2020.

[27] C. Science, D. Learning, and F. N. Detection, "MASTER ' S THESIS Fake News Detection-A Deep Neural Network," no. June 2019.