

Position Estimation at Indoors using Wi-Fi and Magnetic Field Sensors

Ganduri Chandra

Theatro Labs India Pvt Ltd, Bangalore, India

Abstract

Indoor location estimation is important for many applications, including home automation, security, and surveillance, to name a few. This is accomplished via sensors such as cameras and passive infrared (PIR) sensors, as well as radio frequency-based technologies such as Wi-Fi, Bluetooth, and radio frequency identification (RFID). Because of its ease of implementation, Wi-Fi technology is the most extensively used. Aside from Wi-Fi, a growing number of researchers are experimenting with the sensors embedded into mobile phones. To increase the performance of the localization, deep learning, probabilistic, and statistical algorithms are applied to the raw data collected from sensors and Wi-Fi access points. In this paper, we present a sensor-based indoor localization test-bed. We feed data from magnetometer sensors and Wi-Fi access points to Machine Learning (ML) algorithms. With our proposed method, we get a very high accuracy of 98%. Performance analysis of the ML-classifiers' dependence on the size of the test data set is also explored in detail.

Keywords

Extreme Gradient Boosting(XG-Boost), Indoor Localization, Machine learning, Magnetic Field Sensor, Random forest, Wi-Fi RSSI.

1. Introduction

It is well-known that the Indoor Localization Systems (ILS) – also referred to as Indoor GPS – detect the location of a user inside a room with high accuracy. Using this information provided by these ILS, Indoor Navigation Systems (INS) guide a user through optimal paths and with “turn-by-turn” directions. In the working of these systems, the GPS signals from satellites are ineffective due to their poor position resolutions indoors, and attenuation by walls and iron materials etc. So, a variety of ILS and INS use different technologies such as WiFi [1], RFID [2], Bluetooth [3] and Visible light communication [4] etc. Among these, Wi-Fi is the most commonly used method for its pervasiveness.


For a user's location estimation, Wi-Fi based localization methods offer a wide range of solutions: Angle of Arrival (AoA), Time of Arrival (ToA), Time Difference of Arrival (TDoA), Return Time of Flight (RToF), Phase of Arrival (PoA) (see [5] for an account), Particle Filtering [6] and MAP estimation [7]. In both particle filtering and MAP based approaches, the training phase primarily requires a prior knowledge of the building design map, and the new test data will be compared with the existing database for an accurate location estimation.

IPIN 2022 WiP Proceedings, September 5 - 7, 2022, Beijing, China

✉ gjnv.manichandra@gmail.com (G. Chandra)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

Various fingerprinting approaches have recently been studied and deployed in real time [1]. Many works [8] and [9] have demonstrated that fingerprinting-based approaches achieve high precision. In the methods involving Wi-Fi – for instance in all of the above discussed approaches – huge collection of data is required for calibrations, and this one of the major drawbacks. The range-based approaches compute the distances from a user’s end device and Wi-Fi access points, and employ geometrical models (for example, triangulation model) to estimate the user’s location. Several Path loss models, namely Log distance Path loss model, have been investigated in [10] and [11] to estimate distances based on the Received signal strength (RSS) information. As RSS is in general not stable indoors, Channel State Information (CSI) based ranging has been introduced.

Notably, majority of the aforementioned approaches are not stand-alone, i.e., these cannot be implemented independently. More precisely, Wi-Fi based localization requires the use of a network, which raises the concerns about data security. Furthermore, when a small number of access points are installed, using alone a Wi-Fi approach is not very effective. Also, these approaches necessitate prior knowledge of the Geo-graphical locations of the access points deployed indoors. Wireless fingerprinting-based localization is the most popular among all the localization techniques, due to its ease of deployment. The Radio Frequency signals (RF signatures) collected from the access points help in the prediction of the location accurately. Typically, RF measurement values and the Received Signal Strength (RSS) values of WiFi or Bluetooth or any other wireless technologies, are one and the same.

Indoor localization methods which are based on the signatures of magnetic field generally assume that these magnetic fields do not vary much within a smaller range of altitudes; for example, see [12]. This assumption is known to be very reasonable, and applies well to the situation where a magnetic field sensor (to measure the magnetic field signatures) is placed at any point between the surface of the floor and ceiling of a room; note that pillars/beams are usually made of iron bars which are ferromagnetic. More generally, [13] considers a larger range of altitudes for placing a magnetic field sensor, such as the altitudes of buildings; and also other factors, for example buildings are made of multiple vertical and horizontal beams (which are ferromagnetic and hence can influence the measurements by the sensors). One of the purposes of his article is to critically quantify the impact of altitude on the accuracy of traditional Machine Learning classifiers, for position estimation indoors (across the height of a room) using magnetic field data. In particular, we devised small-scale experiments conducted in slightly restricted environment (departmental laboratory) at two different altitudes (2 feet and 4 feet), that aided us in answering few questions such as:

1. Does the magnetic field at indoors vary significantly with altitude?
2. Suppose that the altitude is same for both training and testing an ML model. Then what is the role of the device’s location height whose magnetic field intensity we measure?

The broad goal of the present paper is to provide an ML-based solution to estimate the location information with the data extracted from a realistic embedded test bed. Our contributions in this paper are as follows:

1. We present an efficient method for describing the user’s exact location at Indoors with high resolution.

2. The novel aspect of our method: we used a multi-label and multi-output approach to predict the location of the user's end device.
3. We validate our results via training and testing the data we collect from a realistic test-bed.
4. We also investigate the impact of the sensor's height on the prediction accuracy.

Rest of the paper is structured as follows: Section 2 provides a synopsis of some related works. Section 3 discusses the motivation of the paper. The Experimental setup which includes brief description of the sensors used in this work is presented in Section 4. Section 5 presents the experimental results. Section 6 discusses some concluding remarks.

2. Related Works

In recent past, extensive research has been conducted towards indoor localization and the boosting of the prediction accuracy. In this, many researches have proposed a wide range of novel schemes for Indoor Localization/Navigation which involve the technologies, to name but a few, RF-ID[2], Bluetooth [3] and Wi-Fi [1].

In [2], the authors proposed a 'sensor by proxy' approach that involves sensor fusion of Radio frequency identification measurements to reduce false-off error rates in Localization indoors. In particular, they derived a novel score function, which is used in developing distance based algorithms, with an aim to guarantee maximum coverage using a limited number of sensors indoors. The solutions provided in this work yielded localization with a reasonably good accuracy. But sensor based approaches in general have some drawbacks/ limitations in real time; for instance, limited range of the coverage area when deployed. In [14], some device-free techniques to analyze the user's behavior through Wi-Fi signals at shopping malls have been employed. Usually, Received Signal Strength Information(RSSI) and Channel State Information(CSI) data are used for localization of users indoors, see [1]. The authors of this paper claimed that their proposed method will work efficiently even when the user is not connected to any of the access points. For real time deployment, the approach presented in [14] will work efficiently only if the client devices are embedded with specific hardware: Intel 5300 802.11n Wi-Fi NIC. Device-free Localization Systems with accurate positioning have been proposed in [15]. The work in [15] involves combining the RSSI of the user's end device with a filtering technique for estimating the user's location inside a closed environment. In the same paper, it has been claimed that the approach does not require any additional/dedicated hardware infrastructure.

In [16], a Machine learning based algorithm has been proposed, and a reasonably good accuracy of around 85% has been shown. In the same paper, the sensors utilized were accelerometer, Wi-Fi, light, color and sound, and the approach therein involves translating the pixels from the light sensor to Hue-Saturation-Lightness (HSL). Also, the classifier's used therein were: K-means clustering (distance based approach) and Support Vector Machine (SVM). Recall, considering SVM for localization is not a best suitable solution, since it can be used only for binary classification, i.e., one can only predict the presence or absence of the users.

The work in [17] proved that magnetic field strength can vary in different rooms within buildings. The authors of this paper also concluded that give a place in a building, the magnetic signatures there are unique, and hence the magnetic field data can be used for Localization using

both Machine Learning and Deep Learning (DL) frameworks. In the aforementioned paper, the coefficients extracted after applying Fast Fourier Transform (FFT) to the magnetic field data were fed to Convolutional Neural Networks (CNN) for real time predictions. Hybrid Deep Learning Model (HDLM) based localization systems, which only utilize RSSI heat maps instead of the raw RSSI signals from the access points, have been proposed in [18], resulting in a better localization performance of the Wi-Fi RSSI signal based positioning systems. HDLM is the combination of convolutional neural network and Long Short-Term Memory (LSTM) network. Also, it has been claimed in this paper that HDLM achieves reasonably good accuracy prediction with very less localization error, When compared to the existing solutions that involve the approaches viz fingerprint, trilateration, and Wi-Fi fusion etc.

A Deep Learning based approach has been introduced in [19] to predict the location of a user indoors. Although Deep Learning approaches outperform very well, the bottleneck is that training the network requires huge data and time. In [20], the authors used Long Short-Term Memory. In general, LSTM networks are used for predicting time series data by learning the data collected at previous time instances. For real time deployment, all the neural network based models consume huge amount of time while training and processing, as well as occupy more memory.

In general, in any positioning approach, the evaluation of the accuracy involves the calculation of the Euclidean distance between the true and estimated positions of the users. One major drawback in this is that the floor transitions are being ignored. Dealing with some more challenges in predicting an accurate position of users indoors, the work in [21] introduces a new methodology to measure the positioning error, which considers the length of the user's path that connects the estimated position to the exact position using the visibility graph for the floor map, so as to compute the shortest distance, navigational meshes for the vector maps to identify similar paths with less computation time.

To improve the user localization accuracy using RSSI and the complexity over widely used fingerprinting methods, an efficient compressive sensing based approach have been introduced in [22]. Localization using Least Square Lateration(LSL) method has been introduced in[23]. This paper also discussed a comparative analysis of the accuracy with LSL against that of with Pure lateration(PL), by utilizing different Gaussian noise parameters for varying number of access points and varying distance between the access points and user's end device. Further, it has been proven (in the same paper) that least-square algorithm with curve fitting approach provides significant performance improvements when compared to PL. A novel scheme for localization based on the Angle Difference of Arrival information and triangulation using mm-waves indoors has been proposed in [24]. Also, it has been stated that their proposed algorithm outperforms very well at both vacant as well as office environments, by using a commercial 60-GHZ mm-wave test bed for their experiment. Though the results therein are promising, deploying the same for real time usage will be the bottleneck given that their equipment is expensive.

To study and understand the behaviour of the Earth's magnetic field inside a room, a novel methodology has been introduced in [17, 25, 26]. In these, Dense and spatially referenced samples of the magnetic vector field towards the surface of the ground as well as the free space above, was considered for both the analysis and also the map estimation of the surroundings. The approach followed in [6] utilizes both the Rao-Blackwellized particle filter as well as kalman

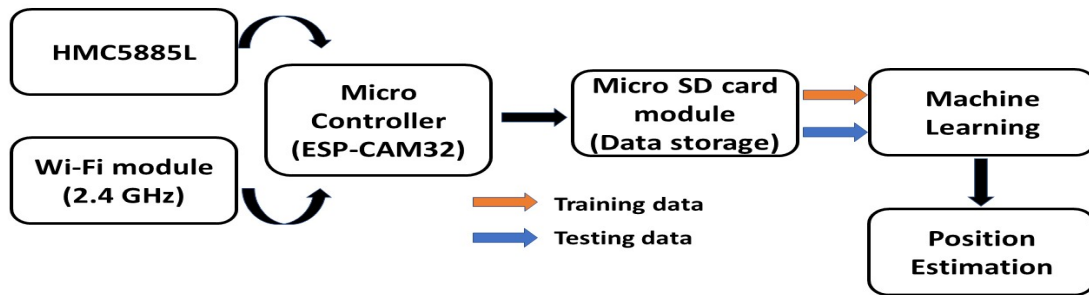


Figure 1: Block level representation of our approach

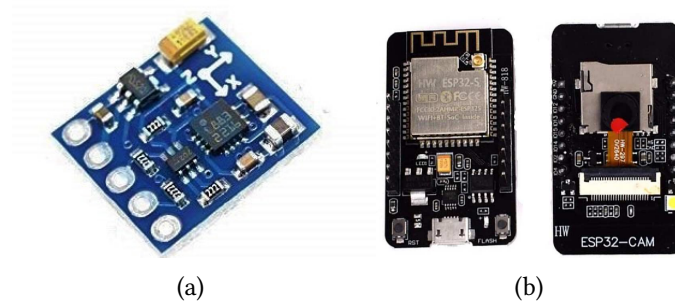


Figure 2: (a) Magnetic sensor (HMC5885L) (b) ESP-CAM32 Micro-controller with inbuilt Wi-Fi and SD card modules

filter, to estimate the pose and calibration parameters.

So far, several research papers have been published in the literature that use either RSSI [27] and [28] or CSI data [1] for Wi-Fi based localization. The bottleneck in such methods is that a large number of access points must be deployed to monitor the user's location, and importantly, the user must be connected to a network when using the widely used triangulation method.

Recent works, such as [29] and [30], to name a few, have demonstrated that using magnetic signatures indoors, and more broadly in the corridors of even large buildings, is quite beneficial in terms of consistency and stability. Furthermore, the sensor signatures do not change over time. In addition, they are unaffected by any changes in the objects, such as furniture in different locations and moving objects such as doors, windows, and user(s), etc. However, some extreme conditions can influence these signatures, for example large-scale reconstructions or the relocation of heavy metallic objects.

3. Motivation

Motivated by the findings mentioned in previous section, we developed a novel machine learning-based approach that makes use of data from both magnetic sensors and Wi-Fi access

points. Interestingly, our approach does not require any network connectivity for localization. Furthermore, in this work, we critically examined the dependence of ML classifiers' location prediction on the altitudes from the surface at which sensors are placed.

We therefore aim to overcome majority of these limitations with a novel idea of combining the utility of both the magnetic field and Wi-Fi Beacon information to describe the indoor position.

Finally, for a further efficient localization indoors, we would like to emphasize that the following can be immensely helpful:

1. A detailed study on the magnetic field data analysis.
2. A study of more recent state-of-the-art

4. Experimental Setup

To begin, we gathered data points spanning an area of 364 sq.ft. In the corridors of a building, this area is subdivided into cells of size 4 sq.ft., resulting in a grid (matrix) of size 7×13 (outdoors). Similarly, while working inside a room (indoors), data points from an area of 180 sq.ft., were divided into cells of 4 sq.ft. each, and resulting in a 5×9 grid. The actual problem is posed as a multi-output and multi-label classification problem, i.e., predicting the co-ordinates (of the user's position) on the grid based on inputs from both the magnetometer sensor (HMC5883L) and the Beacon RSSI values of the selected Wi-Fi access points. We divided the rest of the description of the experiment (its setup and the various elements involved) into four separate subsections for ease of exposition.

4.1. Magnetometer sensor

Honeywell HMC5885L magnetometer sensor is used in this work. This 'low power' module is useful for a variety of general purposes, including measuring the magnetization of a material (for example, ferromagnetic materials, iron bars inside concrete pillars, and so on), measuring the field strength, and measuring the direction of the magnetic field at a point, among others. It is intended for magnetic field sensing with a digital interface for a wide range of real-time embedded applications at a low cost and with high resolution. Next, using the I2C interface, you can easily establish communication between the HMC5885L and microcontrollers. This sensor supports 3V to 5V Input-Output levels on I2C Serial Clock Line (SCL) and Serial Data Line (SDA) pins. The sensor's output is measured in milli-Gauss units.

4.2. Data collection procedure

To collect sensor data and beacon information, we created an Embedded test bed that includes an RF enabled Wi-Fi module and a Magnetometer sensor. Following that, we connected an SD card module to the microcontroller to store the raw data from the sensor into files (.csv format), and these files contain both the magnetic field data (along the x, y, and z planes in reference to the sensor's alignment with the test bed) and the Beacon information of three (nearest, chosen at random) selected Wi-Fi access points. The block level representation of our experimental setup is shown in Fig. 1. We collected magnetic field data and Wi-Fi beacon information in both

outdoor and indoor environments of our research laboratory by placing the sensor at various altitudes.

The data we collected has six attributes: the first three are the data corresponding to the X, Y, and Z axes provided by the Magnetometer sensor, and the remaining three are provided by the Beacon RSSI values of the three selected Wi-Fi access points. Over the 364 sq.ft. area, we collected a total of 54,600 samples (chosen above, outdoors). Remember that this area has 7×13 cells, each with a 4 sq.ft. area, and we collected 600 samples from each of these cells. Similarly, we classified the indoor floor into a 5×9 grid and collected (total of 20,250 samples) 450 samples at each cell of the grid.

4.2.1. Wi-Fi Data

During each scan, the signal strength measurements of all nearby access points can be theoretically obtained. However, not all access points in a location are usually observed/captured in real time. The Service Set Identifier (SSID) of any nearby Wi-Fi access point, as well as its BSSI (MAC address) and Received signal strength Information, are immediately visible in the database obtained by scanning the Wi-Fi network coverage area (RSSI). In our paper, we only considered three access points that were closest to the data collection location.

4.2.2. Magnetic Field Data

The data from the Magnetic field sensor used in this work yields three column vectors each with dimension equal to the number of samples taken:

$$\mathfrak{H} = [\mathfrak{H}_x, \mathfrak{H}_y, \mathfrak{H}_z]_{\text{number of samples} \times 3}.$$

Above, \mathfrak{H}_x , \mathfrak{H}_y and \mathfrak{H}_z denote the position-measurements from the sensor along x , y and z -axes, respectively. Recall, these measurements are noted in mG units.

In general, the raw measurements obtained from the magnetic field sensor can be used in two different ways: 1) to calculate the magnitude of \mathfrak{H} , i.e., the magnitude of each row in \mathfrak{H} , which is a point in 3-dimensional space; 2) to use \mathfrak{H} directly/explicitly. For the purpose of enhancing the accuracy, usually way 2) in the previous line is followed. For each $1 \leq i \leq$ the number of samples, recall:

$$\begin{aligned} \text{Magnitude of } \mathfrak{H}[i] &= \sqrt{(\mathfrak{H}_x[i])^2 + (\mathfrak{H}_y[i])^2 + (\mathfrak{H}_z[i])^2}, \\ \text{where } \mathfrak{H}[i] &= (\mathfrak{H}_x[i], \mathfrak{H}_y[i], \mathfrak{H}_z[i]) \in \mathbb{R}^3. \end{aligned}$$

Figures 3(a), 3(b) and 4 show the plots of the magnitude for the sensor data collected at both outdoor and indoor environments, for few specific columns in the grids in Fig. 5(a), 5(b) and 5(c), respectively. (The latter three figures are explained in the next subsection.)

4.3. Statistical analysis

The 2D-heat maps of the measurements are shown in Fig. 5. Fig. 5(a) corresponds to the experiment outdoors; while the remaining two correspond to indoors and the two different

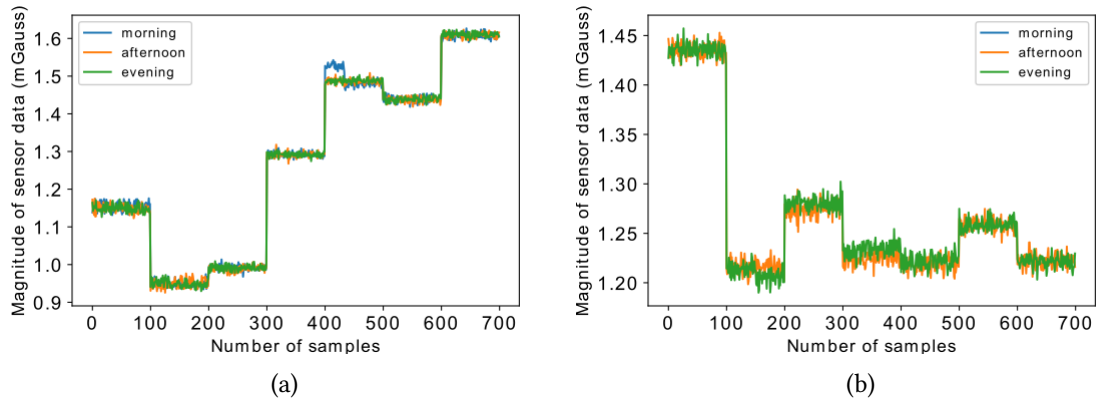


Figure 3: The magnitude plot of magnetic field at few locations in outdoor environment shows that irrespective of time of data collection, the signature remains similar and stable, proving that the field strength is time-invariant. The lag or delay between the patterns can be attributed to the differences in moving the embedded device.

altitudes (2 feet and 4 feet) considered. Furthermore, for a better understanding, we have plotted the 3D-heat map for the experiment outdoors, see Fig. 6.

To verify the consistency of the measurements by the magnetic sensor (embedded on the test bed), the process will be repeated to collect a large number of signatures of the magnetic field in different columns of the grids in Fig. 5. From all of these sub figures, one observes:

- The signatures collected across different columns on the floor can be seen from Fig. 3. Clearly, there is a variation in the patterns in these two figures.
- Consistency: Evidently, the patterns in the plots in Fig. 3 are not varying with the times at which the samples were collected (morning, afternoon and evening).
- At each location indoors, there is a small variation in the signatures as we change the height of the sensors, see Fig. 4.

The collected data at both outdoor and indoor environments is processed to obtain the mean magnitudes – plotted as a heat maps in Fig. 5(a), 5(b) and 5(c) – at every individual cell in the defined grid. The data is then normalised over all the cells, for a better representation of the heat maps. Observe from Fig. 6 (for outdoor environment), the oscillations/non-uniformity in the normalized magnitude values as we move among various cells on the grid.

4.4. Machine learning approach

In our paper, we employ classifiers such as Random Forest(RF) and XG-Boost, since these are purely tree-based ensemble algorithms. Comparatively, in RF classifier, more number of Decision Trees (DT) are applied to multiple sub-samples, so as to increase the prediction performance, to prevent over fitting while training, and also to enhance the stability and accuracy. RF classifier is a bootstrap ensemble for decision trees. Usually, Hyper parameters namely, number of trees and depth etc, are tuned for better prediction performances.

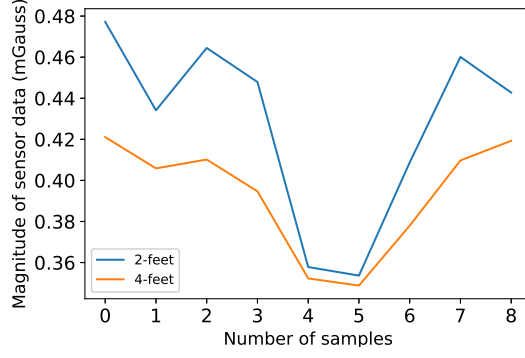


Figure 4: The plot of the magnitudes of the two magnetic signatures at altitudes 2 feet and 4 feet, with some similarity at column-3 in both Fig. 5(b) and 5(c) (at Indoor environment).

The DT predicts the label information of test data based upon calculating entropy of original data set and information gain at each leaf node of the tree. For better understanding, we elaborate the formula of Entropy:

$$\mathfrak{H}(D) = - \sum_{m,n=0}^{M,N} P(m,n) \log P(m,n) \quad (1)$$

Where in, D denotes number of data points in node. m,n denotes number of labels (label m and n correspond to row and column information of room, in our case). $P(m,n)$ denotes probability of data belonging to a label.

Average entropy of a split D : $\{D_i\}_{i=1}^K$ and Information gain (I_g) can be calculated using below equations,

$$A_e = \left[\sum_{i=1}^K \frac{|D_i|}{|D|} \mathfrak{H}(D_i) \right] \quad (2)$$

$$I_g = \mathfrak{H}(D) - A_e = \mathfrak{H}(D) - \left[\sum_{i=1}^K \frac{|D_i|}{|D|} \mathfrak{H}(D_i) \right]. \quad (3)$$

XG-Boost(Xgb) classifier was found effective over off-the-shelf ML classifiers with categorical data. In simple words, Xgb is an algorithm which is an implementation of Gradient Boosted DT's. It is being very dominant in applied machine learning now a days. One of the objectives of Xgb is to minimize the error of cost function $\ell(y, F(x))$, via applying gradient descent algorithm. In this algorithm, DT's are created in a sequential order. Weights, which play an important role in XGBoost, are assigned to all the independent variables, which are then fed into the DT which predicts the outputs. The weight of the variables that is predicted wrong by the DT, is increased and then the variables are fed to the next DT. These individual DT's are then ensemble to give a concrete and more efficient (in terms of accuracy) model. It can work on both the regression and classification problems.

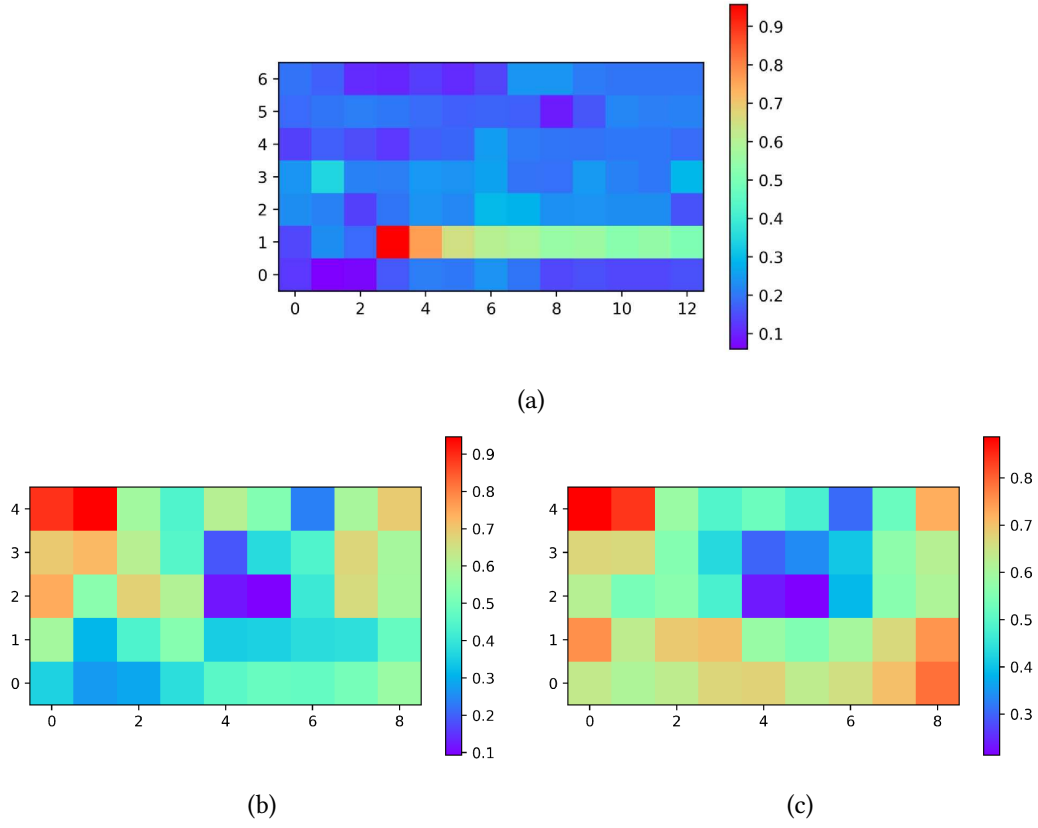


Figure 5: (a) Normalised heat map of the floor at outdoor environment; (b) Magnetic field at 2 feet altitude at Indoor environment; (c) Magnetic field at 4 feet altitude at indoor environment.

The training data set is $D = \{(x_i, y_i)\}_{i=1}^N$, where x_i and y_i corresponding to the input and output attributes (a total of 8 attributes) of data set, $y_i = \{[y_m, y_n]\}_{m,n=1}^{M,N}$. Note that the lengths of y_m and y_n are same. Let the differential loss function for our model be $\ell(y, F(x))$, where $F(x)$ is the prediction corresponding to the input of classifier. We begin by initializing the model with some constant value:

$$F_0(x) = \arg \min_{\gamma} \sum_{i=1}^N \ell(y_i, \gamma) \quad (4)$$

Let K be a positive integer. For each $k \in \{1, \dots, K\}$ and for each $i \in \{1, \dots, N\}$, we compute the entries of the Gradient and Hessian iteratively (using a *for loop*) as follows:

$$r_{ik} = - \left[\frac{\partial \ell(y_i, F(x_i))}{\partial F(x_i)} \right]_{F(x)=F_{k-1}(x)} \quad (5)$$

$$s_{ik} = - \left[\frac{\partial^2 \ell(y_i, F(x_i))}{\partial F(x_i)^2} \right]_{F(x)=F_{(k-1)}(x)} \quad (6)$$

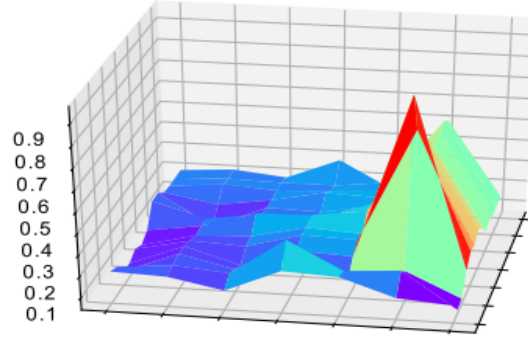


Figure 6: Map showing the variation in magnetic field intensities of cells in a grid(3D- view)

We fit a base learner (Decision tree) for the training set $\left\{x_i, -\frac{r_{ik}}{s_{ik}}\right\}_{i=1}^N$, solving the following optimization problem:

$$\gamma_k = \arg \min_{\gamma} \sum_{i=1}^N \frac{1}{2} s_{ik} \left[-\frac{r_{ik}}{s_{ik}} - \gamma(x_i) \right]^2 \quad (7)$$

$$F_k(x) = v * \gamma_k(x) \quad (8)$$

(v is the learning rate.)

After finding γ_k and thereby $F_k(x)$ as above, we now update (in the *for loop*) the parameter $F_k(x)$:

$$F_k(x) = F_{k-1}(x) + F_k(x) \quad (9)$$

By doing all the above, we obtain the final output:

$$F(x) = F_{(k)}(x) = \sum_{k=0}^K F_k(x) \quad (10)$$

4.5. Experimental Procedure

To begin with, we have collected data inside as well as outside our research laboratory building. Next, we captured the real time data from both the Wi-Fi module and Magnetic sensor, and then stored all of this data in an SD card for further processing. The data extracted from the Magnetic sensor are composed of an individual reading for each Cartesian axis (x, y and z). In real time, the data from the magnetic sensors are anomalous due to sudden changes in the magnetic field, or erroneous sensor readings. The data stored in SD cards has been transferred to a laptop for further data processing, that includes training and testing of ML models.

We have then classified the target attributes into two outputs. The first output corresponds to the row information of the cell in the grid, while the second to the column information. The

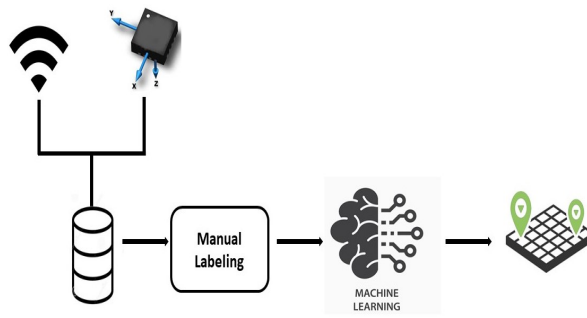


Figure 7: Architecture illustrating the procedures and their flow. First data collection process, followed by database creation, and then feeding to ML classifiers for position estimation.

floor map of outdoor environment can be found in Fig. 5(a). The values 0–12 on the x-axis (as in this figure) represent individual columns of the grid, while the values 0–6 on the y-axis represent individual rows. Similarly, for indoor environment, we considered the total area into a 5x9 grid, and the same can be seen clearly from both Fig. 5(b) and 5(c). Next, the values 0–8 on the x-axis (in both the figures in the previous line) represent individual columns of the grid, while 0–4 on the y-axis represent individual rows.

Now, we present our experimental results, comparing the performances of various ML models used in this paper for the position estimation.

5. Experimental Results

We observed very good testing accuracy running a series of experiments on the collected data set, with varying test sizes.

As can be seen in Fig. 8, the results obtained after training and testing the classifiers with both magnetic field data and Wi-Fi data are depicted (collected from the embedded test bed developed presently). Observe in the same figure that the RF classifier has a very good accuracy of around $> 95\%$. Because we are working with a fixed number of samples, the test size increases as the size of the training data set decreases. Our accuracy results immediately show that we are able to predict with high appreciable accuracy, and more importantly (perhaps remarkably) with only few samples.

The following example is for a quick grasp:

- At 95% test sample size, the number of samples used for training is 2730 (which amounts to the leftover 5%).
- All of these 2730 samples are uniformly distributed across 91 classes, i.e., 30 samples per each class.
- Our prediction accuracy (with only 5% of training data set) is $> 95\%$.

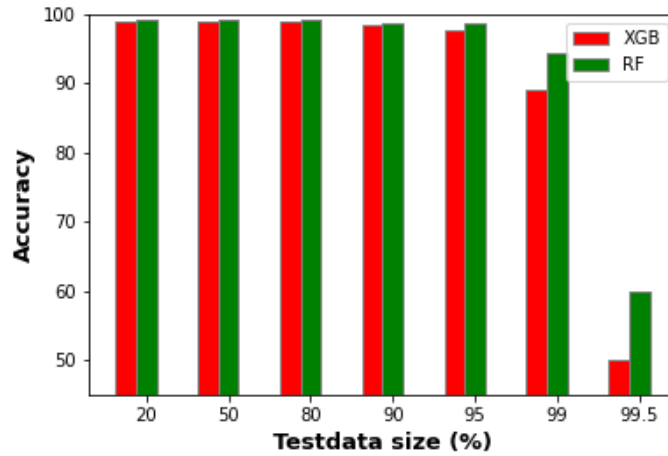


Figure 8: Results obtained after training different ML classifiers with different test sizes at outdoors

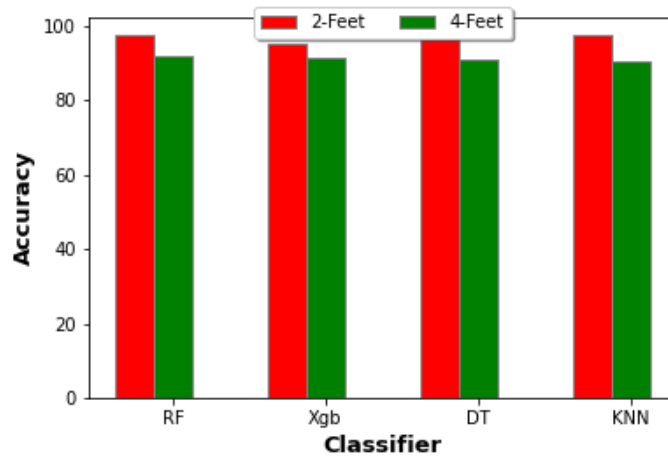


Figure 9: Results obtained with different ML classifiers at indoors

The prediction accuracy was calculated using the formula,

$$Accuracy = \frac{\sum_{m=1}^M (\text{Prediction}(m) == \text{Original label}(m))}{\text{Total no. of samples used for testing}}$$

where, m is a element in a array.

Motivated by the outdoor results, we considered a small number of samples at each cell in an indoor environment at two different altitudes. The results shown in Fig. 9 are the result of applying ML classifiers to two separate data sets collected at two different altitudes within the room. It is clear that all of the classifiers we used performs very well.

6. Conclusions and Future scope

In this paper, we presented a novel approach for estimating the user's position indoors with high accuracy (around 98%). The motivation for using statistical approaches is that deep learning-based models are computationally expensive and require a large amount of memory for real-time deployment. We created a machine learning-based multi-label and multi-output strategy for indoor localization and demonstrated it using our own embedded test bed. However, the motivation for this work is to predict the position of the user indoors as well. In our experiment, we used industry standard sensors on our test bed to evaluate the performance of ML classifiers with very good resolution (in terms of accuracy) for indoor localization. Changes in the size of the training data set were also used to evaluate performance.

We achieved best accuracy of around 98.6% and 97.6% with RF and XGB classifiers, respectively, by training the models with 5% of the original data set. Furthermore, we evaluated the effectiveness of our approach for interior location detection at various altitudes. Our method is simple to implement, and the time required for data collection and prediction is minimal. We mapped a small area within a large building, but given the promising accuracy, this approach can be scaled up to larger infrastructures such as large office buildings and shopping malls, among other things.

Note that this paper does not focus on: materials, configurations of material, and other properties of buildings cause our observations. Indeed, if we can understand the elements playing a role in our observations – such as before – then we can make some predictions on the possible magnetic field variations that might occur in future, in the same buildings worked inside/others. Such predictions would allow for improvements in evaluation of scientific grounding in localization techniques. To begin with, for studying possible links between variables, one can investigate using a building information model to predict the magnetic fields inside it. This approach could help in associating magnetic field variations with specific building elements for instance: steel beams, heating and cooling systems such as Air coolers. In parallel, one can extend the study of the performance of any machine learning model, via conducting the same experiments in different environments in large areas like: a multi-story building constructed by brick and concrete, a warehouse/factory building, a high-rise building made of steel. We would like to explore the above avenues in future, in the light of a reasonable thinking that, there is no guarantee that the results shown in this paper would hold the same for every building, with may be similar structure etc.

References

- [1] X. Wang, L. Gao, S. Mao, S. Pandey, Csi-based fingerprinting for indoor localization: A deep learning approach, *IEEE Transactions on Vehicular Technology* 66 (2017) 763–776.
- [2] Y. Ma, C. Tian, Y. Jiang, A multitag cooperative localization algorithm based on weighted multidimensional scaling for passive uhf rfid, *IEEE Internet of Things Journal* 6 (2019) 6548–6555.
- [3] P. Sthapit, H.-S. Gang, J.-Y. Pyun, Bluetooth based indoor positioning using machine

- learning algorithms, in: 2018 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), 2018, pp. 206–212.
- [4] T. Akiyama, M. Sugimoto, H. Hashizume, Time-of-arrival-based smartphone localization using visible light communication, in: 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2017, pp. 1–7.
 - [5] A. Hilal, I. Arai, S. El-Tawab, Dataloc+: A data augmentation technique for machine learning in room-level indoor localization, in: 2021 IEEE Wireless Communications and Networking Conference (WCNC), 2021, pp. 1–7.
 - [6] B. Siebler, S. Sand, U. D. Hanebeck, Localization with magnetic field distortions and simultaneous magnetometer calibration, *IEEE Sensors Journal* 21 (2021) 3388–3397.
 - [7] H. Zou, C.-L. Chen, M. Li, J. Yang, Y. Zhou, L. Xie, C. J. Spanos, Adversarial learning-enabled automatic wifi indoor radio map construction and adaptation with mobile robot, *IEEE Internet of Things Journal* 7 (2020) 6946–6954.
 - [8] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, J. Huerta, Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems, in: 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2014, pp. 261–270.
 - [9] T. Koike-Akino, P. Wang, M. Pajovic, H. Sun, P. V. Orlik, Fingerprinting-based indoor localization with commercial mmwave wifi: A deep learning approach, *IEEE Access* 8 (2020) 84879–84892.
 - [10] J. Zhang, G. Han, N. Sun, L. Shu, Path-loss-based fingerprint localization approach for location-based services in indoor environments, *IEEE Access* 5 (2017) 13756–13769.
 - [11] H. K. Rath, S. Timmadasari, B. Panigrahi, A. Simha, Realistic indoor path loss modeling for regular wifi operations in india, in: 2017 Twenty-third National Conference on Communications (NCC), 2017, pp. 1–6.
 - [12] H.-S. Kim, W. Seo, K.-R. Baek, Indoor positioning system using magnetic field map navigation and an encoder system, *Sensors* 17 (2017).
 - [13] D. Hanley, A. B. Faustino, S. D. Zelman, D. A. Degenhardt, T. Bretl, Magpie: A dataset for indoor positioning with magnetic anomalies, in: 2017 International Conference on Indoor Positioning and Indoor Navigation, 2017, pp. 1–8.
 - [14] Y. Zeng, P. H. Pathak, P. Mohapatra, Analyzing shopper’s behavior through wifi signals, in: Proceedings of the 2nd Workshop on Workshop on Physical Analytics, Association for Computing Machinery, 2015, p. 13–18. doi:10.1145/2753497.2753508.
 - [15] F. Alam, N. Faulkner, B. Parr, Device-free localization: A review of non-rf techniques for unobtrusive indoor positioning, *IEEE Internet of Things Journal* 8 (2021) 4228–4249.
 - [16] A. Gadhgadhi, Y. Hachaichi, H. Zairi, A machine learning based indoor localization, in: 2020 4th International Conference on Advanced Systems and Emergent Technologies (ICASET), 2020, pp. 33–38.
 - [17] N. Lee, S. Ahn, D. Han, Amid: Accurate magnetic indoor localization using deep learning, *Sensors* 18 (2018).
 - [18] A. Poullose, D. Han, Hybrid deep learning model based indoor positioning using wi-fi rssi heat maps for autonomous applications, *Electronics* 10 (2020).
 - [19] R. Ayyalasomayajula, A. Arun, C. Wu, S. Sharma, A. R. Sethi, D. Vasisht, D. Bharadia, Deep Learning Based Wireless Localization for Indoor Navigation, 2020.

- [20] H. J. Bae, L. Choi, Large-scale indoor positioning using geomagnetic field with deep neural networks, in: ICC 2019 - 2019 IEEE International Conference on Communications (ICC), 2019, pp. 1–6.
- [21] G. M. Mendoza-Silva, J. Torres-Sospedra, F. Potortì, A. Moreira, S. Knauth, R. Berkvens, J. Huerta, Beyond euclidean distance for error measurement in pedestrian indoor location, *IEEE Transactions on Instrumentation and Measurement* 70 (2021) 1–11.
- [22] C. Feng, W. S. A. Au, S. Valaee, Z. Tan, Received-signal-strength-based indoor positioning using compressive sensing, *IEEE Transactions on Mobile Computing* 11 (2012) 1983–1993.
- [23] K. Cengiz, Comprehensive analysis on least-squares lateration for indoor positioning systems, *IEEE Internet of Things Journal* 8 (2021) 2842–2856.
- [24] J. Palacios, G. Bielsa, P. Casari, J. Widmer, Single- and multiple-access point indoor localization for millimeter-wave networks, *IEEE Transactions on Wireless Communications* 18 (2019) 1927–1942.
- [25] R. Shirai, M. Hashimoto, Dc magnetic field based 3d localization with single anchor coil, *IEEE Sensors Journal* 20 (2020) 3902–3913.
- [26] D. Hanley, A. S. D. d. Oliveira, X. Zhang, D. H. Kim, Y. Wei, T. Bretl, The impact of height on indoor positioning with magnetic fields, *IEEE Transactions on Instrumentation and Measurement* 70 (2021) 1–19.
- [27] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, K. Reddy, Recurrent neural networks for accurate rssi indoor localization, *IEEE Internet of Things Journal* 6 (2019) 10639–10651.
- [28] S. Knauth, M. Storz, H. Dastageeri, A. Koukofikis, N. A. Mähser-Hipp, Fingerprint calibrated centroid and scalar product correlation rssi positioning in large environments, in: 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2015, pp. 1–6.
- [29] M. Kwak, C. Hamm, S. Park, T. T. Kwon, Magnetic field based indoor localization system: A crowdsourcing approach, in: 2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2019, pp. 1–8.
- [30] I. Ashraf, S. Hur, Y. Park, Enhancing performance of magnetic field based indoor localization using magnetic patterns from multiple smartphones, *Sensors* 20 (2020).