

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

SEMI-SUPERVISED INDOOR LOCALIZATION
USING RSSI FINGERPRINTS

FURKAN TEKTAŞ
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF COMPUTER ENGINEERING

GEBZE
2018

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

SEMI-SUPERVISED INDOOR LOCALIZATION
USING RSSI FINGERPRINTS

FURKAN TEKTAŞ
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF COMPUTER ENGINEERING

THESIS SUPERVISOR
ASSIST. PROF. DR. YAKUP GENÇ

GEBZE
2018

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YARI GÜDÜMLÜ ÖĞRENME
YÖNTEMLERİYLE
RSSI PARMAKIZI TABANLI
BİNA İÇİ KONUMLANDIRMA SİSTEMİ

FURKAN TEKTAS
YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

DANIŞMANI
DR. ÖĞR. ÜYESİ YAKUP GENÇ

GEBZE
2018

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 04/07/2018 tarih ve 2018/35 sayılı kararıyla oluşturulan jüri tarafından 01/08/2018 tarihinde tez savunma sınavı yapılan Furkan Tektaş'ın tez çalışması Bilgisayar Mühendisliği Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI) : Dr. Öğr. Üyesi Yakup Genç



ÜYE

: Doç. Dr. Hasari Çelebi



ÜYE

: Doç. Dr. Mehmet Kemal Özdemir



ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun

/ /2018 tarih ve 2018/ sayılı kararı.

SUMMARY

Locating a person on a map has many location-based services applications ranging from navigation to gaming. When smartphones are used, ideal location estimation systems should use the existing sensors and infrastructure. GSM or WiFi capabilities of such devices offer such conveniences. Receiver signal strength indicators (RSSI) are frequently used for location estimation based on the fingerprint of the signals received from various sources. Machine learning methods are employed to recognize these fingerprints.

Of course, such a recognition method requires labeling of RSSI fingerprints for a grid of locations. This makes the application of these methods expensive and in many cases impossible. RSSI data is collected for a set of predefined locations and labeled with the location tag which in turn is transformed to a global coordinate system. The accuracy and the amount of data is of critical importance for the estimated location accuracy. A typical algorithm needs the collection of multiple RSSIs at different times and grid locations separated within a few meters requiring many days. The dynamic changes in the environment, such as weather conditions, infrastructure modifications, and people movement, make it necessary to repeat this surveying process.

Eliminating the need for labeled data can make RSSI-based localization methods more accessible. We propose a semi-supervised method for applications where the location to be estimated lies on a constraint path such as a corridor or a winding road. In these cases, the data collection process is reduced to recording the data freely and labeling only for a couple of locations. A new algorithm is proposed to classify the RSSI fingerprints for the unlabeled locations.

The method is applicable to RSSI-based localization using GSM as well as WiFi infrastructures. The results show that the proposed method provides comparable results to supervised algorithms such as KNN and deep learning and much better results than traditional clustering methods such as K-means.

Keywords: RSSI, fingerprint-based localization, indoor localization, semi-supervised learning, clustering, classification.

ÖZET

Harita üzerinde bir kişinin konumlandırılması, navigasyondan oyunlara kadar geniş bir yelpazede pek çok konum tabanlı uygulama alanına sahiptir. Akıllı telefonlar kullanıldığında ideal bir sistemden beklenen, mevcut sensörlerin ve mevcut altyapının kullanılmasıdır. GSM ve WiFi teknolojileri bizlere bu imkânı sunmaktadır. Sinyal gücü seviyesi (RSSI), çeşitli kaynaklardan alınan sinyallerin parmak iziyle çalışan konum tahmini yöntemlerinde sıklıkla kullanılmaktadır. Makine öğrenmesi yöntemleri bu parmak izlerinin tanınmasında kullanılmaktadır.

Hiç şüphesiz, bu tip bir tanımlama yapabilmek için RSSI parmak izlerinin belirli bir konumda nokta nokta etiketlenmesi gerekmektedir. Bu durum, çoğu uygulama için masraflı olmakla beraber, bazı durumlarda da imkânsızdır. RSSI verisi önceden belirlenmiş konumlarda toplanıp, daha sonra küresel koordinat sistemine dönüştürülecek şekilde bir konum etiketi ile etiketlenmektedir. Toplanan verinin miktarı ve doğruluğu, yapılacak olan konum tahmininin başarımı üzerinde önemli bir etkiye sahiptir. Tipik bir algoritma, çok sayıda RSSI verisinin günler boyunca sürecektir, birkaç metre aralıklarla farklı zaman dilimlerinde toplanmasına dayanmaktadır. Ortamdaki hava durumu, yapısal değişiklikler ve insan hareketi gibi dinamik değişimler, veri toplama işleminin tekrarlanmasını gerektirmektedir.

Bu tip etiketli veriye olan ihtiyacı azaltmak, RSSI tabanlı konumlandırma yöntemlerini çok daha erişilebilir kılacaktır. Bu hususta, yapılacak olan konum tahmininin düz veya dolambaçlı koridorlarda yapılması şartıyla yarı güdümlü bir yöntem önermekteyiz. Bu şartlar altında veri toplama işlemini çok daha rahatlatacak ve her konumda yapılan veri toplama işlemini sadece birkaç konuma indirebileceğiz. Bu minvalde RSSI parmak izlerini etiketsiz olarak sınıflandırabilen yeni bir algoritma önermekteyiz.

Bu yöntem RSSI tabanlı konumlandırma için hem GSM hem de WiFi sinyallerini kullanabilmektedir. Alınan sonuçlar, yöntemimizin K-En yakın komşu ve derin öğrenme yöntemleri gibi güdümlü yöntemlerle kıyaslanabilir, K-Ortalama algoritması gibi kümeleme algoritmalarından ise çok daha iyi performans sergilediğini göstermektedir.

Anahtar Kelimeler: RSSI, parmak izi tabanlı konumlandırma, iç mekân konumlandırma, yarı güdümlü öğrenme, kümeleme, sınıflandırma.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor Dr. Yakup Genç for his guidance, advice, and support through my graduate and undergraduate studies. He has taught me countless lessons and steered me in the right direction whenever I got lost.

I deeply thank my family for their patience, encouragement, and trust. Their support has been extremely valuable throughout my life and made this dissertation possible.

I owe my earnest thanks to, my best friend and wife, Şeyma for her incessant support, understanding, encouragement, guidance, and advice. She has stood by me all the time, through all my travails. She helped me in every way she can, prevented me from taking wrong turns, and encouraged me when I needed. I couldn't have completed this thesis without you.

Portions of this research used GPU hardware donated by NVidia under Research Grant #1018720312.

TABLE of CONTENTS

SUMMARY	v
ÖZET	vi
ACKNOWLEDGEMENTS	vii
TABLE of CONTENTS	viii
LIST of ABBREVIATIONS and ACRONYMS	ix
LIST of FIGURES	x
LIST of TABLES	xii
1. INTRODUCTION	1
2. RELATED WORK	5
2.1. CHALLENGES FOR FINGERPRINT-BASED ILBS	6
2.2. CELLULAR FINGERPRINT ANALYSIS	7
2.3. WIFI FINGERPRINT ANALYSIS	9
2.3.1. STACKED AUTOENCODERS	9
2.3.2. CONVOLUTIONAL NEURAL NETWORKS	14
3. RSSI FINGERPRINT DATASETS	16
3.1. UJIINDOORLOC DATASET	16
3.2. IPIN 2017 TRACK 3 DATASET	17
3.3. GTU GSM RSSI FINGERPRINT DATASETS	21
4. MACHINE LEARNING FOR POSITION ESTIMATION	24
4.1. SEMI-SUPERVISED ANALYSIS OF GSM FINGERPRINTS	24
4.2. SEMI-SUPERVISED ANALYSIS OF WIFI FINGERPRINTS	29
5. DEEP LEARNING METHODS FOR WI-FI FINGERPRINT ANALYSIS	33
5.1. PREPROCESSING	33
5.2. STACKED AUTOENCODERS	34
5.3. CONVOLUTIONAL NEURAL NETWORKS	35
5.4. LONG-SHORT TERM MEMORY NETWORKS	43
6. CONCLUSIONS	46
REFERENCES	47
BIOGRAPHY	52

LIST of ABBREVIATIONS and ACRONYMS

<u>Abbreviations</u>	<u>Explanations</u>
<u>and Acronyms</u>	
LBS	: Location-Based Services
CNN	: Convolutional Neural Network
DNN	: Deep Neural Network
FCN	: Fully Connected Network
GNSS	: Global Navigation Satellite Systems
GPS	: Global Positioning System
ILS	: Indoor Localization Solution/System
KNN	: K-Nearest Neighbors (classifier)
LOS	: Line-of-Sight
LSTM	: Long-short Term Memory (Network)
NLOS	: Non-Line-of-Sight
PDR	: Pedestrian Dead Reckoning
RF	: Random Forest Classifier
RSSI	: Received Signal Strength Indicator
SAE	: Stacked Autoencoder
SVM	: Support Vector Machines
WAP	: Wireless Access Point

LIST of FIGURES

<u>Figure No:</u>	<u>Page</u>
1.1: Indoor Localization Systems from the device perspective as described in Xiao et al.	2
2.1: General stacked autoencoder structure.	11
2.2: General structure of a stacked autoencoder-based classifier.	13
2.3: Hadamard Product images of the RSSI fingerprints used in the work of Mittal <i>et al.</i> a, b, and c represent the different RSSI samples in different locations.	15
3.1: The survey for the WiFi RSSI fingerprint dataset of the IPIN 2017 conference's competition is done in these buildings. These buildings have different number of floors which might pose a challenge for RSSI-based algorithms and causes imbalance for floor prediction models.	18
3.2: A sample from a log file of the IPIN17-T3 dataset showing different sensors' measurements. For us, only the WiFi measurements are of interest.	20
3.3: CellInfo application screenshots on manual labeling mode. (a) and (b) labels the current RSSI fingerprints in different locations.	22
4.1: First version of the LocationFinder, online location querying application, architecture which communicates the RSSI localization model through a web API using a REST interface.	25
4.2: RSSI observations over time when phone holds still in hand. Lines represent the signal strength change in each location over time.	26
4.3: Semi-labeled GSM dataset survey plan. User starts walking from the corner labeled as 0 at constant pace and labels fingerprints at each corner. User continues surveying for 7-8 times on the specified path.	28
4.4: LMR and K-Means performance comparison of GSM RSSI fingerprints. Y axis represents the ground-truth labels and X axis represents the predicted labels. Note that K-means cannot preserve the order of cluster centroids, while LMR can. K-means' order of cluster centroid ordered by hand for convenience.	29

4.5: Alcalá Tutorial 2017 dataset coordinates are plotted in X-Y axis. Selected corridor is represented in colors and reference points represented with square markers.	30
4.6: LMR results on a WiFi RSSI fingerprinting Alcalá Tutorial 2017 dataset. LMR successfully clusters WiFi RSSI fingerprints in a corridor while preserving their cluster orders.	31
4.7: The proposed LMR algorithm for semi-supervised classification of RSSI fingerprint data.	32
5.1: CSI feature images used in ConFi where a, b, c, d represents CSI features in different locations.	36
5.2: BF-CNN Multiclass building and floor classifier structure where the first line contains the common layers of both outputs and consecutive layers are building and floor outputs, respectively.	37
5.3: BF-CNN - Multiclass classifier for prediction floor and building at the same time.	38
5.4: CoordNet structure which utilizes the building and floor predictions of the BF-CNN and extracts features using common convolutional layers.	39
5.5: CoordNet – a CNN-based latitude and longitude multitask regressor which learns the common features in the earlier layers and estimates latitude and longitude values at the same time.	40
5.6: E2E-Conv: End-to-end multiclass classifier and regressor based on CNN without temporal data.	41
5.7: Coordinate estimation performance on UJIIndoorLoc dataset of E2E-Conv where lines represents the distance between actual and predicted coordinates.	41
5.8: CNN-based temporal localization model architecture with multi class classifier and multi class regressor, tested using IPIN17-T3 dataset.	43
5.9: Multi-layer LSTM-based temporal localization model architecture with multi class classifier and multi class regressor, tested using IPIN17-T3 dataset.	45

LIST of TABLES

<u>Table No:</u>	<u>Page</u>
2.1: Network structures of Stacked Autoencoder-based models.	12
3.1: The statistics of GTU Labeled GSM Dataset.	22
3.2: The statistics of GTU Semi-labeled GSM Dataset.	23
5.1: Comparison and results of Mai Ibrahim, <i>et al.</i> using different techniques for feeding temporal data into CNN.	37
5.2: The comparison of the proposed multitask learning method with others on UJIIndoorLoc dataset.	42

1. INTRODUCTION

For thousands of years, humans have needed to locate himself/herself on the Earth. In order to do so, they have used various methods like referencing their position with respect to the celestial objects. As technology evolves, several devices including the sextant, the astrolabe, and the compass were invented to fulfill the need for more accurate localization. In the digital era, localization and positioning has become a vital problem in our lives.

In the mid 1960's, The United States' Navy started working on a localization system to track the position of their nuclear missile-carrying submarines. This system was developed to be more robust and stable by the U.S. Department of Defense in the 1970's and became the well-known Global Positioning System (GPS) [1]. Since then, localization researches have concluded a satisfactory point thanks to Global Navigation Satellite Systems (GNSS), like GPS and GLONASS.

The most important drawback of the GNSS is that the satellite signals cannot be used inside the buildings as these signals cannot penetrate the floors and walls well. Since the last decade, with the proliferation of smartphones and mobile connectivity, the need for localization for indoor settings have been ramped. According to [2], people spend most of their time indoors and as a result of that, they use their phone indoor for 70% of the time. Data communication is also performed indoor four times more than outdoor and 74% of the adult mobile phone users are reported to use their phones to access navigational services and to reach location information which depend on localization and positioning systems. Therefore, throughout this thesis, we will focus on indoor localization of humans.

Localization researches have divided into three subgroups: indoor localization, outdoor localization and unifying localization system for both indoor and outdoor. Unlike outdoor localization and positioning, the Indoor Positioning Systems (IPS) still have no de-facto standard. Since satellite signals become obsolete in the indoor settings due to various obstacles including walls, furniture, and moving people which block the Line-of-Sight (LOS) transmission of electromagnetic signals. To overcome indoor positioning problem, researchers concentrated on using several other modalities including RF signals (including cellular, WiFi, and FM signals), cameras for vision-based systems, Bluetooth beacons, acoustic systems, inertial sensors for Dead

Reckoning-based solutions, and wireless signal properties for fingerprinting solutions, and so on.

These modalities have been using in different techniques vary from the most straightforward distance calculation methods to complex fusion models. Since indoor environment setup can vary significantly, each technique has its own pros and cons when comparing to the others. For instance, warehouse robot navigation systems may utilize robot's inertial sensors and use dead-reckoning technique for positioning, yet machineries in the environment make these sensors less reliable.

From a device perspective, Location Based Services (LBS) can be grouped into two main categories: device-free and device-based solutions [3] (see Figure 1.1). If a target carries any type of a device, this solution falls into device-based solution, whereas if the target needs to be localized in a scenario which target may not carry a device on itself, we can name these solutions as a device-free localization solution. While target generally is willing to locate himself/herself in the environment in device-based solutions, device-free solutions are preferred especially when target has no intension for being located. Indoor navigation for humans is a good example of device-based Indoor Localization Solution (ILS) which use the device on the person, especially person's phone, to navigate its owner. On the other hand, some of the most popular device-free ILSs are customer and warehouse asset tracking and intrusion detection which do not depend on any device attached to targets.

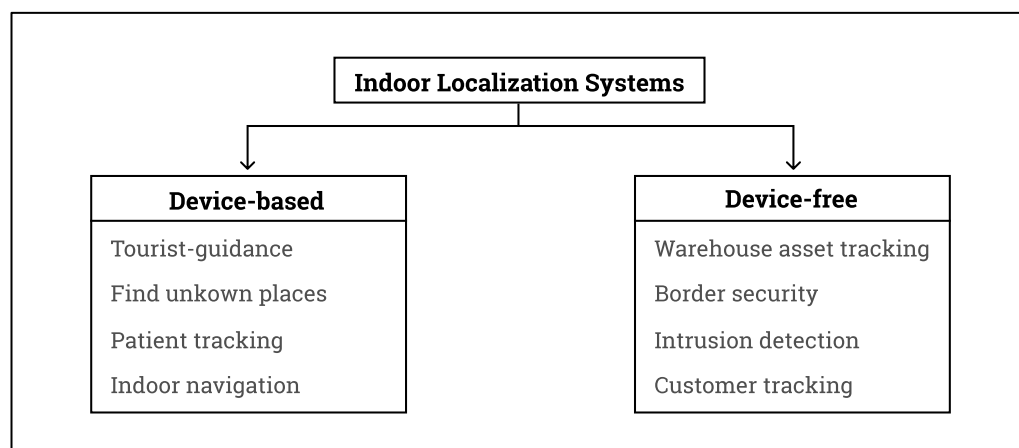


Figure 1.1: Indoor Localization Systems from the device perspective as described in Xiao et al.

There are many criteria when comparing different ILSs. One of the most significant criteria is, without a doubt, its cost. Cost of a localization system often

depends on whether an ILS utilizes the existing infrastructure or requires a new one. Infrastructure-based ILSs may require either modifications to the current infrastructure or using new devices which will not be feasible in some scenarios due to their possible expensive installation costs. On the other hand, infrastructure-free methods, like dead-reckoning and fingerprinting-based methods, employ the current infrastructure and devices. Therefore, for most of the time, infrastructure-free methods are cheaper regarding the installation costs. On the contrary, long-term maintenance costs of the infrastructure-free ILSs may be more expensive depending on the techniques used in the ILS.

Energy efficiency, installation costs, maintenance costs, scalability, and accuracy, are the commonly-used criteria for ILSs [3]. None of these criteria can be considered independently from each other since one affects the others. For instance, acoustic positioning systems can provide high accuracy, yet they are often less scalable than other systems. Inertial sensors of a device are much more scalable than acoustic, but it suffers from the long-term stable positioning accuracy due to the accumulation of sensor errors.

Fingerprinting-based positioning systems employ Received Signal Strength Indicator (RSSI/RSS) which is the average measured signal power for a specific amount of time. RSSI is often reported by the device drivers as a decibel of milliwatts, dBm, as denoted in Equation 1.1 and have a value between 0 and -100. RSSI fingerprinting methods can utilize different modalities including Bluetooth, GSM, CDMA, LTE, or Wi-Fi signals. RSSI is a location dependent property and is affected by the transmission power of a modem and the attenuation of the specified channel.

Fingerprinting-based systems consist of two phases: offline data collection and online localization phase. Data collection phase aims to record every possible RSSI values for each position to create a radio map of the environment which defines the behavior of the RSSI fingerprints of the environment. Creating a radio map is often a labor-intensive task as it requires to cover all positions in all possible conditions such as different human densities and weather conditions during data collection. On the online phase, users query their last recently observed RSSI fingerprints to the localization model based on the radio map created earlier to position themselves in the environment.

$$RSSI (dBm) = 10\log \frac{\text{Received Signal Strength in dBm}}{1mW} \quad (1.1)$$

With the proliferation of machine learning and deep learning methods, fingerprint-based indoor localization made enormous progress. Early methods employed K-Nearest Neighbor (KNN) algorithms [4–6] and Support Vector Machines (SVM) [7,8] to discriminate fingerprints on the radio map. In the meantime, tree-based methods like Random Forest [9] also showed improvement over the existing methods. As deep learning techniques performed quite well in computer vision, pattern recognition, and signal processing areas, indoor localization researches [10,11] also concentrated on utilizing these techniques.

The scope of this theses focuses on the infrastructure-free and device-based indoor localization solutions using fingerprint analysis. We target smartphones users and use commodity smartphones' sensors to build applicable ILS. To navigate users indoor, we aim to a have mean localization error to be less than 10 meters. In this regard, we inspect different techniques and propose new solutions.

2. RELATED WORK

Since the last decade, fingerprinting-based ILBS approaches have come a long way in terms of increasing the accuracy and reducing the cost in all means. ILBS can be divided into different groups from different perspectives. For instance, the methods utilize Angle of Arrival (AoA), Time of Arrival (ToA), Time difference of Arrival (TDoA), Time of Flight (ToF) use the trilateration technique, the technique used in GNSS, and require additional hardware. These methods generally offer 3D position estimation using a modality like ultrasonic, infrared, ultra-wideband (UWB). In addition, other methods which use low-energy Bluetooth beacons, ZigBee devices, WiFi signals, etc. can estimate target's position generally in 2-dimensional space and can be cheaper for mass deployment than 3D position estimation methods.

In localization techniques, fingerprinting may refer four different types of fingerprints as stated in [12]. Which are:

- Visual Fingerprinting
- Motion Fingerprinting
- Signal Fingerprinting
- Hybrid Fingerprinting

Although [12] groups fingerprinting types for outdoor localization, the same taxonomy can be applied for indoor localization, too. Visual fingerprinting methods use device's camera to detect the position of a landmark and infers the device's position with respect to that landmark. Motion fingerprinting utilizes the device's inertial sensors like gyroscope, accelerometer, and compass. The pedestrian dead reckoning methods rely on motion fingerprints. Signal fingerprints expose the signal's behavior from different RF sources including Wireless Access Points (WAPs) and base stations. The hybrid fingerprinting method combines at least two fingerprinting methods to eliminate each method's drawback by combining the powerful features of each.

2.1. Challenges for fingerprint-based ILBS

Indoor localization researches focused on fingerprinting methods due to it leverages the current infrastructure which highly reduces the installation costs. Fingerprinting methods also gain popularity since they are widely adoptable even for unifying localization for both indoor and outdoor.

WiFi RSSI fingerprinting is one of the most important signal fingerprinting methods for indoor localization due to the proliferation of both WiFi infrastructure and smartphones. Since most of the time smartphones connected to wireless modems indoor, WiFi RSSI fingerprinting methods become ubiquitous as they are low cost, energy-efficient and offers high positioning accuracy. Yet, like many of the device-based ILS, RSSI fingerprinting solutions, both WiFi and other RF signals, are vulnerable to:

Multipath effect: The multipath effect is one of the most important problems of ILS since indoor venues contain multiple surfaces with different materials such as doors, humans, walls, furniture, and so on. When radio signals propagate, signals hit these surfaces and reflected just like the light waves. These reflections cause to change in angle, time and strength of the received signal. This situation misleads the device on the target as there may be multiple reflections between the transmitter device and the target device. Therefore, as we don't know which path signal propagated, it named as multipath effect. It does not only affect fingerprinting methods but also deteriorates the performance of other methods like AoA and ToA.

Human density: Since %70 of our bodies are water, human presence in an environment changes the signal propagation as the human body absorbs RF signals. Signal strength can decrease up to 10 dBm when a mobile device blocked by a human body [13].

Environmental changes: Indoor environment setup changes all the time. Furniture can be moved around, replaced, or removed at any time, windows and doors can be opened or closed, human density changes in different hours of a day. These all have an effect on received signal strength. For instance, opening a door can cause either signal to reflect the signal on a different path and alters the multipath effect in the environment or interfere with the signal propagation between device and

transmitter. These environmental setup changes have a dramatic effect on ILS, especially on methods rely on radio maps.

Smart transmitters: As the wireless communication technology advances, the transmission power of WAPs and CDMA stations can vary under different circumstances. For instance, some WAPs transmission power can be altered by a web interface or CDMA base stations change their transmission power according to connected client number in order to accommodate fluctuations in their signals and optimize their power consumption. In this thesis, we did not take such effects into account.

Weather conditions: Since the temperature and humidity affect the signal propagation, seasons, air conditioning, and daily temperature changes can alter the observed signal strength.

Non-Line-of-Sight (NLOS): Unlike outdoors, most indoor spaces are divided into floors, rooms, offices, and halls. Therefore, most of the time RF signal receivers do not have a Line-of-Sight (LOS) path between them and transmitters. Since signals can propagate through floors, walls, and other obstacles, these transmitters can be observable even from a Non-Line-Of-Sight path. Yet, different materials distort the signal propagation at different levels, so that signal strength can change dramatically independent from the distance between the WAP and the device itself.

RSSI calculation differentiation: Virtually all fingerprint-based solutions use dBm for RSSI representation. Since converting milliwatts to dBm requires high precision hardware and this conversion formula vary from manufacturer to manufacturer, RSSI observations can change when different manufacturer's hardware is used. To make fingerprint-based ILBS more robust, researchers generally use different hardware during the radio map survey.

2.2. Cellular Fingerprint Analysis

Cellular signals attract researcher's attention due to their wide-coverage. First publicly available digital cellular network, GSM, began developing in the early 1980s to fulfill the need for unifying and extensible cellular telephone service which offers a decent voice quality and it has been actively using since the early 1990s. GSM has the most widely adopted cellular network which spans every continent, and it has a good coverage in the places where people live. This wide coverage makes GSM and other

cellular networks a good fit for unifying localization systems for both outdoor and indoor.

However, GSM fingerprint solutions do not offer a fine-grained location accuracy, especially for indoors. This is because commodity cell phones report only the connected base station's signal strength with optional six strongest base station's signal strength. This is called narrow fingerprinting [14] and makes GSM based fingerprinting techniques less accurate. Another drawback of the GSM fingerprinting, although GSM signals should have a spatial variation instead of temporal variation, is even a stationary client may observe more than 10dBm temporal variation in signal strength and this variation can be bigger if the client is mobile [15].

GSM-based fingerprinting techniques are not popular for indoor localization. Most of the work which aims to bring applicable solutions with GSM signals focus on outdoor localization since the localization accuracy can be as high as tens of meters, even hundreds in rural areas. [15] proposes an outdoor localization solution using GSM signals. This solution offers a typical localization server-client architecture where localization server stores the radio map. They reported 44-meter localization accuracy in 67th percentile.

While commercial phones can report up to six additional base station's signal strength, [14] employed 29 more base station's signal strength to increase localization accuracy in multi-floor buildings in different indoor settings. Their method uses weighted K-Nearest Neighbor (KNN) classifier and detects floors with the accuracy between 89% to 97%. Their reported floor-level median localization error is between 2.48m to 5.44 meters in multi-floor buildings. They found out that wide signal strength fingerprints, additional RSSI fingerprints from 29 base stations, enable GSM to be used as a widely adoptable indoor localization modality.

On the other hand, CDMA, which was introduced in 1995, is not suitable for signal strength fingerprints as GSM is since CDMA can alter the transmission power of the signals (i.e. cell size) according to the load. Instead, CDMA offers signal delay between the expected signal arrival time and the actual signal arrival time. [16] proposes a new method which uses these signal delays as CDMA fingerprints. They achieve median localization accuracy between 4.5 and 6.7 meters. They also reported floor detection accuracy for GSM, CDMA and WiFi modalities as 84%, 87%, and 100%, respectively.

2.3. WiFi Fingerprint Analysis

All infrastructure-free ILSs try to employ current infrastructure wherever possible to eliminate installation and maintenance costs as well as to support scalability, easily. WiFi infrastructure became a ubiquitous internet infrastructure inside the buildings like offices, campuses, houses, and multi-floor facilities. Due to WiFi signals are not good at the penetration of indoor obstacles like furniture, walls, and floors, earlier works similar to outdoor localization solutions, which rely on a line-of-sight path, like angle-of-arrival techniques, did not perform well in the indoor venues. Thus, WiFi fingerprinting-based techniques became one of the most popular ILS modalities.

In the earlier days of WiFi fingerprint analysis, many researchers tried different machine learning techniques on WiFi RSSI fingerprint analysis. These works tried different representation techniques of RSSI fingerprints including preprocessing steps like eliminating the irrelevant WAPs from the dataset, weighing the most significant WAPs, finding a distance metric between fingerprints, and so on.

The most recent works have focused on deep learning-based techniques, as their performance in other fields exceeded state of the art results by far [17]. In the following pages, we will inspect deep learning-based ILS with keeping our focus on RSSI fingerprint analysis.

2.3.1. Stacked Autoencoders

Stacked Autoencoders (SAEs) are a popular unsupervised learning method. It learns a latent representation of the inputs and helps to reduce the dimensionality of the input data without sacrificing the features inside it. Since RSSI fingerprints are sparse and spatially-related data, we can represent these fingerprints in a much smaller latent space using stacked autoencoders.

The input of the SAE is the signal strength vector for all visible WAPs in a dataset. SAE contains multiple hidden layers, generally, each successive layer has less number of hidden units until the innermost layer. Innermost layer learns the latent representation of the RSSI fingerprint in a much smaller dimension, like $1/4$ or $1/64$ of the input dimension. There are some cases in which vector dimension is not reduced or even increased in other fields, yet there is no such case in the indoor localization

literature. For the sake of indoor localization use cases, we will only cover the general structure used in [18–20].

As can be seen in Figure 2.1, stacked autoencoders have a symmetric structure. The left-hand side is called encoder, which is responsible for encoding the input data, RSSI fingerprint in our case, to lower dimensional embedding vector, i.e., latent representation. The right-hand side is called the decoder, which reconstructs the input data from the latent representation.

To train an SAE network, some non-linear activation functions, like ReLU [21] and hyperbolic tangent, are applied after each hidden layers. Network’s success is examined using its reconstruction accuracy. To calculate reconstruction accuracy, [19,20] used mean squared error (see Equation 2.1) . Mean squared error calculates the squared difference between the input vector and reconstructed in element-wise. In Equation 2.1, X_i denotes i^{th} WAP signal strength, and R_i denotes reconstructed i^{th} WAP signal strength.

$$MSE = \frac{1}{N} \sum_{i=1}^N (X_i - R_i)^2 \quad (2.1)$$

Since RSSI fingerprint samples consist of mostly missing data, they have a sparse nature despite their high-dimensionality. For instance, UJIIndoorLoc dataset’s validation set has 16 AP observations out of total 520 APs on average [22]. This sparsity not only makes compression of RSSI fingerprints possible but also makes compression efficient even in a much smaller space to avoid the curse of dimensionality [23]. To exploit these benefits, several works applied Stacked Autoencoders to reduce the dimensionality of the input space, dramatically.

SAE first proposed in [18] to compress a 520-dimensional RSSI fingerprint to a 128-dimensional embedding vector as can be seen in Table 2.1. They aimed to predict building and floor using the embedded representation of RSSIs to support scalable infrastructure in ILS.

This work is followed by [19] which compresses the same dataset into much smaller, i.e. 8-dimensional, embedding vector (see Table 2.1) to achieve similar accuracy for the same goal as [18].

The latest SAE-based work introduced floor level location estimation as well as building and floor prediction [20]. They convert coordinate estimation problem to a classification problem by enumerating the unique coordinates to use them as a categorical variable. Actually, dataset provided a categorical variable, place identifier, for each unique coordinate, but in this work, researchers enumerated these identifiers within each floor of each building to reduce the maximum number of unique place identifier per floor. Thus, they can represent floor-level coordinates with a categorical variable with 110 different values. This method enabled researchers to predict building, floor and place id at the same time which is a good example of multi-class classification.

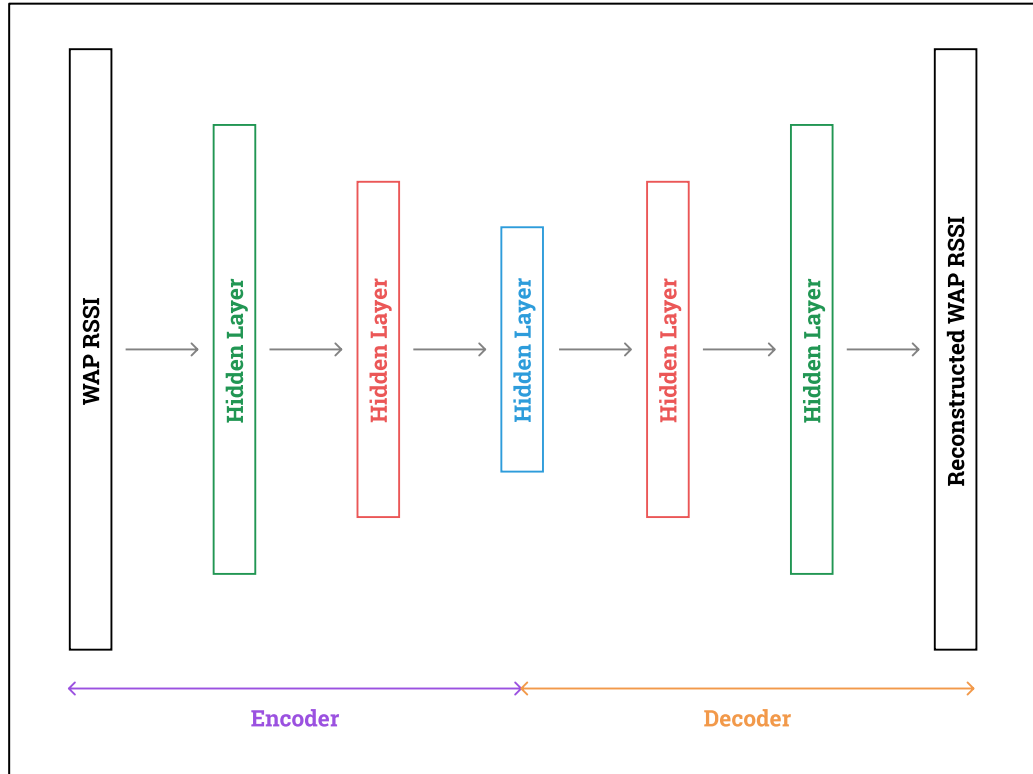


Figure 2.1: General stacked autoencoder structure.

Both [18,20] focused on reducing the dimensionality of the sparse RSSI fingerprint vectors. Dimension reduction using an SAE can make these works valuable but encoding the input vector into a smaller dimensional vector is not a localization system. So that, they added a fully connected network (FCN) after the encoder part by removing decoder part of the SAE, as can be seen in Figure 2.2, to employ encoded features for localization. This FCN contains multiple densely connected layers like the

decoder and optional dropout layers. The difference between the FCN's and decoder's output layer is that the FCN's output constructs an N-hot vector with the desired output size to represent building and floor predictions.

Since SAE-based classifiers contain two different networks, one for learning the latent representation of RSSI fingerprints and the other for classification of these latent representations, these networks' training process needed to be done separately. SAE network is generally trained using mean squared loss function [24] with Adam optimizer [25]. After a satisfactory reconstruction accuracy, decoder network was replaced with an FCN, then the network becomes a classifier, SAE-based classifier in particular.

All of the SAE-based classifiers were trained using Adam optimizer with different loss functions. For instance, [18] preferred categorical cross entropy, whereas [19,20] preferred binary cross entropy loss function. The number of layers and the dimensionality of each layer in the FCN vary and can be seen in Table 2.1.

Table 2.1: Network structures of Stacked Autoencoder-based models.

[18]	$520 \times 256 \times 128 \times 256 \times 520$	SAE for building/floor classification
	$520 \times 256 \times 128 \times 128 \times DO \times 128 \times 15$	SAE-based classifier for building/floor classification
[19]	$520 \times 64 \times 8 \times 64 \times 520$	SAE for building/floor for building/floor classification
	$520 \times 64 \times 8 \times 8$	SAE-based classifier for building/floor classification
[20]	$520 \times 256 \times 128 \times 256 \times 520$	SAE for building/floor classification and location estimator
	$520 \times 256 \times 128 \times 64 \times DO \times 128 \times 118$	SAE-based classifier for building/floor classification and location estimator

Apart from loss functions, the outputs of the classifiers are also different in these networks. An enumeration of the cartesian product of all possible building and floors was used in [18]. The output was expected to be a one-hot vector where only one

element is activated and rest of the elements of the vector are not, i.e. they represented as 1 and 0 in the output vector, respectively. In order to generate an output like this, SoftMax activation layer was placed just before the output layer in [24].

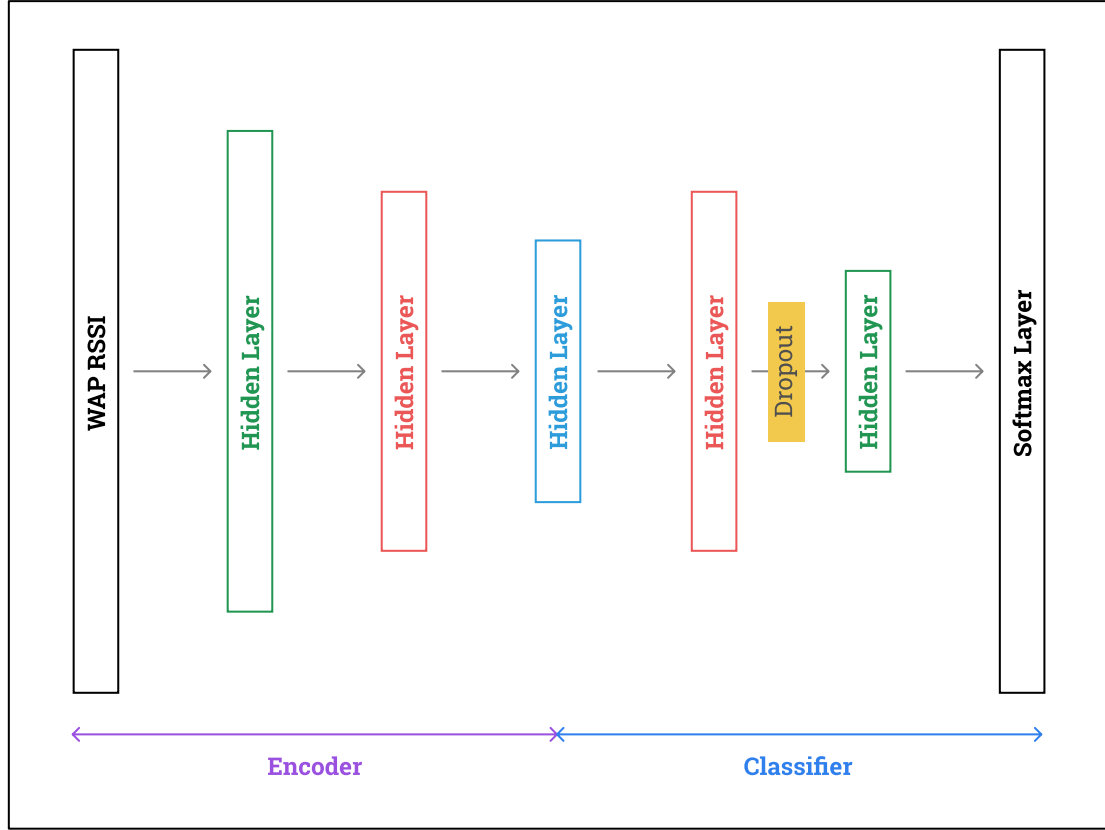


Figure 2.2: General structure of a stacked autoencoder-based classifier.

Other SAE-based localization networks used a different output strategy [19,20]. They converted the same problem to a multi-label classification problem which more than one element of the output vector can be activated at the same time. To do so, they concatenated the one-hot vectors of the expected building, floor, and enumerated place identifier outputs and trained the network with binary cross entropy loss which is commonly used in multi-label classification problems.

Since these classifiers might overfit the training data, they employed random dropout layers between the hidden layers of the FCN. Dropout layers help network to be more generalizable by dropping the connections between densely connected layers randomly on each batch. Since such a regulation prevents model to memorize all possible training inputs, test set performance will be increased by adding one or more dropout layer with appropriate dropout rates.

2.3.2. Convolutional Neural Networks

Convolutional Neural Networks (CNNs) drew a significant attention for their outstanding performance in computer vision [17,26–28], video analysis [29,30], and natural language processing [11,31]. CNNs take locality into account, so that the spatial nature of data can be exploited, easily. For example, in face detection, the presence of eyes, nose and mouth is a very strong cue for face in an image. CNNs have been reported to do improve object recognition results by more than half over the conventional methods [10].

As RSSI fingerprints contain spatial information due to its nature, CNNs are a good fit for employing the spatial information for ILSs. Although CNNs seem convenient for ILS systems, there is not much research has conducted yet. The first work using CNN using WiFi signals is called ConFi [32]. In this work, authors formed a time-frequency matrix of the Channel State Information (CSI) then converted this matrix to an image. Using this generated image, they trained a model with three convolutional layers, followed by a two fully connected layers.

As of July 2018, only two works utilized CNNs to RSSI fingerprint-based ILS were published, which are [33], [34]. The former uses the UJIIndoorLoc dataset as the SAE-based ILSs do, but it differs by using temporal features of dataset. Authors of this paper tried three conventional approaches of feeding temporal data to neural network, which are:

- Concatenating samples into one vector
- Averaging consecutive samples into one vector
- Feeding consecutive samples as a matrix

Their contribution in comparing three different approaches of temporal data usage and being first at showing promising results on temporal data usage are valuable. Our work in this thesis primarily focused on utilizing temporal behavior of RSSI fingerprints to build more robust ILS to temporal fluctuations and sensor noises. So, this work also supports this thesis by showing how valuable the temporal data is.

Another work which was published in 2018 is [34]. Their method involves serious preprocessing steps before training the network. They selected target WAPs in

the survey site beforehand, so that they can collect RSSI fingerprints from only these well-known WAPs to reduce the effects of mobile and temporary WAPs on their solution. In addition, they also eliminated redundant WAPs by calculating Pearson Correlation Coefficient (PCC) [35] using all observations for each WAP after the surveying process. Then they removed WAPs with absolute PCC smaller than 0.3.

$$HP = \sum_{i=1}^N R_i \circ C \quad (2.2)$$

After these preprocessing steps, they calculated Hadamard Product [36] defined in Equation 2.2 where R_i represents normalized RSSI fingerprint at i^{th} location and C represents the PCC coefficient. Then they formed a square matrix using the HP of the RSSI fingerprints, then scaled the values inside this matrix between 0 to 255. If the sizes of matrixes on different locations differ, they filled with the remaining matrix elements with zero. Then they converted these matrixes into grayscale images which can be shown in Figure 2.3.

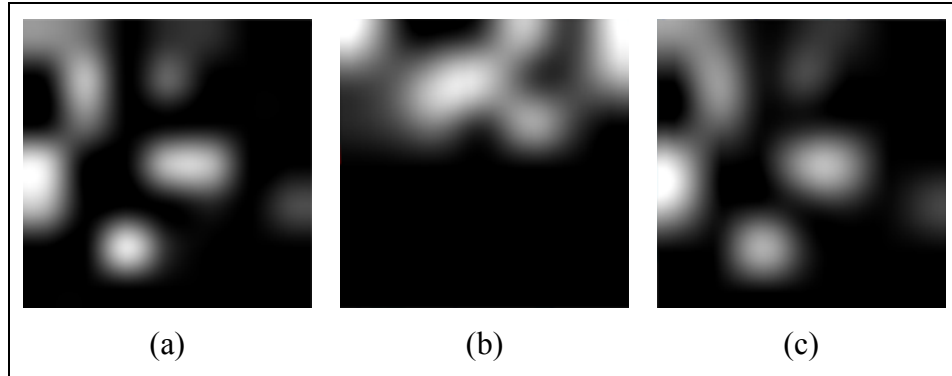


Figure 2.3: Hadamard Product images of the RSSI fingerprints used in the work of Mittal *et al.* a, b, and c represent the different RSSI samples in different locations.

3. RSSI FINGERPRINT DATASETS

Like many research fields in computer science, indoor localization and positioning area also suffered from common evaluation criteria between different techniques. This need gave birth to common datasets including UJIIndoorLoc [37], IPIN Competition datasets [38,39], Tampere University Dataset [40], The Alcalá Tutorial 2017 data set [41].

The Evaluating Ambient Assisted Living initiative (EvAAL) [42] was initiated the indoor localization competition in 2011-2013 [43]. The International Conference on Indoor Positioning and Indoor Navigation took over the indoor localization competition using EvAAL framework in 2014. In the 2017 competition, there were four different tracks on this competition, which were:

- Track 1: Smartphone-based (on-site)
- Track 2: Pedestrian Dead Reckoning Positioning (on-site)
- Track 3: Smartphone-based (off-site)
- Track 4: PDR for warehouse picking (off-site)

In this thesis, due to we are investigating the temporal information effect on fingerprinting-based ILS, we selected two of the publicly available WiFi RSSI fingerprint datasets: UJIIndoorLoc and IPIN 2017 Competition Track 3 datasets. These datasets are multi-building, multi-floor datasets which include timestamp and real-world coordinates. We will dive into details of each dataset we have used in the following sections.

3.1. UJIIndoorLoc Dataset

UJIIndoorLoc is the first publicly available Wi-Fi RSSI fingerprint dataset, which was published by Torres-Sospedra et al. from Universitat Jaume I on the UCI Machine Learning Repository¹ and used as the benchmark dataset on EvAAL-ETRI, Wi-Fi Fingerprint localization competition, at IPIN 2015 conference. UJIIndoorLoc

¹ <http://archive.ics.uci.edu/ml/datasets/UJIIndoorLoc>

consists of three buildings which have least four floors inside Universitat Jaume I campus. It spans almost 110 thousand meter-square and contains the following information for every RSSI fingerprint:

- Building identifier (3 buildings, starting from 0)
- Floor number inside the building (5 floors maximum, starting from 0)
- Unique user identifier
- Device identifier (same for the same device model)
- Place identifier (predefined coordinate identifiers)
- Relative place position (in or out as binary)
- Timestamp (some devices reported erroneous timestamp)
- Latitude and longitude (in meters with UTM)
- Signal strength of 520 different Wireless Access Points (WAPs)

Dataset is represented in tabular form and contains 19337 training and 1111 test samples. Test observations do not include the user identifier, place identifier, and relative place position information. Signal strength information (RSSI) represented in dBm and missing values (i.e., WAPs which are not present during observation) was denoted with 100.

Besides, all of the RSSI samples of UJIIndoorLoc dataset are timestamped, which enables researchers employ temporal behavior of RSSI fingerprints. Since UJIIndoorLoc dataset contains real-world coordinates, in meters with UTM from WGS84, researchers can use regression models to predict each RSSI sample's position [33]. Place identifiers can be considered as a categorical variable for survey coordinates which makes the usage of classifier-based models possible [19]. Dataset aims to predict building and floor of the RSSI sample as well as the real-world coordinates. The correct prediction of both building and floor is named as “success rate” in dataset's paper, and we will follow this naming convention hereinafter.

3.2. IPIN 2017 Track 3 Dataset

A larger data set is made available by EvAAL competition which had the following three tracks in 2015:

- Track 1: Smartphone based positioning
- Track 2: Foot-mounted pedestrian dead reckoning positioning
- Track 3: Wi-Fi fingerprinting in large environments (off-site)

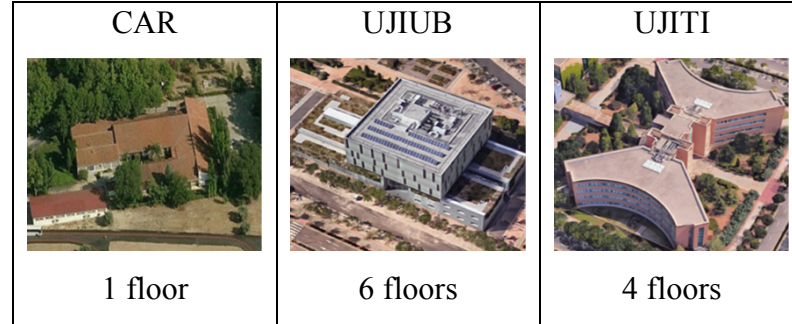


Figure 3.1: The survey for the WiFi RSSI fingerprint dataset of the IPIN 2017 conference's competition is done in these buildings. These buildings have different number of floors which might pose a challenge for RSSI-based algorithms and causes imbalance for floor prediction models.

The track 3 competition has the same purpose with UJIIndoorLoc dataset. In the following year, the track 4, named as indoor mobile robot positioning, was added to the competition and track 2 was renamed to “Pedestrian Dead Reckoning Positioning”.

The International Conference on Indoor Positioning and Indoor Navigation Track 3: Smartphone-based dataset (IPIN17-T3) collected in a natural way as much as possible [44]. Therefore, during the survey, user conducting the survey walked at a constant speed and stopped occasionally. The user also walked laterally or backward in certain coordinates, made turns, even U-turns, and changed floors using stairs. Phone position is also kept natural and for most of the time, user held the phone in front of his/her chest or face. Additionally, user sometimes carried the phone near his/her ear pretending a phone call and hold the phone inside the hand while arms were downwards. Since dataset contain multiple modalities (see Figure 3.2) and the actions of the user are expected to happen in real life, competitors are free to use any modalities they want to challenge these scenarios. Ten different phones used by eight different people throughout the survey in the dataset [43].

Data format of the track 3 has radically changed when compared to UJIIndoorLoc. While UJIIndoorLoc contains only Wi-Fi RSSI fingerprints, 2017

EvAAL competition contains many more modalities in addition to RSSI fingerprints. Available modalities of IPIN17-T3 is listed as follows:

- WIFI RSSI (in dBm) received from a particular AP (one WAP per row)
- The local magnetic field, as measured by the 3-axis magnetometer in the phone
- The phone's acceleration, as measured by the 3-axis accelerometers in the phone
- The phone's rotation, using the 3-axis orthogonal gyroscopes in the phone
- The atmospheric pressure
- Light intensity (in Luxes)
- The sound pressure level (in dB)
- The temperature in degrees (°C)
- Proximity
- Humidity
- The latitude, longitude and altitude estimated from GPS/Glonass (note these are not the ground-truth positions)
- The mobile phone 3D orientation (pitch, roll, and yaw)
- Bluetooth Low Energy 4.0
- Ground-truth position (only in calibration files)

In addition to the new modalities, IPIN17-T3 dataset's format was also changed. IPIN17-T3 dataset contain multiple log files sorted by the timestamp whereas UJIIndoorLoc was in a tabular form. Each log file contains one continuous site survey of a user. The training, validation and evaluation files are available online, while UJIIndoorLoc has only the training and validation files online. Training and validation files have the same log file structure and these files are grouped into buildings survey conducted. Evaluation log files do not have any positioning information and competitors are expected to predict user position (building, floor, latitude and longitude) in every 0.5 seconds. The lack of presence of the positioning information in evaluation sets makes the performing benchmarking on the evaluation set impossible. Researchers also noted that there is no guarantee that evaluation files and training/validation files are created with the same users and phones.

```

LIGH;71.064;81341.477;486.0;3
ACCE;71.073;81341.480;-2.27928;5.44920;7.92002;3
ACCE;71.077;81341.490;-2.12605;5.31513;8.28394;3
LIGH;71.079;81341.493;486.0;3
MAGN;71.080;81341.493;27.88507;-35.45233;-27.05263;3
LIGH;71.083;81341.497;486.0;3
ACCE;71.087;81341.500;-2.03986;5.19063;8.48505;3
LIGH;71.093;81341.507;481.0;3
ACCE;71.097;81341.510;-1.92494;5.12359;8.67659;3
MAGN;71.099;81341.512;27.83928;-35.75843;-27.54077;3
POSI;71.105;3;39.99392464;-0.07380109;5;30
LIGH;71.106;81341.517;499.0;3
ACCE;71.107;81341.520;-1.62806;5.06613;8.25521;3
LIGH;71.115;81341.529;499.0;3
ACCE;71.117;81341.530;-1.64721;5.26724;7.90087;3
MAGN;71.123;81341.533;27.83928;-35.75843;-27.54077;3
LIGH;71.124;81341.537;499.0;3
ACCE;71.127;81341.540;-1.79086;5.55455;7.45076;3
WIFI;71.133;81341.481;WAP113;20:17:FF:FF:F1:13;-69
WIFI;71.133;81341.481;WAP106;20:17:FF:FF:F1:06;-66
WIFI;71.133;81341.481;WAP141;20:17:FF:FF:F1:41;-50
WIFI;71.133;81341.481;WAP310;20:17:FF:FF:F3:10;-64

```

Figure 3.2: A sample from a log file of the IPIN17-T3 dataset showing different sensors' measurements. For us, only the WiFi measurements are of interest.

IPIN17-T3 dataset performance criterion is the accuracy score defined in Equation 3.1 and Equation 3.2. Our results in the following section will follow these criterion, too.

$$Accuracy\ Score = Q_3(Error(Ri, Ei)) \quad (3.1)$$

$$Error(Ri, Ei) = \|Ri, -Ei\| + p_b * b_{fail} + p_f * |f_{pred} - f_{gt}| \quad (3.2)$$

In Equation 3.1, Q_3 represents 3rd quartile of errors for the all samples with a ground-truth label. In Equation 3.2, p_b and p_f represent the building and floor penalties and equal to 50 meters and 15 meters, respectively. b_{fail} is 1 if building is not estimated correctly, otherwise it is taken as 0.

3.3. GTU GSM RSSI Fingerprint Datasets

This part of the work is derived from a joint TUBITAK project between Gebze Technical University, Yildiz Technical University, and, Turkish phone service provider, AVEA in 2014. We investigated the feasibility of GSM signals for indoor localization. Since there was no public indoor GSM RSSI data available in those days, we have collected our own data set. We developed a data collector application, called CellInfo, which collects the GSM RSSI information in a log file with specified time frequency. CellInfo lets user either specify a custom label or choose from the predefined labels for the RSSI samples.

CellInfo was used in surveys in four different buildings to cover all possible environmental conditions. Two of these buildings were shopping malls in the European side of the Istanbul, i.e., Marmara Forum and Atrium malls. These malls differ in size, people density, and architecture. Marmara Forum is much more larger and more crowded than Atrium, while Atrium has narrow corridors and smaller. Third building was the AVEA Multimedia Center in AVEA's research and development center. This building was used for online demonstration of the project. The last building was the Computer Engineering Department building of the Gebze Technical University (GTU). The data collected in the GTU shaped our point of view for RSSI-fingerprint based ILS. In these buildings, two different survey scenarios were performed: labeled rectangular grid survey and semi-labeled survey on a specified path. During the survey, we used Google's Nexus 5 Android phones and compared the correlation of our observations with different brand's cell phones including Samsung Galaxy S3 and General Mobile's 4G.

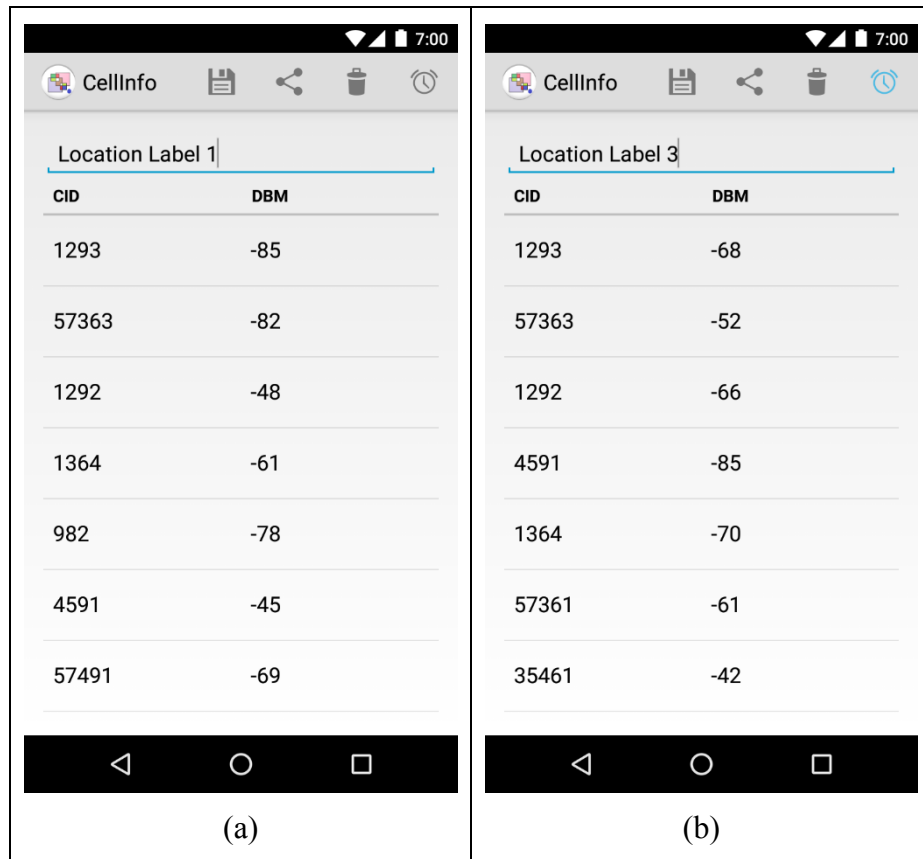


Figure 3.3: CellInfo application screenshots on manual labeling mode. (a) and (b) labels the current RSSI fingerprints in different locations.

Table 3.1: The statistics of GTU Labeled GSM Dataset.

Sample count	22215
Number of observations with unresolved cell tower identifier	30402
Usable sample count ²	11063
Average number of unique cell towers per sample	5.3
Sessions	20

² Usable sample: Sample with no cell tower identifier is unresolved and number of unique cell towers is not less than five.

Table 3.2: The statistics of GTU Semi-labeled GSM Dataset.

Sample count	16676
Number of observations with unresolved cell tower identifier	31013
Usable sample count ³	9973
Average number of unique cell towers per sample	5.8
Sessions	13

³ Usable sample: Sample with no cell tower identifier is unresolved and number of unique cell towers is not less than five.

4. MACHINE LEARNING FOR POSITION ESTIMATION

In this section, we present a semi-supervised method for position estimation using RSSI fingerprints. We first give the details of the method for GSM data followed by the WiFi data. We treat these two sources of RSSIs separately as the infrastructures offer different nature of RSSI signals. The underlying method, however, remains the same.

4.1. Semi-supervised Analysis of GSM Fingerprints

As a widely available modality, GSM signals, can be used either for coarse position estimation or as an auxiliary modality for indoor localization systems. Since GSM is a well adopted technology and cell towers are already deployed around the buildings, we tried to find if GSM fingerprints are suitable for indoor localization using commodity phones. If so, we can even program the SIM cards to send RSSI fingerprints to localization model and query its current location without using an external devices or smartphones.

We used the GSM RSSI fingerprint dataset that we surveyed inside the Gebze Technical University campus, as described in the dataset's section. This dataset holds at most 7 different base station's signal strength collected by off-the-shelf available smart phones, like Nexus 5 and Samsung Galaxy S3.

For predicting the location, we first tried Bayesian machine learning methods on labeled GSM RSSI fingerprints. Though, it performed better than random, we investigate further to increase localization accuracy. We tried a few more different machine learning techniques including SVM and Rotation Forest (RF). RF method provided the best accuracy for the classification of labeled data and SVM performed better when a linear kernel is used.

The model trained using RF offline was integrated in an online localization application called LocationFinder. This application was used to test the localization model's robustness online in the survey site. This application was developed for Android platform and communicates with a REST API (Restful Application Programming Interface) to query phone's current location. REST API was backed with a Java application and predicts location with offline-trained machine learning model.

Weka, a machine learning software and library developed by University of Waikato, was used for both offline training an RSSI localization model and online querying RSSI fingerprints from a mobile device via the REST interface.

After training several RSSI localization models using various machine learning models, we tested each other using LocationFinder in the field. Although GSM has temporal stability, the results were not satisfying and got worse as time goes by.

In order to eliminate the time spent on training and deploying the RSSI localization models, we completely rewrote both the mobile application and server-side API to enable our system to be able to train and query using the same mobile application, instantly.

We also experimented on temporal behavior of RSSI fingerprints using the second version of Location Finder. Using these experiment results, we found out that we could not achieve long-term, reliable and high-accuracy localization model with GSM RSSI fingerprints.

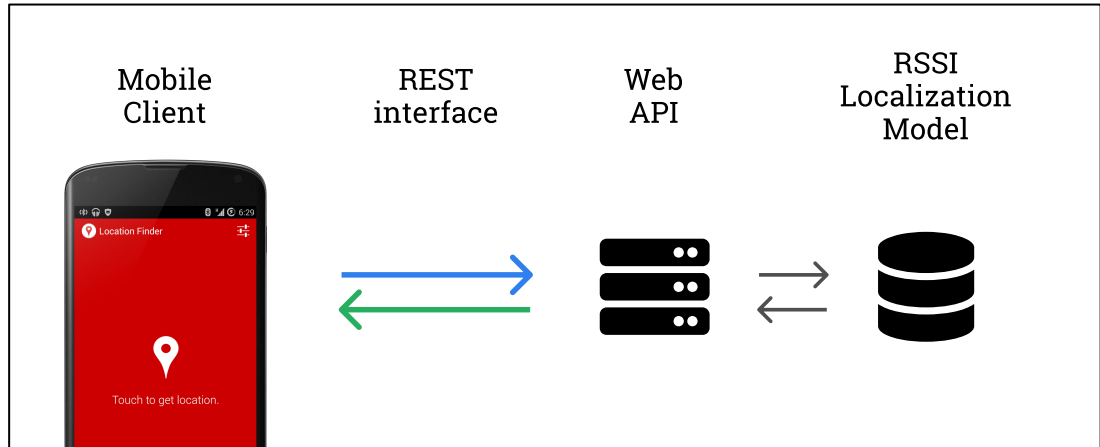


Figure 4.1: First version of the LocationFinder, online location querying application, architecture which communicates the RSSI localization model through a web API using a REST interface.

In Figure 4.2, an example behavior of different cell towers from GTU Labeled GSM dataset within the same session can be seen. The plots show how received signal strength from three nearest cell tower changes over time while the user holds the phone still in his hand near his chest and towards his face, pretending a normal phone usage. We also collected RSSI fingerprints on the same location for 30 minutes with leaving the phone on the ground when there is nobody around it. The behavior of the GSM RSSI fingerprints for both situations seems correlated. Each cell tower's signal has a

fluctuation with different frequency and this causes more than $\pm 5\text{dBm}$ signal strength, as stated in [13]. We could not handle this temporal instability with seven base stations' signal strength at a time, at most.

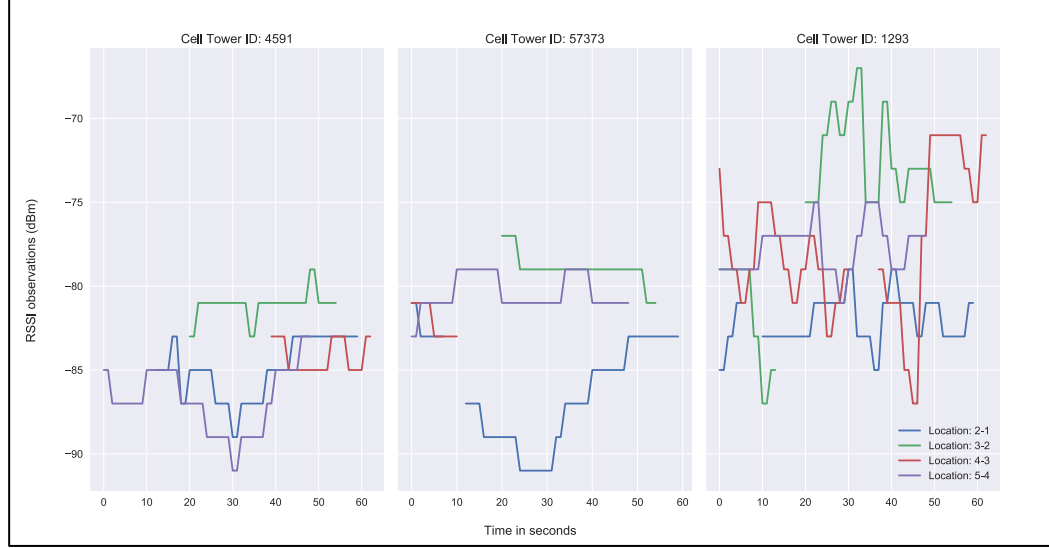


Figure 4.2: RSSI observations over time when phone holds still in hand. Lines represent the signal strength change in each location over time.

In order to use GSM RSSI fingerprints as an auxiliary modality for indoor localization, we developed an algorithm which reconstructs a linear path without a surveying a fine-grained RSSI map. Our incentive was to eliminate surveying process as much as possible under some constraints. This may enable us to collect unlabeled RSSI fingerprints at scale, using a daemon SIM card or smartphone application without a user interaction, and create semi-labeled radio map of the environment with a comparatively small number of labeled fingerprints.

In recent years, crowdsourcing methods on surveying fingerprint database have become more and more popular [45,46]. These methods aim to eliminate as much labor-intensive surveying process as possible and to reduce the installation and maintenance costs of the ILSs. In the proposed algorithm, we employed a similar crowdsourcing technique to create semi-labeled radio map of a non-intersecting corridor. To create this map, we assumed that there is a daemon in the user's mobile phones which sends RSSI fingerprints to our server in the background. These fingerprints do not contain any label and fingerprints were collected on the same linear path. If we have uniformly distributed fingerprints in this specific path or corridor, we

can reconstruct the coarse-grained labels using small number of reference fingerprints from the two edges of this path.

This method, LMR, works like a recursive clustering method, yet it achieves better performance than well-known clustering methods in two ways: separating fingerprints from each other and semantic cluster labeling to reconstruct the path. Current clustering algorithms could not find separate one GSM RSSI fingerprint from another because a simple distance function cannot define the behavior of these fingerprints. Thus, clustering algorithms like K-Means could not simply work on GSM RSSI fingerprints.

Another challenge for GSM RSSI fingerprinting-based method is phones can observe at most seven different cell towers in each observation. These cell towers are selected by their signal strength and only the seven strongest signal emitter base station can be observed at any given time. Since GSM signal strength fluctuates, like we mentioned before, it's unlikely to find a distance function between fingerprint samples without enough data. As we observed the decision-tree-based classification algorithms perform better on identifying the GSM RSSI fingerprints, we employ these algorithms to divide given linear path into sub-segments.

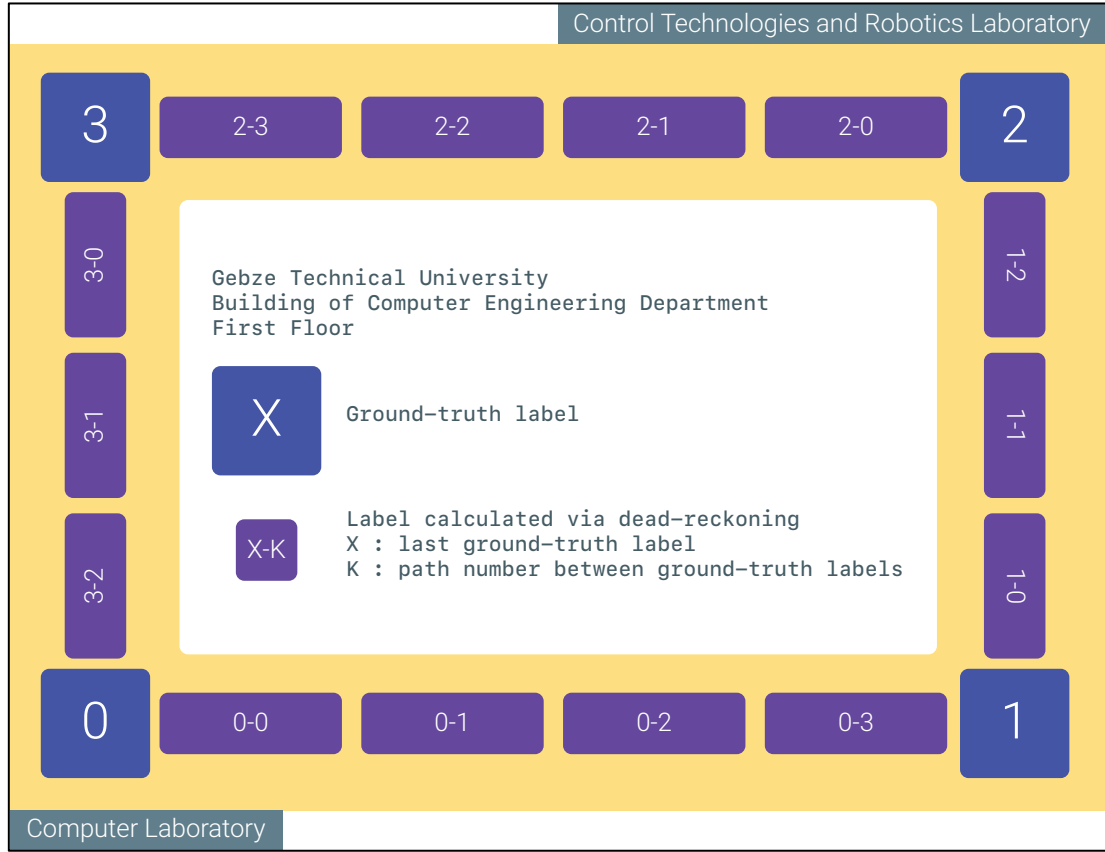


Figure 4.3: Semi-labeled GSM dataset survey plan. User starts walking from the corner labeled as 0 at constant pace and labels fingerprints at each corner. User continues surveying for 7-8 times on the specified path.

Another advantage of LMR over conventional clustering methods is it can preserve the order of cluster centroids with respect to each other. In other words, even some clustering method performs better than LMR, it cannot correctly sort the cluster centroids, yet LMR employs the relationship between cluster centroids and is able to recover the order of clusters with respect the given path. However, no clustering method we tried performed better than LMR.

To test our algorithm, we used the RSSI fingerprints on the corridor between the reference points “2” and “3” in the GTU semi-labeled RSSI dataset. During the survey process, user walked with a constant speed across reference points and we divided the corridor into 6 sub-segments based on the total walking time, as can be seen in Figure 4.3. For 17 days, user walked on this corridor for 56 times to collect training set. In the following two days, user walked 17 times between the reference points. Only the beginning and end of the walks have taken into consideration as reference fingerprints.

Using Rotation Forest algorithm with 85 estimators, LMR successfully constructs the path with 55% accuracy. You can see the reconstruction performance of

K-means and LMR in the Figure 4.4. Note that K-means unable to preserve order of cluster centroids and we manually reordered the cluster indexes for convenience.

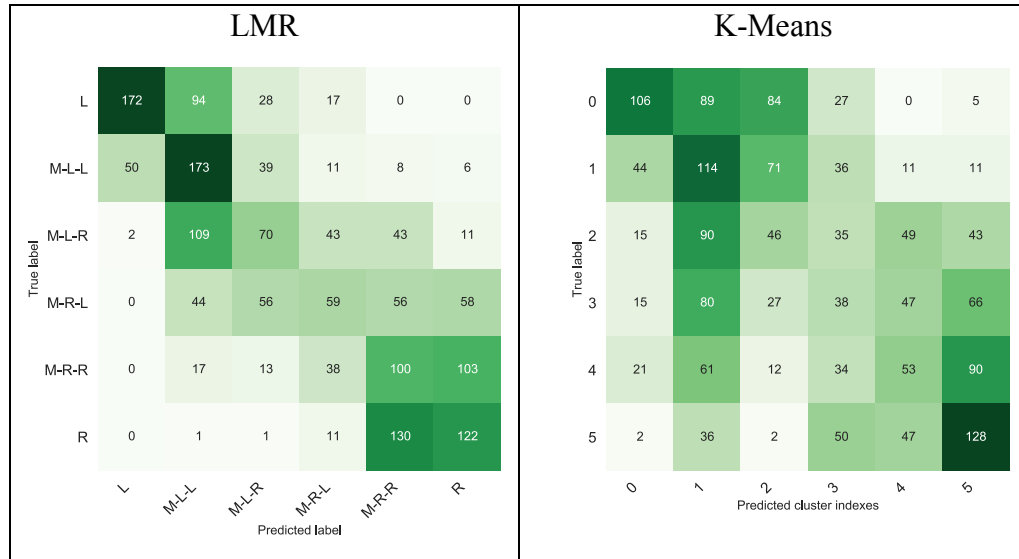


Figure 4.4: LMR and K-Means performance comparison of GSM RSSI fingerprints. Y axis represents the ground-truth labels and X axis represents the predicted labels. Note that K-means cannot preserve the order of cluster centroids, while LMR can. K-means' order of cluster centroid ordered by hand for convenience.

4.2. Semi-supervised Analysis of WiFi Fingerprints

As mentioned before, surveying process is the most laborious work in fingerprinting systems. It causes to increase cost for both installation and maintenance of ILSs. To reduce as much work as possible during the survey, several methods have been proposed [45,46]. We investigated the eligibility of our proposed LMR algorithm that we developed for GSM fingerprints, on the WiFi fingerprints. GSM and WiFi fingerprints are quite different, due to there is at most seven different cell towers with the strongest signal strength are visible in GSM, yet WiFi does not have a similar limitation. Most WiFi fingerprint datasets have hundreds of WAPs while a sample can contain only 5-30 WAPs at most. On the other hand, in a similar environment, number of visible GSM base stations will be much less than the total number of WAPs. Having lots of missing data on each sample may change the performance of the LMR algorithm, dramatically.

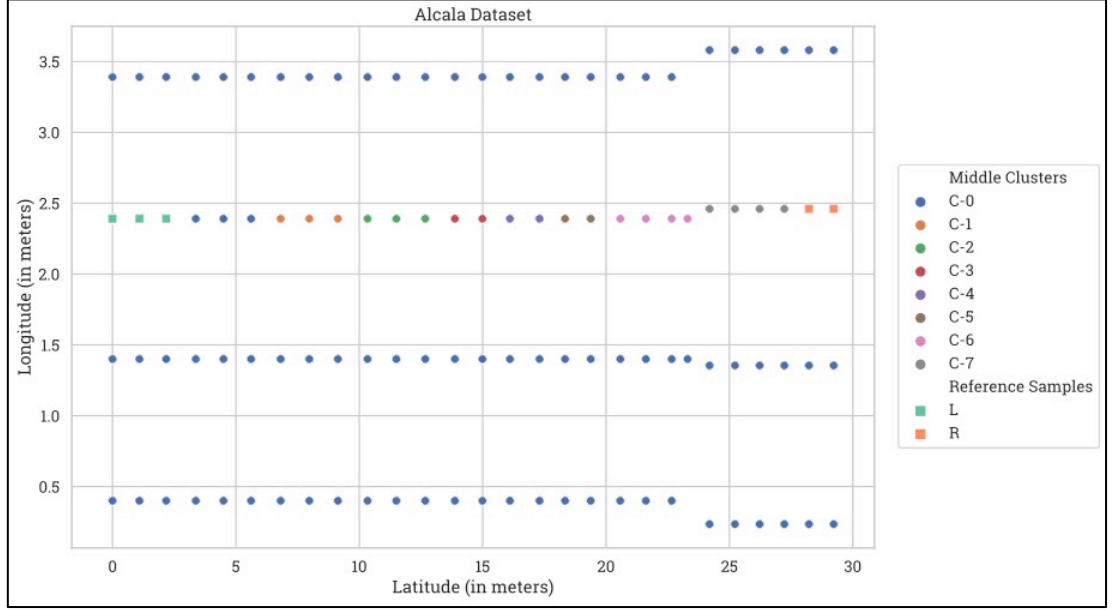


Figure 4.5: Alcalá Tutorial 2017 dataset coordinates are plotted in X-Y axis. Selected corridor is represented in colors and reference points represented with square markers.

To test our algorithm on a WiFi RSSI fingerprints, we used the Alcalá Tutorial 2017 dataset [41] which conforms our requirements well. As explained before, we assume that vast amount of unlabeled data lies on a non-intersecting path (preferably a linear path) and using samples from both edges, we can cluster these unlabeled RSSI fingerprint samples while preserving the order of clusters. Alcalá Tutorial 2017 dataset contains 152 WAPs and have total 670 training samples. There are 4 corridors in the survey site, which we choose the 3rd one from bottom as our testbed (see Figure 4.5).

We picked the samples which lies in the X axis with values smaller than 2.5 as a left reference point and values bigger than 27.5 as a right reference point. We split rest of the samples into 8 clusters based on their latitude values. These clusters are indicated with different colors in Figure 4.5. Each corridor approximately spans 30 meters in width. We plotted the confusion matrix of predicted clusters with respect to ground-truth clusters, like we did with GSM in Figure 4.6. Our cluster predictions are 90% accurate and as can be seen in confusion matrix, LMR's rank-1 accuracy is 100%.

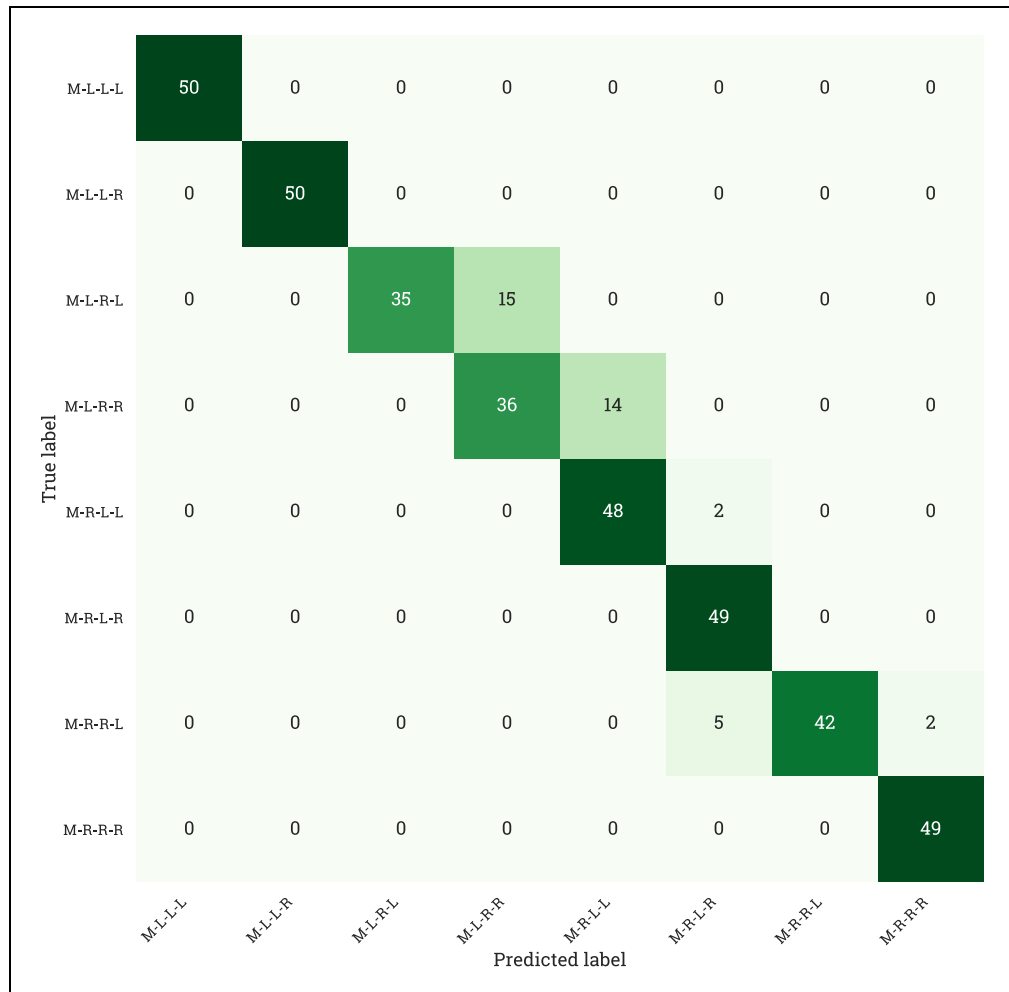


Figure 4.6: LMR results on a WiFi RSSI fingerprinting Alcalá Tutorial 2017 dataset. LMR successfully clusters WiFi RSSI fingerprints in a corridor while preserving their cluster orders.

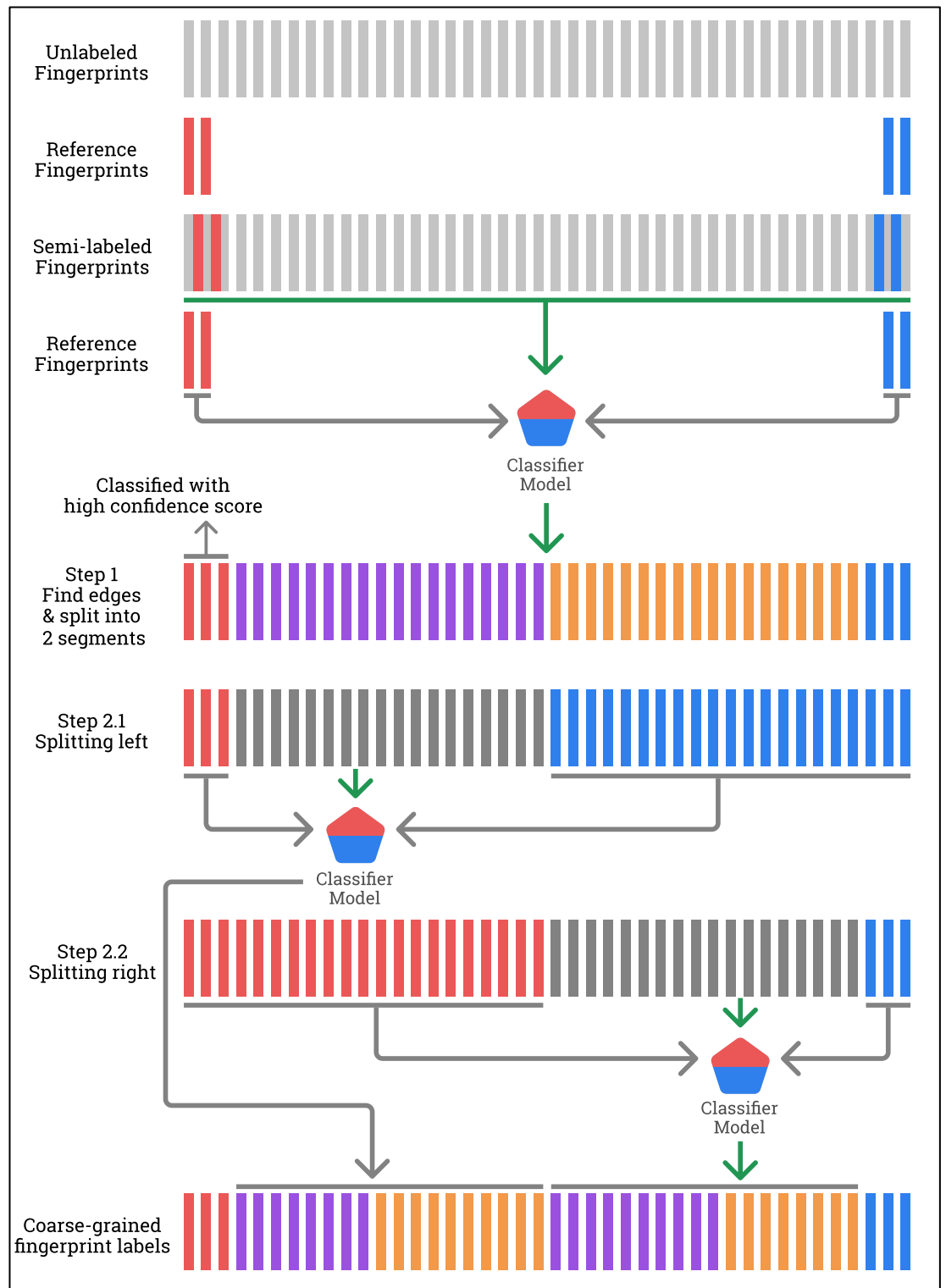


Figure 4.7: The proposed LMR algorithm for semi-supervised classification of RSSI fingerprint data.

5. DEEP LEARNING METHODS FOR WI-FI FINGERPRINT ANALYSIS

For the sake of completeness, this section provides a detailed analysis of deep learning methods on WiFi RSSI fingerprint analysis for location estimation. After a complete review of the literature, we will share our own results.

5.1. Preprocessing

Preprocessing the RSSI fingerprint samples plays a significant role in indoor localization systems. Most of the deep learning methods handle the data in a vector, matrix or tensor form. For RSSI fingerprints, this representation can be a vector for one sample, a matrix for multiple consecutive samples where the number of columns does not change from sample to sample. Therefore, all the WAPs whether their signal is observed or not, should be present in the same column for all samples. As a result, handling the missing WAPs is a key challenge for designing a deep learning-based ILS.

First publicly available dataset, UJIIndoorLoc, represents these missing observations as 100 dBm. All observations in UJIIndoorLoc are between -104 and 0 dBm. 1.7% of all samples are between -45 and 0 dBm and only 2.3% of all samples have less than -95 dBm signal strength [37]. 100 dBm, which indicates a missing WAP, causes a discontinuity in signal strength domain. To overcome this issue, [18] represented missing observations as -110 dBm, which is a lower than the minimum signal strength in all observations. This technique is also used in the following works, too [19,20].

In deep learning, weight initialization determines how well the network is trained, even not trained in some cases. As described in [47], depending on the activation function used and weight initialization, gradient updates can become significantly small or huge. This issue is called as vanishing or exploding gradients in the literature. To avoid this, different weight initialization techniques and activation functions are proposed. As a common sense, weights are initialized with random values using gaussian normal distributions. So, the normalization of the inputs, like signal strengths, in deep learning models is a common regularization pattern in ILS, as well.

Like missing data manipulation, signal strength normalization technique for deep learning-based ILS is also introduced by [18] which uses z-score normalization as defined in Equation 5.1.

$$z = \frac{X - \mu}{\sigma} \quad (5.1)$$

Here X denotes the RSSI observation, μ denotes the mean signal strength and σ denotes the standard deviation of the signal strength. After z-score normalization, RSSI observations have zero mean and unit variance. Z-score normalization is also used in [19,20,33].

Temporal deep learning models are relatively new and more robust to temporal signal fluctuations and noises which can arise from both sensor and the environment itself [33]. In order to use temporal models, observations must be timestamped during survey process. UJIIndoorLoc and IPIN-2017-T3 datasets include timestamp for each RSSI fingerprint sample. While UJIIndoorLoc is in a tabular form and contains a column for timestamp, IPIN17-T3 dataset collected using a specific log file format which every sensor reading is timestamped with absolute and relative timestamps. IPIN17-T3 dataset still has not used in any publication except for the competition participant papers of the IPIN 2017 conference. To the best of our knowledge, this thesis is the first work which employs temporal deep learning models to both datasets mentioned in the previous section.

5.2. Stacked Autoencoders

Stacked Autoencoders (SAE) is the first neural network-based ILS which published first in the late 2017. Several researchers developed SAE-based classifiers by aiming to reduce input dimension while preserving a decent localization accuracy. SAE only creates an embedding from a given RSSI fingerprint, as explained before, so in order to make a localization, a fully connected network (FCN) is placed after the encoder network of the SAE.

First SAE-based ILS was [18], which only predicts the building and floor using a given RSSI fingerprint. Correct prediction of both floor and building is named as

“success rate” as described in the previous section. The networks used in this work can be seen in Table 2.1.

Authors built a network by connecting 128-dimensional encoder output to two consecutive 128-dimensional densely connected layers. They also used dropout layer between the dense layers, stated as “DO” in Table 2.1, to remove random connections in FCN, so that they avoid overfitting the training data and acquired better generalization over the entire dataset.

Similar SAE structure also used in [19,20] with little modifications. These methods achieved similar accuracy in prediction of both building and floor prediction (success rate). Their success rate was %91 and 93%, respectively. Their contributions were building a multilabel classifier, usage of weighed loss function for different outputs, and predicting user’s coordinates using a DNN. Adam optimizer [25] with binary cross entropy error [24] was used to train this network.

However, none of the methods exceeded the baseline accuracy. Baseline method, which is a 1-nearest neighbor classifier, achieves %89.92 success rate with 7.9 meters mean coordinate estimation error [37].

5.3. Convolutional Neural Networks

CNNs are expected to work well in ILS using RSSI fingerprint analysis, yet none of the earlier deep learning methods tried CNN. As of July 2018, two RSSI fingerprint analysis papers have published which reported better scores in terms of localization accuracy than the state-of-the-art results.

The very first method utilizing CNNs into ILS domain is [32]. Since they used Channel State Information (CSI) instead of RSSI fingerprints (see Figure 5.1), we cannot take their result into account when comparing with RSSI fingerprinting-based ILSs. Yet, their contribution on both forming images out of CSI features to build an ILS and comparing their CNN-based model with other DNN models using the same CSI features are valuable. They achieved 1.3-meter average localization error with 0.9-meter standard deviation. However, CSI features can only be acquired using a custom hardware or custom firmware on specific hardware. Thus, CSI-based ILSs cannot be applied to the off-the-shelf phones and devices.

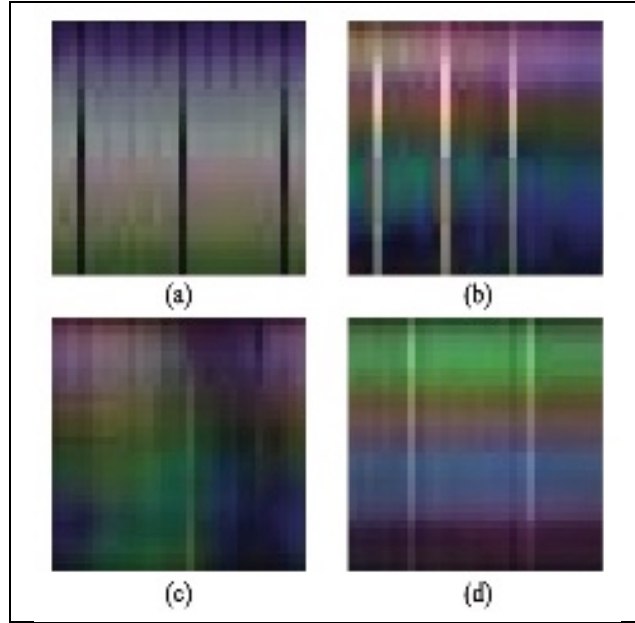


Figure 5.1: CSI feature images used in ConFi where a, b, c, d represents CSI features in different locations.

In [33], researchers have introduced the first use of temporal information handling using CNNs with three different techniques. These techniques are averaging consecutive RSSI fingerprints, concatenation consecutive fingerprints and forming RSSI matrixes as a time series of RSSI fingerprints, just like ConFi. Unlike common evaluation criteria described in dataset’s paper, their results show only one building’s floor-level localization error. According to the claim on the paper, their floor-level localization error for one building represents overall accuracy of their system. It may show how good their model works for this domain, yet it’s unfair to compare their results with other works, since there is no common metric between them.

Following [33], [34] also proposed a CNN-based ILS which uses their own dataset and achieved near CSI-based ILS localization accuracy according to the paper. This work did not compare its performance against public datasets. So that we cannot take their results into consideration during comparison with other works, either although their results seem very promising.

To the best of our knowledge, we presented the first CNN-based ILS which utilizes RSSI fingerprint analysis before the two works mentioned earlier have published. Our aim was to predict user’s location instantaneously as accurate as possible. In order to do so, we proposed two different one-dimensional CNN networks with multitask classifier and multi-task regressor. This type of multi-task learning is

commonly used in other fields such as image understanding and natural language processing [48].

Table 5.1: Comparison and results of Mai Ibrahim, *et al.* using different techniques for feeding temporal data into CNN.

Input	Building Prediction Accuracy	Floor Prediction Accuracy	Localization Mean Error (m)
Single RSS Vector	100%	91.42%	10.25
Averaged RSS Vector	99.91%	96.90%	4.39
Concatenated RSS Time-series	100%	99.66%	3.87
RSS Time-series Image	100%	100%	2.77

We first normalize the RSSI fingerprints using z-score as can be seen in Equation 5.1 in order to distribute signal strength values around the zero with unit variance. Then we normalized each building's RSSI fingerprint locations separately between 0 and 1 where 0 means minimum latitude/longitude value and 1 means maximum latitude/longitude values.

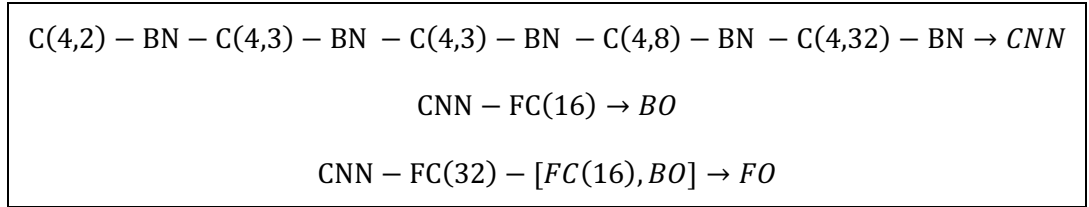


Figure 5.2: BF-CNN Multiclass building and floor classifier structure where the first line contains the common layers of both outputs and consecutive layers are building and floor outputs, respectively.

Multitask classifier (BF-CNN) uses the normalized RSSI fingerprint inputs and predicts one-hot vectors for building and floor, separately. Since these two predictions have many parameters in common, we exposed these parameters in the early convolutional layers. While short filters learn the micro-spatial features like co-existing WAPs within the same kernels and individual values of signal strengths, longer filters learn the large-scale spatial features such as building fingerprints. BF-CNN network has the setup shown in Figure 5.2 and Figure 5.3. The yellow bars on Figure 5.3 indicate CNN layers as described with details in Figure 5.2, green bars indicate batch normalization layers, purple planes indicate densely connected layers

and teal planes indicate the network's output vectors. We also used building's prediction as a feature when predicting floor output.

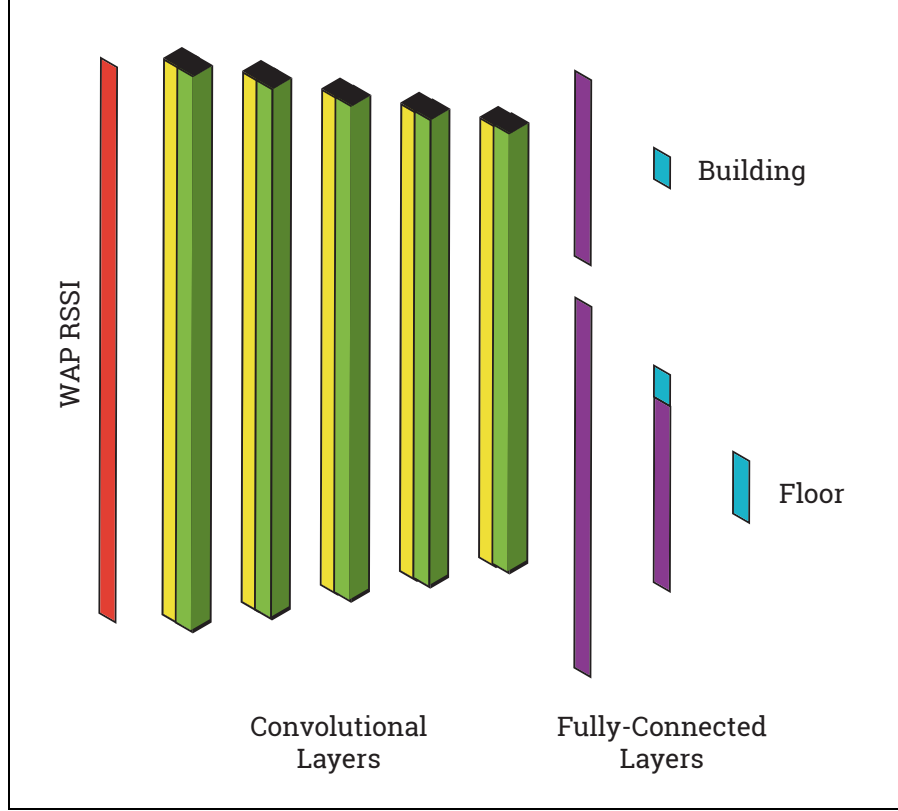


Figure 5.3: BF-CNN - Multiclass classifier for prediction floor and building at the same time.

In Figure 5.2, $C(f, s)$ represents 1-D convolutional layer with f different $s \times 1$ sized filters, N denotes batch normalization layer, $FC(n)$ denotes a fully connected layer with n nodes, BO represents building prediction as a one-hot vector and FO represents floor prediction as a one-hot vector. Floor prediction also uses the building prediction output in its last layer by concatenating. All convolutional and fully connected layers are followed by a sigmoid activation function except for the output layers which have SoftMax activation functions for classification.

We concatenated the same input we fed into BF-CNN with building and floor predictions (one-hot vectors) and train our latitude longitude regressor (CoordNet). Since the coordinates of each building normalized within itself, we believe that feeding building and floor predictions to the CoordNet would be useful. CoordNet follows the same structure as BF-CNN does in learning common features of latitude and longitude

in the early layers of the network. CoordNet’s structure can be seen in Figure 5.4 and Figure 5.5 where we followed the same taxonomy of BF-CNN.

Like SAE-based works, we indicated missing values as -110, too. Since our method contain two different neural networks, we needed to train them separately. To train BF-CNN, we choose AdaDelta optimizer [49] with the learning rate of 1.0 and categorical cross entropy loss [24] for both outputs. This network is trained with batch size of 30 for 20 epochs.

After this training process, we achieved 100% building and 91.27% floor prediction accuracy which yields 91.27% of success rate as described in UJIIndoorLoc dataset’s evaluation criteria [37].

To estimate absolute location coordinates, we use Adam optimizer [25] with mean squared error loss function (see Equation 2.1) for both latitude and longitude estimation. We set batch size to 75 during training and network is trained for 25 epochs. On the test set, we achieve the best performance of 7.36 meters mean coordinate estimation error. Since the training process is stochastic, different training runs yield different results. For this reason, we have trained our entire network at least 10 times recording the test performances. The worst performance was 13.6 meters and average error was 8.5 meters.

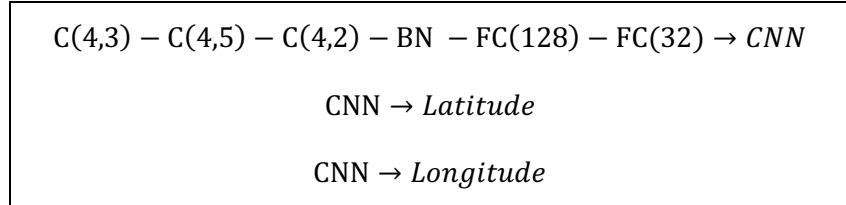


Figure 5.4: CoordNet structure which utilizes the building and floor predictions of the BF-CNN and extracts features using common convolutional layers.

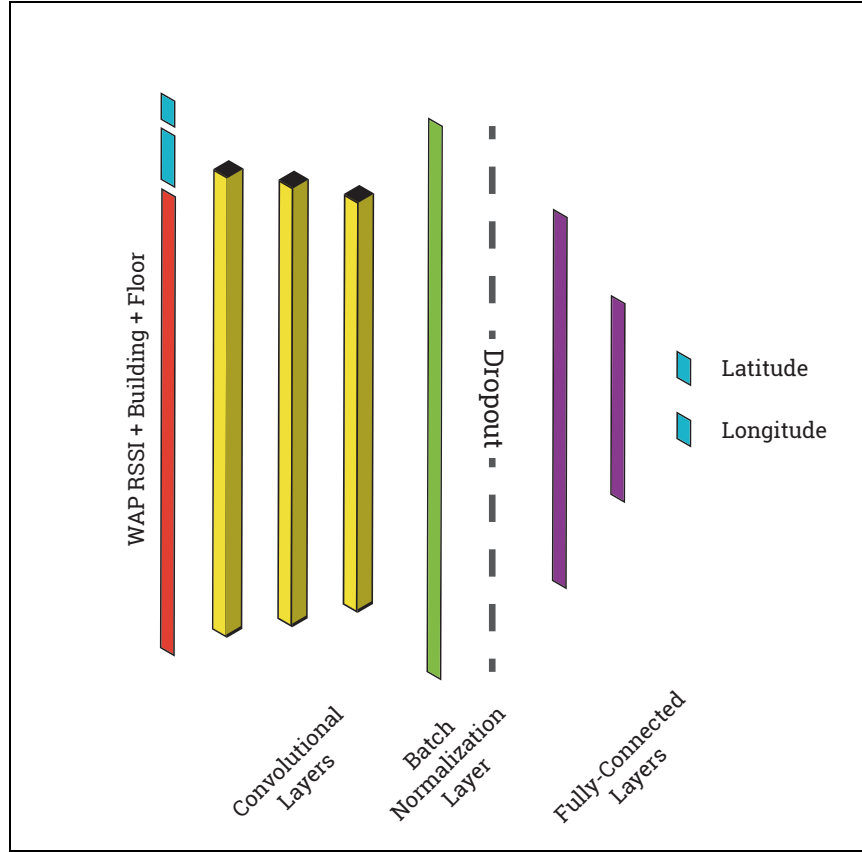


Figure 5.5: CoordNet – a CNN-based latitude and longitude multitask regressor which learns the common features in the earlier layers and estimates latitude and longitude values at the same time.

We tested our method against UJIIndoorLoc dataset and followed the dataset’s own performance criterion which is calculating both the success rate and mean location estimation error among all test fingerprints. Note that, although [33] also show their results on the same dataset, they did not follow the same performance criterion with others. We showed their results with others to show the importance of temporal data usage. Proposed method, which uses the same input except [33], shows better performance in terms of localization accuracy.

Then we combined these two networks, BF-CNN and CoordNet, into one convolutional network and train it the whole network together. This network, E2E-Conv, provide better success rate over the BF-CNN and more accurate coordinate estimation than CoordNet, as can be seen in Figure 5.7. This is because, E2E-Conv extracts features at the convolutional layers better since all the outputs have so much information in common. This end-to-end network enabled us to exploit these features, better than two separate convolutional networks. E2E-Conv’s structure can be seen in Figure 5.6.

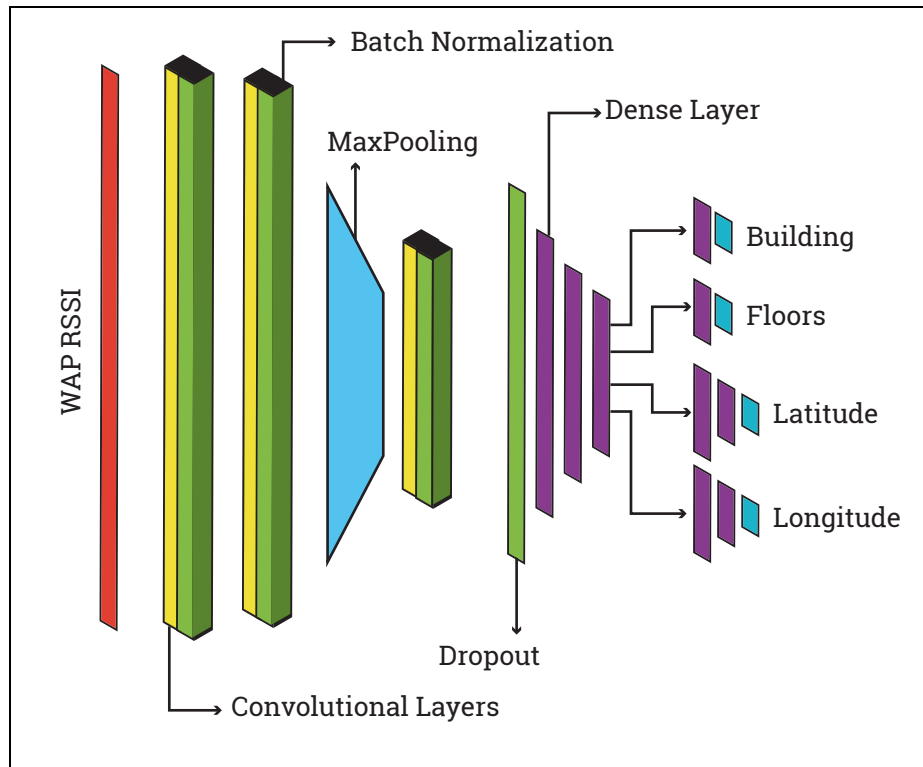


Figure 5.6: E2E-Conv: End-to-end multiclass classifier and regressor based on CNN without temporal data.

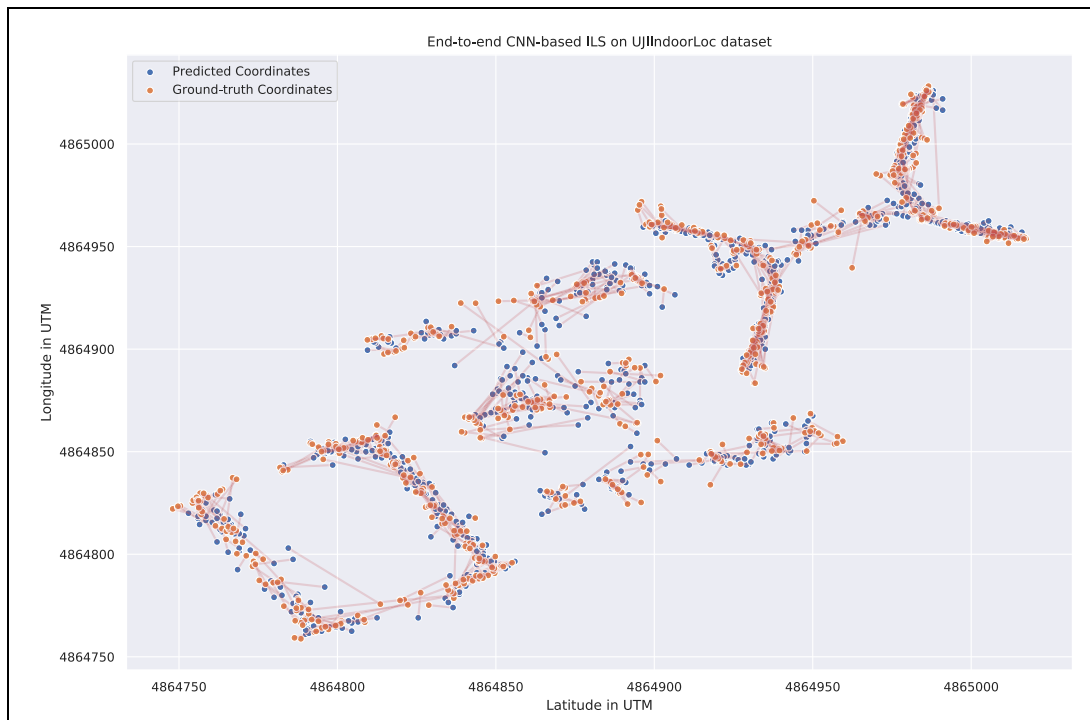


Figure 5.7: Coordinate estimation performance on UJIIndoorLoc dataset of E2E-Conv where lines represents the distance between actual and predicted coordinates.

Table 5.2: The comparison of the proposed multitask learning method with others on UJIIndoorLoc dataset.

Method	Success Rate	Estimation Error
[37] - <i>Baseline</i>	89.92%	7.9m
[18]	92%	N/A
[19]	93.33%	N/A
[20]	91.18%	9.29m
[33] ⁴	100.00%	2.77m
CoordNet	91.27%	7.36m
E2E-Conv	93.42%	7.58m

We also propose an end-to-end trainable CNN to predict building, floor, latitude and longitude at the same time which also employs temporal behavior of the RSSI fingerprints. We tested this network’s performance on the IPIN17-T3 dataset. To the best of our knowledge, this work is the first deep learning-based ILS on this dataset. We achieved less than 7 meters of localization accuracy with 100% building estimation and %92.13 floor estimation accuracy. We used 8 filters on convolutions with the following kernel sizes: 2x2, 5x3, 3x3. We used %20 random dropout after the 128 and 64-sized fully connected layers to avoid overfitting the training data. We used 10 consecutive RSSI observations to train this network, which can be seen in Figure 5.8.

⁴ The results show one building’s location estimation error

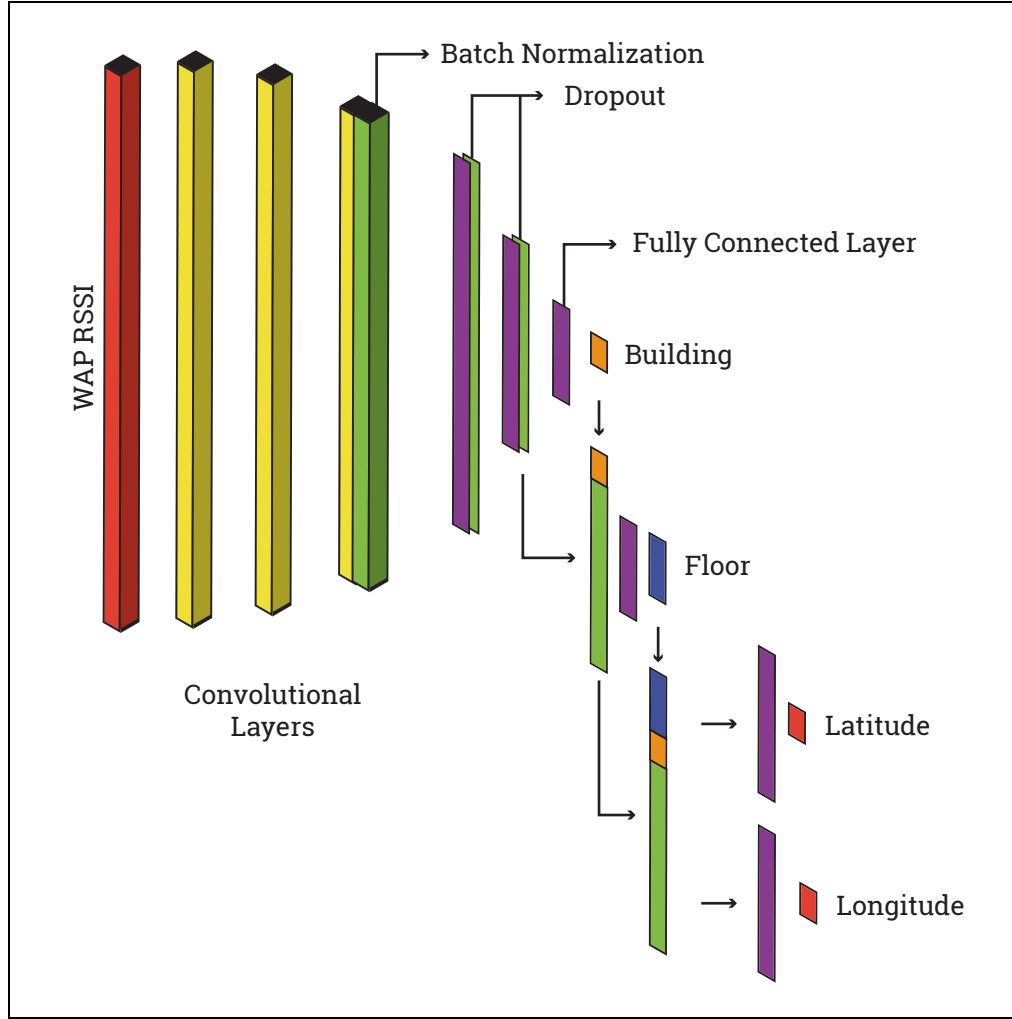


Figure 5.8: CNN-based temporal localization model architecture with multi class classifier and multi class regressor, tested using IPIN17-T3 dataset.

5.4. Long-Short Term Memory Networks

In addition to CNN-based, we propose the first long-short term memory (LSTM) network in this field. This proposed method on the IPIN17-T3 dataset, which to the best of our knowledge, is the first deep learning-based work in this dataset and first LSTM usage for temporal behavior extraction in the WiFi RSSI Fingerprint-based ILS literature.

In the IPIN17-T3 dataset, there are 578 different WAPs in the training, test, and evaluation datasets. Dataset contains 2817 training, 1112 test samples. Due to our model exploits temporal behavior of the data, we formed RSSI fingerprint matrixes using these samples.

We also tried floating windows on time sequences of the RSSI fingerprints to augment data. However, this approach did not work well. This is because earlier windows are far away from the reference coordinates. Using tens of seconds earlier RSSI fingerprint readings to estimate coordinates confuses our model and reduces our accuracy. Our method is end-to-end trainable network which predicts building, floor, latitude, and longitude at the same time using the common feature extractor module, like we proposed in BF-CNN and CoordNet. The network structure can be seen in the Figure 5.9.

We used 128 hidden units in the first layer and 64 hidden units in the second layer of LSTM cells. We fed LSTM's output to a fully connected network with the size of 64, 16, and 64 dense connections, respectively. This network acts like a feature extractor for our multi-class classifier and regressor. Then building prediction is made using these features by 16-sized dense layer. One-hot vector of the building prediction is concatenated with the same features to predict floor after a fully connected network made up by 32, 16, and 16-sized dense layers. Then both building and floor predictions' one-hot vectors are concatenated with the same features and latitude and longitude are estimated separately using 32, 32, 16, and 16 sized fully connected dense layers. Last layers of the latitude and longitude regressors use linear activation, while floor and building predictions use SoftMax activation functions. Rest of the layers use sigmoid activation function. We also used different weights for building, floor, latitude and longitude estimations. The weight of the floor estimation is two times bigger than building estimation and 5 times smaller than latitude and longitude estimations.

We achieved 100% building and 93.01% floor accuracy in this network. This network's location estimation error on the third quartile is 6.20 meters by taking building and floor penalties into account, as described in the Equation 3.1 and Equation 3.2. We used 20% dropout in both recurrent outputs and between the LSTM layers to avoid overfitting. This proposed method learns less than 100K parameters in total for all its outputs. The power of LSTM networks can be proved with this work when we compared this one to our CNN-based similar network on the same dataset. This network performs better with learning less parameters than our CNN-based proposed end-to-end network.

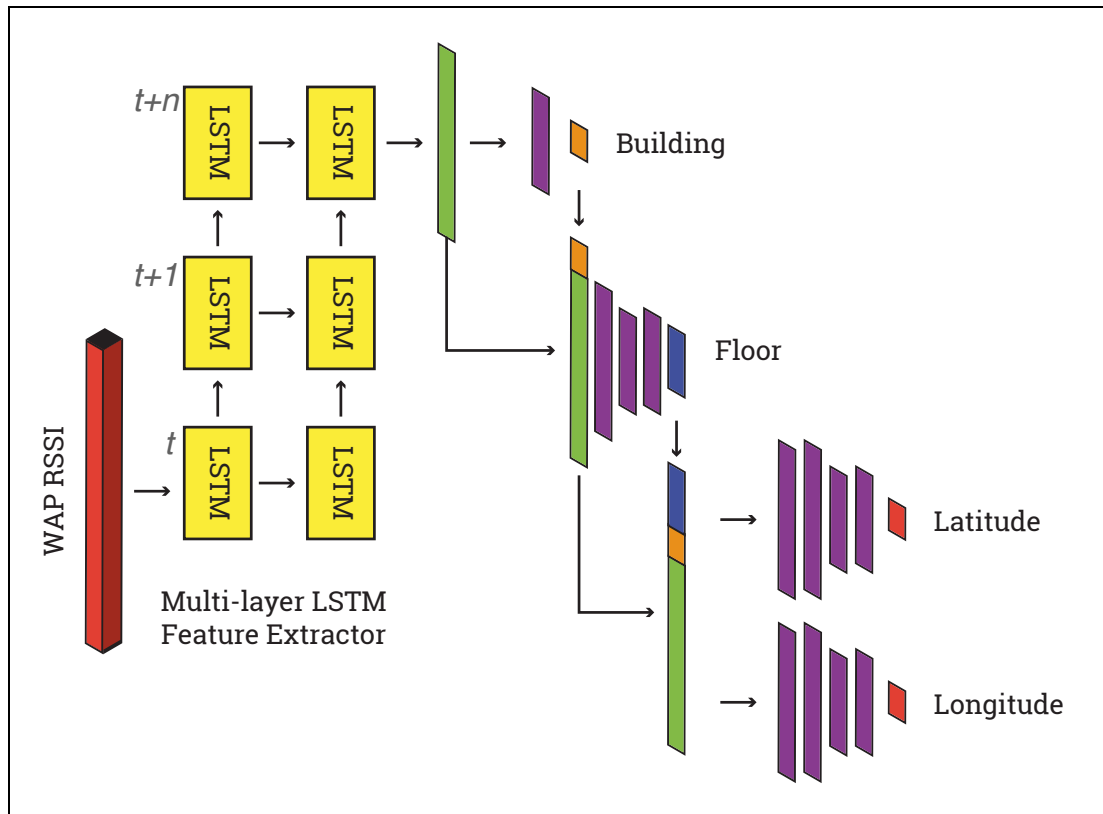


Figure 5.9: Multi-layer LSTM-based temporal localization model architecture with multi class classifier and multi class regressor, tested using IPIN17-T3 dataset.

6. CONCLUSIONS

We have presented a new method for estimating the location of a device collecting RSSI data from a GSM or a WiFi infrastructure. Unlike many methods in the literature, the proposed method does not require fully labeled RSSI fingerprints for training. Instead, given that the data is collected while the device is located on a constrained path such as a road or a corridor, one does not need to label the entire data for training. Only beginning and end location needs a label which is much more practical than labeling all the data.

Experimental results show that the method works for GSM as well as WiFi RSSI fingerprint analysis. As the GSM data is lower dimensionality, performance on WiFi RSSI fingerprints is better. To the best of our knowledge, there is no other semi-supervised method for localization.

The advances in deep learning methods for computer vision and natural language processing are also applicable to supervised location estimating using RSSI fingerprints. We have introduced the using convolutional structures that provides better results than the state of art in the literature.

To the best of our knowledge, our proposed multi-layer long-short term memory (LSTM) network for WiFi RSSI fingerprint localization is the first LSTM network in the literature. We implemented two temporal deep neural networks, one uses convolutional neural networks, the other uses LSTM network. Both of them perform better than state-of-the-art results and LSTM performs better while learning less parameters. Successive RSSI measurements are more reliable for position estimation since it eliminates sensor errors and exploits temporal behavior of the RSSI data.

Generative adversarial networks can also be used for the semi-supervised location estimation problem. Further research in this direction will be pursued. In particular, enforcement of map information within the training of network as a part of constraint satisfaction will be explored.

REFERENCES

- [1] Web 1, (2018), https://www.nasa.gov/directorates/heo/scan/communications/policy/GPS_History.html, (Accessed on: 1/6/2018).
- [2] Web 2, (2018), <http://www.pewinternet.org/2013/09/12/location-based-services/>, (Accessed on: 21/6/2018).
- [3] Xiao J., Zhou Z., Yi Y., Ni L. M. (2016) “A Survey on Wireless Indoor Localization from the Device Perspective.” *ACM Computing Surveys* 49, 1–31.
- [4] Ma J., Li X., Tao X., Lu J. (2008) “Cluster filtered KNN: A WLAN-based indoor positioning scheme.” In *2008 IEEE International Symposium on A World of Wireless, Mobile and Multimedia Networks, WoWMoM2008*, Newport Beach, CA, USA, 25–35.
- [5] Sun Y., Liu M., Meng M. Q. H. (2014) “WiFi signal strength-based robot indoor localization.” *2014 IEEE International Conference on Information and Automation, ICIA 2014* 250–256.
- [6] Yang L., Chen H., Cui Q., Fu X., Zhang Y. (2015) “Probabilistic-KNN: A novel algorithm for passive indoor-localization scenario.” *IEEE Vehicular Technology Conference 2015*, 1.
- [7] Miura H., Sakamoto J., Matsuda N., Taki H., Abe N., Hori S. (2006) “Adequate RSSI Determination Method by Making Use of SVM for Indoor Localization.” In *Knowledge-Based Intelligent Information and Engineering Systems, Gabrys B, Howlett RJ, Jain LC, eds. Springer Berlin Heidelberg, Berlin, Heidelberg*, 628–636.
- [8] Oussar Y., Ahriz I., Denby B., Dreyfus G. (2011) “Indoor localization based on cellular telephony RSSI fingerprints containing very large numbers of carriers.” *EURASIP Journal on Wireless Communications and Networking 2011*, 81.
- [9] Jedari E., Wu Z., Rashidzadeh R., Saif M. (2015) “Wi-Fi based indoor location positioning employing random forest classifier.” *2015 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2015* 13–16.
- [10] Jia Deng., Wei Dong., Socher R., Li-Jia Li., Kai Li., Li Fei-Fei. (2009) “ImageNet: A large-scale hierarchical image database.” In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 248–255.
- [11] Collobert R., Weston J., Bottou L., Karlen M., Kavukcuoglu K., Kuksa P. (2011) “Natural Language Processing (almost) from Scratch.” 1,.

- [12] Vo Q. D., De P. (2016) “A survey of fingerprint-based outdoor localization.” *IEEE Communications Surveys and Tutorials* 18, 491–506.
- [13] Xia S., Liu Y., Yuan G., Zhu M., Wang Z. (2017) “Indoor Fingerprint Positioning Based on Wi-Fi: An Overview.” *ISPRS International Journal of Geo-Information* 6, 135.
- [14] Otsason V., Varshavsky A., LaMarca A., de Lara E. (2005) “Accurate GSM Indoor Localization.” In, 141–158.
- [15] Laitinen H., Lähteenmäki J., Nordström T. (2001) “Database correlation method for GSM location.” *IEEE Vehicular Technology Conference* 4, 2504–2508.
- [16] ur Rehman W., de Lara E., Saroiu S. (2008) “CILoS: A CDMA Indoor Localization System.” *Proceedings of the 10th International Conference on Ubiquitous Computing* 104–113.
- [17] Krizhevsky A., Sutskever I., Hinton G. E. (2012) “ImageNet Classification with Deep Convolutional Neural Networks.” *Advances In Neural Information Processing Systems* 1–9.
- [18] Nowicki M., Wietrzykowski J. (2017) “Low-effort place recognition with WiFi fingerprints using deep learning.” *Advances in Intelligent Systems and Computing* 550, 575–584.
- [19] Kim K. S., Wang R., Zhong Z., Tan Z., Song H., Cha J., Lee S. (2017) “Large-Scale Location-Aware Services in Access: Hierarchical Building/Floor Classification and Location Estimation using Wi-Fi Fingerprinting Based on Deep Neural Networks.”
- [20] Kim K. S., Lee S., Huang K. (2017) “A Scalable Deep Neural Network Architecture for Multi-Building and Multi-Floor Indoor Localization Based on Wi-Fi Fingerprinting.” *the 9Th International Congress on Ultra Modern Telecommunications and Control Systems*.
- [21] Glorot X., Bordes A., Bengio Y. (2011) “Deep sparse rectifier neural networks.” *AISTATS '11: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics* 15, 315–323.
- [22] Torres-Sospedra J., Moreira A., Knauth S., Berkvens R., Montoliu R., Belmonte O., Trilles S., João Nicolau M., Meneses F., Costa A., Koukofikis A., Weyn M., Peremans H. (2017) “A realistic evaluation of indoor positioning systems based on Wi-Fi fingerprinting: The 2015 EvAAL-ETRI competition.” *Journal of Ambient Intelligence and Smart Environments* 9, 263–279.
- [23] Bellman R., Bellman R. E., Collection K. M. R. (1961) “Adaptive Control

Processes: A Guided Tour,” 1st edition, Princeton University Press.

- [24] Goodfellow I., Bengio Y., Courville A. (2016) “Deep Learning,” 1st edition, MIT Press.
- [25] Kingma D. P., Ba J. (2015) “J. ADAM: a method for stochastic optimization.” In Proc. International Conference for Learning Representations, 1–15.
- [26] Wu S., Zhong S., Liu Y. (2017) “Deep residual learning for image steganalysis.” Multimedia Tools and Applications 1–17.
- [27] Simonyan K., Zisserman A. (2014) “Very Deep Convolutional Networks for Large-Scale Image Recognition.” 1–14.
- [28] Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. (2015) “Going deeper with convolutions.” Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 07–12–June, 1–9.
- [29] Karpathy A., Toderici G., Shetty S., Leung T., Sukthankar R., Li F. F. (2014) “Large-scale video classification with convolutional neural networks.” Proc. IEEE CVPR 1725–1732.
- [30] Simonyan K., Zisserman A. (2014) “Two-Stream Convolutional Networks for Action Recognition in Videos.” 1–11.
- [31] Kim Y. (2014) “Convolutional Neural Networks for Sentence Classification.”
- [32] Chen H., Zhang Y., Li W., Tao X., Zhang P. (2017) “ConFi: Convolutional Neural Networks Based Indoor Wi-Fi Localization Using Channel State Information.” IEEE Access 5, 18066–18074.
- [33] Ibrahim M., Torki M., Elnainay M. (2018) “CNN based Indoor Localization using RSS.”
- [34] Mittal A., Tiku S., Pasricha S. (2018) “Adapting Convolutional Neural Networks for Indoor Localization with Smart Mobile Devices.” Proceedings of the 2018 on Great Lakes Symposium on VLSI 117–122.
- [35] Benesty J., Chen J., Huang Y., Cohen I. (2009) “Pearson Correlation Coefficient.” In Noise Reduction in Speech Processing Springer Berlin Heidelberg, Berlin, Heidelberg, 1–4.
- [36] Styan G. P. H. (1971) “Hadamard Products and Multivariate Statistical Analysis,” 1st edition, McGill University Press.

- [37] Torres-Sospedra J., Montoliu R., Martinez-Usó A., Avariento J. P., Arnau T. J., Benedito-Bordonau M., Huerta J. (2014) "UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems." In IPIN 2014 - 2014 International Conference on Indoor Positioning and Indoor Navigation, 261–270.
- [38] Goal C. (2017) "IPIN 2017 Track 3 Competition and Dataset Details." 1–8.
- [39] Torres-Sospedra J., Jiménez A. R., Knauth S., Moreira A., Beer Y., Fetzer T., Ta V. C., Montoliu R., Seco F., Mendoza-Silva G. M., Belmonte O., Koukofikis A., Nicolau M. J., Costa A., Meneses F., Ebner F., Deinzer F., Vaufreydaz D., Dao T. K., Castelli E. (2017) "The smartphone-based offline indoor location competition at IPIN 2016: Analysis and future work." *Sensors (Switzerland)* 17,.
- [40] Lohan E., Torres-Sospedra J., Leppäkoski H., Richter P., Peng Z., Huerta J. (2017) "Wi-Fi Crowdsourced Fingerprinting Dataset for Indoor Positioning." *Data* 2, 32.
- [41] Web 41, (2018), <http://indoorlocplatform.uji.es>, (Accessed on: 1/6/2018).
- [42] Barsocchi P., Chessa S., Furfari F., Potorti F. (2013) "Evaluating ambient assisted living solutions: The localization competition." *IEEE Pervasive Computing* 12, 72–79.
- [43] Torres-Sospedra J., Jiménez A., Moreira A., Lungenstrass T., Lu W.-C., Knauth S., Mendoza-Silva G., Seco F., Pérez-Navarro A., Nicolau M., Costa A., Meneses F., Farina J., Morales J., Lu W., Cheng H., Yang S.-S., Fang S.-H., Chien Y.-R., Tsao Y. (2018) "Off-Line Evaluation of Mobile-Centric Indoor Positioning Systems: The Experiences from the 2017 IPIN Competition." *Sensors* 18, 487.
- [44] Torres-Sospedra J., Jiménez A. R., Knauth S., Moreira A., Beer Y., Fetzer T., Ta V. C., Montoliu R., Seco F., Mendoza-Silva G. M., Belmonte O., Koukofikis A., Nicolau M. J., Costa A., Meneses F., Ebner F., Deinzer F., Vaufreydaz D., Dao T. K., Castelli E. (2017) "The smartphone-based offline indoor location competition at IPIN 2016: Analysis and future work." *Sensors (Switzerland)* 17, 1–17.
- [45] Lohan E., Torres-Sospedra J., Leppäkoski H., Richter P., Peng Z., Huerta J. (2017) "Wi-Fi Crowdsourced Fingerprinting Dataset for Indoor Positioning." *Data* 2, 32.
- [46] Yang D., Xue G., Fang X., Tang J. (2012) "Crowdsourcing to Smartphones : Incentive Mechanism Design for Mobile Phone Sensing." 18th Annual International Conference on Mobile Computing and Networking (MobiCom 2012) 173–184.

- [47] Hochreiter S., Frasconi P. (2001) “Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies.” In A Field Guide to Dynamical Recurrent Networks, 237–244.
- [48] Caruana R. (1997) “Multitask Learning.” Machine Learning 28, 41–75.
- [49] Zeiler M. D. (2012) “ADADELTA: an adaptive learning rate method.” arXiv preprint arXiv:1212.5701.

BIOGRAPHY

Furkan Tektas was born in 1991 in Istanbul, Turkey. He earned a Bachelor of Science degree in Computer Engineering from Gebze Technical University in 2015. Afterwards, he started pursuing his graduate research at the Institute of Natural and Applied Sciences of the same university. He is also a research assistant in Gebze Technical University Institute of Information Technologies.