COMPARATIVE ANALYSIS OF OFFLINE DATABASES FOR Wi-Fi FINGERPRINTING FOR POSITIONING ALGORITHMS IN REAL ENVIRONMENT

by

NS Muhammad Waleed Pasha

Supervisor

Dr. Mir Yasir Umair

A thesis submitted to the faculty of Electrical Engineering Department Military College of Signals, National University of Sciences and Technology, Rawalpindi as part of the requirements for the degree of MS in Electrical Engineering

June 2021

Abstract

The demands for location-based services (LBS) are increasing day by day for indoor environments. Due to the inability of the Global Positioning System (GPS) signals to penetrate through surfaces like roofs, walls, and other objects in indoor environments, various alternative methods for user positioning have been proposed. Among them, the Wi-Fi fingerprinting approach has sparked significant interest in Indoor Positioning Systems (IPS) because it eliminates the need for line-of-sight measurements and achieves higher performance even in complex indoor environments. For indoor positioning, offline and online are the two phases of the fingerprinting method. Different authors have highlighted the problems in the offline phase as it deals with huge datasets and validation of the Fingerprints without pre-processing of the datasets become a concern. Efficient systems are important to minimize delays, complexity and the associated additional costs. Positioning accuracy of Wi-Fi indoor positioning systems highly depends upon offline databases. Therefore, development of robust Wi-Fi fingerprints is performed to improve the positioning results. Machine learning has been used for the model training in the offline phase whereas the locations are estimated in the online phase. Machine learning algorithms are a natural solution for winnowing through large datasets and determining the significant fragments of information for localization, creating precise models to predict an indoor location. Large training sets are important for improving results in machine learning problems. Therefore, an existing WLAN fingerprinting-based multi-story building location database has been used with 21049 samples, divided into 19938 training and 1111 testing samples. The proposed model uses mean and median filtering as pre-processing techniques applied to the database to improve accuracy by reducing the effect of environmental dispersion, as well as machine learning algorithms (*k*NN, W*k*NN, FS*k*NN, and SVM) for estimating the position. The proposed SVM with median filtering algorithm gives a reduced mean positioning error of 0.7959m and an improved efficiency of 92.84% as compared to all variants of the proposed method for 108703m^2 area.

On the basis of the results of this research, it can be concluded that it is possible to use Wi-Fi fingerprinting for indoor positioning to obtain a state-of-the-art accuracy.

Declaration

I certify that the work *"Comparative Analysis of Offline Databases for Wi-Fi Fingerprinting for Positioning Algorithms in Real Environment"* exhibited in this thesis has not been submitted in support of any other award or educational qualification either at this institution or elsewhere.

Acknowledgements

I am grateful to Allah Almighty who has bestowed me the strength and the passion to accomplish this thesis and I am thankful to Him for His mercy and benevolence, without whose consent I could not have indulged myself in this work.

I would especially like to thank my supervisor Dr. Mir Yasir Umair, Assistant Professor, Military College of Signals (NUST), for the amazing support he has provided me throughout the course of my thesis. I am very grateful for the teachings of faculty members of Electrical Engineering Department and especially the Head of Electrical Engineering Department for their everlasting encouragement that has fueled my sense of continued determination over the years. I would also like to thank Mr. Faizan Rao, Electrical Engineering Department, Military College of Signals (NUST) for his support during my research work.

I would like to show gratitude to my parents and siblings for all their care, love, and support through my times of stress and excitement.

Lastly, I express my gratitude towards my good friend Madeeha Niazi, without her motivation and continuous support, I would have not completed my masters degree on a high note.

Dedication

I dedicate my work to,

my parents who always encouraged my higher education, for their prayers, love and motivation and sacrifices all along.

my supervisors and friends for their support and patience during all the phases of my MS.

Table of Contents

List of Figures

List of Tables

List of Acronyms

Global Positioning System	GPS
Wireless Fidelity	Wi-Fi
Indoor Positioning Systems	IPS
Wireless Local Area Network	WLAN
k-Nearest Neighbor	*k*NN
Weighted *k*-Nearest Neighbor	W*k*NN
Feature Scaling Based *k*-Nearest Neighbor	FS*k*NN
Support Vector Machine	SVM
Line of Sight	LoS
Radio Frequency	RF
Radio Frequency Identification	RFID
Frequency Modulation	FM
Ultra-Wideband	UWB
Infrared	IR
Access Points	APs
Reference Points	RPs
Received Signal Strength Indicator	RSSI
Time of Arrival	ToA
Time Difference of Arrival	TDoA
Roundtrip Time of Flight	RToF
Received Signal Phase	RSP

Chapter 1

INTRODUCTION

The following chapter gives the general introduction to basic indoor positioning terminologies. Furthermore, it involves the need for indoor positioning and the motivation for selecting this topic. It also gives an outline of the next chapters.

1.1 Localization

Finding the location of a person can be defined as localization [1]. Different systems were developed in ancient times for the localization and navigation of ships at sea, but some of them were also capable of land navigation. In modern days, tracking an object can be considered equivalent to that. Currently, the availability of GPS on our smartphones help us to localize our present location. Self localization and aided localization are the two types of localization [2]. The innate ability of a person to locate his current position using natural abilities like sight, sense, hearing, etc., is defined as self localization while making use of the electronic devices to perform localization for a person is known as aided localization. Aided Localization is further classified into outdoor and indoor localization [2].

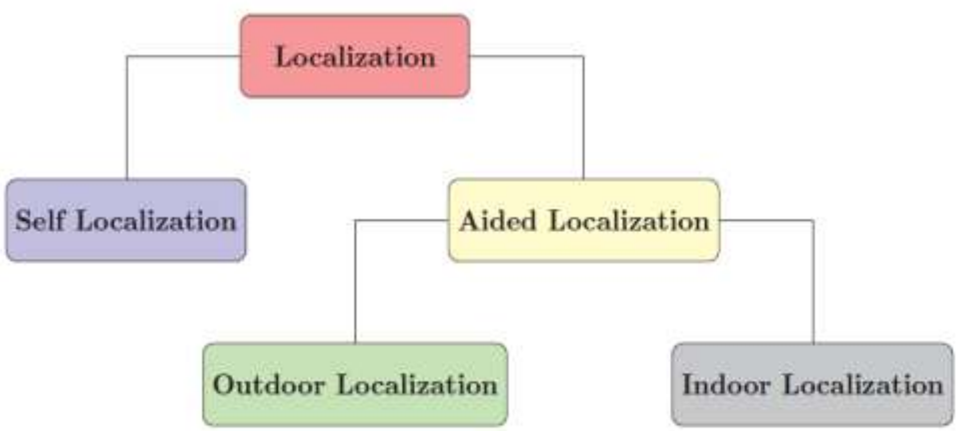

Figure 1.1: Types of localization

In the last decade, the development in the field of positioning and localization have evolved massively. The expansion of modern communication technologies has resulted in a widespread positioning services. In outdoor environments, adequate services have been provided by GPS for positioning and localization. GPS works finest with Line-of-Sight (LoS), however, it is not suitable for indoor positioning, because the signals do not penetrate through hard surfaces and are attenuated and dispersed by the roof, walls, and other objects, therefore different localization systems for indoor environments have been proposed and developed by different researchers [1].

1.2 Indoor Localization

For the past 20 years, indoor localization has gained significance which resulted in the development of multiple indoor positioning systems having their pros and cons for both commercial and research purposes. Figure 1.2 shows the taxonomy of indoor localization systems to differentiate between a variety of signal-based indoor positioning systems.

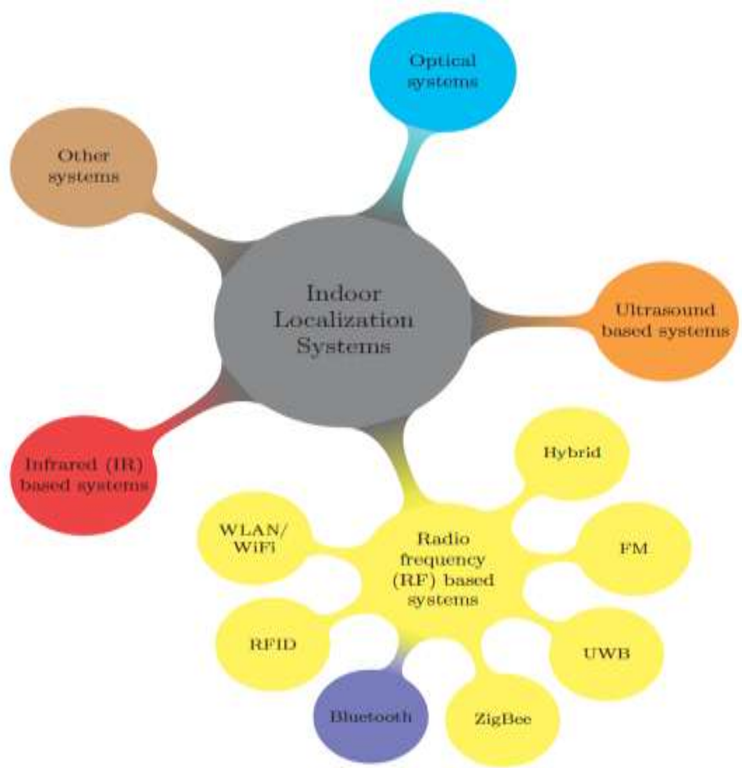

Figure 1.2: Taxonomy of indoor localization systems

In Chapter 2, the measurement methods and signal technologies used in indoor localization systems are thoroughly discussed. A vigorous research has been done in the field of Radio Frequency (RF) and multiple systems have been developed using this technology. A radio wave is an excellent source of information for indoor positioning as it penetrates through walls, roofs, human bodies, and other objects. RF is further divided into the narrow band (RFID, Bluetooth, WLAN/Wi-Fi, FM) and the wideband-based technologies (UWB).

Localization can be used for many purposes, such as identifying combat troops, gathering marketing data, monitoring endangered species, directing self-driven vehicles, robot movement, indoor position for firefighters, hospitals, and malls, to provide navigation aid. While others mandate indoor localization to create better markets for the customers, to find the exact location of products placed in warehouses, to automatically detect object location, to detect medical equipment in hospitals, to detect firefighters in a fired building, to locate police-trained dogs for finding explosives in buildings and to find tagged maintenance equipment.

Indoor localization systems also play a key role in automatic object detection or product tracking based on their locations. Detecting the location of a baggage in a large indoor area, location detection of any product lost in a shop, equipment spread over a large factory or farm. Similarly, position monitoring and emergency alerts in a building engulfed by fire, keeping track of a patient's activity in intensive care, monitoring a suspicious person's actions, and many more. Several localization techniques have been developed for positioning and navigation purposes [3, 4].

1.3 Wi-Fi Fingerprinting

Different localization systems for indoor environments have been proposed and developed by the researchers with their pros and cons. The most commonly adopted method is the Fingerprinting method. Fingerprinting provides better precision, though the practical implementation is relatively arduous, however, the working is very simple or least complex as opposed to other localization techniques. Also, no new equipment is needed and it can be introduced using the existing infrastructure.

The measurement of signal power from an access point (AP) to a receiver that can be sampled in the WLAN environment without any additional requirement can be defined as a received signal strength indicator (RSSI). RSSI-based fingerprint positioning method uses location-dependent features and the position is estimated using these features. Offline and online are the two phases involved in the fingerprint-based indoor positioning. During the offline step, a database is developed that has fingerprints in it where RSSI values are collected from the APs at predetermined reference points (RPs) over a fixed time. The fingerprints stored in the database consist of the reference point position and every single RSSI value collected from each access point measured in dBm. To successfully locate a fingerprint, it is important to apply some pre-processing techniques on the RSSI readings because of the noise present in the environment. In the online phase, RSSI readings from APs at random RPs are taken by the mobile users in the form of queries. Their suitable location would ultimately be determined by the machine learning algorithms through fingerprint matching. Mean position error is attained as a result of the difference between the actual and predicted location of a user in motion [3]. The baseline working of the Wi-Fi fingerprinting method is depicted in Figure 1.3.

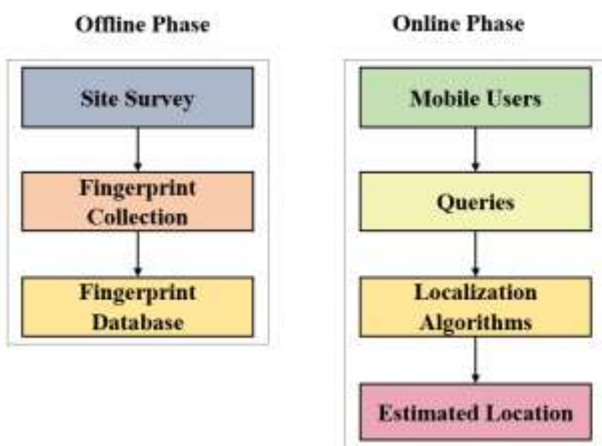

Figure 1.3: Wi-Fi fingerprinting method

1.4 Problem Statement

- Indoor positioning systems focused on Wi-Fi fingerprinting have become very common in recent years. However, the precision and robustness of these systems is a challenge because of the Wi-Fi signal propagation.
- In the literature, various solutions for improving the accuracy and robustness of an IPS have been suggested, however, they rely on equipment costs, microprocessor computing failure, building plan, and implementation.
- Due to geographical uncertainty, an unexpected error in localization is caused by the environmental dispersion.

1.5 Motivation

If there is any emergency, fire in a building, or any other such situations, the security and law enforcement staff need to know exactly how many people are affected and their exact locations to safely rescue them. The basic knowledge about the number of people in the building or a large number of travelers at an airport is not enough to effectively deal with such a situation. The need is to know the individuals' precise positions because this can be a matter of life and death.

Therefore, radio systems with positioning capabilities have emerging applications in home defense, law enforcement, emergency response, and defense command and control. Indoor location and positioning systems have become very common in recent years as a result of emerging technologies and developments. The Internet of Things (IoT), Automation, Directions and Navigations, Robotics, Self-driven Vehicles, etc. All these technologies require the implementation of localization systems within them.

1.6 Indoor Localization Use Cases

Indoor localization can be deployed to the following:

- Museums.
- Private Homes.
- Context Detection and Situational Awareness.
- Medical Care.
- Police and Firefighters.
- Guiding of Vulnerable People.
- Gym and Fitness Centers.
- Environmental Monitoring.

1.7 Summary of Contributions:

The main contributions of the work in this thesis are as follows:

- A solution has been proposed for indoor localization where mean and median filtering techniques are used as pre-processing techniques with machine learning algorithms (*k*NN, W*k*NN, FS*k*NN, and SVM) to enhance the localization accuracy and efficiency of an IPS.

- Large training sets are key for obtaining better results in machine learning problems. Therefore, we have used the largest database available online created by authors in [5] and further processed it by applying the mean and median filtering.
- Outliers were removed from the database after the proposed pre-processing techniques and then the position of a mobile user was estimated by the machine learning algorithms.
- Moreover, to validate the superior performance of the proposed solution, a comparative analysis was carried out where different machine learning algorithms were compared with one another with mean, median, and without filtering.
- The results have shown that the proposed SVM with median filtering algorithm outperformed other investigated machine learning algorithms with mean and median filtering.

1.8 Thesis Organization

There are five chapters of this thesis. Chapter 1 justifies the importance, motivation, and relevance of the research to the national needs. Objectives and summary of contributions are also discussed. Chapter 2 briefly discusses the related work. Chapter 3 explains the proposed model consisting of pre-processing techniques and the machine learning algorithms. Chapter 4 presents simulation results where the investigated machine learning algorithms are compared with one another using different pre-processing techniques. Chapter 5 summarizes the thesis work along with future recommendations.

Chapter 2

LITERATURE REVIEW

An analysis of previous advances in indoor positioning has been provided in this chapter which covers the most popular methods and matching algorithms used for indoor positioning. Performance matrices are explored with a particular emphasis on fingerprinting-based localization algorithms. The applications of indoor positioning systems are also discussed in detail.

2.1 Localization Methods

The generally used localization methods for indoor positioning are as follows [2]:

- Trilateration & Triangulation.
- Fingerprinting.
- Proximity.
- Dead Reckoning.

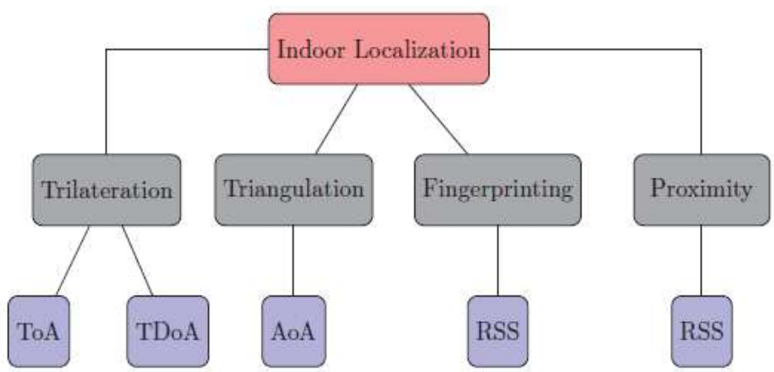

Figure 2.1: Indoor localization methods

In the following subsections, each of these methods will be discussed for better understanding.

2.1.1 Triangulation

In triangulation, the location of a target is determined using the geometric properties of triangles. It has two types:

- Lateration.
- Angulation.

2.1.1.1 Lateration

In lateration, the distance from several reference points is calculated for determining the location of an object. Ranges are also calculated by using this method. Time of arrival (ToA) or time difference of arrival (TDoA) measurement method is used for this approach and using the relationship that gives the distance to the signal velocity multiplied by the time traveled is used to calculate the distance [1] [4]. Well-known used Lateration techniques are as follows:

- Time of Arrival (ToA).
- Time Difference of Arrival (TDoA).
- RSS (Received Signal Strength).
- RToF (Roundtrip Time of Flight).
- RSP (Received Signal Phase.

ToA The calculation of overall signal travel time from the transmitter to the receiver is the basis of ToA theory. This implies that the distance between the mobile target and the unit of measurement is directly proportional to the propagation time. As shown in Figure 2.2, ToA must be calculated with at least three reference points to approximate the position of an object in 2D. The ToA-based positioning systems measure unidirectional time of propagation and then the distance between the measurement unit and the transmitting unit is evaluated. With various signaling methods, such as direct-sequence spread-spectrum (DSSS) or ultra-wide-band (UWB) measurements, ToA may be used. ToA has issues with indoor environments because it is not always possible to guarantee LoS and multipath fading is a natural occurrence [2] [4].

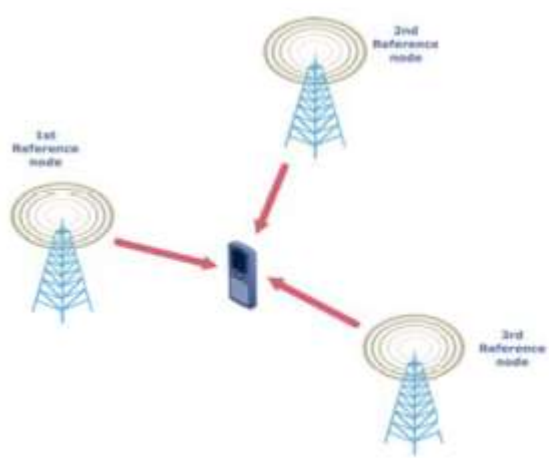

Figure 2.2: Time of arrival (ToA) technique

TDoA Rather than an absolute time of arrival, this method determines the relative position of the cell transmitter by comparing the times at which the signal crosses different units of measurement. The transmitter must be based on a hyperboloid with a continuous difference in range between two measurement units in order to calculate TDoA. Two transmitters at established positions and receiver positioned on hyperboloid may be used for a TDoA calculation. The location of the target can be estimated in 2D, as shown in Figure 2.3, through two or more TDoA measurements' intersection. Usage of correlation method is the standard method for calculating estimates of TDoA. This technique does not offer the guarantee of LoS indoor connectivity [1] [4].

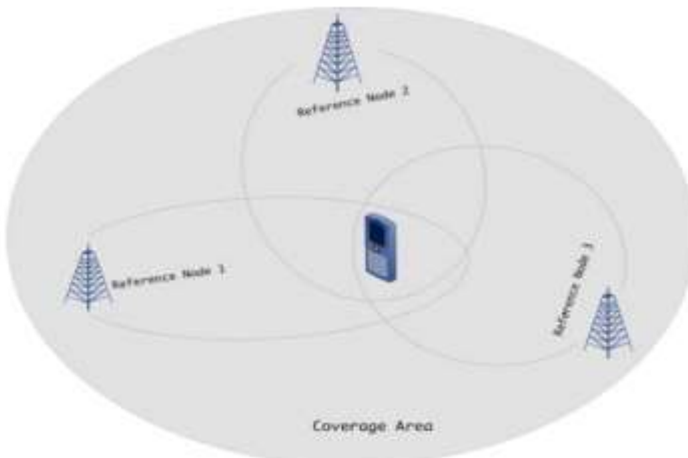

Figure 2.3: Time difference of arrival (TDoA) technique

RSSI Environmental factors affect the radio transmission and the reliability of the ToA & AoA measuring methods is influenced by RSSI. A substitute method for calculating the distance of the receiver from the transmitter is based on attenuation of the signal power. This method measures the path loss portion of the signal which results from propagation. The discrepancy between the RSSI of the transmitter and the RSSI of the receiver can be translated into range approximation with the aid of theoretical and empirical methods, as shown in Figure 2.4, where LS1, LS2, and LS3 represent the path loss that is measured. Since each indoor atmosphere has its own disruption features, path-loss models are often specific for the site which can enhance accuracy for pre-measured RSSI contours oriented at the receiver or multiple measurements at multiple base stations [1, 2].

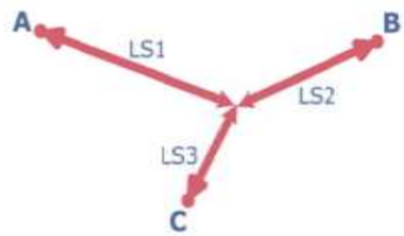

Figure 2.4: Localization based on received signal strength (RSS)

RToF This approach tests flight time from the transmitter to the unit of measurement and from the unit of measurement to the transmitter. RToF approach also appears in literature as Round-Trip Time (RTT) and Two Way Ranging (TWR). In this way, the criterion for getting synchronized is not as rigorous as in the ToA measuring system. In this case, the measuring unit is a typical radar. The radar signal being measured is received by a target transponder and the units of measurement measure the full round-trip time propagation. The delay issue with the respondent must be known by the unit of measurement. This is not a significant concern in long or medium-range systems, but it must be considered in short-range systems [2] [4].

RSP To estimate the range this approach uses phase difference. It is also called the Arrival Phase process. Better localization results are obtained through this approach when used with ToA / TDoA or RSSI methods for indoor localization systems. However, the downside of this approach is the need of the LoS path which leads to more errors in indoor environments if not fulfilled. Ambiguous calculations of carrier-phase are still an issue to be solved [1] [4].

2.1.1.2 Angulation

The orientation of an object is determined using estimated angles relative to different reference points in the angulation measurement method. Usually, this technique is implemented with the AoA approach [1].

AoA The intersection of all pairs of angle direction lines created by the circular distance between the base station and the mobile target can be determined in AoA to determine the position of the target that is chosen, as shown in Figure 2.5. At least two reference points and two measured angles to obtain the target's 2D position, are normally allowed by these methods. An antenna array or a directional antenna are used to calculate the AoA. It is also known as path finding [2] [4].

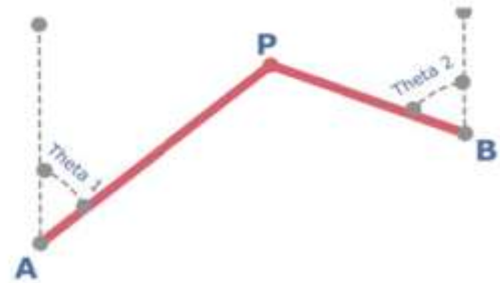

Figure 2.5: Localization based on angle of arrival (AoA) measurement

2.1.2 Fingerprinting

The RSSI based fingerprint positioning method uses location-dependent features and the position is estimated by using these features. Offline and Online are the two stages involved in the fingerprint-based indoor positioning as demonstrated in Figure 2.6.

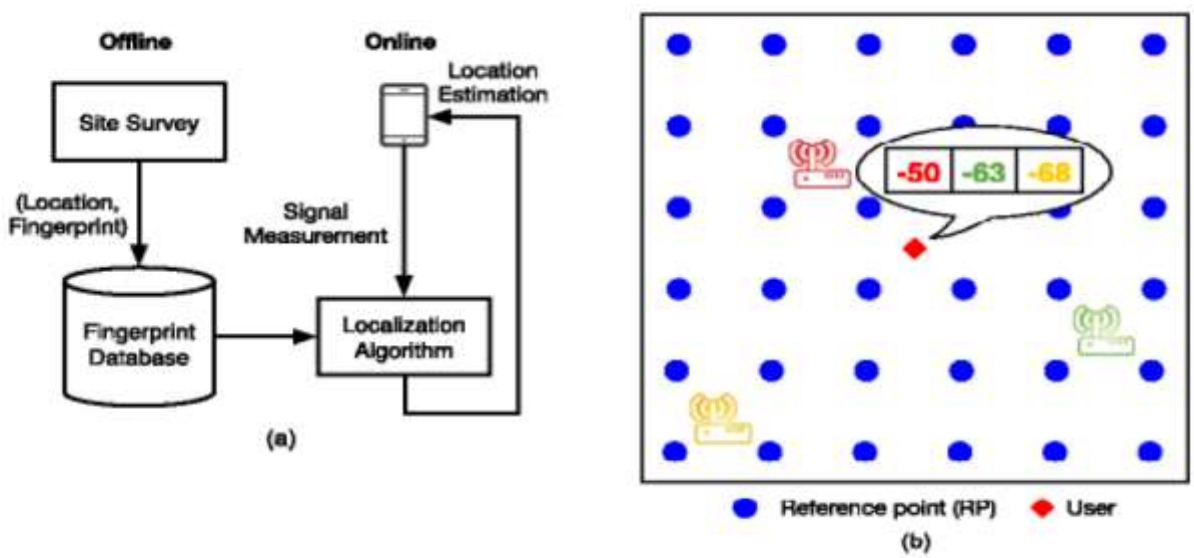

Figure 2.6: Fingerprinting workflow of training and online phase

In the offline stage, a fingerprint database is developed and in the online stage, the target position is estimated. In a fingerprint database, the RSSI values collected from access points at predetermined reference points over a fixed time are stored. In this database, each fingerprint effectively contains the information of location and RSSI values obtained from the access points around that location. RSSI fingerprints are obtained from several APs at each RP in the offline process. In an indoor environment, RP is the location that needs to be monitored and the fingerprint is updated or inserted into the offline database at each RP. To successfully identify the fingerprint, it is important to apply some filtering techniques on the RSSI readings because of the noise present in the environment. In the online phase, the RSSI value vector at an unknown location is checked by the background service running on a mobile user. The most likely location of mobile users would ultimately be determined using the positioning algorithms through RSSI value comparison between the offline and online phase. Fingerprinting aims to greatly increase the precision and accuracy of traditional signal strength lateration techniques. While RSSI is the most common signal technology used for fingerprinting, there are other systems that identify audio signals or visual images [1] [4].

2.1.3 Proximity

A mobile device's location only through its existence in a special area is determined by the proximity approach. Details about the symbolic relative location are provided by the proximity-based algorithms. This approach works by simply redirecting the position of an anchor. The proximity measuring device is simple to set up, but the precision of the process is determined by the placement of anchor points and the range of the signals. The localization systems based on proximity are generally focussed on signaling techniques such as Infrared (IR) Identification and Radio Frequency defined as RFID. Physical touch detection, automatic recognition systems, and mobile wireless positioning systems are specific examples of proximity-based localization systems [2].

2.1.4 Dead Reckoning

In dead reckoning, the location is calculated by using knowledge of the previously determined locations and the predicted velocities over time. One concern with the use of this method is the cumulative inaccuracy; thus, over time, variance in the fix position rises. A concept called Pedestrian Dead Reckoning (PDR) is used in the context of indoor applications suggesting that external sensors such as accelerometers are connected to the user's body [2] [4].

Table 2.1 summarizes the various algorithms and measurement methods used for indoor localization in terms of certain primary output metrics.

Table 2.1: Summary of different methods used in indoor localization systems

Method	**Measurement type**	**Accuracy**	**Coverage**	**LoS/NLoS**	**Multipath effect**	**Cost**
Proximity	RSS	Low to high	Good	Both	No	Low
Direction	AoA	Medium	Good	LoS	Yes	High
Time	ToA, TDoA	High	Good	LoS	Yes	High
Fingerprinting	RSS	High	Good	Both	No	Medium
Dead reckoning	Acceleration, Velocity	Low to medium	Good	NLoS	Yes	Low

2.2 Information Source-based Indoor Localization Technologies

Different signal techniques may be used to build indoor localization systems. Those technologies are described below:

- Infrared (IR) Localization Systems.
- Ultrasonic (US) Localization Systems.
- Radio Frequency (RF) Localization Systems.
- Optical Localization Systems.
- Other Localization Systems.

2.2.1 IR Localization Systems

Infrared radiation (IR)-based systems use the infrared spectral field to find applications to detect or track objects or individuals. They are readily available on different devices, such as cell phones, PDAs, and TVs. IR-based device mechanism is based on the use of LOS contact between the transmitter and the receiver unless the environment interferes with the optical sources. Because of their small size and lightweight, they are beneficial but also have protection and privacy problems, which require high hardware with maintenance costs [1, 2].

Active Badge System [6] is an example of an infrared-based localization system. Table 2.2 lists some of the indoor localization systems that use infrared technologies.

Table 2.2: Indoor localization systems based on IR technology

Name	Year	Accuracy	Coverage	Principle	Target illumination
Active Badges	1999	6m	Scalable	Proximity	Signal tranmission
Lee and Song	2007	dm	$36m^2$	IR camera	Retro reflective
Ambiplex	2011	20-30cm	10m	AoA	Natural IR radiation
Kinect	2011	1cm	3.5m	Structured light	Passive

2.2.2 Ultrasonic Localization Systems

Ultrasonic positioning devices use ultrasonic waves to determine the distance between the source of the sound and the mobile unit. Typically, such devices have a large number of ultrasonic receivers, and communication between them is accomplished using IR or RF waves. ToA sound signal data from source to the receiver is used by these systems to estimate the distance from source for receivers. Devices based on ultrasound have very high accuracy. These systems allow a good choice of indoor location even at low cost, ease of installation, and higher precision. A downside of ultrasonic localization systems is that multi-path reception is always impaired and can be difficult to implement on a large scale [1, 2].

Active Bat [7], Cricket [8], Losnus [9], and Alloula [10] are several examples of sound-based indoor localization systems. These devices have applications in smart mapping, surveillance, wireless sensor networks (WSN) and have cm-level precision. Table 2.3 summarizes some of the ultrasonic based indoor localization systems.

Table 2.3: Ultrasonic based indoor localization systems

Name	Year	Accuracy	Carrier Frquency	Principle	Application
Active Bat	1997	3cm	40kHz	Multilateration	Smart tracking
Cricket	2005	1-2cm	40kHz	Multilateration	Smart tracking
Losnus	2010	1cm	35-65kHz	Multilateration	WSN
Alloulah	2010	3cm	20-50kHz	Multilateration	Monitoring
Sato	2011	4cm	40kHz	Multilateration	Human motion

2.2.3 Radio Frequency (RF) Localization Systems

Because of the property of radio waves, localization systems based on radio frequency (RF) technology are now commonly used to traverse obstacles such as buildings, human beings, and other objects. As a result, these systems have more coverage and need fewer equipment to deploy. The technologies using RF localization are as follows:

- RFID.
- Bluetooth.
- WLAN/Wi-Fi.
- FM.
- ZigBee.
- UWB.
- Hybrid [1] [4].

2.2.3.1 RFID

Indoor localization solutions use radio frequency identification (RFID), which is one of the most advanced tools for identifying persons or objects. A basic device would consist of a reader with an antenna searching its surroundings for active transceivers or passive tags on a continuous basis. Radio signals are one way to wirelessly relay data from RFID tags to the reader. The proximity approach is the most common technique for localization, in which the device detects that a human is currently wearing the RFID tag. RSSI can also be used in applications requiring coarse-range localization. Using the measuring methods ToA and AoA, RFID-based localization systems have proven difficult to build. It is also possible to apply fingerprint implementation based on pre-measured signal maps to localize the RFID scheme. In many applications, such as people position, automotive assembly industry, warehouse management, supply chain network, etc., RFID-based localization systems are used because the software operates without sight requirements [2] [4].

[11], [12], and the “ways4all” method created by [13] are some examples of RFID-based localization systems. The table 2.4 provides a summary of several RFID based localization systems.

Table 2.4: RFID based indoor localization systems

Name	Year	Tag Range	Accuracy	Principle	Application
Dziadak	2005	2m	in meters	Proximity	Burried asset detection
Seco	2010	30m	1.5m	RSSI, FP	Person/object location
Peng	2011	100m	1-3m	RSSI + IMU	Pedestrian navigation
Kimaldi	2011	13m	room-level	Proximity	Hospital
Kiers	2011	11-30cm	dm	Proximity	Navigation of blind

2.2.3.2 ZigBee

ZigBee is a popular wireless standard technology for short- and medium-range communications. This can be considered as a low-rate Personal Area Wireless Network (WPAN). The specification is intended for applications that require low power consumption and have no high throughput of data. Indoor environments usually have a ZigBee signal range of 20-30 m. For the measurement of the distance between two ZigBee nodes, RSSI is the standard term used. One downside is that the proposed localization system is vulnerable to interference from the other sources of signal because ZigBee operates in the unlicensed industrial, scientific, and medical (ISM) band, which would interfere with the communication of radio [2, 3].

[14] and [15] are two examples of ZigBee-based indoor localization research. Table 2.5 lists a number of ZigBee based localization systems.

Table 2.5: Indoor localization systems based on ZigBee technology

Name	Year	Calibration	Accuracy	Principle	Application
Tadakamadla	2006	Minimal	3m	RSSI	Context, LBS
Larranaga	2010	Yes	3m	RSSI	WSN, tracking
MyBodyguard	2011	No	Proximity	Fingerprinting	Tracking

2.2.3.3 WLAN/Wi-Fi

Because of the low network expense and lack of need for LoS, it is standard practice to use a Wi-Fi-based indoor positioning system. Without any extra modification of the hardware or software, any computer with Wi-Fi compatibility can be easily found. Although the measurement methods ToA, AoA and TDoA have different frameworks, they are commercially available and are also based on the concept of measuring signal intensity received. Introducing a localization scheme using WLAN technology has many advantages. Many of them have readily accessible access points in indoor environments, no special hardware specifications, and a range of 50-100 meters, which makes it more attractive than Bluetooth or RFID.

RADAR [16] is an example of a Wi-Fi-based localization program. It was created as a user-friendly location and monitoring application that addresses all location and tracking problems. It is a comfortable indoor localization solution since it was implemented entirely in software. The basic idea behind RADAR was to use signal strength as a feature of receiver location and transmitter map. The approach to fingerprinting has an offline and an online process [2] [4]. Table 2.6 shows a summary of several WLAN/Wi-Fi based localization systems.

Table 2.6: Indoor localization systems based on WLAN/Wi-Fi technology

Name	Year	Accuracy	Calibration	Principle	Application
Bahl	2000	5m	Yes	FP	Offline training
Gunther	2004	5-15m	No	RTT	-
Chen	2005	2-4m	Yes	FP & FRID	Offline training
Wong	2008	2m	No	AoA	-
Ekahau	2009	7m	Yes	FP	Offline training
Gansemer	2010	2.1m	Yes	FP	Offline training
Hansen	2011	4cm	Yes	FP	Dynamic model

2.2.3.4 Frequency Modulation (FM)

FM radios are a well-established broadcasting system, and the circumstance that they are used in most homes and vehicles makes them a good option as FM radio transmission audio signals can be used for indoor navigation (and positioning). Fingerprinting (approach) methods have been found to be more feasible for FM radio-based positions than ToA and TDoA approaches [3, 4].

There hasn't been much work done on improving indoor-based localization using FM radio signals, however [17], which is based on the RSSI fingerprinting concept for an office setting, contains some of the applied work. Table 2.7 shows some of the FM based localization systems.

Table 2.7: Indoor localization systems based on FM technology

Name	Year	Calibration	Accuracy	Principle	Application
Papliatseyeu	2009	Yes	4.5m	Fingerprinting	Indoor navigation
Popleteev	2011	Yes	5m	Fingerprinting	Employee tracking
Moghtadaiee	2011	Yes	3m	Fingerprinting	Employee tracking

2.2.3.5 Bluetooth

Bluetooth is identical to ZigBee; a Wireless Protocol Bluetooth is a patented Bluetooth Special Interest Group (SIG) technology. Bluetooth works at 2.4 GHz in the ISM band. The most significant benefit of using Bluetooth on an application is that virtually any Wi-Fi-enabled handheld phone, smartphone, personal digital assistant (PDA), or laptop now has a Bluetooth module. This technology also offers another advantage by the use of the Bluetooth protocol for information communication in the form of providing high security, low cost, low power, and small size. Each Bluetooth tag has a unique identifier that can be used to track down a Bluetooth user. The use of Bluetooth has one possible disadvantage in the form of latency of the Bluetooth system that could make it unsuitable for real-time positioning applications. This is because, with every location detection, the device discovery process has to be run, which in turn increases localization latency and power consumption [3, 4].

The Real-Time Navigational Assistance (URNA) system [18] was one of the first localization applications to use the Bluetooth technology standard. The goal was to allow Bluetooth-enabled mobile devices to share location-based information. It is based on the Proximity principle. Table 2.8 shows some of the Bluetooth based localization systems.

Table 2.8: Indoor localization systems based on Bluetooth technology

Name	Year	Calibration	Accuracy	Principle	Application
Aalto	2004	No	20m	Proximity	Advertising
Bargh	2008	Yes	room-level	Fingerprinting	LBS
ZONITH	2011	No	room-level	Proximity	Employee tracking

2.2.3.6 Ultra-wideband

UWB is a high-bandwidth, short-range radio system with strong multi-path resistance properties. Since other conventional wireless solutions, such as RFID and WLAN/Wi-Fi, do not have such high accuracy, UWB is widely used for location systems with high precision measurements (20-30 cm). Radio wave generators and receivers capable of capturing propagated and scattered waves will be used in a basic UWB stimulation localization framework. UWB signals have the ability to travel through walls, windows, and other barriers, making them ideal for indoor environments where ranging is not constrained by LoS and inter-room ranging is feasible. The problem with UWB is that it is a costly technology that is inefficient for large-scale deployment [1-4]. [19], [20], and [21] all have localization systems dependent on UWB technologies. Table 2.9 shows some of the UWB based localization systems.

Table 2.9: Indoor localization systems based on UWB technology

Name	Year	Noise Radar or IR (Pulse Duration)	Accuracy	Principle	Application
Stoica	2006	IR(750ps)	4cm	ToA	Sensor networks
Fischer	2010	IR(200ps)	4cm	ToA, RTT	Industrial
Segura	2010	IR(2ns)	20cm	TDoA	Mobile robot
Kroell	2010	Pseudo noise	4cm	FP	Office
UBISENSE	2011	IR(very short)	<15cm	TDoA, AoA	Automation

2.2.3.7 Hybrid

For finding a mobile user, hybrid localization schemes use a variety of technologies. One of the most important services provided by a localization system is the ability to locate a mobile user, and since some location solutions are primarily developed for indoor and GPS-based positioning systems, using an indoor and outdoor hybrid system will be extremely advantageous. This is how the concept of a mixed localization approach arose. Navizon, Xtify, Devicescape, and SkyHook are examples of hybrid localization systems that have been developed [3, 4].

2.2.4 Optical Positioning Systems

The key sensor cameras are optical indoor positioning devices. Optical positioning devices may also be used in conjunction with mechanical or remote sensors. The AoA method is used exclusively in optical indoor localization systems that use camera-based device architectures.

Camera-based indoor localization systems have benefited from advancements in CCD technologies, processing speed, and image perception [1] [4].

Table 2.10 summarizes how optical localization systems can be classified based on their primary mode of reference.

Table 2.10: Optical indoor localization systems

Name	Reference	Coverage	Accuracy	Camera Positioning	Camera Cost
Hile	Floor plan	Scalable	30cm	cam., SR 4000	900 $
Ido	Images	Scalable	30cm	cam., IEEE 1394	-
Mulloni	Coded markers	Scalable	m-dm	cam., call phone	Low
Popescu	Projection	25m^2	cm	camera	1500 $
DEADALUS	None	m-km	0.04mm	obj., Guppy F80	High

2.2.5 Other Systems

Some other systems can be used for indoor positioning which can be explicitly built devices with a particular application. They include:

- Inertial Navigation Systems (INS).
- Magnetic Localization.
- Infrastructure Based Localization Systems [4].

In the following subsections, each of these systems will be discussed for better understanding.

2.2.5.1 INS

An INS consists of an Inertial Measurement Unit (IMU) and the main components are a processing unit. But it also makes use of complementary sensors to provide localization information. An INS is an electronic tool used to estimate location, velocity, and direction from the IMU. The standard IMU consists of three accelerometers, three gyroscopes, and/or one magnetometer, arranged orthogonally [4].

Table 2.11 shows the summary of INS based localization systems.

Table 2.11: INS based localization systems.

Name	Mounting body part or device	Comlimentary Sensors	Accuracy	Local Reference	IMU Sensors
Kemppi	Waist/pocket	3 Accelerometers & Gyroscopes	17m	Map, beacon	Accelerometer
Seitz	Phone	3 Accelerometers & Magnetometers	5m	WLAN, RSSI	Bosch, BMA150
Kligbeil	Waist	3 Accelerometers, Magnetometers, Gyroscopes & 1 Barometer	1-6m	GPS, US, RF, CSS	Accelerometer
Jimenez	Foot	3 Magnetometers & Gyroscopes	1m	RFID, RSSI	MTI-G, Xsens

2.2.5.2 Magnetic Localization

Magnetic and electromagnetic fields are now being used to create localization technologies. Permanent magnets or coils of AC or DC as a source of magnetic fields can be used in these devices [4].

Table 2.12 shows some of the magnetic localization systems.

Table 2.12: Magnet localization systems

Name	Year	Coverage	Accuracy	Principle	Application
Haverinen	2009	280m	1mm	Fingerprinting	Robot localization
InfraSurvey	2011	200m	1m	AC magnetic field	Caves, mines, tunnels
Q-Track	2011	23m	50cm	Near field	NLoS office & industry
Arumugam	2011	50m	20cm	DC field, coils	American football

2.2.5.3 Infrastructure Based Localization Systems

In addition to using current building facilities or incorporating the new technology into building materials, localization systems do not depend upon any of the technologies mentioned above. Common examples are Power Lines, Floor Tiles, and Fluorescent Lamps [3, 4].

Table 2.13 shows some of the infrastructure based localization systems.

Table 2.13: Infrastructure based localization systems

Name	Year	Coverage	Accuracy	Principle	Application
Stuntebeck	2008	Building	1-3m	Power lines	Location aware homes
SensFloor	2011	$50m^2$	dm	Floor tiles	Assistance for elderly
Nishikata	2011	Building	10cm	Fluorescent lamps	Robot guidance
Weber	2011	40m	4m	Leaky feeder	Indoor localization

Figure 2.7 shows a comparison of various indoor location systems in terms of accuracy and coverage.

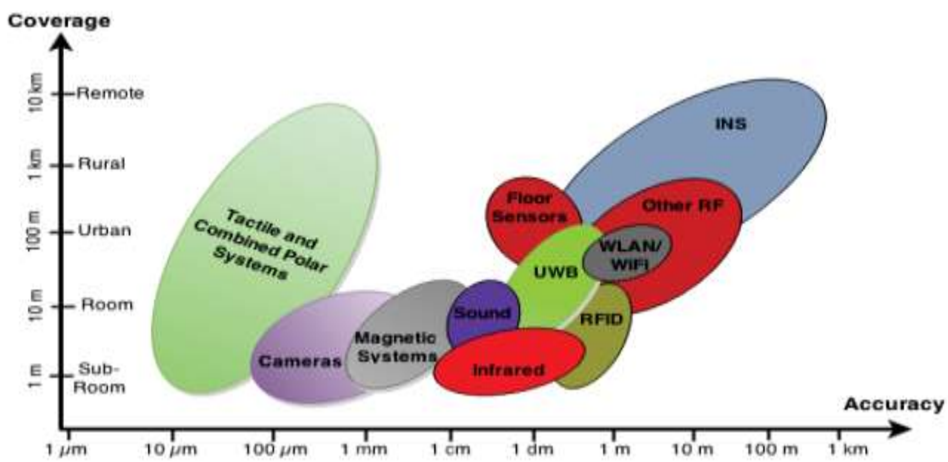

Figure 2.7: Overview of indoor technologies according to their accuracy and coverage

2.3 Matching Algorithms

The nearest neighbor (deterministic) and Bayesian (probabilistic) are the two most common matching algorithms. In the prior approach, a clear distinction between deterministic and probabilistic techniques is that the RSSI values are interpreted as single values, and the latter use distributions of probability for representation. The probabilistic methods contain more information about the range of the signal strength but are highly complex, while the deterministic methods are simple to implement and process [22].

2.3.1 Probabilistic (Bayesian)

The probabilistic method is based on Bayes Theorem. The location is estimated from signal strength vector collected in the online phase $S = (S_1, S_2 ..., S_k)$, and a set of locations $X = \{X_1, X_2, ..., X_m\}$, that maximizes the conditional probability shown in eq. 2.1.

$$P\ (X_i \mid S) \tag{2.1}$$

According to Bayes theorem:

$$P\ (X \mid S) = P\ (S \mid X)\ P(X) / P(S) \tag{2.2}$$

The Horus system [23] is a probabilistic system, as opposed to RADAR. The radio map saves the histogram of the samples obtained from each AP for the signal power. The process of position estimation is a mixture of two techniques, i.e., a discrete estimator of space and a continuous estimator.

2.3.2 Deterministic

The deterministic matching algorithm uses deep learning, data processing, artificial neural networks, and other techniques to align fingerprint data in real time. Since this approach relies on signal intensity measurements, the accuracy of fingerprint data obtained offline has a significant impact [1].

2.3.3 *k*NN

For supervised machine learning, the *k*-Nearest Neighbors (*k*NN) algorithm can be deployed to unravel the problems related to both classification and regression. It determines the location based on the distance between patterns and reference patterns present in the database. Distance can be calculated using a variety of methods, such as Manhattan distance or Euclidean distance. Depending on their reciprocal distance, it decides the best matching fingerprints and averages the position of the same *k* patterns.

The *k*NN algorithm was initially used in RADAR for indoor positioning that takes *k* RPs with the least signal distance between the unspecified RSSI vector of the online user and the identified location databases [24, 25].

2.3.4 W*k*NN

Weighted *k*NN is a modified version of *k*NN. Weight is calculated based on the weighting function which is the reciprocal of the distance between the neighbors. The choice of the hyper parameter *k* is one of the many problems that influence the *k*NN algorithm's efficiency. The algorithm may become more vulnerable to outliers when the value of *k* is very small. Too many points may be included in the field from other groups when *k* is very high. Another problem is the way to incorporate class names. The simplest method is for majority voting, but if the closest neighbors are far different from each other and the closest neighbors display the object's class more precisely, it could be a problem [26].

In [26], a new approach is suggested to improve the accuracy of the W*k*NN algorithm by 33.82 % by changing the weight of the neighboring reference nodes obtaining a mean position error of 0.9 m in a relatively small indoor environment with few RPs. Junhuai Li et al. [27] proposed the Improved Fuzzy C-Means (IFCM) algorithm for the division of the region in the offline training process, and the Pearson Correlation Coefficient (PCC) based weighted *k*-nearest neighbor (W*k*NN) algorithm in the online positioning process, achieving a mean positioning error of 2.53 m. However, the proposed model did not address the optimization of AP deployment approach based on the real environment in order to optimize fingerprint discrimination in each area. An improved adjacent RPs filtering system for Wi-Fi-based indoor localization. was suggested in [28]. To enhance their selection phase, the physical distances between the testing point and the neighboring RPs are used to cluster *k*'s nearest neighbors. A mean position error of 2.6 m was obtained using the proposed algorithm that outperformed the standard *k*NN, W*k*NN, and TPIC algorithms. Moreover, no pre-processing techniques were applied in the proposed model.

2.3.5 FS*k*NN

Assigning different weights to signal variances at different RSSI rates is the basic concept behind FS*k*NN when determining the similarity between dual RSSI values. It needs to create an RSSI-value-based FS model for comparison of similarities. For this model to run legally, it must first be optimized databases, depending on which simulated annealing (SA) is used to actuate the RSSI-level-based scaling weights in the offline phase. In the online process, between an instantaneous RSSI vector and any reference fingerprint in the database, the effective signal distance calculation is performed, and afterward, the restoration of the first *k* RPs causing the least active signal distances is carried out, and the same distances are then used for the estimated area.

Authors suggested a feature scale k-nearest neighbor (FS*k*NN) algorithm in [29] to enhance localization accuracy when a new FS RSSI model is created, including a scaling RSSI level for measuring efficient signal distances between an MS signal vector and fingerprints in the radio map that achieve an average position error of 1.7 m. However, multipath fading was a major concern in their findings which can vary the RSSI values.

2.3.6 SVM

The Support Vector Machine (SVM) has become popular due to its favorite classification/regression effect, a relatively new multivariate statistical approach. A classic SVM is a support-vector network that can be used in supervised learning models. This model is a non-probabilistic binary classifier that can be used to characterize the hyperplane that separates the classes in the training set. It gives a maximized margin. The side of the hyperplane on which a previously unobserved data point falls can be used to estimate its expected mark.

SVM is a powerful supervised learning model that excels at dealing with high-dimensional data sets. It is particularly useful for addressing memory use because it promotes estimation using support vectors. SVM's main appeal is that a high degree of precision is guaranteed for a few training points. These training points are support vectors that can categorize any new data point in the network. SVMs are capable of multiclass classification as well as binary classification. Non-linear classification can also be carried out by SVMs, which can help to find the hyperplane of a non-linear operating input vector. For example, an input variable can be mapped into a high-dimensional space for functionality [30].

Authors in [31] suggested a method of extraction of fingerprint features known as the Fisher score-stacked sparse autoencoder (Fisher-SSAE) method building a hybrid localization model to prevent major coordinate errors of localization accredited to subregional errors of localization. A mean position error of 2.09 m was obtained in combination with three localization algorithms, support vector regression (SVR), random forest regression (RFR), and multiplayer perceptron classification (MPC). However, in this analysis, the moving target was not taken into account.

2.4 Performance Metrics of IPS

IPSs use various localization techniques which are incredibly different in terms of precision, cost, accuracy, innovation, scalability, robustness, and safety. Low-cost IPS can be used in certain applications, whereas high-precision IPS may be used in others, such as medicinal tracing, industrial environmental control, and blind indoor positioning. Diverse performance metrics of IPSs are depicted in this field.

2.4.1 Accuracy

The term accuracy is characterized in the Joint Committee for Metrology Guides (JCGM), as the similarity of agreement between a calculated value and an actual value of a deliberate [32]. The average Euclidean distance between the measured and actual location is the precision of IPS in this way. The precision is an incredibly difficult area for some field analysts. Some compromises among accuracy and other performance metrics that be required [32], given the fact that the accuracy of an IPS is an essential driver for most applications.

2.4.2 Coverage Area

The area covered by IPS is called the area of coverage. Every IPS operates in a different range all of the time. The most powerful systems are those that shelter a large amount of data [33]. Inclusion levels can normally be three; local, scalable, and global for positioning systems. A well-defined, finite area is referred to as local coverage that cannot be extended as a single room or building, while a scalable coverage is taken as a system's capacity to expand the space by including equipment. Then again, for example, GPS, a device that has a worldwide region is referred to as global coverage. In present days, current IPSs range from 5 m to 50 m. In this way, it is difficult to offer a system that has a range of more than 60 m [33].

2.4.3 Availability

The percentage of time the positioning service is available for use with the required precision and fairness is known as accessibility. As IPS Credibility is the assurance then it can be set in the IPS yield. Much like congestion in communications and scheduled factors, the availability may be limited by irregular factors, such as daily maintenance. There are three levels of availability, i.e., less availability (< 95%), consistent availability (> 99%) and higher availability (> 99%) [33].

2.4.4 Scalability

A positioning system can determine the location of objects around the world, within a metropolitan area, on campus, in a specific building, or a solitary room. Additionally, the quantity of targets system may be traceable with a specific number of infrastructures or may be restricted above a given time [34]. If the scalability scales in one of the two measurements: geology and number of customers, then the scalability of an IPS implies device ensures general positioning efficiency. The quantity of the size of clients means that there is a rise in the number of units found per geographic region by time.

2.4.5 Cost

The cost of the IPS is measured at different scales, and these measures are money, time, space, and energy. This is triggered at various frame stages: system installation and maintenance, network elements, and gadget location [35]. For the establishment and maintenance of the system, the cost includes the costs needed for the establishment and any costs that are necessary to save the functional framework, while for the procurement and preparation of components, the costs for infrastructure components and the location of gadgets may include expenditures, space, and energy to use the equipment. For example, IPSs that make use of existing facilities, such as the network, are more cost-effective. Some positioning systems, such as passive RFID stickers, are energy-efficient, while others use a lot of energy. This energy can be considered as a basic asset in IPSs to keep away from the disruption of the facilities and have greater portability arrangements.

2.4.6 Privacy

People who use IPSs need to be protected, and having a clear command about how individual client data is gathered and used is important. The aim is to improve the privacy of consumers, for a particular intent, confidential tools must be implemented and held to protect information from intrusion, theft, and misappropriation. Regrettably, until now, the safety factor of IPSs has been overlooked in much of the presumed work in the area of indoor positioning [36].

Chapter 3

PROPOSED MODEL

This chapter gives complete overview of the proposed indoor positioning model in real time environment. Mean and median filtering techniques are used as pre-processing techniques with machine learning algorithms (*k*NN, W*k*NN, FS*k*NN, and SVM) to enhance the localization accuracy. Large training sets are key for obtaining better results in machine learning problems. Therefore, we have used a WLAN fingerprinting-focused multi-building and multi-floor location database created by authors in [5] and further processed it by applying the mean and median filtering on the database. In the online phase, the machine learning algorithms are applied when a query needs to be matched with the offline database to find the best match for locating the user in a given area. This model helps in reducing the impact of environmental factors and minimizing the mean position error.

3.1 Introduction

RSSI localization techniques are based on calculating the signal intensity of a user to multiple access points located at different locations, then evaluating the distance between the user and the access points. Trilateration techniques may be used to measure the user's calculated location according to the APs' predetermined position.

Traditional fingerprinting is also RSSI-based, but it simply relies on signal strength recording from multiple access points in range, and this information is stored in an offline database with the user device's established co-ordinates. That data, however, may be deterministic or probabilistic. The current RSSI vector is then compared to the stored database as a fingerprint by an unknown location, it occurs during the online phase, and the nearest match is served as the estimated user location.

The database has been created with two android applications, CaptureLoc and ValidationLoc. Both applications are used as services of reference map, providing spatial details about the interiors of buildings as well as the localization of training reference points [5]. The proposed model suggests to use mean and median filtering as pre-processing techniques with the investigated machine learning algorithms to enhance the efficiency of an indoor positioning system. Figure 3.1 shows the block diagram of the proposed model.

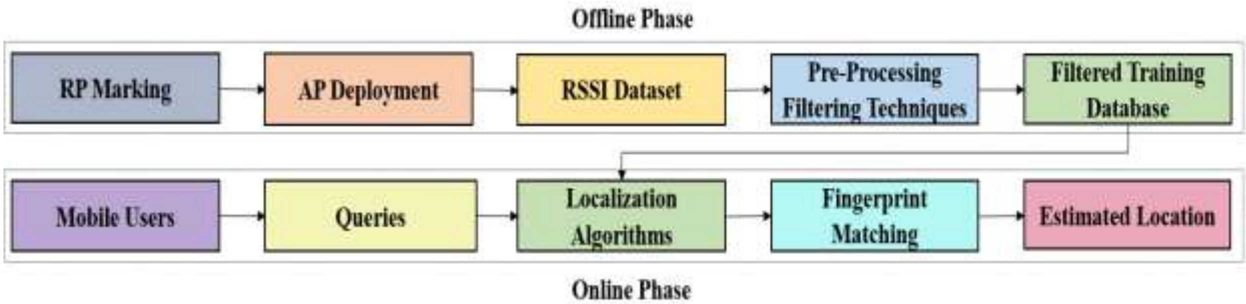

Figure 3.1: Proposed model block diagram

3.2 Pre-Processing Techniques

In the following subsections, the pre-processing techniques of mean and median filtering will be discussed in detail and how these techniques behave to the extreme values in the datasets.

3.2.1 Mean Filtering

Mean filtering is one of the pre-processing techniques that is used to minimize the noise in the RSSI database. It plays a vital role in indoor positioning system as it takes average of the recorded RSSI samples thus, minimizing the effect of environmental factors. Mean filtering is applied on the database before the online phase where machine learning algorithms are used [1-3]. Mean $\bar{X}$ can be defined as:

$$\bar{X} = \frac{\sum_{i=1}^{N} X_i}{\mathrm{N}} \tag{3.1}$$

where $\sum$ X is the sum of all the x values and N is the number of x values.

Consider a case when an extreme value ‘y’ is added to the data set due to noise. Eq. (3.1) then becomes:

$$\overline{X_y} = \frac{\sum_{i=1}^{N-1} X_i + \mathrm{y}}{\mathrm{N}} \tag{3.2}$$

Eq. (3.2) shows that if y > 0 then, $\overline{X_y} > \bar{X}$. However, if y < 0 then, $\overline{X_y} > \bar{X}$. This proves that mean is affected when an extreme value is added to the data set. Therefore, some sort of filtering is required to tackle the extreme values in the data set.

3.1.2 Median Filtering

Median filtering is used to remove outliers from the recorded RSSI values. By applying median filtering, the RSSI values on a current reference point from a particular access point are arranged in ascending order and median is calculated. If the total number of RSSI values is odd then the central value is taken as the median, however, if the number is even then median is the average of the two central values.

If 'n' is odd, then median is given by:

$$\text{Median} = \frac{(n+1)}{2}^{\text{th}} \text{ term} \tag{3.3}$$

However, if 'n' is even, we have:

$$\text{Median} = ((\frac{n}{2})^{\text{th}} \text{ term} + (\frac{n}{2} + 1)^{\text{th}} \text{ term}) / 2 \tag{3.4}$$

Consider two cases when an extreme value 'y' is added to the data set due to noise. Eq. (3.3) and Eq. (3.4) then becomes:

$$\text{Median}_y = \frac{(n+1)}{2}^{\text{th}} \text{term} \tag{3.5}$$

$$\text{Median}_y = ((\frac{n}{2})^{\text{th}} \text{ term} + (\frac{n}{2} + 1)^{\text{th}} \text{ term}) / 2 \tag{3.6}$$

From Eq. (3.5) and Eq. (3.6), we can observe that median remains unchanged when an extreme value is added to the data set. The samples are arranged in either ascending or descending order and extreme values are never used while calculating median. Therefore, median filtering is immune to outliers.

3.3 Machine Learning Algorithms

In the following subsections, machine algorithms will be discussed in detail. The pseudo codes of the proposed machine learning algorithms have also been given for understanding the working of the proposed model.

3.3.1 *k*-Nearest Neighbor (*k*NN) Algorithm

The *k*NN algorithm is a supervised algorithm for machine learning that can solve problems with both classification and regression. It determines the location through the distance between patterns and reference patterns present in the database. A range of formulas, e.g., Manhattan distance or

Euclidean distance, may be used for distance measurement. Depending on their reciprocal distances, it decides the best matching fingerprints and averages the position of the same k patterns. Before that, the reference points relating to these k fingerprints are used to find the predicted position [25]. In order to find the difference between the RSSI vector being assessed and one of the fingerprints, the vectors thus serve as the main component in pattern matching algorithms. We can also say the similarity between the two matching objects is essential. Euclidean distance can be calculated according to Eq. (3.7).

Let (RSS_1, RSS_2, · · ·, RSS_N) represent an RSS vector reported by a mobile station (MS) for localization, where RSS_x represents the RSS value received from the x^{th} AP by the MS. The distance can be calculated as:

$$d_m = \sqrt{\sum_{n=1}^{N} (RSSI_{m,n} - r_n(t))^2} \tag{3.7}$$

where the mean RSS value received at the m^{th} RP ($1 \leq m \leq$ M) and n^{th} AP ($1 \leq n \leq$ N) with M and N representing the total number of RPs and APs, respectively is represented by $RSSI_{m,n}$. The fingerprint associated with the m^{th} RP is represented by the m^{th} row, which has known location in terms of coordinates.

The pseudo code of kNN algorithm for the proposed model is shown below:

Algorithm 1: *k*NN Algorithm

Input:

- Pre-processed training dataset **T**.
- Distance defining function **D**.
- An integer ***k*.**

Output:

- Estimated location.
- Mean position error.

Steps:

For a testing dataset **X**, for which we want to predict its output variable dataset **Y**:

1. Calculate all the distances of this testing dataset **X** with the other fingerprints of the training dataset **T**.
2. Take ***k*** fingerprints from the training dataset **T** close to **X** using the distance calculation

function **D.**

3. Take the values of **Y** from the ***k*** fingerprints taken and calculate the mean of **Y** deductions.
4. Return the values calculated in step 3 as the values that were predicted by *k*NN for testing dataset **X**.

3.3.2 Weighted *k*-Nearest Neighbor (W*k*NN) Algorithm

Weighted *k*NN is a modified version of *k* nearest neighbors. Weight is calculated based on the weighting function which is the reciprocal of the distance between the neighbors. The choice of the hyper parameter *k* is one of the many limitations that influence the *k*NN algorithm's efficiency. The algorithm may become more vulnerable to outliers when the value of *k* is very small. Too many points may be included in the field from other groups when *k* is very high. Another problem is the way to incorporate class names. The simplest technique is for majority voting, however if the closest neighbors are far different from each other and the closest neighbors display the object's class more precisely, it could be a problem [26].

Assuming that there are *M* RPs and *N* APs, the signal strength vector of i^{th} RP is $RSS_i = RSS_{i1}, RSS_{i2}, \ldots, RSS_{ij}, \ldots, RSS_{iN}$, where $i = 1, 2, \ldots, M$ and $j = 1, 2, \ldots, N$ and the fingerprinting database is formed by all the vectors. Suppose $RSS_{un} = RSS_1, RSS_2, \ldots, RSS_j, \ldots, RSS_N$ is the RSSI vector measured from all the APs on the unknown points. The distances between RSS_{un} and all the RSS_i are calculated with Eq. (3.8) in order to obtain the location of this unknown point (UP), where Manhattan distance (or sum of absolute differences, SAD) is represented by $q = 1$ and Euclidean distance (or sum of the squared differences, SSD) is represented by $q = 2$ respectively.

$$D_i = \left(\sum_{j=1}^{N} |\mathrm{RSS}_{ij} - \mathrm{RSS}_j|^q \right)^{1/q}, \quad i = 1,2\ldots\ldots M \tag{3.8}$$

The minimum distance of k is then chosen for the next step in all D_i. An UP's coordinate can be represented as:

$$C_{un} = \frac{1}{k} \sum_{t=1}^{k} C_t \tag{3.9}$$

The coordinates of UP and the corresponding RP are respectively denoted by C_{un} and C_t. The W*k*NN algorithm assigns a weight according to the distance value to each coordinate. The weight W_i of the i^{th} selected RP can be calculated as:

$$W_i = \frac{\frac{1}{D_i}}{\sum_{j=1}^{k}(\frac{1}{D_j})} \ , \quad i = 1, 2 \ldots\ldots k \tag{3.10}$$

For WkNN, Eq. (3.9) is updated as:

$$C_{un} \ = \ \sum_{i=1}^{k} W_i \, C_i \tag{3.11}$$

For the weighted *k*NN, any function may be used as a kernel function whose value decreases as the distance increases. Inverse distance function is the simplest function used for this purpose. The pseudo code of W*k*NN algorithm for the proposed model is shown below:

Algorithm 2: W*k*NN Algorithm

Input:

- Pre-processed training dataset **T**.
- Distance defining function **D**.
- An integer ***k*.**

Output:

- Estimated location.
- Mean position error.

Steps:

For a testing dataset **X**, for which we want to predict its output variable dataset **Y**:

1. Calculate all the distances of this testing dataset **X** with the other fingerprints of the training dataset **T**.
2. Predict the class of the query point, using distance-weighted voting.
3. Take the values of **Y** from the ***k*** fingerprints taken and calculate the mean of **Y** deductions.
4. Return the values calculated in step 3 as the values that were predicted by W*k*NN for testing dataset **X**.

3.3.3 Feature Scaling Based *k*-Nearest Neighbor (FS*k*NN) Algorithm

FS*k*NN algorithm introduces RSSI level-based scanning in order to calculate the effective signal difference between various signal vectors during the corresponding synchronization [25] [29]. d'_m

shows the fingerprint attached to the m^{th} RP and an appropriate signal distance between the online applications in order to measure it by:

$$d'_m = \sqrt{\sum_{l=1}^{L} (RSSI_{m,l} - RSSI_l)^2 * w(RSSI_l)} \tag{3.12}$$

Quantity of the effective RSSI distance is being shown by scaling the weight function $w(.)$. At the RSSI level of $RSSI_l$, one unit of RSSI shift is comparable and it must also be noted that its esteem can differ from the actual RSSI. The effective distance of the signal is computed by Eq. (3.12) in such a way that the relation between the actual signal distance and the geometrical distance is explained in a better way. In a complex indoor environment, it is hard to provide a closed-form expression for $w(.)$ In such a simulation model, the scenario is treated by dividing the entire RSSI plane equally into n ($n \geq 1$) intervals. A single fixed value as a scaling weight for each interval is then calculated through precise measurements and repeated tuning. If the value of n is equal to 1, the FS*k*NN model will be degraded to the *k*NN model. RSSI-value-to-scaling-weight plotting is therefore known as a type of a step function, such as:

$$w(x) = \sum_{i=1}^{N} a_i \; X_i(x) \tag{3.13}$$

where x represents the RSSI value and the scaling weight at RSSI vector (x) for actual difference of the signal is represented by $w(x)$. Let A_i is the i^{th} RSSI break $1 \leq i \leq n$, for that A_i interval with α_i as coefficient. The sign function x_i (x) of the same interval A_i, is expressed as:

$$X_i(x) = \begin{cases} 1, & x \in A_i \\ 0, & x \notin A_i \end{cases}, \quad 1 \leq i \leq n \tag{3.14}$$

If a value of RSSI (x) lying in the interval A_i is collected by the mobile station then, from Eq. (3.14), X_i (x) will be equal to one while all other values, i.e., X_i (x) having position $j \neq i$, equals to zero. After that the accomplished outcome $w(x)$ from Eq. (3.14) (equivalent α_i) is used to measure the distance of effective signal in Eq. (3.13) as a scaling weight for RSSI value x.

The pseudo code of FS*k*NN algorithm for the proposed model is shown below:

Algorithm 3: FS*k*NN Algorithm

Input:

- Pre-processed training dataset **T**.
- Distance defining function **D**.
- An integer ***k*.**

Output:

- Estimated location.
- Mean position error.

Steps:

Step 1: **Offline phase**

- For all RPs on radio map, save Reference files (R_f)
- **w(.)**: tuning coefficient
- $\boldsymbol{R_{f/2}}$: Representation
- **C**: Evaluation
- $\boldsymbol{\alpha_i}$: Optimization

Step 2: **Online phase**

- $\boldsymbol{d_m}$: Euclidean distance × $w(R_{fi})$
- **D**: sorting in ascending order and select "k" R_f
- **P(x,y)**: taking average of set k

Representation, evaluation and optimization are the three stages of FSkNN model. In representation, a filtered training set prepared fingerprints with known coordinates for locating a mobile user in an unknown location. The testing set has been used to evaluate iteratively the localizing efficiency of modified coefficients during various iterations. The sum of distance errors denoted by cost based on testing set was calculated using Eq. (3.15) for each set of coefficients obtained in the evaluation phase.

$$Cost = \sum_{i=1}^{m} \sqrt{(x_i - x'_i)^2 + (y_i - y'_i)^2} \tag{3.15}$$

where m is the total no of features and x_i and y_i are the actual coordinates of the ith element in the testing set. The RPs present in the testing set were taken as anonymous locations in the evaluation process. Larger sum results in greater accuracy of the positioning. The optimal solution is, thus, obtained by using coefficients to make cost = 0. In optimization, new coefficients were searched by simulated annealing (SA) for obtaining better accuracy. Coefficients were changed randomly during each iteration. This whole process continued until the iteration number reached a pre-set maximum number.

3.3.4 Support Vector Machine (SVM) Algorithm

SVM has become popular due to its classification/regression effect, a relatively new multivariate statistical approach. A classic support SVM is a support-vector network that can be used for supervised learning models. This model is a non-probabilistic binary classifier that can be used to characterize the hyperplane that separates the classes in the training set. The projected mark of a formerly overlooked data point can be evaluated by the side of the hyperplane it lands on. [30].

SVM is a powerful supervised learning model that can effectively manage high-dimensional data sets [30]. The training points are support vectors that can categorize any new data point in the network. SVMs are capable of multiclass classification as well as binary classification. Non-linear classification can also be carried out by SVMs, which can help find the hyperplane of a non-linear operating input vector. A high-dimensional space for features may be mapped from an input variable.

A linear support vector regression challenge may be designated as a restricted optimization problem defined as [30].

$$\min_{w, b, \varepsilon} f(w, b, \varepsilon) = \frac{1}{2} w^T w + C \sum_{i=1}^{n} \varepsilon_i$$

$$\text{Subject to } y_i (w^T x_i + b) - 1 + \varepsilon_i \geq 0, \quad i = 1, 2 \ldots\ldots n \tag{3.16}$$

where w is the standard hyperplane vector, b is the hyperplane offset control parameter, ε controls the width of the ε-insensitive zone, used to fit the training data. The number of support vectors used to create the regression function may be affected by the value of ε. The degree of precision is defined by the value of epsilon of the approximated function. It is completely dependent on the target values in the training set. We cannot predict a positive outcome if epsilon is greater than the range of target values. Overfitting can be expected if epsilon is zero. The use of epsilon for a certain precision guarantees the precision in the training set only. But we must select a marginally smaller epsilon to maintain a certain overall precision. Therefore, $\varepsilon = 0.1$ was chosen as it gives the best results as reported in [30]. The degree of the penalty for the violation is defined by the parameter $C > 0$. Besides, the parameter C is a hyperparameter that is chosen either by cross-validation or by Bayesian optimization.

Support Vector Regression (SVR) is a regression model that is unlike any other. The SVM algorithm is used to predict a continuous variable. Other linear regression models try to minimize the error between the expected and real values, while SVR attempts to match the best line under a predefined or threshold error value. The pseudo code of SVM algorithm for the proposed model is shown below:

Algorithm 4: SVM Algorithm

Input:

- Pre-processed training dataset **T**.
- Distance defining function **D**.
- Parameters: Epsilon ($\boldsymbol{\varepsilon}$), Tolerance ($\boldsymbol{C}$).

Output:

- Estimated location.
- Mean position error.

Steps:

1. Choose kernel: kernel type (linear, gaussian).
2. Form Correlation matrix: $\overline{\boldsymbol{K}}$
3. Vector of values corresponding to training set: $\overline{\boldsymbol{Y}}$
4. Evaluate kernel for all pairs of points in the training set and add the regularized results in the matrix.
5. Model training to get contraction coefficients $\boldsymbol{\alpha} = \{ \boldsymbol{\alpha}_i \}$
6. To estimate the unknown value, $\widetilde{\boldsymbol{Y}}$, for a test point $\widetilde{\boldsymbol{X}}$, take inner product of $\boldsymbol{\alpha}$ and the correlation matrix $\overline{\boldsymbol{K}}$.

Chapter 4

RESULTS AND DISCUSSION

In the previous chapter, the working of the proposed model was explained. RSSI database has been constructed in real time environment by physically taking readings at each reference point. This chapter provides the simulation results of the pre-processing techniques used with the investigated machine learning algorithms.

4.1 Simulation Setup

The system parameters for an indoor simulation environment have been listed in Table 4.1. Total area of the real world environment is 108703m^2 along with 933 RPs and 520 APs. The RSSI values are the negative interger values that are measured in dBm (-100 dBm is considered as very weak signal while 0 dBm is an extremely good signal). It is very important to deploy access points at suitable positions so that the RSSI signal from every AP is received at the current RP. When no signal is received at the given RP then that signal is simply replaced by -100 dBm in the database. Each reading represents the real-world coordinates by means of three values, the longitude and latitude coordinates and the building floor.

Table 4.1: Simulation parameters

Parameters	Values
No. of RPs	933
No. of APs	520
Total sample points	21049
Training samples	19938
Testing samples	1111
k in kNN	3
k in WkNN	3
k in FSkNN	5
ε in SVM	0.1
Area	108703 m^2

Figure 4.1 shows the map of the Jaume i university (UJI) Campus where red, green, and blue refer to the multi-floor TI, TD, and TC buildings, respectively. A reference point is represented by the blue dot on the interior of a TI building.

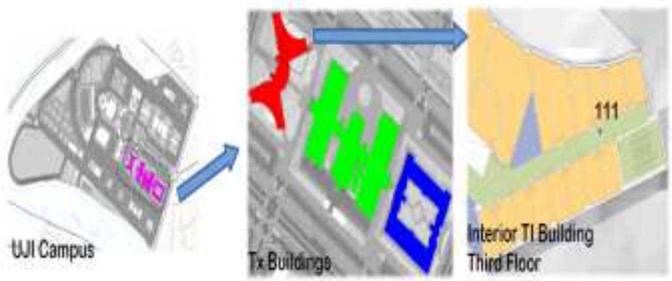

Figure 4.1: Map of UJI campus

More than 20 people used 25 separate mobile devices to gather data [5]. An android application *CaptureLoc* was developed in [5] to take readings for the offline phase. All the required information is gathered by this application and then sent to a centralised server where it is stored permanently. Due to the challenge of the WLAN signal propagation [37], this phase is automatically repeated 10 times for every captured spot. To choose the user identifier and the position where the capture is taken, the user interface framework is essential. Figure 4.2 shows the user-device interaction. Capturing is done (red circle) on the left when the button, *Send Fingerprint*, starts the collect and send procedure. Four errors are reported on the right side as a result of the capturing phase when the location is not captured correctly.

Figure 4.2**:** User-device interaction using *CaptureLoc*

All of the closed spaces of the three buildings (offices, labs, and classrooms) were considered useful areas for collecting in order to construct the training set. Then, for any of the closed spaces considered, an RP is selected inside each space, as well as at least one RP outside each space (i.e., in corridors). The inside point is located in the middle of the closed room, while the outside point is located in front of the door. One RP was chosen for each entry if there are several accesses (door). A graphical example of the positioning and location of the RPs has been shown in Figure 4.3. Red points are the points within closed spaces where RPs taken outside the door are blue points (outside the spaces).

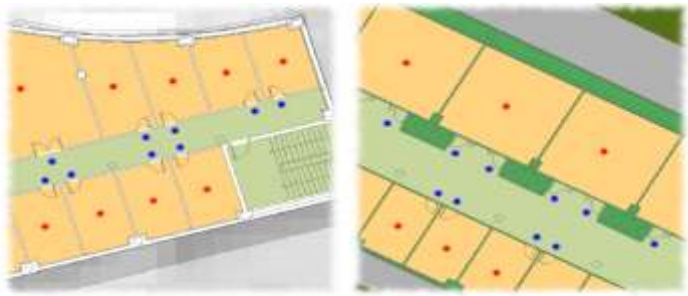

Figure 4.3**:** RP positioning

ValidateLoc, an android application created in [5], to further collect the validation points. Operation phase is performed by the application by sending the necessary information from a centralized server (only APs detected and RSSI levels measured) and it gets a point inside a building (due to its longitude, latitude, floor) from the server. The application validates the location from the user. If the location is correct, the Wi-Fi fingerprint will be sent to the server and is stored with a predicted location. The application would otherwise ask the user for the exact location and the information is submitted to be processed on the server side. An execution example of *ValidateLoc* is shown in Figure 4.4. The first picture displays the location and queries the user if the location is right. The second image tells the user of the successful entry of the fingerprint validation into the server. The blue point represents the expected location whereas, the position assigned to the fingerprint is represented by the green dot.

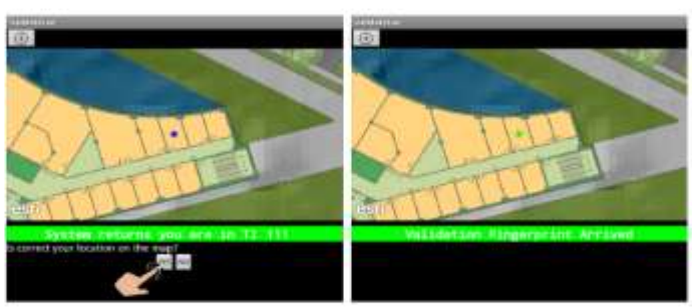

Figure 4.4: *ValidateLo*c phase

4.2 Simulation Results

The simulation results of the pre-processing techniques used with the investigated machine learning algorithms i-e., *k*NN, W*k*NN, FS*k*NN, and SVM are compared with one another. MATLAB software (R2016a) is used for simulation purposes.

The standard metric for performance evaluation of IPS algorithms are localization accuracy and precision. Localization accuracy is defined as the mean position error diverged from actual location whereas distribution of positioning errors is considered as positioning precision [38].

The cumulative distribution function (CDF) of *k*NN, W*k*NN, FS*k*NN, and SVM algorithm with mean, median, and without filtering is shown in Figure 4.5, Figure 4.6, Figure 4.7, and Figure 4.8, respectively.

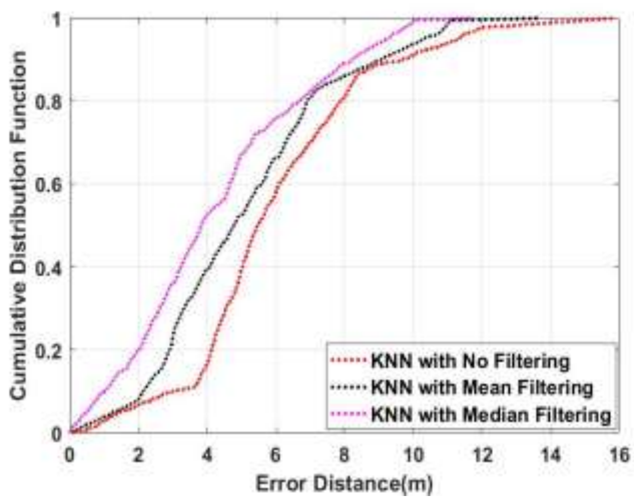

Figure 4.5: CDF plot of *k*NN algorithm with mean, median and without filtering

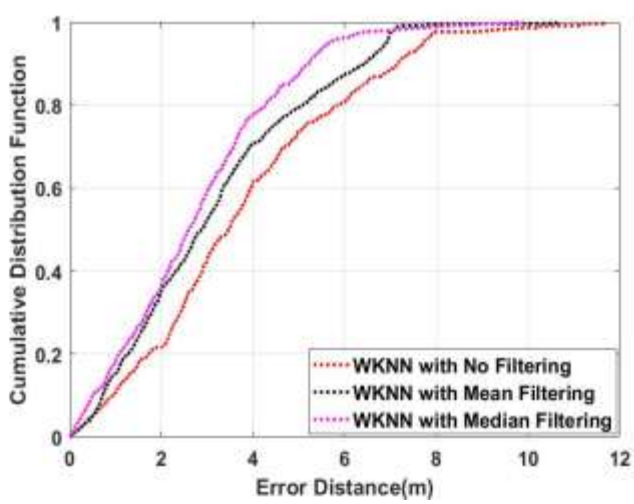

Figure 4.6: CDF plot of W*k*NN algorithm with mean, median and without filtering

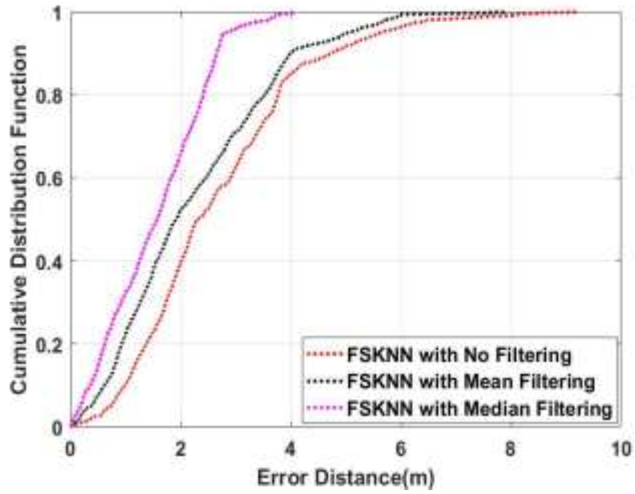

Figure 4.7: CDF plot of FS*k*NN algorithm with mean, median and without filtering

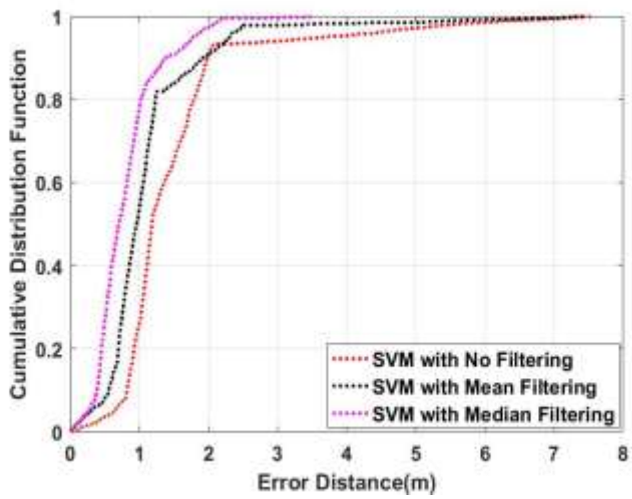

Figure 4.8: CDF plot of SVM algorithm with mean, median and without filtering

*k*NN using median filtering outperforms *k*NN with mean filtering and without filtering as shown in Figure 4.5. The mean position error obtained with median filtering is 4.2581m as compared to the mean position error of 5.0896m and 5.9638m with mean and without filtering, respectively. It is clear that median filtering when used with un-filtered *k*NN, the mean position error is improved by 28.6%.

W*k*NN using median filtering outperforms W*k*NN with mean filtering and without filtering according to Figure 4.6. The mean position error with median filtering is 2.7604m as compared to the mean position error of 3.1511m and 3.7602m with mean and without filtering, respectively. It is clear that median filtering when used with un-filtered W*k*NN, improves the mean position error is by 26.59%.

FS*k*NN using median filtering outperforms FS*k*NN with mean filtering and without filtering according to Figure 4.7. The mean position error obtained with median filtering is 1.5461m as compared to the mean position error of 2.2361m and 2.6743m with mean and without filtering, respectively. It is clear that median filtering when used with un-filtered FS*k*NN, improves the mean position error by 42.19%.

SVM using median filtering outperforms SVM with mean filtering and without filtering according to Figure 4.8. The mean position error obtained with median filtering is 0.7959m as compared to the mean position error of 1.1139m and 1.4791m with mean and without filtering, respectively. It is clear that median filtering when used with un-filtered SVM, improves the mean position error by 46.2%.

The mean positioning error of various machine learning algorithms with and without filtering is shown in Table 4.2. The cumulative distribution function (CDF) of different algorithms using mean and median filtering as pre-processing techniques is shown in Figure 4.9 and Figure 4.10. Results show that when median filtering is used with the machine learning algorithms, it outperforms the mean and no filtering.

Table 4.2: Mean Position Error Comparison

Algorithms	No Filtering	Mean Filtering	Median Filtering
SVM	1.4791m	1.1139m	0.7959m
FS*k*NN	2.6743m	2.2361m	1.5461m
W*k*NN	3.7602m	3.1511m	2.7604m
*k*NN	5.9638m	5.0896m	4.2581m

Applying median filtering on *k*NN, W*k*NN, FS*k*NN, and SVM improves the efficiency in comparison with mean filtering by 16.34%, 12.4%, 30.86%, and 28.55%, respectively. From

Figure 4.9, it is clear that SVM with mean filtering outperforms other machine learning algorithms as it improves the mean position error by 78.12% in comparison with *k*NN.

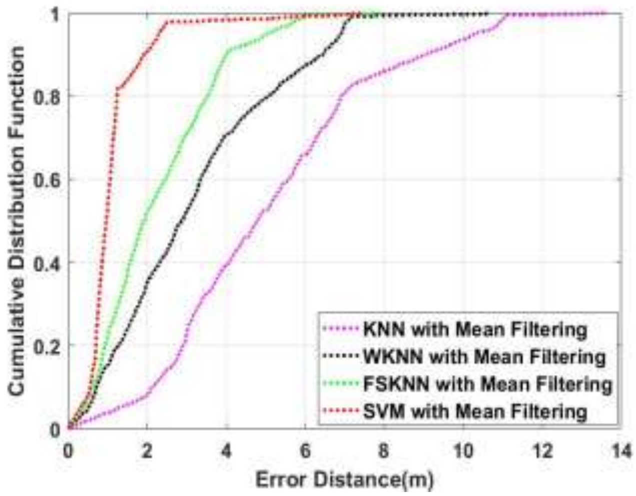

Figure 4.9: CDF comparison of machine learning algorithms with mean filtering

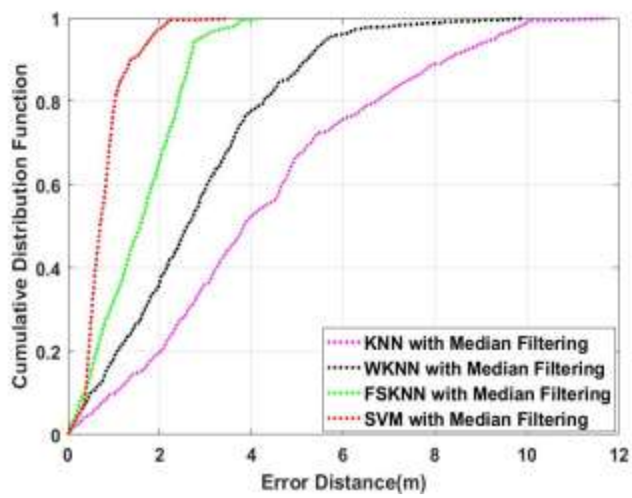

Figure 4.10: CDF comparison of machine learning algorithms with median filtering

Figure 4.10 shows that SVM with median filtering outperforms other machine learning algorithms as it improves the mean position error by 81.31% in comparison with *k*NN.

Overall, the proposed SVM with median filtering algorithm gives the best results in terms of mean position error as it outperforms other machine learning algorithms which are using both mean and median filtering, as depicted in Figure 4.11. It can also be seen from the above results that machine learning algorithms perform better when they are used with the pre-processing techniques of mean and median filtering.

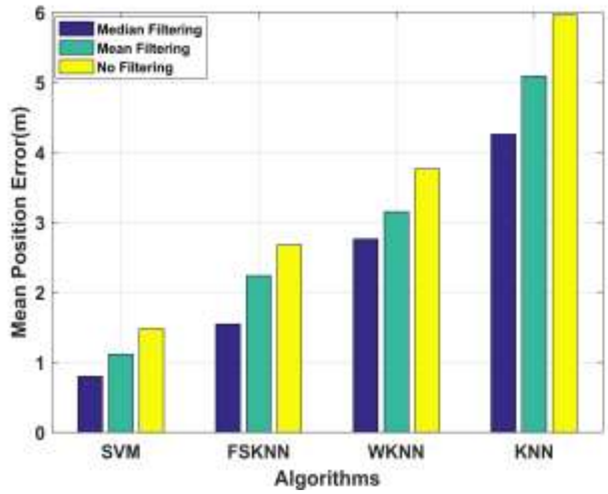

Figure 4.11: Mean position error comparison of machine learning algorithms with mean, median, and without filtering

Determining the value of '*k*' parameter is very important when it comes to *k*NN, W*k*NN, and FS*k*NN as it effects the accuracy of the proposed model. If smaller values of *k* are chosen, the model will learn to predict locally. However, if larger values of *k* are chosen then the model will learn to predict globally. Increasing the value of *k* will improve the mean position error until it becomes constant. The larger values of *k* provide more smoothing which might or might not be desirable [39].

Figure 4.12 depicts the performance of machine learning algorithms for different values of *k* using mean and median filtering.

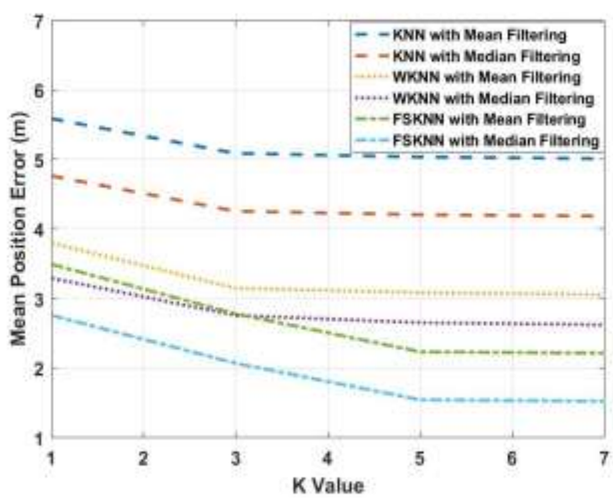

Figure 4.12: Performance of machine learning algorithms using mean and median filtering with varying *k*

It is obvious from Figure 4.12 that k=3 is a suitable choice for both kNN and WkNN algorithms. Moreover, for both mean and median filtering, the mean position error becomes constant when k>3. In case of FSkNN, the mean position error becomes constant when the value of k>5 for both mean and median filtering. Therefore, k=5 can be a suitable choice for FSkNN.

Table 4.3 depicts the efficiencies of the investigated machine learning algorithms with mean, median, and without filtering.

Table 4.3: Efficiency Comparison

Algorithms	No Filtering	Mean Filtering	Median Filtering
SVM	86.69 %	89.57 %	92.84 %
FSkNN	75.93 %	79.87 %	85.68 %
WkNN	66.15 %	71.64 %	74.85 %
kNN	54.19 %	57.25 %	61.67 %

Figure 4.13 shows the efficiency comparison of investigated machine learning algorithms with mean, median, and without filtering in terms of histograms. It is clear that machine learning algorithms perfrom efficiently when they are used with the pre-processing techniques. SVM with mean and median filtering outperforms every other machine learning algorithm with the efficiencies of 89.57 % and 92.84 %, respectively.

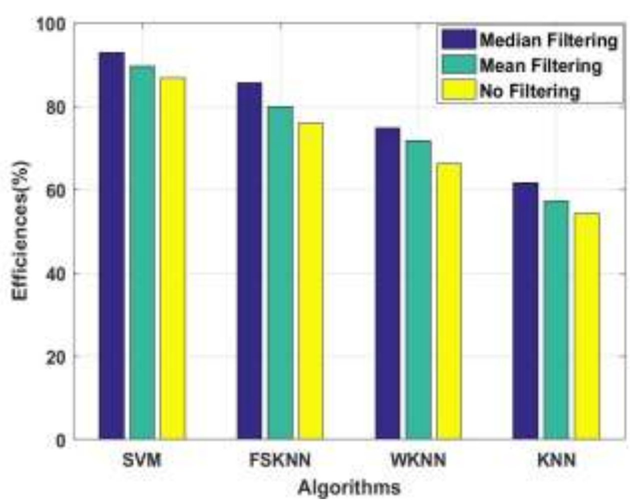

Figure 4.13: Efficiency comparison of machine learning algorithms using mean and median filtering

From the above results and discussion, SVM algorithm outperformed all the other variants of the proposed model as it provided better regularization and generalization capabilities by handling non-linear data efficiently [30]. Similarly, median filtering gave better results as compared to mean filtering because it was immune to the outliers in the dataset [1-3]. Therefore, the proposed SVM with median filtering algorithm enhances the performance of an indoor positioning system.

Chapter 5

CONCLUSION AND FUTURE WORK

The simulation results of the proposed model were seen in the previous chapter. So, the outcomes of the proposed method and potential future work are discussed in this chapter.

5.1 Conclusion

An effective solution for improving the positioning accuracy and efficiency of an IPS was presented in this thesis work. For better outcomes in machine learning problems, large training sets are important. As a result, we used the largest database available online and pre-processed the data using mean and median filtering techniques. To increase the performance of the proposed model, the processed data was fed into machine learning algorithms for training. The aim of the study was to develop an effective indoor positioning approach that not only makes use of the largest database available, but also enhances an IPS's accuracy by reducing the impact of environmental factors. When median filtering was applied on *k*NN, W*k*NN, FS*k*NN, and SVM, the efficiency was improved in comparison with mean filtering by 16.34 %, 12.4 %, 30.86 %, and 28.55 %, respectively. The proposed SVM with median filtering algorithm outperformed other investigated machine learning algorithms with mean and median filtering obtaining a mean position error of 0.7959 m and exceptional efficiency of 92.84 % achieving the research objective.

5.2 Future Work

The positioning of a mobile user can be explored in the future using reinforcement learning algorithms. Furthermore, the proposed approach can be extended to neural network algorithms such as Dynamic Nearest Neighbor and decision tree algorithms such as the Random Forest algorithm to reduce the effect of environmental factors. To minimize positioning errors, Spearman, Minkowski, Chebyshev, and Manhattan distances can be used instead of Euclidean distance with the existing machine learning algorithms. Much larger datasets can also be used to improve the overall efficiency of the proposed model.

REFERENCES

[1] S. Xia, Y. Liu, G. Yuan, M. Zhu and Z. Wang, “Indoor fingerprint positioning based on Wi-Fi: an overview,” *International Journal of Geo-Information*, vol. 5, no. 6, pp. 135, 2017.

[2] W. Sakpere, M. Adeyeye-Oshin and N. B. W. Mlitwa, “A state-of-the-art survey of indoor positioning and navigation systems and technologies,” *South African Computer Journal*, vol. 29, no. 3, pp. 145-197, 2017.

[3] R. F. Brena, J. P. García-Vázquez, C. E. Galván-Tejada, D. Muñoz-Rodriguez, C. Vargas-Rosales et al., “Evolution of Indoor Positioning Technologies: A Survey,” *Journal of Sensors*, vol. 2017, pp. 1-21, 2017.

[4] S. He and S. H. G. Chan, “Wi-Fi Fingerprint-Based Indoor Positioning: Recent Advances and Comparisons,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.

[5] J. Torres-Sospedra, R. Montoliu, A. Martinez-uso, J. P. Avariento, T. J. Arnau et al., "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems", *Proceedings of the 5th International Conference on Indoor Positioning and Indoor Navigation (IPIN 2014)*, Busan, Korea, pp. 261-270, 2014.

[6] R. Want, A. Hopper, V. Falcao, and J. Gibbons, “The active badge location system,” *ACM Transactions on Information Systems*, vol. 10, no. 1, pp. 91–102, 1992.

[7] A. Ward, A. Jones, and A. Hopper, “A new location technique for the active office,” *IEEE Personal Communications*, vol. 4, no. 5, pp. 42–47, 1997.

[8] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, “The Cricket location-support system,” *Proceedings of the 6th annual international conference on Mobile computing and networking*, Boston, Massachusetts, USA, pp. 32-43, 2000.

[9] H. Schweinzer and M. Syafrudin, “LOSNUS: An ultrasonic system enabling high accuracy and secure TDoA locating of numerous devices,” *2010 International Conference on Indoor Positioning and Indoor Navigation*, Zurich, Switzerland, pp. 1-8, 2010.

[10] M. Alloulah and M. Hazas, “An efficient CDMA core for indoor acoustic position sensing,” *In Proceedings of 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Zurich, Switzerland, 2010.

[11] F. Seco, C. Plagemann, A. R. Jiménez and W. Burgard, “Improving RFID-based indoor positioning accuracy using Gaussian processes,” *2010 International Conference on Indoor Positioning and Indoor Navigation*, Zurich, Switzerland, pp. 1-8, 2010.

[12] X. Liu, J. Peng and T. Liu, "A novel indoor localization system based on passive RFID technology," *Proceedings of 2011 International Conference on Electronic & Mechanical Engineering and Information Technology*, Harbin, China, pp. 4285-4288, 2011.

[13] M. Kiers, W. Bischof, E. Krajnc, and M. Dornhofer, “Evaluation and improvements of an RFID based indoor navigation system for visually impaired and blind people,” *In 2011 International Conference on Indoor Positioning and Indoor Navigation*; Guimaraes, Portugal, 2011.

[14] X. Hu, L. Cheng, and G. Zhang, “A Zigbee-based localization algorithm for indoor environments,” *Proceedings of 2011 International Conference on Computer Science and Network Technology*, Harbin, China, pp. 1776-1781, 2011.

[15] J. Larranaga, L. Muguira, J. M. Lopez-Garde, and J. I. Vazquez, “An environment adaptive zigbee-based indoor positioning algorithm,” *Proceedings of 2010 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–8, 2010.

[16] F. Potortì, F. Palumbo and A. Crivello, “Sensors and Sensing Technologies for Indoor Positioning and Indoor Navigation,” *Sensors*, vol. 20, no. 20, pp. 5924, 2020.

[17] V. Moghtadaiee, A. G. Dempster, and S. Lim, “Indoor localization using FM radio signals: A fingerprinting approach,” *2011 International Conference on Indoor Positioning and Indoor Navigation*, Guimaraes, Portugal, pp. 1-7, 2011.

[18] L. Aalto, N. Gothlin, J. Korhonen, and T. Ojala, “Bluetooth and WAP push based location-aware mobile advertising system,” *Proceedings of the 2nd international conference on Mobile systems, applications, and services*, Boston, MA, USA, pp. 49-58, 2004.

[19] G. Fischer, O. Klymenko, D. Martynenko, and H. Luediger, “An impulse radio UWB transceiver with high-precision TOA measurement unit,” *2010 International Conference on Indoor Positioning and Indoor Navigation*, Zurich, Switzerland, pp. 1-8, 2010.

[20] P. Tome, C. Robert, R. Merz, C. Botteron, A. Blatter, and P. Farine, “UWB-based local positioning system: From a small-scale experimental platform to a large-scale deployable system,” *2010 International Conference on Indoor Positioning and Indoor Navigation*, pp. 1-10, 2010.

[21] S. Hosseinzadeh, M. Almoathen, H. Larijani and K. Curtis, “A neural network propagation model for LoRaWAN and critical analysis with real-world measurements,” *Big Data and Cognitive Computing*, vol. 1, no. 1, pp. 7-17, 2017.

[22] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, “Modern WLAN Fingerprinting Indoor Positioning Methods and Deployment Challenges,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 3, pp. 1974-2002, 2017.

[23] A. Sayed. et al. “2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),” *IEEE Signal Processing Letters*, vol. 14, no. 6, pp. 431-431, 2007.

[24] S. Subedi and J. Pyun, “A Survey of Smartphone-Based Indoor Positioning System Using RF-Based Wireless Technologies,” *Sensors*, vol. 20, no. 24, pp. 7230, 2020.

[25] L. Wang, “Research and implementation of machine learning classifier based on KNN,” *IOP Conference Series Materials Science and Engineering*, vol. 677, no. 5, pp. 052038- 052043, 2019.

[26] C. Li, Z. Qiu and C. Liu, “An improved weighted k-nearest neighbor algorithm for indoor positioning,” *Wireless Personal Communications*, vol. 96, no. 2, pp. 2239-2251, 2017.

[27] J. Li, X. Gao, Z. Hu, H. Wang, T. Cao et al., “Indoor localization method based on regional division with IFCM,” *Electronics*, vol. 8, no. 5, pp. 559-578, 2019.

[28] W. Xue, X. Hua, Q. Li, K. Yu and W. Qiu, “Improved neighboring reference points selection method for Wi-Fi based indoor localization,” *IEEE Sensors Letters*, vol. 2, no. 2, pp. 1-4, 2018.

[29] D. Li, B. Zhang and C. Li, “A feature-scaling-based k-nearest neighbor algorithm for indoor positioning systems,” *IEEE Internet of Things Journal*, vol. 3, no.4, pp. 590-597, 2016.

[30] H. Dai, “Research on SVM improved algorithm for large data classification,” *IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, Shanghai, China, pp. 181-185, 2018.

[31] Z. Wang, Z. Wang, L. Fan and Z. Yu, “A hybrid Wi-Fi fingerprint-based localization scheme achieved by combining fisher score and stacked sparse autoencoder algorithms,” *Mobile Information Systems*, vol. 2020, pp. 1-14, 2020.

[32] F. Subhan, S. Ahmed, K. Ashraf and M. Imran, “Extended gradient RSSI predictor and filter for signal prediction and filtering in communication holes,” *Wireless Personal Communications*, vol. 83, no. 1, pp. 297-314, 2015.

[33] S. Han, C. Zhao, W. Meng, and C. Li, "Cosine similarity based fingerprinting algorithm in WLAN indoor positioning against device diversity," *2015 IEEE International Conference on Communications (ICC)*, London, pp. 2710-2714, 2015.

[34] A. E. M. El Ashry and B. I. Sheta, "Wi-Fi based indoor localization using trilateration and fingerprinting methods," *IOP Conference Series: Materials Science and Engineering*, vol. 610, pp. 012072, 2019.

[35] M.S. Choi and B. Jang, "An Accurate Fingerprinting based Indoor Positioning Algorithm," *International Journal of Applied Engineering Research*, vol. 12, no. 1, pp. 86–90, 2017.

[36] Z. Farid, R. Nordin, and M. Ismail, "Recent Advances in Wireless Indoor Localization Techniques and System," *Journal of Computer Networks and Communications,* vol. 2013, pp. 12, 2013.

[37] Y. Wang, C. Xiu, X. Zhang and D. Yang "Wi-Fi indoor localization with CSI fingerprinting-based random forest," *Sensors,* vol. 18, no. 9, pp. 2869, 2018.

[38] Z. Yuan, X. Zha, and X. Zhang, "Adaptive Multi-Type Fingerprint Indoor Positioning and Localization Method Based on Multi-Task Learning and Weight Coefficients K-Nearest Neighbor," *Sensors*, vol. 20, no. 18, pp. 5416, 2020.

[39] Y. Zhao, "An Improved Indoor Positioning Method Based on Nearest Neighbor Interpolation," *Network and Communication Technologies*, vol. 6, no. 1, pp. 1, 2021.