RF BASED INDOOR POSITIONING SYSTEM

Doctoral Thesis

Sinem BOZKURT KESER
Eskişehir, 2017

RF BASED INDOOR POSITIONING SYSTEM

Sinem BOZKURT KESER

DOCTORAL THESIS

Graduate School of Sciences
Computer Engineering Program
Supervisor: Assoc. Prof. Dr. Serkan GÜNAL
Second Supervisor: Assoc. Prof. Ahmet YAZICI

Eskişehir
Anadolu University
Graduate School of Science
December, 2017

This thesis work is supported by the Scientific Research Projects Commission of The Scientific and Technological Research Council of Turkey (TÜBITAK) [grant number: 1130024].

FINAL APPROVAL FOR THESIS

This thesis titled "RF Based Indoor Positioning System" has been prepared and submitted by Sinem Bozkurt Keser in partial fulfillment of the requirements in "Anadolu University Directive on Graduate Education and Examination" for the Degree of Doctor of Philosophy (PhD) in Computer Engineering Department has been examined and approved on 22/12/2017.

		Committee Members	Signature
Member (Supervisor)	:	Assoc. Prof. Dr. Serkan GÜNAL	…………..
Member	:	Prof. Dr. Rifat EDİZKAN	…………..
Member	:	Assoc. Prof. Dr. Cihan KALELİ	…………..
Member	:	Assoc. Prof. Dr. Kemal ÖZKAN	…………..
Member	:	Assist. Prof. Dr. Alper BİLGE	…………..

……………………………

Director

Graduate School of Science

ABSTRACT

RF BASED INDOOR POSITIONING SYSTEM

Sinem BOZKURT KESER

Department of Computer Engineering
Anadolu University, Graduate School of Sciences, December 2017
Supervisor: Assoc. Prof. Dr. Serkan GÜNAL
Second Supervisor: Assoc. Prof. Dr. Ahmet YAZICI

Although the Global Positioning System is a publicly recognized technology for positioning in the outdoor environment, it is ineffective in the indoor environment. For this reason, the search for effective solutions to indoor positioning still continues. Within the scope of this dissertation, it is aimed to develop radio frequency (RF) based, high-accuracy and low-cost indoor positioning approaches based on the fingerprint method. For this purpose, in addition to the existing indoor positioning datasets in the literature, a new dataset has been constituted and made available to researchers. In terms of selected performance criteria, the most suitable algorithm for three different indoor environments is determined by multi-criteria optimization technique. Hybrid fingerprints are defined using a combination of WiFi received signal strength and magnetic field measurements. It has been observed that the positioning accuracy is improved when the proposed hybrid fingerprint dataset is used with different classification algorithms. F-score weighted indoor positioning algorithm combining WiFi received signal strength and magnetic field measurements is proposed. It has been observed that the accuracy of the proposed algorithm is higher than that of the conventional algorithms. In addition, an improved indoor positioning approach has been proposed that uses WiFi signal strength and magnetic field fingerprints for more precise locating. With this approach, high accuracy position estimation can be done.

Keywords: Indoor Positioning Systems, Fingerprint Based Positioning, WiFi Received Signal Strength, Magnetic Field.

ÖZET

RF TABANLI İÇ ORTAM KONUMLANDIRMA SİSTEMİ

Sinem BOZKURT KESER

Bilgisayar Mühendisliği Anabilim Dalı
Anadolu Üniversitesi, Fen Bilimleri Enstitüsü, Aralık 2017
Danışman: Doç. Dr. Serkan GÜNAL
İkinci Danışman: Doç. Dr. Ahmet YAZICI

Küresel Konumlama Sistemi, dış ortamlarda konumlandırma için herkes tarafından kabul gören bir teknoloji olmasına karşın iç ortamlarda etkisiz kalmaktadır. Bu nedenle, araştırmacıların iç ortamlarda konum belirlemek için etkili çözüm arayışları devam etmektedir. Bu tez çalışması kapsamında, parmak izi yöntemini temel alan radyo frekansı (RF) tabanlı, yüksek doğruluğa sahip ve düşük maliyetli iç ortam konumlandırma yaklaşımları geliştirilmesi hedeflenmiştir. Bu doğrultuda, literatürde var olan iç ortam konumlandırma veri kümelerine ilave olarak yeni bir veri kümesi oluşturulmuş ve araştırmacıların kullanımına sunulmuştur. Seçilen performans kriterleri açısından üç farklı iç ortam için en uygun algoritma, çok-kriterli optimizasyon tekniği ile belirlenmiştir. WiFi alınan sinyal gücü ve manyetik alan ölçümleri bir arada kullanılarak hibrid parmak izleri tanımlanmıştır. Önerilen hibrid parmakizi veri kümesi, farklı sınıflandırma algoritmalarıyla birlikte kullanıldığında konumlandırma doğruluğunun iyileştiği görülmüştür. WiFi alınan sinyal gücü ve manyetik alan ölçümlerini bir araya getiren F-skor ağırlıklı iç ortam konumlandırma algoritması önerilmiştir. Önerilen algoritmanın sağladığı doğruluğun geleneksel algoritmalardan daha yüksek olduğu gözlenmiştir. Ayrıca, daha hassas konum belirleme amacıyla, WiFi alınan sinyal gücü ve manyetik alan parmak izlerini kullanan, geliştirilmiş bir iç ortam konumlandırma yaklaşımı önerilmiştir. Bu yaklaşım ile yüksek hassasiyette konum tahmini yapılabilmiştir.

Anahtar Sözcükler: İç Ortam Konumlandırma Sistemleri, Parmakizi Tabanlı Konumlandırma, WiFi Alınan Sinyal Gücü, Manyetik Alan.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Ahmet YAZICI and Serkan GÜNAL for their valuable help and support during my PhD studies. I would have never been able to finish my dissertation without their guidance. They gave innovative ideas and important lectures during this dissertation.

Furthermore, I would like to thank my family and husband Can KESER for their support and patience during my dissertation.

Sinem BOZKURT KESER
Eskişehir, 2017

22/12/2017

STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES

I hereby truthfully declare that this thesis is an original work prepared by me; that I have behaved in accordance with the scientific ethical principles and rules throughout the stages of preparation, data collection, analysis and presentation of my work; that I have cited the sources of all the data and information that could be obtained within the scope of this study, and included these sources in the references section; and that this study has been scanned for plagiarism with "scientific plagiarism detection program" used by Anadolu University, and that "it does not have any plagiarism" whatsoever. I also declare that, if a case contrary to my declaration is detected in my work at any time, I hereby express my consent to all the ethical and legal consequences that are involved.

....

Sinem BOZKURT KESER

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AOA	: Angle of Arrival
AP	: Access Point
BLE	: Bluetooth Low Energy
dBm	: Decibel Milliwatt
GPS	: Global Positioning System
GSM	: Global System for Mobile Communications
IPS	: Indoor Positioning Systems
LBS	: Location Based Services
LOS	: Line of Sight
MF	: Magnetic Field
MU	: Mobile Unit
NLOS	: Non Line-of-Sight
PDR	: Pedestrian Dead Reckoning
RF	: Radio Frequency
RFID	: Radio-frequency Identification
RP	: Reference Point
RSS	: Received Signal Strength
TDOA	: Time Difference of Arrival
TOA	: Time of Arrival
WiFi	: Wireless Fidelity
WLAN	: Wireless Local Area Network

1. INTRODUCTION

Indoor positioning systems (IPS) include technologies and methods to calculate mobile unit (MU) position within a structure in a closed area. IPS are used to estimate user position in closed areas such as airports, shopping centres, train stations, hospitals, to inform promotions and discounts at stores to the users, and automatically directing a visually impaired user in a closed area. The main problem in IPS is to perform positioning in a cheapest and accurate way. There is no standardized system for solving indoor positioning problem like as Global Positioning System (GPS) for outdoor environment [1]; therefore studies on this area are still in the process of evolution.

Various technologies are utilized to support IPS such as Global System for Mobile Communications (GSM) [2], Radio-frequency Identification (RFID) [3], ultrasonic [4], Bluetooth Low Energy (BLE) [5], Wireless Local Area Network (WLAN) [6], magnetic field (MF) [7], and so on. GSM-based system uses existing infrastructure, but it does not give reasonable accuracy for indoor areas. RFID-based and ultrasonic-based IPS are also having reasonable accuracies, but they need the installation of extra sensors. BLE-based IPS have short operating range as well as poor predictability. Therefore, it is recommended that BLE is used as a supplementary technology in an IPS. WLAN-based IPS is the most widely deployed system when compared with other systems. It is inexpensive, infrastructure free, and it has ubiquitous availability and easier deployment inside buildings. So, several algorithms are proposed in this context.

Indoor positioning algorithms are divided into five main categories such as triangulation, proximity, pedestrian dead reckoning (PDR), vision analysis and fingerprinting as it can be seen in Figure 1.1.

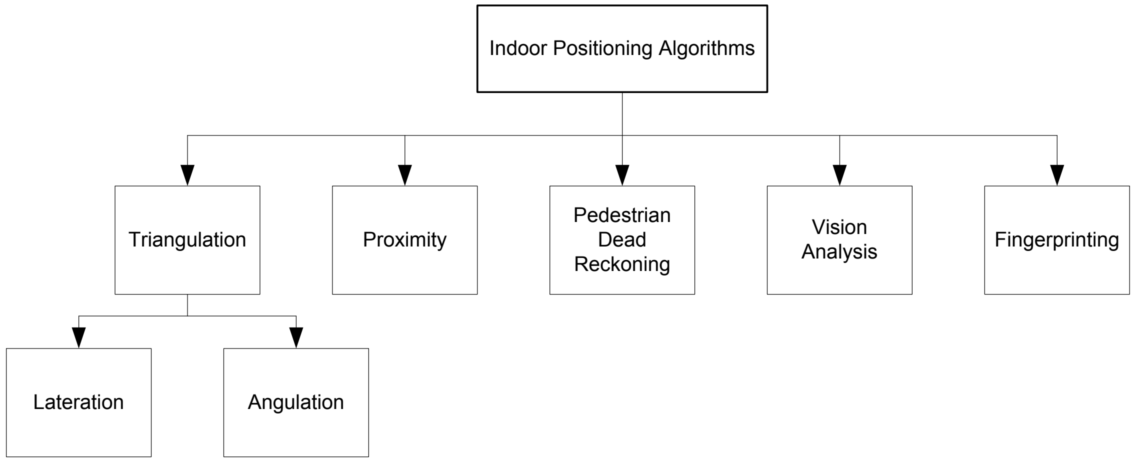


Figure 1.1. *Classification of Indoor Positioning Algorithms*

Triangulation is a geometric-based method that uses signal parameters to calculate the MU position [8]. It is divided into two subcategories such as lateration and angulation based on the signal parameters. Distance measurements are used to calculate the MU position in lateration method. Time of Arrival (TOA) and Time Difference of Arrival (TDOA) are two most popular lateration methods. In TOA, the distance between the transmitter of the Access Point (AP) and the receiver of the MU is calculated using the velocity and the travel time of the transmitted signal [9]. At least three APs are needed for positioning in this method. The calculated distances between each AP and the MU are utilized as the radius of the three propagation circles of the signals transmitted from the APs. At the end, the intersection point of these three circles is the estimated position of the MU. TOA requires time synchronization between the APs and the MU. This synchronization is eliminated using TDOA method. In TDOA, arrival time differences of the signals obtained from different APs at the receiver are utilized as distance measurements in place of travel time [10]. Angle of Arrival (AOA) of the signal is used to calculate the MU location in the angulation method [11]. In AOA, intersection of virtual lines from different transmitters is used to calculate the MU position. It does not require any time synchronization between the transmitters and receivers. But, it requires more complex hardware (an array of antennas) to determine the angle between the transmitters and receivers. All the triangulation methods suffer from Non Line-of-Sight (NLOS) conditions; therefore give erroneous results for indoor positioning. In proximity, the MU location is estimated as the antenna position which receives the strongest signal from the MU [12]. Therefore, a dense grid of antennas with known positions is used. This method is generally used in RFID technology. It requires additional hardware, and has low resolution and poor accuracy. So, it is impractical for indoor positioning. In PDR, the position of the MU is calculated using the previously calculated position, speed, and the direction of the MU [13]. Since current position is relative to the previous position, the errors are cumulative. In addition to this, the sensors in the most smartphones do not provide very accurate data. Therefore, PDR is not solely adequate for an IPS. Vision analysis has the high complexity, since it requires creating the large image database and the real-time communication between the server and the MU [14]. Therefore, it gives undesirable solution for indoor positioning. Wireless Fidelity (WiFi) Received Signal Strength (RSS) based fingerprinting method is widely adopted approach due to the relatively high accuracy and modest cost. It utilizes the existing WLAN infrastructure. RADAR [15] and HORUS [16] are typical IPS based upon WiFi-RSS based fingerprinting method. MF-based positioning is another approach to solve indoor positioning problem [7]. In this method, LOS is not required in order to estimate the MU position. It is easy to obtain MF measurements using today's smartphones. When MF signals are static, fingerprinting method can be applied to construct MF-based fingerprinting method for indoor positioning [17].

1.1. Problem Statement and Technical Challenges

The main problem in IPS is to obtain a reasonable positioning accuracy in a cost-effective manner. Therefore, WiFi-RSS based fingerprinting method is the most employed method by the researchers. But, it has some challenges such as WiFi-RSS values suffer from multipath effect which leads to erroneous position estimate. Therefore, it can be enhanced using supplementary technologies such as BLE or MF as mentioned above. But, using BLE with WiFi-RSS based fingerprinting method is not a good choice. Because both of the technologies operate in the same frequency (2.4GHz), therefore signal inference is inevitable. On the other hand, MF has some advantages such as MF does not suffer from NLOS conditions or multipath effects in indoors whereas it has short operating range, and sensitivity to certain materials. However, MF strength diminishes rapidly with distance. So, MF-based fingerprinting method is best utilized as a supplementary method to WiFi-RSS based fingerprinting method for indoor positioning.

1.2. Research Aim and Objectives

The aim of this study is to design a cost-effective radio frequency (RF)-based IPS which adopts fingerprinting method. Our objective is to enhance the IPS positioning performance. For this purpose, various positioning algorithms are utilized with the publicly accessible indoor positioning datasets to determine the most appropriate algorithms in terms of selected performance metrics. A multi-criteria optimization technique is defined to obtain the most appropriate algorithm for a given dataset. In another approach, the test environment is divided into clusters to construct cluster-based classification algorithms. By using this approach, the positioning accuracy is improved. Since using solely WiFi-RSS based fingerprinting method is not adequate to obtain reasonable accuracy for indoor positioning, we shed light on the advantages of WiFi-RSS and MF measurements at the same time and counteract their drawbacks. For this purpose, we handle MF-based fingerprinting method as a supplementary solution with the WiFi-RSS based fingerprinting method. Therefore, hybrid fingerprints are defined to improve the positioning performance. Then, several positioning algorithms are applied with hybrid fingerprint database to solve the indoor positioning problem. In another study, we propose an f-score weighted indoor positioning algorithm integrating WiFi-RSS fingerprints with MF fingerprints to enhance IPS performance in terms of accuracy. The proposed f-score weighted indoor positioning algorithm has better accuracy performance than the conventional algorithms. Thus far, these algorithms solve indoor positioning problem as a classification task. Since more precise position estimates are more preferred, and then we propose an enhanced indoor positioning algorithm using WiFi RSS and magnetic field fingerprints. This final method calculates the position in terms of x and y coordinates to obtain more precise location.

1.3. Outline

The rest of the thesis is organized as follows: Literature review for fingerprint-based positioning is given in Section 2. The methods and positioning approaches proposed by this thesis and their evaluation are explained in Section 3. Finally, Section 4 outlines the conclusion and future work.

2. BACKGROUND AND RELATED WORK

Fingerprint-based positioning is one of the most exploited methods in indoor positioning because of its inexpensive cost, relatively high accuracy, simplistic design, and easier deployment [18]. It contains mainly two phases: fingerprint mapping (offline) phase and positioning (online) phase. In the fingerprint mapping phase, signal measurements obtained from each reference point (RP) are stored into the database named as fingerprint map with the known RP coordinates. In the positioning phase, a positioning algorithm is applied to estimate the position of the MU by comparing online fingerprint measurement with the fingerprints in the fingerprint map. There is also one internal step named as preprocessing which used to optimize the IPS by removing redundant information, by selecting most informative information, or by dividing whole experimental area into sub-areas using clustering. The fingerprint-based positioning is illustrated in Figure 2.1.

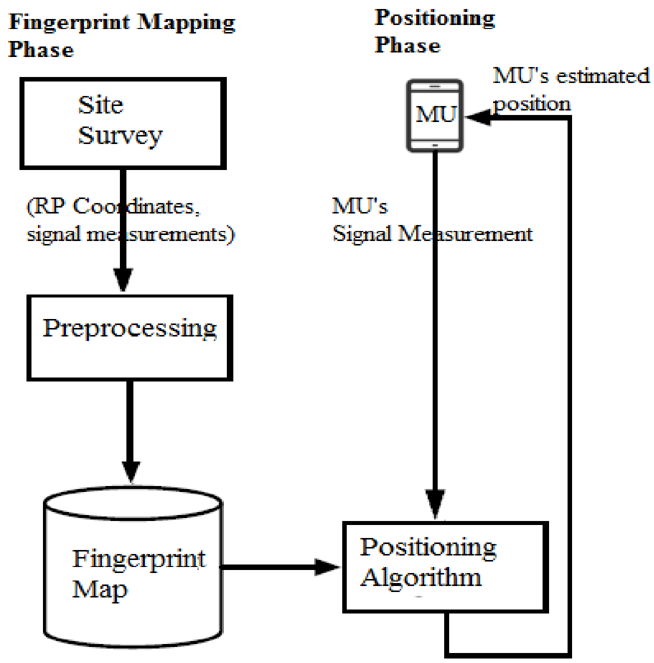


Figure 2.1. *Fingerprint-based positioning*

In the following subsections, prior works and the necessary background and methods employed in the fingerprint-based positioning algorithms are presented.

2.1. Fingerprint-based Mapping

Fingerprint-based mapping is started with dividing test area into equi-sized grids; then sensor measurements are collected from the center of each grid, and stored as a fingerprint in the database. There are various studies which adopt fingerprint-based indoor positioning in the literature. These studies generally store the WiFi-RSS values for their database [19, 20]. Recent studies recognize the efficiency of the MF and store the samples from the magnetometer to construct their database [21, 22]. Among these databases, the publicly available databases are limited [19, 20, 22]. And these databases contain one type of sensor measurement. A multi-sensor fingerprint database that includes WiFi-RSS and MF for indoor positioning is proposed in [23]. The processes of obtaining the fingerprint-based maps are given below.

2.1.1. Radio Map

Radio map is constructed by dividing the experimental area into equi-sized grids [15]. The centre of each grid represents the reference points (RPs) where WiFi-RSS (measured in decibel-milliwatt (dBm)) values of the radio signals transmitted by APs are collected. These WiFi-RSS values are stored into the radio map as a fingerprint with the known coordinates of RPs. The fingerprint at the i^{th} RP in the radio map is stored as

$$FP_i = \left\{ lb_i, xCoord_i, yCoord_i, (MAC_{i,1}, RSS_{i,1}), \ldots, (MAC_{i,k_i}, RSS_{i,k_i}) \right\} \tag{2.1}$$

where FP_i is the fingerprint information at RP_i, lb_i is the label of the RP_i, $xCoord_i$ and $yCoord_i$ are the x and y coordinates of RP_i, $MAC_{i,j}$ and $RSS_{i,j}$ are the MAC address and WiFi-RSS values of the j^{th} AP received at RP_i, and k_i is the number of available APs at RP_i. An example of the radio map is given in Table 2.1.

Table 2.1. *An example of a radio map*

RP Label	x Coordinate	y Coordinate	MAC1	…	MACn
…		…….			
1	1.2	1.2	-82	…	-83
1	1.2	1.2	-82	…	-83
1	1.2	1.2	-86	…	-83
1	1.2	1.2	-86	…	-82
2	3.6	1.2	-87	…	-81
2	3.6	1.2	-88	…	-77
2	3.6	1.2	NaN	…	-76
2	3.6	1.2	NaN	…	-76
…		…….			

2.1.2. Magnetic Map

Magnetic map is established using the same procedure as the radio map construction. Each fingerprint in the magnetic map contains the x, y, and z values of MF strength values (measured in microTesla or μT) which are obtained by a magnetometer sensor on a mobile device [22]. The magnetic fingerprint at the i^{th} RP in the magnetic map is represented as

$$MFP_i = \{lb_i, xCoord_i, yCoord_i, global_{i,x}, global_{i,y}, global_{i,z}\} \tag{2.2}$$

where MFP_i is the magnetic fingerprint information at RP_i, lb_i is the label of the RP_i, $xCoord_i$ and $yCoord_i$ are the x and y coordinates of RP_i, $global_{i,x}$, $global_{i,y}$ and $global_{i,z}$ are the $global$ x, y, and z values of magnetic field strength in relation to the world coordinates at RP_i. An example of the magnetic map is given in Table 2.2.

Table 2.2. *An example of a magnetic map*

RP Label	x Coordinate	y Coordinate	Global x	Global y	Global z
…		…….			
1	1.2	1.2	8.76	-5.03	-2.52
1	1.2	1.2	8.76	-5.15	-2.88
1	1.2	1.2	8.86	-5.05	-2.88
1	1.2	1.2	8.86	-4.73	-2.90
2	3.6	1.2	-71.02	-2.68	-0.22
2	3.6	1.2	-70.98	-2.75	-0.26
2	3.6	1.2	-71.01	-2.78	-0.09
2	3.6	1.2	-70.96	-2.78	-0.07
…		…….			

2.2. Preprocessing

The aim of the preprocessing is to optimize the IPS performance in terms of accuracy and computation time. This can be done by removing redundant information, by selecting most informative information, or by dividing whole experimental area into sub-areas using clustering. Therefore, the applied methods for preprocessing step in this thesis can be classified as follows: fingerprint filtering, AP selection, and fingerprint map clustering. These methods are explained in the following subsections.

2.2.1. Fingerprint Filtering

Fingerprint filtering is applied for two signal types such as WiFi RSS and MF measurements separately. In this subsection, filtering methods are explained in terms of each signal type.

Radio map contains the RSS values from the APs for each RP. When collecting RSS data to form the radio map, some APs are not detected during each scan. Therefore, NaN values are occurred in the radio map. Since the WiFi-RSS level values are ranged from -100dBm to 0dBm, generally NaN values are replaced with -100dBm in the literature. Then, minimum (except -100dBm) and maximum RSS values for each AP are replaced with the minimum and maximum values of all instances in the radio map. This filtering approach is illustrated in Figure 2.2 and Figure 2.3 as before filtering and after filtering.

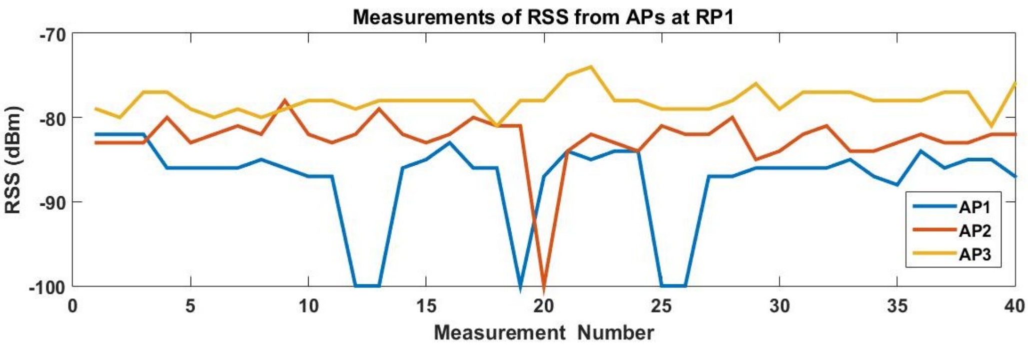


Figure 2.2. *Measurements of RSS values from three APs in first RP*

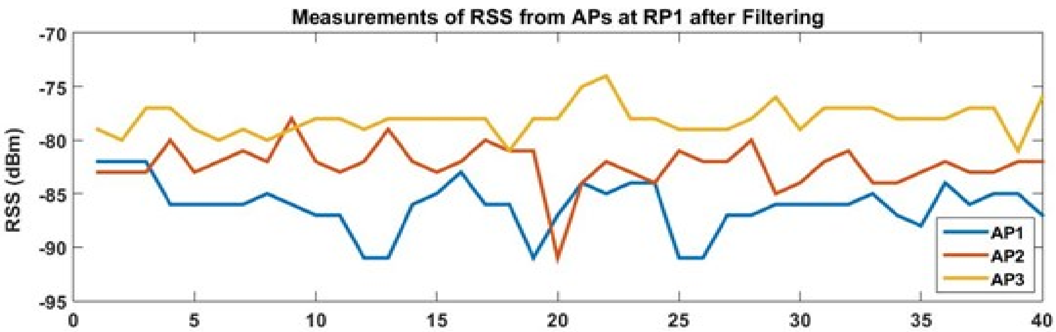
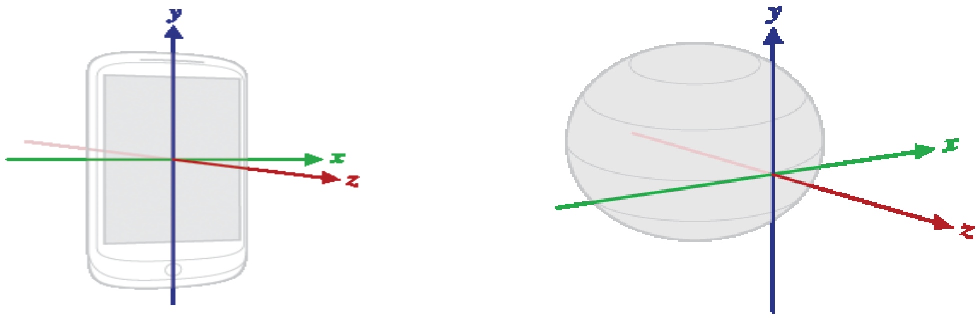


Figure 2.3. *Measurements of RSS values from three APs in first RP after filtering approach*

As seen in the Figure 2.2 and 2.3, the interval between the minimum and maximum values of WiFi RSS values from APs in a RP comes closer, and the -100dBm values are replaced with the lowest detected signal measurement in the dataset. There are also other filtering approaches are applied frequently for WiFi RSS values such as median filtering [24] and neighbourhood mean filtering [25] in the literature. Median filtering is a non-linear filtering method which replaces NaN values in the radio map with the median value of the neighbourhood, whereas in neighbourhood mean filtering, the NaN values are replaced with the average value of the neighbourhood. Since the performance of these filtering approaches are depended on the positioning algorithm, there is not an exact inference is obtained which filtering method is best suit for the radio map.

Magnetometer sensor returns the x, y and z values of the MF strength values in relation to the device orientation, therefore, the magnitude of each axis may differ as the device's orientation changes, even when it stays in the same position. So, they must be converted to the world coordinates before stored in the magnetic map. These orientations are represented in the Figure 2.4.



(a) (b)

Figure 2.4. *Local axis for device coordinates (a), global axis for world coordinates (b)*

The accelerometer and gyroscope included on the mobile phone can be used to convert the device orientation to world coordinates. Yaw (ϕ) and pitch (θ) angles obtained from the accelerometer are integrated with heading angle (ψ) from the MF and gyroscope sensors using Kalman Filter to obtain the orientation of the mobile device. The orientation angles are used to construct rotation matrix as seen in Eq. 2.3.

$$
\begin{aligned}
R_x(\phi) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix} \\
R_y(\theta) &= \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix} \\
R_z(\psi) &= \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned}
\tag{2.3}
$$

Then, the local magnetic field strength measurements are multiplied with the rotation matrix which includes orientation angles to calculate the global magnetic field strength measurements as follows.

$$B_p = R_x(\phi) R_y(\theta) R_z(\psi) B_e \tag{2.4}$$

where B_e and B_p are the magnetic field strength vector in device and world coordinate orientations, respectively. Figure 2.5 is constructed to demonstrate changes of the x, y and z components of all MF data in time in device coordinate orientations.

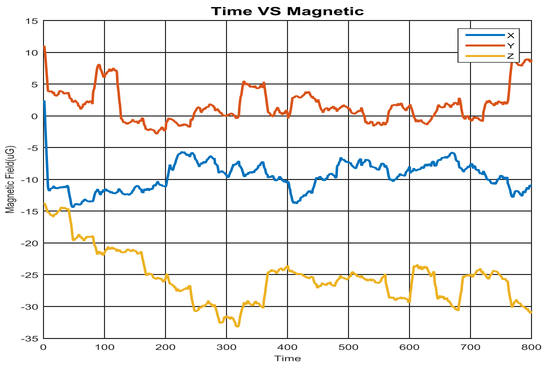


Figure 2.5. *Magnetic field x, y, and z components in time*

2.2.2. AP Selection

The dimension of the radio map is increased when using all the detected APs in the experimental area. But, same accuracy results may be obtained using less APs after applying AP selection methods. Performing AP selection methods before positioning, less complex models are constructed and the overfitting is decreased. Besides, computation time of the IPS is reduced. Various AP selection methods are proposed in the literature. For example, MaxMean method that is proposed in [26] ranks APs in descending order of their average signal-strength values, and selects the kth strongest APs to reduce the computation time. In [27], the most discriminating APs are selected using the information entropy by the Info Gain method for enhancing computation time. In [28], Principal Component Analysis is applied to reduce the computation complexity besides reserving all the APs's information. In [29], AP selection method is based on the minimizing correlation between selected APs. In [30], an AP selection strategy named as ResidualRanking applied only in the online phase to choose the APs which least sensitive to the dynamic environment conditions in an indoor positioning.

2.2.3. Fingerprint Map Clustering

With the increasing number of RPs in experimental area, the size of the fingerprint map continually expanding. This expansion effects the accuracy and computation time of the IPS negatively. Therefore, clustering is applied into IPS to divide the experimental area into sub-areas to improve the system performance. Clustering is done by evaluating fingerprint map based on similarity degree of the signal measurements. Performing clustering prior to positioning has some advantages. Firstly, it diminishes the negative effects on positioning accuracy caused by signal measurement deviations. Secondly, it decreases the computation time since cluster-specific models are constructed for positioning on each sub-area.

Various methods which are based on clustering are developed in the literature. In [31], median and K-means clustering algorithms are applied separately into the fingerprint map. K-means clustering is combined with KNN algorithm to improve the performance of the classical KNN approach in [32]. Experimental results demonstrate that KNN with K-means clustering reduces the required data for positioning and average distance errors. Support vector machine-based clustering (SVM-C) is proposed to reduce the positioning mean error in [33]. In [34], the authors reveal that combining different metrics for different steps of the cluster-specific positioning algorithms enhance the IPS performance.

2.3. Classification Algorithms for Positioning Purpose

Classification algorithms are used to determine the MU position in the positioning phase. They can be categorized into two groups such as deterministic algorithms and probabilistic algorithms. The deterministic algorithms use original or mean values of signal measurements observed at a RP whereas the probabilistic algorithms use distributions of signal measurements at a RP. In the literature, various deterministic and probabilistic algorithms have been used to construct model for the positioning. Since it is not the objective of this thesis to list all the existing algorithms, only the applied deterministic algorithms such as k-nearest neighbour, decision tree, support vector machine, artificial neural network, and extreme learning machine, and probabilistic algorithms such as Bayesian approach, and Maximum Likelihood Estimation are presented below. They are selected due to their efficiency, and wide usage in indoor positioning.

2.3.1. K-nearest Neighbour (KNN) Algorithm

KNN algorithm is one of the simplest algorithms to estimate the position of the MU by using fingerprint map [15]. It considers K RPs to calculate approximate position of the MU. The aim of the KNN is to compare the signal measurements in the fingerprint map with the observed measurement of the MU, and to choose the K RPs with the closest signal measurements. Euclidean, Manhattan, Chebyshev distance functions are used in the comparison. The most common RP in the closest K RPs is determined as the MU position. In general, KNN with $K = 3$ and $K = 4$ can achieve better accuracy than NN [35]. However, if the density of the fingerprint map is high, then NN can perform as well as the more complicated algorithm [36].

In the literature, several algorithms are proposed based on KNN. In a study, a feature-scaling-based KNN (FS-KNN) is proposed [37]. In the FS-KNN, RSS-level-based scaling weights are defined to calculate the signal differences when computing the similarity between signal vectors. The experimental results demonstrate that FS-KNN outperforms classical KNN algorithm in terms of positioning accuracy and precision. Jffreys&Matusita distance is replaced with Euclidean distance in KNN algorithm to improve the positioning accuracy and stability [38]. Cluster-specific KNN algorithms are proposed in the literature in order to enhance the performance of classical KNN algorithm. Cluster filtered KNN (CKF) algorithm that utilizes hierarchical clustering to divide the nearest neighbours of RPs is proposed in [39]. In [40], CKF is enhanced by using k-means clustering algorithm instead of hierarchical clustering algorithm. In another study, KNN is integrated with fuzzy c-means clustering (KNN-FCM) [41]. KNN-FCM divides k-nearest neighbours into several clusters and chooses one cluster to calculate the MU's position. KNN with affinity propagation clustering algorithm is utilized for positioning and outperforms other mentioned cluster-specific KNN algorithms [42]. The proposed method in [42] is improved by removing isolated RPs using semi-supervised affinity propagation clustering algorithm in [43].

2.3.2. Decision Tree (DT) Algorithm

Decision tree algorithm constructs tree structure to build models by using the fingerprint map [44]. A DT contains root, decision, and leaf nodes. The root node has zero or more outgoing edge without incoming edge. The decision node contains two or more branches with one incoming edge. It is represented with the function of any attribute of the fingerprint map. It breaks down the fingerprint map into smaller subsets. Each leaf node is assigned a classification decision. The sensor values and reference point labels are utilized in the decision and leaf nodes, respectively in the indoor localization problem. The position of the MU is calculated by traversing the tree from the root node down to a leaf node, according to the output of the condition or the function along the path. Since the goal is to construct the optimal DT by minimizing the generalization error, several heuristic methods are needed for solving the problem such as Iterative Dichotomizer (ID3) [45], C4.5 [46], Classification and Regression Tree (CART) [47], and so on. The C4.5 algorithm is generally used method to find the optimal tree. It takes the fingerprint map as input and generates a tree using the divide-and-conquer algorithm.

In the literature, several algorithms are proposed based on DT. The work in [48] analyse the parameters of the DT and find the optimal DT for their IPS. The intervals of RSS values represent the nodes of the DT in [49]. Then, the proposed DT algorithm is compared with NN, Bayesian approach, and ANN in terms of accuracy and computation time, and experimental results demonstrate that the proposed DT algorithm is better than other algorithms in terms of both performance metrics. In another study, the proposed DT algorithm is compared with NN, and MLP in terms of accuracy and experimental results demonstrate that the DT is more accurate than other algorithms [50]. In [51], RSS combination of each AP is selected using Random Forest algorithm by recursively creation of DT. Multiple weighted DTs using boosting algorithm is proposed in [52] in order to enhance the computational complexity and accuracy of the IPS. The proposed algorithm in [53] combines information theory, clustering analysis, and a DT algorithm in order to increase the accuracy of the IPS while decreasing the power consumption. A DT based localization algorithm specific to context awareness is presented in [54].

2.3.3. Artificial Neural Network (ANN)

ANN is a mathematical model which simulates the functions of the human brain [55]. It consists of interconnected nodes and directed links in order to process information. The connections between these nodes contain weights which store the knowledge of this model. ANN algorithm is frequently used for indoor positioning field, since it is robust to noise and interference. Multi-layer perceptron (MLP) is one of the extensively used neural network topology that incorporates an input layer with input nodes and output layer. MLP was used for WLAN-based IPS by [55] for the first time. The MLP has a feed-forward layered structure. In addition to this, the MLP has the advantage of having hidden layers. The input layer and output layer represent input variables and output variables, respectively. The hidden layer is responsible for the capacity of the MLP and represents the connections between the input and output layers.

The classification process of MLP is a nonlinear mapping from a list of attributes (sensor measurements) into MU's location. The MLP performance is enhanced by tuning the parameters such as number of hidden layers, the number of nodes in each hidden layer, initial weights to start the training, the activation function, learning rate, and the momentum rate. Since there is no explicit algorithm to select optimal parameters, therefore, they can be selected empirically.

Various algorithms are proposed based on ANN structure in the literature. A multilayer feed-forward network with a secant activation function is used for positioning in [55]. Modular MLP (MMLP) is proposed in [56] to decrease the uncertainty arising from unavailable signals at positioning phase in IPS. In [57], the robustness of ANN algorithm to noise and interference is proved using three different environments. In addition to this, The ANN is compared with probabilistic model, and better results in terms of accuracy are obtained with the ANN. A cascade-connected ANNs and space partitioning are utilized in the proposed positioning algorithm in [58] to enhance the performance of the IPS in terms of accuracy. In [59], a multilayer feed-forward back-propagation based model with hyperbolic tangent sigmoid activation function is developed. The optimal parameters of the ANN are selected using genetic algorithm to improve the system performance in [59]. In a study, a discriminant-adaptive neural network (DANN) is proposed [60]. The DANN considers the redundant information as noise and extract the useful information from the available APs into discriminative components (DCs). Then, these DCs are inserted into neural network for updating weights. Since the network is trained only using the discriminative information, the IPS performance is improved. In addition to this, experimental results demonstrate that the DANN outperforms MLP in terms of positioning accuracy.

2.3.4. Extreme Learning Machine (ELM)

ELM is a new learning algorithm based on single-hidden layer feed forward neural network (SLFN) architecture [61]. It has faster training speed since it chooses the input weights randomly and calculates the output weights of SLFN analytically [62].

Various algorithms are proposed based on ELM structure in the literature. ELM algorithm is used to estimate the position by taking the advantage of signal strength and signal quality in [63]. The proposed model based on ELM is superior to KNN in terms of accuracy. The work in [64], the ELM is compared with KNN and MLP in terms of accuracy, and the ELM outperforms other algorithms. An indoor localization algorithm based on an online sequential extreme learning machine (OS-ELM) is proposed in [65]. OS-ELM is compared with ELM and it can provide higher accuracy and faster learning speed under dynamic environment conditions than ELM. Weighted version of ELM (WELM) and Signal Tendency Index (STI) are integrated in STI-WELM to construct an efficient and robust IPS [66]. STI is presented in order to handle the device heterogeneity and environmental changes in indoors. A feature adaptive online sequential extreme learning machine (FA-OSELM) is proposed in [67] in order to handle the changes of APs numbers in indoor areas. Robust ELM (RELM) considering close to mean (CTM) and small residual (SR) constraints is proposed to improve the robustness of the IPS [68]. Experimental results show that RELM achieves good performance in terms of accuracy, repeatability, and worst case error. A constraint online sequential extreme learning machine (COSELM) is proposed to deal with fluctuation of wireless signals over time [69]. Experimental results demonstrate that COSELM outperforms OS-ELM in terms of accuracy and computation time. There are also new methods are proposed recently in the literature which integrate deep learning with ELM to classify the unlabelled data in indoor positioning field [70, 71].

2.3.5. Support Vector Machine (SVM)

SVM is a non-parametric supervised learning algorithm which is used in the IPS by training the support vectors on the fingerprint map [72]. SVM is based on statistical learning theory and structural risk minimization. In the training phase, a decision boundary that separates the samples belonging to different classes is determined at an optimal level. The aim of the SVM is to obtain the optimal separation hyperplane to distinguish the classes from each other, i.e. to maximize the distance between the support vectors of different classes.

Various algorithms are proposed based on SVM structure in the literature. SVM algorithm is used in two versions such as classification and regression in [73]. Experimental results demonstrate that SVM as a classifier outperforms WKNN, Bayesian approach, and MLP in terms of accuracy. And, its accuracy results nearly same as WKNN results when its regression version is utilized for indoor positioning. The proposed positioning algorithm in [74] divides experiment area into sub-areas according to RSS features, and then applies SVM models in each sub-area to estimate the position. Since SVM has higher training time, Least Squares SVM (LS-SVM) is proposed in order to reduce the training time of SVM [75]. The work in [76] transforms LS-SVM into multiple binary classification problems by introducing axial decoupled LS-SVM (AD-LS-SVM). Experimental results demonstrate that AD-LS-SVM is superior to LS-SVM, SVM, and KNN in terms of accuracy and computation time.

2.3.6. Bayesian Approach

Naïve Bayes is a simple bayesian approach but effective probabilistic classification method for indoor positioning [16]. It utilizes the samples with known class labels to calculate the likelihood of new sample with unknown label belong to any of the existing classes. Bayesian Network (BN) is another bayesian approach method which is also frequently used in the literature for calculating MU's position [77]. The BN is a directed acyclic graph (DAG) which consists of nodes and edges.

Various algorithms are proposed based on Bayesian approach in the literature. In [78], Naïve Bayes algorithm is utilized considering user's orientation to deal with the blocking effect of human body. Improved Naïve Bayes Simple (INBS) learning algorithm is proposed in [79] in order to solve zero probability problems in WiFi-based fingerprinting approach. Domain Clustering (DC) based algorithm based on Naïve Bayes classifier is proposed in [80] to enhance the IPS performance in terms of accuracy. In [81], Bayesian-based location estimation system is proposed for indoor positioning. Bayesian hierarchical model is utilized for positioning in [82]. In [83], probabilistic fingerprinting approach is presented to reduce the computation time of traditional probabilistic fingerprinting algorithms.

2.3.7. Maximum Likelihood Estimation (MLE) Algorithm

Maximum likelihood estimation (MLE) is one of the most popular algorithms which take into account the standard deviation of the measurements [84]. It gives higher accuracy when compared with the other algorithms [85]. Additionally, the calculated likelihood values for different sensor types are useful for constructing hybrid solutions for the indoor positioning problem.

Various algorithms are proposed based on MLE structure in the literature. MLE is applied in [86] to estimate the MU's location. Experiment results in [87] prove that MLE is more accurate than WKNN, SVM, and MLP to estimate the MU's position. Maximum likelihood function is chosen for positioning in [88]. Maximum likelihood-based fusion algorithm which integrates WiFi IPS with a pedestrian dead reckoning system is presented in order to improve the positioning accuracy [89].

2.4. Datasets

Fingerprint-based indoor positioning is started with constructing datasets (fingerprint maps). Various datasets are proposed which are publicly available in the literature. In the following subsections, the datasets which are used in this study are described briefly.

2.4.1. KIOS Dataset

KIOS dataset is constructed by collecting RSS data at KIOS Research Centre which is a $560m^2$ typical office environment that consists of offices, labs, a conference room and corridors [19]. 9 APs and 5 different mobile devices were used for data collection (HP iPAQ hw6915 PDA with Windows Mobile, an Asus eeePC T101MT laptop running Windows 9, an HTC Flyer Android tablet and two other Android smartphones (HTC Desire, Samsung Nexus S)). Training dataset is constructed by collecting RSS measurements from all 9 APs, at 105 distinct reference locations by carrying all 5 devices at the same time. There are 20 fingerprints per reference points, so total number of fingerprints in training data is 2100. Besides, test data are collected 2 weeks later by walking along a predefined route 10 times at the same time with all devices. There are 96 locations on this route most of which different from the RPs. There are ten fingerprints are collected per each test location, so total number of fingerprints in test data is 960. This database is used to solve device calibration problems in indoor positioning [19]. The experimental setup of KIOS dataset is given in Fig. 2.6.

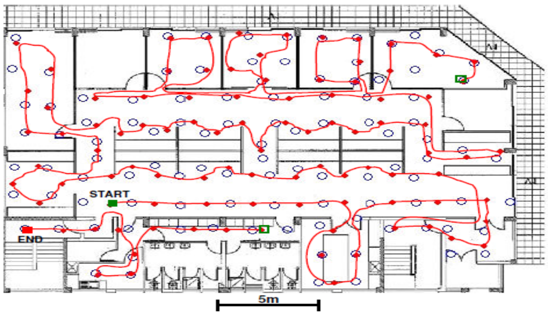


Figure 2.6. *Experimental setup at KIOS Research Centre [20]*

As seen in Fig. 2.6, blue circles represent the training RPs, and red points represent the test RPs. Test data are collected by following a route which is represented with a red line in Fig. 2.6.

2.4.2. UJIIndoorLoc Dataset

UJIIndoorLoc database is the biggest and publicly available database in the literature [20]. Data were collected from a surface of $108703m^2$containing 3 buildings with 4 or 5 floors depending on the building. There are 933 RPs and 520 different wireless access points (WAPs) including in the database. More than 20 users using 25 different mobile devices collect fingerprints. There are 19938 fingerprints for training and 1111 fingerprints for testing are recorded. Test data are collected 4 months later after training data. This database could be used to make comparisons among different methods in indoor positioning. The experimental setup of UJIIndoorLoc dataset is given in Fig. 2.7.

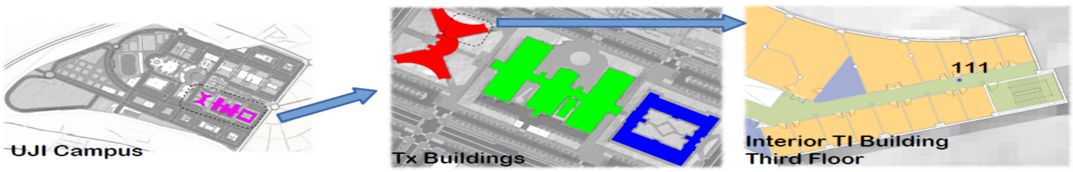


Figure 2.7. *Experimental setup at UJI University Campus [21]*

As seen in Fig. 2.7, left subfigure shows the UJI University campus, the centre figure shows the three buildings of the School of Technology and Experimental Sciences, and right figure is a zoom inside the third floor of the TI building. The RP which is numbered with 111 is an example point in the right figure [20].

2.4.3. RFKON Dataset

System architecture and experimental setup of our IPS are explained in this section [23]. The system architecture of our indoor positioning system (IPS) contains two major units named as Gezkon, and Konsens. Gezkon is a mobile application and it is responsible for collecting WiFi-RSS and MF strength values from the test area. Konsens is a server which is used to estimate the position of mobile device and also responsible for updating and calibrating of RFKON database. The communication between Konsens and Gezkon is achieved by sensor nodes through Data Distribution Service (DDS) layer. Konsens maintains sensor nodes. The system architecture of our IPS is shown in Figure 2.8.

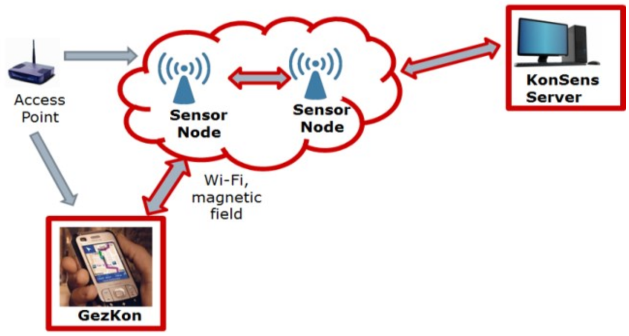


Figure 2.8. *System architecture of our indoor positioning system (IPS)*

The layout of the test bed for RFKON database is shown in the Figure 2.9. In this area, real-world indoor localization experiments are conducted to evaluate the performance of the proposed f-score weighted indoor positioning algorithm. The test bed is the Teknopark in the Eskisehir Osmangazi University. The area of the test-bed is around $800m^2$. The area is broken into $2.4m \times 2.4m$ size grid squares and the center of the each grid is reported as RP. There are five sensor nodes which are represented with small squares deployed at the locations on the first floor. The green bolded sub-region is used to collect sensor measurements from the area.

Figure 2.9. *Experimental setup at Eskisehir Osmangazi University, Teknopark*

2.4.4. Other Datasets

There are also other datasets which are publicly accessible in the literature. For example, WiFi RSS, magnetic field strength, and inertial sensors measurements including accelerometer and orientation values are collected to construct the database in [90]. The measurements are collected by two users each wear both smartphone and smartwatch. In another database, WiFi and Bluetooth RSS measurements are collected with 28 Android phone users for three weeks in [91]. WiFi RSS data are collected with a mobile robot in indoor and outdoor environments by recording the robot location using its odometer in [92]. UJIIndoorLoc database [20] is enhanced by collecting magnetic field sensor measurement in the same environment in [93]. Miskolc IIS Hybrid dataset is recently uploaded dataset which includes data acquired from WLAN card, magnetometer, and Bluetooth interface [94].

2.5. Performance Metrics

IPS developers recognize the wellness of their own positioning systems considering performance metrics. Several metrics are introduced in the literature to evaluate the performance of the IPS. These metrics are defined as follows:

- Accuracy

  Accuracy is the most important performance metric. It is defined as the mean distance error that exists between the predicted and actual position of the MU. The IPS with higher accuracies is preferred.

- Precision

  Precision is the percentage of positive predictions, i.e. the measure of how good predictions are with respect to false positives (FP). In terms of positioning purpose, precision considers how often the system works, and the consistency between the results of the IPS. Cumulative distribution function (CDF) is used to measure the precision of an IPS. When two IPS have same accuracy values, then the system which reached the highest probability values faster is preferred.

- Recall

  Recall is the ratio of correctly classified positive instances to the total instances in a positive class, i.e. the measure of how good the predictions are with respect to false negatives (FN).

- F-score

  F-score metric is established to optimize accuracy with precision and recall. It is the harmonic mean of precision and recall. Thus, it considers both false positives (FP) and false negatives (FN).

- Complexity

  The complexity of an IPS is measured with the time that it takes to calculate the MU's position. This time is high for more complex system. Since the shortage of battery life in mobile devices, it is recommended to keep the complexity of an IPS as low as possible.

- Robustness

  A robust IPS can be work even some signals are not seen or are distributed because the indoor environment structure is changed.

- Scalability

  Scalability is affected by the size of the indoor area. When the size of an indoor area is huge, IPS requires extra calculation and extra communication infrastructure to cover the indoor area.

- Cost

  The cost of an IPS depends on required infrastructure, time, money, space, and so on.

In the context of this thesis, the proposed methods are evaluated in terms of accuracy, precision, and time complexity.

3. PROPOSED METHODS

Indoor positioning still remains unsolved today since there is no standardized solution like as GPS for outdoor environment. Therefore, various solutions are proposed in the literature for indoor positioning problem. In this section, we present our own solutions. The proposed methods in the context of this thesis use the existing infrastructure of the indoor area, so they are inexpensive. We start to develop methods using WiFi signals, and enhanced the methods by combining MF measurements to achieve better positioning accuracies. All the methods adopt fingerprint-based positioning approach which is explained in the Section 2. The proposed methods in this thesis are introduced in the following subsections.

3.1. A Multi-criteria Decision Strategy to Select a Machine Learning Algorithm for Indoor Positioning System

Several machine learning (ML) algorithms are applied in indoor positioning field. In this method, k-nearest neighbor (KNN), support vector machine (SVM), decision tree (DT), naïve bayes (NB) and bayesian networks (BN) which are explained briefly in Section 2.3 are compared. In the experiments, UJIIndoorLoc, KIOS and RFKON datasets that are given in Section 2.4 are used. The experiments are performed into two categories. In the first category, the selected positioning algorithms are applied directly using all the attributes of each dataset. In the second category, a preprocess phase is employed and feature selection and extraction methods are used to eliminate redundant attributes to reduce the dimension of each dataset. In addition to these, ensemble learning algorithms, namely adaBoost and bagging, are used to enhance the performance of the selected algorithms such as DT and KNN. Then, Experimental results are reevaluated using a multi-criteria decision strategy to select the most appropriate algorithm. The analytical hierarchy process (AHP) is applied for the multi-criteria decision process.

3.1.1. Preliminaries

IPS can be evaluated using several performance criteria, and an appropriate ML algorithm is selected considering the values for each criterion. Multi-criteria decision strategies can be used concurrent evaluation of various performance criteria. In following subsections, preliminaries are given for these topics.

The performance criteria in indoor positioning

Computation time, accuracy, precision, recall, f-score and some other metrics related to the performance of ML algorithms may be considered to evaluate the performance of an IPS.

Accuracy is one of the most widely used performance criteria for an indoor positioning system and a ML algorithm. In the area of ML, it is measured based on the percentage of correctly classified instances over total instances. Although, accuracy is easy to use, understand, and compute with less complexity, it induces suboptimal solutions when dealing with uneven class distributions and produces less discriminating values. To overcome the limitations incurred from accuracy, precision and recall are defined. Precision is the percentage of positive predictions, i.e. the measure of how good predictions are with respect to false positives (FP). Recall is the ratio of correctly classified positive instances to the total instances in a positive class, i.e. the measure of how good the predictions are with respect to false negatives (FN). To optimize accuracy with precision and recall, the f-score metric is established that is a harmonic mean of precision and recall. Thus, it considers both FP and FN.

Computation time is critical performance criterion of a ML algorithm. It depends on both the size of the problem and the complexity of the ML algorithm. It is calculated by the sum of the values of training time and test time. The training time is the time taken to build the training model, and the test time is the time to predict the position using the training model. Since, the test time is negligible compared the training time, the computation time of a ML algorithm is usually measured based on the training time.

Multi-criteria Decision Strategy

The performance criteria can be used to select a ML algorithm for a specific indoor positioning system. As mentioned before, there exists performance metrics such as accuracy, precision, recall, f-score or computation time. In order to evaluate multiple criterions at the same time a multi-criteria decision strategy [95] is required. In the literature, AHP [95] is used in many areas for concurrent evaluation and decision making.

In the multi-criteria decision making, first of all the selected performance criteria are normalized. The maximized criterion is normalized using Eq. (3.1),

$$x_{ij}^{'} = 1 - \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}} \tag{3.1}$$

and, the minimized criterion is normalized using Eq. (3.2)

$$x_{ij}^{'} = \frac{x_{ij} - x_{\min}}{x_{\max} - x_{\min}} \tag{3.2}$$

where x_{ij} is the normalized value, $x_{ij}^{'}$ is the raw value of the j^{th} algorithm i^{th} criteria, and $x_{\min}$, and $x_{\max}$ is the minimum and maximum score of the selected criteria in the experimental results, respectively. Then, selected criteria are ready to be integrated. Secondly, various criteria can be integrated using user preferences and corresponding relative weights. The AHP procedure for relative weight calculation from the user preferences is explained through an example with three criteria that is used in this study. In the AHP, the terms equal, moderate, strong, very strong and extreme importance are represented by the numbers 1, 3, 5, 7 and 9, respectively to make a pairwise comparison. And, the interval values 2, 4, 6, and 8 are also used to make judgements between any two criteria. The pair wise comparison of the criteria is given in Table 3.1. In this table, the diagonal elements are set to 1. Each criterion is placed in the Table 3.1 according to their importance. The upper triangular part of this table is constructed according to user preferences. Each element of the lower triangle is set to the inverse of these pairwise comparisons automatically. As seen in Table 3.1, a 'very strong' linguistic term is selected to judge the accuracy over the computation time, which is enumerated with 7. Then, the importance of the computation time over the accuracy is set to 1/7 automatically.

Table 3.1. *Normal pairwise comparison of criteria*

	Accuracy	F-score	Computation Time
Accuracy	1	3	7
F-score	1/3	1	5
Computation Time	1/7	1/5	1

The user preferences in Table 3.1 are transformed into the relative weights as follows:

Construct a matrix *A* from Table 3.1 as

$$A = \begin{bmatrix} 1.00 & 3.00 & 7.00 \\ 0.33 & 1.00 & 5.00 \\ 0.14 & 0.20 & 1.00 \end{bmatrix}$$

Then, the following equation

$$\frac{a_{ij}}{\sum_{i=1}^{n} a_{ij}} \tag{3.3}$$

is applied to each cell of matrix A to obtain the following matrix D:

$$D = \begin{bmatrix} 0.68 & 0.71 & 0.54 \\ 0.22 & 0.24 & 0.38 \\ 0.10 & 0.05 & 0.06 \end{bmatrix}$$

After this step, the weight vector W is created by the arithmetic mean of each row of the D matrix.

$$W = \begin{bmatrix} 0.65 \\ 0.28 \\ 0.07 \end{bmatrix}$$

Each entry of the weight vector W corresponds to the relative weights. In this example, the relative weights $w_1 = 0.65, w_2 = 0.28$ and $w_3 = 0.07$ correspond to the accuracy, f-score, and the computation time, respectively, where $\sum_{i=1}^{3} w_i = 1$. After obtaining the relative weight for each criterion, a linear combination is applied to the aggregation of the multi-criteria with their weights. According to the above equations, the final test result of the j^{th} algorithm TR_j can be calculated by Eq. (3.4).

$$TR_j = \sum_{i=1}^{n} w_i x_{ij}^{'}, j = 1...m \tag{3.4}$$

where n is the number of selected criteria and m is the number of selected algorithms.

3.1.2. Proposed Method

There are various ML algorithms that used for fingerprint-based indoor positioning systems. The ML algorithms show different performance depending on the datasets. In order to increase the expected performance, a proper ML algorithm should be selected for a given indoor positioning system considering multiple criteria. In this study, a multi-criteria decision strategy is applied to find the most appropriate ML algorithm for any given indoor positioning system considering the user preferences. The flowchart of the proposed multi-criteria decision strategy is given in Fig. 3.1.

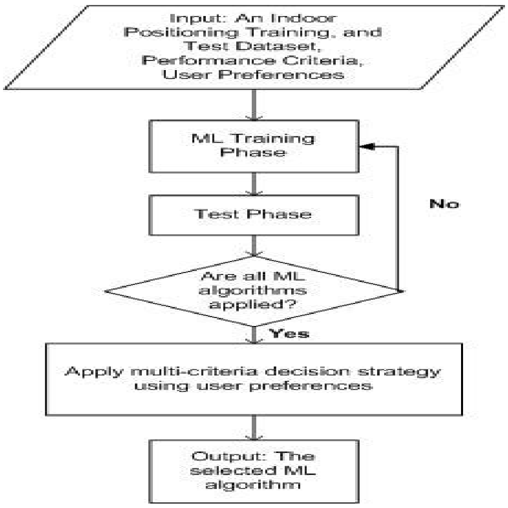


Figure 3.1. *The flowchart of the proposed multi-criteria decision strategy*

In order to run the overall algorithm, an indoor positioning training and test datasets, performance criteria, and user preferences should be defined as an input. Then, in the training phase, each ML algorithm is applied to the indoor positioning system as in Fig. 3.2. This step is performed in two ways. In the first way, classification is done with preprocessing using Correlation-based Feature Selection (CFS), Chi Square Selection (CHI), Filtered Attribute Selection (FILT), Gain Ratio Selection (GAIN), and Principal Component Analysis (PCA) to remove redundant APs. In the second way, classification is performed without preprocessing algorithms. Decision Tree (DT), Naïve Bayes (NB), Bayes Net (BN), Sequential Minimal Optimization (SMO), and Nearest Neighborhood (NN) are utilized as classifiers. And, two ensemble learning algorithms (AdaBoost (AB) and Bagging (BAG)) are used in order to improve the performance of the DT and NN.

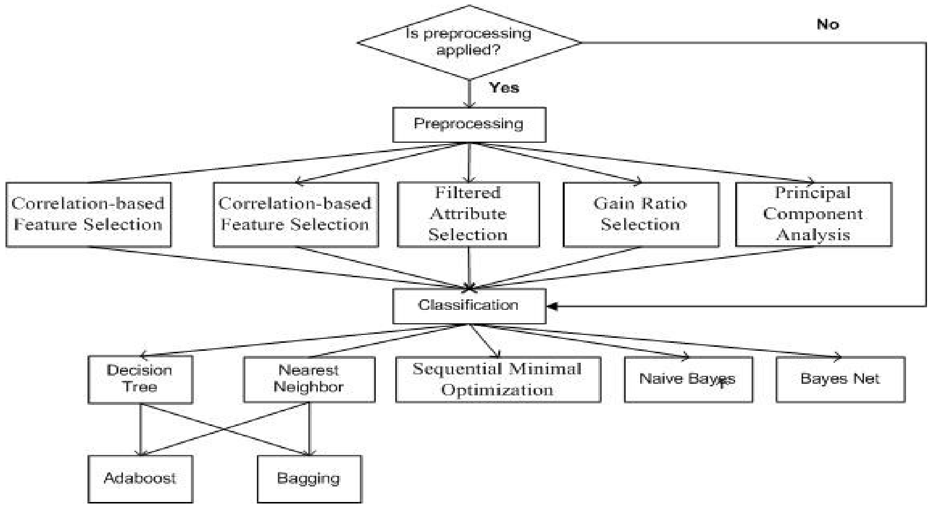


Figure 3.2. *The block diagram of the ML training phase*

In the test phase, the test data are classified using the training model which obtained from the training phase. In this step, all performance metrics are calculated and returned to the final step. After applying all ML algorithms, finally, the multi-criteria decision strategy as mentioned in Section 3.1.2 is applied to find the most appropriate algorithm for the indoor positioning problem using the selected performance criteria and user preferences. In this final step, firstly, the selected performance metrics are normalized. Then, they are weighted according to user preferences with AHP procedure. Finally, they are aggregated with Eq. (3..4) to obtain the final result for each algorithm. The algorithm that gives the minimum value using Eq. (3.4) gives the most appropriate algorithm for the given indoor positioning dataset.

3.1.3. Experimental Results

In this subsection, seven different classifiers (Decision Tree (DT), Naïve Bayes (NB), Bayes Net (BN), Sequential Minimal Optimization (SMO), Nearest Neighborhood (NN), AdaBoost (AB), Bagging (BAG)), and five preprocessing algorithms Correlation-based Feature Selection (CFS), Chi Square Selection (CHI), Filtered Attribute Selection (FILT), Gain Ratio Selection (GAIN), Principal Component Analysis (PCA) from WEKA are tested. The analysis has been performed on a Windows 7 operating system with Intel® CoreTM i7-4510U CPU, 2.00 GHz Processor and 8.00 GB RAM.

Classification results without preprocessing

In this part, selected classifiers are applied to the all existing datasets without any preprocessing. Fig. 3.3 shows the accuracy results of each classifier without removing any attributes from the datasets.

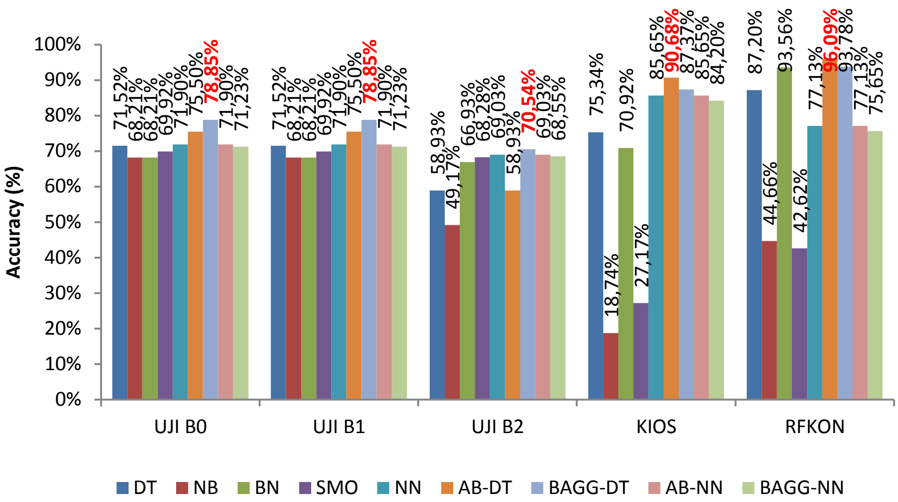


Figure 3.3. *Accuracy result of raw data*

As seen in Fig. 3.3, BAGG-DT has resulted highest accuracy, for UJI datasets. And, AB-DT gives the best accuracy results, 90.68% for KIOS and 96.09% for RFKON datasets. It can be deduced from Fig. 3.3 that ensemble learning algorithms are improve accuracy results of applied classifiers. Computation time results of each classifier using all attributes are given in Fig. 3.4. In this experiment, since the value of minimum computation time is more important, any algorithm that has bigger than 100sec of computation time is assumed to have 100sec computation time.

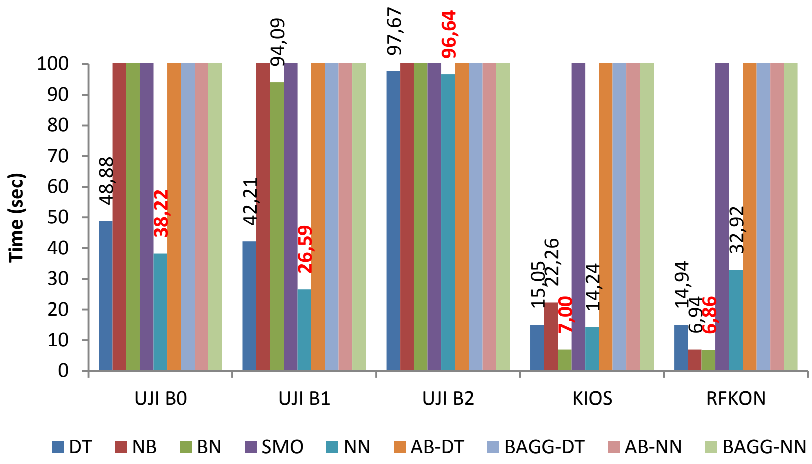


Figure 3.4. *Computation time of raw data*

Computation time results from Fig. 3.4 show that NN is superior to other algorithms for all UJIIndoorLoc datasets and BN gives the bests for KIOS and RFKON datasets.

Classification Results with Preprocessing

Feature selection methods are applied in IPS to remove redundant or irrelevant attributed from the fingerprint map. After removing redundant attributes, the remaining attributes are adequate for positioning. So, the training time of the IPS is decreased. And, the irrelevant attributes do not contain any useful information; the generalization performance of the applied algorithm for positioning is enhanced. In this section four feature selection (Correlation-based, Chi-square, Filtered, and Gain Ratio) methods and an extraction method (Principal Component Analysis) are used in preprocessing step for each classifier. Linear Forward Selection is selected as a search method in Correlation-based feature selection algorithm. Ranker is applied as a search algorithm to order the attributes for Chi squared, Filtered, Gain ratio attribute evaluators and Principal Components. The threshold value for ranker is selected as zero.

Analysis using RFKON dataset

RFKON dataset contains 27 attributes, i.e. APs initially. The number of attributes is reduced after preprocessing. The accuracy results of classifiers after removing redundant APs are given in Fig. 3.5.

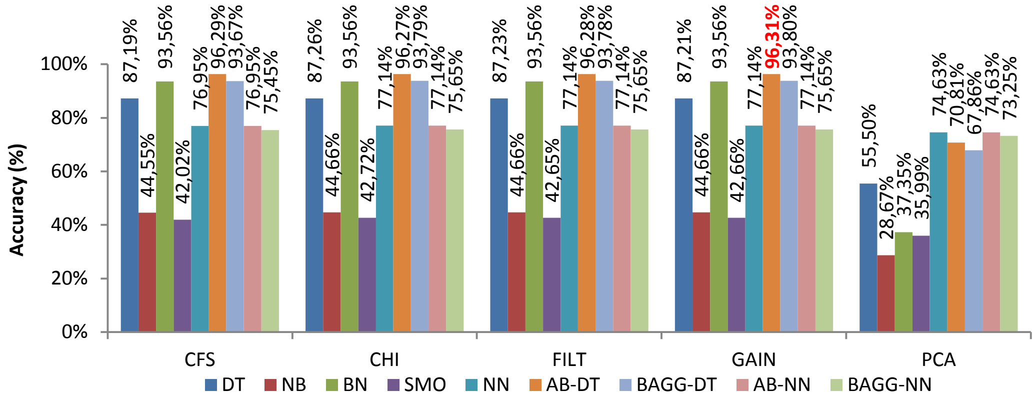


Figure 3.5. *Accuracy results of RFKON after preprocessing*

Fig. 3.5 reveals that best accuracy result is obtained from AB-DT considering 24 attributes instead of considering all attributes in the dataset after utilizing Filtered attribute evaluator. The best accuracy result (96.31%) is nearly same as the best accuracy result (96.09) that obtained considering all attributes. To show the improvement of preprocessing step over the computation time Fig. 3.6 is constructed.

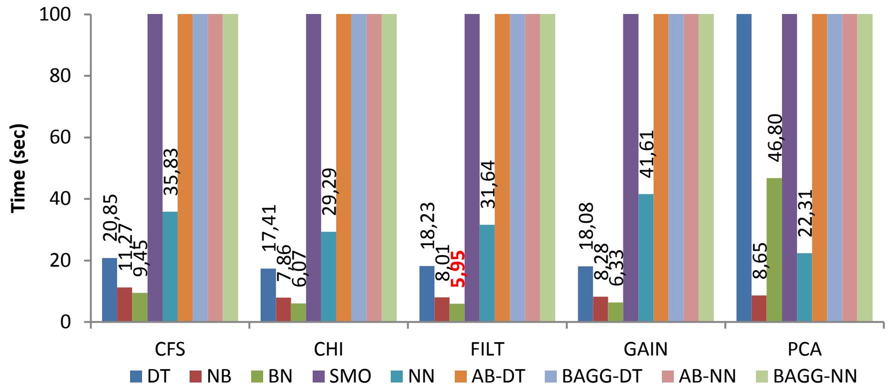


Figure 3.6. *Computation time results of RFKON after preprocessing*

It can be deduced from Fig. 3.6 that, Filtered attribute evaluator (FILT) reduces the computation time of constructing training model of BN from 6.86 sec to 5.95 sec.

Analysis using KIOS dataset

KIOS dataset contains 70 attributes before utilizing preprocessing step. The irrelevant attributes are removed from the database after this step. The accuracy results of classifiers after removing redundant APs are given in Fig. 3.7.

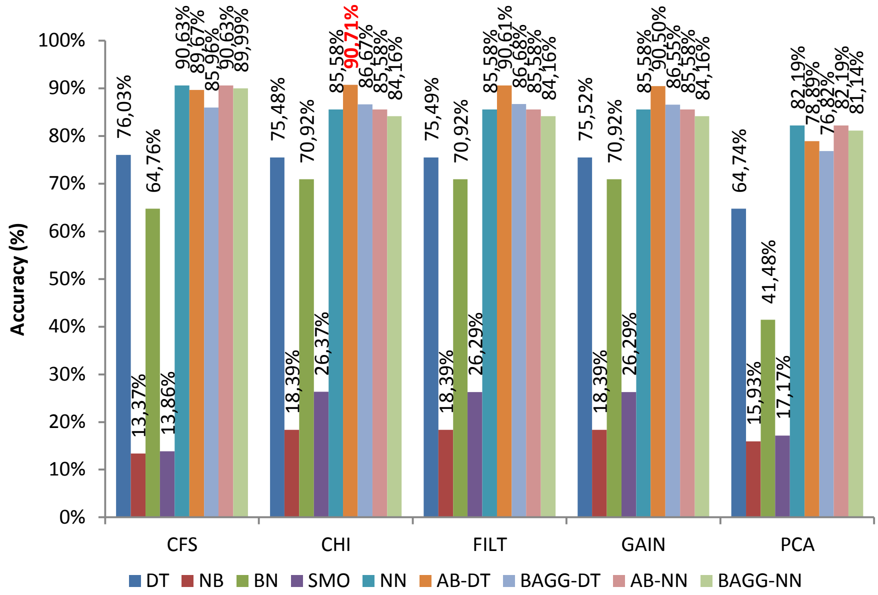


Figure 3.7. *Accuracy results of KIOS after preprocessing*

As seen in Fig. 3.7, better classification performance is achieved by AB-DT reducing number of attributes from 70 to 40 after applying Chi-square attribute evaluator. The best accuracy result (90.71%) is nearly same as the best accuracy result (90.68) that obtained considering all attributes. Computation time results of KIOS dataset are given in Fig. 3.8.

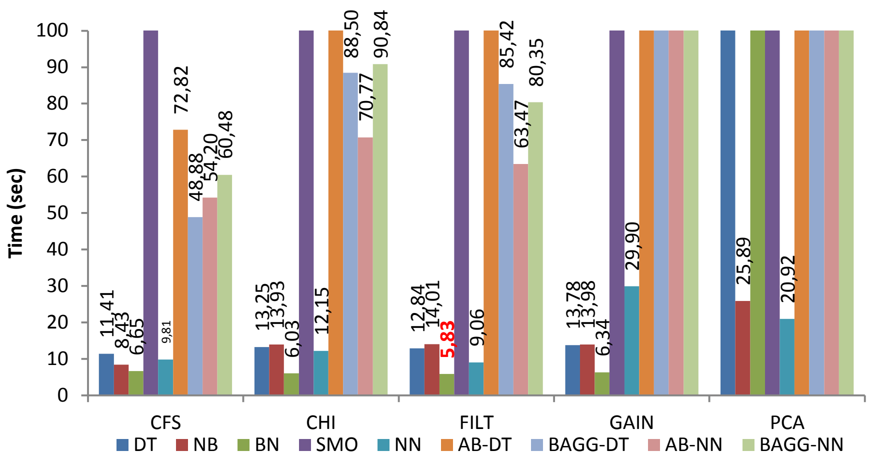


Figure 3.8. *Computation time results of KIOS after preprocessing*

Fig. 3.8 reveals that, Filtered attribute evaluator (FILT) with BN has resulted into lowest computation time. Filtered attribute evaluator (FILT) reduces time taken to build training model of BN from 7 sec to 5.83 sec.

Analysis using UJI B0 dataset

UJI B0 contains the data obtained from the first floor of UJIIndoorLoc dataset and includes 521 attributes before utilizing preprocessing step. There are 130 attributes remained after this step that is a considerable reduction of the database size. The accuracy results of classifiers after this step are given in Fig. 3.9.

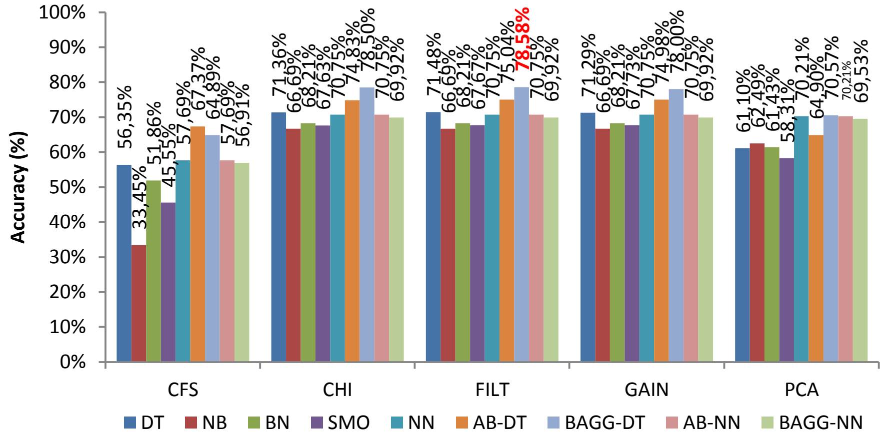


Figure 3.9. *Accuracy results of UJI B0 after preprocessing*

Fig. 3.9 reveals that best accuracy result is obtained from BAGG-DT selecting 130 attributes from the dataset using FILT instead of considering all attributes in the dataset. The best accuracy result is 78.58%. To show the improvement of preprocessing step over the computation time Fig. 3.10 is constructed.

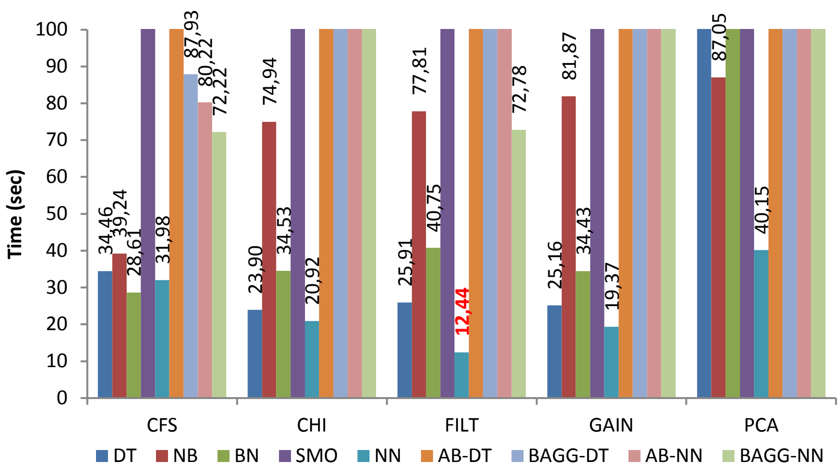


Figure 3.10. *Computation time results of UJI B0 after preprocessing*

It can be deduced from Fig. 3.10 that, Filtered attribute evaluator (FILT) reduces the computation time of constructing training model of NN from 38.22 sec to 12.44 sec.

Analysis using UJI B1 dataset

UJI B1 contains the data obtained from the second floor of UJIIndoorLoc dataset and includes 521 attributes before applying preprocessing step. There are 138 attributes remained after this step that is a considerable reduction of database size. The accuracy results of classifiers after this step are given in Fig. 3.11.

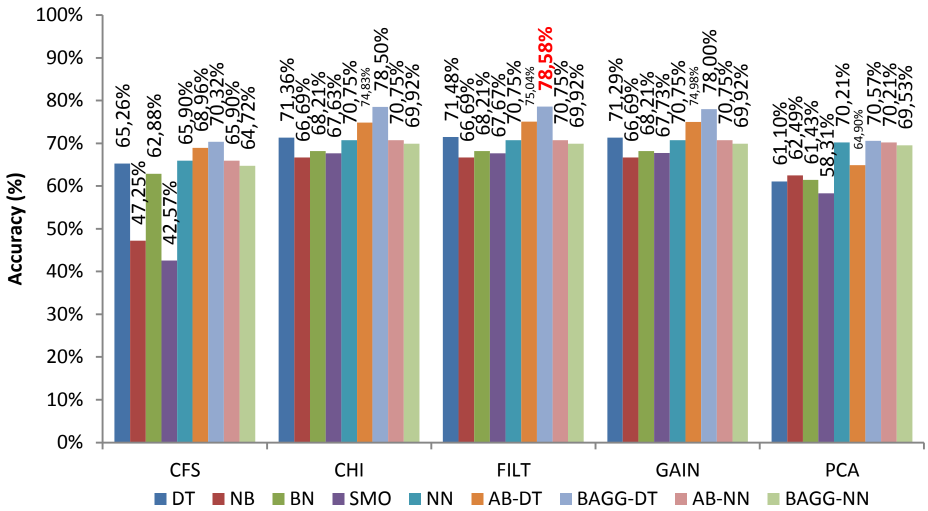


Figure 3.11. *Accuracy results of UJI B1 after preprocessing*

Fig. 3.11 reveals that best accuracy result is obtained from BAGG-DT with FILT considering only 138 attributes. The best accuracy result (78.58%) is nearly same as the best accuracy result (78.85) that obtained considering all attributes. To show the improvement of preprocessing step over the computation time Fig. 3.12 is constructed.

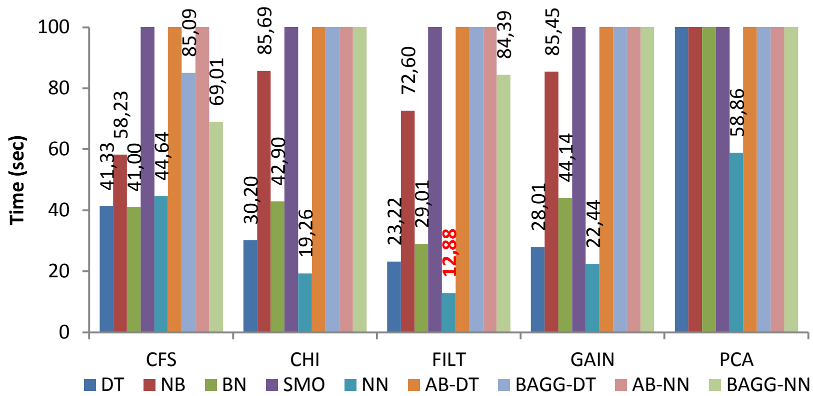


Figure 3.12. *Computation time results of UJI B1 after preprocessing*

It can be deduced from Fig. 3.12 that, Filtered attribute evaluator (FILT) with NN stands as a better performer with a computation time of 12.88 sec. FILT reduces the computation time of constructing training model of NN from 26.59 sec to 12.88 sec.

Analysis using UJI B2 dataset

UJI B2 contains the data obtained from the second floor of UJIIndoorLoc dataset and includes 521 attributes before applying preprocessing step. There are 100 attributes remained after this step that is a remarkable reduction of database size. The accuracy results of classifiers after this step are given in Fig. 3.13

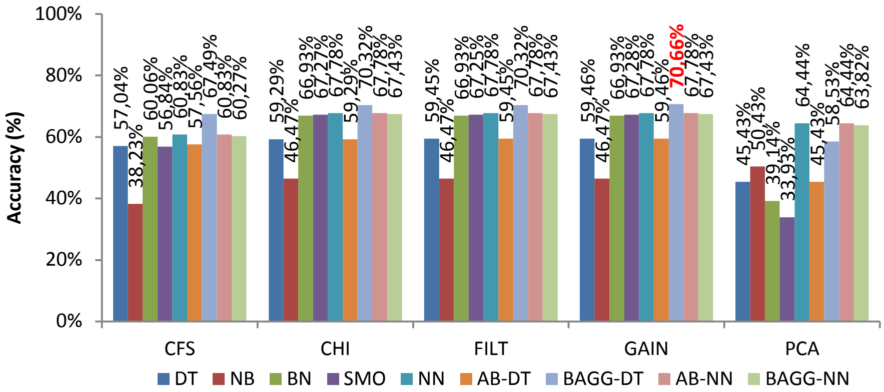


Figure 3.13. *Accuracy results of UJI B2 after preprocessing*

Fig. 3.13 reveals that best accuracy result is obtained from BAGG-DT with filtered attribute evaluator considering only 100 attributes. The best accuracy result

(70.66%) is nearly same as the best accuracy result (70.55) that obtained regarding all attributes. To show the improvement of preprocessing step over the computation time Fig. 3.14 is constructed.

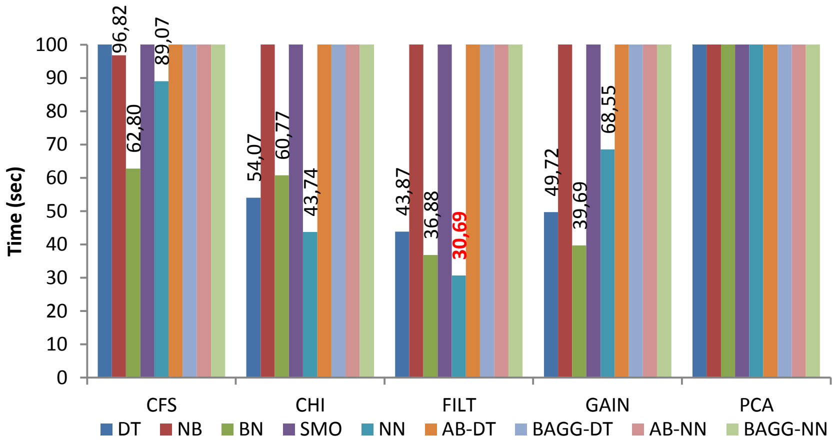


Figure 3.14. *Computation time results of UJI B2 after preprocessing*

Specifically, as shown in Fig. 3.14, Filtered attribute evaluator (FILT) with NN succeeds to obtain low computation time. Filtered attribute evaluator (FILT) reduces the computation time of constructing training model of NN from 96.64 sec to 30.69 sec. The summary of all the experiments including best results is shown in the Table 3.2 below.

Table 3.2. *Best results obtained from all experiments*

Dataset	Raw Data			
	Number of Attributes	Accuracy (%)	F-score (%)	Computation Time (sec)
RFKON	70	AB-DT (96.09)	AB-DT (96.00)	BN (6.86)
KIOS	27	AB-DT (90.68)	AB-DT (91.00)	BN (7.00)
UJI B0	521	BAGG-DT (78.85)	BAGG-NN (79.00)	NN (38.22)
UJI B1	521	BAGG-DT (78.85)	BAGG-NN (79.00)	NN (26.59)
UJI B2	521	BAGG-DT (70.54)	BAGG-NN (71.00)	NN (96.64)
Dataset	Preprocessing			
	Number of Attributes	Accuracy (%)	F-score (%)	Computation Time (sec)
RFKON	42	GAIN AB-DT (96.31)	GAIN AB-DT (96.00)	FILT BN (5.95)
KIOS	24	CHI AB-DT (90.71)	CHI AB-DT (91.00)	FILT BN (5.83)
UJI B0	130	FILT BAGG-DT (78.58)	FILT BAGG-DT (79.00)	FILT NN (12.44)
UJI B1	138	FILT BAGG-DT (78.58)	FILT BAGG-DT (79.00)	FILT NN (12.88)
UJI B2	100	GAIN BAGG-DT (70.66)	GAIN BAGG-DT (71.00)	FILT NN (30.69)

According to Table 3.2, it can be concluded that there is not a single ML algorithm for any indoor positioning system. For example, AB-DT is the best algorithm for RFKON dataset in terms of accuracy whereas BN is the best algorithm in terms of computation time. In order to find a ML algorithm considering multiple user preferences, a multi-criteria decision strategy should be applied.

Multi-criteria algorithm selection strategy

In this subsection, the best ML algorithm for each dataset is determined considering the user preferences. They play a significant role to judge one criterion over another one as mentioned before. In order to show, the multi-criteria decision approach two set of user preferences are given as in Table 3.3.

Table 3.3. *The user preferences for the selected criteria*

Set 1				Set 2			
	Accuracy	F-score	Computation Time		Accuracy	F-score	Computation Time
Accuracy	1	5	9	Accuracy	1	1/2	1/9
F-score	1/5	1	3	F-score	2	1	1/7
Computation Time	1	1/3	1	Computation Time	9	7	1

As mentioned before, lower triangular part should be filled after the user defines the relative importance in the upper triangular. According to user preferences in Set 1, the relative importance of accuracy over f-score and computation time are 5 and 9, respectively, and the relative importance of f-score over computation time is 3. According to these user preferences, the calculated relative weights for each criterion are given in Table 3.4.

Table 3.4. *The relative weights for the selected criteria*

	w_1 (Accuracy)	w_2 (F-score)	w_3 (Computation Time)
Set 1	0.76	0.16	0.08
Set 2	0.08	0.13	0.79

The best algorithm in terms of the selected criteria for the specific dataset is obtained using the multi-criteria decision strategy as seen in Table 3.5.

Table 3.5. *Final performance comparisons of all experiments*

	Accuracy + F-score + Computation Time Set 1	Accuracy + F-score + Computation Time Set 2
RFKON	GAIN AB-DT (96.31%, 96.00%, 153.97sec)	FILT BN (93.56%, 94.00%, 5.95sec)
KIOS	CHI AB-DT (90.71%, 91.00% , 127.59sec)	CFS NN (90.63%, 91.00% , 9.81sec)
UJI B0	FILT BAGG-DT (78.58%, 80.00%, 126.60sec)	FILT NN (70.75%, 71.00%, 12.44sec)
UJI B1	FILT BAGG-DT (78.58%, 79.00%, 141.83sec)	FILT NN (70.75%, 71.00%, 12.88sec)
UJI B2	GAIN BAGG-DT (70.66%, 72.00%, 258.71sec)	FILT NN (67.78%, 68.00%, 30.69sec)

The best ML algorithm for the RFKON dataset considering the user preferences in Set 1, i.e. highest accuracy is GAIN AB-DT. Since the computation time is the dominated criteria in Set 2, and the best ML algorithm for the RFKON dataset is FILT BN. In addition to dominated criterion, the other selected criteria also play a role in order to select the best ML algorithm. For example, for the KIOS dataset, the ML algorithm with the best computation time is FILT BN. The proposed approach finds out CFS NN due to effects of the relative weights of other criteria. The best ML algorithms in Table 3.5 can be changed with respect to user preferences.

3.2. A Hybrid Approach for Indoor Positioning

Performance of IPS can be enhanced by constructing cluster specific classification algorithms. Therefore, a hybrid method that utilizes both clustering and classification algorithms is proposed in this method. After selecting most discriminative APs from the WiFi-RSS based fingerprint map, Expectation Maximization (EM) algorithm is applied to divide the whole area into sub-clusters. Then, decision tree algorithm is utilized to develop a classifier models for each sub-cluster [96].

3.2.1. Preliminaries

IPS can be effected various situations such as using all detectable APs in the fingerprint map and size of the indoor area. As the number of APs and size of the test area are increased, fingerprint map is getting huge. Therefore, the computational complexity of the IPS is also increased. To overcome this situation, various methods are proposed in the literature. Feature selection and clustering techniques are frequently applied methods before positioning among them. In the following subsections, preliminaries are given for feature selection, and clustering. In addition to these, applied positioning algorithm (Decision Tree) is described.

Information Gain Based Feature Selection

Information gain (InfoGain) based feature selection method is the most commonly used feature selection method in the ML that is based on the entropy [97]. Information gain of each feature is calculated using Eq. 3.5.

$$IG(f) = -\sum_{c,\bar{c}} P(c)\log(c) + \sum_{f,\bar{f}} P(c \mid f)\log P(c \mid f) \tag{3.5}$$

where f is the feature (access point for fingerprint map) and c is the class. Information gain based feature selection is applied in the proposed method to determine the most important APs in the fingerprint map.

Expectation Maximization (EM) Clustering Algorithm

Clustering algorithms assign similar data to same cluster without the prior knowledge about the data's characteristics. Since the data's labels' are not known, these algorithms are also called as unsupervised learning algorithms [98]. EM algorithm is a clustering algorithm that assigns data to particular clusters by computing one or more probability distributions. It then maximizes the overall probability of the data belonging to a certain cluster [99]. EM algorithm consists of two steps: determination of expectation and maximization of expectation iteratively. To handle the size of the indoor area problem, EM algorithm is applied in the clustering step to divide the indoor area into sub-areas.

Decision Tree Classifier Algorithm

Decision Tree (DT) predicts an output by tracking the decisions in the tree from the root node down to a leaf node according to the outcome of the tests along the path. DT algorithm is detailed in Section 2.3.2 [44]. In this study, C4.5 that is a benchmark tree is applied in the classification step for performing positioning.

3.2.2. Proposed Method

The proposed method is started with preparing the collected WiFi-RSS measurements from all accessible APs at each RPs in the experimental area for constructing fingerprint map. This can be done by replacing NaN values with the minimum values in the fingerprint map. Then, train and test dataset are reorganized in order to comparable among each other. In the next step, minimum and maximum values for each APs are replaced with the minimum and maximum values in the whole fingerprint map as mentioned in Section 2.2.1. Most important APs are selected using InfoGain based feature selection algorithm. Then, EM algorithm is applied to divide experimental area into optimum number of sub-areas. Finally, DT algorithm is applied for each cluster for positioning purpose. The flowchart of the proposed method is given in Figure 3.15.

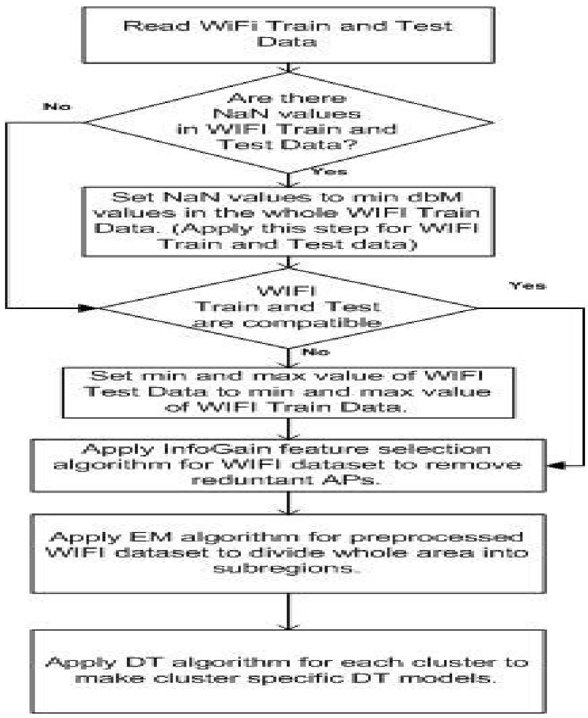


Figure 3.15. *The flowchart of the hybrid approach for indoor positioning*

3.2.3. Experimental Results

RFKON database that is described briefly in Section 2.4.3 is used to show the effectiveness of the proposed method. In experiments, firstly "InfoGain based Feature Selection" algorithm is applied to determine the number of most important APs. The optimum number of APs for RFKON dataset is given in Figure 3.16.

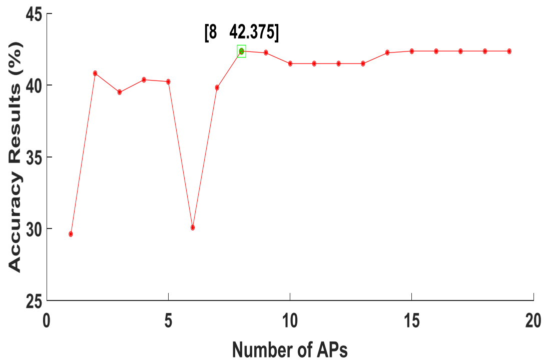


Figure 3.16. *Optimum number of APs determination using InfoGain based feature selection*

There are 24 APs in the database before applying InfoGain based feature selection algorithm. As seen in Figure 3.16, we obtain best accuracy results after selecting 8 APs using InfoGain based feature selection algorithm. This is an important improvement of reducing computational time.

In the clustering phase, EM algorithm is applied. In experiment, different number of clusters are tried to select the best number of clusters. Among the attempted numbers of clusters, five clusters give best accuracy results. Each RP in RFKON database is assigned to a cluster using EM algorithm as seen in Table.3.6.

Table 3.6. *EM Clustering Assignments*

Cluster Name	RP Number
Cluster0	7, 8, 9, 10
Cluster1	19, 20
Cluster2	11, 12, 13
Cluster3	14, 15, 16, 17, 18
Cluster4	1, 2, 3, 4, 5, 6

After applying EM clustering algorithm, the whole area is divided into sub-clusters. Then, DT algorithm is utilized to form cluster-specific classifier models for each cluster. The accuracy and computation time of the positioning are enhanced using this hybrid algorithm. The comparison of the applied hybrid algorithm with DT algorithm is given in Table 3.7.

Table 3.7. *Comparison of accuracy results*

Algorithm	Number of APs	Accuracy Results (%)
DT	19	42.25
DT with InfoGain	8	42.37
Proposed Method	8	66.42

As seen in Table 3.7, the accuracy of DT algorithm without preprocessing is 42.25%. When we apply InfoGain-based feature selection before DT classifier; the number of APs is reduced to 8, and the accuracy is 42.375%. As a result of Table 3.7, the hybrid algorithm enhances the accuracy about %25 using 8 APs. To demonstrate the applied hybrid algorithm improvement on the decision tree size, Table 3.8 is constructed.

Table 3.8. *Decision tree size*

Train database	Number of leaves	Size of the tree
19 APs	34	67
8 APs	35	69
Cluster0	3	5
Cluster1	8	15
Cluster2	6	11
Cluster3	2	3
Cluster4	5	9

As seen in Table 3.8, reducing the number of APs using InfoGain algorithm does not make an improvement on the tree size solely. But, after utilizing EM clustering algorithm, the size of the decision tree is diminished. This causes lower computational time in the classification step in addition to improvement on the accuracy results as seen in Table 3.8.

3.3. A Hybrid Fingerprint Based Indoor Positioning with Extreme Learning Machine

WiFi-based indoor positioning is preferred frequently by the researchers due to massive deployment in indoor area, and wide usage of WiFi enabled devices. But, WiFi based indoor positioning methods have some drawbacks such as WiFi signals deteriorate over time which lead to inaccurate position estimates. Therefore, the accuracy of only WiFi signals may not be adequate for some applications and can be enhanced using other sensor measurements such as MF. In the proposed hybrid fingerprint based indoor positioning with extreme learning machine method, the method which is given Section 3.2 is enhanced by considering two sensor types such as WiFi-RSS and MF measurements concurrently. This can be performed by constructing hybrid database which contains WiFi-RSS and MF sensor data. The accuracy of IPS is improved by taking advantages of these sensor types. Besides, significant improvements are acquired in terms of computation time using 'ReliefF' feature selection and 'k-means' clustering algorithms [100].

3.3.1. Preliminaries

In the following subsections, the methods and the algorithms that are used in the proposed hybrid fingerprint based indoor positioning with extreme learning machine method are explained.

Constructing Hybrid Fingerprint Map

In the literature, fingerprint-based positioning is generally started with collecting one type of sensor measurements. But, in recent years hybrid fingerprint maps are generated to take the advantage of more than one sensor measurement at the same type. Therefore, in this study a hybrid fingerprint map is constructed which contains WiFi-RSS and MF values. An instance in the hybrid fingerprint map is given in Equation 3.6.

$$[RP_i, x_i, y_i, RSS_{i,1}, RSS_{i,2}, ..., RSS_{i,19}, global_{i,x}, global_{i,y}, global_{i,z}] \tag{3.6}$$

where RP_i is the i^{th} RP label, x_i, y_i are the x and y coordinates of i^{th} RP, $RSS_{i,j}$ is the signal strength measurement obtained from j^{th} AP, and $global_{i,x}, global_{i,y}, global_{i,z}$ are the MF global x, y, and z values at the i^{th} RP.

ReliefF Feature Selection Algorithm

Feature selection algorithm is applied to reduce the dimension of the hybrid fingerprint map. ReliefF is applied in the proposed method before positioning to remove the redundant APs in the hybrid fingerprint map. ReliefF is the extended version of statistical Relief method from two-class problems to multi-class problems [101]. The feature selection process is done by constructing model with choosing an instance closeness of the instances in the same class, and distantness of the instances in different classes.

$$\mathrm{Re}\,lief_i = \frac{\sum_{j=1}^{m} -diff(x_{ij}, nearestins\tan ce_{sameclass_{i,j}}) + diff(x_{ij}, nearestins\tan ce_{differentclass_{i,j}})}{m} \quad (3.6)$$

where m is the size of randomly selected subset of the fingerprint map, $x_{i,j}$ is the j^{th} value of i^{th} instance, $diff(x_{ij}, nearestins\tan ce_{sameclass_{i,j}})$ is the difference between the $x_{i,j}$ and the nearest instance in same class, and $diff(x_{ij}, nearestins\tan ce_{differentclass_{i,j}})$ is the difference between the $x_{i,j}$ and the nearest instance in different class. It is expected that $x_{i,j}$ and $nearestins\tan ce_{sameclass_{i,j}}$ are very close each other, and $x_{i,j}$ and $nearestins\tan ce_{differentclass_{i,j}}$ are far away from each other for discriminative attributes.

K-means Clustering Algorithm

K-means is the basic clustering algorithm which divides the fingerprint map into K subareas where K represents the number of clusters [102]. The division is based on the highest similarity between the instances in the intra cluster, and the lowest cluster between the inter cluster. Clustering is done by minimizing the sum of square distances between the instances and the relevant cluster centroid. This algorithm is performed as follows: Select randomly centroid of K clusters initially. Then, the instances in the fingerprint map are assigned to nearest cluster. The centroid of each cluster is recalculated using the mean of all the instances in each cluster. This recalculation repeats until the centroid of each cluster is not change or defined number of iteration is reached.

3.3.2. Proposed Method

The aim of the proposed hybrid fingerprint based indoor positioning with ELM is to use advantages of WiFi and MF sensor measurements at the same time. For this purpose, hybrid fingerprint map is constructed. Then, the IPS performance is enhanced by removing attributes using ReliefF and dividing the whole dataset to subareas using

K-means clustering. Finally, positioning is performed using cluster-specific ELM algorithm. The flowchart of the proposed method is given in Fig. 3.17.

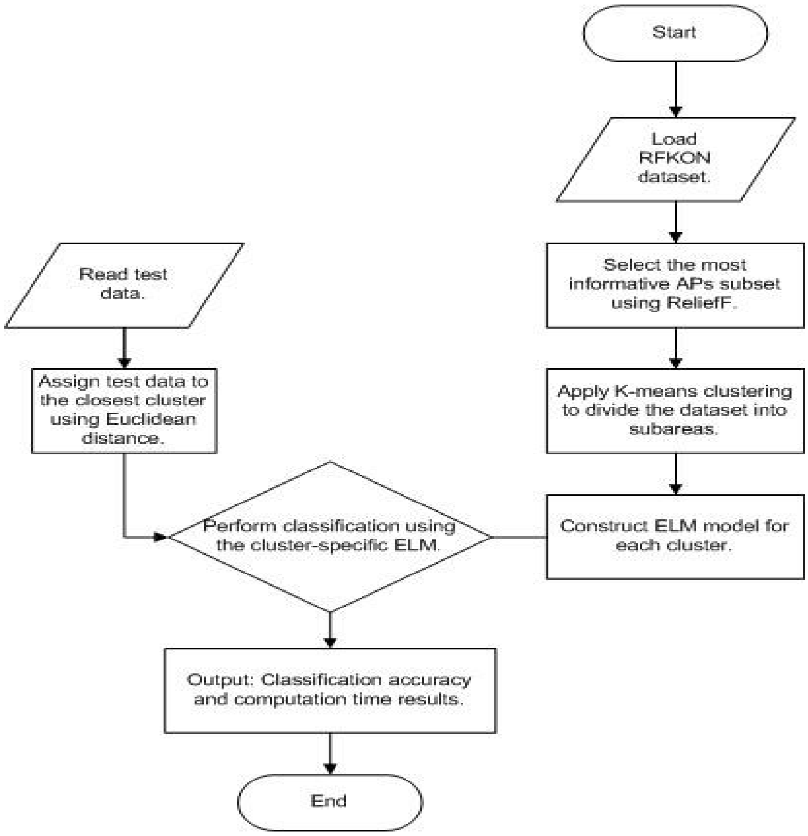


Figure 3.17. *Flowchart of the hybrid fingerprint based indoor positioning with ELM*

3.3.3. Experimental Results

In experiments, RFKON dataset is utilized. The outputs of the each step of the proposed method are as follows: Firstly, ReliefF is applied to determine the number of most informative APs in the RFKON dataset. After applying ReliefF, the number of APs in the RFKON dataset is decreased from 24 to 19. Then, k-means algorithm is applied to assign each RP to a cluster. Table 3.9 gives these assignments.

Table 3.9. *K-means Clustering Results*

Cluster Name	Reference Points
$Cluster_1$	1, 2, 3, 4
$Cluster_2$	5, 6, 7, 8, 9, 10
$Cluster_3$	11, 12, 13, 14, 15, 16, 17, 18, 19, 20

The number of clusters in K-means algorithm is chosen experimentally. $K = 3$ gives best accuracy among the attempted cluster numbers in the experiments. The experiments are started with comparing ELM algorithm which uses only one type of sensor measurement and hybrid fingerprint map. The comparison results are given in Table 3.10.

Table 3.10. *The accuracy and the computational time results of the proposed method*

Sensor Type	Accuracy Results (%)	Computational Time (ms)
WiFi	36	29.01
MF	29	20.16
WiFi + MF	55	20.62

As seen in Table 3.10, the accuracy results of positioning using ELM with WiFi or MF data are not adequate, and they are enhanced when using WiFi and MF sensor data at the same time. In addition to this, the computational time result of ELM algorithm using WiFi and MF data in a hybrid fingerprint nearly same as other types of positioning. Then, the results of the proposed hybrid fingerprint based indoor positioning with ELM is given in Table 3.11.

Table 3.11. *Experimental results of the proposed method*

		Sensor Type		
		WiFi	MF	WiFi + MF
$Cluster_1$	Accuracy Results (%)	72	96	92
	Computational Time (ms)	0.57	0.53	0.56
$Cluster_2$	Accuracy Results (%)	56	82	90
	Computational Time (ms)	1.19	1.08	1.89
$Cluster_3$	Accuracy Results (%)	53	24	66
	Computational Time (ms)	4.72	4.32	4.35

As seen in Table 3.11, the proposed method improves the accuracy and the computation time of the IPS significantly. According to the inference from the studies in the literature, WiFi based fingerprint positioning success is better in large-scale indoor area, whereas MF success is better in small-scale regions due to its nature. This inference is verified when comparing the experimental results in Table 3.10 and 3.11. Before clustering, accuracy result of positioning using WiFi data is better than MF data. But, the accuracy of positioning with MF data is better after clustering process. Besides, the proposed method enhances the accuracy results using the advantages of both sensor types. Also, dividing whole datasets into sub-areas using clustering reduces the computational time. This is an important contribution when considering the battery requirements of mobile devices.

3.4. Integration of Classification Algorithms for Indoor Positioning System

IPS performance depends on various criteria. Selection of appropriate algorithm and sensor measurement type for positioning are critical issues. Various algorithms with different sensor types are applied in the literature. The aim of the proposed method is to integrate more than one sensor type and positioning algorithm concurrently to enhance the IPS performance in terms of accuracy. In the proposed method WiFi and MF data are used for positioning with DT, MLP, and BN, simultaneously. The selected algorithms are integrated using majority voting method.

3.4.1. Preliminaries

Fingerprint-based positioning method is adopted in this study. In the following subsections, the phases of this method are explained briefly.

Training Phase

Training phase is the first phase of fingerprint-based positioning method. In this phase, WiFi RSS values and MF sensor values are combined in order to form a hybrid fingerprint for the fingerprint map. The i^{th} instance of the fingerprint map is given in Eq. (3.7).

$$FP_i = (RP_i, x_i, y_i, \{(MAC_{i,1}, RSS_{i,1}), ..., (MAC_{i,k}, RSS_{i,k})\}, \{global_{i,x}, global_{i,y}, global_{i,z}\}) \quad (3.7)$$

where RP_i is the i^{th} reference point (RP) label, x_i, and y_i are the x and y coordinates, $MAC_{i,j}$ and $RSS_{i,j}$ are the MAC address and RSS values of the j^{th} AP, k is the number of APs, $global_{i,x}, global_{i,y}$, and $global_{i,z}$ are the MF strength values at the RP_i.

Positioning Phase

In the positioning phase, the measurement of the mobile device is compared with the fingerprints in the database. Different classifier algorithms such as DT, MLF, and BN are integrated using majority voting method in this phase to estimate the position.

3.4.2. Proposed Method

Indoor positioning algorithm which is proposed in this study is started with constructing hybrid fingerprint database. The database integrates WiFi RSS and MF measurements to construct a fingerprint for a RP. After choosing any number of base classifiers; majority voting method is utilized to integrate the base classifiers. Majority voting method combines the outputs of the classification algorithms to estimate the final position. The final position is the one that is mostly predicted by the base classifiers. In this study, DT, MLP, and BN are selected as base classifiers [103]. Fig. 3.18 shows the flowchart of the proposed hybrid indoor positioning algorithm.

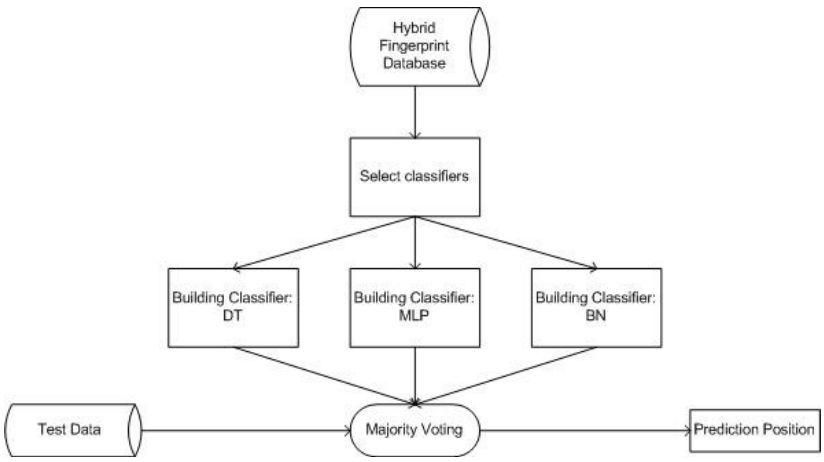


Figure 3.18. *The flowchart of the proposed hybrid indoor positioning algorithm*

As seen in Fig. 3.18, three classification algorithms such as DT, MLP, and BN are integrated using majority voting method instead of using any single classifier. Majority voting method is one of the widely used methods to combine classifiers when the base classifiers give class labels as outputs. It improves the performance of the classifiers through voting. One of the important advantages of majority voting method is that there is no need to adjust parameter if the base classifiers have been trained. After integrating different classifiers, the majority voting method constructs a classifier which is superior to other classifiers.

3.4.3. Experimental Results

In this study, DT, MLP, and BN classifiers are selected as the base classifiers. All parameters of the DT and BN classifiers are selected as default values in WEKA. The "training time", "momentum" ,"learning rate" parameters of MLP change from "500"to

“100”, “0.2” to “0.4”, and “0.3” to “0.1” for faster the execution of classifier. The average distance error of each algorithm is given in Table 3.12.

Table 3.12. *Average Distance Error of Algorithms (m)*

	WiFi Database	MF Database	Hybrid (WiFi + MF) Database
DT	4.21 m	4.78 m	4.21 m
MLP	2.48 m	3.54 m	1.40 m
BN	4.22 m	4.85 m	1.06 m
Hybrid Algorithm (Majority Voting)	2.15 m	4.69 m	1.23 m

As seen in the Table 3.12, MLP and BN algorithms performance are improved using more than single type of measurement. In addition to this improvement, the system whole performance is also enhanced by taking multiple sensor measurements for the fingerprint map and multiple classifier approach with “majority voting method”.

3.5. An F-score Weighted Indoor Positioning Algorithm Integrating WiFi with Magnetic Field Fingerprints

In this method, an f-score weighted indoor positioning algorithm is proposed. The proposed algorithm integrates WiFi and MF sensor measurements to consider advantages of both sensor types simultaneously.

3.5.1. Preliminaries

The algorithms used for positioning are given in the following subsections.

Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation (MLE) is one of the most popular algorithms which take into account the standard deviation of the measurements [84]. It gives higher accuracy when compared with the other algorithms [85]. Additionally, the calculated likelihood values for different sensor types are useful for constructing hybrid solutions for the indoor positioning problem. In MLE algorithm, the probability of obtaining the mobile device fingerprint $F^{'}$ at the i^{th} RP whose fingerprint F_i is given by

$$p(F' \mid F_i) = \frac{1}{\left(\prod_{i=1}^{n} s_i\right)(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i' - \overline{x_i}}{s_i}\right)^2}$$

$$F_i = \left(\overline{x_i}, s_i\right) \tag{3.8}$$

$$F' = (x_1', x_2', \ldots, x_n')$$

where $\overline{x_i}$ and s_i are the mean and the standard deviation of the signal measurements at the i^{th} RP, x_i' is the signal measurement at an unknown location, and n is the dimension of the fingerprint map. The mean and the standard deviation are calculated by

$$\overline{x_i} = \frac{1}{m}\sum_{j=1}^{m} x_i^j, s_i = \sqrt{\frac{1}{m-1}\sum_{j=1}^{m}\left(x_i^j - \overline{x_i}\right)^2} \tag{3.9}$$

where m is the number of collected measurements for the RP_i, x_i^j is the j^{th} measurement from AP_i for the radio map. For the magnetic map, x_i^j is magnetic field strength value.

The computational cost of Eq. (3.8) can be reduced by taking its natural log

$$-\ln\left(\prod_{i=1}^{n} s_i\right) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i' - \overline{x_i}}{s_i}\right)^2 \tag{3.10}$$

Since the second term of Eq. (3.10) is constant, it can be ignored. Eq. (3.10) is rewritten by multiplying -1 and defining $c = \left(\prod_{i=1}^{n} s_i\right)$ as

$$g(x_i') = c + \frac{1}{2}\sum_{i=1}^{n}\left(\frac{x_i' - \overline{x_i}}{s_i}\right)^2 \tag{3.11}$$

Now, the Eq. (3.8) is converted to Eq. (3.11) [104]. For the mobile device fingerprint F' Eq. (3.11) is calculated for each RP, and then the label of the RP is returned by calculating Eq. (3.12).

$$\arg\min_i g(x_i') \tag{3.12}$$

The wellness of the MLE model can be evaluated using a separated test data. Model evaluation is done as follows: after calculating the estimated position of each test data using Eq. (3.12), confusion matrix is generated with the estimated position labels and actual position labels. The confusion matrix is a basis for calculating the terms such as true positive (TP), false negative (FN), true negative (TN), and false positive (FP) [105]. The binary case representation of the confusion matrix is given in Table 3.13.

Table 3.13. *Confusion matrix binary case representation*

	Positive (Estimated)	Negative (Estimated)
Positive (Actual)	TP	FN
Negative (Actual)	FP	TN

Then, f-score which is one of the performance metrics can be calculated using the confusion matrix as follows:

$$f - score = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3.13}$$

In the literature, the f-score values are generally used to evaluate the train model performance using the test data. In this study, the f-score values are used to evaluate the model (Eq. (3.12)) performance for each signal type. Besides, these f-score values are used as the weights of each signal type in the positioning phase to enhance the IPS performance.

3.5.2. Proposed Method

Traditional fingerprint-based positioning algorithms are started with constructing database by collecting measurements from the experimental area. The measurements generally contain only WiFi-RSS values obtained from the APs in the region. RFKON database contains both WiFi-RSS and MF strength values for each RP. The radio map and the magnetic map are utilized as inputs for the proposed f-score weighted indoor positioning algorithm. The pseudo code of the algorithm is given in Algorithm 1.

Algorithm 1. The pseudo code of the f-score weighted indoor positioning algorithm

Training-Testing Phase:

Inputs: Radio Map, Magnetic Map.

Outputs: WiFi-RSS-based model (MLE_{WiFi}), MF-based model (MLE_{MF})

1) Normalize each instance in radio map and magnetic map using min-max normalization procedure.
2) Split the radio map and the magnetic map as train data (60%) and test data (40%).
3) Use Equation (3.9) to obtain $\overline{x_i}$, and s_i for the train data of each signal type.
4) Calculate the RP labels for both test data type using Eq. (3.12) separately.
5) Apply Equation (3.13) to calculate the f-score values of each signal type per RP using the calculated RP labels, and the actual RP labels. The f-score values are stored as weight of each sensor type as $weight_{WiFi,i}$, and $weight_{MF,i}$.
6) Construct WiFi-RSS-based model (MLE_{WiFi}), MF-based model (MLE_{MF}) as follows:

$$MLE_{WiFi,i} = (lb_i, xCoord_i, yCoord_i, \overline{x}_i, s_i, weight_{WiFi,i}) \tag{3.14}$$

$$MLE_{MF,i} = (lb_i, xCoord_i, yCoord_i, \overline{x}_i, s_i, weight_{MF,i}) \tag{3.15}$$

Positioning Phase:

Inputs: MLE_{WiFi}, MLE_{MF}, WiFi New Test Data, MF New Test Data.

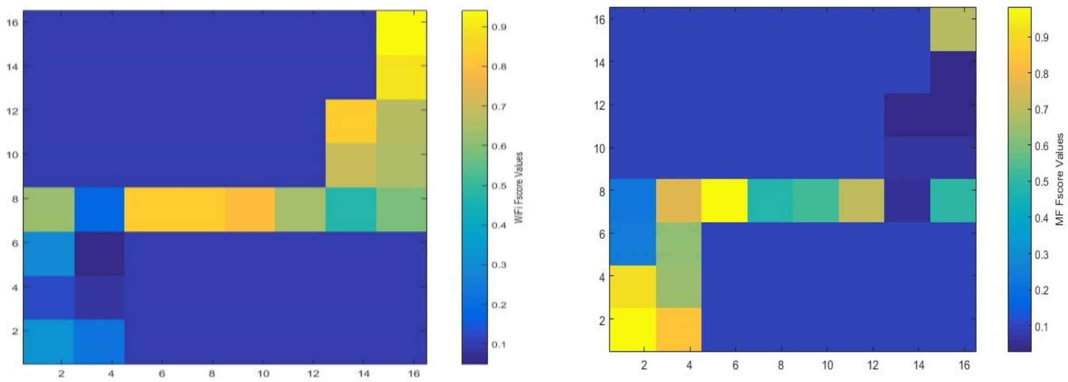
Outputs: Estimated position.

1) Apply MLE_{WiFi} with Equation (3.11) using WiFi New Test Data to obtain likelihood values of each RP.
2) Apply MLE_{MF} with Equation. (3.11) using MF New Test Data to obtain likelihood values of each RP.
3) Normalize likelihood values using max-min normalization method.
4) Use Equation (3.16) to calculate final position.

$$\arg\min_i \; g(x'_{wiFi,i})' \times weight_{WiFi,i} + g(x'_{MF,i})' \times weight_{MF,i} \tag{3.16}$$

3.5.3. Experimental Results

WiFi-RSS values or MF values may provide better positioning accuracy depend on the structure of indoor area. If there are different fingerprint maps for each signal type, the f-score values can be used to understand the quality of accuracy for each map. In our experimental area, f-score values are obtained for radio map and magnetic map as in Figure 3.19. Heat maps show the quality of positioning in the given RPs.



(a) (b)

Figure 3.19. *Heat map of f-score values for WiFi (a) and MF (b) data using MLE algorithm*

As seen in the Figure 3.19.a, the f-score values are high in the right-above sub area which means that WiFi-based positioning has good accuracy. But, the f-score values are worse in the left-below area. On the other hand, f-score values have higher values for this region in Figure 3.19.b. It means that MF-based positioning has good accuracy in that region. The f-score values are almost the same in the middle of the figures. It means that the accuracy of each method is nearly same. Therefore, applying each sensor type with f-score weight values concurrently in the proposed positioning algorithm can be enhanced the performance of the IPS.

The proposed f-score weighted indoor positioning algorithm is compared with KNN that is used in RADAR [15], and NB that is used in Horus [16]. The experiments are performed by comparing KNN, NB, and MLE algorithms using only one sensor measurement. Therefore, firstly, proposed algorithm is compared with KNN, NB, and MLE algorithms with WiFi-RSS data. Then, the comparison is done when applying KNN, NB, and MLE algorithms with MF data. Figure 3.20 depicts the distribution of localization error for compared algorithms.

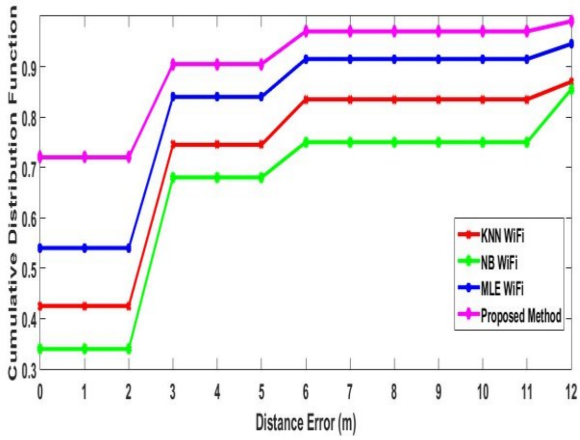


Figure 3.20. *Performance comparison of f-score weighted indoor positioning algorithm with KNN, NB, and MLE that uses only WiFi-RSS data*

As seen in Figure 3.20, the cumulative distribution function for the proposed algorithm is obviously superior to other algorithms when WiFi-RSS sensor measurements are utilized.

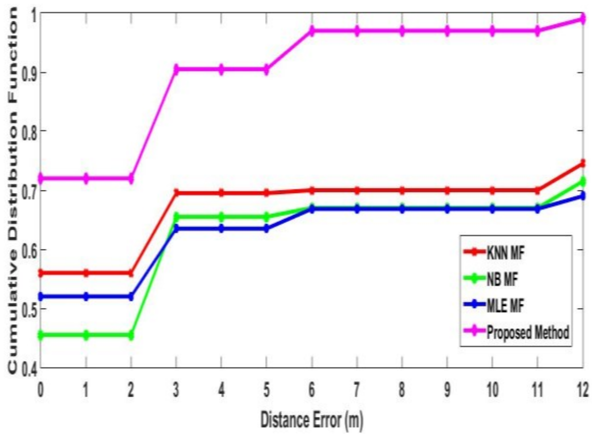


Figure 3.21. *Performance comparison of f-score weighted indoor positioning algorithm with KNN, NB, and MLE that uses only MF data*

In Figure 3.21, the precision results of proposed method are better than MLE, KNN, and NB when MF strength measurements are used. According to Figure 3.20 and Figure 3.21, it is deduced that the proposed algorithm outperforms other algorithms whatever which type of sensor measurement is used. The positioning precisions of the applied algorithms are given in details in Table 3.14.

Table 3.14. *Detailed experimental results of the positioning algorithms*

Method	Precision (%)	
	<3m	<6m
Proposed Algorithm	91%	97%
MLE WiFi	84%	92%
KNN WiFi	75%	84%
NB WiFi	64%	75%
MLE MF	64%	67%
KNN MF	69%	70%
NB MF	66%	67%

In Table 3.14, the proposed algorithm results positioning error less than 3m for 91% of test data, while MLE WiFi, KNN WiFi, NB WiFi, and MLE MF, KNN MF, NB MF are 84%, 75%, 64%, and 64%, 69%, and 66%, respectively. The proposed algorithm also results positioning error less than 6-m for 97% of test data while MLE WiFi, KNN WiFi, NB WiFi, and MLE MF, KNN MF, NB MF are 92%, 84%, 75%, and 67%, 70%, and 67% respectively. The proposed algorithm can effectively integrate the MF and WiFi signals for more accurate positioning.

The proposed algorithm is also compared with MLE, KNN, and NB which use hybrid fingerprint map for positioning. These comparison results are given in Figure 3.22.

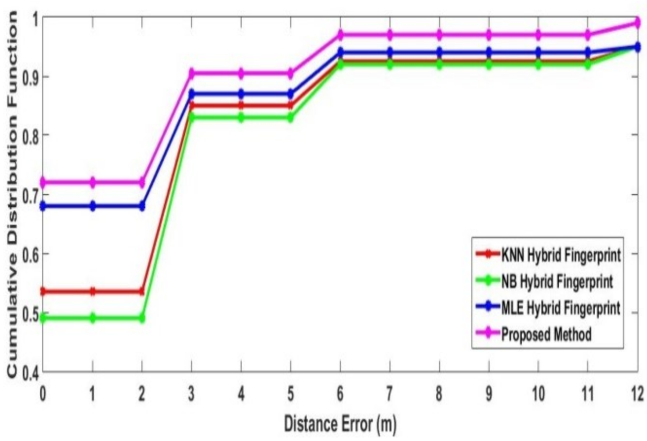


Figure 3.22. *Performance comparison of f-score weighted indoor positioning algorithm with KNN, NB, and MLE that uses hybrid fingerprint data*

As seen in Figure 3.22, all algorithms are enhanced when using the hybrid fingerprint map, but again the proposed algorithm is the best when compared with other algorithms. The detailed results of this comparison are given in Table 3.15.

Table 3.15. *Detailed experimental results of the positioning algorithms*

Method	Precision (%)	
	<3m	<6m
Proposed Algorithm	91%	97%
MLE Hybrid Fingerprint	87%	94%
KNN Hybrid Fingerprint	85%	93%
NB Hybrid Fingerprint	83%	92%

According to Table 3.15, MLE, KNN, and NB result positioning error less than 3m are enhanced to 87%, 85%, and 83%, respectively. And, also MLE, KNN, and NB result positioning error less than 6m are improved to 94%, 93%, and 92%, respectively. However, the situation which the proposed algorithm results with 91% and 97% for positioning error less than 3m and 6m respectively are superior to other algorithms still remains.

3.6. An Enhanced Approach of Indoor Positioning Algorithm using WiFi Received Signal Strength and Magnetic Field Fingerprints

In this method, a precise indoor positioning algorithm is proposed. The proposed method uses WiFi RSS and magnetic field measurements at the same to construct a fingerprint map. Then, positioning is performed to obtain the estimated position in terms of x and y coordinates.

3.6.1. Preliminaries

The algorithms used in the proposed method are given in the following subsections.

WiFi RSS and Magnetic Field based Fingerprint Database

The fingerprint database contains WiFi RSS and magnetic field signal measurements. The measurements are obtained at each RP in the experimental area. All measurements at each RP is averaged to eliminate the noisy data. The representation of the fingerprint map is illustrated in Fig. 3.23.

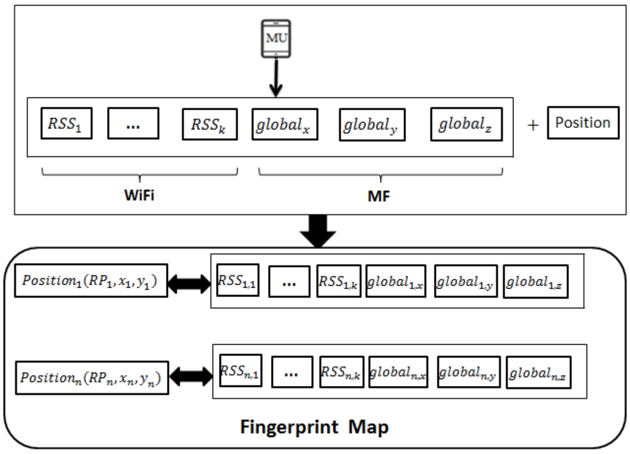


Figure 3.23. *The representation of the fingerprint map*

3.6.2. Proposed Method

After constructing fingerprint map, the positioning is performed by applying the proposed positioning algorithm. The proposed positioning algorithm is illustrated in Fig. 3.24.

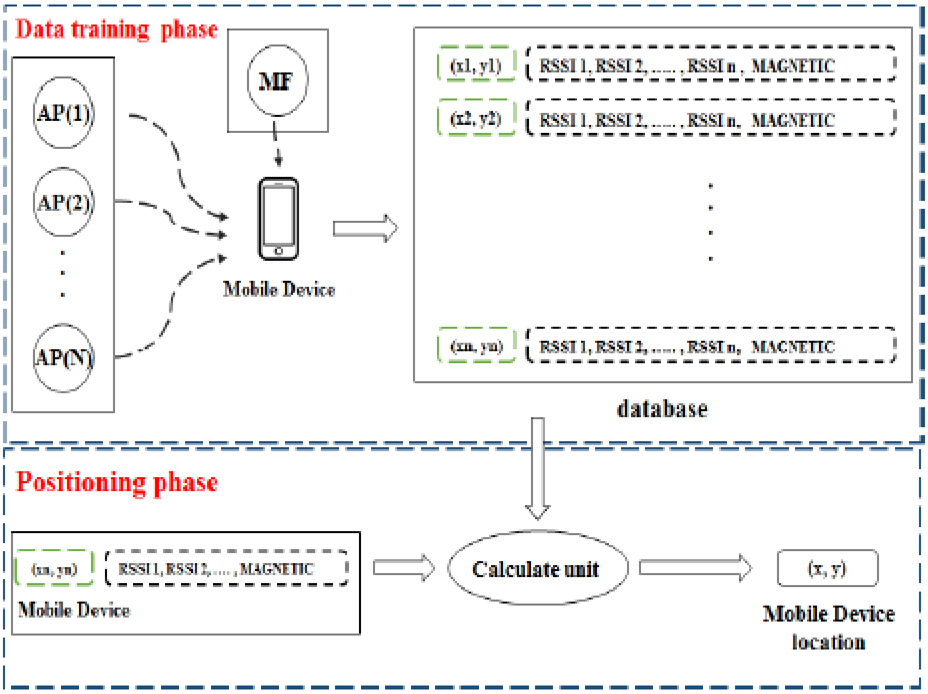


Figure 3.24. *The proposed method*

The proposed positioning algorithm is performed into two steps. In the first step, fingerprint map is constructed using average values of WiFi RSS and magnetic field measurements. Then, positioning is started by comparing test data with each instance in the fingerprint map in the calculate unit. Then, K closest points are selected using distance metric. In this step, the K and the distance metric are important parameters that change the estimated error. Several distance metrics are applied in the proposed algorithm, but spearman's rank order correlation gives the best result among them. Spearman's rank order correlation is calculates as follows: Rank the two vectors which are used in the comparison. The number '1' is given to the biggest number in a vector, '2' to the second biggest value and so on. The smallest value in the vector will get the lowest ranking. The rank of any equal values is calculated by averaging of the rank values. This should be done for both sets of measurements. After calculating the rank values, then following equation is used to calculate the distance between two vectors:

$$dist_{trainData_i,testData} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n\left(n^2 - 1\right)} \tag{3.17}$$

where d_i is the difference between the ranks of the two values on each vector, i.e. the rank of the second value is subtracted from the rank of the first value, and n is the number of instances in the column vector. Finally, K closest points are integrated with a weight to calculate the mobile unit position. The proposed algorithm is explained briefly in Algorithm 2.

Algorithm 2. The pseudo code of the enhanced approach of indoor positioning algorithm using WiFi RSS and MF fingerprints

Training Phase: **Inputs:** WiFi RSS and magnetic field measurements. **Outputs:** Fingerprint map 1) Normalize each instance in radio map and magnetic map using min-max procedure. 2) Get the average values of each measurements to construct fingerprint map.
Positioning Phase: **Inputs:** $trainData, testData, K, r.$ **Outputs:** Estimated position. 1) Find the K closest points by comparing each instance in the fingerprint map with the test data using spearman's rank-order correlation as given in Eq. (3.17). 2) Each point in the closest list must satisfy the following condition: $$dist_{trainData_i,testData} < r, i = 1,...,K$$ 3) Weight of each point is calculated as follows: $$w_j = \frac{1/dist_{trainData_j,testData}}{\sum_{i=1}^{K} 1/dist_{trainData_i,testData}}$$ 4) The estimated position is calculated as follows: $$(x, y) = \sum_{j=1}^{K} w_j (x_j, y_j)$$

3.6.3. Experimental Results

The experimental area is given in Fig. 3.25 where red points represents the RPs for training phase, and the blue points represents the RPs for positioning phase.

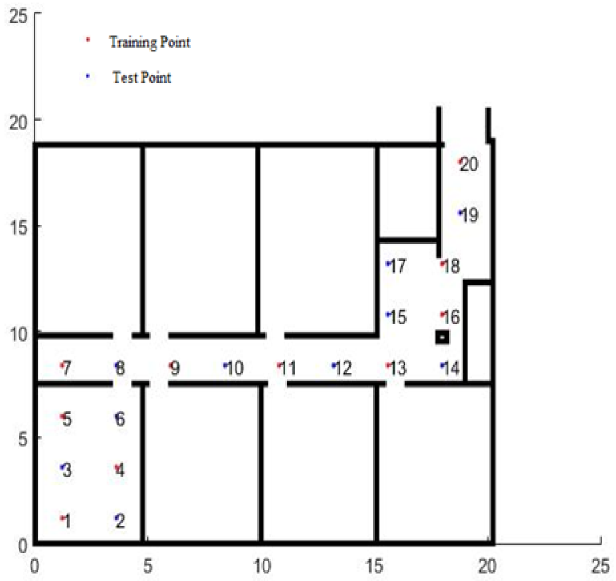


Figure 3.25. *Experimental area*

As seen in the Fig. 3.25, the numbers 1, 4, 5, 7, 9, 11, 16, 13, 18, and 20 represent the training points. And, the remaining ones are the test points. The training phase is performed by collecting measurements from the training points. Then, positioning phase is started with collecting test data from the test points which are labelled with the numbers 2, 3, 6, 8, 10, 12, 14, 15, 17, and 19. In the positioning phase the parameters K and $,r$ are setted to 3 and 0.60, respectively. The parameter r is used to restrict the search area. After, performing positioning phase the Fig. 3.26 is constructed to show the estimated position of the test points

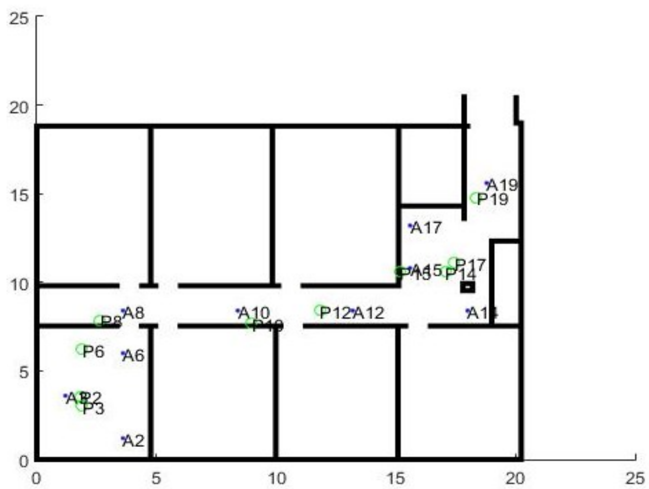


Figure 3.26. *Experimental results*

In the Fig. 3.26, the actual coordinates and estimated coordinates of each test point are illustrated with blue point and green small circles, respectively. The letter ‘A’ represents the actual point, and the letter ‘P’ represents the predicted point in the figure. Finally, the minimum, average, and maximum distance error are obtained as 0.47, 1.56 and 2.96, respectively.

4. CONCLUSIONS and FUTURE WORKS

The main problem in IPS is to obtain a reasonable positioning accuracy in a cost-effective manner. Therefore, the main objective of this study is to construct a cost-effective radio frequency-based indoor positioning system which adopts fingerprinting method. Our objective is to enhance the indoor positioning system performance. For this purpose, various positioning algorithms are utilized with the publicly accessible indoor positioning datasets to determine the most appropriate algorithms in terms of selected performance metrics. A multi-criteria optimization technique is defined to obtain the most appropriate algorithm for a given dataset. Since, WiFi Received Signal Strength based fingerprinting method suffers from multipath effect which leads to erroneous position estimate, it can be enhanced using supplementary technologies such as magnetic field. Magnetic field has some advantages such as it does not suffer from NLOS conditions or multipath effects in indoors whereas it has short operating range, and sensitivity to certain materials. However, magnetic field strength diminishes rapidly with distance. Therefore, we handle magnetic field-based fingerprinting method as a supplementary solution with the WiFi Received Signal Strength based fingerprinting method. So, hybrid fingerprints are defined to improve the positioning performance. Then, several positioning algorithms are applied with hybrid fingerprint dataset to solve the indoor positioning problem. Then, we propose an f-score weighted indoor positioning algorithm integrating WiFi Received Signal Strength fingerprints with magnetic field fingerprints to enhance indoor positioning system performance in terms of accuracy. The proposed f-score weighted indoor positioning algorithm has better accuracy performance than the conventional algorithms. Thus far, these algorithms solve indoor positioning problem as a classification task. Since more precise position estimates are more preferred, and then we propose an enhanced indoor positioning algorithm using WiFi received signal strength and magnetic field fingerprints. This final method calculates the position in terms of x and y coordinates to obtain more precise location.

In future works, we plan to develop a robust indoor positioning system that aggregates various indoor positioning systems focused on satisfying system requirements such as latency, accuracy, and so on.

REFERENCES

[1] G. M. Djuknic and R. E. Richton, "Geolocation and assisted gps," *Computer* 34, 2, 123-125, 2001.

[2] A. Varshavsky, E. Lara, J. Hightower, A. LaMarca and V. Otsason, "Gsm indoor localization," *Pervasive Mob. Comput*, 3, 6 698-720, 2007.

[3] B. Kyoungsoo and J. Yoo, "RFID Based Indoor Positioning System Using Event Filtering," *Journal of Electrical Engineering & Technology*, 12, 335-345, 2017.

[4] U. Yayan, H. Yucel and A. Yazıcı, "A Low Cost Ultrasonic Based Positioning System for the Indoor Navigation of Mobile Robots," *Journal of Intelligent & Robotic Systems*, 78, 2015.

[5] R. Faragher and R. Harle, "Location Fingerprinting With Bluetooth Low Energy Beacons," *IEEE Journal on Selected Areas in Communications*, 33, 2418-2428, 2015.

[6] T. Roos, P. Myllymäki, H. Tirri, P. Misikangas and J. Sievänen, "A probabilistic approach to WLAN user location estimation," *Int. J. Wireless Inf. Networks*, 9, 155–164, 2002.

[7] J. Haverinen and A. Kemppainen, "Global indoor self-localization based on the ambient magnetic field," *Robotics and Autonomous Systems*, 57, 1028–1035, 2009.

[8] H. Liu, H. Darabi, P. Banerjee and J. Liu, "Survey of wireless indoor positioning techniques and systems", *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, 37, 6, 1067–1080, 2007.

[9] M. Heidari and K. Pahlavan, "A new statistical model for the behavior of ranging errors in TOA-based indoor localization," *IEEE Wireless Communications and Networking Conference (WCNC '07)*, 2564–2569, 2007.

[10] M. Bocquet, C. Loyez, and A. Benlarbi-Dela¨ı, "Using enhanced-TDOA measurement for indoor positioning," *IEEE Microwave and Wireless Components Letters*, 15, 10, 612–614, 2005.

[11] C.-H. Lim, B. P. Ng, and D. Da, "Robust methods for AOA geolocation in a real-time indoor WiFi system," *Journal of Location Based Services*, 2, 2, 112–121, 2008.

[12] C. L. Mak, M. A. Hon, W. M. Lau and W. M. Cheung, "Refined Wi-Fi fingerprinting with tag-less proximity-based positioning technique," *2014 International Conference on Indoor Positioning and Indoor Navigation*, 758-761, 2014.

[13] L. Li and X. Lin, “Apply pedestrian dead reckoning to indoor Wi-Fi positioning based on fingerprinting,” *IEEE International Conference on Communication Technology*, 206–210, 2013.

[14] D. Focken and R. Stiefelhagen, ”Towards vision-based 3-D people tracking in a smart room”, *4th IEEE Intl Conference on Multimodal Interfaces*, 2002.

[15] P. Bahl, and V. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," *IEEE INFOCOM'2000*, 2000.

[16] M. Youssef and A. Agrawala, "The Horus WLAN location determination system," *3rd international conference on Mobile systems, applications, and services*, 205-218, 2005.

[17] H. Xie, T. Gu, X. Tao, H. Ye, and J. Lv, “MaLoc: A practical magnetic fingerprinting approach to indoor localization using smartphones,” *ACM UbiComp*, 243–253, 2014.

[18] S. He and S. H. G. Chan, "Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons," *IEEE Communications Surveys & Tutorials*, 18, 466-490, 2016.

[19] C. Laoudias, R. Pich´e and C. G. Panayiotou, "Device self calibration in location systems using signal strength histograms,” *Journal of Location Based Services*, 7, 165-181, 2013.

[20] J. Torres-Sospedra, R.l Montoliu, A. Martínez-Usó, T. J. Arnau, Joan P. Avariento, M. Benedito-Bordonau and J. Huerta, “UJIIndoorLoc: A New Multi-building and Multi-floor Database for WLAN Fingerprint-based Indoor Localization Problems,” *Fifth International Conference on Indoor Positioning and Indoor Navigation*, 261-270, 2014.

[21] J. Chung, M. Donahoe, C. Schmandt, I.-J. Kim, P. Razavai and M. Wiseman, “Indoor location sensing using geo-magnetism,” *9th international conference on Mobile systems, applications, and services. ACM*, 141–154, 2011.

[22] J. Torres-Sospedra, D. Rambla, R. Montoliu, O. Belmonte and J. Huerta, “Ujiindoorloc-mag: A new database for magnetic field based localization problems,” *IEEE 2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 1–10, 2015.

[23] S. Bozkurt Keser, U. Yayan, A. Yazici and S. Gunal, "A priori verification and validation study of RFKON database," *International Journal of Computer Science: Theory and Application*, 5, 20-27, 2016.

[24] C.-W. Lee, T.-N. Lin, S.-H. Fang, Y.-C. Chou, "A novel clustering-based approach of indoor location fingerprinting", *IEEE 24th Annual International*

Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), 8–11, 2013.

[25] M. S. Choi, and B. Jang, "An Accurate Fingerprinting based Indoor Positioning Algorithm", *International Journal of Applied Engineering Research*, 12(1), 86-90, 2017.

[26] M. A. Youssef, A. Agrawala and A. U. Shankar, "WLAN location determination via clustering and probability distributions," *First IEEE International Conference on Pervasive Computing and Communications* , 2003.

[27] Y. Chen, Q. Yang, J. Yin, and X. Chai, "Power-efficient access point selection for indoor location estimation," *IEEE Transactions on Knowledge and Data Engineering*, 877–888, 2006.

[28] S. H. Fang, T. N. Lin and P. C. Lin, "Location Fingerprinting in a Decorrelated Space", I*EEE Transactions on Knowledge and Data Engineering*, 20(5), 685–691, 2008.

[29] A. S. Al-Ahmadi, T. A. Rahman, M. R. Kamarudin, M. H. Jamaluddin and A. I. Omer, "Single- Phase Wireless LAN Based Multi-floor Indoor Location Determination System," *IEEE 17th International Conference on Parallel and Distributed Systems (ICPADS)*, 1057- 1062, 2011.

[30] W. Min, and Z. Chunkai, "ResidualRanking: A robust access-point selection strategy for indoor location tracking," *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 5035-5040, 2009.

[31] N. Swangmuang and P.V. Krishnamurthy, "On clustering RSS fingerprints for improving scalability of performance prediction of indoor positioning systems", 1*st ACM International Workshop on Mobile Entity Localization and Tracking in Gps-Less Environments*, 61–66, 2008.

[32] G. Sithole, and S. Zlatanova, "Position, Location, Place and Area: An Indoor Perspective," *ISPRS Annals of Photogrammetry Remote Sensing & Spatial Information Sciences*, 89–96, 2016.

[33] C. Song, J. Wang, and G. Yuan, "Hidden Naive Bayes Indoor Fingerprinting Localization Based on Best-Discriminating AP Selection," *ISPRS Int. J. Geo-Inf.*, 2016.

[34] G. Caso, L. De Nardis and M. G. Di Benedetto, "A mixed approach to similarity metric selection in affinity propagation-based WiFi fingerprinting indoor positioning", *Sensors*, 15(11), 27692-27720, 2015.

[35] S. Saha, K. Chauhuri, D. Sanghi, and P. Bhagwat, “Location determination of a mobile device using IEEE 802.11b access point signals,” *Department of Computer Science and Engineering, Tech. Rep.*, 2003.

[36] V. Honkavirta, T. Perala, S. Ali-Loytty, and R. Piché, "A comparative survey of WLAN location fingerprinting methods," *IEEE 6th Workshop on Positioning, Navigation and Communication*, 243-251, 2009.

[37] D. Li, B. Zhang, B., and C. Li, "A Feature-Scaling-Based k-Nearest Neighbor Algorithm for Indoor Positioning Systems", *IEEE Internet of Things Journal*, 3(4), 590-597, 2016.

[38] Z. Guowei, X. Zhan, and L. Dan, "Research and improvement on indoor localization based on RSSI fingerprint database and K-nearest neighbor points", *2013 International Conference on Communications, Circuits and Systems (ICCCAS)*, 2013.

[39] W. Jing, H. Zhao, C. Song, X. Lin, and X. Shen, "Community detection based reference points clustering for indoor localization in WLAN," I*nternational Conference on Wireless Communications and Signal Processing (WCSP '13)*, 1–6, 2013.

[40] B. Altintas and T. Serif, "Improving RSS-based indoor positioning algorithm via K-Means clustering," *11th European Wireless Conference 2011—Sustainable Wireless Technologies (EuropeanWireless)*, 1–5, 2011.

[41] Y. Sun,Y.Xu, L.Ma, and Z.Deng, "KNN-FCM hybrid algorithm for indoor location in WLAN," *2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS '09)*, 251–254, 2009.

[42] Z. Tian, X. Tang, M. Zhou, and Z. Tan, "Fingerprint indoor positioning algorithm based on affinity propagation clustering," *EURASIP Journal onWireless Communications and Networking*, 2013.

[43] X. Hu, J. Shang, F. Gu, and Q. Han, "Improving Wi-Fi indoor positioning via AP sets similarity and semi-supervised affinity propagation clustering", *International Journal of Distributed Sensor Networks,* 11(1), 2015.

[44] D. Sánchez-Rodríguez, P. Hernández-Morera, J.M. Quinteiro, and I. Alonso-González, "A low complexity system based on multiple weighted decision trees for indoor localization", *Sensors*, 14809–14829, 2015.

[45] J.R. Quinlan, "Induction of decision trees", *Mach. Learn*, 81–106, 1986.

[46] J.R. Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann: San Francisco, CA, USA, 1993.

[47] L. Breiman, J. Friedman, R.A. Olshen, and C.J. Stone, "Classification and Regression Trees", Chapman and Hall/CRC: New York, NY, USA, 1984.

[48] S. B. Keser and U. Yayan, "A case study of optimal decision tree construction for RFKON database," *2016 International Symposium on INnovations in Intelligent SysTems and Applications (INISTA)*, 1-6, 2016.

[49] J. Yim, I. Ko, J. Do, J. Joo, and S. Jeong, "Implementation of a prototype positioning system for LBS on U-campus", *Journal of Universal Computer Science*, 14 (14), 2381-2399, 2008.

[50] O.M. Badawy, and M. Hasan, "Decision tree approach to estimate user location in WLAN based on location fingerprinting", *24th National Radio Science Conference*, 1-10, 2007.

[51] Y. Mo, Z. Zhang, Y. Lu, W. Meng, and G. Agha, "Random forest based coarse locating and kpca feature extraction for indoor positioning system", *Mathematical Problems in Engineering*, 2014.

[52] D. Sánchez-Rodríguez, P. Hernández-Morera, J.M. Quinteiro, and I. Alonso-González, "A low complexity system based on multiple weighted decision trees for indoor localization", *Sensors*, 15(6), 14809-14829, 2015.

[53] Y. Chen, Q. Yang, J. Yin and X. Chai, "Power-efficient access-point selection for indoor location estimation," *IEEE Transactions on Knowledge and Data Engineering*, 18(7), 877-888, 2006.

[54] H. Y. Noh, J. H. Lee, S. W. Oh, K. S. Hwang, and S. B. Cho, “Exploiting indoor location and mobile information for contextawareness service,” *Information Processing & Management*, 48(1), 1-12, 2012.

[55] R. Battiti, “Location-aware computing: a neural network model for determining location in wireless lans,” *Universita degli Studi di Trento, Tech. Rep. DIT-0083*, 2002.

[56] U. Ahmad, U. Nasir, M. Iqbal, "In-building localization using neural networks", *IEEE international conference on engineering of intelligent systems*, 1–6, 2006.

[57] S. Ngah, "Artificial neural network based model for local position systems", *Journal of Computer Science*, 97-106, 2009.

[58] H. Mehmood, N.K. Tripathi, and T. Tipdecho, "Indoor positioning system using artificial neural network", *Journal of Computer Science*, 1219-1225, 2010.

[59] H. Mehmood, and N. K. Tripathi, "Optimizing artificial neural network-based indoor positioning system using genetic algorithm", *International Journal of Digital Earth*, 6(2), 158-184, 2013.

[60] S. H. Fang and T. N. Lin, "Indoor Location System Based on Discriminant-Adaptive Neural Network in IEEE 802.11 Environments," *IEEE Transactions on Neural Networks*, 19(11), 1973-1978, 2008.

[61] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications", *Neurocomputing*, 489–501, 2006.

[62] R. Rajesh and J. Prakash, “Extreme learning machines - a review and state-of-the-art,” *International Journal of Wisdom Based Computing*, 35–49, 2011.

[63] T. Mantoro, A. Olowolayemo, S. O. Olatunji, M. A. Ayu, A. Osman, and M. Tap, "Extreme learning machine for user location prediction in mobile environment", International Journal of Pervasive Computing and Communications, 7(2), 162-180, 2011.

[64] W. Xiao, P. Liu, W.-S. Soh, and Y. Jin, “Extreme learning machine for wireless indoor localization,” *11th Int. Conf. Inf. Process. Sens. Netw.*, 101–102, 2012.

[65] H. Zou, X. Lu, H. Jiang, and L. Xie, "A fast and precise indoor localization algorithm based on an online sequential extreme learning machine", *Sensors*, 15(1), 1804-1824, 2015.

[66] H. Zou, B. Huang, X. Lu, H. Jiang, and L. Xie, "A robust indoor positioning system based on the procrustes analysis and weighted extreme learning machine", *IEEE Transactions on Wireless Communications*, 15(2), 1252-1266, 2016.

[67] X. Jiang, J. Liu, Y. Chen, D. Liu, Y. Gu, and Z. Chen, Z, "Feature Adaptive Online Sequential Extreme Learning Machine for lifelong indoor localization", *Neural Computing & Applications*, 27(1), 2016.

[68] X. Lu, H. Zou, H. Zhou, L. Xie, and G. B. Huang, "Robust extreme learning machine with its application to indoor positioning", *IEEE transactions on cybernetics*, 46(1), 194-205, 2016.

[69] Y. Gu, J. Liu, Y. Chen, and X. Jiang, "Constraint online sequential extreme learning machine for lifelong indoor localization system", *2014 International Joint Conference on Neural Networks (IJCNN)*," 732-738, 2014.

[70] Y. Gu, Y. Chen, J. Liu, and X. Jiang, " "Semi-supervised deep extreme learning machine for Wi-Fi based localization", *Neurocomputing*, 282-293, 2015.

[71] J. Liu, Y. Chen, M. Liu, and Z. Zhao, "SELM: semi-supervised ELM with application in sparse calibrated location estimation", *Neurocomputing*, 74(16), 2566-2572, 2011.

[72] Y. Tian, B. Denby, I. Ahriz, P. Roussel, R. Dubois, and G. Dreyfus, "Practical indoor localization using ambient RF", *Instrumentation and Measurement Technology Conference (I2MTC)*, 1125-1129, 2013.

[73] M. Brunato and R. Battiti, “Statistical learning theory for location fingerprinting in wireless lans,” *Comput. Netw.*, 47(6), 825–845, 2005.

[74] M. Vahidnia, M. Malek, N. Mohammadi, and N. Alesheikh, "A Hierarchical Signal-Space Partitioning Technique for Indoor Positioning with WLAN to Support Location-Awareness", *Mobile Map Services. Wireless personal communications*, 69(2), 2013.

[75] J. A. Suykens, and J. Vandewalle, "Least squares support vector machine classifiers", *Neural processing letters*, 9(3), 293-300, 1999.

[76] W. Yanhua, W. DongH, and Z. Yan, "Axial decoupled LS-SVMs for indoor positioning using RSS fingerprints", *Control Conference (CCC)*, 3920-3925, 2015.

[77] D. Madigan, E. Einahrawy, R. P. Martin, W.-H. Ju, P. Krishnan, and A. S. Krishnakumar, “Bayesian indoor positioning systems,” I*EEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM ’05)*, 1217–1227, 2005

[78] M. Kessel, and M. Werner, "SMARTPOS: Accurate and precise indoor positioning on mobile phones," *The First International Conference on Mobile Services, Resources, and Users*, 158–163, 2011.

[79] W. Zhang, "INBS: An Improved Naive Bayes Simple learning approach for accurate indoor localization," *2014 IEEE International Conference on Communications (ICC)*, 148-153, 2014.

[80] W. Zhang, X. Hua, K. Yu, W. Qiu and S. Zhang, "Domain clustering based WiFi indoor positioning algorithm," *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 1-5, 2016.

[81] S. Ito, and N. Kawaguchi, "Bayesian based location estimation system using wireless LAN", *Third IEEE International Conference on Pervasive Computing and Communications Workshops*, 273–278, 2005.

[82] D. Madigan, E. Elnahrawy, R. P. Martin, W. H. Ju, P. Krishnan, and A. S. Krishnakumar, "Bayesian indoor positioning systems", *IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies,* 1217–1227, 2005.

[83] I. Bisio, F. Lavagetto, M. Marchese, and A. Sciarrone, "Smart probabilistic fingerprinting for Wi-Fi-based indoor positioning with mobile devices", *Pervasive Mob. Comput.*, 107-123, 2016.

[84] J. D. Paola and R. A. Schowengerdt, "A detailed comparison of back propagation neural network and maximum-likelihood classifiers for urban land

use classification," *IEEE Transactions on Geoscience and Remote Sensing*, 981-996, 1995.

[85] S. Bozkurt, G. Elibol, S. Gunal and U. Yayan, "A comparative study on machine learning algorithms for indoor positioning," *2015 International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, 1-8, 2015.

[86] N. Pritt, "Indoor location with Wi-Fi fingerprinting," *2013 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 1-8, 2013.

[87] M. Brunato and R. Battiti, "Statistical learning theory for location fingerprinting in wireless LANs," *Comput. Netw.*, 47(6), 825–845, 2005.

[88] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O'Dea, "Relative location estimation in wireless sensor networks", *IEEE Trans. Signal Process"*, 2137–2148, 2003.

[89] L. H. Chen, E. H. K. Wu, M. H. Jin and G. H. Chen, "Intelligent Fusion of Wi-Fi and Inertial Sensor-Based Positioning Systems for Indoor Pedestrian Navigation," *IEEE Sensors Journal*, 14(11), 4034-4042, 2014.

[90] P. Barsocchi, A. Crivello, D. La Rosa, and F. Palumbo, "A multisource and multivariate dataset for indoor localization methods based on WLAN and geo-magnetic field fingerprinting", *Indoor Positioning and Indoor Navigation (IPIN)*, 1-8, 2016.

[91] K. Nahrstedt, and L. Vu, "CRAWDAD Dataset", uiuc/uim (v. 2012-01-24); University of Illinois Movement (UIM): Champaign, IL, USA, 2012.

[92] P. Ramviyas, C. Sergio, B. Fredrik, and O. Petter, "CRAWDAD Dataset kth/rss (v. 2016-01-05)" Computer Vision and Active Perception Lab., Centre for Autonomous Systems, School of Computer Science and Communication, Royal Institute of Technology (KTH): Stockholm, Sweden, 2016.

[93] J. Torres-Sospedra, R. Montoliu, G. M. Mendoza-Silva, O. Belmonte, D. Rambla, and J. Huerta, "Providing Databases for Different Indoor Positioning Technologies: Pros and Cons of Magnetic Field and Wi-Fi Based Positioning. Mobile Information Systems", 2016.

[94] Z. Tóth and J. Tamás, "Miskolc IIS hybrid IPS: Dataset for hybrid indoor positioning," *26th International Conference Radioelektronika (RADIOELEKTRONIKA)*, 408-412, 2016.

[95] T. L. Saaty, "*The Analytic Hierarchy Process, Planning, Priority Setting, Resource Allocation*", McGraw-Hill, New York, NY, USA, 1980.

[96] S. Bozkurt Keser, U. Yayan, "A Hybrid Approach for Indoor Positioning", *International Journal of Intelligent Systems and Applications in Engineering (IJISAE)*, pp. 162–165, 2016.

[97] A. Abusara, M. S. Hassan, and M. H. Ismail, "RSS fingerprints dimensionality reduction in WLAN-based indoor positioning", *IEEE 2016 Wireless Telecommunications Symposium (WTS)*, 2016.

[98] A. G. Karegowda, A. S. Manjunath, and M. A. Jayaram, " Comparative study of attribute selection using gain ratio and correlation based feature selection", *International Journal of Information Technology and Knowledge Management*, 271-277, 2010.

[99] I.H. Witten, E. Frank, and M.A. Hall" Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques", *Elsevier*, 2011.

[100] S. B. Keser, , A. Yazici, and S. Günal, S, "A hybrid fingerprint based indoor positioning with extreme learning machine", *Signal Processing and Communications Applications Conference (SIU)*, 1-4, 2017,

[101] I. Kononenko, "Estimating Attributes: Analysis and Extensions of RELIEF", *European Conference on Machine Learning*, 171-182, 1994.

[102] J. A. Hartigan, and A. W. Manchek, "Algorithm AS 136: A k-means clustering algorithm", *Journal of the Royal Statistical Society,Series C (Applied Statistics)*, 100-108, 1979.

[103] A. Yazıcı, S. B. Keser, and S. Gunal, "Integration of classification algorithms for indoor positioning system," *2017 International Conference on Computer Science and Engineering (UBMK)*, 267-270, 2017.

[104] N. Pritt, "Indoor positioning with maximum likelihood classification of Wi-Fi signals," *2013 IEEE SENSORS*, 1-4, 2013.

[105] F. Tom. "ROC graphs: Notes and practical considerations for researchers," *Machine learning*, 1-38, 2004.