الجمهوريـة الجزائريـة الديمقراطيـة الشعبيـة

وزارة التعليم العالـي والبحـث العلمـي

جـــامعــة قـــاصـدي مربـــاح ـــورقلة
كلية التكنولوجيات الحديثة للمعلومات والإتصال
قســـم الإلــكـــترونـيك والإتــصالات
المرجع: ....

MEMORY

In order to obtain the
Academic Master
Domain: Science and Technology
Field: Telecommunication
Specialty: Telecommunication Systems

Theme:

Towards Indoor Localization Guided by Machine Learning

Presented by: **ZERROUKI Fatma Zahra**
OUARGLI Lamia

Publicly presented
On: 06/22/2024
In front of the jury:

Mr. CHENINA Hachemi	President	MCB-UKM Ouargla
Mr. ELAGGOUNE Hocine	Supervisor	MCB-Univ-Guelma
Mr. MESSAADIA Mourad	Assistant Supervisor	PhD-Univ-Bahreïn
Mr. BENATHMAN Khaled	Examiner	MAA-UKM Ouargla

Academic year: 2023 /2024

بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

والْحَمْدُ لِلَّهِ رَبِّ الْعَالَمِينَ

والصَّلاَةُ وَالسَّلامُ عَلَى مَنْ بُعِثَ رَحْمَةً لِلعَالَمِينَ وَعَلَى آلِهِ وَصَحْبِهِ أَجْمَعِينَ.

Dedication ---

In the name of Allah, my Creator, the Facilitator of my affairs, and the Protector of my destiny. All praise and gratitude are due to You. The journey has not been short, nor was it meant to be. The dream was not close, and the path was not paved with ease. But I did it. I dedicate this success to my ambitious self first, and then to everyone who strived with me to complete this journey. May you remain my support forever. I dedicate this success to the one who Allah has honored with dignity and respect, to the one who taught me to give without expecting anything in return, to the one whose name I carry with pride. Oh, you who have been my support and still are, all praise be to Allah who has extended your life for me to be your second graduate, my father. You who cleared the thorns from my path to pave the way for me to knowledge. May your life be sweet, O master of men, and may my life be sweet, my dear father, Zrouki Abdel Fattah. To my angel in life, the light of my eyes, and the most precious thing I possess, my darling and the garden of my heart, who stayed up late and was with me in all my circumstances, situations, and pressures. To the woman who made me an ambitious girl who loves challenges. To the one whose prayers were the secret of my success and whose tenderness was the balm of my wounds. My role model, my teacher, and my friend of days, my dear mother, Kadri Nora, and her epitome. To my sisters, each by name and position: Hiba, Aaya, Batoul, Jouwayriya. To all my family and loved ones. To Mr. Majouja Jamal, thank you for all the guidance and valuable information you have given me. May Allah reward you for me with all good. I send a greeting whose letters are shy heads, and a greeting full of love and pride to every martyr who gave his soul for the homeland to live. To beloved Palestine, to those who made my soul happy with their closeness. To Taqwa Inqal, Ayah Ayah, Allah is the light and connection. To the prayer hall of the students of Khadija bint Khuwaylid, and all thanks to Allah who enabled me to reach this moment. Praying to Allah Almighty to benefit me from what I have learned and to teach me what I do not know and make it an argument for me, not against me.

ZERROUKI Fatma Zahra

إهداء

الحمدلله حتى يبلغ الحمد منتهاه الحمدلله لولا خالقنا وتيسيره ماوصلنا اللهم نحمدك على خيرك ونعمك اللهم كما اتممت فبارك .

يامن سهرت ليالي طوال تمسح عرق جبينها ,تصارع الام ظهرها ,منك استوحيت قوتي يا قلبي النابض ياعسل حياتي , جنتي امي فاطمة بلكحل كل ماوصلت اليه لك وبفضلك حفظك الله لي وشرفني بتشريف امراة عظيمة مثلك .

الى من تحمل حرارة الصيف ومشقة العمل , الى صاحب القلب الكبير فخري وعزتي ابي سندي ورقلي بوحفص مجهوذاتك ياعزيز قلبي سبب وصولي هنا ,اسعدك الله وجعلني مصدر فخرك .

الى الاعز والاحب الى عضدي وسندي اخوتي مراد ,سعد,محمد .الى فراشتي الجميلة اميرتنا ربحة الى أبناء عمي عزوتي عبد الرؤوف ,حذيفة,عبد الباري ,زكريا طاب بكم العمر وطبتم لي عمر شكرا لوجدكم في حياتي .

وتيني مصدر الهامي وقوتي اغلى الاحبة اختي مريم اسعدك الله وانار قلبك ,شكرا على دعمك المستمر

الى عمي مختار يا داعمي شكرا لأنك مختلف بطريقة تجعلني أدعو الله دائماً ألا يذيقني ألم فقدك او حُزنك

اخوالي علي ومحمد الفخر والعزة ادامكما الله لي ,زوجة خالي محمد وعائلته الملجئ الامن .

اختي الكبرى والملجئ الامن الداعم والسند زوجة عمي عائشة دمتي لي خير اخت افخر بقربها .

عمتي زهراء نور حياتي ياراحتي بعد التعب يامصدر تفاؤلي دعائك جعلني منا عليه الان كلي فخر لانك عمتي

الى سمية صاحبة اطيب قلب و الروح اللطيفة, زوجة خالي دهيبة حنونة القلب دمتم لي.

خالتي امي الثانية الصدر الرحب عائشة وابناءها .

إلى زهرات حياتي نعيمة , فيروز , فاطمة , سلاف .

إلى كل عائلة ورقلي صغيرها وكبيرها، شكراً لكم على الحب والدعم الذي لم ينقطع.

سلاف، رفيقة عمريِ يميني أنتِ الصديقة المختلفة الناس ناس إنما أنتِ الكتف

أمان قلبي ونور عيني وطمأنينة روحي . الى نور دربي ودليلي , رانيا وأسماء، رفيقتا السكن،اخواتي ذكرياتي الجميلة .

استاذي الكريم العقون حسين والأستاذ الفاضل مراد مساعدية جزيتم عنا خير الجزاء .

وإلى جميع الأهل والأصدقاء وكل من ساندني ، ووقف إلى جانبي ، وكان سبباً في تفوّقي بعد الله تعالى..

اللهم اجعلها بداية خير .

اللهم انصر فلسطين واهل فلسطين وفرج همهم .

لامية ورقلي

Congratulations!

Acknowledgments --

Praise be to Allah, the Almighty, who has enabled us to complete this Memory.

First and foremost, we would like to express our gratitude to our advisors, **Mr. Hocine Elaggoune** *and* **Mr. Mourad Messaadia**, *who have encouraged and driven us to complete this work within the set deadlines.*

We extend our sincere thanks to the president and members of the jury who have honored us by accepting to evaluate this work. We acknowledge each by name:

Mr. CHENINA Hachemi: *President of the jury.*

Mr. BENATHMAN Khaled: *Examiner.*

Our thanks also go to the educators and teachers who perform their duties selflessly, seeking no reward except from Allah.

Lastly, we extend our heartfelt thanks to all those who have supported us or contributed to the completion of this work.

ZERROUKI F.Z & OUARGLI L

ملـخص:

لقد أصبح تحديد الموقع الجغرافي في داخل مبنى مسألة هامة في العديد من المجالات، مثل الملاحة الداخلية (*Indoor Geolocation*)، ولوجستيات المستودعات، أو حتى التسويق المستهدف في مراكز التسوق.

تُظهر تقنيات تحديد الموقع الجغرافي التقليدية، مثل نظام تحديد المواقع العالمي(**GPS**) ، حدودها في البيئات الداخلية بسبب فقدان الإشارة الناتج عن هياكل المباني. وبالتالي، توفر شبكات **WI-FI** وأجهزة استشعار الهواتف الذكية بديلاً واعدًا لتحديد الموقع الجغرافي داخل المباني.

يهدف هذا البحث في درجة الماستر إلى تحليل الأعمال القائمة واقتراح حل دقيق وقوي لتحديد الموقع الجغرافي الداخلي من خلال دمج إشارات شبكات **WI-FI** المحيطة وأجهزة استشعار الهواتف الذكية (مثل مقياس التسارع، الجيروسكوب، الميغناتومتر، إلخ). الهدف الرئيسي هو تصميم خوارزمية دمج البيانات التي تجمع بفعالية المعلومات الواردة من هذه المصادر المختلفة للحصول على تحديد موقع دقيق في الوقت الحقيقي.

الكلمات المفتاح: تحديد الموقع الجغرافي ، ويفي ، الهواتف الذكية، أجهزة الاستشعار، خورزمات.

Abstract :

Indoor geolocation has become a major focus in various fields, such as indoor navigation, warehouse logistics, and targeted marketing in shopping centers. Traditional geolocation technologies, like GPS, show their limitations in indoor environments due to signal loss caused by building structures. As a result, Wi-Fi networks and smartphone sensors present a promising alternative for indoor geolocation.

This master's thesis aims to analyze existing work and propose a precise and robust indoor geolocation solution by integrating signals from surrounding Wi-Fi networks and smartphone sensors (such as accelerometers, gyroscopes, magnetometers, etc.). The primary objective is to design a data fusion algorithm that effectively combines information from these different sources to achieve accurate real-time localization.

Keywords: *Indoor Geolocation; Smartphone; Wi-F; Deep Learning; LSTM (Long Short-Tem Memory); BiLSTM (Bidirectional Long Short-Tem Memory).*

Résumé :

La géolocalisation indoor est devenue un enjeu majeur dans de nombreux domaines, tels que la navigation en intérieur (indoor geolocation), la logistique des entrepôts, ou encore le marketing ciblé dans les centres commerciaux. Les technologies de géolocalisation traditionnelles, comme le GPS, montrent leurs limites en environnement intérieur en raison de la perte de signal relatif aux structures des édifices. Ainsi, les réseaux Wi-Fi et les capteurs des smartphones offrent une alternative prometteuse pour la géolocalisation indoor.

Ce stage de master vise à analyser les travaux existants et proposer une solution de géolocalisation indoor précise et robuste en intégrant les signaux des réseaux Wi-Fi environnants et les capteurs des smartphones (tels que l'accéléromètre, le gyroscope, le magnétomètre, etc.). L'objectif principal est de concevoir un algorithme de fusion de données qui combine efficacement les informations provenant de ces différentes sources pour obtenir une localisation précise en temps réel.

Mots clés : *Géolocalisation intérieur; Smartphone; Wi-Fi; Apprentissage Approfondi; LSTM ; BiLSTM.*

Contents

List of Figures

List of Tables

Acronyms

AOA	Angle Of Arrival
AP	Access Point
APD	Access Point Deployment
BLE	Bluetooth Low-Energy
BSS	Basic Service Set
BSA	Basic Service Area
BiLSTM	Bidirectional Long Short-Tem Memory
CSI	Channel State Information
ESS	Extended Service Set
GPS	Global Positioning System
IPS	Indoor Positioning System
J48	Decision Trees
LSTM	Long Short-Tem Memory
MAC	Mount Access Controls
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MSE	Mean Squared Error
NB	Naïve Bayse
NLOS	Non-Line-Of-Sight
RF	Random Forest
RFID	Radio Frequency Identification
RMSE	Root Mean Squared Error
RN	Reference Node: in the context of network-based localization, a reference node, also known as an anchor node or beacon node, is a node with a known position used to help determine the positions of other nodes in the network. For example, in Wi-Fi-based indoor localization, reference nodes are often Wi-Fi access points with known locations.

RSS	Really Simple Syndication
RSSI	Really Simple Syndication Indicator
STA	Station
TDoA	Time Difference Of Arrival THR Threshold
ToA	Time Of Arrival
UWB	Ultra-Wide Band
Wi-Fi	Wireless Fidelity
WLANs	Wireless Local Area Networks
WM	Wireless Medium
WSMM	Wi-Fi Signal Measurement Mechanism

Chapter 01. General introduction

1.1. Context

Indoor localization, which refers to the ability to precisely determine the position of an object or person within a building, is a rapidly growing field with a wide range of potential applications, including indoor navigation, asset management, and security [1]. The use of machine learning in this context paves the way for more accurate and robust solutions by leveraging data collected from various sensors such as Wi-Fi signals, inertial sensors, cameras, and Bluetooth beacons.

The integration of machine learning into indoor localization offers a promising solution to the challenges posed by traditional methods. By utilizing data from diverse sensors, machine learning algorithms can significantly enhance the precision and robustness of indoor positioning systems. This advanced approach not only improves the accuracy of location estimates but also adapts to the dynamic and complex nature of indoor environments.

In this work, we focus on exploring indoor localization guided by machine learning, exploiting data collected from various sensors, particularly Wi-Fi signals. This approach provides significant benefits for a wide range of applications and industries, offering improved accuracy and adaptability in indoor positioning systems.

1.2. Problematic

Achieving high accuracy and precision in indoor environments is difficult due to the complex nature of signal propagation. Factors such as multipath propagation, signal attenuation, and interference from building materials and other electronic devices can significantly degrade the quality of the localization data.

In addressing this problem, the integration of machine learning techniques into indoor localization systems offers a promising avenue. Machine learning algorithms can learn from vast amounts of data to improve localization accuracy, adapt to changing environments, and fuse data from multiple sensors effectively.

1.3. Objectives

The primary objective of this research is to develop and evaluate machine learning-based techniques for indoor localization that leverage data from various sensors to provide accurate, reliable, and scalable positioning solutions. To achieve this overarching goal, the specific objectives of the study are as follows:

Data Collection and Preprocessing and Development of Machine Learning Models...

By proactively addressing this issue can help overcome the challenges associated with machine learning-guided indoor localization and develop more accurate, reliable and robust solutions.

1.4. Contributions

This research on indoor localization guided by machine learning aims to make several key contributions to the field, for example:

Comprehensive literature review, dataset creation and new machine learning models....

1.5. Hardware and software implemented for conducting our experiments

Our work was carried out using a computer with an Intel(R) Core(TM) i5-4300M CPU @ 2.60GHz 2.90GHz processor and 4.00 GB of RAM. Smartphone model: OPPO A31 with an accelerometer, gyroscope, and compasses The programming and simulation software used is Python 3.12.4.

1.6. Memory organization

The thesis manuscript is structured around three chapters: after introducing the state of the art in Chapter 1, Chapter 2 is dedicated to our proposed approach based on modeling and machine learning. Chapter 3 focuses on the implementation and performance evaluation of the proposed approach. Finally, we conclude the thesis with a general conclusion and future perspectives.

Chapter 02. *State of the art*

2.1. Introduction

Wireless signals have been widely used for indoor positioning. The most popular signals include those used in cellular phones and WLAN. Pervasive adoption of WLAN enables mobile devices to use WLAN signals for indoor positioning. In principle, WLAN signals can be used for various types of positioning such as one time, continuous, and even 3D tracking, and positioning with low-performance mobile platforms (e.g., mobile phones, PDAs, and laptops). Moreover, WLAN infrastructure is easy to deploy in most indoor environments. With such infrastructure, no additional deployment cost is necessary for an access point. Therefore, WLAN presents a good candidate for indoor positioning [2].

An indoor positioning system (IPS) is a system to locate objects or people inside a building using lights, radio waves, acoustic, Wi-Fi, or other signals collected by mobile devices. There are several commercial systems available, but they typically use expensive infrastructures that require the installation of specific hardware in the building. Existing indoor geolocation systems generally require floor plans or cumbersome calibration processes in order to enable an RSS approach. In some cases, proprietary and specific software capable of creating maps with this data is used, which is not directly compatible with popular mapping systems. They cannot be achieved by users to generate their own maps and RSS distributions [3].

2.2. Wi-Fi Localization Techniques

Wi-Fi location systems have gained significant importance across various domains, such as indoor navigation, warehouse logistics, and targeted marketing in shopping centers. Traditional geolocation technologies, like GPS, are limited in indoor environments due to signal loss caused by building structures. Wi-Fi networks and smartphone sensors offer a promising alternative for indoor geolocation. This chapter provides an overview of Wi-Fi-based localization techniques, associated challenges, and recent advancements in the field.

Wi-Fi-based localization techniques can be categorized into several primary methods:

2.2.1. Really Simple Syndication (RSS)

RSS, which stands for Really Simple Syndication, is a web feed that allows users and applications to access updates to websites in a standardized, computer-readable format. It is widely used for delivering regularly changing web content from blogs, news sites, and other online publishers.

RSS-based localization requires trilateration or alteration with N points. In this approach, the Received Signal Strength (RSS) at the device is used to estimate the absolute distance between the user's device and at least three reference points. Basic geometry and trigonometry are then applied to determine the device's position relative to a reference point, as shown in (Fig 2. 1) [4][5].

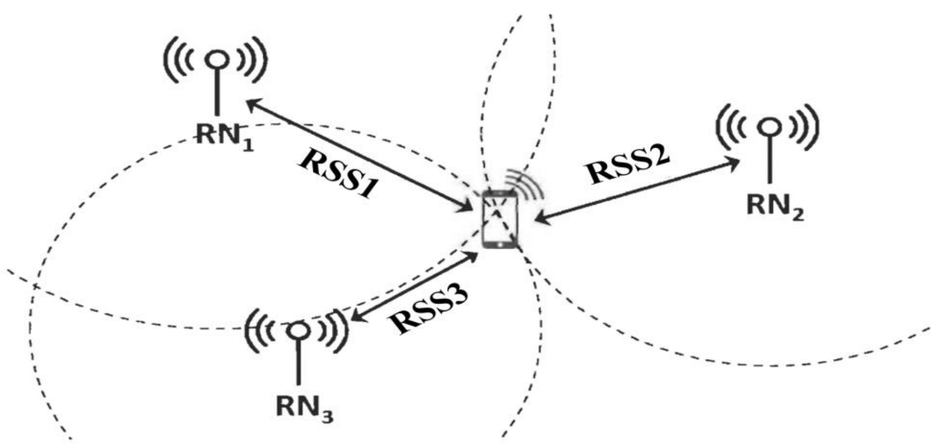

Fig 2. 1. RSS Radio Propagation Technology

The RSS indicates (RSSI) the strength of the received radio signal, which is mainly measured in dBm ($1dBm = 1.259\ milliwatts$). The closer the receiver is to the transmitter, the stronger the signal, because the signal strength is attenuated as it propagates through the air, and the attenuation is distance proportional. One assumes that, with constant environmental circumstances, the nearer the source, the stronger the processed signal. On the other hand, the following limitations of $RSSI$ include: it is affected by physical objects such as furniture and walls, which cause significant fluctuations of signal strength and thus affect the accuracy of the distance reading; other factors in the environment, like temperature and humidity variations, affect the propagation of a signal. Besides, unequal power settings for different devices can give different RSSI readings and, thereby, different measurements. However, despite such drawbacks, RSSI finds application in indoor geolocation for asset tracking, indoor navigation, and systems for location-based advertising [5].

$$RSSI(d) = -10nlog10(d) + A \qquad (2.1)$$

$$n = path\ loss\ exponent, A = environment\ constant$$

Where: RSSI refers to the Received Signal Strength Indicator, usually indicated in dBm. Is the path loss exponent, and depending on the variability of the environments, typical values usually are 2 to 4 for indoor environments?

Represents the distance of the location from the transmitter to the receiver, measured typically in meters. A stand for the value of the RSSI at a reference distance, which is usually 1 m, and expressed in dBm.

Two methods have been used to estimate the location of a smartphone based on the RSS technique.

a. Pseudo-Range Measurement Method:

This method is based on a known radio propagation analysis relationship. It employs trilateration to determine the locations of smartphones from the estimated pseudo-ranges between a smartphone and multiple $\boldsymbol{BSs/WAPs}$ (See Fig 2. 1) [1].

To estimate the pseudo-range between smartphones and $\boldsymbol{BSs/WAPs}$, equation (2.2) is used:

$$p_i = p_0 * 10\left(\frac{RSS_{i0} - RSS_i}{10_i^{ni}}\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots (2.2)$$

Where $\boldsymbol{p_i}$ is the pseudo-range between the smartphones and the $\boldsymbol{BSs/WAPs}$, $\boldsymbol{p_0}$ is the calibrated pseudo-range estimated at zero distance $\boldsymbol{RSS_{i0}}$ is the measured signal power value for

p_0, RSS_i is the measured signal power for the received **BSs/WAPs** signals, and ni is the path loss exponent calculated/calibrated for the received **BSs/WAPs** signals.

b. RSS Fingerprinting:

RSS-based localization systems are typically implemented in $802.11\ wireless$ local area networks to determine the user's location by measuring frames sent from different access points (APs) [1]. The principle of this method is to create a map of an area based on the RSS signal strength fingerprints. The area is divided into sub-zones, each with its own signal strength fingerprint. Thus, smartphone geolocation is based on matching the pre-stored RSS values in a database [1].

In this method, both offline and online phases should be conducted to calculate the smartphone's location. These steps, along with their geolocation process, are shown in Fig 2. 2.

During the offline phase, a radio map (Database) for the signal strengths at different points (Reference Nodes) of the main $\boldsymbol{BSs}$ around an area must be recorded. In the online phase, a matching process between the real-time RSS and the recorded data from the predefined radio map is used to estimate the smartphone's location.

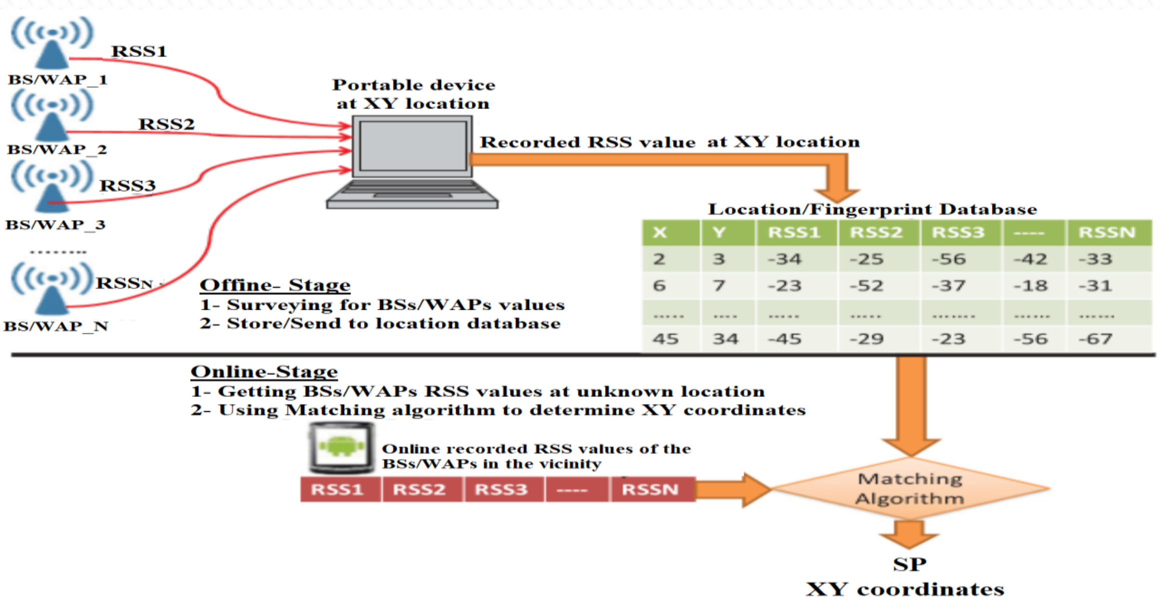

Fig 2. 2. Determining Smartphone Position Using the RSS Fingerprinting Technique[1].

2.2.2. Time of Arrival (ToA)

The concept behind The Time of Arrival method is determining the distance the signal has travelled from the transmitter to the receiver by guessing the time the signal has taken. This is basically and mainly carried out by the electromagnetic waves that transmit at the speed of light. This distance is thus calculated by multiplying the travel time by the speed of light. The ToA methods are pretty accurate if the precise arrival time of signals is determined. The problems generally encountered in implementing ToA methods in indoor environments are due to the multipath propagation, which results in multiple signal reflections, each suffering varied times of arrival in the measurement process. Also, the synchronization of the transmitter to the receiver has to be maintained for accurate ToA measurement, which could be quite a task indoors. These shortcomings aside, the ToA technique is very broadly used in applications that require high-precision distance measurements [6].

As given in (Fig 2. 3), the distance between the sensor and the target can be determined so long as the ToA is measured. Then, the intersection of a circle and a distance determined by the sensor are the positions of the measured target. One wants to get a certain single point of articulation that ought to have at least three circles, which means at least three sensors for a 2-D positioning system[7].

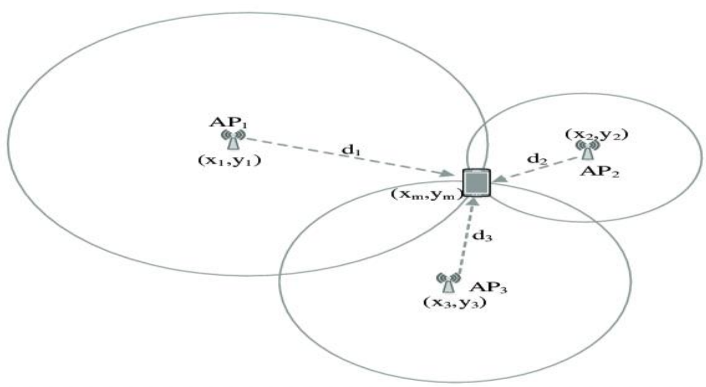

Fig 2. 3. Localization based on time of arrival (ToA) measurement

2.2.3. Time Difference of Arrival (TDoA)

TDoA is based on the simple principle of differential signal transmission to multi-receivers, allowing for the calculation of the hypothetical position of any source. The difference in arrival time of the different APs is used to measure the distance difference between APs and the device being tagged. As opposed to ToA, TDoA is not supposed to be synchronized concerning time between transmitter and assurer but relies only on synchronization between the receivers, which reduces the requirements. TDoA is also effective in a mean-path environment since it uses advanced signal-processing technologies. This can simply be put into indoor positioning technologies that will offer users accurate location information important in navigation for complex environments like a mall or an airport [6].

The figure (Fig 2. 4) is the diagram where it shows the different uses of the three different reference nodes for achieving the 2-D position of any target. It also shows the hyperboles. These arise from the results of the measurement of the user's position (black point) obtained from the nodes of references[7].

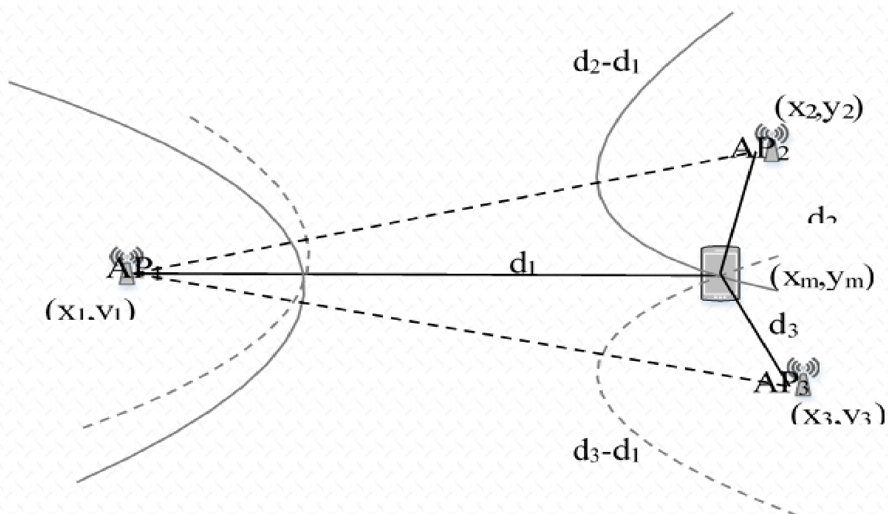

Fig 2. 4. Localization based on time difference of arrival (TDoA) measurement

2.2.4. Angle of Arrival (AoA)

The AoA is an advanced technique in determining the relative angle of signal arrival concerning a given reference direction. One will achieve this by placing a representative number of antennas in an array, thus enabling the automatics of the system to detect varieties between received signals from different directions, be it phase or strength variations. AoA-based localization systems can, therefore, estimate the position of a tag device by analyzing the intersection points of hypothetical paths at certain angles. The accuracy applied using AoA in an indoor scenario immensely lies not only in the quality and configuration of the antenna array but also in technologies applied for the processing of signals. Nonetheless, several challenges affect the accuracy and applicability of AoA [1].

Of course, with this comes many different demands on hardware: frequently quite sophisticated and elaborate antenna arrays that are unwieldy to install, require large spaces, and therefore shouldn't be inexpensive. Aside from this, indoor conditions mean certain additional specific problems: critical reflections and diffractions from walls and other obstacles distort angle measurements and usually detract from accuracy. After all those complications, AoA seems to be a constructive localization approach hosting various unique strengths elaborated from application areas like indoor navigation, asset tracking, and location-based services. Its utility, however, depends on the use case and the environment: every scenario bears peculiar strengths and limitations.

In Angle of Arrival (AoA) localization, the position of the target is determined at the intersection of multiple pairs of angular direction lines, each extending from a base station or beacon station to the moving target . As illustrated in Figure (Fig 2. 5), AoA can use at least two known reference nodes, A and B, and two angles (θ1, θ2) to obtain the 2-D position of the target P [1].

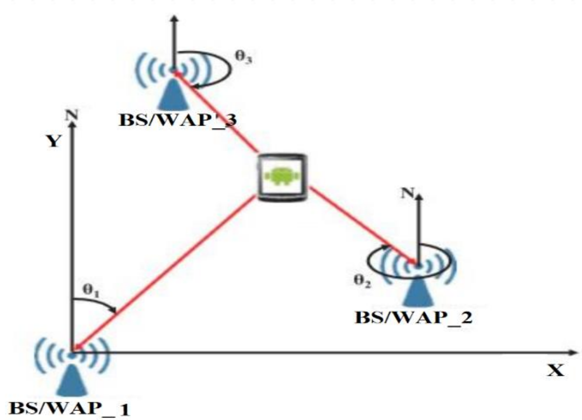

Fig 2. 5. Localization based on angle of arrival (AOA) measurement

The definition of the smartphone's position for 2D coordinates is expressed in Equation (2.3):

$$(x_i - x_{sp})\sin(\theta_i) = (y_i - y_{sp})\cos(\theta_i) \dots\dots\dots\dots\dots\dots (2.3)$$

where $\boldsymbol{x_i}$ and $\boldsymbol{y_i}$ are the **XY** coordinate values of the **BS / WAP**, $\boldsymbol{\theta_i}$ is the AoA for the received **WAP** signals, and $\boldsymbol{x_{sp}}$ and $\boldsymbol{y_{sp}}$ are the **XY** coordinate values of the smartphone's position.

2.3. Wi-Fi technologies

Wi-Fi (Wireless Fidelity) refers to a family of wireless networking protocols based on the IEEE 802.11 standards, designed for local area networking of devices and internet access over short distances using radio waves.

2.3.1. IEEE 802.11 standards

IEEE802.11 is a set of international standards that focus on the specifications of wireless LANs. In 1999, the IEEE802.11b that operated at 2.4GHZ frequency was introduced at a maximum speed of 11Mbps. Followed by IEEE802.11g in 2003, with a predecessor's frequency, but with a maximum 54Mbps transmission speed. The IEEE802.11n adopted in 2009 operates on both 2.4Ghz and 5Ghz frequency bands, achieving a maximum data transfer rate of 600Mbps. In 2013, The IEEE802.11ac standard has emerged based on 2.4Ghz and 5Ghz frequency bands and provides a maximum speed of 1.3Gbps. Last, the IEEE802.11ax that has been launched in 2019, At all 2.4Ghz and 5Ghz frequency bands it also provides data transfer speeds up to 9.6Gbps. Below is a (**Table 2. 1.**) containing differences between the above standards [8].

Table 2. 1. Differences between Wi-Fi standards

Wi-Fi Generation	IEEE Standard	Release Date	2.4 GHz	5 GHz	Maximum Data Rate
Wi-Fi	802.11	1997	Yes	No	2 Mbps
Wi-Fi 1	802.11b	1999	Yes	No	11 Mbps
Wi-Fi 2	802.11a	1999	No	Yes	54 Mbps
Wi-Fi 3	802.11g	2003	Yes	No	54 Mbps
Wi-Fi 4	802.11n	2009	Yes	Yes	600 Mbps
Wi-Fi 5	802.11ac	2013	No	Yes	6.93 Gbps
Wi-Fi 6	802.11ax	2019	Yes	Yes	4x 802.11ac

Wi-Fi network architecture covers the overall design and layout of Wi-Fi networks that conforms to the IEEE 802. 11 standards. This architecture is very crucial for wireless communication and networking in homes, offices and other public places. And here are the most important components [8] :

- **Basic Architecture:** The organization of IEEE 802.11 Wi-Fi standard involves several building blocks. Some of the important components and functions are the following.
- **Station (STA):** As another example, any network with IEEE 802.11conforming medium access control.
- **Mount Access Controls** (**MAC**): and physical layer (PHY) which works as an interface for communicating with the wireless medium (WM).
- **Access Point (AP):** A window would serve as an important educational and promotional tool for all space stations in the international space partnership that have such capabilities and allow people to be there. Radio power distribution services for their branch stations by the wireless medium.
- **Basic Service Set (BSS):** A BSS (Basic Service Set in IEEE802.11 theory) is the fundamental constructor of the wireless network. 802.11 wireless LAN. BSS, a seem to be any number of station group is a babel.
- **Basic Service Area (BSA):** It is the conceptual space within which, the members who I have invited, believe, in and love the Basic Service Unit. Set may communicate.

The figure (Fig 2. 6) shows a diagram of the Expanded Wi-Fi Service System (ESS).

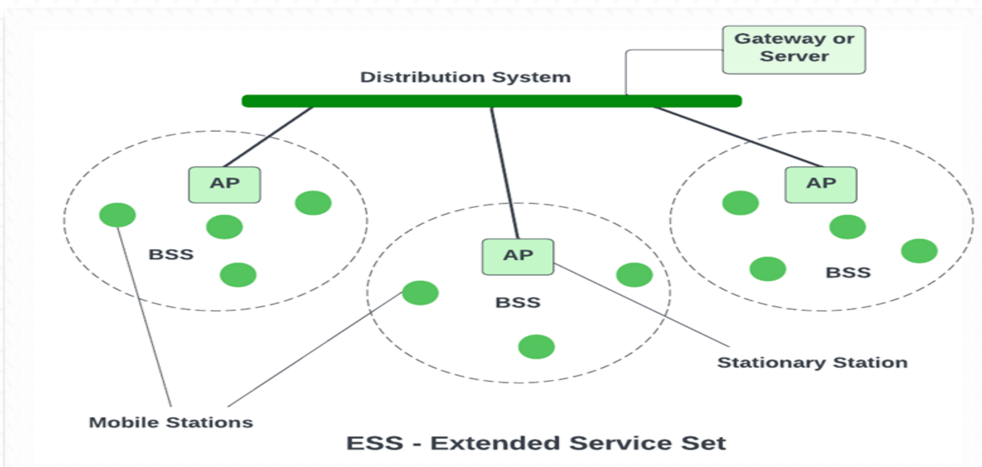

Fig 2. 6. Basic Service Set (BSS) of IEEE802.11 WLAN

2.3.2. Signal measurement mechanisms

The Wi-Fi Signal Measurement Mechanism (WSMM) is used to calculate the strength of the Wi-Fi signal in an area. There are two common mechanisms used to measure [8]:

- **Wi-Fi signal strength:** There are two common mechanisms used to measure Wi-Fi signal strength.
- **Received Signal Strength Indicator (RSSI)**: RSSI is a power measurement of the received Wi-Fi signal at the source. It is measured in units of decibels (dB) where the power of the received signal (referred to as the subject) is compared to another signal of a specific power (defined as reference). RSSI is calculated as follows:

$$RSSI = 10 * log10\,(signal\ power/Reference\ power) \quad (2.4)$$

2.3.3. Channel State Information (CSI)

CSI is the quality metric for the noise motherboard connectors. The value indicates the strength of the transmission signal and errors in data transmission with respect to received data packages. CSI is being used to check the channels conditions and choose the data transmission based on them.

In most wireless systems, including IEEE 802.11 and UWB, smaller coherence bandwidth of the wireless channel bigger compared to the bandwidth of the signal which is what signals the channel selective in frequency (i.e., each frequency may have its own amplitude and phase behaviour). Further, in several antenna transceivers, the channel frequency responses for each antennae pair can differ greatly; distance and signal wavelength). Even though RSS has been done on used due to its simplicity and lower hardware requirements, it computes only an average amplitude estimate over an average the entire signal bandwidth and the integrated signal over all the antennas. This makes RSS sensitive to multipath effects and influence and yield high variability through time.

On the other hand, channel impulse response (CIR) or its Fourier Transform Pair; The Channel Frequency Response (CFR) That is normally reported to higher layers in the form of channel state information (CSI), the granularity of which is higher than that of the RSS it can Capture Amplitude and Phase responses of the channel in various frequencies and between disparate transmitter-receiver antenna pairs. In general, the CSI is a complex amplitude, and then in polar form it can be expressed as:

$$H(f) = |H(f)|\, e^{j\angle H(f)} \qquad (2.5)$$

Where: |H(fi)| is the amplitude (or magnitude) response The phase response of the frequency fi of is represented by the channel. Nowadays, there are plenty of IEEE 802.11 NICs cards, Give subcarrier-level channel measures for Orthogonal Frequency Division Multiplexing (OFDM) systems that can translated into the richness of multipath information, more stable Measurement and localization to a finer accuracy [9].

2.3.4. Challenges and solutions in indoor environments

Indoor-related Wi-Fi technologies suffer from critical impairments that include high attenuation, multipath distortion, interference, non-line-of-sight (NLOS) propagation, fast changes of channel conditions, and shadowing. These are caused by the signal weakening by walls, objects, and reflections, and multipath effects created by the high-density electronic interferences. Moreover, the characteristics of the unpredictable route taken by the signals under NLOS conditions and the dynamic nature of spaces indoors by movement of both people and objects and outdoors by moving vehicles lead to fluctuating channel conditions, which results in decreased signal visibility, especially within cluttered office and home environments.

Some of the solutions implemented against those challenges include using multiple antennas to minimize the problem of multipath fading and, consequently, further increase the precision of the signal. This is due to the reason that the best choice of channels and power control overcome the interference without compromising the strength of the signal such that it avails efficient communication. Other advanced techniques employed include diversity reception and spatial multiplexing in managing multipath fading to minimize fading and provide a reliable signal. In addition, more integration of location technologies, like BLE and ultrasound, can provide better location accuracy, which would be sufficient for efficiently tracking items and inventory management within a warehouse. The collected strategies aim to achieve optimal Wi-Fi performance in complex indoor environments to ensure consistent connectivity and accurate positioning [9].

2.3.5. Wi-Fi Network Architecture

Wi-Fi network architecture refers to the structure and organization of components that enable wireless communication within a defined area. Understanding this architecture is essential for implementing effective indoor localization systems. Key components and concepts include:

a. Deploying Access Points

Some recent challenges of broadband delivery in Wi-Fi connectivity include streetlight Wi-Fi infrastructure delivery. In terms of coverage, one of the biggest things to worry about is the Wi-Fi and location extension capability. For these purposes, APs must be located in strategic locations. Surveying is done to crosscheck weak coverage areas, demands, and effective placement for each access point. In most of the cases, the access points are mounted on the ceiling because "they scatter in all directions, hence giving a wider coverage and in most of the cases uniform coverage with minimal interference with objects like furniture and tables.".

The central region of the entire coverage area is the usually central position in which access points are installed so that the whole space can fully be covered. Physical obstructions must also be avoided, such as walls, ceilings, and furniture, because these will block wireless signals. It is thus so that the access points are located in strategic positions to reduce the number of such obstructions. In significant spaces, too, there should be several access points to avoid drop-outs, the speed of access, a process that will limit the traffic in the network and facilitate a positive user experience. The efficiency of the entire project is equally determined by the height at which it will be mounted. More often, this height may be between 6 to 8 feet because that height promotes good signal dispersal. Generally, with these considerations that include site surveys, centralized placement, no obstacles, multiple access points wherever required, and at the right height, broadband connectivity can be increased many times over using the infrastructure of streetlights. This can allow a complete and reliable Wi-Fi setup [10].

b. Channel Selection

That of course is the reason that increasing the efficiency of Wi-Fi lies in the ability to choose the right channel: it is one of the principal steps in the fight against interference. Knowledge of the frequency spectrum from 2 million to 200 million and the ability to seize control of it for own needs. A basic feature which bands 4GHz and 5GHz support is crucial. Within the 2. 14 channels are available at 4GHz band; limitations arising from interferences and decreased data transfer rates are usually a result of channel overlaps. On the other hand, the 5GHz band offers a wide range of the spectrum with 23 channels despite being a few not efficient. Wireless LAN utility applications arise as inevitable to provide an understanding of channel use and evaluate occupancy. With this knowledge on hand, networks are therefore in a position to deploy channels as they intend to in a given optimised manner. Moreover, there's more to signal processing than Wi-Fi basics; applying procedures like particle filters enhances Wi-Fi competence in complex residential spaces. In regards to non-Linear conditions and different distributions they perform better than the traditional techniques like Kalman filters and clearing the way to improved networks. These principles and tools complement each other improving the chances of better Wi-Fi management and also guarantee proper performance in diverse environments thus reinforcing the need for inclusion in today's network enhancement activities [12].

The Kalman filter: that filter created back in 1960 by R. E. Thatcher is a method, for estimating the parameters of noisy systems. This filter is particularly skilled at estimating these parameters with error even when dealing with systems that are not fully understood. It is especially adept at handling processes, like combining motion data with Wi Fi data in indoor settings. By using models and considering possibilities particle filters can approximate machine movement and improve these estimations by giving more weight to observations. This allows them to handle distributions and tackle nonlinearities better than traditional Kalman filters do [12].

c. Network Management

This concept is very much essential in enhancing the Quality of Service and in the sustainability of networks. Monitor Network Performance: It equally important to monitor the performance of the network so as to check for signs of trouble and correct them in their earliest possible stages. It may also call for a network management that includes real-time updates of traffic flow, packet losses among other capabilities [13].

- **Optimize Network Configuration:** Optimization of the network configuration can thus be seen as the cornerstone in the improvement of the performance of the complex network. This, for instance, entails configuring it for the correct number of access points and the right channel, applying a QoS policy.
- **Implement Network Security:** In order to have a strong security mechanism in the network any intrusion and vindictive attack has to be prohibitive. It entails the use of firewalls, the use of encrypted codes, and restricted admittance to critical data.
- **Use Network Management Tools:** Network diagnostics utilities are said to assist in easing the task of carrying out activities involved in managing networks since they provide an interface through which these tasks can be performed. These tools can also

facilitate in discovering the current and providing recommendations for the enhanced network.

d. Indoor Localization and Wi-Fi Network Architecture

For indoor localization, the Wi-Fi network architecture plays a pivotal role. The placement and density of APs affect the accuracy of localization. Techniques such as RSS fingerprinting, trilateration, and advanced machine learning models rely on the signals received from these APs. Understanding the architecture allows for optimizing the network setup to improve localization precision and robustness.

2.4. Smartphone sensors

Some sensors an electronic device that measures the changes in electrical or physical signals and the response is in digital form of which it automatically generates. The various enhancements in smartphone devices can result in both internal effects within them, or externally in the surrounding environment. The number of sensors in a conventional smart phone amount to 14. There are some of which are more useful for the indoor localization although others are not (See Table 2. 2.)[10].

Table 2. 2. Smartphone sensors and their description

Sensor type	Description
Accelerometer	Measuring the changing rate of the device acceleration
Ambient light	Measuring the ambient visible light's intensity
Barometer	Measuring the atmospheric pressure
Bluetooth	Communicating with nearby Bluetooth beacons or other Bluetooth-enabled devices
Camera	Capturing the scenery via the front or back lenses
Cellular	Communicating with nearby cell towers
FM	Receiving information from nearby radio towers
GNSS	Receiving the satellite signals to compute the latitude and longitude of the phone
Gyroscope	Measuring the changing rate of the device's tilting angle
Magnetometer	Measuring the ambient magnetic field strength
Microphone	Capturing the ambient acoustic noise
NFC	Communicating with nearby RFID tags
Proximity	Measuring distance to the nearest object within 10 cm
Wi-Fi	Communicating with nearby access points or other Wi-Fi-enabled devices

Although there are some specific key findings about smartphone sensors that were pointed out, these sensors, apart from GNSS, were intended primarily for other functions and not necessarily for indoor positioning. Therefore, these are the ones. The sensor is quite compact. Over time, the sensors become smaller as they need to be squeezed into the diminishing body of the phone. The challenge is that their sensitivity goes down with this (i.e., the SNR goes down as well). The measures are noisy. These involve transmitting and receiving noise from the sensors located one directly on top of the other and the interference of other devices that are conductive, such as the battery. The sensors are heterogeneous. The number of smartphone manufacturers and the range of sensor suppliers create a headache in getting the data normalized across devices. The internal architecture is challenging. The building material is determined by the huge ferrous metal components, such as metal bars, steel rebar's, and reinforced concrete, which cause some sensors to misread. The condition indoors is continuous. The building generally has diverse users walking around inside when active, which tends to degrade some of the wireless signal-based devices [10].

On the whole, the majority of indoor positioning systems are geared toward two objectives, that is to say, tracking and navigation. Tracking is possible, at any particular moment when it is necessary, to detect the position of the person or object and then pinpoint it. The aim of this setup is to find the best option available for users when they are navigating (See Fig 2. 7).

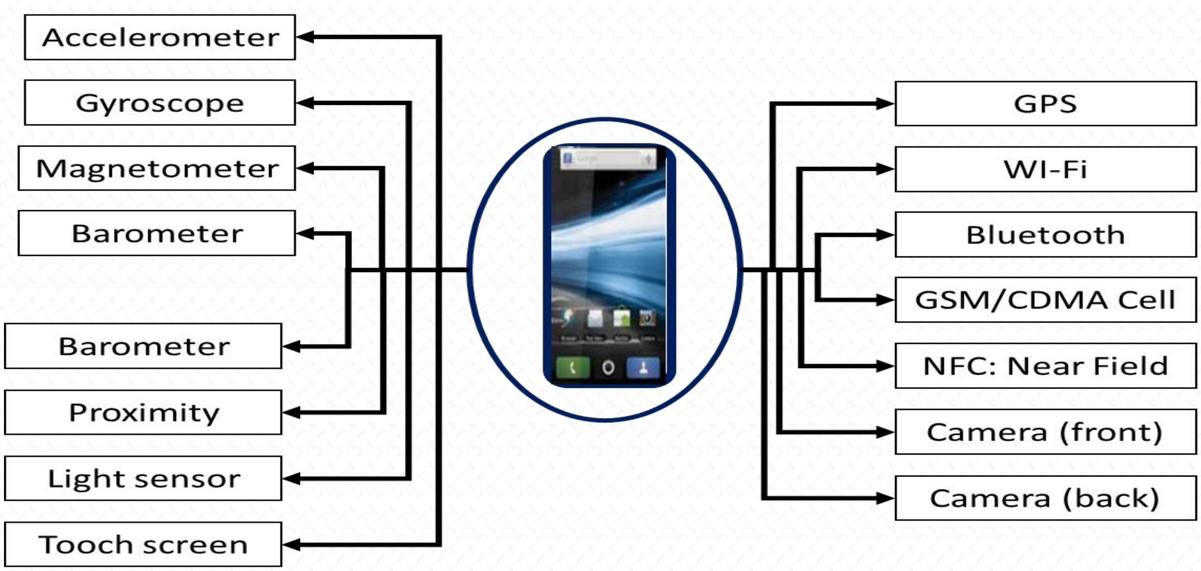

Fig 2. 7. Different sensors inside smartphone

2.5. Conclusion

This work, which is a survey, has comprehensively explored the current on indoor geolocation comprising various approaches and technologies. We went into the techniques of RSS, ToA, TDoA, and AoA in the context of Wi-Fi-based, discussing their principles, applications, and inner obstacles. The testing activities covered other options including GPS, Bluetooth, UWB, and RFID, all of which have their own advantages and disadvantages for use indoors. Moreover, I split Wi-Fi technology down into its roles and functions such as the IEEE 802. The 11 standards and signal detection methods, also the structure of the Wi-Fi networks are the key elements of the Wi-Fi. The study however, addressed the major issues that these technologies are facing, including but not limited to, attenuation, multipath effect, and interference in intricate indoor settings.

It was also evident in the assessment of smartphone sensors such as accelerometers, compasses, gyroscopes, and GPS which showed how they significantly contribute and equally limit the enhancement of indoor geolocation accuracy. In sum, the research shows the complexity and technical diversity because of indoor geolocation, yet, it also demonstrates the innovation and integration when reducing and improving indoor geolocation services. In this way, we master the technology and its mutual engagement and at the same time we set the foundations for the future progress of this exponentially changing discipline.

Chapter 03. Proposed Approach

3.1. Introduction

In today's data-based world, the ability to integrate information from multiple sources is critical to informed decision-making. Data integration algorithms play a vital role in this process, integrating data from sensors, diverse databases or human experts to create a more comprehensive and accurate image. However, the success of any data consolidation approach depends on its performance.

This part of the note explores different aspects of evaluating data integration algorithms. We will delve into basic standards that measure the effectiveness of data integration technology, ensuring that it meets real-world application requirements. Why evaluate data integration performance?

Data integration algorithms are often deployed in critical applications where reliable information is of paramount importance. Here's why the rigorous evaluation process is so important: Confidence in results: The evaluation helps to ensure that conjugated data accurately reflect the real state of the system under surveillance.

Identifying weaknesses: Evaluation reveals potential weaknesses in algorithm performance, such as sensitivity to noisy data or slow processing times. This allows for meaningful improvement and improvement.

Comparison and selection: Evaluation methods facilitate comparison of algorithms integrating different data for a particular application. This allows for the objective selection of the most appropriate approach based on its strengths and weaknesses.

By comprehensively evaluating data integration algorithms, we can ensure that they provide the reliability and accuracy required to succeed in their intended applications.

3.2. Indoor localization model based on machine learning

In recent years, with the rise of artificial intelligence, cloud computing, and big data technologies, autonomous navigation and localization technologies have rapidly evolved. Specifically, the development of machine learning has introduced new solutions for RF signal-based localization [14].

3.2.1. Long short-term memory (LSTM)

Long Short-Term Memory (LSTM) networks are a variation of recurrent neural networks (RNNs). They were proposed by Jürgen Schmidhuber and his team in the early 1990s. While traditional RNNs struggle with long-term dependencies due to the vanishing gradient" problem, LSTMs are specifically designed to overcome this issue [15].

a) The architecture of LSTM

LSTM networks are an excellent variant of RNNs, inheriting the features of most RNN models while addressing the vanishing gradient problem caused by the gradual reduction in the backpropagation process. LSTMs are particularly well-suited for handling tasks that are highly correlated with time series data, such as machine translation, dialogue generation, and encoding/decoding. Therefore, we have selected LSTM as one of the two deep learning models for our Wi-Fi localization application.

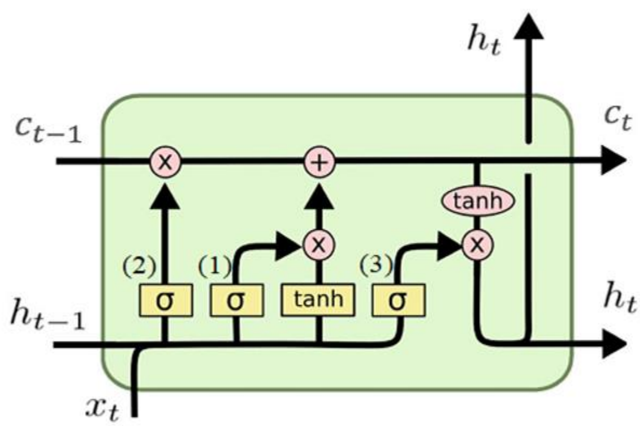

Fig 3. 1. Diagram of an LSTM Memory Cell

The basic unit of the LSTM model is depicted in Figure (Fig 3. 1), which includes the forget gate, the input gate, and the output gate. In the forget gate, the portion of the cell state C_t that is forgotten is determined by the input X_t, the previous cell state C_{t-1}, and the previous hidden state h_{t-1}. In the input gate, the remaining vector of the cell state C_t is contributed by the value produced by the sigmoid and $tanh$ functions applied to X_t . The hidden state h_t is updated based on the new cell state C_t and the output together.

b) Forget Gate

The introduction of the forget gate marked a key improvement in LSTM architecture. This gate empowers the LSTM to selectively erase information from its internal state, allowing it to adapt to changing sequences and learn long-term dependencies. This capability proved crucial for tasks like mastering complex grammars.

$$\begin{cases} f_t = \sigma_g\left(W_f x_t + U_f h_{t-1} + b_f\right) \\ i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i) \\ o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o) \\ \tilde{c}_t = \sigma_c(W_c x_t + U_c h_{t-1} + b_c) \\ c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \\ h_t = o_t \odot \sigma_h(c_t) \end{cases} \quad (3.1)$$

Where the initial values are $c_0 = 0$ and $h_0 = 0$ and the operator $\odot$ denotes the Hadmard product (element-wise product). The subscript t indexes the time step.

The basic unit of the LSTM model is depicted in Figure (Fig 3. 1), which includes the forget gate, the input gate, and the output gate. In the forget gate, the portion of the cell state C_t that is forgotten is determined by the input X_t, the previous cell state C_{t-1}, and the previous hidden state h_{t-1}. In the input gate, the remaining vector of the cell state C_t is contributed by the value produced by the sigmoid and $tanh$ functions applied to X_t. The hidden state h_t is updated based on the new cell state C_t and the output together.

Letting the superscripts d and h refer to the number of input features and number of hidden units, respectively:

- $x_t \in \mathbb{R}^d$: input vector to the LSTM unit
- $f_t \in (0,1)^d$: forget gate's activation vector
- $i_t \in (0,1)^h$: input/update gate's activation vector
- $o_t \in (0,1)^h$: output gate's activation vector
- $h_t \in (-1,1)^h$: hidden state vector also known as output vector of the LSTM unit
- $c_t \in (-1,1)^h$: cell input activation vector
- $x_t \in \mathbb{R}^h$: cell state vector
- $W \in \mathbb{R}^{hxd}, U \in \mathbb{R}^{h\mathrm{xh}}$ and $b \in \mathbb{R}^h$: weight matrices and bias vector parameters which need to be learned during training

c) Activation function

σ_g: sigmoid function

σ_c: hyperbolic tangent function

σ_h: hyperbolic tangent function or, as the peephole LSTM paper [16][17]suggest, $\sigma_h = x$.

d) Peephole Connections:

- **Fine-Tuning the Flow** Gers and Schmidhuber identified the need for the LSTM to have finer control over its internal flow, particularly when dealing with tasks requiring precise timing. Previously, this control was limited by the open output gate. To address this, peephole connections were introduced. These connections, depicted by blue lines in (Fig 3. 2)**,** provide a direct path for the cell state to influence the gates. This enhancement simplifies the learning process for tasks involving precise timing. As an additional refinement, the output activation function was deemed unnecessary for the tasks explored at the time and was therefore omitted.

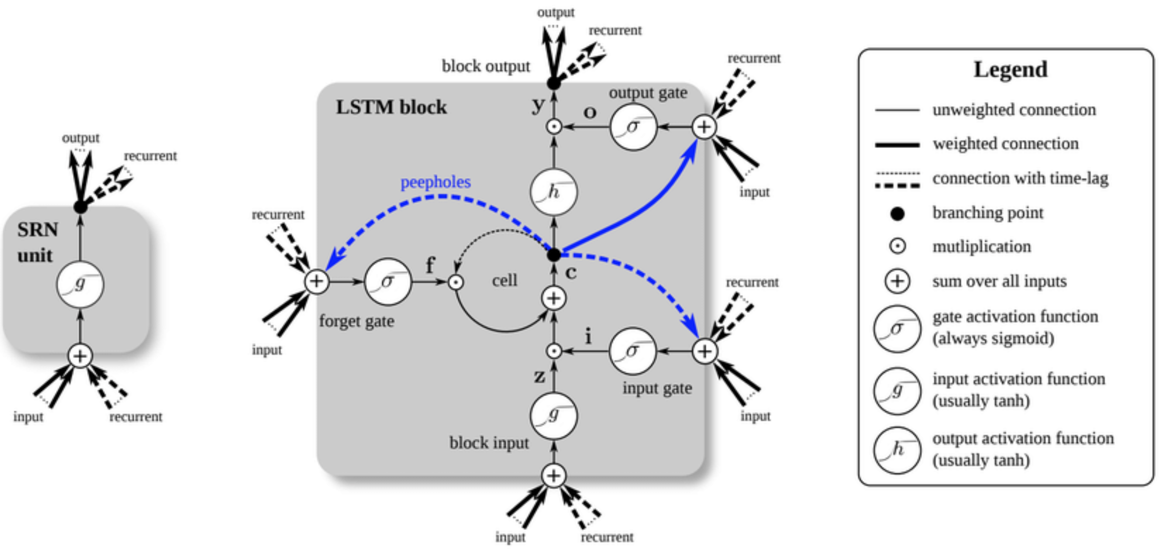

Fig 3. 2. Detailed schematic of the Simple Recurrent Network (SRN) unit (left) and a Long Short-Term Memory block (right) as used in the hidden layers of a recurrent neural network

- **Full Gradient Training** Building upon the LSTM architecture described earlier, Graves and Schmidhuber introduced a crucial advancement: full Backpropagation Through Time (BPTT) training for LSTM networks. This method allowed for verification of LSTM gradients using finite differences, leading to more reliable implementations. They demonstrated the effectiveness of this approach on the TIMIT benchmark dataset.
- **Variations on a Theme** While the vanilla LSTM remains the most widely used architecture, numerous variations have emerged.

e) Training:

An RNN using LSTM units can be trained in a supervised fashion on a set of training sequences, using an optimization algorithm like gradient descent combined with backpropagation through time to compute the gradients needed during the optimization process, in order to change each weight of the LSTM network in proportion to the derivative of the error (at the output layer of the LSTM network) with respect to corresponding weight.

A problem with using gradient descent for standard RNNs is that error gradients vanish exponentially quickly with the size of the time lag between important events. This is due to $\lim_{n\to\infty} W^n = 0$ if the spectral radius of W is smaller than 1 [18].

However, with LSTM units, when error values are back-propagated from the output layer, the error remains in the LSTM unit's cell. This "error carousel" continuously feeds error back to each of the LSTM unit's gates, until they learn to cut off the value.

f) CTC score function:

An RNN Many applications use stacks of LSTM RNNs and train them by connectionist temporal classification (CTC) to find an RNN weight matrix that maximizes the probability of the label sequences in a training set, given the corresponding input sequences. CTC achieves both alignment and recognition [19][20].

3.2.2. Bidirectional Long short-term memory (BiLSTM)

Bi-LSTM (Bidirectional Long Short-Term Memory) is a type of recurrent neural network (RNN) that processes sequential data in both forward and backward directions. It combines the power of LSTM with bidirectional processing, allowing the model to capture both past and future context of the input sequence.

The results showed that the BiLSTMs are effected particularly useful in situations it is relevant when the context of the input is required. It is very useful for international students to understand the complexities of the informational environment, organize their interactions with it and avoid publication of materials containing controversial information. works like sentiment classification. In unidirectional LSTM It indicates that information flows from the backward to the forward and hence perfected accordingly. On the opposite for in Bi-directional LSTM information moves both ways with the help of two arrows, which permits to go from backward to forward and from forward to backward. hidden states. Thus when it comes to understanding the context, Bi-LSTMs do so far better. These coupled with BiLSTMs were used to amplify the chunk of input information. usable to the network. RNN with LSTM and Structure This kind of neural networks is called Recurrent Neural Network (RNN), and one of the improved versions of RNN is the LSTM. RNN with BiLSTM. In other words, BRNN will do the following: Basically, it follows out such a technique applied to partition the neurons of a normal RNN. bidirectional ways. As mentioned earlier, there are two types of theories and one is for the backward states or negative. time direction, and another for future condition or positive. time direction. The reversed direction states inputs are independent of the results of these two states. The structure or It is called BiLSTM and is illustrated in the following (Fig 3. 3). By utilizing two time [21].

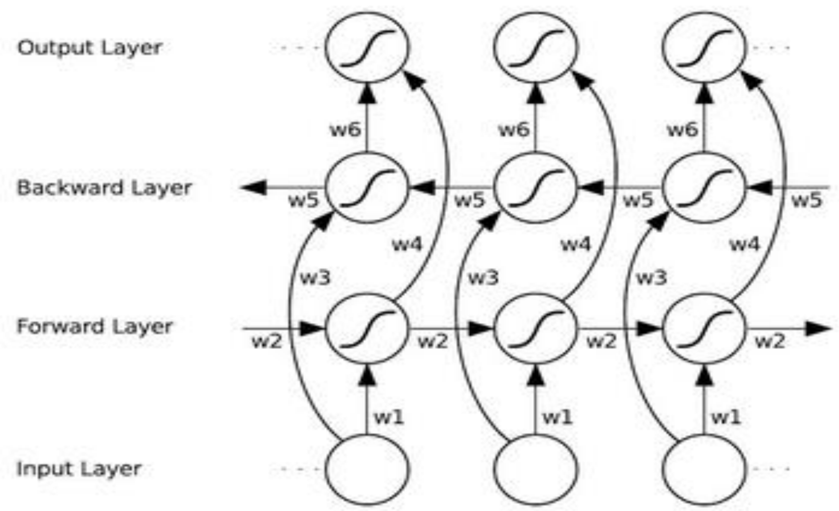

Fig 3. 3. Bidirectional LSTM

The instructions, use data prior to the current and data in the subsequent part of the current. time frame can be used. In its place it is better to use a certain another unique time frame. whereas standard RNN is the model which requires the delays to enable for future data to be incorporated**.**

3.2.3. Random Forest (RF)

A Random Forest leverages the power of multiple decision trees, reducing overfitting through random training sample selection. This makes it a versatile tool for tasks like computer vision, pattern recognition, and various machine learning applications. Look at (Fig 3. 4). This illusory decision tree tries to predict whether someone will use a random forest based on their statistical anxiety and years of programming. Rectangles are called “nodes”. Dotted circuits are model predictions with sample size of those who meet each requirement listed below appropriate predictions [22][23].

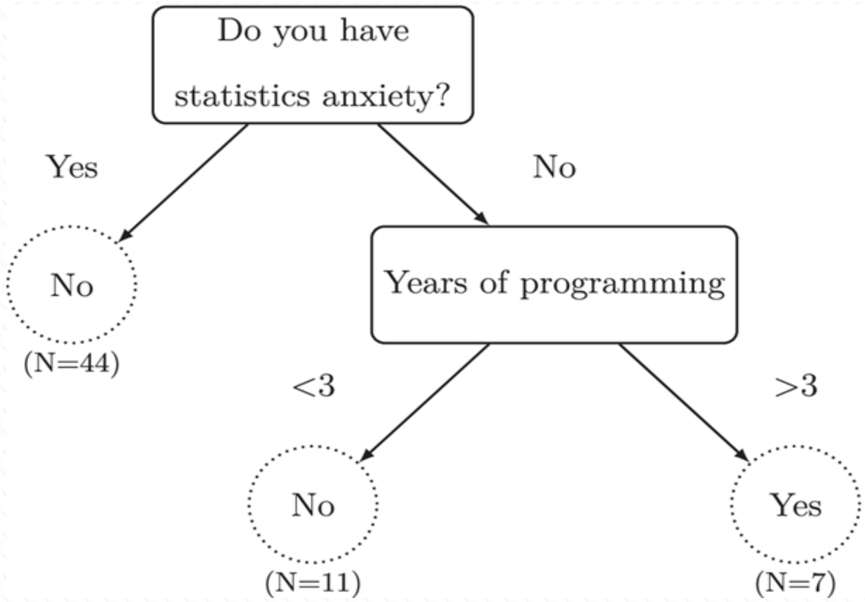

Fig 3. 4. Example of a Decision-Tree

Imagine a forest of decision trees, each with a slightly different perspective on the data. These trees work together to make predictions, like a team of experts voting on the best answer [23].

3.2.4. Decision Trees

Decision trees are tree-like classification tools. It frequently divides data based on characteristics to create homogeneous subsets until the stop standard is met. J48 is a popular algorithm for this task and uses effective guidelines for building decision trees. The decision tree is shown in (Fig 3. 5), which includes category, result, and salary as the root and internal nodes, and the classes for the leaf nodes are pass, fail, HOD, teacher, and clerk [24].

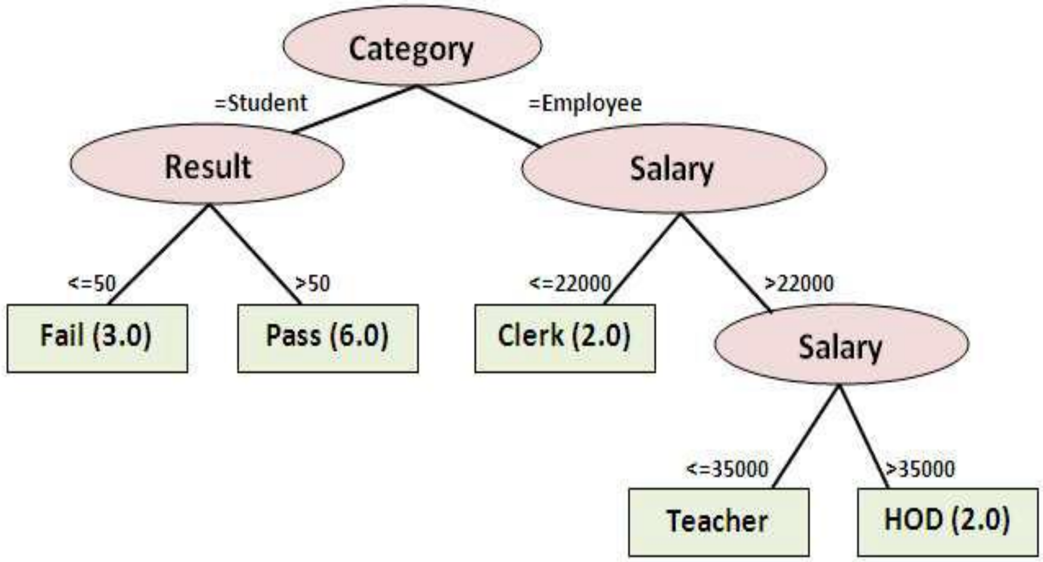

Fig 3. 5. Decision tree visualization

3.3. Conclusion

In this chapter, we presented two indoor localization models based on machine learning (LSTM and BiLSTM) and a Wi-Fi network that we developed. This setup allowed us to collect data and acquire various types of information, such as Channel State Information (CSI) and Received Signal Strength (RSS).

Chapter 04. Implementation and Validation

4.1. Introduction

In the field of internal localization research, the development of robust and effective algorithms is necessary to ensure the effectiveness of its application and the validity of its scientific results. Chapter 3 addresses methodological challenges related to the validation and validation of the algorithm proposed by us, using a combination of experimental assessment and simulated analysis, seeks to evaluate its performance across a variety of internal environments and highlight its accuracy, strength and reliability.

4.2. Data Collection

a) Methods for Collecting Data in Different Indoor Environments

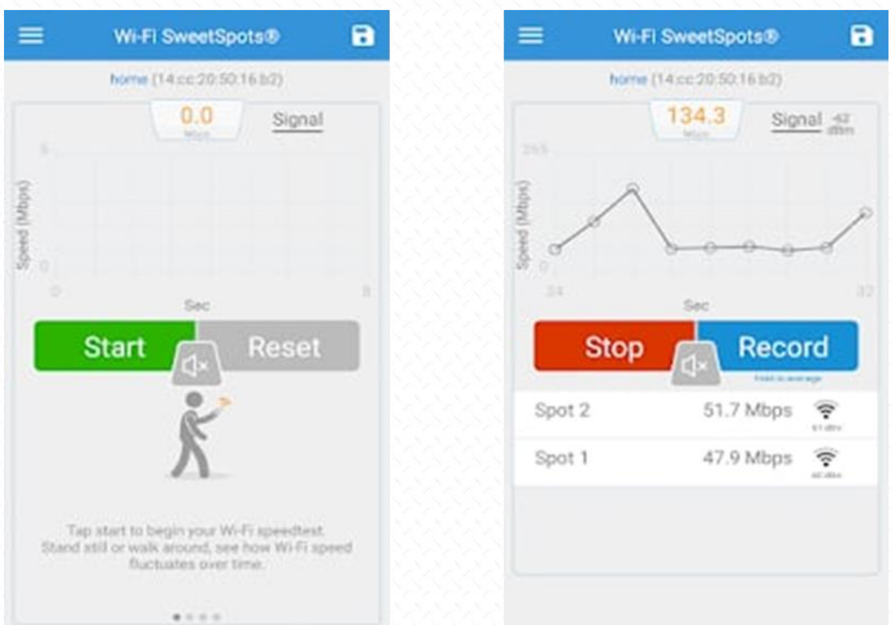

Fig 4. 1. Wi-Fi Analyzer" software"

We used the "Wi-Fi Analyzer" software, which is accessible on the Android software Store, to gather real-world RSSI (Received Signal Strength Indicator) data. Within a 12 square meter region of our house, data was gathered. There are a total of 12 measurement sites because this region was methodically divided into grids of one square meter.

b) Procedure

- Grid Layout: A grid comprising one square meter for each cell was created by mapping out the 12 square meter space.
- Measuring Points: The Wi-Fi Analyzer app was used to capture the Wi-Fi signal strength at each of the 12 grid points.
- Data Recording: For additional analysis, the gathered RSSI values at each location were jotted down and combined into an Excel spreadsheet (p12.xlsx).

Table 4. 1. A snippet of the collected data in Excel

Timestamp (UTC)	Point 1 (dBm)	Point 2 (dBm)	...	Point 12 (dBm)
2024-01-01 12:00:00.000	-45	-47	...	-50
2024-01-01 12:01:00.000	-44	-46	...	-49
...	...	...	...	...

For the data collection process, we selected a real-world case study - the interior of a house, specifically a single room, as shown in Figure (Fig 4. 2). The decision to focus on a single room was intentional, as it allowed us to conduct a thorough exploration of our approach for approximating points and evaluating the accuracy of the models. The room we chose provided a realistic and representative environment for our study. By concentrating our efforts on a single, enclosed space, we were able to carefully control the variables and ensure that the data collected would be meaningful and applicable to our research goals.

The dimensions of the selected room are provided in the following figure. As you can see, the room has a rectangular layout with clearly defined boundaries and features, such as walls, a door, and potentially other furniture or objects. This level of detail in the physical environment was crucial for our analysis, as it allowed us to map the space accurately and understand the relationships between different points within the room. Additionally, the figure illustrates the way we divided the room's area into smaller sections or zones. This strategic partitioning of the space enabled us to explore the accuracy of our models at a granular level, testing their ability to precisely locate and identify specific points or areas within the larger room. By analysing the performance of our approach across these defined zones, we could gain valuable insights into the strengths and limitations of our methods, ultimately helping to refine and improve our overall research. Overall, the selection of a real-world room as the focus of our data collection was a deliberate and thoughtful decision, designed to provide a comprehensive and realistic testbed for evaluating our approximation techniques and the accuracy of the resulting models.

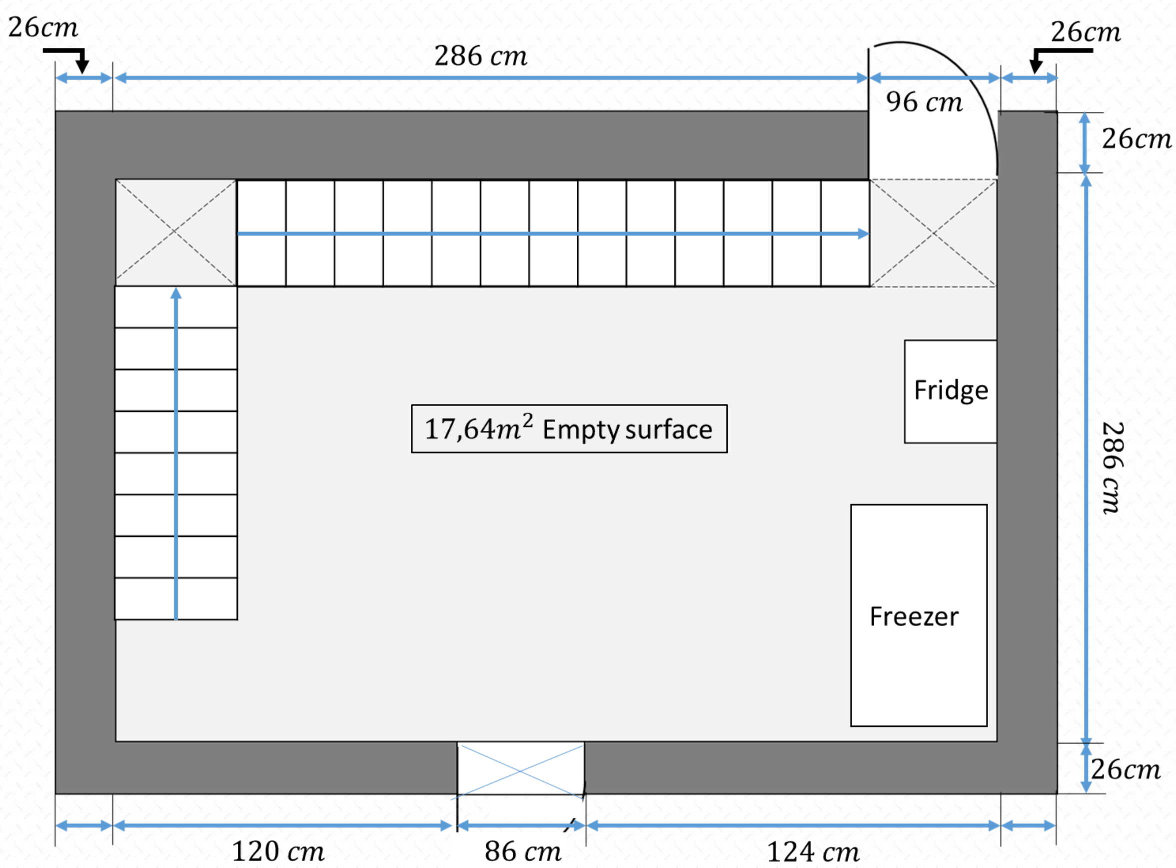

Fig 4. 2. Interior design of the room where the measurement was made

As depicted in Figure (Fig 4. 3), the room's area has been strategically divided into 12 equal square sections, with each individual square measuring 1 square meter in size ($1\ m^2$). This intentional partitioning of the space serves a crucial purpose in our research approach. By breaking down the room into these precise, 1-square-meter zones, we can establish a highly granular framework for evaluating the accuracy of our approximation techniques. Targeting an accuracy level of 1 meter is a particularly challenging objective, as it requires our models to pinpoint locations within the room with a high degree of precision. This level of detail in our spatial analysis will allow us to rigorously test the capabilities of our approximation approach and compare its performance against other methods. By assessing the model's ability to correctly identify the specific square or zone within which a given point is located, we can gain invaluable insights into the strengths and limitations of our techniques.

Furthermore, the consistent 1-square-meter size of each zone provides a standardized unit of measurement that enables us to make meaningful comparisons and draw clear conclusions about the relative accuracy of our approach. This standardization helps to minimize potential confounding factors and ensures that our evaluation of the models is as objective and reliable as possible. Overall, the strategic division of the room's area into these 12 equal squares is a deliberate design choice that supports the core objectives of our research. By focusing on a challenging 1-meter accuracy target and establishing a well-defined spatial framework, we aim to rigorously test and benchmark our approximation approach against other methods, ultimately identifying the most effective solution for our application.

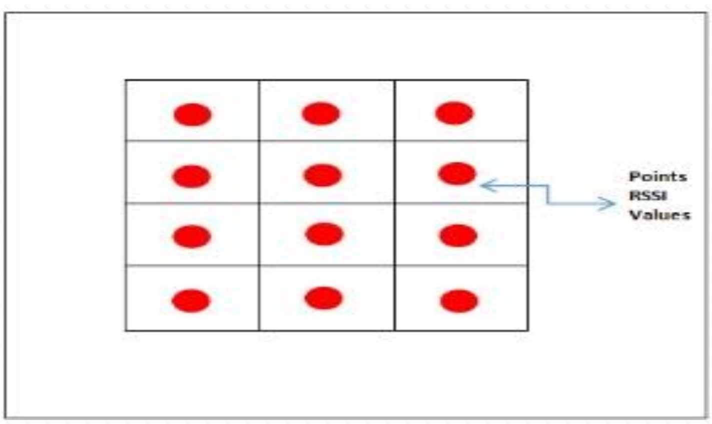

Fig 4. 3. Split measurement space

Finally, for each square we will measure the RSSI (Red dots in the figure) during several days and several times.

Real data collection**:** we used smartphone equipped with WI-FI SEEWTSPOTS. The data are recorded in MICROSOFT EXCEL table**.**

The collection data was conducted during more than 20 days. The adopted timestamp step is 30 minutes. During this period, we collected more than 500 measures related to the 12th places defined before. As shown in the following figure "Fig 4. 2" and Table 4. 2. the collected data shows a non-stability of the RSSI signal. For example, for the place 1 ($P1$) and during a period for one day 12/04/2024 we can see the value of RSSI changing between $[-47dBm, -60dBm]$. The same conclusion can be done for the other places with different RSSI values and ranges of variation. For more readability the table is presented as a $heatmap$. According to this hight variation in the measures we are obliged to use a forecasting method to predict the value of RSSI which will allow as to geolocation the smartphone.

Table 4. 2. Example of the table of the collected RSSI values

Timestamp	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Days Hours	(dBm)											
12/04/2024 15:00	-60	-61	-59	-61	-57	-57	-59	-56	-56	-55	-55	-54
12/04/2024 15:30	-57	-56	-57	-54	-50	-51	-56	-56	-54	-56	-55	-55
12/04/2024 16:00	-56	-53	-54	-55	-54	-54	-55	-55	-57	-57	-56	-53
12/04/2024 16:30	-56	-55	-55	-51	-51	-51	-52	-55	-55	-55	-55	-55
12/04/2024 17:00	-57	-55	-55	-52	-53	-53	-54	-55	-55	-57	-55	-57
12/04/2024 17:30	-54	-54	-53	-57	-57	-55	-55	-51	-53	-57	-56	-55
12/04/2024 18:00	-60	-50	-52	-52	-51	-52	-53	-55	-56	-56	-55	-55
12/04/2024 18:30	-49	-50	-52	-53	-49	-50	-53	-52	-54	-55	-52	-52
12/04/2024 19:00	-50	-52	-53	-50	-51	-51	-52	-52	-53	-55	-55	-57
12/04/2024 19:30	-50	-50	-50	-51	-51	-52	-53	-54	-54	-54	-54	-54
12/04/2024 20:00	-55	-52	-52	-52	-52	-52	-52	-52	-52	-54	-54	-54
12/04/2024 20:30	-49	-49	-50	-51	-52	-53	-54	-51	-53	-54	-54	-53
12/04/2024 21:00	-51	-51	-56	-56	-54	-52	-53	-52	-52	-51	-51	-52
12/04/2024 21:30	-51	-51	-52	-53	-53	-52	-52	-53	-52	-52	-53	-53
12/04/2024 22:00	-54	-54	-54	-53	-53	-53	-53	-53	-54	-54	-54	-55
12/04/2024 22:30	-49	-49	-52	-54	-54	-54	-54	-53	-53	-53	-54	-52
12/04/2024 23:00	-50	-46	-47	-49	-51	-51	-53	-53	-55	-53	-53	-54
12/04/2024 23:30	-47	-48	-48	-49	-50	-51	-51	-52	-51	-51	-52	-54
13/04/2024 0:00	-46	-47	-45	-47	-49	-49	-52	-52	-53	-51	-53	-54

4.3. Model building

Our approach is based on proposing an architecture of a BiLSTM model and another one of LSTM with the same parameters, then we proceed with the comparison of both models. These models are commonly used for time series forecasting and other sequence-to-sequence tasks.

The model is a Long Short-Term Memory (LSTM) neural network model. The input shape of the model is defined as ($sequence\_length, df.shape[1]$), where $sequence\_length$ is the length of the input sequence, and $df.shape[1]$ is the number of features (columns) in the input data. The model starts with an LSTM layer with 60 units and a '$relu$' activation function, which is designed to capture the long-term dependencies in the input sequence. The output layer of the model is a single Dense layer with $df.shape[1]$ units, which corresponds to the number of features in the input data. This output layer is responsible for predicting the next value in the sequence, given the input sequence. The model is compiled with the '$adam$' optimizer and 'mse' ($mean\ squared\ error$) loss function, which are common choices for regression problems. Finally, the model is trained using the $fit()$ method, with X_train and y_train as the input data and target, respectively. The training is run for 500 epochs with a batch size of 1, which means that the model will see the entire training set 500 times during the training process.

The following figure shows the architecture of proposed model, where the input is represented by $x_1, x_2, .., x_T$ with $T = 60$ related to the number of units.

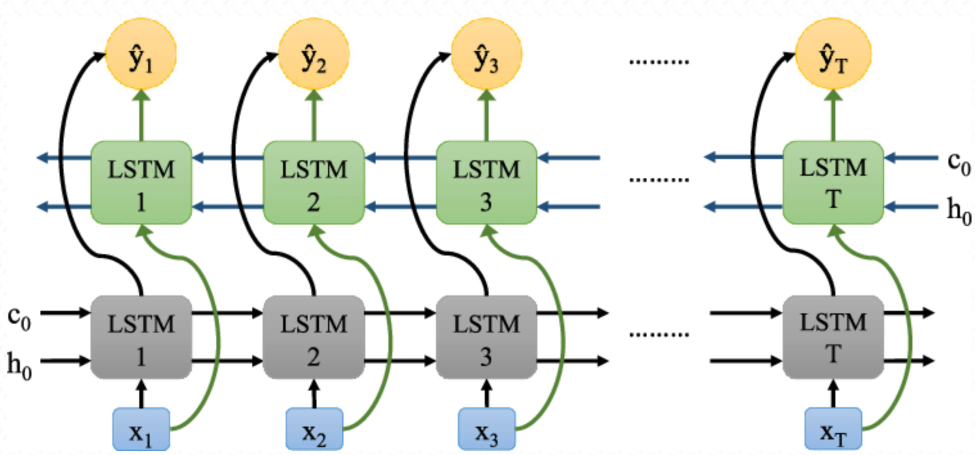

Fig 4. 4. Example of the table of the collected RSSI values

The proposed model is a Bidirectional Long Short-Term Memory (BiLSTM) neural network model. For our comparison we will keep the same parameters. The input shape of the model is defined as ($sequence\_length$, $df.shape$ [1]), where $sequence\_length$ is the length of the input sequence, and $df.shape[1]$ is the number of features (columns) in the input data.

The key difference from the previous model is the use of the Bidirectional layer, which is a wrapper around the LSTM layer. The Bidirectional layer allows the LSTM layer to process the input sequence in both the forward and backward directions, capturing information from both contexts. This can often lead to improved performance compared to a unidirectional LSTM.

The Bidirectional LSTM layer has 60 units and uses the '$relu$' activation function. This layer is responsible for encoding the input sequence into a hidden representation that can capture both the short-term and long-term dependencies in the data. The output layer of the model is a single Dense layer with $df.shape[1]$ units, which corresponds to the number of features in the input data. This output layer is responsible for predicting the next value in the sequence, given the input sequence.

The model is compiled with the '$adam$' optimizer and 'mse' ($mean\ squared\ error$) loss function, which are common choices for regression problems. The training process is the same as the previous model, with the model being trained for 500 epochs with a batch size of 1

In general, the key difference between this model and the previous one is the use of a Bidirectional LSTM layer, which allows the model to capture information from both the forward and backward directions of the input sequence, potentially leading to improved performance.

4.4. Model Evaluation

After model building, we proceed with the process of training and evaluating the two different types of recurrent neural network models: A Long Short-Term Memory (LSTM) model and a Bidirectional Long Short-Term Memory (BiLSTM) model.

The first step is to build and train the LSTM model using the $build\_and\_train\_lstm()$ function, which likely defines the LSTM model architecture and trains it on the provided training data, X_train and y_train. This LSTM model is then stored in the $lstm\_model$ variable. Similarly, the $build\_and\_train\_bilstm()$ function is used to build and train the BiLSTM model, which is then stored in the bilstm_model variable. The BiLSTM model is a variant of the LSTM that processes the input sequence in both the forward and backward directions, potentially allowing it to capture more contextual information.

After training the models, the we proceed to the predictions on the test data, X_test. The $lstm\_model.predict(X\_test)$ and $bilstm\_model.predict(X\_test)$ calls generate the predicted values for the LSTM and BiLSTM models, respectively. These predicted values are stored in the $lstm\_predicted\_scaled$ and $bilstm\_predicted\_scaled$ variables. It's important to note that the predicted values are initially in the scaled format, as they were produced using the same data scaler that was applied to the training data. To obtain the actual, unscaled predicted values, the code applies the inverse transformation using the $scaler.inverse\_transform()$ function, storing the results in the $lstm\_predicted$ and $bilstm\_predicted$ variables. Finally, the original, unscaled test targets, y_test, are also rescaled using the same scaler and stored in the $y\_test\_rescaled$ variable. This is likely done to ensure consistency when comparing the predicted values to the true target values.

The purpose of this actions is to train and evaluate the performance of both the LSTM and BiLSTM models on the built dataset. The predicted values from these models are then analyzed and compared to assess their relative strengths and weaknesses in accurately forecasting the target variable. The results compare the performance of a Long Short-Term Memory (LSTM) model and a Bidirectional Long Short-Term Memory (BiLSTM) model on a specific dataset or task. The metrics presented give us a comprehensive understanding of the models' performance and allow us to draw some meaningful conclusions.

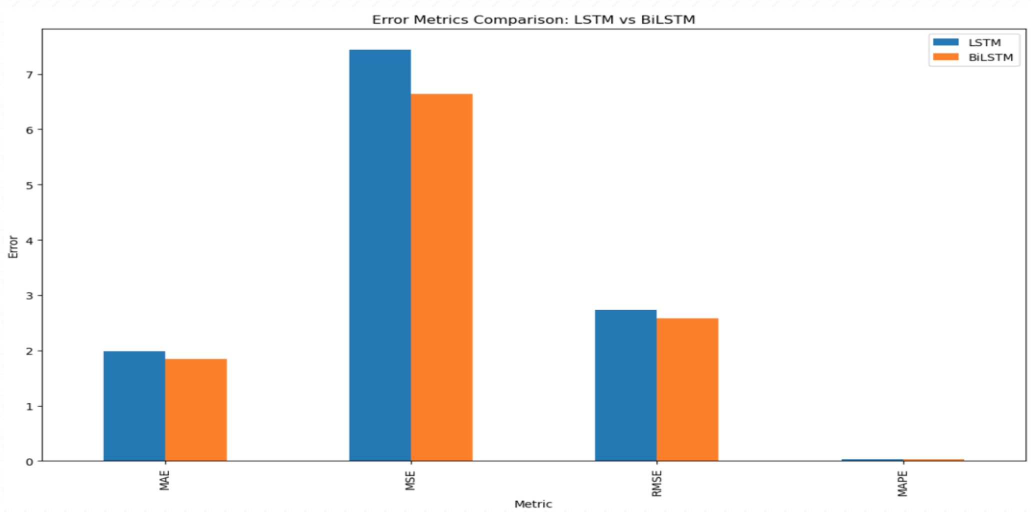

Fig 4. 5. Error measure for LSTM and BiLSTM

The LSTM model achieved a Mean Absolute Error (MAE) of 1.9821, which indicates that on average, the model's predictions deviate from the true values by approximately 1.98 units. The Mean Squared Error (MSE) of 7.4418 suggests that the model's overall squared error is around 7.44, and the Root Mean Squared Error (RMSE) of 2.7279 provides a more interpretable measure of the average magnitude of the model's errors, which is approximately 2.73 units. Additionally, the Mean Absolute Percentage Error (MAPE) of 0.03700006 shows that the model's predictions, on average, deviate from the true values by around 3.7% in relative terms (Fig 4. 5).

Compared to the LSTM model, the BiLSTM model demonstrates improved performance across all the reported metrics. The BiLSTM model achieved a lower MAE of 1.8477, indicating that its predictions are, on average, closer to the true values by around 0.13 units compared to the LSTM model. The BiLSTM model also has a lower MSE of 6.6365 and a lower RMSE of 2.5761553967468673, suggesting that the overall magnitude of its errors is smaller than the LSTM model. Finally, the BiLSTM model has a lower MAPE of 0.03453, meaning that its predictions, on average, deviate from the true values by around 3.45% in relative terms, which is a slight improvement over the LSTM model.

4.5. Result discussion

The obtained results suggest that the BiLSTM model, which utilizes bidirectional information processing, outperforms the traditional LSTM model in this dataset. The improved performance of the BiLSTM model can be attributed to its ability to capture contextual information from both the forward and backward directions of the input sequence, leading to more accurate predictions compared to the unidirectional LSTM model.

In this analysis, we investigated the performance of Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) models for the task of indoor geolocation. The study was conducted in the context of a room divided into 12 equal squares, with the goal of predicting the location of an object within the room. The room was divided into 12 equal squares, and data was collected for each square. LSTM and BiLSTM models were trained on this data, and their performance was compared.

The results of the analysis showed that the BiLSTM model outperformed the LSTM model in terms of prediction accuracy. The 12 graphics of the LSTM and BiLSTM models, as well as the forecasting results, were plotted, and the BiLSTM model demonstrated a better ability to capture the temporal dependencies in the data. The superior performance of the BiLSTM model can be attributed to its ability to consider both the forward and backward sequences of the input data. This allows the model to better capture the contextual information and learn the underlying patterns in the data.

The accurate prediction results obtained from the BiLSTM model suggest that it is a more suitable choice for indoor geolocation tasks, where understanding the spatial and temporal relationships within the data is crucial for making accurate predictions.

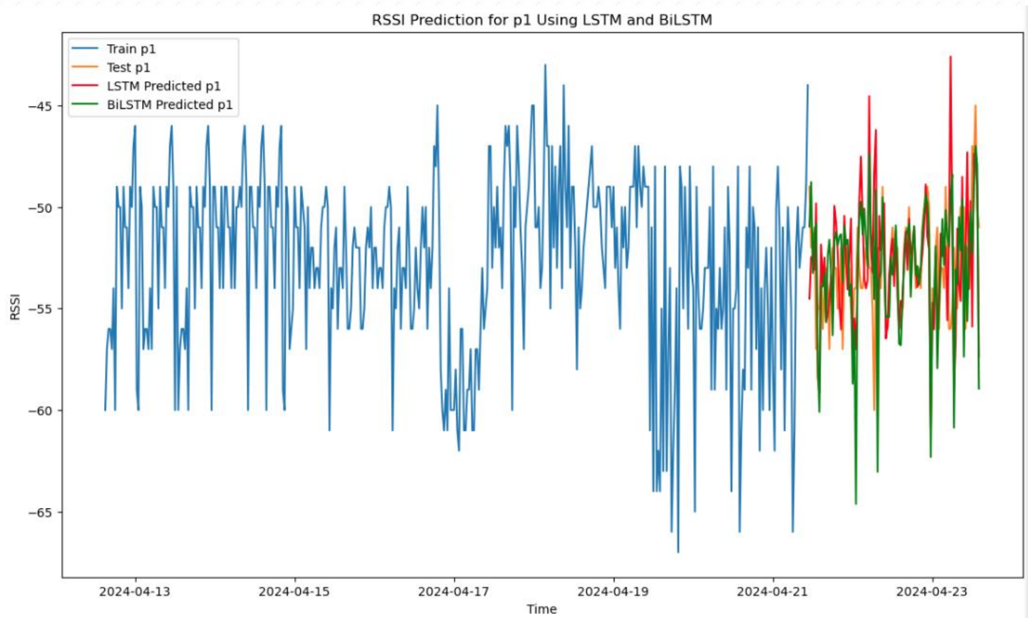

Fig 4. 6. RSSI prediction for P1 using LSTM and BiLSTM

The figure (Fig 4. 6) shows the RSSI measures of P1. The forecasting graphic shows the actual values in blue, the LSTM predictions in red, and the BiLSTM predictions green over the 10-day period. The LSTM Error and BiLSTM Error metrics are also provided, indicating that the BiLSTM model has a lower average error compared to the LSTM model. The same analysis will be done for P3 in figure 3.9.

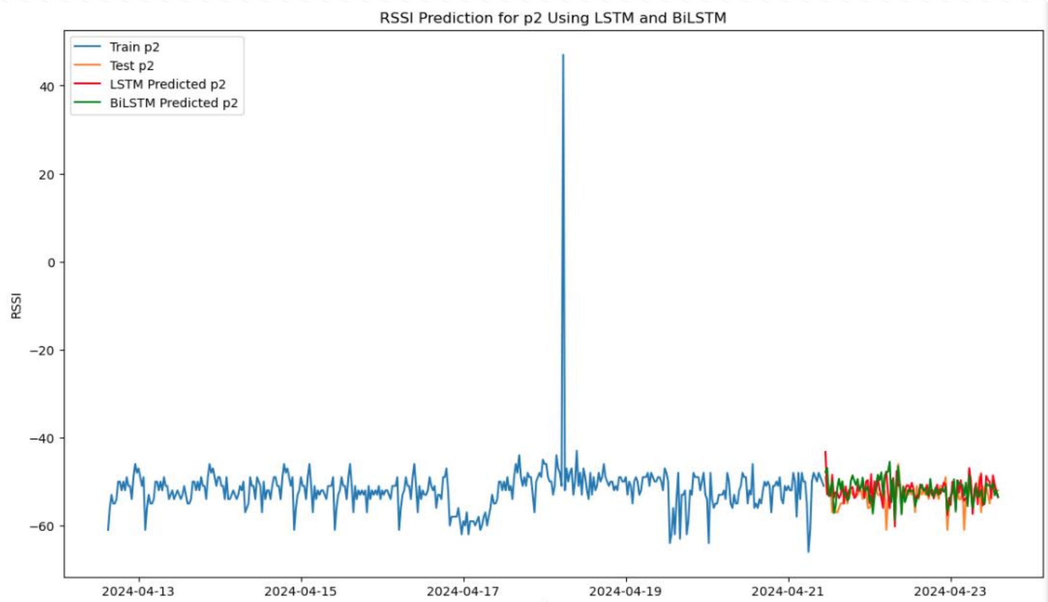

Fig 4. 7. RSSI prediction for P2 using LSTM and BiLSTM

The figure (Fig 4. 7) shows the RSSI measures of P2. The forecasting graphic shows the actual values in blue, the LSTM predictions in red, and the BiLSTM predictions green over the 10-day period. The LSTM Error and BiLSTM Error metrics are also provided, indicating that the BiLSTM model has a lower average error compared to the LSTM model. However, P2 shows a unique pic which can be considered as outlier and should be analyzed during long periods.

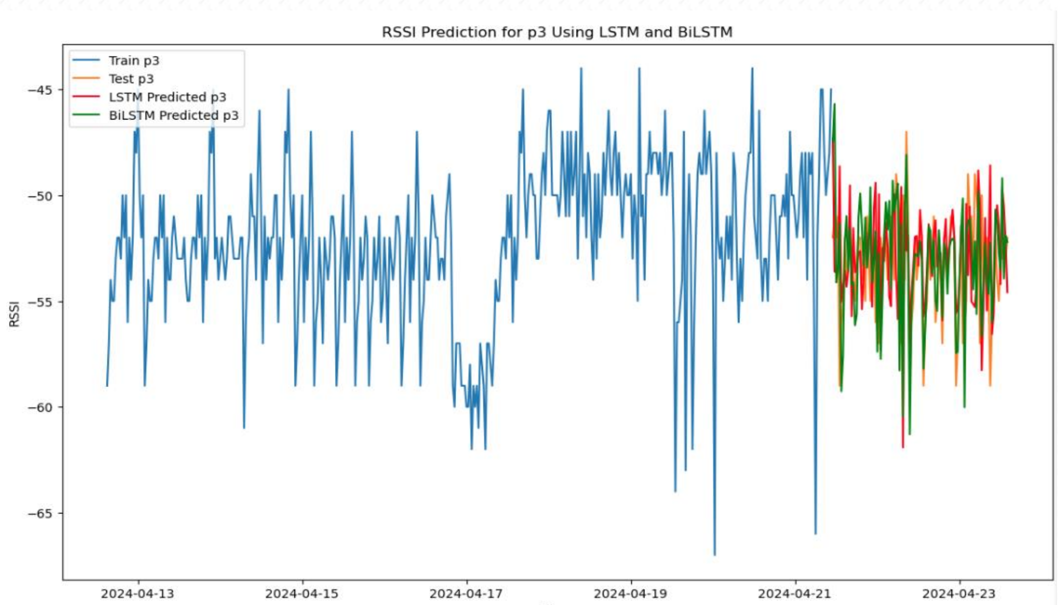

Fig 4. 8. RSSI prediction for P3 using LSTM and BiLSTM

The figure (Fig 4. 8) shows the RSSI measures of P3. The forecasting graphic shows the actual values in blue, the LSTM predictions in red, and the BiLSTM predictions green over the 10-day period. The LSTM Error and BiLSTM Error metrics are also provided, indicating that the BiLSTM model has a lower average error compared to the LSTM model.

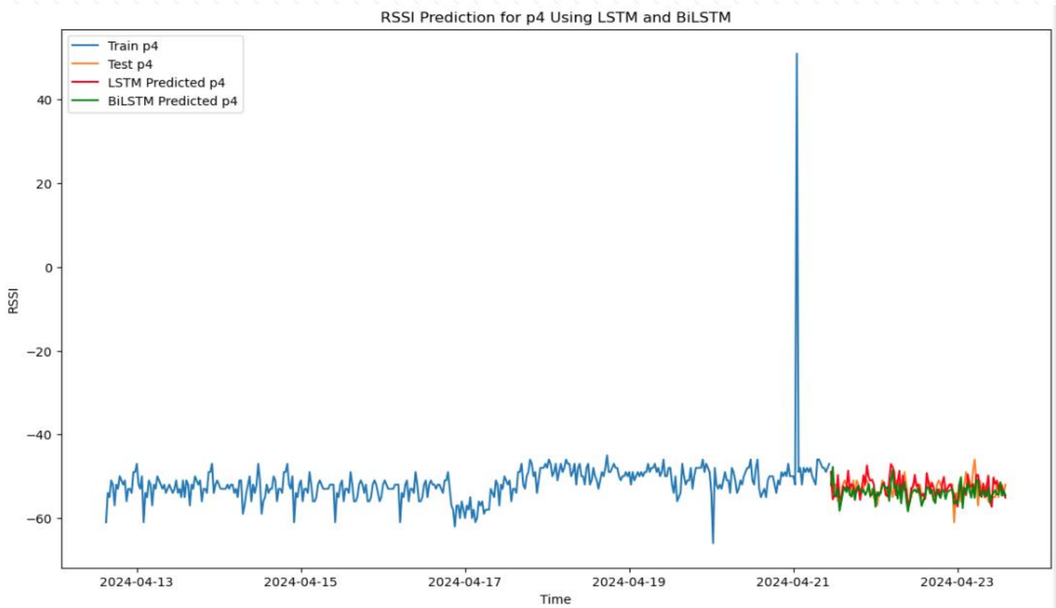

Fig 4. 9. RSSI prediction for P4 using LSTM and BiLSTM

The figures (Fig 4. 9, Fig 4. 10, Fig 4. 12, and Fig 4. 13) show the RSSI measures of respective P4, P5, P7 and P8. The forecasting graphic shows the actual values in blue, the LSTM predictions in red, and the BiLSTM predictions green over the 10-day period. The LSTM Error and BiLSTM Error metrics are also provided, indicating that the BiLSTM model has a lower average error compared to the LSTM model. However, these figures show rare pics which can be considered as outliers. The redundancy on these high pics can be considered as outliers related to the quality of Wi-Fi device and should be analyzed for a longer period.

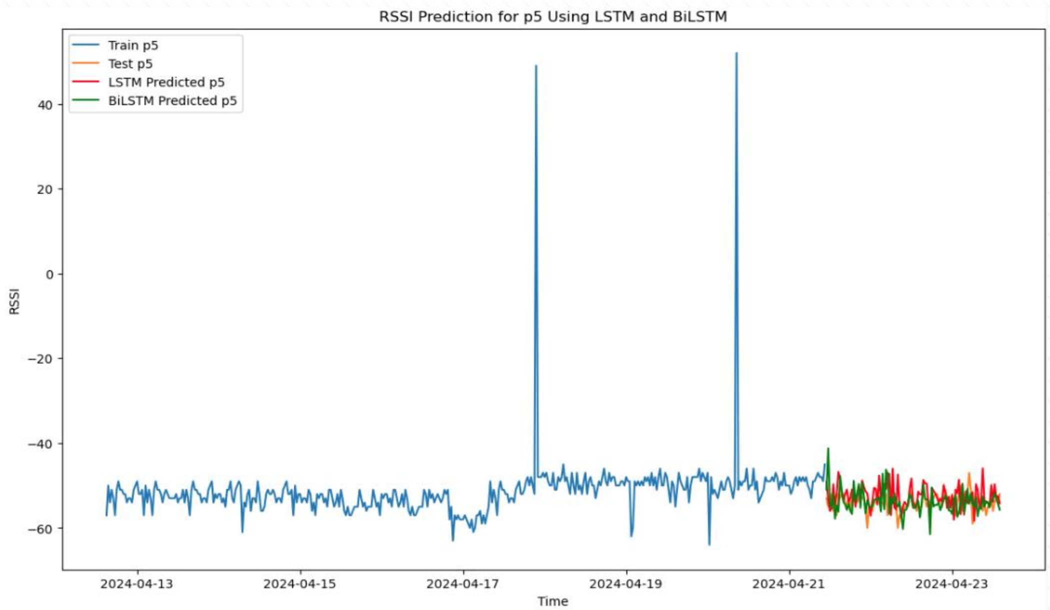

Fig 4. 10. RSSI prediction for P5 using LSTM and BiLSTM

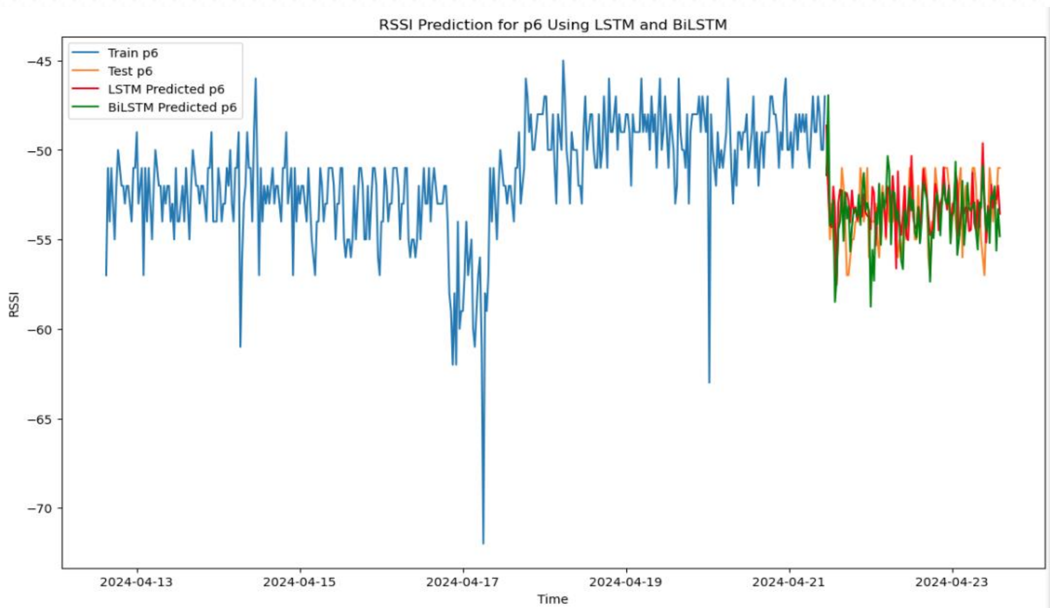

Fig 4. 11. RSSI prediction for P6 using LSTM and BiLSTM

The figures (Fig 4. 11, Fig 4. 14, Fig 4. 15, Fig 4. 16, and Fig 4. 17) show the RSSI measures of respective P6, P10, P11 and P12. The forecasting graphic shows the actual values in blue, the LSTM predictions in red, and the BiLSTM predictions green over the 10-day period. The LSTM Error and BiLSTM Error metrics are also provided, indicating that the BiLSTM model has a lower average error compared to the LSTM model.

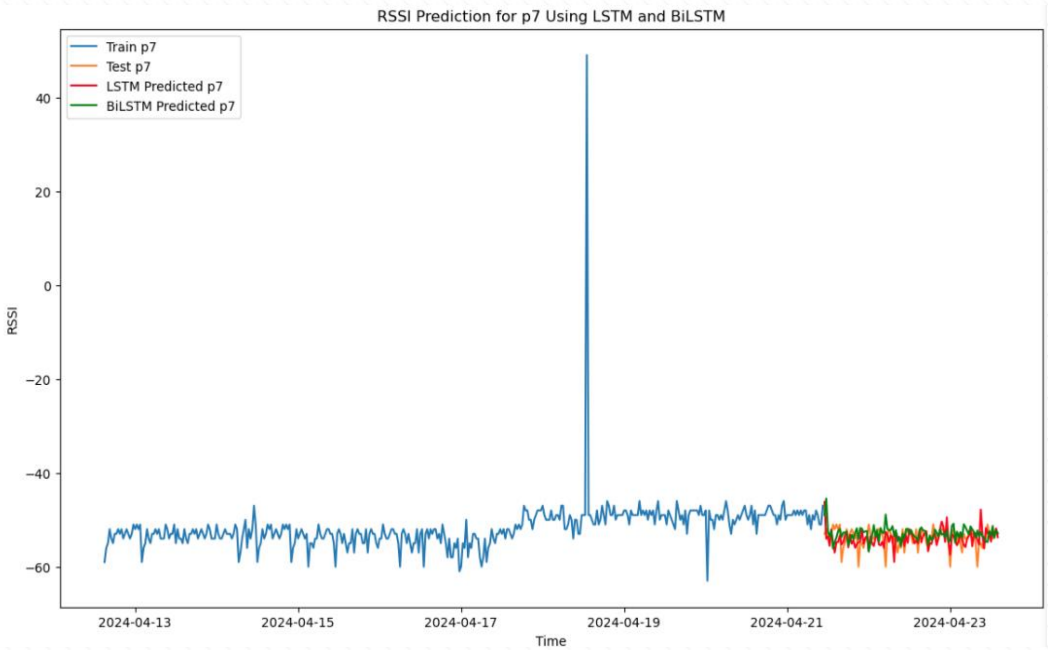

Fig 4. 12. RSSI prediction for P7 using LSTM and BiLSTM

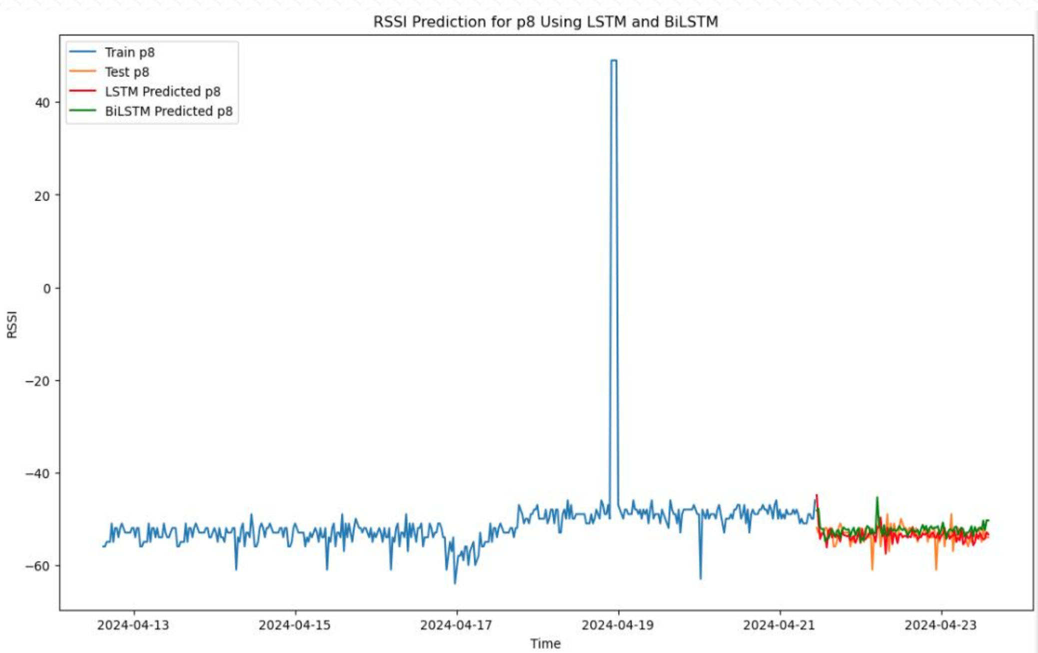

Fig 4. 13. RSSI prediction for P8 using LSTM and BiLSTM

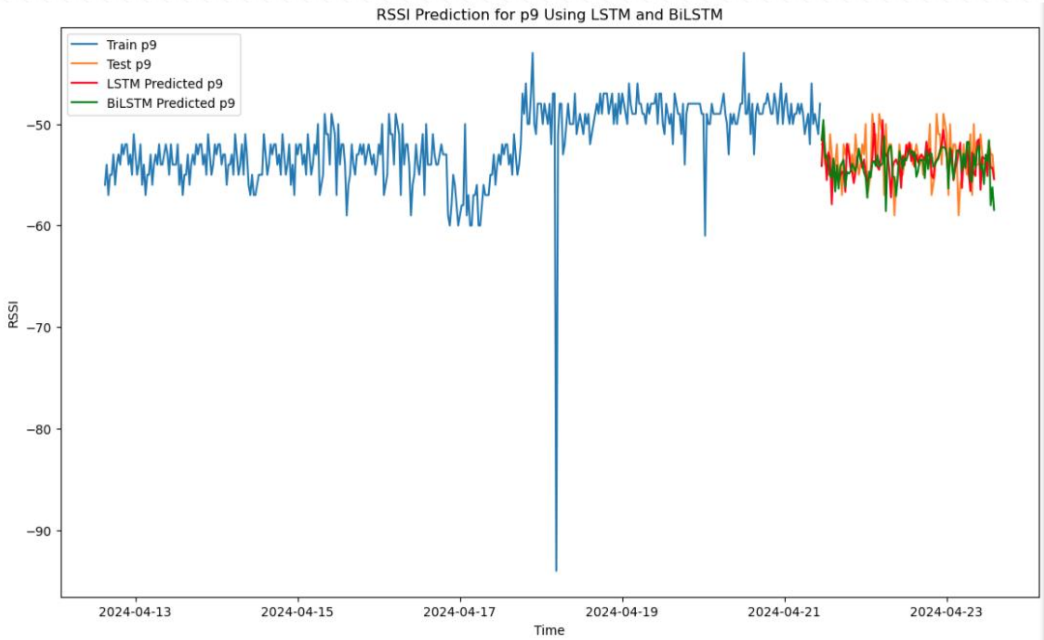

Fig 4. 14. RSSI prediction for P9 using LSTM and BiLSTM

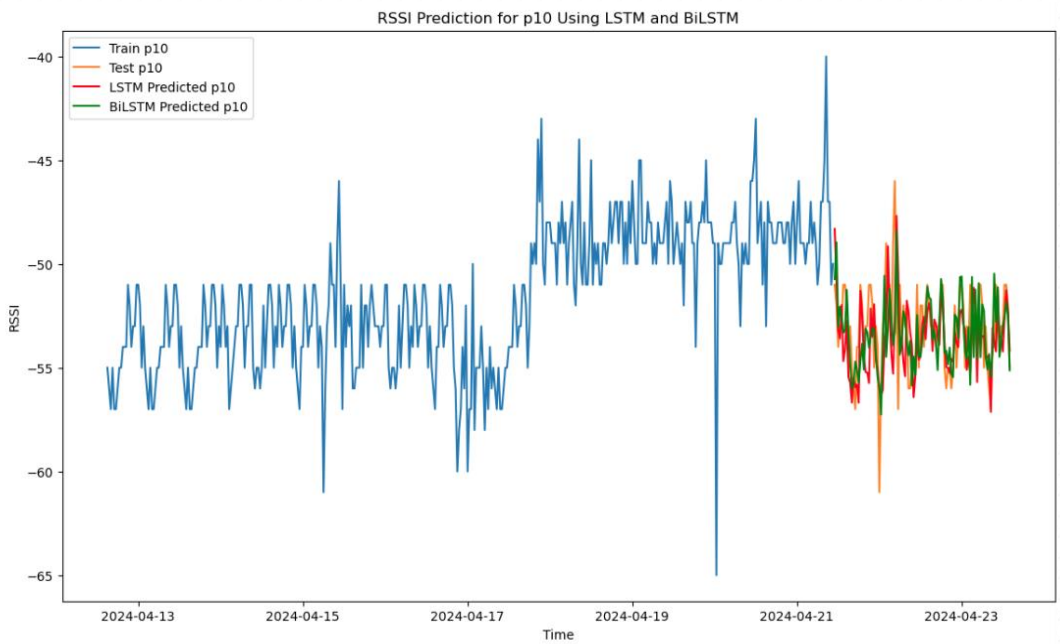

Fig 4. 15. RSSI prediction for P10 using LSTM and BiLSTM

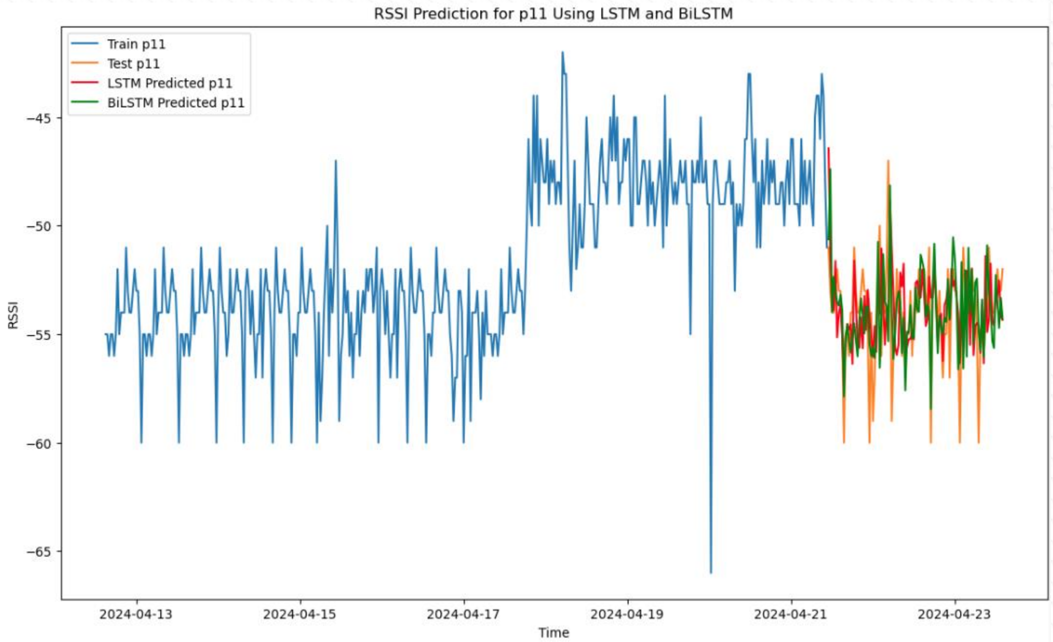

Fig 4. 16. RSSI prediction for P11 using LSTM and BiLSTM

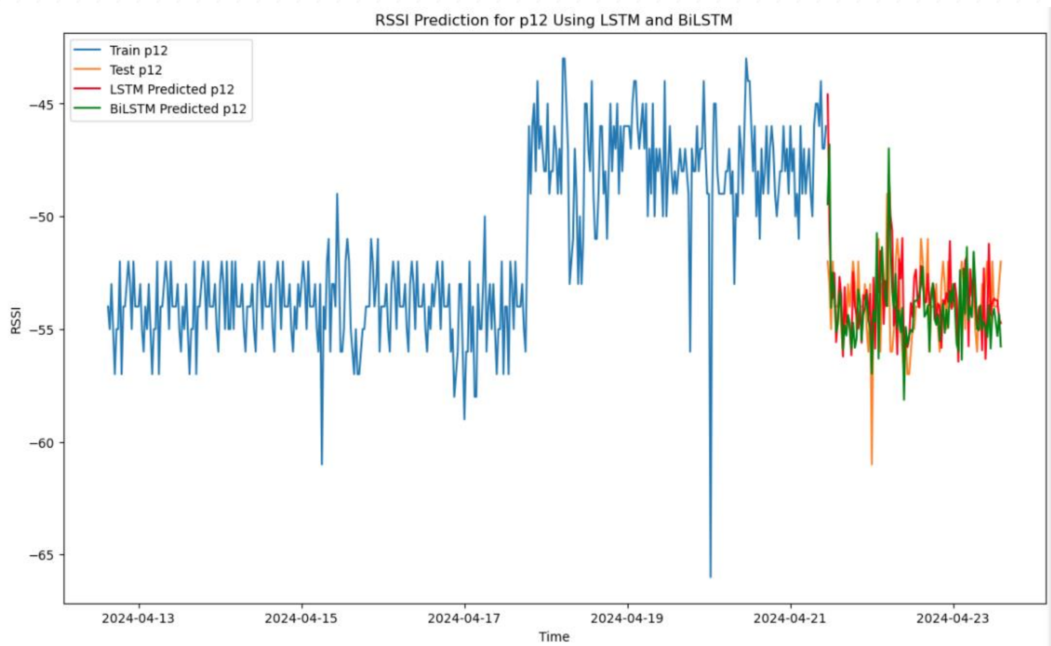

Fig 4. 17. RSSI prediction for P12 using LSTM and BiLSTM

4.6. Conclusion

In conclusion, the analysis conducted in this study highlights the advantages of the BiLSTM model over the LSTM model for indoor geolocation tasks. The BiLSTM model's ability to capture bidirectional dependencies in the data resulted in better predictive performance, making it a valuable tool for applications that require accurate indoor positioning.

Future research could explore the use of more advanced techniques, such as convolutional layers, to further improve the performance of these models in indoor geolocation tasks.

Chapter 05. Conclusion et perspectives

Typical LSTM and BILSTM effectiveness were assessed to predict the power of Wi-Fi (RSSI) signal using the real RSSI data collected from an internal environment. The data was divided into training and testing groups, and each model predicted RSSI values in the test group. The performance of the models was assessed using the error measures (MAE, MSE, RSSE, MAPE). The results showed that both models performed well, with a little BILSTM over LSTM. A simulation using Python was also performed to further validate the results. This study provides a promising approach to predicting the accuracy of RSSI in internal environments.

The obtained results show that the BiLSTM performance is better than LSTM and shows a better accuracy in term of indoor position prediction. However, the forecasting plots show the existence of outliers in the measures and these outliers can generate high error during the prediction. During our simulation we tested two different Wi-Fi hotspots from two different generations and we discovered that the used technology impacts the variability of the RSSI signal. This point should be explored in the future by simulating different Wi-Fi hotspots and comparing results according to their characteristics. Also, as future work we need to envisage a simulation for long period to analyze the pics and explore if they can be considered as outlier or seasonal effects

BIBLIOGRAPHIE

[1] GHERIB, K. *GEOLOCALISATION DES PERSONNES EN ENVIRONNEMENT INDOOR VIA LA FUSION DES CAPTEURS DE SMARTPHONE* (Doctoral dissertation, UNIVERSITÉ KASDI MERBAH OUARGLA). https://dspace.univ-ouargla.dz/jspui/bitstream/123456789/29135/1/Theme-GhribKamel.pdf

[2] Chen, P., Zheng, X., Gu, F., & Shang, J. (2020). Path distance-based map matching for Wi-Fi fingerprinting positioning. *Future Generation Computer Systems*, *107*, 82-94. https://doi.org/10.1016/j.future.2020.01.053

[3] Wahab, N. H. A., Sunar, N., Ariffin, S. H., Wong, K. Y., & Aun, Y. (2022). Indoor positioning system: A review. *International Journal of Advanced Computer Science and Applications*, *13*(6).

[4] Hara, S., Zhao, D., Yanagihara, K., Taketsugu, J., Fukui, K., Fukunaga, S., & Kitayama, K. I. (2005, May). Propagation characteristics of IEEE 802.15. 4 radio signal and their application for location estimation. In *2005 IEEE 61st Vehicular Technology Conference* (Vol. 1, pp. 97-101). IEEE. https://doi.org/10.1109/VETECS.2005.1543257

[5] Zafari, F., Gkelias, A., & Leung, K. K. (2019). A survey of indoor localization systems and technologies. *IEEE Communications Surveys & Tutorials*, *21*(3), 2568-2599. https://doi.org/10.1109/COMST.2019.2911558

[6] Abd Elgwad, M., & Sheta, B. I. (2019, September). Wi-Fi based indoor localization using trilateration and fingerprinting methods. In *IOP Conference Series: Materials Science and Engineering* (Vol. 610, No. 1, p. 012072). IOP Publishing. https://iopscience.iop.org/article/10.1088/1757-899X/610/1/012072/pdf

[7] Subedi, S., & Pyun, J. Y. (2020). A survey of smartphone-based indoor positioning system using RF-based wireless technologies. *Sensors*, *20*(24), 7230. https://doi.org/10.3390/s20247230

[8] https://www.youtube.com/watch?v=xEPWRk4z5FY&t=294s

[9] http://www.ee.ic.ac.uk/pcheung/teaching/DE2_EE/Lecture%208%20-%20Frequency%20Responses%20%28x1%29.pdf

[10] Nguyen, K. A., Luo, Z., Li, G., & Watkins, C. (2021). A review of smartphones-based indoor positioning: Challenges and applications. *IET Cyber-Systems and Robotics*, *3*(1), 1-30. https://doi.org/10.1049/csy2.12004

[11] Singh, D. (2020). Channel Scanning and Access Point Selection Mechanisms for 802.11 Handoff: A Survey. https://scholarcommons.scu.edu/cgi/viewcontent.cgi?article=1014&context=cseng_mstr

[12] Liu, X., Krishnamachari, B., & Liu, H. (2010, April). Channel selection in multi-channel opportunistic spectrum access networks with perfect sensing. In *2010 IEEE symposium on new frontiers in dynamic spectrum (DySPAN)* (pp. 1-8). IEEE. https://doi.org/10.1109/DYSPAN.2010.5457885

[13] Gimenez-Guzman, J. M., Marsa-Maestre, I., de la Hoz, E., Orden, D., & Herranz-Oliveros, D. (2023). Channel selection in uncoordinated ieee 802.11 networks using graph coloring. *Sensors*, *23*(13), 5932. https://doi.org/10.3390/s23135932

[14] Bencharif, L. (2020). *Localisation en intérieur basée sur l'apprentissage automatique : exploitation des informations sur l'état des canaux Wi-Fi* (Doctoral dissertation, Université du Québec à Trois-Rivières). https://depot-e.uqtr.ca/id/eprint/9546/1/eprint9546.pdf

[15] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

[16] Gers, F. A., & Schmidhuber, E. (2001). LSTM recurrent networks learn simple context-free and context-sensitive languages. *IEEE transactions on neural networks*, *12*(6), 1333-1340. https://doi.org/10.1109/72.963769

[17] Gers, F. A., Schraudolph, N. N., & Schmidhuber, J. (2002). Learning precise timing with LSTM recurrent networks. *Journal of machine learning research*, *3*(Aug), 115-143. https://www.jmlr.org/papers/volume3/gers02a/gers02a.pdf

[18] Hochreiter, S., Bengio, Y., Frasconi, P., & Schmidhuber, J. (2001). Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.

[19] Fernández, S., Graves, A., & Schmidhuber, J. (2007). Sequence labelling in structured domains with hierarchical recurrent neural networks. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence, IJCAI 2007*. https://mediatum.ub.tum.de/doc/1289940/document.pdf

[20] Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006, June). Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning* (pp. 369-376). https://doi.org/10.1145/1143844.1143891

[21] Sharfuddin, A. A., Tihami, M. N., & Islam, M. S. (2018, September). A deep recurrent neural network with bilstm model for sentiment classification. In *2018 International conference on Bangla speech and language processing (ICBSLP)* (pp. 1-4). IEEE. https://doi.org/10.1109/ICBSLP.2018.8554396

[22] https://www.irit.fr/~Zoubir.Mammeri/Cours/Introduction_IEEE802-11.pdf

[23] Mishina, Y., Murata, R., Yamauchi, Y., Yamashita, T., & Fujiyoshi, H. (2015). Boosted random forest. *IEICE TRANSACTIONS on Information and Systems*, *98*(9), 1630-1636. https://doi.org/10.1587/transinf.2014OPP0004

[24] Liu, Y., & Yang, S. (2022). Application of Decision Tree-Based Classification Algorithm on Content Marketing. *Journal of Mathematics*, *2022*(1), 6469054. https://doi.org/10.1155/2022/6469054

ملخص:

لقد أصبح تحديد الموقع الجغرافي في داخل مبنى مسألة هامة في العديد من المجالات، مثل الملاحة الداخلية (*Indoor Geolocation*)، ولوجستيات المستودعات، أو حتى التسويق المستهدف في مراكز التسوق.

تُظهر تقنيات تحديد الموقع الجغرافي التقليدية، مثل نظام تحديد المواقع العالمي(GPS) ، حدودها في البيئات الداخلية بسبب فقدان الإشارة الناتج عن هياكل المباني. وبالتالي، توفر شبكات WI-FI وأجهزة استشعار الهواتف الذكية بديلاً واعدًا لتحديد الموقع الجغرافي داخل المباني.

يهدف هذا البحث في درجة الماستر إلى تحليل الأعمال القائمة واقتراح حل دقيق وقوي لتحديد الموقع الجغرافي الداخلي من خلال دمج إشارات شبكات WI-FI المحيطة وأجهزة استشعار الهواتف الذكية (مثل مقياس التسارع، الجيروسكوب، الميغناتومتر، إلخ). الهدف الرئيسي هو تصميم خوارزمية دمج البيانات التي تجمع بفعالية المعلومات الواردة من هذه المصادر المختلفة للحصول على تحديد موقع دقيق في الوقت الحقيقي.

الكلمات المفتاح: تحديد الموقع الجغرافي ، ويفي ، الهواتف الذكية، أجهزة الاستشعار، خورزمات.

Abstract :

Indoor geolocation has become a major focus in various fields, such as indoor navigation, warehouse logistics, and targeted marketing in shopping centers. Traditional geolocation technologies, like GPS, show their limitations in indoor environments due to signal loss caused by building structures. As a result, Wi-Fi networks and smartphone sensors present a promising alternative for indoor geolocation.

This master's thesis aims to analyze existing work and propose a precise and robust indoor geolocation solution by integrating signals from surrounding Wi-Fi networks and smartphone sensors (such as accelerometers, gyroscopes, magnetometers, etc.). The primary objective is to design a data fusion algorithm that effectively combines information from these different sources to achieve accurate real-time localization.

Keywords: *Geolocation; Smartphone; Wi-F; Deep Learning; Inertial Measurement Unit (IMU) sensors.*

Résumé :

La géolocalisation indoor est devenue un enjeu majeur dans de nombreux domaines, tels que la navigation en intérieur (indoor geolocation), la logistique des entrepôts, ou encore le marketing ciblé dans les centres commerciaux. Les technologies de géolocalisation traditionnelles, comme le GPS, montrent leurs limites en environnement intérieur en raison de la perte de signal relatif aux structures des édifices. Ainsi, les réseaux Wi-Fi et les capteurs des smartphones offrent une alternative prometteuse pour la géolocalisation indoor.

Ce stage de master vise à analyser les travaux existants et proposer une solution de géolocalisation indoor précise et robuste en intégrant les signaux des réseaux Wi-Fi environnants et les capteurs des smartphones (tels que l'accéléromètre, le gyroscope, le magnétomètre, etc.). L'objectif principal est de concevoir un algorithme de fusion de données qui combine efficacement les informations provenant de ces différentes sources pour obtenir une localisation précise en temps réel.

Mots clés : *Géolocalisation; Smartphone; Wi-F; Apprentissage Approfondi; Capteurs d'unités de mesure inertielle (UMI).*