DEVICE-FREE DETECTION OF HUMAN MOVEMENT IN OUTDOOR ENVIRONMENTS USING WI-FI CHANNEL STATE INFORMATION

Nayan Gokull Dharmaraj

218021P

Master of Science (Major Component Research)

Department of Electronic and Telecommunication Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

May 2025

DEVICE-FREE DETECTION OF HUMAN MOVEMENT IN OUTDOOR ENVIRONMENTS USING WI-FI CHANNEL STATE INFORMATION

Nayan Gokull Dharmaraj

218021P

Thesis submitted in partial fulfillment of the requirements for the degree
Master of Science (Major Component Research)

Department of Electronic and Telecommunication Engineering
Faculty of Engineering

University of Moratuwa
Sri Lanka

May 2025

DECLARATION

I declare that this is my own work and this Thesis does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 29/07/2025

The supervisors should certify the Thesis with the following declaration.

The above candidate has carried out research for the Master of Science (Major Component Research) Thesis under our supervision. We confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Prof. S A D Dias

Signature of the Supervisor: Date: 29/07/2025

Name of Supervisor: Dr. K T Hemachandra

Signature of the Supervisor: Date: 30/07/2025

DEDICATION

To my beloved parents.

ACKNOWLEDGEMENT

I would like to deeply express my acknowledgment to each and every individual who provided support to ensure the completion of this Masters research.

I would like to extend my sincere gratitude to my superviosrs Prof. S A D Dias, Professor, Department of Electronic and Telecommunication, University of Moratuwa and Dr. Kasun T. Hemachandra, Senior Lecturer, Department of Electronic and Telecommunication, University of Moratuwa, for providing their valuable time, unconditional support and guidance throughout the research.

Special thank you goes out to Mr. Tilan Wickramarachchi, Mr. Nipun Hearth, Mr. Dilhan Senuruk, Mr. Sobedan Sridaran, Mr. Vignesh Veerappa, Mr. Imaad Aarif and Ms. Senuri Pussewala for their support for taking part in my experimentation and supporting me with research related activities.

I would like to Acknowledge the Research Coordinator and the Examiner Panel who reviewed my progress and provided valuable feedback to enhance my research outcomes.

Finally, i would like to share my gratitude to all the friends, colleagues from Dialog Axiata PLC and family for supporting me throughout this research.

ABSTRACT

Real-time pedestrian sensing at crossings is essential for Intelligent Transportation Systems (ITS). World Health Organization (WHO) reports that vehicular accidents claim nearly 3,500 lives daily and result in 20 to 50 million injuries annually. Traditional vision-based solutions and wearable devices attempts to solve this challenge however face privacy, cost and deployment challenges. While Wi-Fi channel state information (CSI) has shown promise in indoor environments, its potential for dynamic outdoor environments remains largely unexplored. This paper presents a device-free, CSI-based system for detecting human movement and walking direction in crossings and alert oncoming vehicles to enhance pedestrian safety. A two-stage classification framework is adopted with Amplitude and Phase-sensitive features filtered with a adaptive noise suppression and evaluated using classical and deep learning models. Tested across four outdoor locations and seven subjects, the system demonstrates **95.9%** accuracy for movement detection and up to **62.4%** for directional classification. Results demonstrate the feasibility of deploying low-cost commercial off-the-shelf (COTS) hardware such as ESP32 and Raspberry Pi hardware for scalable, privacy-preserving ITS applications.

Keywords: Outdoor CSI, Smart Pedestrian Crossing using CSI, Smart Crosswalk application using CSI

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
ANN	Artificial Neural Networks
AP	Access Point
BiLSTM	Bi-directional LSTM
CARM	CSI-based human Activity Recognition and Monitoring system
CFO	carrier frequency offset
CFR	Channel Frequency Response
CNN	Convolutional Neural Network
ConvLSTM	Convolutional-LSTM
COTS	commercially off-the-shelf
CSI	Channel State Information
DNN	Deep Neural Networks
DWT	Discrete Wavelet Transform
EMF	electromagnetic force
ENTC	Department of Engineering and Telecommunication
EoL	End of Life
ESP32	Espressif Systems 32
ETSI	European Telecommunications Standards Institute
FFT	Fast Fourier Transform
GPU	graphics processing unit
HAR	Human Activity Recognition
HMM	Hidden Markov Model
I/O	Input and Output
IFFT	Inverse Fast Fourier Transform
ISAC	Integrated Sensing and Communication
ITF	International Transport Forum
ITS	Intelligent Transportation Systems
LOOCV	Leave-One-Out Cross validation
LSTM	Long Short-Term Memory
NIC	Network Interface Card
NLOS	Non-Line-Of-Sight
OFDM	Orthogonal Frequency-Division Multiplexing

Abbreviation	Description
PCA	Principal Component Analysis
PHY	physical layer
PRI	Pedestrian's Risk Index
RF	Random Forest
RFID	Radio Frequency Identification
RNN	Recurrent Neural Networks
RSSI	Received Signal Strength Indicator
RSU	Road-side Unit
SoC	System-on-Chip
STA	Station
STFT	Short-Time Fourier Transform
STO	sampling time offset
SVM	Support Vector Machine
ToC	Time-To-Collision
Tx-Rx	Transmitter-Receiver
UART	Universal Asynchronous Receiver/Transmitter
UNO	United Nations Organization
UoM	University of Moratuwa
USB	Universal Serial Bus
UWB	Ultra-Wideband
V2I	Vehicle-to-Infrastructure
V2P	Vehicle-to-Pedestrian
V2V	Vehicle-to-Vehicle
V2X	Vehicle-to-Everything
VAM	VRU Awareness Message
VRU	Vulnerable Road User
WHO	World Health Organization

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Research Overview

Real-time Human Activity Recognition (HAR) in various environments has emerged as a critical technology for enhancing public safety and operational efficiency. Its wide range of applications in Intelligent Transportation Systems (ITS) include smart pedestrian crossings, crowd monitoring and emergency response coordination. The importance of such systems is highlighted by global statistics from the World Health Organization (WHO) , which report that vehicular accidents claim nearly 3,500 lives daily and result in 20 to 50 million injuries annually [21]. Alarmingly, more than 90% of these incidents occur in low- and middle-income countries, often due to distracted driving, poor vehicle maintenance, inadequate road infrastructure due to socioeconomic factors. The International Transport Forum (ITF) 's Road Safety Annual Report 2024 further emphasizes the global initiative to improve road safety through strategic and cost-effective interventions. Most of the 35 surveyed countries have committed with the United Nations Organization (UNO) 's dedicated goal of reducing road fatalities up to 50% by 2030 [22].

To support this, effective HAR in ITS particularly for applications like smart pedestrian crossings aims to improve situational awareness and deliver timely alerts to drivers which in turn enhances pedestrian safety. Existing HAR systems commonly rely on technologies such as vision-based systems (e.g. CCTV cameras and thermal imaging), wearable sensors (e.g. smartphones, smartwatches), Radio Frequency Identification (RFID) tags [23] and network-based triangulation.

Vision-based systems, for instance are intrusive, cannot operator under Non-Line-Of-Sight (NLOS) condition and are highly susceptible to lighting and weather conditions while demanding significant computational costs. Wearable devices while considered accurate, depend on user compliance and limited battery life, making them impractical for pedestrian behavior sensing in ITS. Similarly, RFID and network-based solutions suffer from design constraints and accuracy issues, coupled with similar privacy concerns to vision-based systems and have significant reliability in specialized infrastructure, thus restricting their broader applicability for ITS.

Recently, Wi-Fi Channel State Information (CSI) has emerged as a promising alternative given that there were up to 66% Wi-Fi based device-free detection [24]methods taking advantage of non-intrusive, device-free sensing with results demonstrating better performances compared to traditional Received Signal Strength Indicator (RSSI) methods [25, 26]. CSI-based HAR utilizes fluctuations in wireless signal propagation caused by human motion to detect and classify activities. While this approach

has shown high accuracy in indoor studies such as WiDir [10] and WiCount [27], which benefit from consistent multipath conditions due to static reflectors like walls and furniture, it remains largely underexplored in outdoor conditions. Consequently, its applicability to dynamic real-world pedestrian scenarios is still not well established.

1.2 Objectives

- RO1: Determine a suitable acquisition and preprocessing hardware platform for CSI data.
- RO2: Devise an algorithm to statistically analyze CSI for movement in noisy outdoor environments.
- RO3: Enhance the algorithm to identify movement by humans in opposing directions.
- RO4: Design a deployable system architecture for detecting pedestrian movement on a crossing.

1.3 Research Contribution

This research addresses a critical gap in the current landscape of HAR technologies by shifting the focus from controlled indoor CSI applications to outdoor environments such as pedestrian detection within ITS applications. Unlike the existing literature, which largely confines CSI-based analysis to indoor environments, this study develops and evaluates a dedicated framework to detect pedestrian movement at crossings in complex, real-world outdoor settings.

Contrary to indoor settings, where CSI information is a strong isolated deterministic feature due to the rich multipath reflections from ceilings, walls and furniture against motion, outdoors settings produces reflections which are sparse and mixed with dynamic low and high frequency noise caused by wind, animals and birds, vehicles and other environmental clutter. To overcome these challenges in outdoor CSI sensing, adaptive filtering across the CSI subcarriers is adopted and extracts statistical features that inherently amplify the motion-related information of movement and diminish the noise introduced by the dynamic settings. This enables the design of a robust and adaptable outdoor CSI-based HAR system for pedestrian crossings, optimized for lightweight and low-cost deployment on commercially off-the-shelf (COTS) hardware platforms.

In contrast to more resource-intensive alternatives such as vision and RSSI-based systems, the proposed solution becomes scalable and privacy-preserving as summarized in Table 1.1. Hence the study investigates CSI signal behavior across for motion

and directional cues across multiple subjects and diverse outdoor environments, offering novel insights into environmental and user-specific variability in outdoor environments.

The scope of this study primarily focuses on detecting movement in a channel while classifying direction of movement of one pedestrian only. Multi-person detection scenarios are beyond the current implementation discussed and is identified as future directions for this research. Nonetheless, this work will contribute to the ecosystem of vehicular communication system by establishing the practicality of integrating CSI-based pedestrian sensing as an input that could behave as the Road-side Unit (RSU) to systems such as hybrid DSRC/Wi-Fi frameworks [28] for fulfilling Vehicle-to-Everything (V2X) systems. This integration would support real-time alerts about Vulnerable Road User (VRU) improving pedestrian safety, aligning with Integrated Sensing and Communication (ISAC) objectives which can be targeted for ITS.

TABLE 1.1: COMPARISON BETWEEN EXISTING AND PROPOSED PEDESTRIAN DETECTION APPROACH

Attribute	**Vision-Based**	**RSSI-Based**	**Radar-Based**	**PIR / Ultra-sonic**	**Related CSI Work**	**This Work**
Privacy	No	Yes	Yes	Yes	Yes	**Yes**
Truly Outdoor	Yes	No	Yes	Limited	No	**Yes**
Cost	High	Low	High	Low	Low	**Low**
Accuracy	High	Low	High	Low	High	**High**
Real-time Capable	Yes	Yes	Yes	Yes	Yes	**Yes**
NLOS Capability	No	Yes	Yes	No	Yes	**Yes**

Note. Related CSI Works [8, 11, 13, 29]

1.4 Methodology

The Methodology of the research is categorized into 4 key milestones:

1. Identifying and setting up CSI and implementation platforms.

2. Treating collected CSI data and engineer relative features to achieve objectives.

3. Conduct Exploratory data analysis on the treated CSI data in dynamic environments and subjects.

4. Develop and evaluate classification models for movement and direction as a lightweight system.

The research is conducted in phases to achieve the specified milestones. The overview shown in Fig 1.1 outlines the literature survey and understanding the components of achieving the objectives. Supported by clear data collection and extraction strategy, a system is designed to enable the classification and achieving the milestones outlined above.

1.5 Organization of the Thesis

This thesis is organized into five chapters. **Chapter 1** introduces the research, the motivation and objectives along with the key contributions. **Chapter 2** compiles a comprehensive literature survey on enabling technologies, applicability of CSI for HAR systems covering ITS and pedestrian safety applications. **Chapter 3** details the methodology, including CSI extraction, data collection, hardware setup, preprocessing and feature engineering techniques and the proposed classification framework. **Chapter 4** discusses experimentation and results covering testing strategies, model evaluations and exploratory analyses on the variability of CSI data. **Chapter 5** concludes the thesis by summarizing findings, presenting recommendations and proposing a sustainable ITS application.

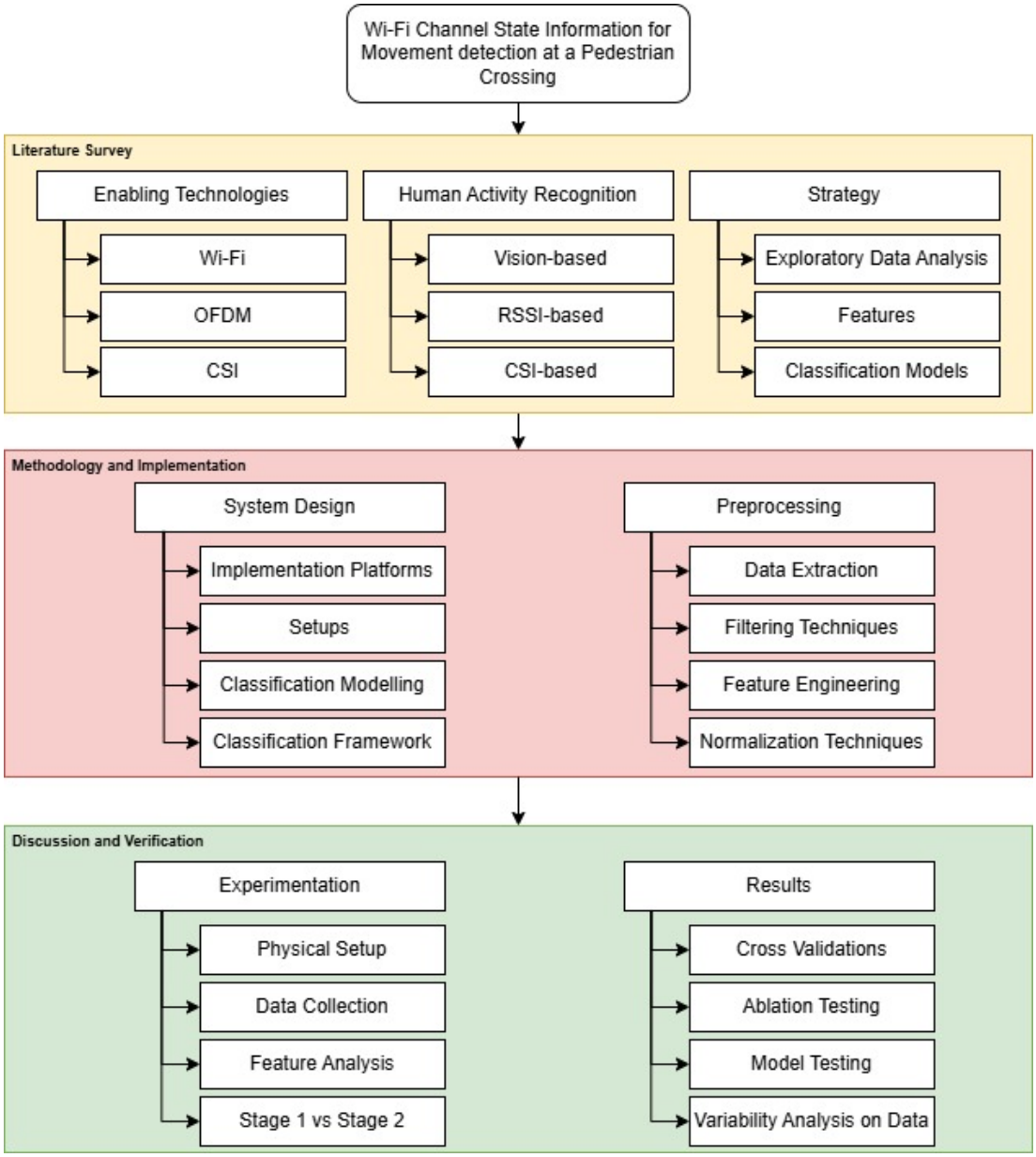


Fig. 1.1: Research Overview

CHAPTER 2

LITERATURE SURVEY

2.1 Intelligent Transport Systems (ITS)

Intelligent Transport systems include state-of-the-art wireless, sensing and automated technologies which can seamlessly integrate vehicles, users and infrastructure [30] to create an advanced and sustainable network. The objective of ITS is ensuring optimal use of innovative technologies to enhance safety while improving efficiencies. Key areas of ITS include sensing, control, information processing, data analytics and communication which enables ITS systems to deploy real-time monitoring and management applications.

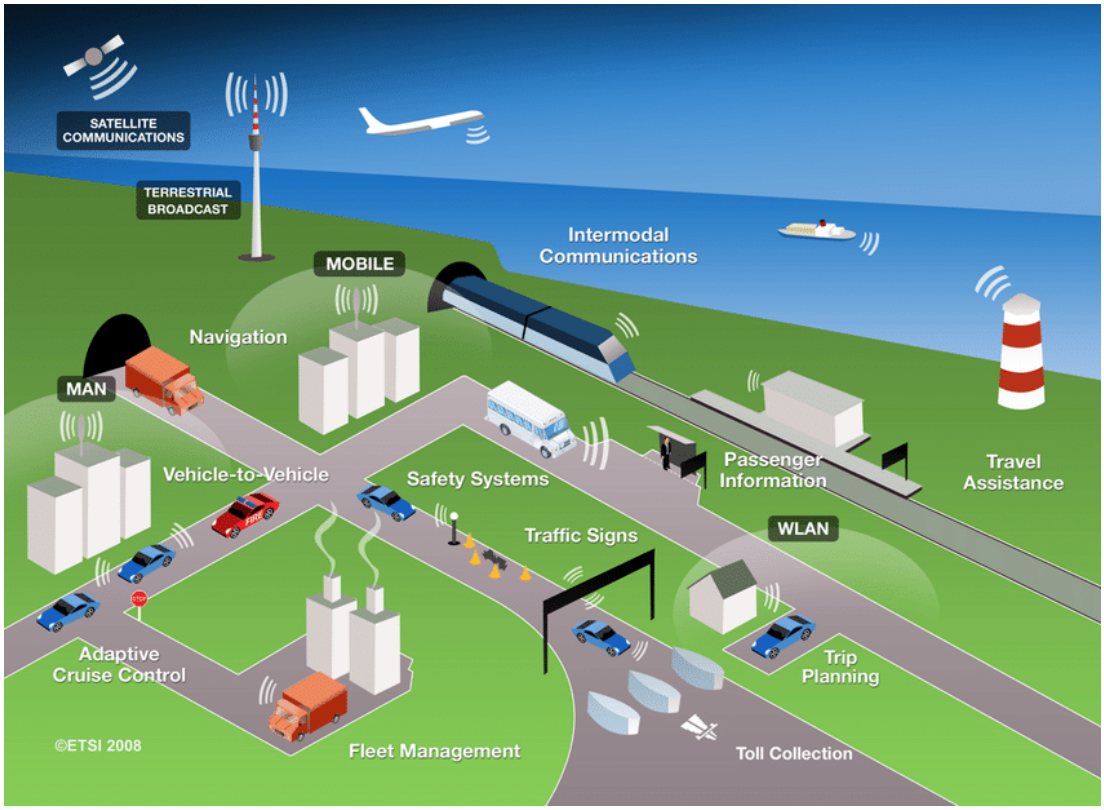


Fig. 2.1: Intelligent Transport System applications at a Glance [1]

A key component that enable the application of ITS is V2X communication which encompasses Vehicle-to-Vehicle (V2V) , Vehicle-to-Infrastructure (V2I) and Vehicle-to-Pedestrian (V2P) systems. These allow information such as speeds, positions and trajectories to be shared between infrastructure to support analytics and enable proactive decision making. Some of the applications are illustrated in Fig. 2.1.

ITS plays a significant role in promoting sustainable transportation by optimizing

traffic flow and reducing idle times which in turn lower fuel consumption and decrease greenhouse gas emissions. Additionally, ITS supports the integration of multimodal transportation options such as public transit, cycling and walking, encouraging shifts towards more sustainable modes of travel. In the context of developing countries, ITS aims to create opportunities to address transportation challenges associated with rapid urbanization and limited infrastructure. Implementing ITS solutions can lead to improved traffic management, enhanced road safety and increased accessibility to overall transportation services. However, the deployment of ITS in these regions requires careful consideration of cost-effectiveness, scalability and adaptability to local contexts.

Some ITS applications are listed below:

1. Advanced Traffic Management Systems [31].

2. Automatic Vehicle Guidance [32].

3. Adaptive Cruise Control [33].

4. Vulnerable Road User Safety Systems [34].

2.2 Pedestrian Crossings

The literature review on pedestrian crossings, setups and behavior of pedestrians was conducted to obtain insights related to design features such as minimum width, expected pedestrian patterns forming the feasibility of a smart pedestrian crossing. Crosswalks, commonly referred to as Zebra or Pedestrian crossings, form a critical part of the road infrastructure as shown in Fig. 2.2. They are designed to allow pedestrians to move in an orderly manner between roadways providing a safety framework for pedestrians within transportation systems.

Within the framework, traditional crossings rely on passive indicators such as visual markings and signage as shown in Fig. 2.3 placed around the crossing. These roadsigns offer limited safety to pedestrians and no support to the vehicles that approach the crossing in complex urban and sub-urban settings especially in low light conditions. With urbanization, the activity of pedestrians increases drastically which presses the need to deploy responsive safety applications for pedestrian crossings [35]. Developments of smart pedestrian crossings as a subset of human activity and behavioral detection systems have been an emerging ITS application [36, 37] which can actively detect presence or movement and provide real-time alerts to vehicles. Such systems can couple as sensing inputs to the RSU proposed in [28] and create cooperative V2X systems providing adaptive responses under dynamic conditions.

Fig. 2.2: A Typical four-lane Pedestrian Crossing in Colombo, Sri Lanka

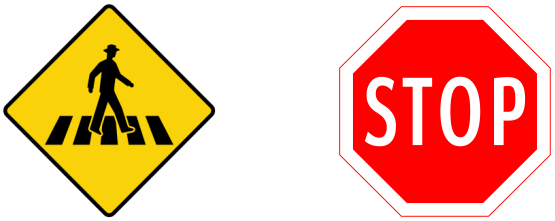


Fig. 2.3: Typical Roadsigns in Sri Lanka (Vienna Convention)

2.2.0.1 Pedestrian Walking Speeds

Average walking speeds of general healthy pedestrians are stated to be roughly between 1.2 to 1.3 meters per second [38] as summaried in Table 2.1. The speeds may vary based on the age, gender and exceptional conditions such as heavy traffics [39].

TABLE 2.1: PEDESTRIAN WALKING SPEEDS SUMMARY ACROSS LITERATURE

Source	Location	Mean (m/s)	Standard Deviation (m/s)
[38]	Sri Lanka	1.33	0.18
[39]	China	1.22	0.41
[40]	India	1.21	–

2.2.0.2 Pedestrian Crossing Design and layout

The distinguishable feature of the design of a typical pedestrian crossing is the alternating design of white and dark stripes on the surface of the road as shown in 2.4. Until December 2016, all pedestrian crossings in Sri Lanka were yellow which were

then colored white to match the international standard [41]. Typical road widths in Sri Lanka range between 5 to 7 meters [38] and extend up to 9 meters or more including the shoulders and kerbs for a standard two-lane carriageway making up about 65% of the road networks as stated in the national road master plan [40, 42, 43]. This seems to be consistent for other countries like the United Kingdom [44].

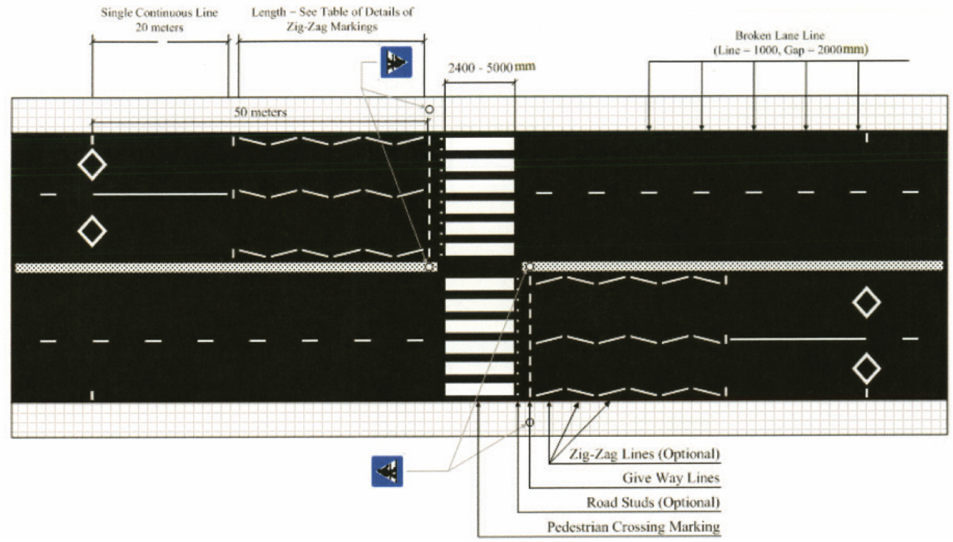


Fig. 2.4: Typical Pedestrian crossing design layout in Sri Lankan roads [2]

2.3 Human Activity Recognition (HAR)

Human Activity Recognition (HAR) refers to the process of identifying and recognizing human behaviors in varying conditions from sensor data using classification modeling. HAR has become a popular research component in recent years as illustrated in [24]. HAR enables context-aware computing, where machines adapt their responses based on a user's current state or motion pattern. Its applications are widespread across domains such as healthcare, surveillance, elderly care, smart homes applications and intelligent transportation systems (ITS). Traditionally, HAR systems are dominated by vision-based systems and wearable devices. However, radio frequency based systems such as RSSI and CSI have become popular. These studies have become especially valuable in monitoring vulnerable individuals such as the elderly, assisting autonomous systems in V2X systems for data-driven decision making and enhancing safety in ITS applications.

2.3.1 Device-free HAR for Pedestrian Detection in ITS

Integration of HAR into pedestrian detection within ITS introduces a distinct set of requirements and constraints. In outdoor and semi-controlled environments, the systems must account for dynamic background interference, natural lighting variability

and unpredictable pedestrian behavior [29, 45]. As studies have indicated, many road fatalities involve human errors or insufficient awareness of pedestrian intent and behavior, particularly at crossings and intersections [21]. This emphasizes on making pedestrian based device-free HAR systems which contributes directly to building safer and responsive environments.

The core objective of a device-free pedestrian based HAR is to enable real-time and accurate detection of human presence and motion, particularly at crossings while maintaining privacy. This is vital for the safety of VRUs, where timely alerts and adaptive responses from infrastructure to vehicles can prevent collisions and enhance traffic efficiency. Recent research has evaluated various technologies including vision-based systems [3, 4, 6, 45], radio frequency-based systems such as RSSI [7, 46, 47] and Wi-Fi CSI [8, 9, 11, 29, 48] to determine their feasibility and performance in real-world conditions.

2.3.2 Vision and Image based Systems

Vision-based Human Activity Recognition (HAR) leverages camera systems to detect and monitor human behavior. An extensive survey conducted on vision based system is available in [45] covering major aspects of the HAR systems. Traditionally used in surveillance and assistive technologies, these systems have increasingly been adopted in ITS especially for their pedestrian detection and traffic behavioral analysis capabilities. Their ability to provide rich spatial and contextual information by defining regions of interests or single shot detection [49] including posture, trajectory and localization data is beneficial for applications like vehicular collision avoidance, pedestrian crossing safety and even autonomous navigation in V2X systems.

One of the focused studies on pedestrian safety includes the study in [3] on using CCTV infrastructure to detect and analyze behaviors at intersections marked risky for pedestrians and vehicle. The system utilizes object detection techniques such as YOLO [50] to map trajectories to identify the behavioral patterns from the video feeds samples of which are found in Fig. 2.5. The system was setup in Osan, South Korea where the near-miss interactions between vehicles and pedestrians at pedestrian crossings are quantified and recorded. The study suggests the applicability of re-purposing publicly installed CCTV cameras for intelligent mobility monitoring without the need to install new infrastructure.

This comprehensive study contributes to a metric for development of ITS in pedestrian safety called the Pedestrian's Risk Index (PRI) . The paper measures variables such as Time-To-Collision (ToC) and trajectories to allow for data-driven decision making.

While effective in extracting contextual risk with accuracies ranging in 90% as shown in Fig. 2.6, the system still faces practical deployment challenges, including

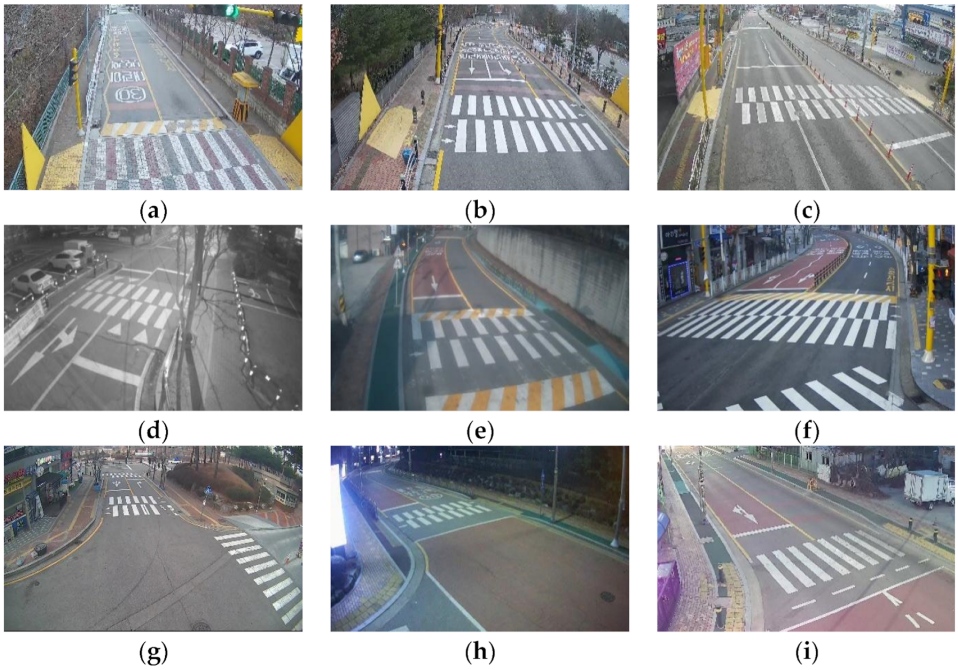


Fig. 2.5: Actual CCTV Views from locations labeled (a)-(i) in South Korea in [3]

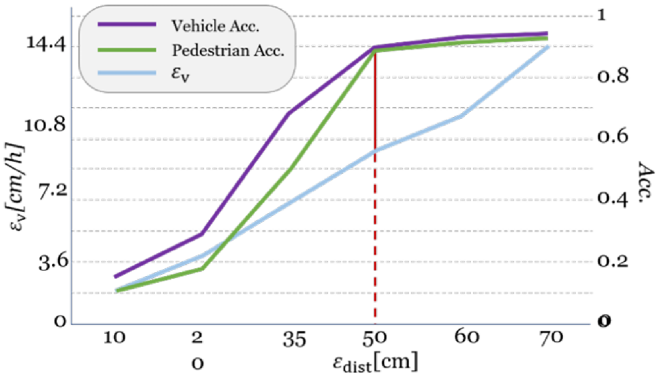


Fig. 2.6: Accuracy Graph illustrated in the Study [3]

sensitivity to occlusion, lighting and high computational requirements for real-time processing. Additionally, privacy concerns related to continuous video surveillance may impact its scalability. The study suggests the consideration on upgrading existing CCTV deployments in South Korea which. This is a viable option across most developed countries however, alternatives may be required for developing countries where public CCTV infrastructure is currently lacking.

Another compelling example of a vision-based system is presented in [4], where the authors proposed a deep neural network framework for tracking pedestrians and vehicles using computer vision and video data collected from street level. The study uses a multi-stage system that integrates popular object detection and visual feature extraction algorithms along with identification networks to ensure consistent object tracking

across video frames captured. The approach also uses Kalman filters and mapping to convert image-plane detections to count pedestrians and vehicles in a region of interest while also estimating ToC.

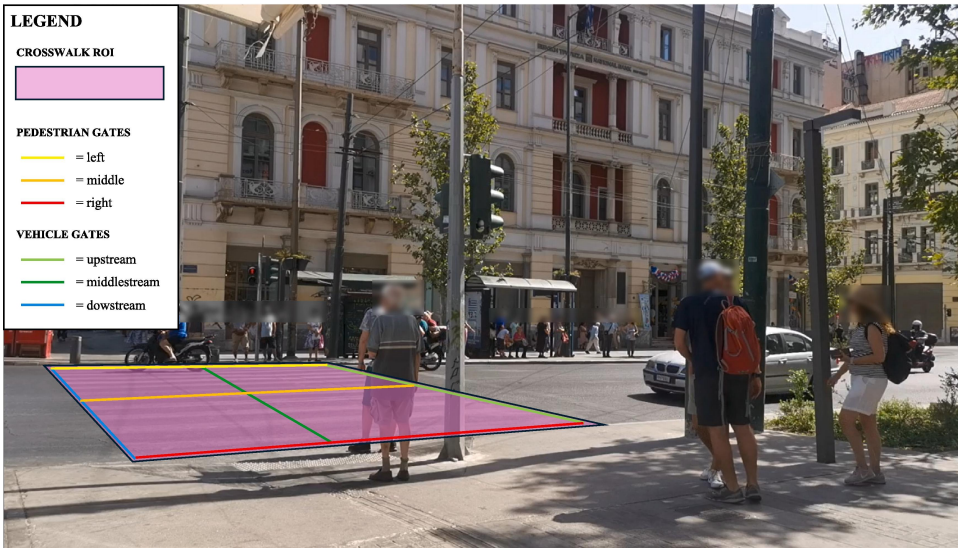
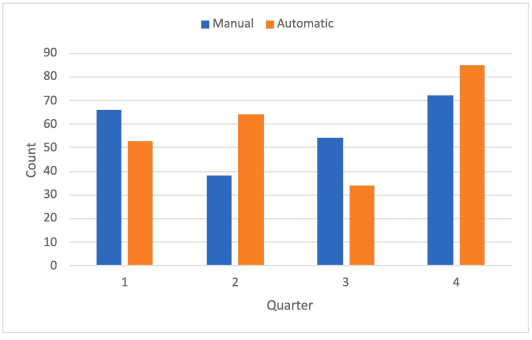
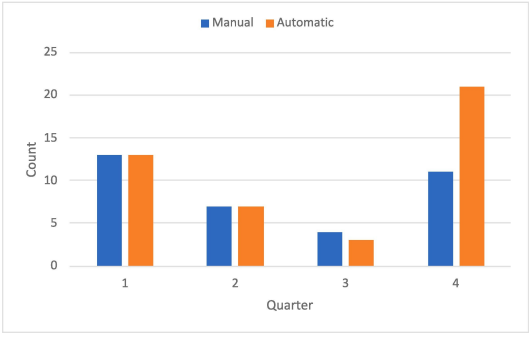


Fig. 2.7: Pedestrian crossing region of interest marked for pedestrian and vehicular detection [4]

The implementation was carried out by recording videos from the streets of central Athens, Greece of which a frame is shown in Fig. 2.7. Leveraging low-cost smartphone cameras and tripods as deployment sensing hardware, the setup achieved robust performance in real-world conditions identifying both legal and illegal pedestrian crossings and counting them to flag as high-risk pedestrian to vehicle interactions as shown in Fig. 2.8a and Fig. 2.8b. The output of such system can be utilized to identify high risks and be used in extended applications of V2I.



(a) Pedestrian crossing: manual vs. automated counts



(b) Vehicle crossing: manual vs. automated counts

Fig. 2.8: Comparison between manual (true count) and automated system outputs for pedestrian and vehicle crossings reported in [4]

This vision-based approach faces several inherent limitations. Occlusions, light-

ing variability and shadows significantly degrade performance, particularly in dense urban settings as outlined in the literature [24]. Moreover, the reliance on camera perspectives limits scalability and adaptability to diverse road layouts. These studies also outlines that they experienced tracking discontinuities and identity switches under severe occlusion or abrupt motion when camera shakes due to nearby vibrations while also acknowledging calibration sensitivity in mapping of video frames which can impact the real-world accuracy. Similar studies and research is also discussed in [5] by using a Deep ConvNet achieves performances up to 85.7%. However the setup utilized as shown in Fig. 2.9 demonstrates a typical setup for computer vision based system needed to achieve the expected results.

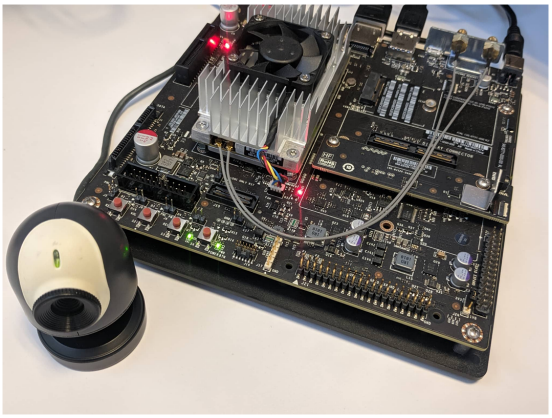


Fig. 2.9: Typical Experimental setup for vision based systems compromising of expensive dedicated graphics processing unit (GPU) and a webcam to capture and process live video [5]

One recent benchmark in multimodal roadside solution "R-LiViT" introduces a large-scale, high-resolution dataset integrating synchronized LiDAR, RGB and thermal imaging data from a roadside unit [6]. The approach specifically focuses on the detection and tracking of VRUs such as pedestrians and cyclists, particularly at urban intersections and crossings where occlusions, lighting variability and cluttered scenes pose significant challenges for perception systems to other system. Data of over 10,000 LiDAR frames and 2,400 aligned RGB and thermal images collected across 150 traffic scenarios one frame of which is shown in Fig. 2.10 and used for evaluation. The authors also offer benchmark evaluations using state-of-the-art detection models like YOLOv8 [50] providing detailed performance metrics under both day and night conditions.

Although the model demonstrates strong detection and tracking performance, the system's dependence on multi-sensor LiDAR setups introduces critical limitations. The need for precise calibrations, secure mounting and high-bandwidth data throughput increasing both deployment complexity and cost. These limits the feasibility of such systems in developing regions or large-scale deployments where cost-efficiency

Fig. 2.10: Visualization of Pedestrian crossing data in R-LiViT with left being original, middle being thermal image and the right being both image overlaid [6]

and simplicity are essential.

Furthermore, most studies acknowledge the privacy concerns of visual data [24, 51, 52] which continue to challenge the widespread deployment of such systems. Regulatory and ethical considerations regarding image storage, facial data and behavioral profiling also factor in which creates the need for encrypted anonymization and data governance policies that increases system complexity and hinders feasibility overall.

2.3.3 Received Signal Strength Indicator based Systems

In recent years increasing amounts of research have emerged leveraging the use of radio frequency based sensing techniques due to their low-cost, privacy-preserving passive nature. RSSI is a coarse wireless signal metric that has been widely explored for device-free human sensing applications [53] due to its simplicity and accessibility.

RSSI captures the aggregate signal power received from the wireless channel, providing a scalar representation of the quality which has gained popularity in cost-effective and computationally less demanding deployments due to their ability to detect macro-level motion patterns such as presence, movement and density changes [46].

The study, WiDet [7] proposes a real-time passive indoor pedestrian detection system that utilizes Wi-Fi RSSI measurements. Unlike conventional approaches that depend on manually generated features such as moving averages or wavelet statistics, WiDet transforms RSSI time series into two-dimensional wavelet coefficient maps as show in Fig. 2.11, which are then processed using a deep Convolutional Neural Network (CNN) . This approach allows for robust learning of signal fluctuation patterns associated with motion information. Their use of data augmentation such as for simulating walking speed changes and wireless interference effects further enhances the generalization of the model. The system, implemented on off-the-shelf Raspberry Pi boards and USB Wi-Fi adapters, achieves a detection accuracy of 95.5% over 163 recorded walking events within indoor Corridors. Additionally, the authors introduce a dynamic time warping (DTW)-based method to infer walking direction across transceiver pairs, achieving direction classification rates exceeding 82% in their corridor deployment.

While WiDet demonstrates the high feasibility of utilizing RSSI-based motion de-

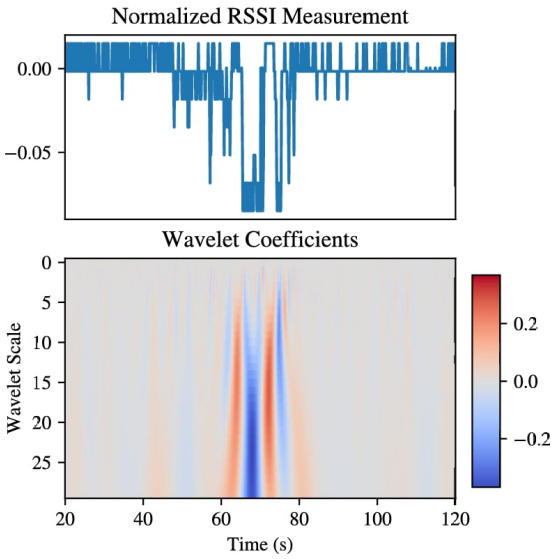


Fig. 2.11: RSSI signal wavelet coefficient map [7]

tection with COTS hardware with great accuracies shown in Fig. 2.12, its effectiveness is tied to controlled indoor settings within constrained environmental variability. Furthermore, the reliance on wavelet preprocessing and CNN-based compensation mechanisms undermines the limitations of RSSI's resolution in capturing fine-grained signal dynamics. By extending directional inference to such conditions without infrastructure constraints, this work builds upon RSSI-based baselines like WiDet while addressing their inherent limitations through the exploitation of fine-resolution CSI features.

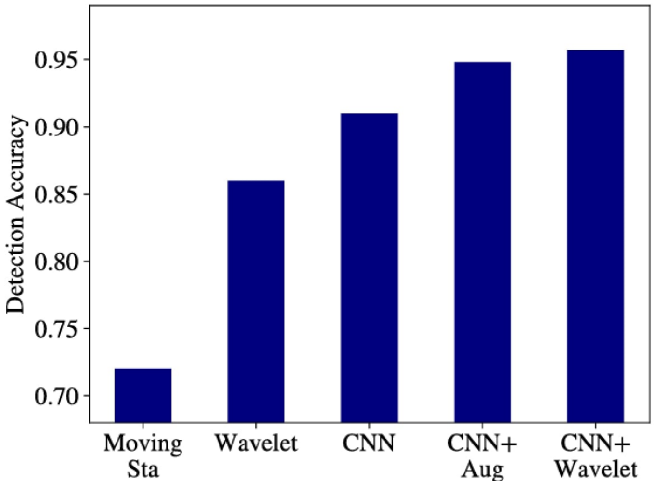


Fig. 2.12: Detection accuracy across multiple approaches in [7]

Recent studies on RSSI especially using Wi-Fi at a pedestrian crossing for ITS application were highly limited. This could be potentially due to the inherent limitations in precision of using RSSI due to noise susceptibility, multipath fading, shadowing effects and signal interference [54]. As RSSI is highly sensitive to environmental

changes as well and offer poor spatial resolution, this makes them inadequate for applications requiring activity recognition or directional detection especially outdoors where multi path fading and shadowing effects are significant.

2.3.4 Channel State Information based Systems

Wi-Fi CSI has emerged as a strong alternative sensing system capable of capturing fine-grained information of the wireless channel. By measuring amplitudes and phase changes across multiple subcarriers, the dynamics of an environment can be recognized being useful for various applications.

With the COTS hardware availability for extracting CSI steadily increasing, it has slowly become a common sensing method. The related studies on utilizing CSI has been captured comprehensively in multiple surveys especially for HAR and ITS related applications in [24, 55–57].

2.3.4.1 Indoor CSI systems

One of the most influential study on using CSI for HAR, CSI-based human Activity Recognition and Monitoring system (CARM) [8], influenced by fine-grained localization studies such as Spotfi [48] introduces the framework for modeling the relationship between human body movement and Wi-Fi CSI signals. An overview of the system utilized in CARM is shown in Fig. 2.13. The study proposes two key models: the CSI-speed model, which correlates body-part velocities with distortions in the data and the CSI-activity model which links these variations to specific motion patterns moving away from traditional statistical metrics used in studies like [].

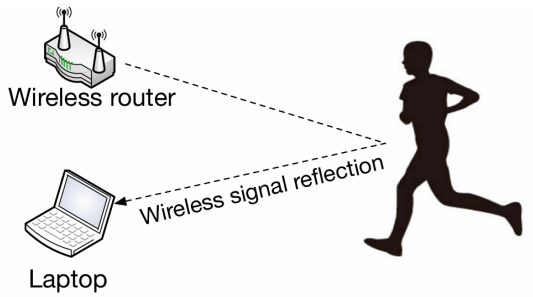
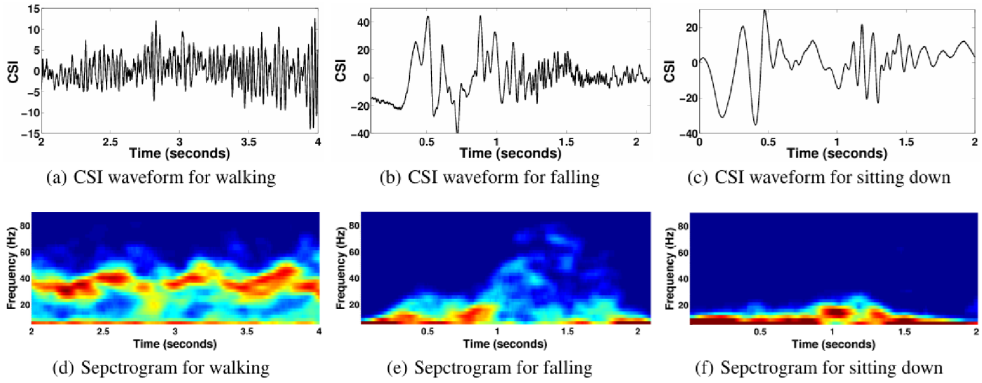


Fig. 2.13: CARM's System shown in [8]

The system utilizes Principal Component Analysis (PCA) to denoise raw CSI measurements collected using Intel 5300 Wi-Fi card and access points of COTS (NETGEAR and TP-Link) routers. It also employed Discrete Wavelet Transform (DWT) to extract multi-resolution features, which were classified using Hidden Markov Model (HMM) to capture temporal transitions between sequential activities. Figure 2.14

shows some waveforms for walking, falling and sitting down setting the expectations for similar activities for this research.



(a) CSI waveform for walking (b) CSI waveform for falling (c) CSI waveform for sitting down

(d) Sepctrogram for walking (e) Sepctrogram for falling (f) Sepctrogram for sitting down

Fig. 2.14: Waveforms and spectrograms for different activities [8]

CARM was able to recognize eight general human activities between Running, Walking, Sitting, Opening refrigerator, Falling, Boxing, Pushing one hand, Brushing teeth and a Empty class with over 96% accuracy as shown in Fig. 2.15. The system demonstrated certain degree of resilience to environmental variation as shown in Fig. 2.16 and packet losses while operating at sampling rates such as 2500 Hz.

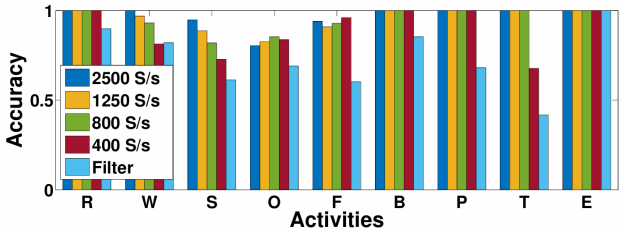


Fig. 2.15: Impact of Sampling Rates on Accuracy [8]

Note. Graph extracted directly from the study [8].
Running (R), Walking (W), Sitting (S), Opening refrigerator (O), Falling (F), Boxing (B), Pushing one hand (P), Brushing teeth (T) and Empty (E).

The study's systemic segmentation, understanding of CSI, modeling approach, testing on multiple environments, testing against different activities sets a strong precedent for this research.

Damodaran and Schäfer [9] presents a comparative study of two CSI-based device-free human activity recognition (HAR) systems implemented in a residential indoor setting. Their setup utilizes Intel 5300 NICs with modified firmware to extract fine-grained CSI features in both LOS and NLOS environments. Five activities(WALK, SIT, STAND, RUN and EMPTY) are recorded and analyzed. The first classification approach combined DWT for denoising, PCA for dimensionality reduction and a diverse feature set comprising spectral (e.g., Power Spectral Density and Haar wavelet

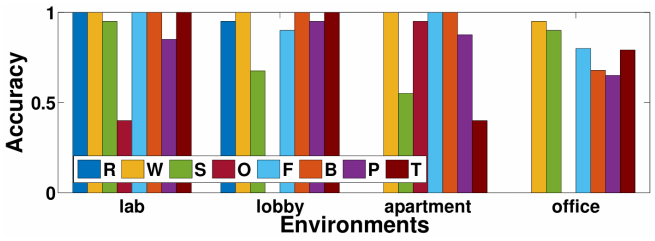


Fig. 2.16: Impact of Environments on Accuracy [8]

Note. Graph extracted directly from the study [8].

transforms) and statistical (e.g., skewness, kurtosis, moments) descriptors were calculated. The classification was then carried out using a linear Support Vector Machine (SVM) . The second approach employed Long Short-Term Memory (LSTM) networks with minimal preprocessing, demonstrating promising results despite its reduced complexity.

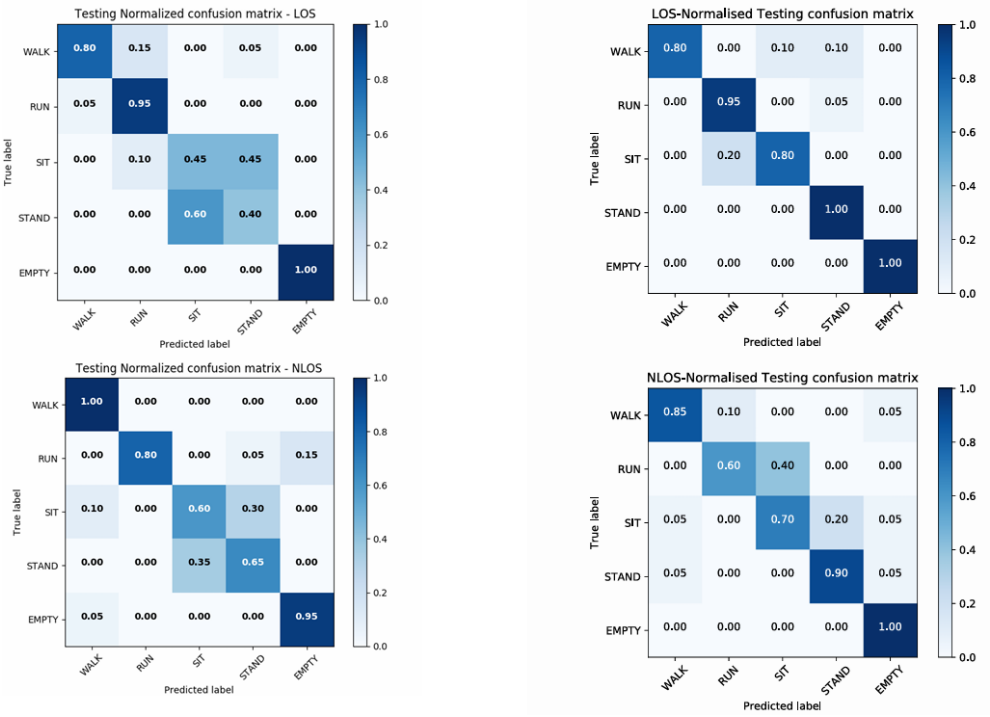
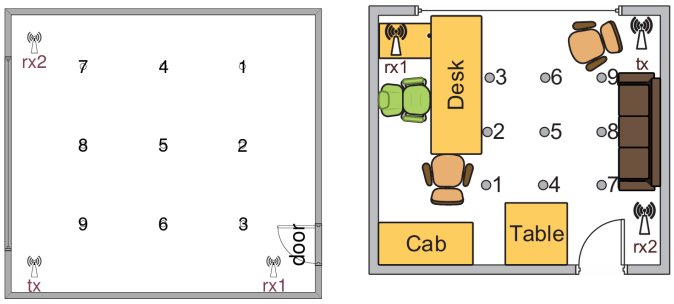


Fig. 2.17: Confusion matrices for SVM (left side) vs LSTM (right side) [9]

Their experiments showed that both SVM and LSTM performed well for high-level classification tasks such as presence detection and movement. SVM slightly outperformed LSTM in precision and recall under both LOS and NLOS conditions as shown in Fig. 2.17. However, LSTM's smaller preprocessing pipeline showed competitive performance. A key strength of this work is its thorough comparison of statistical and neural models. The paper also highlights practical concerns such as computational

overhead, training times and the effect of multipath-rich indoor environments. Importantly, it validates that CSI amplitude alone without phase can offer strong discriminative features for HAR. The adoption of both LOS and NLOS settings also contributes valuable insights for CSI-based deployment in variable conditions.

Another study, WiDir [10] introduces a device-free system for estimating human walking direction using CSI phase information as a critical context in CSI motion related sensing systems. Unlike conventional wearable or vision-based approaches, WiDir capitalizes on the multipath-induced phase shifts across subcarriers using a Fresnel zone model. It interprets subcarrier phase shifts through multi-frequency analysis to determine whether a subject is walking towards or away from a given receiver pair, termed as the "Fresnel direction" in the paper. To overcome the limitation of single-plane Fresnel which limits directional classification to just one axis (side-by-side), WiDir forms 2D Fresnel zones by deploying multiple receivers hence enabling it to compute local walking angles within Cartesian space. The study reports that this technique provides a sub-10° median absolute angular error in controlled indoor settings.



(a) Room A: an empty room (b) Room C: an office room

Fig. 2.18: Setup and Room Types used in [10]

Key takeaway from the study includes understanding the use of phase information for directional classification while the need for adaptive filtering to preserve CSI information integrity during denoising. WiDir demonstrated performance and robustness across different rooms as shown in Fig. 2.18. These findings showcase the potential use of phase information from CSI in directional motion analysis for this research.

2.3.4.2 Outdoor CSI systems

Recent efforts to use CSI for outdoors in a study [11] proposes a system that predicts outdoor human flow using Wi-Fi CSI with linear SVM. The study targets activity types including walking, running and cycling involving varying numbers of individuals passing a fixed channel as shown in Fig. 2.19 aiming to address the privacy-preserving limitations of vision-based systems. The study discuses on the custom antenna deployment strategy with a 1.5m separation and diagonal alignment to maximize the CSI cov-

erage field which could be ideal for a pedestrian crossing, as well as the derivation of eigenvalue-based features from amplitude and sanitized phase components as the feature set. Notably, features are extracted from auto-correlation and variance–covariance matrices of CSI across 3×3 MIMO links from their setup as shown Fig. 2.20 to capture and classify the temporal dynamics of the signal induced by humans walking between the channel.

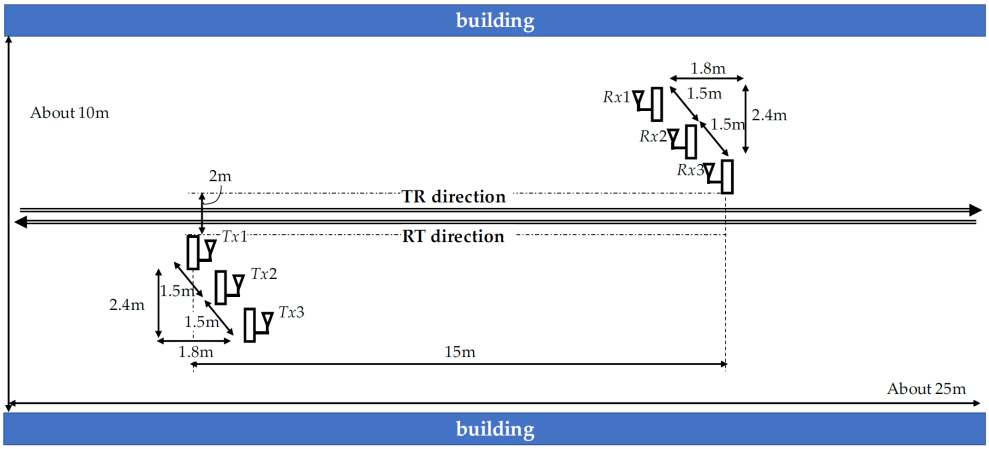


Fig. 2.19: Experimental Setup of [11] for the Study of Human flow detection

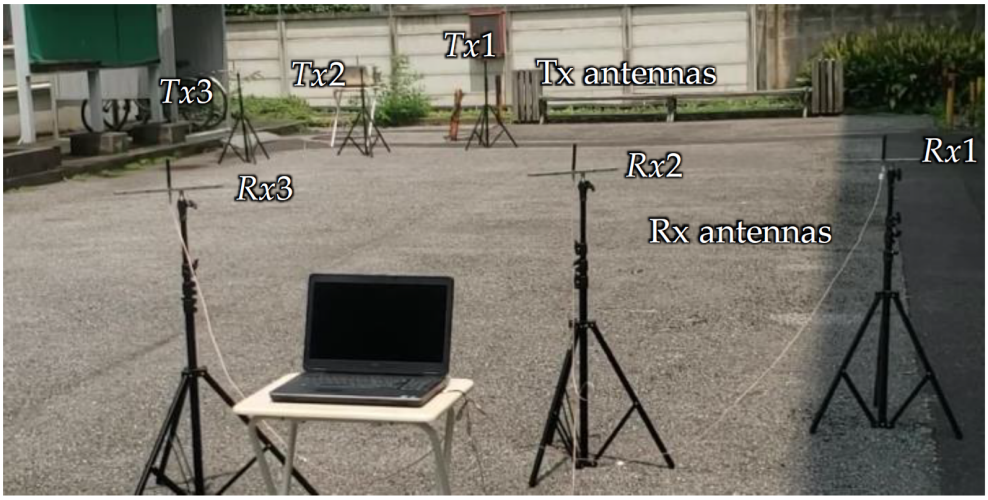


Fig. 2.20: Actual setup of [11] using 3x3 MIMO links to the Wi-Fi device connected to the Laptop

Their experimental results conducted along a campus road show a maximum classification accuracy of 100% for single-direction pedestrian flow and 99.5% when human traffic from both directions. This performance is especially relevant for applications such as smart pedestrian crossings and flow sensing in urban ITS settings. The study's methodology of using sanitized phase features and antenna placement for maximizing

signal coverage provides valuable insights that can be extended to directional pedestrian detection in more complex, uncontrolled outdoor environments.

However, despite the impressive reported results suggesting feasible CSI sensing in outdoor conditions, the deployment in Fig. 2.19 and Fig. 2.20 shows the setup situated between two buildings, likely enhancing signal multipath effects. Furthermore, the system's reliance on evaluating using a fixed set of participants within a single environment, limits the ability of the model to generalize to unseen data, highlighting a critical gap not addressed in this work.

Another interesting outdoor CSI study investigates the feasibility of CSI-based pedestrian monitoring in both indoor and outdoor environments [12], explicitly targeting detection of pedestrian presence and movement directions. Using an Intel 5300 NIC setup with three omnidirectional antennas, their system applies a multi-stage signal processing flow incorporating Hampel filtering, linear interpolation, Kalman smoothing and wavelet denoising to extract usable features from raw CSI signals. The setup of used for this study can be found in Fig. 2.21.

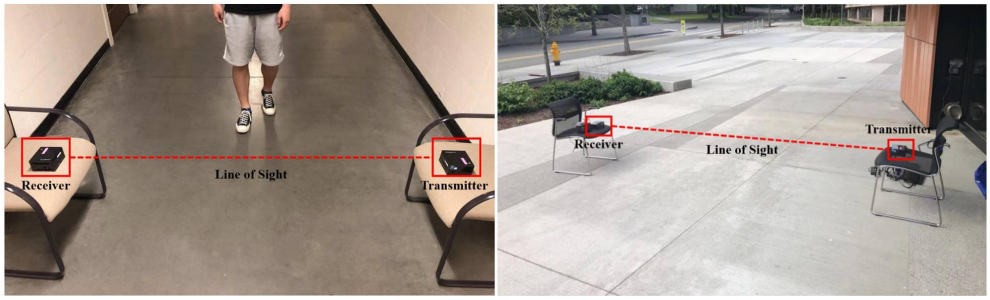


Fig. 2.21: Indoor and Outdoor Setup of the Study in [12]

The pedestrian presence detection is done using amplitude fluctuation values across 30 subcarriers, while directional classification is achieved using a Fresnel zone-based model to detect subcarrier phase shifts. Their experiments span both indoor lab spaces and an sidewalk testbed with a 3.5 metre LoS Tx-Rx pair, showing maximum detection accuracy of 99.23% indoors and 95.45% outdoors at a 100 Hz sampling rate.

Notably, their outdoor direction classification accuracy reached 92.21% and 93.51% for the two opposing movement directions. While the results are promising, the system is tested at a single location. As such, the robustness under real-world variability with diverse subjects coupled with environmental occlusions and cross-location validation remains unexplored. However, it establishes a baseline for the feasibility of applying CSI in outdoor settings while motivating further research across multi-subject and multi-environment scenarios.

A low-cost and modular radio tomography system for road user detection and classification proposed in [13] uses both Wi-Fi CSI and Ultra-Wideband (UWB) data. The system targets applications in ITS and V2X applications, specifically for detecting

cyclists and classifying vehicles using COTS devices (ESP32 and ARM Cortex-M3). Their setup involves a single-link configuration over a 4 metre cycle path and a 7 metre road segment, using directional antennas as shown in the hardware setup in Fig. 2.22 and embedded ML pipelines (ANN, RF and SVM) implemented on-device.

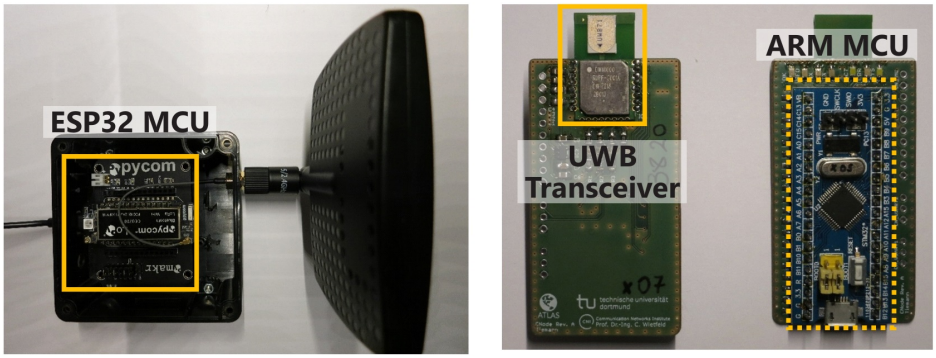


Fig. 2.22: WLAN CSI (left) and UWB (right) transceiver modules for radio tomography system used in [13]

Using statistical features derived from subcarrier amplitude and RSSI, the system achieves a binary bicycle detection accuracy of up to 100% and a multi-class classification accuracy exceeding 98% for vehicular differentiation. The focus of the study remains on distinguishing between idle, cyclist and motorized vehicles, leaving a gap in addressing pedestrian presence and movement. However, the study demonstrates the potential of using statistical features of amplitudes to achieve reliable accuracies on low-cost embedded platforms. Additionally, the comparative evaluation of multiple classifiers between Artificial Neural Networks, Random Forest (RF) and SVM offers a valuable reference for evaluating trade-offs, a practice that can be extended in this research for evaluating classifications under outdoor conditions.

A notable contribution to outdoor CSI-based pedestrian sensing is presented in [14], which proposes a device-free approach for estimating pedestrian counts at signalized crossings using amplitude and phase information from the CSI of ESP32 devices. Setup of the study is shown in Fig. 2.23.

Fig. 2.23: Environments of the Setup used in [14]

Unlike traditional push-buttons or vision-based systems, their study addresses privacy, cost and scalability concerns directly utilizing COTS hardware. The authors have

designed two Tx-Rx pair setup with ESP32 devices, transmitting UDP packets at 25 Hz of 51 usable OFDM subcarriers information.

The Amplitude and Phase information are preprocessed using techniques such as phase unwrapping, sanitization and filtering then fed into two deep learning architectures—MLP and 1D-ConvNet running on a Raspberry Pi 4B. The 1D-ConvNet achieved a counting accuracy of 79% for up to 12 pedestrians across multiple outdoor environments, with 95% of predictions within a ±2 error margin as shown in Fig. 2.24.

Fig. 2.24: Confusion Matrix of Cross-validation in [14] with 1D-ConvNet for pedestrian counting

This study is particularly relevant as it showcases the feasibility of real-time CSI processing in outdoor deployments using low-cost COTS hardware such ESP32 and Raspberry Pi. The CSI unwrapping, phase sanitization and edge inferencing techniques detailed in this study directly inform and motivate this research.

2.4 Summary and Benchmark

While significant progress has been made in vision-based [3, 4, 6, 58], RSSI-based [7, 54] and CSI-based [8, 10, 12, 14] HAR systems as summarized in Table. 2.2, each

study presents certain limitations in the context of real-time, cost effective outdoor pedestrian detection for ITS applications.

Vision-based systems, despite their high spatial and contextual resolution, face challenges related to privacy, occlusion, lighting condition requirements and high computational costs, limiting scalability in resource constraint settings [3, 6].

Basic pedestrian detection approaches such as laser beams, Passive Infrared (PIR) sensors and ultrasonic sensors can indicate presence of pedestrians crossing. However, PIR sensors are prone to high false positives, as they operate by detecting changes in infrared radiation (heat signatures). In environments like Sri Lanka, where ambient temperature fluctuations are frequent and significant, these sensors may mistake non-human heat sources or fail to detect pedestrians when background temperatures approach human body temperature [59, 60]. Ultrasonic sensors, which rely on reflected sound waves for detection, typically provide short-range, binary presence outputs and are highly sensitive to environmental factors such as wind, surface textures and ambient noise [61]. Laser-based systems, while offering more direct detection, raise regulatory and safety concerns related to eye safety and frequent exposure to skin, making them less practical for pedestrian detection applications.

RF based sensing options such RSSI and CSI systems offer low-cost, non-intrusive passive sensing but systems utilizing only RSSI lack the spatial information to capture directional or fine-grained motion characteristics, particularly in noisy outdoor conditions [7, 54]. In contrast to Vision, PIR and Ultrasonic sensing, Wi-Fi based approach is largely unaffected by environmental temperature variations or lighting conditions, as it relies on analyzing the distortions in Wi-Fi signals caused by human movement, rather than Line of sight changes, temperature or acoustic reflections.

However, Wi-Fi RSSI is a scalar measure of the total received power that represents the aggregated signal strength and lacks frequency-selective fading and phase information. This limitation constrains its ability to capture multipath dynamics, increasing susceptibility to false positives in varying environments, including shadowing effects hence reducing its effectiveness in inferring between pedestrian type, presence or direction especially in dynamic outdoor conditions. Meanwhile, CSI provides the complex channel response for each OFDM subcarrier, preserving amplitude and phase variations that encode Doppler and path-specific changes induced by pedestrian movement [54, 62, 63]. These characteristics offer a higher theoretical information ceiling for classification tasks such as pedestrian crossing detection while enabling further capabilities such as specialized alerts, system scalability and behavior learning for ITS.

Hence, CSI-based approaches show great promise in utilizing multipath effects and phase shifts for precise activity recognition, particularly indoors [8, 10]. Moreover, by demonstrating the use of CSI as a sensing system, opens the pathway for integrating customized alerts based on specific pedestrian types (e.g., wheelchairs, pets, children). Furthermore, the system's inherent compatibility of using Wi-Fi enables messaging

and cloud connectivity enabling integration with intelligent traffic management platforms proving the versatility for broader ITS aligning to the goals of ISAC.

However, many CSI related studies still rely on NICs like Intel 5300 which require bulky desktops or laptops setups and solely tested within controlled lab settings and rarely validated across diverse outdoor environments and subjects. Most promising study in [14] demonstrates the feasibility of using CSI from ESP32 devices in outdoor environments along with the aspects of edge deployment and domain generalization across locations for shared participants.

These gaps motivate the design of a lightweight, outdoor-ready CSI system using scalable hardware for ITS applications aiming for movement detection accuracies of over 95% to stay competitive to the state-of-the-art systems currently available.

TABLE 2.2: SUMMARY OF RELAVANT HAR STUDIES ACROSS VISION, RSSI AND CSI BASED SYSTEMS

Study	**HAR Type**	**Environment**	**Results(%)**
[3]	Vision	Intersections	> 90
[4]	Vision	Crossings	–
[58]	Vision	Crossings	–
[6]	LiDAR Vision	Intersections	–
[59]	PIR - Detection	Crossings	up to 100
[7]	RSSI	Indoor	95.5
[9]	CSI	Indoor	>90
[8]	CSI	Indoor	96.1
[11]	CSI - Flow detection	Sidewalks	99.5–100
[12]	CSI - Flow detection	Sidewalks	92–99
[13]	CSI - Counting	Intersections	>98
[14]	CSI - Counting	Crossings	up to 95

2.5 Feature Engineering Trends in CSI-Based Sensing

Wi-Fi CSI research has categorized the features into well defined groups that capture how human motion creates disturbances in the channel which is reflected in the CSI frames. Comprehensive studies and surveys have summarized the specific features used in vast studies and how are engineered in [24, 55–57]. Consequently, the system's classification performance is deeply rooted by the quality of its input features.

Statistical feature engineering in studies [9, 11, 64, 65] have demonstrated that simple descriptors such as mean, variance, peak-to-peak amplitude and Doppler spreads over fixed windows can achieve reliable performance across various human activity recognition tasks.

Beyond statistical feature sets, more complex and custom feature engineering approaches have emerged on capturing motion related information from the CSI se-

quences as a frequency-domain representations such as Short-Time Fourier Transform (STFT) and Wavelet Transforms. These methods have shown to isolate motion signatures effectively at different scales [8, 9] often resulting with better accuracies. Additionally, studies such as [10] introduce Fresnel zone-based models, capturing the phase shifts across subcarriers to estimate the walking direction, while others studies such as [11, 65] explore eigenvalue based features and Doppler spread.

These custom feature engineering methods are often computationally heavier but contribute to enhancing directionality and robustness to multipath effects for motion detection and analysis.However, due to nature of low computational cost and simple extraction of the statistical based features used in [64], they can be considered to be particularly well-suited for real-time sensing scenarios and resource-constrained deployments such as smart pedestrian crossings.

2.6 Enabling Technologies

Use of various innovative technologies to achieve efficiency for sensing, processing, control and communicate, forms the foundation of ITS applications. Developing sustainable methods to utilize technologies that could behave as integrated sensing and communication together plays a vital role in V2X and ITS applications [66].

2.6.1 Wireless Fidelity (Wi-Fi)

Wi-Fi refers to the family of wireless communication protocols defined under the IEEE 802.11 standard, which covers multiple generations of updates aimed at improving throughput, coverage and spectral efficiency. As summarized in Table 2.3, these standards operate across the 2.4GHz, 5GHz and the more recent 6GHz frequency bands. In most regions, Wi-Fi consists of 11 overlapping channels in the 2.4,GHz band, although this number can extend to 13 or 14 depending on policies locally. Wi-Fi transmission rates improves significantly each generation, ranging from as low as 2 Mbps in legacy systems to over 45 Gbps in modern generations such as Wi-Fi 7 with outdoor communication ranges in open spaces typically between 100 to 250 meters [67].

Beyond its widespread use for wireless communication, Wi-Fi has become a powerful medium for passive sensing due to its cost efficiency and access to the physical layers [68]. This illustration of signal overlaps shown in Fig. 2.25 demonstrates the potential congestion in the 2.4 GHz Wi-Fi spectrum which can lead to significant inter-channel interference, affecting signal quality and stability that should be factored when collecting CSI for movement sensing applications in uncontrolled outdoor settings. The sharp increase of Wi-Fi-capable embedded devices and the availability of open-source firmwares have enabled fine-grained signal analysis, including the extraction of CSI frames. These properties position Wi-Fi as a viable alternative for device-

TABLE 2.3: EVOLUTION OF IEEE 802.11 WI-FI STANDARDS

Wi-Fi Generation	Adopted Year	Frequency Bands (GHz)	Max Data Rate
Wi-Fi 0 - (802.11)	1997	2.4	2 Mbps
Wi-Fi 1 - (802.11b)	1999	2.4	11 Mbps
Wi-Fi 2 - (802.11a)	1999	5	54 Mbps
Wi-Fi 3 - (802.11g)	2003	2.4	54 Mbps
Wi-Fi 4 - (802.11n)	2009	2.4 / 5	600 Mbps
Wi-Fi 5 - (802.11ac)	2013	5	6.9 Gbps
Wi-Fi 6 - (802.11ax)	2019	2.4 / 5	9.6 Gbps
Wi-Fi 6E - (802.11ax)	2021	6	9.6 Gbps
Wi-Fi 7 - (802.11be)	2024	2.4 / 5 / 6	46.1 Gbps

Note. Information collated from IEEE standards [67]

free human activity recognition [69], particularly in applications where privacy, cost and deployment scale are critical concerns.

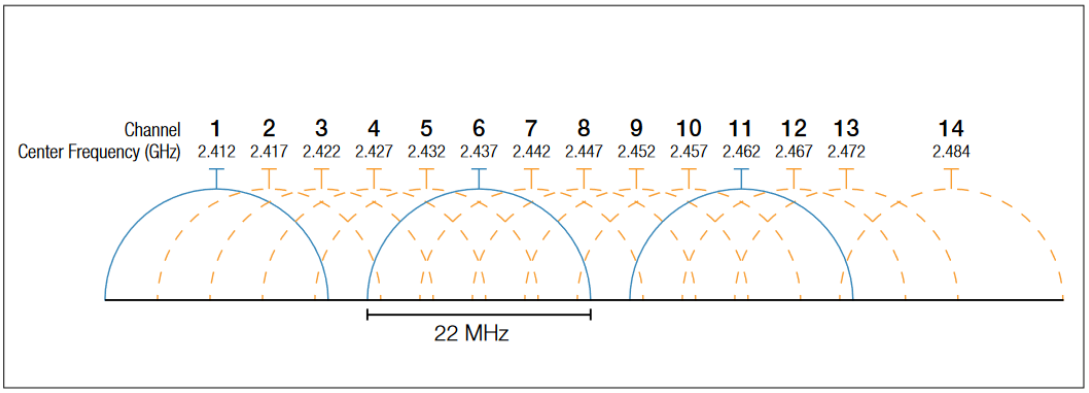


Fig. 2.25: Spectrum view of 2.4GHz bands divided into overlapping channels [15]

2.6.2 Orthogonal Frequency-Division Multiplexing

Orthogonal Frequency-Division Multiplexing (OFDM) is the core modulation technique adopted in modern Wi-Fi standards including IEEE 802.11a/g/n/ac and ax [67]. It enables high-speed and robust wireless communication by dividing a high-rate data stream into multiple lower-rate parallel streams, each carried over as its own subcarrier. These subcarriers are orthogonally spaced in the frequency domain, ensuring their transmissions do not interfere with each other. This approach significantly improves spectral efficiency while offering strong resilience against inter-symbol interference and frequency-selective fading, which are commonly observed in practical wireless communications.

The OFDM signals are formed in the frequency domain, with each subcarrier corresponding to a unique frequency bin specified in the standard. This composite signal is then transformed into the time domain using the Inverse Fast Fourier Transform (IFFT) enables simultaneous transmission of all subcarriers. The orthogonality of the subcarriers ensures that the peak of each one aligns with the zero-crossings of all others, eliminating inter-carrier interference and preserving signal clarity [16].

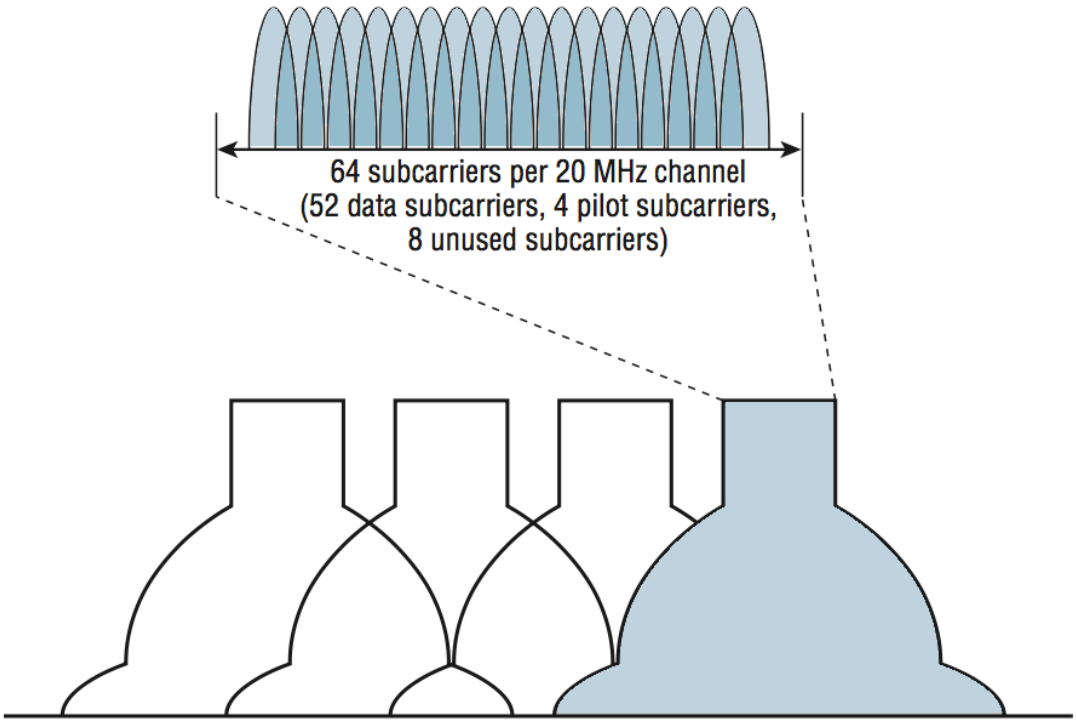


Fig. 2.26: Orthogonal subcarrier structure of an OFDM signal [16]

As illustrated in Fig. 2.26, in a standard 20 MHz Wi-Fi channel, OFDM defines 64 subcarriers, out of which 52 are active—48 are used for data transmission and 4 are reserved as pilot subcarriers. These pilot tones carry known reference symbols that allow the receiver to estimate and equalize the channel, forming the basis for Channel State Information (CSI) extraction. The modulation scheme used on each subcarrier can vary depending on Wi-Fi configurations and may include BPSK, QPSK, or QAM formats [70].

In the context of this research, OFDM serves as a enabler for CSI-based sensing. Human movement in the channel introduces variations in the multipath components captured within the OFDM signals, resulting in changes on the subcarriers components which can be measured to demonstrate the signal's experience in the channel with the CSI frames. Thus, providing the granularity required to classify activities in real-time compared to scaler measurements such as received signal strength indicator.

2.6.3 Channel State Information (CSI)

CSI refers to the fine-grained measurements captured in the wireless signal's channel frequency response at the receiver for transmitted signals across subcarriers and antennas for each data frame. Enabled by OFDM this information reflects the signal's experiences from the environment due to the principles such as reflection, diffraction and scattering caused by dynamic objects such as humans movements and objects [71].

CSI is generated at the receiver for each data frame is first amplified, sampled and processed using Fast Fourier Transform (FFT) to convert the data from time domain into frequency domain. The Wi-Fi module then estimates the Channel Frequency Response (CFR) which effectively becomes the CSI frames grouping into an information matrix. This collection of CSI frames encapsulates information such as the Amplitude and Phase across the subcarriers of each antenna providing a "snapshot" of the channel for each frame over time.

The CFR is depicted with the following equation[55]:

$$H(f,t) = \sum_{i=1}^{N} a_i(t) e^{-j2\pi f \tau_i(t)} \tag{2.1}$$

where $a_i(t)$ and $\tau_i(t)$ denote the amplitude attenuation and propagation delay of the i-th multipath component at time t and f denotes subcarrier frequency.

The collection of CSI matrix can be depicted with the equation:

$$y(f,t) = H(f,t) \cdot x(f,t) + n \tag{2.2}$$

where y is the received signal of the transmitted signal x where CSI matrix is generated for H along with some noise vector n. A simple illustration can be found in Fig. 2.27.

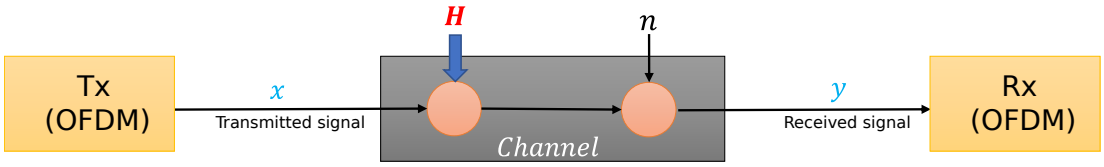


Fig. 2.27: Illustration on the effects of H (CSI) in signal propagation through a wireless Channel

Given that CSI is captured as a pair of complex numbers [55, 72] at H in Eq. 2.1 and Eq. 2.2, the amplitude $A(f,t)$ and phase $\phi(f,t)$ information can be calculated by:

$$A(f,t) = |H(f,t)| \quad , \quad \phi(f,t) = \angle H(f,t) \tag{2.3}$$

Amplitude and Phase information form the fundamental features for a HAR system across each subcarrier. From these, more sophisticated features can be derived to

address complex sensing and recognition classification challenges.

2.7 Implementation Platforms

2.7.1 Commercially available Wi-Fi CSI capable devices

A literature survey was conducted to identify COTS Wi-Fi devices capable of CSI extraction for sensing and classification applications.

2.7.1.1 Intel 5300

Intel's 5300 Network Interface Card (NIC) is widely used in CSI based applications [24, 55]. It has been foundational in early researches on Wi-Fi CSI-based sensing due to its ability to extract raw physical layer Channel State Information from up to 30 subcarriers (effective 90 CSI frames) across three antennas as shown in Fig. 2.28 using custom firmware such as the Linux 802.11n CSI Tool [73].

Its support for IEEE 802.11a/b/g/n and MIMO capabilities [74] as described in Table 2.4 makes it ideal for a wide range of human activity recognition experiments, particularly for studying spatial diversity and fine-grained multipath propagation effects. Despite its widespread adoption in in popular research such as SpotFi [48] and CARM [8]for Indoor Localization, the 5300 NIC has considerable limitations, including lack of support for newer Wi-Fi standards, large form factor as it needs to be accompanied with a desktop or laptop with a compatible PCIe slot.

Fig. 2.28: Intel 5300 NIC with its extended 3-Antenna setup using an adapter [17]

However, Intel has discontinued support and production of these devices since 2016 [75]. As such, it remains a valuable tool for experimentation but its legacy nature and end of life motivates the shift towards more modern and low-cost alternatives.

TABLE 2.4: INTEL WIFI LINK 5300 SPECIFICATIONS

Feature	Description
Bandwidth	Up to 450 Mbps of bandwidth
Range	Up to 2× greater range
Antenna Configuration	MIMO and support for up to three antennas
Standards Compliance	IEEE 802.11a/b/g and n compliant
PAN Capabilities	WiFi Personal Area Network capabilities
Security	Advanced security via 802.11i
Software	Intel® PROSet v12.0 WLAN Software
Compatibility	PCI-e and Support for Cisco Compatible Extensions* v4

2.7.1.2 Atheros Chipsets

Studies related to [73, 76, 77] describe the usage of Atheros NIC namely the models AR9380 and AR9390. The Atheros family of NICs are viable alternatives to the Intel 5300 NIC. A key advantage of Atheros NICs is their compatibility with more recent Wi-Fi standards (e.g., 802.11ac), depending on the chipset from the family and their ability to be deployed on lower-power, router hardware. However, these devices generally offer fewer subcarriers than the Intel 5300 (e.g., 56 for Atheros versus 90 for Intel 5300) as shown in the comparison Table 2.6 and are often limited to single-input single-output (SISO) or dual-antenna configurations, affecting spatial resolution [71]. The most commonly used Atheros based routers includes the TP-LinkWDR as seen in Fig. 2.29 however they are currently in the End of Life (EoL) status.

Fig. 2.29: TL-WDR3600 Router containing a Atheros NIC [18]

Despite these constraints, their successful use in HAR applications and localization tasks particularly in distributed edge setups [78] makes Atheros NICs a compelling option for CSI-based sensing applications however, they pose certain degree of difficulties for outdoor deployments due to general availability for indoor COTS routers and specialized circuitry along with a desktop or a laptop to enable the use of the NIC.

2.7.1.3 Espressif Systems 32 (ESP32)

ESP32 System-on-Chip (SoC) shown in Fig. 2.30 developed by *Espressif Systems* has been gaining popularity in recent CSI-based sensing researches due to its affordability, compact form factor and open access to physical layer (PHY) Channel State Information via official SDKs [79] and third-party tools [80]. Unlike NIC based systems that require a desktop to function, the ESP32 operates as a standalone platform capable of both transmitting and receiving Wi-Fi packets while logging CSI data directly and extend its capabilities to function fully standalone as suggested in [81] to run classifiers onboard.

Fig. 2.30: ESP32-WROOM-32 SoC module

CSI extraction on ESP32-WROOM-32 leverages the `HT-LTF` portion of IEEE 802.11n packets, supported by Espressif's "esp-csi" framework and community-developed alternatives tools. These tools enable access to CSI amplitude information across 64 subcarriers, sampled at rates up to 1000 Hz, depending on serial port configuration, network and environmental conditions.

The function block of the ESP32 is shown in Fig. 2.31 which higlights its built-in Bluetooth and Wi-Fi capabilities. However, in the context of CSI extraction, the ESP32 lacks MIMO support as most modules operate with a single antenna. While this limits the fine-grained subcarrier information obtainable in the likes of NICs such as Intel 5300 or Atheros, which support multiple antennas and broader subcarrier access by design, ESP32 offers significant advantages in terms of ease of deployment, cost and portability. These factors make it a practical alternative for device-free sensing, especially in ITS applications. The platform has seen increasing adoption in HAR, gesture recognition and low-latency edge inference tasks across both indoor and outdoor environments[24, 55–57, 82].

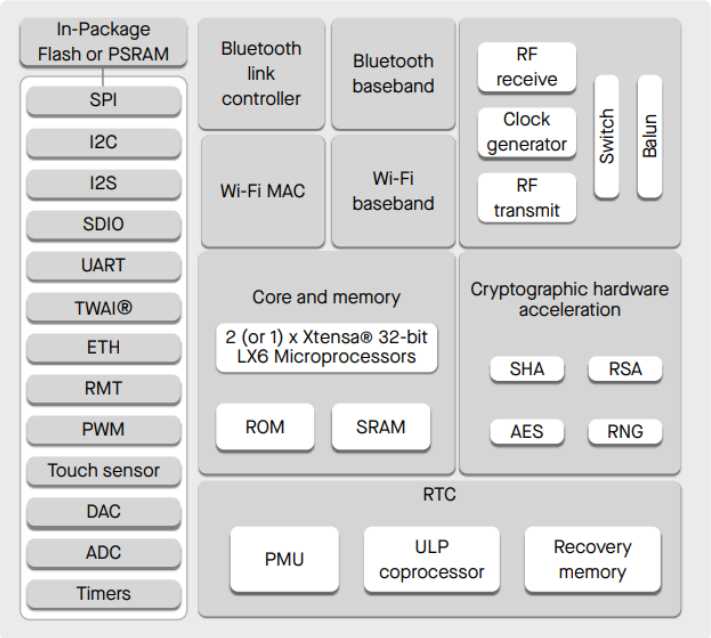


Fig. 2.31: ESP32 Function block [19]

TABLE 2.5: ESP32-WROOM-32 SPECIFICATIONS

Feature	Description
Chipset	ESP32-D0WDQ6 dual-core SoC (Xtensa LX6 @ 240 MHz)
Wireless Standards	IEEE 802.11 b/g/n (2.4 GHz only)
Maximum Data Rate	150 Mbps (802.11n, 20 MHz bandwidth)
CSI Access	HT-LTF based CSI logging (via third-party tools)
Antenna Options	PCB or external antenna
Interfaces	UART, SPI, I2C, GPIO and more
Flash	4MB embedded flash (WROOM-32 module)
Operating Voltage	MCU at 3.3 V while the SoC operates at 5V
Power Supply	USB or external supply via LDO
Form Factor	18 mm × 25.5 mm (module), approx. 10g

Note. Data extracted from [19]

2.7.1.4 Raspberry Pi

The Raspberry Pi is a low-cost, single-board computer widely used in research and prototyping in embedded and edge computing applications [83, 84]. Its compact form factor, Linux operating system and diverse Input and Output (I/O) make it a feasible platform for deploying lightweight sensing and processing tasks.

Raspberry Pi offers various models from ultra compact boards such as the Pi Pico, Zero and Nano to generational variants such as the Pi 3B, Pi 4, 4B and the recently introduced Pi 5. Raspberry Pi systems are able to function fully standalone and with the recent release Pi5 shown in Fig. 2.32 and its inclusion of a onboard GPU, the

TABLE 2.6: COMPARISON OF CSI-CAPABLE WIRELESS DEVICES

Feature	INTEL 5300	ATHEROS	ESP32
Commonality	Very	Moderate	Very
Standalone	No	No	Yes
Dimension L×B (cm)	30×20	30×20	5.0×3.0
Weight	∼1.0kg	∼1.0kg	∼10g
Power Supply	Specialized	Specialized	USB Powered
Cost	$20 / 3500 LKR	$̃20 / >3500 LKR	$6 / 1200 LKR
No. of Subcarriers	90	56	52
Resolution	8 bit	10 bit	8 bit
Implementation	Kernel Level	Kernel Level	User
Sampling Rate	Up to 1000 Hz	Up to 1000 Hz	Up to 1000 Hz
Antennas	3	Up to 3	1
Complexity	Very	Very	Relatively Simple

single-board computer's functionality can be utilized for complex computations.

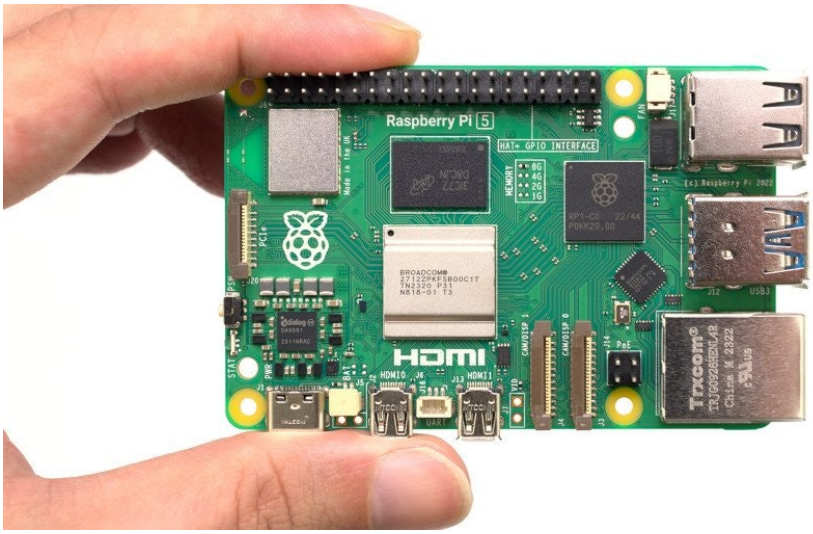


Fig. 2.32: Raspberry Pi 5 shown with a hand for scale

Although the Pi 3B can be functional it is limited by its Universal Serial Bus (USB) 2.0 ports, which operate on a shared bus architecture, causing bottlenecks which became apparent during high-speed serial communication when packet loss rates were observed. In contrast, the 4B being the 4th Generation of PI includes two USB 3.0 ports as compared in Fig. 2.33 with a dedicated bus that significantly improves data bandwidth and the reliability for serial logging at higher serial baud rates. This setup minimizes the chances of CSI frame loss at the Universal Asynchronous Receiver/Transmitter (UART) communication link and ensures reliable logging and inferencing during real-time data operations.

In addition, the 4B offers enhanced CPU and RAM, which are beneficial for exe-

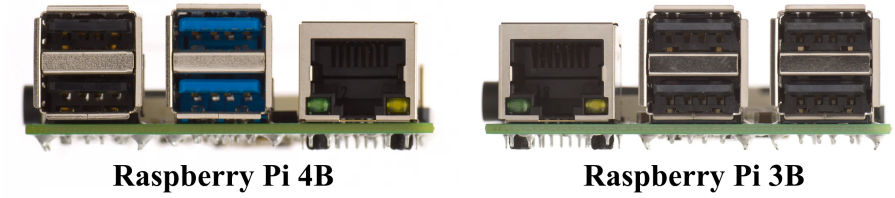


Fig. 2.33: USB-Ports of Raspberry Pi 4B vs 3B

cuting preprocessing and machine learning systems completely onboard. A full comparison between Raspberry Pi 3B and 4 is available at [85]. The Pi's widespread availability, computational capabilities and USB-C power assists to a compact form factor for the deployment. The release of the Raspberry Pi 5 improves on the capabilities [86] from its previous generation, the features of 4B was able to meet the research requirements as such an upgrade to Model Pi 5 for this research had no significant advantage.

2.7.1.5 Summary on Implementation Platforms

While Raspberry Pi devices have recently gained support and traction for CSI extraction for HAR applications through the Nexmon tool [87, 88], this research uses an ESP32-WROOM-32 based sensing system where the ESP32 functions as a low-cost sensing unit capable to be integrated into a RSU coupled with a Raspberry Pi that serves as the edge processing unit for classification tasks with CSI information for the availability, implementation, cost and performance balanced reasons for the platform. This design separates data collection from the relatively intensive computations for classification by modularizing the system and allowing capacity for future upgrades for V2X alert generation and cooperative ITS applications.

2.8 Classification Models

The objective of detecting activity or presence of human movements in the Wi-fi channel requires to have a classification model capable of interpreting these induced signal variations. These tasks can be considered a multi-class classification problem where the model can effectively learn from the complex patterns embedded within complex subcarriers of CSI frames. Two popular categories for such classification models are Statistical Models and Neural Networks.

2.8.1 Statistical Models

Statistical models have remained a popular choice in CSI-based HAR systems due to their relatively lower computational costs and training times. These models rely on statistical features such as mean, variance, skewness or Doppler spreads over a fixed length window to learn decision boundaries between each activity class.

2.8.1.1 Random Forests (RF)

Random Forests are a family of statistical learning models (ensemble method) that construct multiple decision trees during training and outputs the class that obtains the majority vote of the predictions across all the trees in the model as illustrated in Fig. 2.34. Their robustness to overfitting due to the binary decision making and ability to handle noisy feature sets make them particularly suitable for CSI data where noisy environmental data variations can be quite common.

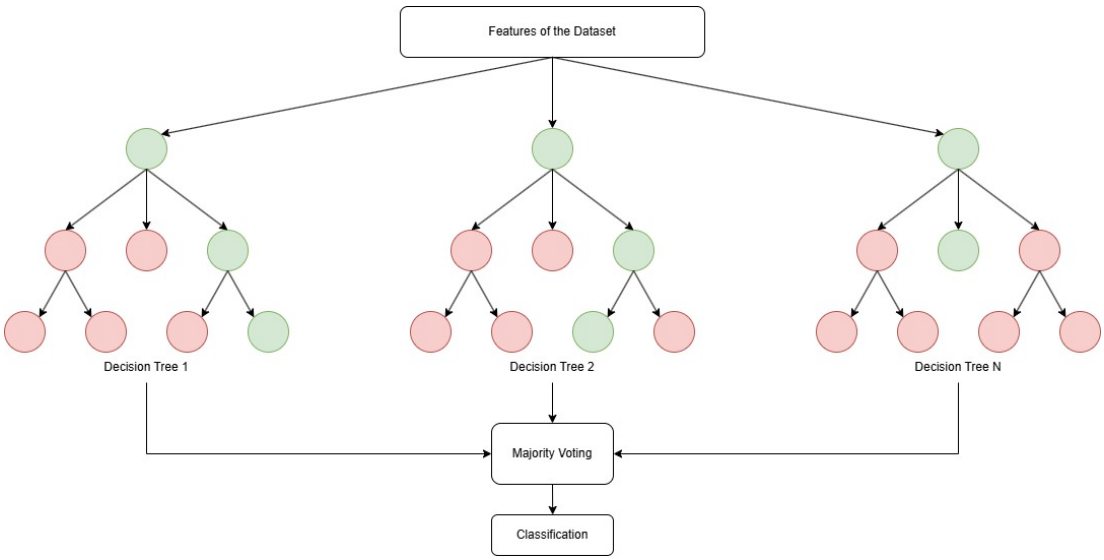


Fig. 2.34: Illustration of a Random Forest Decision Tree

Studies such as [13] have used RF classifiers on features derived from CSI, UWB amplitudes and RSSI for their classification tasks. The approach was implemented on embedded hardware and achieved high performance for distinguishing between bicycles and vehicles. Similarly, [65] employed RF as part of a comparative evaluation with SVM and a neural network for CSI-based gait recognition demonstrating its competitive accuracy with significantly lower resource requirements.

2.8.1.2 Support Vector Machines (SVM)

Support Vector Machines are another set of satistical models that operate by finding the optimal hyperplane which separates data points belonging to different classes in

the high-dimensional feature space as illustrated in the Fig 2.35. Several mathematical functions for statistical analysis are available known as "kernels" which allows SVMs to effectively learn non-linear decision boundaries, making them powerful tools for CSI-based classification tasks involving complex movement signatures.

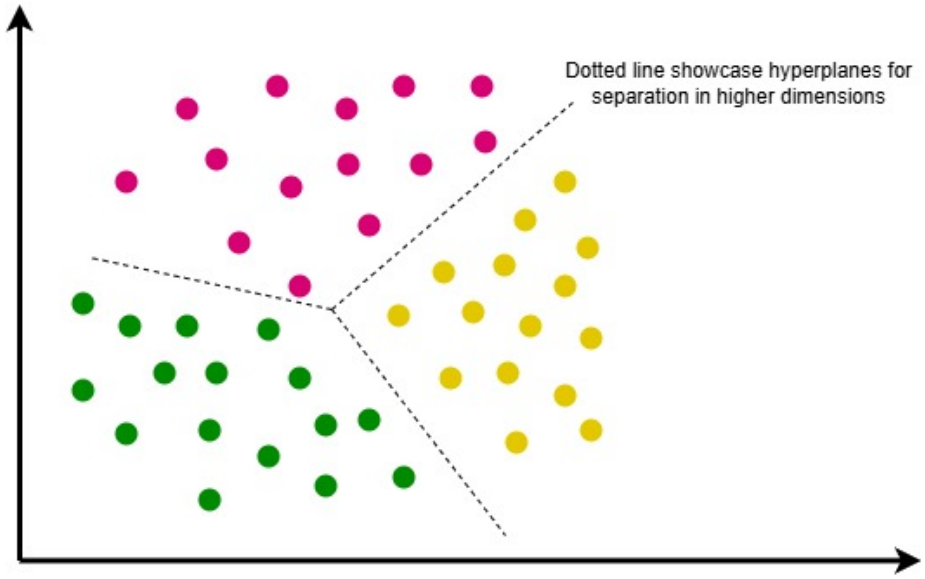


Fig. 2.35: Illustration of a Support Vector Machine with its hyperplane separation

SVMs have been widely adopted in CSI sensing similar to Random forests. For instance, [11] utilizes a SVM to classify outdoor flow of humans from a summarized feature set of eigenvalues over fixed windows from the statistical information of the amplitude and phase data. In [9], a linear SVM outperformed the LSTM model in classifying five indoor activities using these statistical features. The WiHACS study [64] also implemented SVM classification over PCA reduced and energy-based amplitude features, showing strong recognition even through walls.

2.8.2 Neural Networks

To overcome the limitations of the capabilities of the statistical models in interpreting intricate patterns in data, the concept of Neural Networks was introduced as a way of mathematical approximations inspired by the human brain. These models simulate networks of interconnected hidden neurons identical to the functioning of a human brain and can progressively learn relationships across wide variations of inputs. This makes them effective for classifying non-linear, high-dimensional information such as the ones present in CSI data.

Hence, their ability to generalize patterns and learn meaningful information from noisy data has made neural networks increasingly popular for tasks such as movement and gesture detection in HAR systems.

2.8.2.1 Convolutional Neural Networks

Convolutional Neural Networks (CNN) are widely used in applications such as computer and vision-based recognition system due to the capability to process high resolution information. A simple CNN consisting of a input layer, hidden layer and a output layer passed with a fully connected layer as illustrated in Fig. 2.36 can process 2D inputs and compute spatial relationship between the data points [3]. The study in [14] utilizes a four layer stacked CNN architecture to achieve results up to 93.9% in counting pedestrians waiting at a crossing. Other related studies have suggested that CNN layers are capable of extracting features from their input and become input layers that feed features to subsequent layers creating deeper neural networks.

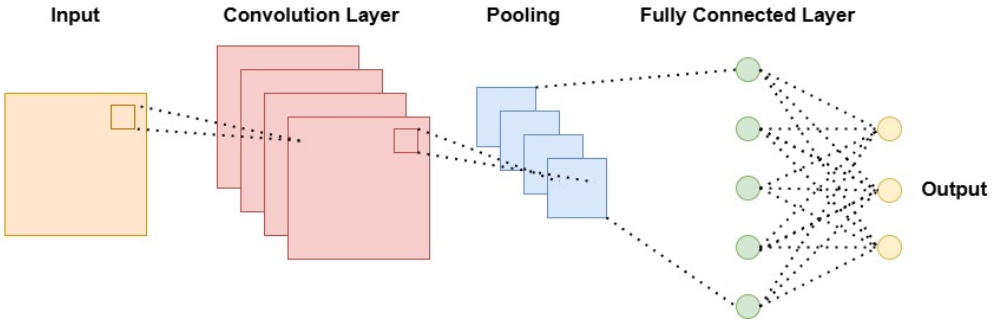


Fig. 2.36: Illustration of a Convluation Neural Network processing of an image

CNNs are a specialized subset of neural networks from a wide group of models such as Artificial Neural Networks (ANN) and Deep Neural Networks (DNN) . ANNs are often considered as the building blocks of a neural network system typically containing a simple single hidden layer architecture[13]. An increase to the count of hidden layers however improves the capabilities of the neural network creating DNNs. These are more commonly found in studies for HAR system, such as in [27, 29] which utilizes complex architectures for counting or positioning people indoors and outdoors due to the intricate nature of the CSI information.

2.8.2.2 Long Short-Term Memory

Long Short-Term Memory (LSTM) networks are a subset of Recurrent Neural Networks (RNN) designed to capture temporal patterns in sequential data while overcoming the vanishing gradient problem common in traditional RNNs [20]. LSTM introduces memory cells as gates such as the input gate, forget gate and output gate which regulates the flow of information within the neurons, allowing the model to selectively remember or forget information over time. An illustration of LSTM architecture can be found in Fig. 2.37. In Wi-Fi CSI-based sensing tasks, LSTMs are particularly effective for modeling human movements over time, as they can capture evolving temporal patterns within amplitude and phase sequences. Studies [8, 9, 82] have recorded the

use of these LSTMs and shown that they are prominent for time-series detection of activities such as walking.

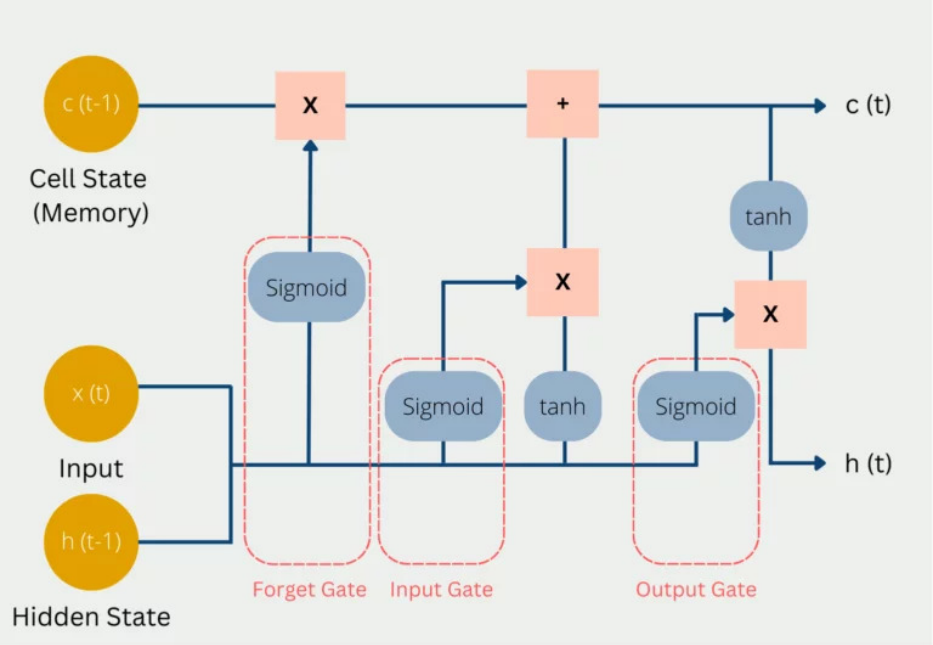


Fig. 2.37: Illustration of a Long Short Term Memory network with Gates from [20]

2.8.2.3 Bi-directional LSTM and Convolutional-LSTM

Bi-directional LSTM (BiLSTM) networks extends the capability of the LSTM by processing the temporal input sequence in both forwards and backwards directions. This enables the neural network to capture intricate patterns in data from both directions making it especially useful in scenarios where understanding the full sequence is critical. By leveraging complete temporal information, BiLSTMs can often outperform unidirectional models in classification accuracy for extreme sequential tasks [78].

Similarly, hybrid approaches that combine two or more types of neural networks such as using CNN-RNN and Convolutional-LSTM (ConvLSTM) architectures, provides the spatial feature extraction capabilities of CNNs with the temporal sequence recognition capabilities of RNNs. One closely related study that demonstrates the hybrid architecture of integrating a spatio-temporal model available in [71]. Such hybrid architecture can be well-suited for CSI applications, where both spatial patterns across subcarriers and temporal evolution of signal serves as meaningful indicators for accurate movement and direction detection.

2.9 ETSI Guidelines for VRU Detection and Reaction Timing

As this research focuses on real-time pedestrian detection for ITS applications, it is important to understand performance benchmarks and regulatory conditions for deployments. The report published by the European Telecommunications Standards Institute (ETSI) [89] outlines a performance guideline for VRU detection systems for ITS applications. It specifies an end-to-end latency budget of approximately 1.75 seconds, which includes pedestrian detection within the critical zone (e.g. the crossings), risk evaluation, message transmission and driver reflex time as illustrated in Fig. 2.38.

These timing constraints emphasize the need for efficient sensing and classification system, such as the utilizing CSI proposed in this work and setting a guideline for a lightweight system able to sense and prepare alert while still allowing sufficient time for the driver to react. This type of alerting is covered under VRU Awareness Message (VAM) within the ETSI TS standards.

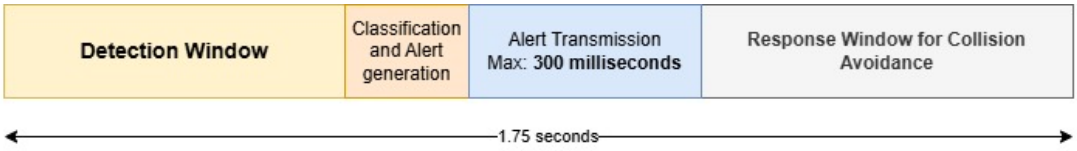


Fig. 2.38: Illustrated End-to-End Time Window for VRU applications defined by ETSI

CHAPTER 3

METHODOLOGY AND IMPLEMENTATION

Following the review of CSI based sensing and classification systems for device-free non-intrusive HAR and ITS application, this chapter describes the research methodology and implementation adopted. The system architecture proposed for this research to create a smart pedestrian crossing system is shown in Fig. 3.1.

The methodology primarily focuses on the CSI extraction, processing of CSI data and implementing a robust classification model for pedestrian movement detection that can couple with various output applications. Potential output applications include, Smart Phone popups for drivers within the vicinity, in-vehicle alerts such as dashboard notifications and alarms as well as display notifications on Active Road Signs.

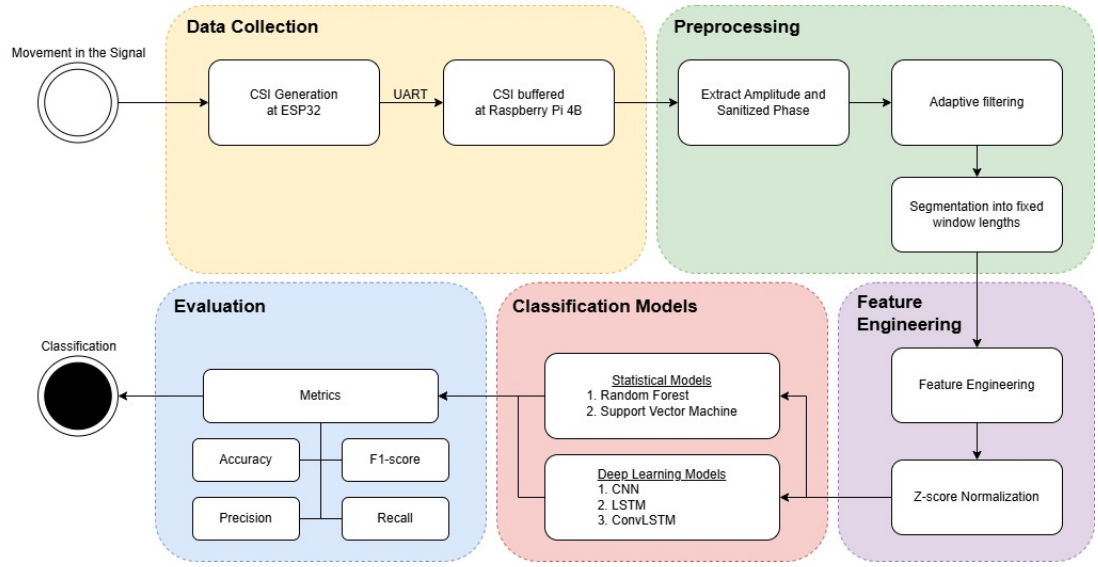


Fig. 3.1: Illustrated overview of the research methodology for implementing a smart pedestrian crossing

3.1 CSI Extraction using ESP32

3.1.1 ESP32-CSI-Tool

Extraction of CSI frames from the physical layer of the ESP32 is possible via tools available in [79–81]. All three tools effectively recommend similar approach and provide detailed guides to ensure a quick start-up and experimentation. For this research, Steven's ESP32-CSI-Tool [81] was selected due to the convenient startup guide and provisioned monitoring scripts for examining CSI frames live for easier experimentation. Open-source MinGW32 compiler allowed access to the CSI-tool's SDK to

configure the ESP32 device as specified in the toolkit.

The device configuration process involved flashing a ESP32 with the access point firmware provided within the tool to enable extraction and transmission of CSI frames via the serial port of the device. Within the same configuration window are the parameters to setup the SSID and password of the Wi-fi access point, allowing transmitter devices (STA) to connect and initiate CSI generation. This configuration allows potential MIMO setup where multiple STAs can connect to a single ESP32 AP. Once basic configurations are completed, the ESP32 device will start generating CSI data at the AP and begin transmission via the UART interface once a connection is established between the STA-AP devices.

3.1.2 Sampling Rate

To achieve higher sampling rates the UART baud rates for the communication needs to be set higher. The default baud-rate in the SDK is set to 115200 which potentially limits the throughout required.

Baud-rate is defined as the number of bits transmitted per second. Below equation is a general representation of the calculation.

$$R = \frac{B}{10} \tag{3.1}$$

where B is the UART baud rate in bits per second (bps) and R is the resulting throughput in bytes per second as a fraction of 10 due to the overhead of UART communication using 8N1 format (start and stop bit combined with 8bits of data) between ESP32 and Raspberry Pi.

Assuming CSI packets arrive at a frequency f (in Hz) and each packet has a size of S bytes, the required data rate is:

$$R_{\text{req}} = f \times S \tag{3.2}$$

Substituting Equation (3.1) into Equation (3.2), the minimum baud rate required is:

$$B_{\text{min}} = 10 \times f \times S \tag{3.3}$$

and if each CSI frame is approximately $S = 204$ bytes assuming 52 usable subcarriers represented with 16 bit signed integer where each subcarier is represented as a real and imaginary pair while the expected CSI sampling frequency, $f = 100$Hz then:

$$B_{\text{min}} = 10 \times 100 \times 204 = 204,000bps \tag{3.4}$$

Since the default baud rate of 115200 bps seems insufficient as per the results

obtained in Eq. 3.4 as this will create potential loss of CSI frames from being logged. The highest baud rate of 921600 bps is selected to ensure maximum buffer for UART transmission and reliable data logging.

The CPU operational frequency of ESP32 is set to its maximum value of 240Mhz to enable maximum support. While the ESP32 can support up to 52 usable subcarriers (excluding null and pilot at HT40 [90]), only 26 alternative subcarriers reliably logged during preliminary testing. This selective logging was not only a result of hardware constraints but also supports the trade-off to maintain a lightweight, resource-efficient architecture.

3.2 Platform Setup and System Configurations

3.2.1 Field Setup

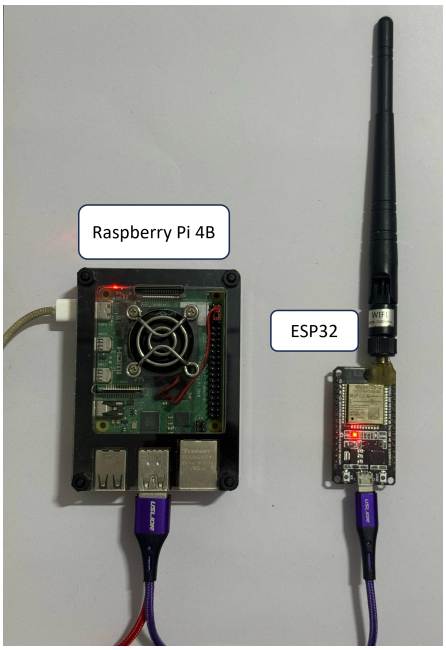
The ESP32 Transmitter-Receiver (Tx-Rx) pair was mounted to respective standard issue tripods using double-sided tape and zip ties. This was intentional to ensure the prevention of any external influence by metallic objects on the electromagnetic force (EMF) (Wi-Fi) which can distort the CSI information while care was taken to keep the antennas of the devices unobstructed. Figure 3.2a shows the receiver of the experimental setup while the hardware configuration used for this research is shown in Fig 3.2b.

To mimic a realistic pedestrian crossing environment, the Tx and Rx stands were placed on opposite sides of the 5 meter crossing. Each stand was placed with an offset of 1.2 meters from the point of walking trajectory as shown in Fig. 3.2c. This corresponds to the dimensions of a standard pedestrian crossing width, typically 2.4 meters wide.

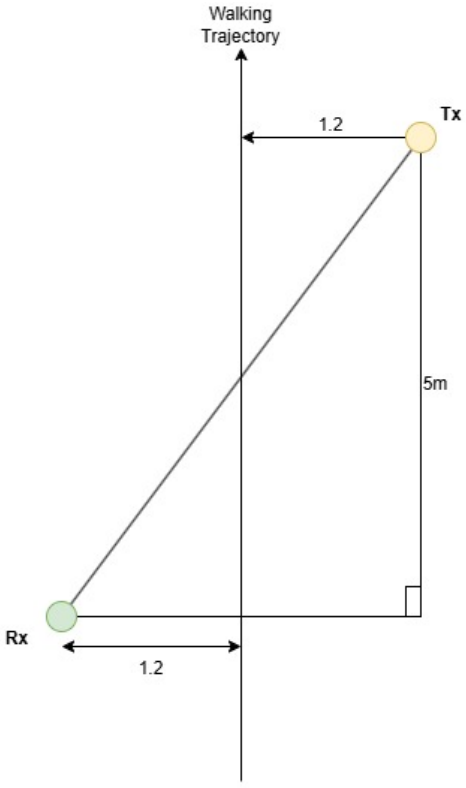
Transmitter ESP32 device is configured using a custom Station (STA) script which continuously streams UDP data packets towards the receiver that has its firmware configured as the Access Point (AP) from [81]. This AP-STA setup combination allows the deployment of multiple STA devices to create a MIMO application. However, for this research a single Tx-Rx pair is used to evaluate the feasibility of the system. Notably, a multi-STA configuration would require synchronized clocks between the devices to ensure timesteps are aligned to extract meaningful temporal information from the CSI data.

The height of the Tx-Rx pair was fixed at 1.4 meters above ground level aligning to a average human heights ranging between 1.2 and 1.6 meters. As studies have indicated [71] the majority of scattering and non-line-of-sight (NLOS) paths originate from the torso region hence this height was decided.

(a) Hardware setup deployed at a location



(b) Hardware setup of the system



(c) Top-down geometrical view of the setup

Fig. 3.2: Hardware deployment and geometrical configuration of the CSI-based pedestrian detection field setup

3.2.2 CSI Data logger application

A simple flask web application was developed to run on the Raspberry Pi which connects to a smartphone via a personal Hot-spot allowing the server to be accessed via designated IP address and the Port. First time setup involves using remote access to the raspberry pi using applications like RealVNC viewer installed on both the smartphone and the raspberry pi to verify the device's accessibility and to execute the flask service to start the logging server.

The application's interface consists of two simple UI buttons that can trigger the "Start" and "Stop" logging function which triggers a popup to type in the file name when a logging session is completed. A discard button is also included to delete any unwanted session data that was collected potentially caused by collection errors or unpredicted disturbances from the environment such as when vehicles or animals stray into the channel. A sample interface of the application can be found in Fig. 3.3.

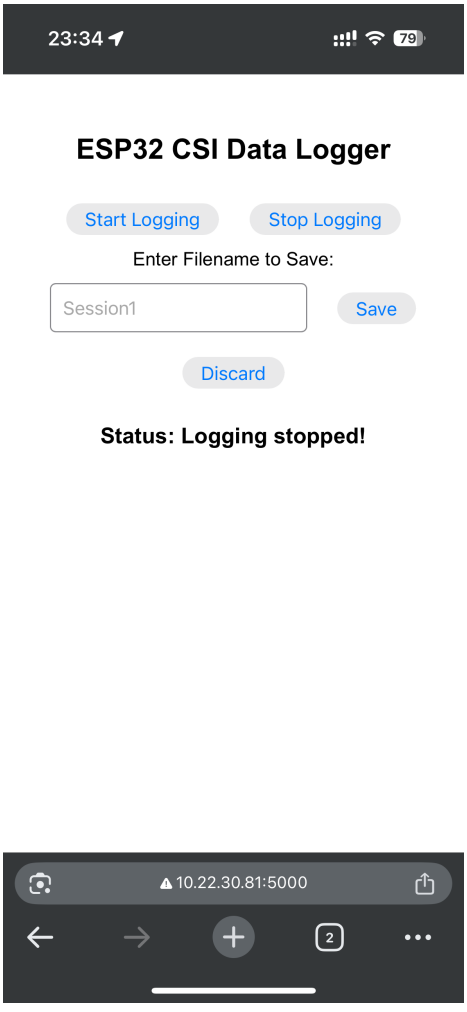


Fig. 3.3: Interface of the flask app used for CSI logging

3.2.3 Packet Loss Analysis

With the experimentation setup finalized for data collection, preliminary tests were conducted on the ESP32 devices to ensure they extracted data at 100 Hz sampling rate. Both STA and AP were programmed to log and validate the packet counts received over a fixed time period. The tests showed that there were considerable levels of packet loss especially when the distance between the Tx-Rx was increased. This could cause issues in skipping important CSI frames and needed to be addressed.

A packet was considered lost when (i) the CSI frame did not have the full array of information (all 64 columns of data in the CSI frame) or (ii) if there were missing rows such as logging fewer than 1000 frames for data collecting over a 10 second period.

To verify, a RSSI validation was conducted using a test code to measure the signal strength at the receiver. Results confirmed that when the distance increases between the devices, the RSSI drops considerably which could lead the observed packet loss. To mitigate this issue, a +2dB Antenna was fixed to the ESP32 to check if the performances will improve. Graph shown in Fig. 3.4 and Fig. 3.5 show the results obtained.

It is worth noting that, the potential of missing data could happen between UART communication and logging independent of the signal tests that were conducted. However, as the final system will be setup to use 5 meter gap between the Tx-Rx pair, the packet loss can be considered negligible and the EPS32 device attached with the +2dB gain antenna was used as shown in the Fig. 3.2b.

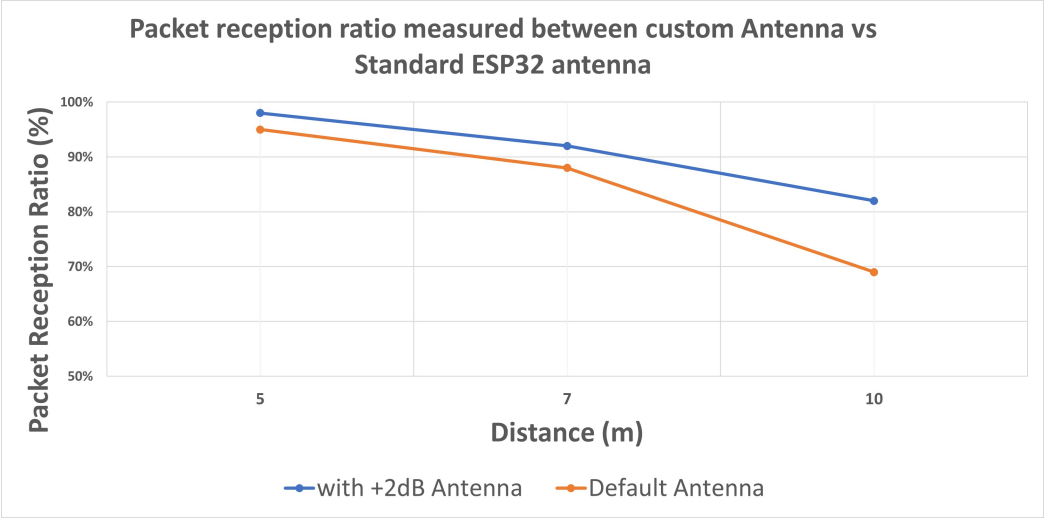


Fig. 3.4: Packet loss observed between ESP32 devices with and without +2dB antenna

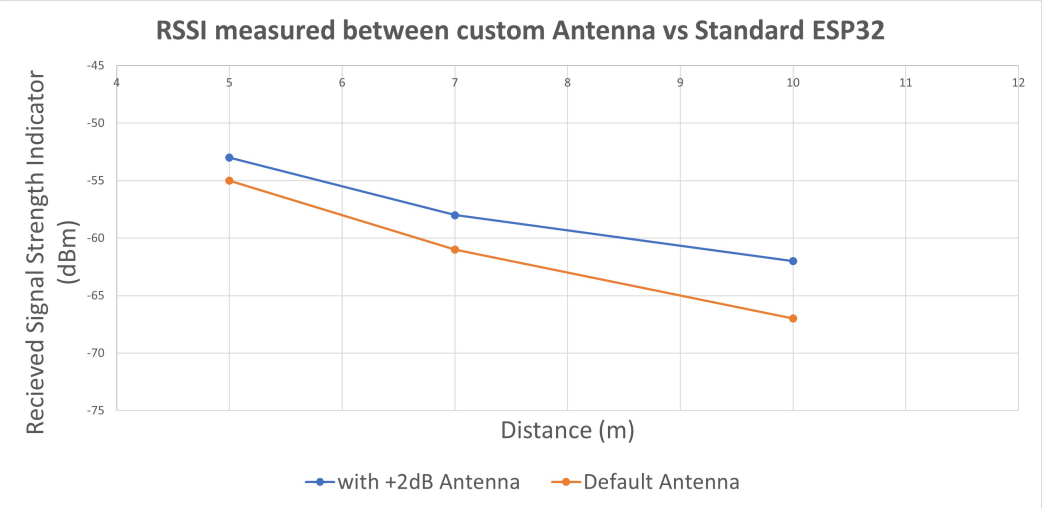


Fig. 3.5: RSSI observed between ESP32 devices with and without +2dB antenna

3.3 Data Collection

Data collection for CSI based sensing required extreme caution as a simple unexpected variation can disrupt the entire sequence of data degrading the performance of the overall system. Data collection process was conducted to capture data between diverse sets of individuals and and phased out to obtain them from varying environments across several days to ensure generalization.

At each location, participants were asked to walk individually in the trajectory set as a fixed path in the LOS line between the transmitter and the receiver as illustrated in Fig. 3.6 while CSI was continuously recorded for each walk separately. The data were collected across diverse group of individuals with varying body builds and gaits. As CSI is highly sensitive to the sized and shape of the subject in motion between the channel, this diversity in the data set including a female can help build a robust classification system. A full breakdown of the information regarding each individual can be found in Table 3.1.

TABLE 3.1: MEASUREMENTS OF SUBJECTS PARTICIPATED

Person ID	Height (cm)	Collar Width (cm)	Waist Size (cm)
A	170	40	81
B	165	42	79
C	160	40	86
D	187	44	106
E	172	42	86
F	180	38	82
G	158	40	101

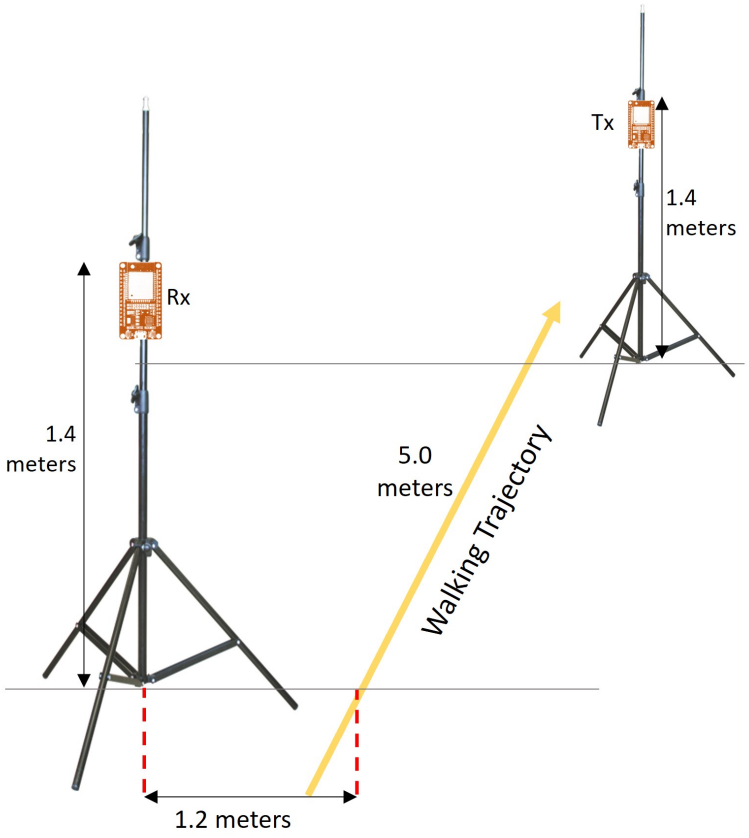


Fig. 3.6: Illustration of the setup and walking trajectory for data collection mimicking an actual pedestrian crossing

The collected data were categorized and labeled into three classes for each person:

- **Empty:** No human crossing between Tx and Rx (background CSI only).
- **Rx2Tx:** Subject walking from the receiver side toward the transmitter.
- **Tx2Rx:** Subject walking from the transmitter side toward the receiver.

In total, the dataset consists of approximately 4000+ labeled CSI sequences between the three classes distributed evenly. Data collection was performed at four separate outdoor locations as shown in Table 3.2, some of which are shown in Fig. 3.7 with subjects walking individually. CSI was recorded at 100 Hz and will be later segmented into uniform windows for training and classification.

The data collected for certain individuals are not available across all environments. This gap however, is practical for the setup as the final system for the movement detection in a crossing must generalize for subjects across environments for "unseen" data. The overall data structure and distribution is visualized in Fig. 3.8.

TABLE 3.2: EXPERIMENTAL ENVIRONMENTS AND TRANSMITTER-RECEIVER DISTANCES

Location ID	Location	Tx-Rx Distance (m)
1	ENTC – UoM	5.0 ± 0.4
2	Civil – UoM	5.0 ± 0.4
3	Wattala	5.0 ± 0.4
4	Battaramulla	5.0 ± 0.4

Note. Department of Engineering and Telecommunication (ENTC) , Department of Civil Engineering (Civil) and University of Moratuwa (UoM)

(a) ENTC department, University of Moratuwa

(b) Civil department, University of Moratuwa

(c) Battaramulla Housing Scheme

Fig. 3.7: Three out of the four deployment locations used for CSI data collection

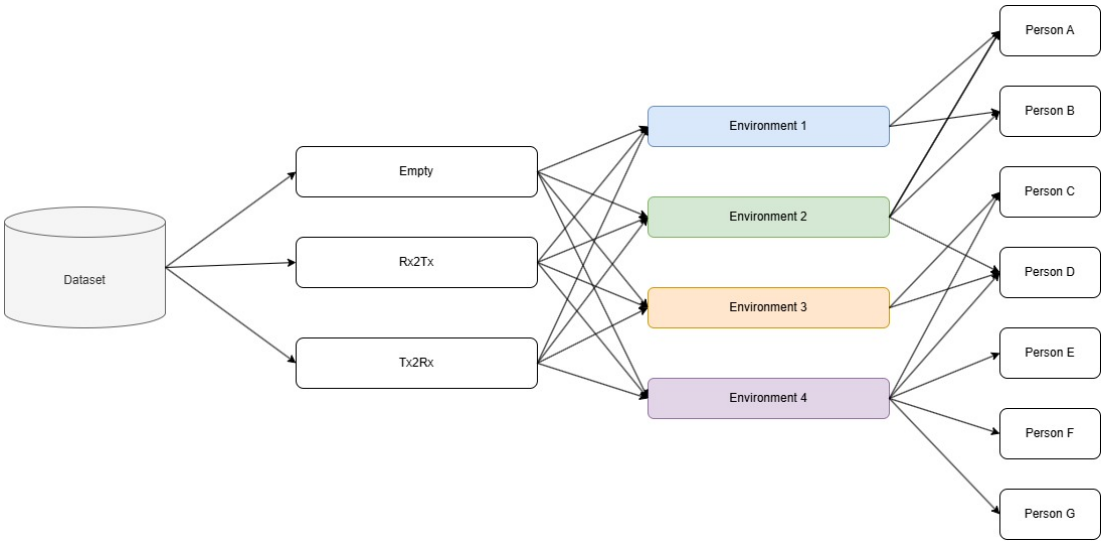


Fig. 3.8: Data Structure used in the System

3.4 Preprocessing

CSI information is a capture of both LOS propagated signal directly between the channel and NLOS signals which are the main reflected signals that holds the motion induced information as studies have shown. Capturing this NLOS components is the key to recognition various activities. Eq. 3.5 shows the representation of a CSI signal's superposition of both static and dynamic components of the propagated signal.

$$H(f,t) = \underbrace{\sum_{n \in \mathcal{P}^s} a_n e^{-j2\pi f \tau_n}}_{H_{\text{static}}(f)} + \underbrace{\sum_{m \in \mathcal{P}^d} a_m(t)\, \xi(f)\, e^{-j2\pi f \tau_m(t)}}_{H_{\text{dynamic}}(f,t)} \tag{3.5}$$

However, for an outdoor environment, the dynamic component of the signal would not only consist of the motion from the subject of activity but also noisy information induced by uncontrollable objects within the environments such as from foliage, animals and birds, winds and other variable transient conditions. This requires a stringent preprocessing mechanism to isolate the activity from environmental noise to best represent the motion and create relatable features for classification.

3.4.1 Amplitude and Phase Visualization

Visualizing CSI signals plays a vital role in understanding the underlying characteristics of motion induced information in the CSI frames. CSI information generated the ESP32 consists of the imaginary value followed by the real value of its respective subcarrier in the information frame as shown in Fig. 3.9.

Using tools such as Python, MATLAB and GNU Octave, the raw CSI data logged from the ESP32 devices can be calculated into their respective amplitude and phase

```
[ 2 -29 2 -30 4 -26 4 -24 4 -23 ]
[ -12 26 -12 25 -13 22 -13 21 -11 20 ]
[ -18 21 -17 20 -18 17 -16 15 -15 15 ]
```

Fig. 3.9: Snippet of raw CSI values across first 5 subcarriers for three time steps.

components and plotted across time and subcarriers.

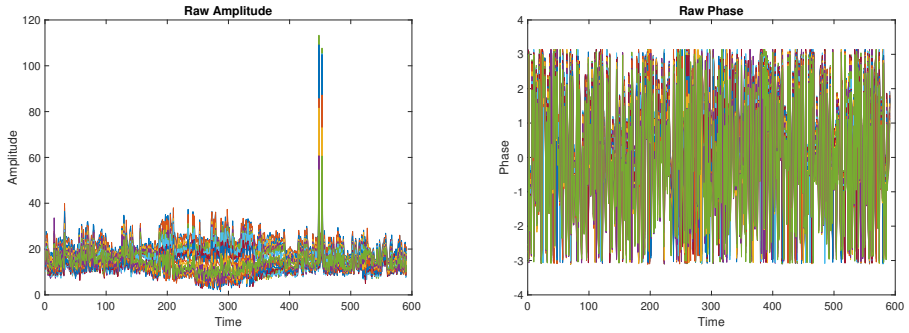


Fig. 3.10: Visualization of Raw Amplitude and Phase for a single activity of Tx2Rx graphed using MATLAB

Figure 3.10 visualizes raw amplitude and phase corresponding to a single instance of a subject walking from Tx to Rx, labeled as the *Tx2Rx* movement class. The amplitude plot reveals clear temporal fluctuations across subcarriers, which indicates the presence of movement for the duration. Meanwhile, the phase information appears more erratic due to noise and hardware constraints which will require further preprocessing before moving to feature extraction.

These initial exploratory plots validate the presence of motion patterns in the CSI data and highlights the importance of denoising and filtering to isolate meaningful features for classification.

3.4.2 Phase unwrapping and Sanitization

Raw phase information from the CSI data is wrapped within the range $-\pi, \pi$ as noted in the ESP32 CSI datasheet [19], making it difficult to interpret or extract meaningful patterns over time or across subcarriers similar to the amplitude graph. To visualize and extract features, a phase unwrapping step is applied. This process is performed along both time steps (row-wise) and the subcarriers (column-wise) of the phase information.

However, even after unwrapping, the phase data remains affected by hardware imperfections such as carrier frequency offset (CFO) between the central frequencies of the OFDM, sampling time offset (STO) due to asynchronous clock and random phase shifts. These artifacts make it difficult to interpret the motion information in the phase which is essential for classification.

To address this issue, a popular phase sanitization algorithm is applied similar to [14]. This method estimates and removes the linear phase distortions caused by CFO and STO, isolating the motion related components of the phase. The sanitized phase provides a better visualization as shown in Fig. 3.11.

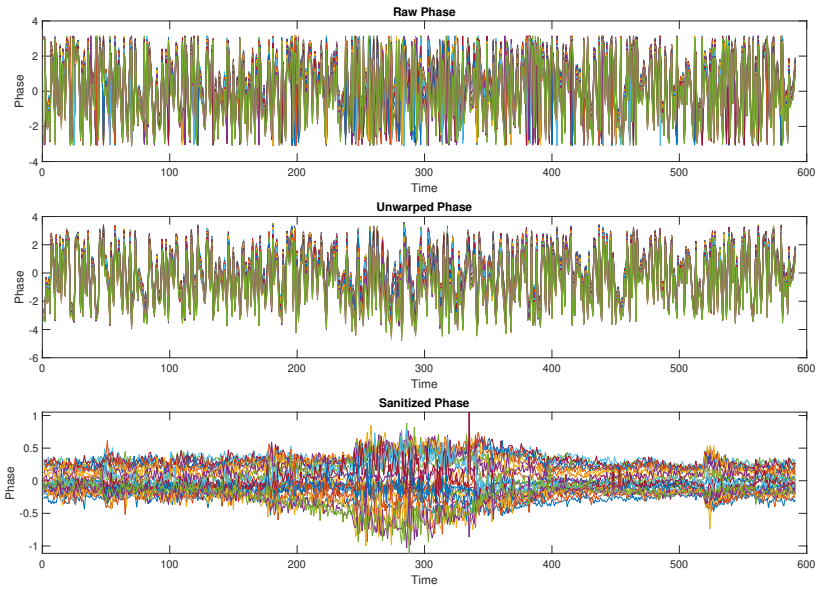


Fig. 3.11: Visualization of Raw Phase vs Unwrapped Phase vs Sanitized Phase for a single activity of Tx2Rx graphed using MATLAB

3.4.3 Filtering techniques

CSI data particularly in outdoor conditions is highly prone to dynamic noise from the environment from passing vehicles, animals and trees etc. Since CSI is a representation of the superposition of all such disturbances, extracting motion related feature becomes challenging. This can be attributed to the lower multipath components in outdoor settings such as at a pedestrian crossings compared to indoor settings as illustrated in Fig. 3.12. Therefore effective preprocessing and filtering techniques must be adapted to ensure movement related information are meaningfully isolated for classification.

3.4.3.1 Hampel Outlier Filtering

Due to the dynamic nature of outdoor environments, CSI data contains noise from non-human sources. To suppress these while preserving motion patterns, a Hampel filter was applied to both amplitude and sanitized phase information. A threshold of $\sigma = 3$ over a sliding window of 11 time steps (selected due to the sampling rate set at 100Hz) was selected empirically to identify spikes as a standard outlier detection. The

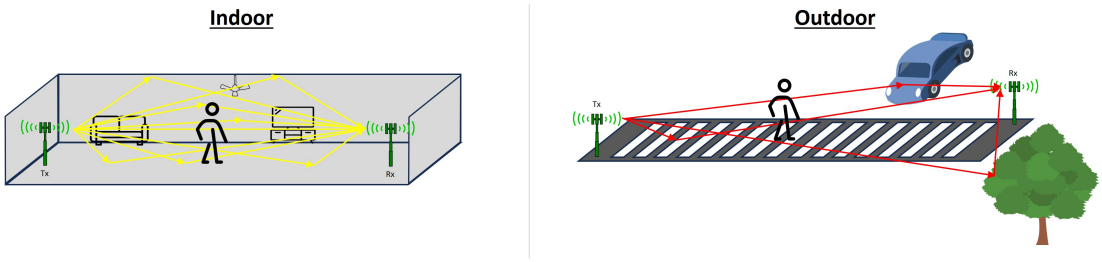


Fig. 3.12: Illustration of the Intensity of Indoor vs Outdoor CSI multipaths

indication of the Hampel outlier detection is shown in Fig. 3.13. The filter was applied such that it would calculate the median value and replace the outlier data point instead of removing it to ensure the CSI signal integrity is maintained and that no data points are removed causing information loss in the CSI data.

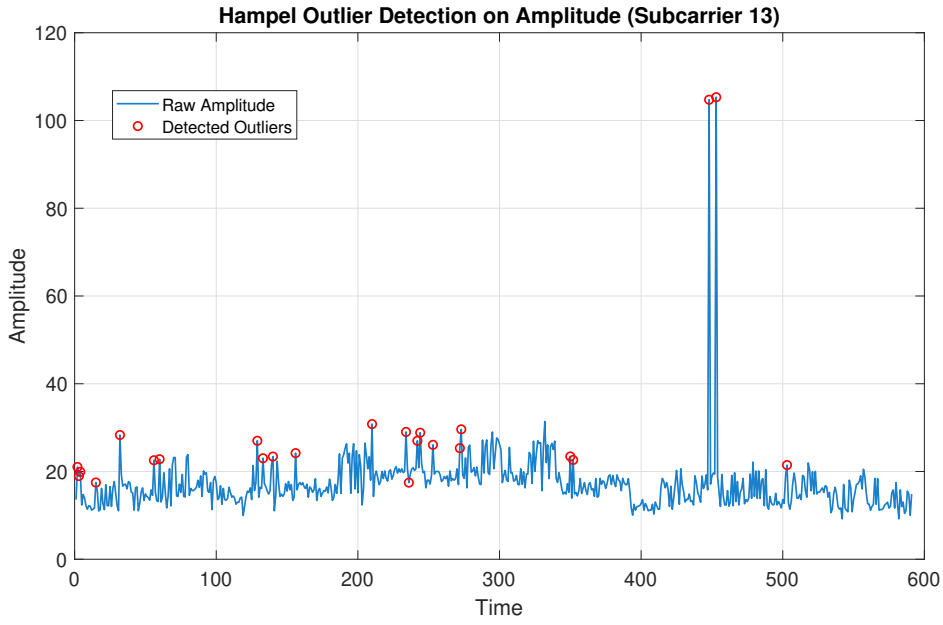


Fig. 3.13: Visualization of Amplitude overlaid with Hampel outlier of 13th subcarrier graphed using MATLAB

3.4.3.2 Kalman Denoising

Once outliers have been removed, further refinement of the signal was considered to remove any residual environmental noise while preserving the structure of motion, a Kalman filter was used as an adaptive smoothing technique. Kalman filters are considered adaptive due to the nature of capturing relative gain within a provided window by estimating the underlying signal state. Kalman filters are also well known for the real-time operability and supports interpolation of data points, however for this research

Kalman will be strictly used for denoising as illustrated in the Fig. 3.14.

$$
\begin{aligned}
\hat{x}_t &= \underbrace{\hat{x}_{t-1}}_{\text{prior estimate}} + \underbrace{K_t\left(z_t - \hat{x}_{t-1}\right)}_{\text{measurement correction}} \\
K_t &= \frac{P_t}{P_t + R} \\
P_t &= (1 - K_t)P_t + Q
\end{aligned} \tag{3.6}
$$

From the eq. 3.6, $\hat{x}_t$ denotes the estimated signal at time step t, calculated using the prior estimate and a measurement correction from the previous time step. The Kalman gain K_t controls the measurement correction estimate, while P_t denotes the error covariance that changes over time. However for implementation, the process noise variance $Q = 10^{-3}$ and measurement noise variance $R = 10^{-2}$ were empirically chosen to get effective subcarrier level smoothing without removing movement information in the CSI signal.

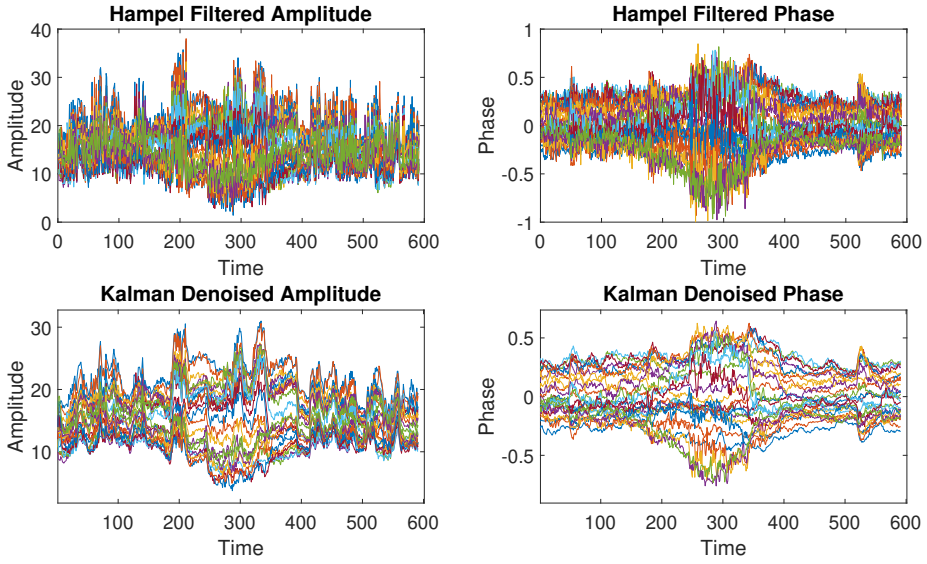


Fig. 3.14: Visualization of Kalman filtered Amplitude and Phase information against Hampel graphed using MATLAB

Alternatively, a butterworth filter was initially considered based on the assumption that motion related information such as walking could be identified in the lower frequency components (between 1Hz to 5Hz) of the CSI signal hence applying a low-pass filter was expected to isolate this information. However, it was observed that due to the noise by slow moving objects in the environment such as swaying of tress and wind gets captured and remains unfiltered. Hence, butterworth filtering was dismissed confirming with an ablation test on performance of filters attached in the Appendix A.

3.5 Exploratory Data Analysis

With the ability to visualize collected CSI information, a series of exploratory data analyses was conducted to better understand the dynamics formed by CSI to collect insights. These patterns and insights will better inform on the development of the classification problem for the pedestrian crossing developed on the basis of visualizing denoised amplitude and phase components across subjects and environments.

3.5.1 Subject Variability Analysis

Initially, the CSI data was visualized to better understand how each participant's CSI signature varies between each other based on the multi-path reflections. For this, Fig. 3.15 was plotted of person C and D collected from location 3. The amplitudes between the subjects show distinguishable transients however person D has significantly more variations in motion, potentially due to the physique and their movement intensity such as walking speed. Similarly, phase shifts reflect the signature of movement in descriptive ways which can be leveraged by a machine learning model.

However, the overall observation from the CSI data highlights that despite being highly distinguishable between subjects, the features in their raw form would create difficulties in generalizing across all participants. This creates a need to extract useful features that assist the classification problem and help generalize as the model is expected to classify for "unseen" data such as for the application of a pedestrian crossing.

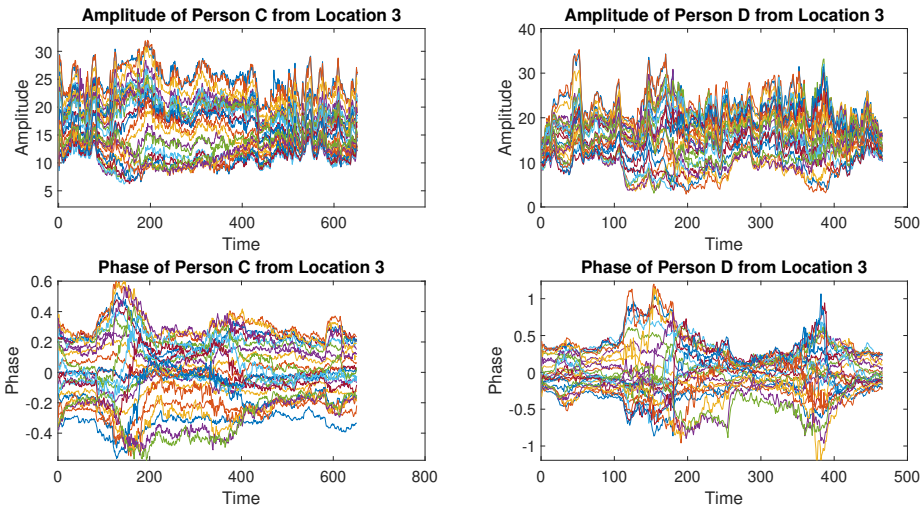


Fig. 3.15: Visualization of Denoised CSI information between subject C and D

3.5.2 Environment Variability Analysis

Similar to the Subject variability, CSI signatures are analyzed by visualizing data between locations of the same person as shown in Fig. 3.16. The plots clearly highlight the domain-specific nature of CSI where amplitude and phase information of the same person creates different transient patterns at different locations.

For instance, amplitude signals from location 2 show more pronounced fluctuations and higher variance compared to location 1. This poses a potential challenge in diversifying a classification model that can generalize across different outdoor settings.

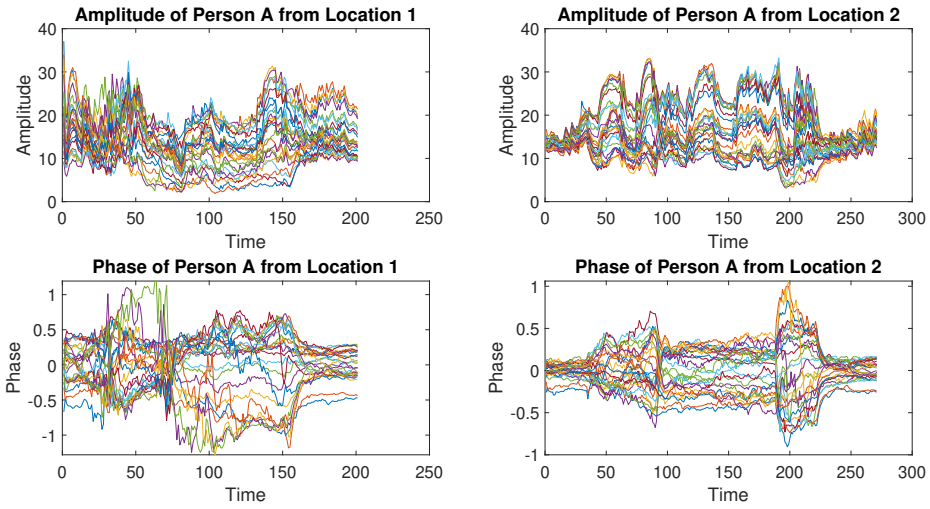


Fig. 3.16: Visualization of Denoised CSI information between same subject from Location 1 and 2

This further emphasizes the need for feature engineering that can capture and isolate relative motion from CSI data across participants and environments.

3.5.3 Feature Engineering

With initial testing, the information contained within amplitude and phase for determining the three classes were tested. However, results showed accuracies of roughly 89% when these two original components are used for testing. This could be due to the dynamic noise components that are still superposed in the CSI frames from the outdoor environment. This separates the direct implementation of indoor CSI setups for outdoor environments.

Hence statistical features which can suppress additive noise and isolate movements are extracted from amplitude and sanitized phase signals for each sliding window once Hampel and Kalman denoising are applied. A window length of 50 samples with a 10% overlap was used to align within the 100 Hz sampling rate. The following features for ($f1$ to $f8$) were selected based on their ability to capture movement and direc-

tion related variations across time and subcarriers from insights learned from related studies:

- **Mean of Amplitude (f_1)**: Gives the average magnitude level of the signal within a window. As a person moves through the channel, the overall amplitude of the window often modules due to added multipath components signifying movement.
- **Variance of Amplitude (f_2)**: Measures how much the amplitude changes over time. Higher variance corresponds to the level of activity in the channel caused by movement.
- **Standard Deviation of Phase (f_3)**: Captures how values are deviated for the sanitized phase. Relative motion often disturbs the phase more than elements from a static environment, making this a useful feature.
- **Mean Absolute Difference of Phase (f_4)**: Checks how fast the phase is changing over time. It helps identify dynamic behavior in the signal which can signify motion.
- **Temporal Centroid (f_5)**: The energy distribution of the rate of change in phase across subcarriers over the window. This is a feature that suggests where the "center of energy" is within the window as a directional cue for the model related to the direction between Tx and Rx.
- **Centroid Differential (f_6)**: The differential between the previous f_5 window to the current f_5 window value across subcarriers to determine change in value that provides directional information.
- **Mean Instantaneous Frequency Differential (f_7)**: Measures how fast the phase is changing over time per subcarrier. Higher values should indicate when the subject is moving towards the receiver while lower is a indication that the person is moving away..
- **Mean Inter-Subcarrier Correlation (f_8)**: Calculates the mean value of the correlation between all 26 subcarriers of phase within the window. Positive values should typically indicate Rx2Tx movement while negative values should correspond to Tx2Rx direction.

For this research, statistical features $f1$ to $f4$ are calculated for all 26 subcarriers resulting in a 104-dimensional $[26 \times 4]$ input sequence per window for Stage 1. For Stage 2, a 79-dimensional $[(26 \times 3) + 1]$ input sequence is computed consisting of $f5$ to $f7$ for all subcarriers and a global mean value $f8$ of the window as a snapshot of

the CSI data. This design captures both amplitude and phase changes in a relatively compact format that helps to suppress residual noise still contained within the CSI sequences while still being computationally light.

3.5.4 Dimensionality Normalization Strategy

To ensure consistent scaling across subcarriers and subjects which is a common practice used for classification model training in machine learning, a Z-score normalization was applied to all extracted features to support models convergence while keeping the computational cost lower as the feature set of the system was already relatively simpler.

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{3.7}$$

Each feature vector x was then normalized as:

$$z = \frac{x - \mu}{\sigma} \tag{3.8}$$

where, x refers to the i^{th} feature value across all training samples for a given feature dimension, N is the total number of training samples, μ is the mean and σ is the standard deviation. The resulting z is the normalized feature value. This z-score normalization ensures that each feature dimension contributes equally to the model by removing scale differences which needs to be maintained separately for live deployments and tuned when more training datasets can be collected.

These normalization step shown in Eq. 3.7 and Eq. 3.8 removes bias due to variations in subject and environment specific motion patterns from the environment which can create scaling differences in the training data. The mean μ and standard deviation σ of each feature dimension are computed using only the training data and will be applied on the CSI test data to ensure realistic generalization of the system across subjects and environments.

3.5.5 The Two Stage Classification Framework

To achieve the expected performance with a lightweight classification system for pedestrian movement detection using Wi-Fi CSI, a classification framework was necessary. During preliminary testing for movement detection, the system was assumed to be a multi-class classification problem to detect between Empty, Rx2Tx and Tx2Rx classes. However, the performance of classification between the three classes were not within acceptable levels particularly in distinguishing between the direction classes.

With deeper analysis, it became apparent that the system could benefit from efficiently handling the detection problem as a two-stage binary classification. First stage

focuses on "Empty vs Movement" for the pedestrian detection covering the primary objective of movement detection and second stage can stretch the model's capability to identify the direction of movement as the data captured is inherently segmented as "Rx2Tx vs Tx2Rx" if movement has been identified in the first stage. Figure. 3.17 illustrates the methodology of the framework.

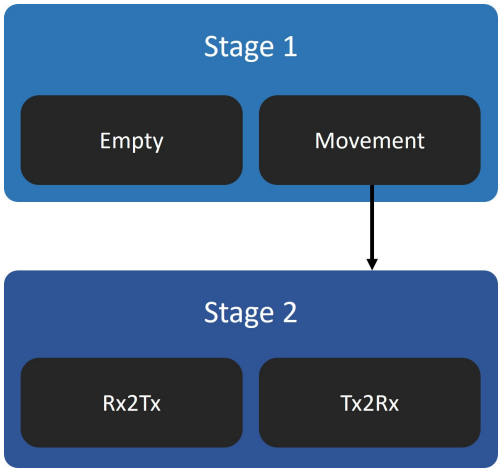


Fig. 3.17: Illustration of the Two-stage classification framework proposed in the research

This two-stage design offers several practical advantages:

- **Separates Classification Task:** By isolating movement detection from directional classification, models at each stage becomes more focused. Stage 1 trains as a binary classifier for movement while Stage 2 handles the directional detection. This modular approach effectively allows for faster response time from the system which can meet ETSI requirements for VAM.

- **Primary Interests:** Many applications such as warning pedestrians could primarily require only the binary presence detection to announce oncoming traffic where direction classification may not be of interest.

- **Improves Efficiency:** Directional classification is only performed when movement is detected, avoiding unwanted computation during idle periods which can save energy and hardware degradation.

- **Passive Noise Reduction:** Stage 1 effectively eliminates noisy or irrelevant CSI segments for direction detection which could potentially introduce false positives in the direction classifications improving overall reliability.

3.6 Classification models

For this research, four classification models were designed and tested, guided by the suggestions from related studies to identify the most accurate and lightweight model suitable for the deployment of pedestrian movement detection. The four models evaluated are as follows:

- Statistical Models: Random Forest and Support Vector Machines .
- Neural networks: Long Short-term Memory and 2-Dimensional Convolutional LSTM (Conv2D-LSTM).

3.6.1 Statistical Models as Baselines

RF and **SVM** were designed using MATLAB Simulink software as baselines for evaluation. The input data consists of windows segmented to form $[features \times timestep]$. However, these models cannot handle sequential inputs hence each window is flattened to $[feature \times 1]$ by averaging across time before being fed into the models for classification. Table. 3.3 summarizes the architecture and the hyperparameter settings decided based on empirical tests balancing the models for a lightweight yet effective classification system.

TABLE 3.3: HYPERPARAMETER CONFIGURATION FOR RF AND SVM MODELS

Model	Hyperparameter and Settings
RF	• Number of Trees: 200 • Tree Depth: Default set by MATLAB • Minimum Leaf Size: 10 • Input: Windowed segments of 104-dimensional feature vector • Out-of-Bag Prediction: Disabled
SVM	• Kernel Function: Linear • Feature Standardization: Enabled • Input: Windowed segments of 104-dimensional feature vector • Kernel Scale: Default

3.6.2 Deep Neural Networks

3.6.2.1 Long Short-term Memory

Due to the inherent time-series nature of identifying movement in a crossing, the LSTM was a suitable option for its ability to learn temporal patterns from the input data. Features extracted from the collected CSI frames are set up in fixed-length windows structuring them naturally for a time-series classification.

LSTM functions in two modes, the *Sequence* and *Last*. In sequence mode, the output of the layer from each $timestep$ is passed preserving the time-series nature while the last mode evaluates the entire length of the input sequence and provides a single output as a summary of that window. As such fig. 3.18 is designed such that it is structured to first learn the time-series patterns of the window looking for motion related information while passing the information to 'last' layers to then converge into a classification.

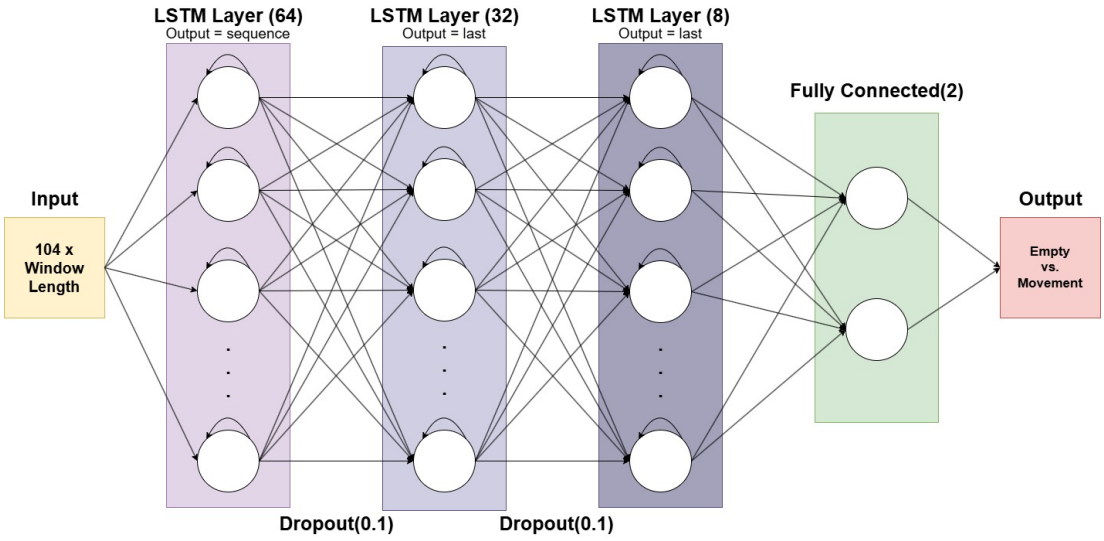


Fig. 3.18: Implementation Architecture of LSTM model for the system

3.6.2.2 Convolutional LSTM

Convolutional-LSTM was considered for the setup due to its ability to identify spatial information and pass it to a LSTM as a sequence. Since the features are arranged in an image-like matrix with resolution 26×4 over fixed window lengths, this model is suitable for the evaluation on the feasibility of deployment for CSI in detecting movement.

As shown in Fig. 3.19, the model is designed to use a Conv2D layer to extract the spatial patterns in the CSI input sequence and then pass it to a LSTM layer operating

in 'last' mode to predict for any change over time as the classification of movement in a provided sequence.

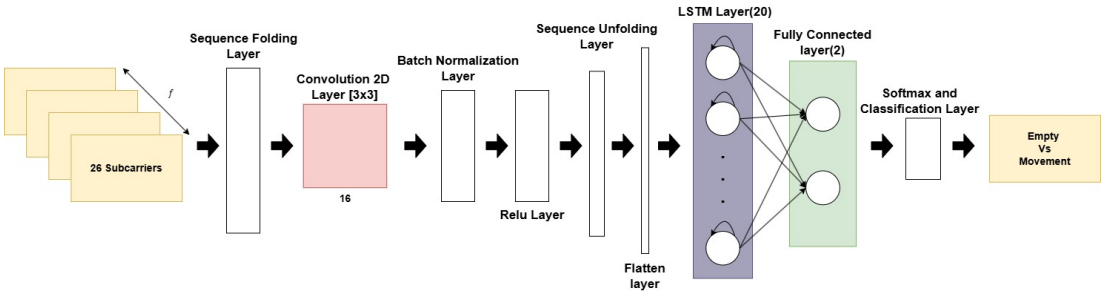


Fig. 3.19: Implementation Architecture of Conv2D-LSTM model for the system

TABLE 3.4: HYPERPARAMETER CONFIGURATION FOR DEEP NEURAL NETWORKS

Model	**Hyperparameter and Settings**
LSTM	• Input: [104 × timesteps] sequences • Training Epochs: 20 • Mini-Batch Size: 32 • Optimizer: Adam • Gradient Threshold: 0.5
Conv2D-LSTM	• Input: [26 × 4 × 1 × timesteps] • Batch Normalization + ReLU activation • Training Epochs: 20 • Mini-Batch Size: 32 • Optimizer: Adam • Gradient Threshold: 0.5

Table 3.4 summarizes the hyperparameter settings for both LSTM and Conv2D-LSTM neural networks used for evaluation. All values were chosen based on empirical tests to balance the models output performances.

3.7 Implementation Framework

A complete overview of the final architecutre framework of implementation is illustrated in Fig. 3.20, guided by the application for a smart pedestrian crossing to detect human movement at a crossing.

This framework serves as the foundation for the evaluation phase, where its robustness and generalizability are evaluated to conclude the feasibility and deployment of the proposed smart pedestrian crossing application.

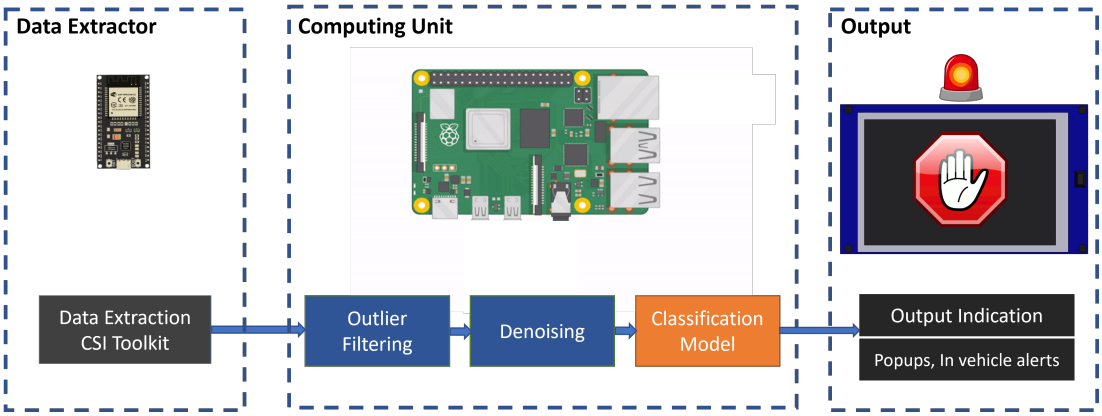


Fig. 3.20: Implementation Architecture for the Proposed System

CHAPTER 4

EXPERIMENTAL EVALUATION AND RESULTS

4.1 Dataset Summary

TABLE 4.1: OVERVIEW OF THE DATASET

Location ID	Persons	Tx-Rx Separation (m)	Average Walking Speed (m/s)
1	A, B	5.0 ± 0.4	1.3
2	A, B, D	5.0 ± 0.4	1.4
3	C, D	5.0 ± 0.4	1.6
4	C, D, E, F, G	5.0 ± 0.4	1.9

The dataset consists of CSI sequences recorded at 100 Hz sampling rate for each class label across seven subjects from four diverse locations with an average walking speed of 1.6 meters per second. It includes CSI data of participants who appear exclusively in locations such as Location 1 and 4 (as summarized in Table 4.1 enabling realistic diversity to evaluate the generalization capability of the final models.

4.2 Evaluation strategy

4.2.1 Leave-One-Out Cross Validation strategy over K-fold and train/test splits

As part of the experimentation and results, cross validation technique was the go-to strategy when evaluating the system. As this research primarily focuses on the potential of deploying a smart pedestrian system, it was essential to ensure that the model is able to detect random "Persons" from unseen data. As such, the Leave-One-Out Cross validation (LOOCV) method was chosen to verify the generalization capability of the model per subject in folds, P.

This provides a significant advantage over the traditional Train/Test splits which can leak knowledge of a person from the training set to the test dataset and create unreliable evaluation. Similar flaw can be observed with the K-fold cross validation where the subject specific information might be shared across folds due to the nature of CSI data and this research. Hence, LOOCV will provide realistic assessment of the system for a real-world deployment of a device-free smart pedestrian crossing for movement detection.

4.2.2 Model Performance Evaluation Criteria

To assess the effectiveness of the proposed classification models in detecting human motion and direction from CSI information, four widely accepted evaluation metrics are used: **accuracy**, **precision** in Eq. 4.1, **recall** in Eq. 4.2 and **F1-score** in Eq. 4.3. These metrics are particularly relevant for binary classification tasks such as Empty vs. Movement detection as they provide useful insight into the performances and help identify class imbalance or misclassifications which can significantly affect the final evaluation.

Given that a matrix with true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) are generated for each fold, the evaluation metrics are defined as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.2}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.3}$$

4.3 Results and Discussion

The following section presents the experimental results of the proposed methodology and implementation framework to deploy a CSI-based HAR system for a smart pedestrian crossing as an ITS application.

4.3.1 Leave One Out Cross Validation (LOOCV) on Persons

The following section covers the results obtained from conducting up to 5 runs of cross validation across folds for each model where the best three scores are selected and averaged.

4.3.1.1 Stage 1

Stage 1 part of the proposed system is a binary classification model which detects between *Empty* and *Movement* classes. This forms a $P-1$ cross-subject testing where P is the total number of person participated. Total of 7 folds were evaluated for each model and the scores were averaged across the subjects and summarized in Table 4.2.

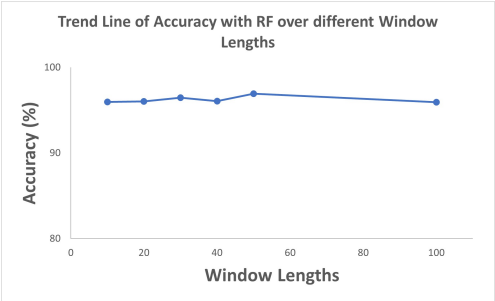
The performance results show that the random forest has slightly outperformed other models, achieving the highest accuracy of 96.9 with strong consistency denoted by its lowest standard deviation across folds. This is consistent with the expectations

TABLE 4.2: STAGE 1 PERFORMANCE OF CLASSIFICATION MODELS

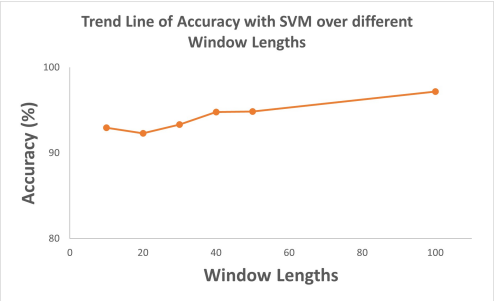
Models	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
RF	96.9 ± 2.77	97.9	97.1	97.4
SVM	94.8 ± 4.97	95.8	96.7	95.9
LSTM	94.6 ± 4.92	97.4	93.7	95.4
Conv2D-LSTM	95.2 ± 4.00	96.1	95.8	95.8

as the input features are derived statistically forming a scalar feature map that reflect effectively with these models. However, Conv2D-LSTM also demonstrates remarkable performance with the computed features across folds given the input sequence is time-series information but it is worth noting that deep neural networks typically perform better provided it can learn with much larger datasets.

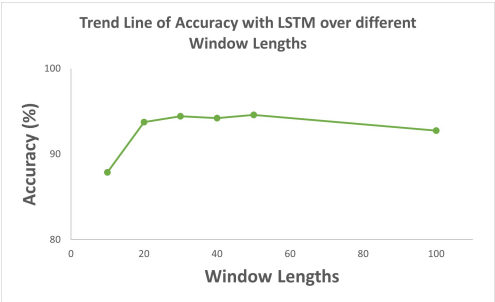
Different window lengths were also evaluated for each model across all folds P. The results of which determines the final window length for deployment within the expectations set by the ETSI TS report.



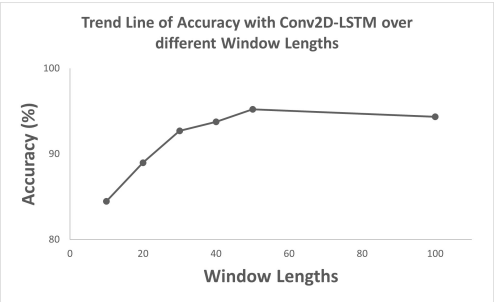
(a) RF



(b) SVM



(c) LSTM



(d) Conv2D-LSTM

Fig. 4.1: Movement Detection accuracy of various window lengths for each model averaged over all subjects

Figure 4.1 shows the comparison across the four models under evaluation. RF maintains consistent performance across all window lengths while SVM shows gradual increase. In contrast, LSTM and Conv2D-LSTM require relatively longer windows to reach comparative accuracy which clearly reflects their nature of requiring much longer

sequences of data to capture temporal context better.

Based on this evaluation, Window length of 50 is chosen for pedestrian detection across models, however RF with window length 10 would still give promising results as a trade-off between accuracy and low latency application.

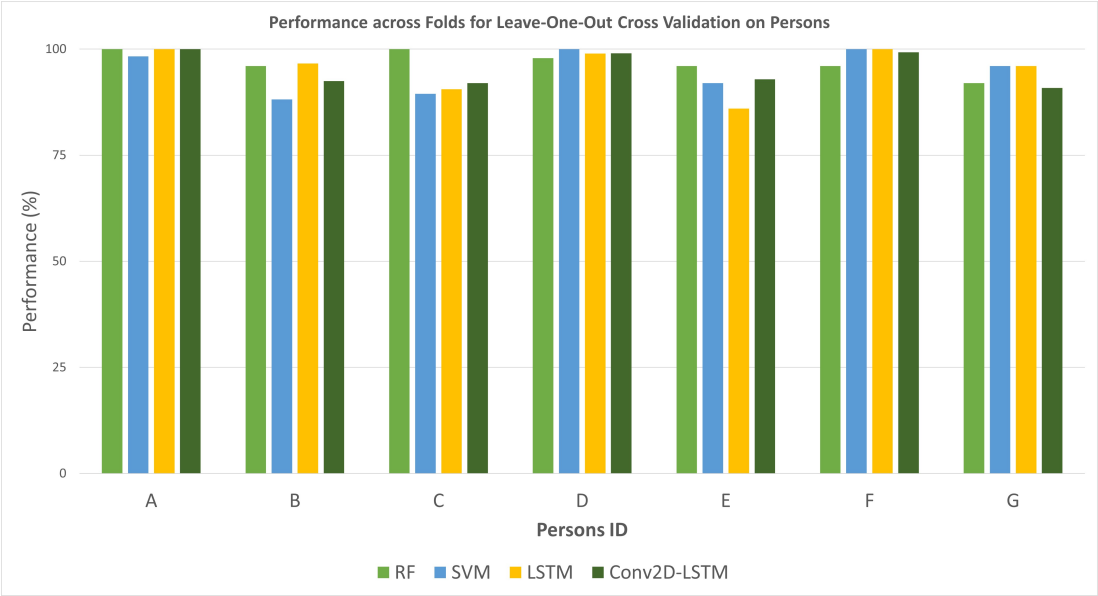
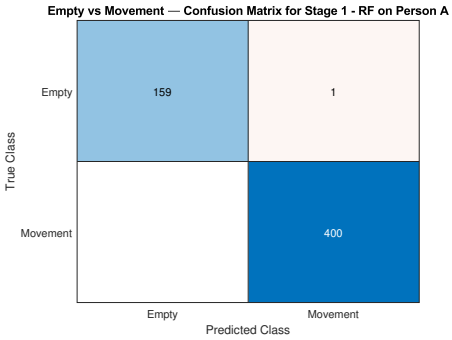


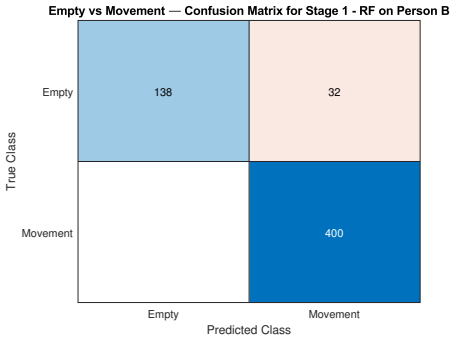
Fig. 4.2: Movement Detection performance across Persons using Leave-One-Out Cross Validation

The summary across persons for each model for the window length of 50 is shown in Fig. 4.2 which highlights the Stage 1 model's ability to generalize to "unseen" data of people. The observable variations between the accuracies can be attributed to the different signatures of CSI related to the participants physical attributes.

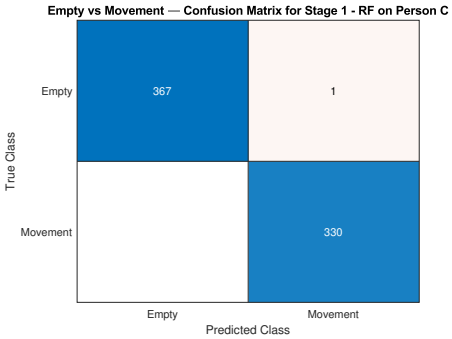
For instance, with the Tx-Rx pair mounted at 140cm for sensing, **Person C and G** being the shortest subjects with heights 160cm and 158cm respectively, received higher variations in test accuracies between models while **Person D and F** being the tallest subjects achieve consistent test accuracies. Although height is one significant factor, collar width and waist size which reflect the overall build of the person also influences CSI's discriminative power. However, **Person E** records lowest accuracies across models corresponding to their smaller physique influencing weaker multipath components and potential human errors in data. Furthermore, confusion matrices in Fig. 4.3 shows generalization for movement on unseen subjects such as for **Person A** of a general physique, highlighting the need to include training data of diverse body types for the application such as a pedestrian crossings.



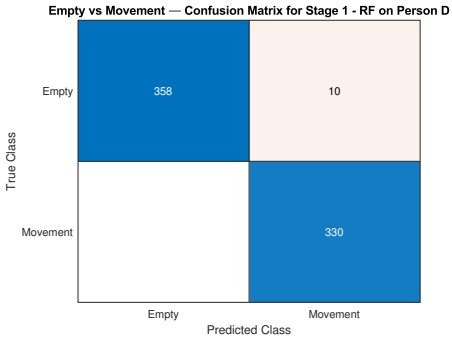
(a) Person A



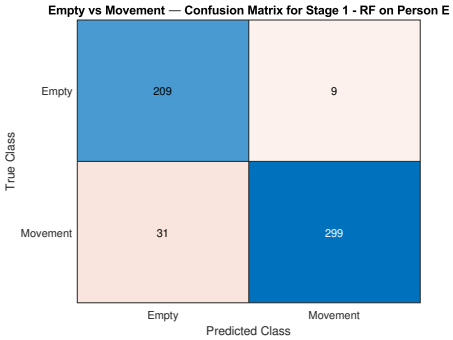
(b) Person B



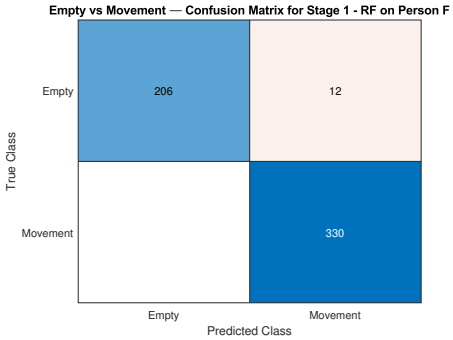
(c) Person C



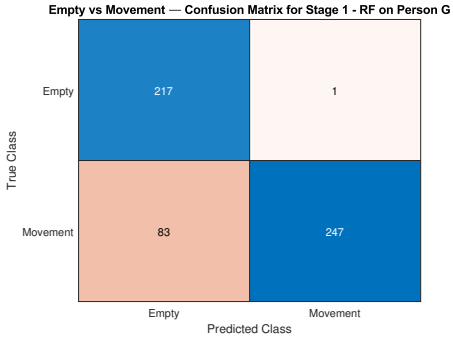
(d) Person D



(e) Person E



(f) Person F



(g) Person G

Fig. 4.3: Confusion matrices for Stage 1 (Empty vs Movement) classification using Random Forest across all participants

4.3.1.2 Stage 2

Stage 2 section of the proposed system is a binary classification model based on the two-stage framework which extends the scope to detects directional $Rx2Tx$ and $Tx2Rx$ classes. This also forms a $P-1$ person cross-subject testing where P is the total number of person in the dataset. Total of 7 folds were evaluated for each model and the scores were averaged across the subjects and summarized in Table. 4.3.

TABLE 4.3: STAGE 2 PERFORMANCE OF CLASSIFICATION MODELS

Models	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1 Score (%)**
RF	51.1 ± 3.15	51.1	55.0	52.6
SVM	47.1 ± 5.98	47.7	52.9	50.0
LSTM	52.9 ± 7.84	54.4	53.6	52.8
Conv2D-LSTM	62.4 ± 6.86	60.3	72.8	65.6

The directional classification performance across the models show difficulty in distinguishing between the two directional classes using the provided feature set of CSI data. This constraint can be attributed to the CSI information from subcarriers forming a symmetrical pattern when a person is moving along the crossing and passing the midway point making it difficult for models including deep learning to clearly separate the opposing directional information.

However, while LSTM shows marginal improvement over the statistical models, Conv2D-LSTM stands out due to its capability to map spatio-temporal patterns as seen across folds in Fig. 4.4 while achieving the highest accuracy of 62.4% but with a higher standard deviation across folds denoting the constraint between folds.

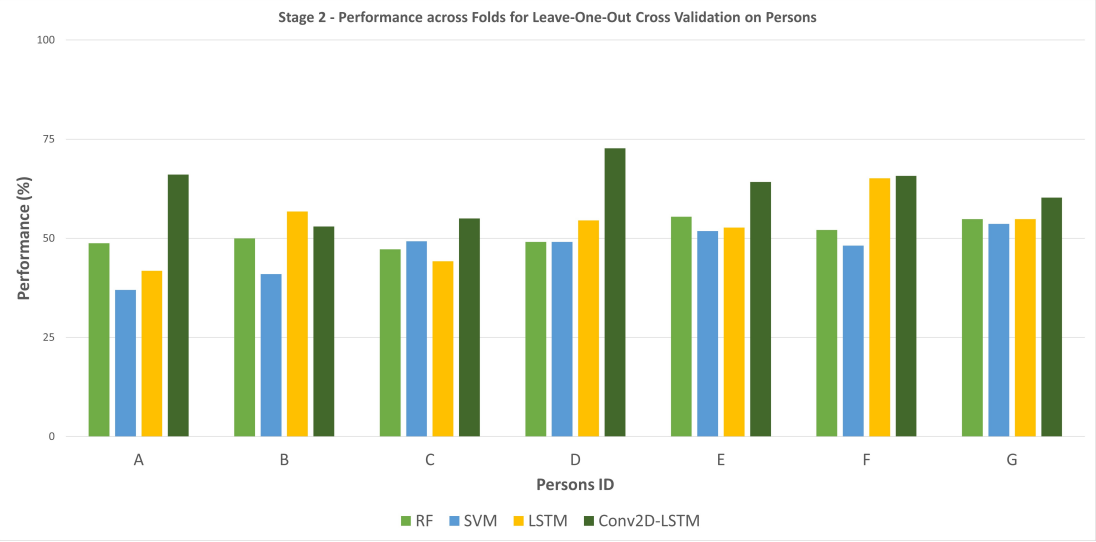


Fig. 4.4: Stage 2 Performance across Persons using Leave-One-Out Cross Validation

4.3.2 Leave One Out Cross Validation (LOOCV) on Environments

Related works [8, 48] demonstrate the nature of CSI being highly domain specific, where models trained for identifying between different actions, cannot generalize between the same actions in multiple locations. With that understanding, running a Leave-One-Out cross validation to generalize between environments across unseen subjects can be considered too strict. This can be witnessed with the results seen in Fig. 4.5 which highlights the domain-specific nature of CSI.

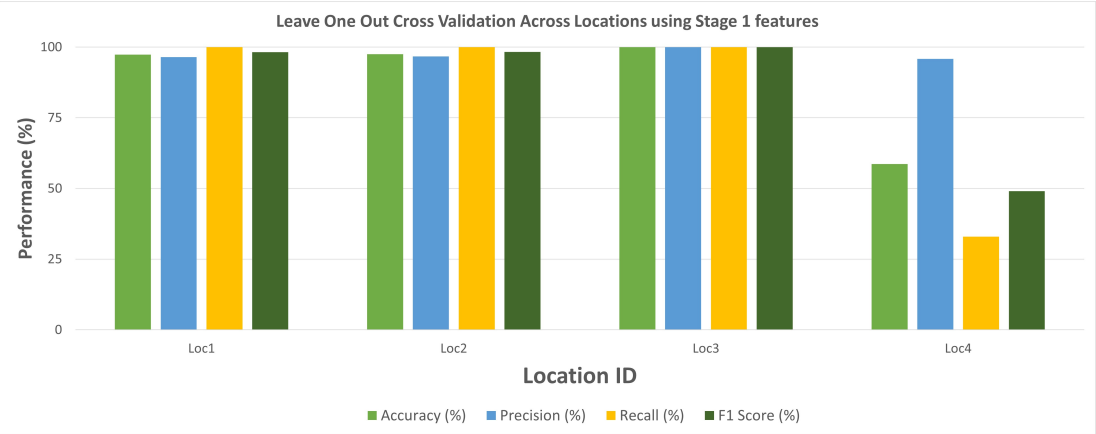


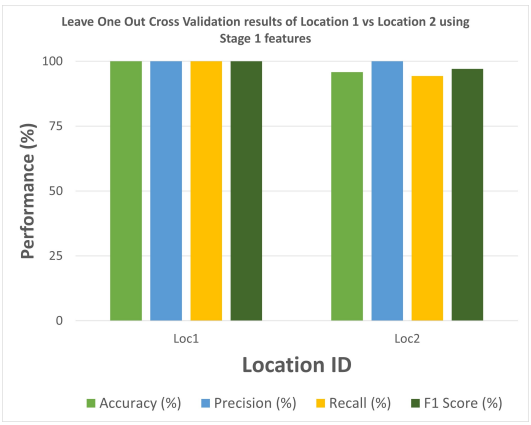
Fig. 4.5: Cross validation results across all environments using Stage 1 features

Location 1,2 and 3 show high performance due to the shared CSI information between Person A,B and D which is captured in the features and the Z-score normalization of the data. However, Location 4 which is highly imbalanced with data of Person E,F,G that are completely unseen by the model in other locations. This evaluation hence becomes a cross-domain analysis of both Environments and Subjects evaluated for generalization.

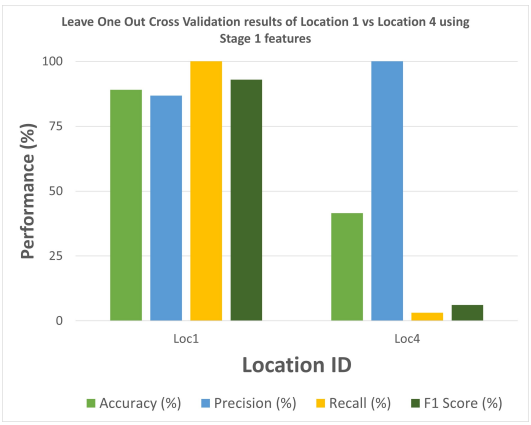
This outcome is consistent with related studies of [8, 48, 65] where the performances are seen to be lower for data trained of the same individuals tested on different environments have shown lower performances. However, Fig. 4.6a illustrates how the proposed framework is able to generalize without losing performance across environments where the training data consists of the same subjects while Fig. 4.6b confirms the cross-domain constraint of CSI based sensing in entirely unseen environments.

4.4 Latency Analysis

A key contribution of this research is the implementation of a CSI-based pedestrian detection using compact, energy-efficient and COTS hardware. The system utilizes ESP32 devices for CSI extraction at 100 Hz and a Raspberry Pi 4B for filtering, feature extraction, and two-stage classification framework at the edge.



(a) Cross validation results between Location 1 and Location 2



(b) Cross validation results between Location 1 and Location 4

Fig. 4.6: Cross Validation results across environments of Random forest using Stage 1 features

A custom script utilizing the End to End system pipeline described in Fig. 3.20 is tested using MATLAB's tic and toc function to simulate the real-world operation. Operating at 50 CSI packets per detection, the system achieves a near-real-time response with an average latency across Stage 1 and Stage 2 of approximately 175 milliseconds. This demonstrates an End-to-End detection time of 675 milliseconds (500ms packet window plus 175ms) as illustrated in Fig. 4.7.

Random Forest-based Stage 1 detection demonstrates reliable accuracy (96.9%) even with 50-packet window, enabling identification of vulnerable road users within the 175 ms. This performance aligns with the expectations outlined in [89], which specifies the total conflict risk analysis window of 1.75 seconds that includes detection, threat evaluation in a conflict zone (e.g. pedestrian crossing), VAM transmission and providing sufficient time for the driver to react to detected VRUs in a critical zone.

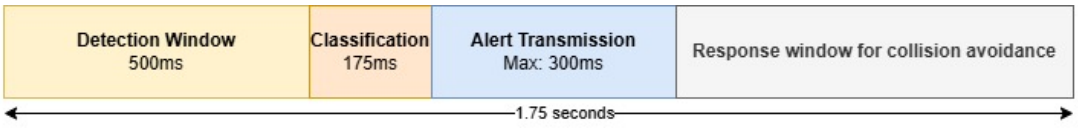


Fig. 4.7: Overall timing map illustration for End to End Latency meeting ETSI requirements

4.5 Discussion

4.5.1 Limitations of the Proposed CSI-based Method

While the proposed research has demonstrated promising outcomes in outdoor pedestrian detection using CSI sensing, the research consists of certain limitations:

- **Limited to Single Pedestrian Detection:** The current implementation evaluates performance involving only one pedestrian crossing the channel at a given moment. Multi-pedestrian scenarios of real-world crossings were not examined which presents a limitation for deployment within the proposed framework.

- **Directional Classification Accuracy Limitation:** Although the system achieved robust accuracy (96.9%) in detecting pedestrian movement (Stage 1), directional detection (Stage 2) demonstrated limited accuracy (62.4%). This indicates that environmental noise masks subtle effects of movement between directions, hence the difficulty of reliably detecting movement direction in real-world outdoor conditions.

- **Sensitivity to New Environments and Subjects:** The classification accuracy is also impacted by environmental factors such as wind, foliage, moving vehicles and varying pedestrian characteristics and other static reflectors different between the locations of deployment. Due to limited number of participants and locations, the model's generalization across diversified outdoor settings can remain uncertain. Thus, extensive data collection encompassing varied environmental conditions (weather, time of day etc.) and a larger subject pool is essential to improve generalization and robustness.

4.5.2 Alternative Techniques for Stage 2

Several hardware-oriented improvements could enhance the robustness and accuracy of the proposed CSI-based classification system. Employing directional antennas would improve the signal-to-noise ratio (SNR) by directing antenna beams precisely onto the pedestrian crossing, potentially reducing unwanted multi-path reflections and environmental interference by vehicles, foliage etc., while strengthening the directional cues embedded in the CSI data.

As the directional classification could not achieve desired results, an alternate approach such as utilizing the localization potential of CSI as supported by the study in [91]. Given that CSI data collected are structured as $Rx2Tx$ and $Tx2Rx$ as a single sequence of movement, they can be repurposed as spatial zones. The approach of continuous estimation of "Zones" will involve taking fixed window length of CSI sequences and training a model to estimate the subject's relative position from the receiver along the crossing as illustrated in Fig. 4.8.

When the subject's position changes from one zone to another, this can assist to estimate the direction based on confidence levels provided by a deep learning model. This spatial modeling for localization could potentially mitigate symmetry that was observed with the current approach proposed for Stage 2.

Alternatively, Integrating static reflectors with the RSU setup such as a metallic

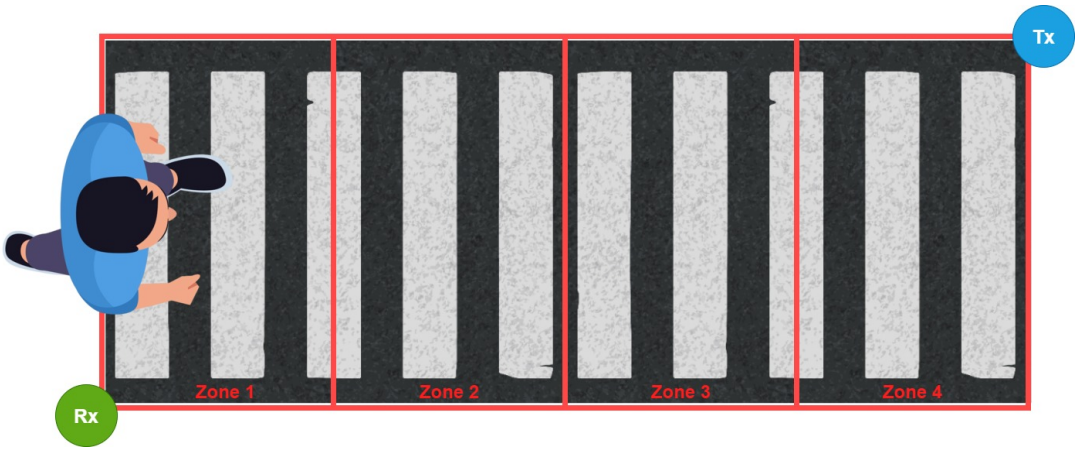


Fig. 4.8: Zone based alternate approach for Stage 2 classification

surface that could reflect corresponding wavelength (12.5cm) of the 2.4GHz Wi-Fi signals, could provide stable and predictable multi-path signatures at the receiver, improving the consistency of CSI data under dynamic noisy outdoor conditions. Additionally, adopting a MIMO antenna arrangement could substantially improve spatial resolution capturing richer CSI features and thus enhancing directional classification performance. Future exploration of such a these method might offer greater robustness and accuracy in practical deployments, effectively overcoming existing limitations in directional classification (62.4%) achieved in the current research. Due to the time constraints and the scope of this research, these methods are recommended as a future work for better directional detection.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

This research presents a comprehensive understanding of Wi-Fi Channel state information for an device-free detection of human movement in diverse outdoor conditions targeting a smart pedestrian crossing as an ITS application. A lightweight modular architecture backed by a two-stage classification system is proposed and designed to evaluate the capability of detecting pedestrian movement and direction across a crossing using adaptive filters, 26 subcarriers and effective feature engineering.

The framework specifically focuses on the design for real-world deployment using COTS hardware such as ESP32 and Raspberry Pi 4B with potential for future upgrades. The system achieves a Stage 1 classification accuracy of 96.9% using random forest demonstrating the feasibility for use in safety applications achieving detection timings of average 175 ms. With the hardware designed for low power consumption (under 200 milliwatts for ESP32 and up to 5 Watts for raspberry pi to compute) within a compact form factor allows the system to be powered using solar panels or rechargeable batteries making the system well suited for scalable ITS applications.

Furthermore, the research conducts robust cross validation across diverse subjects and environments benchmarking classification models and the potential to generalize CSI for HAR applications. Overall the work lays a strong foundation for a deployable CSI based pedestrian detection system highlighting the potential for a non-intrusive, lightweight detection system viable as an ITS application.

5.2 Recommendations and Future Work

The research establishes a foundation for developing an affordable and scalable pedestrian movement detection system using CSI sensing with COTS devices to stay relatively competitive. Using ESP32's low-cost architecture, the approach demonstrates significant potential for future deployments where all processing could occur entirely on the ESP32 hardware, thus removing the need for additional devices such as Raspberry Pi 4. Studies such as in [56], have shown promising results in utilizing the dual-core architecture of ESP32 devices to concurrently perform real-time CSI acquisition and inference tasks. In addition, the proposed two-stage classification framework of an initial stage for movement detection followed by a direction detection provides functional advantage over simple people detection methods. This structured approach ensures robust pedestrian recognition, making it particularly valuable for ITS such as smart pedestrian crossings.

Future work can investigate multi-antenna MIMO-OFDM sensing to enhance Stage 2 classification accuracy. This can also extend the capability to detect walking directions for multiple pedestrians simultaneously with the inclusive of Wi-Fi 5GHz bands which contain additional overlapping frequencies providing higher resolutions of CSI. Applying transfer learning with transformer models can be a potential area of exploration which may significantly improve domain adaptation challenges faced with statistical and deep neural networks. Together, these advances will broaden the system's applicability to complex real-world settings.

REFERENCES

[1] S. Hess, G. Segarra, K. Evensen, A. Festag, T. Weber, S. Cadzow, M. Arndt, and A. Wiles, "Towards standards for sustainable its in europe," *ITS World Congress*, 01 2009.

[2] "Motor traffic (signs, signals, symbols and road markings) regulations no. 02 of 2015," Ministry of Transport and Civil Aviation, Democratic Socialist Republic of Sri Lanka, Colombo, Sri Lanka, Gazette Extraordinary 1940/21, November 2015, accessed: 2025-05-03. [Online]. Available: https://documents.gov.lk/view/extra-gazettes/2015/11/1940-21_E.pdf

[3] B. Noh, H. Park, S. Lee, and S.-H. Nam, "Vision-based pedestrian's crossing risky behavior extraction and analysis for intelligent mobility safety system," *Sensors*, vol. 22, no. 9, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/9/3451

[4] R. Ventura, S. Roussou, A. Ziakopoulos, B. Barabino, and G. Yannis, "Using computer vision and street-level videos for pedestrian-vehicle tracking and behaviour analysis," *Transportation Research Interdisciplinary Perspectives*, vol. 30, p. 101366, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590198225000454

[5] C. Kyrkou, "Yolopeds: Efficient real-time single-shot pedestrian detection for smart camera applications," *CoRR*, vol. abs/2007.13404, 2020. [Online]. Available: https://arxiv.org/abs/2007.13404

[6] J. Mirlach, L. Wan, A. Wiedholz, H. E. Keen, and A. Eich, "R-livit: A lidar-visual-thermal dataset enabling vulnerable road user focused roadside perception," 2025. [Online]. Available: https://arxiv.org/abs/2503.17122

[7] H. Huang and S. Lin, "Widet: Wi-fi based device-free passive person detection with deep convolutional neural networks," *Computer Communications*, vol. 150, pp. 357–366, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140366419301331

[8] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-free human activity recognition using commercial wifi devices," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 5, pp. 1118–1131, 2017.

[9] N. Damodaran and J. Schäfer, "Device free human activity recognition using wifi channel state information," in *2019 IEEE SmartWorld, Ubiquitous Intelligence*

Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), 2019, pp. 1069–1074.

[10] D. Wu, D. Zhang, C. Xu, Y. Wang, and H. Wang, "Widir: walking direction estimation using wireless signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ser. UbiComp '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 351–362. [Online]. Available: https://doi.org/10.1145/2971648.2971658

[11] M. Ogawa and H. Munetomo, "Wi-fi csi-based outdoor human flow prediction using a support vector machine," *Sensors*, vol. 20, p. 2141, 04 2020.

[12] Z. Pu, Q. Zhang, Y. Zhuang, Y. Lv, and Y. Wang, "A device-free wi-fi sensing method for pedestrian monitoring using channel state information," 08 2020, pp. 207–220.

[13] M. Haferkamp, B. Sliwa, and C. Wietfeld, "A low cost modular radio tomography system for bicycle and vehicle detection and classification," in *2021 IEEE International Systems Conference (SysCon)*, 2021, pp. 1–7.

[14] R. Sandaruwan, I. Alagiyawanna, S. Sandeepa, S. Dias, and D. Dias, "Device-free pedestrian count estimation using wi-fi channel state information," in *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 2610–2616.

[15] Tektronix. (2014) Wi-fi: Overview of the 802.11 physical layer and transmitter measurements. Accessed: 2025-05-03. [Online]. Available: https://www.tek.com/en/documents/primer/wi-fi-overview-80211-physical-layer-and-transmitter-measurements

[16] Extreme Networks, "OFDM and OFDMA Subcarriers – What Are the Differences?" 2018. [Online]. Available: https://www.extremenetworks.com/resources/blogs/ofdm-and-ofdma-subcarriers-what-are-the-differences

[17] BPlus Technology Inc., "Discontinued mp2w-5300 adapter," http://www.bplus.com.tw/Adapter/Discontinued_MP2W_5300.html, accessed: 2025-05-12.

[18] TP-Link Technologies Co., Ltd., "Td-w8970 | 300mbps wireless n gigabit adsl2+ modem router," https://www.tp-link.com/nordic/home-networking/dsl-modem-router/td-w8970/, accessed: 2025-05-12.

[19] Espressif Systems, *ESP32 Series Datasheet*, version 4.9 ed., April 2025, 2.4 GHz Wi-Fi + Bluetooth® + Bluetooth LE SoC. [Online]. Available: https://www.espressif.com/sites/default/files/documentation/esp32_datasheet_en.pdf

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, pp. 1735–1780, 11 1997.

[21] WHO, "Road traffic injuries," World Health Organization, Dec. 2023, accessed December 13, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries

[22] I. T. F. (ITF), *Road Safety Annual Report 2024*. OECD/ITF, 2024, no. IRTAD. [Online]. Available: https://www.itf-oecd.org/sites/default/files/docs/irtad-road-safety-annual-report-2024.pdf

[23] C. Dhiman and D. K. Vishwakarma, "A review of state-of-the-art techniques for abnormal human activity recognition," *Engineering Applications of Artificial Intelligence*, vol. 77, pp. 21–45, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197618301775

[24] M. Ghosian Moghaddam, A. Asghar Nazari Shirehjini, and S. Shirmohammadi, "Device-free human activity recognition: A systematic literature review," *IEEE Open Journal of Instrumentation and Measurement*, vol. 4, pp. 1–34, 2025.

[25] F. Li, M. Al-qaness, Y. Zhang, B. Zhao, and X. Luan, "A robust and device-free system for the recognition and classification of elderly activities," *Sensors*, vol. 16, p. 2043, 12 2016.

[26] H. Zhang, H. Du, Q. Ye, and C. Liu, "Utilizing csi and rssi to achieve high-precision outdoor positioning: A deep learning approach," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6.

[27] S. Liu, Y. Zhao, and B. Chen, "Wicount: A deep learning approach for crowd counting using wifi signals," in *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*, 2017, pp. 967–974.

[28] T. Wickramarachchi, D. Dias, T. Samarasinghe, and N. Gokull, "Evaluation of dsrc/wi-fi hybrid communications for intelligent transport systems," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2022, pp. 3509–3514.

[29] M. Miyazaki, S. Ishida, A. Fukuda, T. Murakami, and S. Otsuki, "Initial attempt on outdoor human detection using ieee 802.11ac wlan signal," in *2019 IEEE Sensors Applications Symposium (SAS)*, 2019, pp. 1–6.

[30] "Intelligent transportation system," https://www.sciencedirect.com/topics/engineering/intelligent-transportation-system, accessed: 2025-05-03.

[31] L. Figueiredo, I. Jesus, J. Machado, J. Ferreira, and J. Martins de Carvalho, "Towards the development of intelligent transportation systems," in *ITSC 2001. 2001 IEEE Intelligent Transportation Systems. Proceedings (Cat. No.01TH8585)*, 2001, pp. 1206–1211.

[32] A. Broggi, *Automatic Vehicle Guidance: The Experience of the ARGO Autonomous Vehicle*, ser. G - Reference,Information and Interdisciplinary Subjects Series. World Scientific, 1999. [Online]. Available: https://books.google.lk/books?id=K1e55e8wiEUC

[33] P. Bhatia, "Vehicle Technologies to Improve Performance and Safety," University of California Transportation Center, Tech. Rep., Mar. 2003, accessed: 2025-05-13. [Online]. Available: https://escholarship.org/uc/item/4zw4m05k

[34] S. Lobo, A. Festag, and C. Facchi, "Enhancing the safety of vulnerable road users: Messaging protocols for v2x communication," in *2022 IEEE 96th Vehicular Technology Conference (VTC2022-Fall)*, 2022, pp. 1–7.

[35] C. Mallawaarachchi and N. Amarasingha, "A study on pedestrian crossings in colombo suburbs," 01 2017, pp. 57–62.

[36] P. Porouhan and W. Premchaiswadi, "Proposal of a smart pedestrian monitoring system based on characteristics of internet of things (iot)," in *2020 18th International Conference on ICT and Knowledge Engineering (ICTKE)*, 2020, pp. 1–4.

[37] S. Srinivasan, R. Raman, C. B. Thacker, and A. Shrivastava, "Smart crosswalk management with vehicle-to-pedestrian communication," in *2023 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, 2023, pp. 992–997.

[38] T. Rengarasu, H. Jayawansa, and G. Perera, "Estimation of pedestrian walking speeds at controlled cross walks in sri lanka -a pilot study," 03 2012.

[39] S. XIE, S. Wong, T. Ng, and W. Lam, "Pedestrian crossing behavior at signalized crosswalks," *Journal of Transportation Engineering, Part A: Systems*, vol. 143, 05 2017.

[40] B. R. Kadali and P. Vedagiri, "Evaluation of pedestrian crossing speed change patterns at unprotected mid-block crosswalks in india," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 7, no. 6, pp. 832–842, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2095756418302381

[41] Daily Mirror, "From yellow to white," *Daily Mirror*, 2016, accessed: 2025-05-02. [Online]. Available: https://www.dailymirror.lk/caption_story/From-yellow-to-white/110-120174

[42] "National road master plan 2021–2030," Road Development Authority, Ministry of Highways, Democratic Socialist Republic of Sri Lanka, Colombo, Sri Lanka, Tech. Rep., 2021, accessed: 2025-05-03. [Online]. Available: https://rda.gov.lk/images/publications/pdf/National_Road_Master_Plan_2021-2030/Natinal%20Road%20Master%20Plan%202021-2030%20Main%20Report.pdf

[43] "Urban development authority planning and development regulations 2020," Ministry of Urban Development and Housing, Democratic Socialist Republic of Sri Lanka, Colombo, Sri Lanka, Government Regulation, 2020, issued under the Urban Development Authority Law No. 41 of 1978 and the Amendment Act No. 4 of 1982. [Online]. Available: https://documents.gov.lk/view/extra-gazettes/2015/11/1940-21_E.pdf

[44] Department of Transport and The Welsh Office and The Scottish Office and The Department of the Environment for Northern Ireland, "The design of pedestrian crossings," Department for Transport, London, UK, Local Transport Note LTN 2/95, April 1995, third impression 2005. [Online]. Available: https://assets.publishing.service.gov.uk/media/5a7d5cc0e5274a3356f2bc27/ltn-2-95_pedestrian-crossings.pdf

[45] D. R. Beddiar, B. Nini, M. Sabokrou, and A. Hadid, "Vision-based human activity recognition: a survey," *Multimedia Tools Appl.*, vol. 79, no. 41–42, p. 30509–30555, Nov. 2020. [Online]. Available: https://doi.org/10.1007/s11042-020-09004-3

[46] L. Arrotta, G. Civitarese, X. Chen, J. Cumin, and C. Bettini, "Multi-subject human activities: A survey of recognition and evaluation methods based on a formal framework," *Expert Systems with Applications*, vol. 267, p. 126178, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417424030458

[47] L. Gong, W. Yang, D. Man, G. Dong, M. Yu, and J. Lv, "Wifi-based real-time calibration-free passive human motion detection," *Sensors*, vol. 15, no. 12, pp. 32 213–32 229, 2015. [Online]. Available: https://www.mdpi.com/1424-8220/15/12/29896

[48] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using wifi," ser. SIGCOMM '15. New York, NY, USA:

Association for Computing Machinery, 2015, p. 269–282. [Online]. Available: https://doi.org/10.1145/2785956.2787487

[49] C. Kyrkou, "Yolopeds: efficient real-time single-shot pedestrian detection for smart camera applications," *IET Computer Vision*, vol. 14, no. 7, p. 417–425, Oct. 2020. [Online]. Available: http://dx.doi.org/10.1049/iet-cvi.2019.0897

[50] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics yolov8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[51] R. Korbmacher and A. Tordeux, "Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24 126–24 144, 2022.

[52] C. Premebida, O. Ludwig, and U. Nunes, "Lidar and vision-based pedestrian detection system," *Journal of Field Robotics*, vol. 26, pp. 696–711, 09 2009.

[53] T. Sarsodia, U. R. Bhatt, R. Upadhyay, and V. Bhat, "A survey on different application areas based on rss (received signal strength) and possible hardware and software tools for the collection of rss," in *2022 IEEE International Conference on Current Development in Engineering and Technology (CCET)*, 2022, pp. 1–6.

[54] M. I. M. Ismail, R. A. Dzyauddin, S. Samsul, N. A. Azmi, Y. Yamada, M. F. M. Yakub, and N. A. B. A. Salleh, "An rssi-based wireless sensor node localisation using trilateration and multilateration methods for outdoor environment," 2019. [Online]. Available: https://arxiv.org/abs/1912.07801

[55] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, Jun. 2019. [Online]. Available: https://doi.org/10.1145/3310194

[56] S. M. Hernandez and E. Bulut, "Wifi sensing on the edge: Signal processing techniques and challenges for real-world systems," *IEEE Communications Surveys Tutorials*, vol. 25, no. 1, pp. 46–76, 2023.

[57] J. Yang, "Awesome wifi csi sensing: A curated list of papers and resources on wifi csi sensing," https://github.com/Marsrocky/Awesome-WiFi-CSI-Sensing, 2022.

[58] T. G. Balbuzanov and B. I. Evstatiev, "Pedestrian presence detection system based on image processing," in *2019 IEEE 25th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, 2019, pp. 110–113.

[59] T. Choubisa, R. Upadrashta, S. Panchal, A. Praneeth, H. Ranjitha, K. Senthoor, A. Bhattacharya, S. Anand, M. Hegde, A. Kumar, P. V. Kumar, M. S. Iyer, A. Sampath, T. Prabhakar, J. Kuri, and A. N. Singh, "Challenges in developing and deploying a pir sensor-based intrusion classification system for an outdoor environment," in *2016 IEEE 41st Conference on Local Computer Networks Workshops (LCN Workshops)*, 2016, pp. 148–155.

[60] A.-M. Căilean, C. Beguni, S.-A. Avătămăniței, M. Dimian, and V. Popa, "Design, implementation and experimental investigation of a pedestrian street crossing assistance system based on visible light communications," *Sensors*, vol. 22, no. 15, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/15/5481

[61] A. Rudyk, A. Semenov, S. Baraban, O. Semenova, P. Kulakov, O. Kustovskyj, and L. Brych, "Influence of environmental factors on the accuracy of the ultrasonic rangefinder in a mobile robotic technical vision system," *Electronics*, vol. 14, no. 7, 2025. [Online]. Available: https://www.mdpi.com/2079-9292/14/7/1393

[62] H. Zhang, H. Du, Q. Ye, and C. Liu, "Utilizing csi and rssi to achieve high-precision outdoor positioning: A deep learning approach," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6.

[63] F. Abuhoureyah, W. Yan Chiew, A. S. Bin Mohd Isira, and M. Al-Andoli, "Free device location independent wifi-based localisation using received signal strength indicator and channel state information," *IET Wireless Sensor Systems*, vol. 13, no. 5, pp. 163–177, 2023. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/wss2.12065

[64] T. Z. Chowdhury, C. Leung, and C. Y. Miao, "Wihacs: Leveraging wifi for human activity classification using ofdm subcarriers' correlation," in *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2017, pp. 338–342.

[65] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Wiwho: Wifi-based person identification in smart spaces," in *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, 2016, pp. 1–12.

[66] "Integrated sensing and communication emerging technology initiative," IEEE Communications Society. [Online]. Available: https://isac.committees.comsoc.org/

[67] IEEE Standards Association. (2024) The evolution of wi-fi technology and standards. [Online]. Available: https://standards.ieee.org/beyond-standards/the-evolution-of-wi-fi-technology-and-standards

[68] S. Tan, Y. Ren, J. Yang, and Y. Chen, "Commodity wifi sensing in ten years: Status, challenges, and opportunities," *IEEE Internet of Things Journal*, vol. 9, no. 18, pp. 17 832–17 843, 2022.

[69] E. Soltanaghaei, R. A. Sharma, Z. Wang, A. Chittilappilly, A. Luong, E. Giler, K. Hall, S. Elias, and A. Rowe, "Robust and practical wifi human sensing using on-device learning with a domain adaptive model," in *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*, ser. BuildSys '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 150–159. [Online]. Available: https://doi.org/10.1145/3408308.3427983

[70] Y. Li, "Ofdm for wireless communications: techniques for capacity improvement," in *ICCT'98. 1998 International Conference on Communication Technology. Proceedings (IEEE Cat. No.98EX243)*, vol. 2, 1998, pp. 5 pp. vol.2–.

[71] Z. Yang, K. Qian, C. Wu, Y. Zhang, G. Zhang, Y. Zheng, and G. Chi, "Hands-on wireless sensing with wi-fi: A tutorial," https://tns.thss.tsinghua.edu.cn/wst/, 2022, accessed: 2025-05-01.

[72] S. ten Brink, "Enhancing wi-fi communication with effective csi approximations," *Research Outreach*, no. 137, 2023. [Online]. Available: https://researchoutreach.org/articles/enhancing-wi-fi-communication-with-effective-csi-approximations/

[73] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *Computer Communication Review*, vol. 41, p. 53, 01 2011.

[74] GGS Data AB, "5300 series laptop accessories specification sheet," http://www.ggsdata.se/laptops/bilder/accessories/5300.pdf, n.d.

[75] Intel Corporation, "Intel® wifi link 5300," https://www.intel.com/content/www/us/en/support/products/70971/wireless/legacy-intel-wireless-products/intel-wireless-series/intel-wifi-link-5300.html, n.d., accessed: 2025-05-03.

[76] B. S. da Silva, G. T. Laureano, A. S. Abdallah, and K. V. Cardoso, "Widmove: Sensing movement direction using ieee 802.11n interfaces," in *2018 IEEE Canadian Conference on Electrical Computer Engineering (CCECE)*, 2018, pp. 1–4.

[77] S. Kato, T. Fukushima, T. Murakami, H. Abeysekera, Y. Iwasaki, T. Fujihashi, T. Watanabe, and S. Saruwatari, "Csi2image: Image reconstruction from channel state information using generative adversarial networks," *IEEE Access*, vol. PP, pp. 1–1, 03 2021.

[78] A. Zhuravchak, O. Kapshii, and E. Pournaras, "Human activity recognition based on wi-fi csi data -a deep neural network approach," *Procedia Computer Science*, vol. 198, pp. 59–66, 2022, 12th International Conference on Emerging Ubiquitous Systems and Pervasive Networks / 11th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050921024509

[79] Espressif Systems, "Esp-csi: Applications based on wi-fi channel state information," https://github.com/espressif/esp-csi, 2025.

[80] M. Atif, M. Shapna, H. Ko, and B. Yoo, "Wi-esp—a tool for csi-based device-free wi-fi sensing (dfws)," *Journal of Computational Design and Engineering*, vol. 7, pp. 644–656, 10 2020.

[81] S. M. Hernandez and E. Bulut, "Lightweight and Standalone IoT Based WiFi Sensing for Active Repositioning and Mobility," in *21st International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM) (WoWMoM 2020)*, Cork, Ireland, Jun. 2020.

[82] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A survey on behavior recognition using wifi channel state information," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 98–104, 2017.

[83] Raspberry Pi Foundation. (2025) Raspberry pi official website. [Online]. Available: https://www.raspberrypi.com/

[84] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your csi: A channel state information extraction platform for modern wi-fi chipsets," in *Proceedings of the 13th International Workshop on Wireless Network Testbeds, Experimental Evaluation & Characterization*, ser. WiNTECH '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 21–28. [Online]. Available: https://doi.org/10.1145/3349623.3355477

[85] L. Hattersley, "Raspberry pi 4 vs raspberry pi 3b+," *The MagPi Magazine*, 2019, accessed: 2025-05-03. [Online]. Available: https://magazine.raspberrypi.com/articles/raspberry-pi-4-vs-raspberry-pi-3b-plus

[86] G. Halfacree. (2023) Raspberry pi 5 review — hands-on with the most powerful raspberry pi yet. Accessed: 2025-05-07. [Online]. Available: https://www.hackster.io/news/raspberry-pi-5-review-hands-on-with-the-most-powerful-raspberry-pi-yet-57efaf61b10f

[87] J. Schäfer, B. R. Barrsiwal, M. Kokhkharova, H. Adil, and J. Liebehenschel, "Human activity recognition using csi information with nexmon," *Applied Sciences*, vol. 11, no. 19, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/19/8860

[88] T. Li, C. Shi, P. Li, and P. Chen, "A novel gesture recognition system based on csi extracted from a smartphone with nexmon firmware," *Sensors*, vol. 21, p. 222, 12 2020.

[89] "Intelligent Transport Systems (ITS); Vulnerable Road Users (VRU) awareness; Part 3: Specification of VRU awareness basic service; Release 2," European Telecommunications Standards Institute (ETSI), Tech. Rep. TS 103 300-3 V2.2.1, February 2023. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/103300_103399/10330003/02.02.01_60/ts_10330003v020201p.pdf

[90] Espressif Systems, "ESP-IDF Programming Guide: Wi-Fi APIs," 2024, accessed: 2025-05-08. [Online]. Available: https://docs.espressif.com/projects/esp-idf/en/stable/esp32/api-guides/wifi.html

[91] X. Wang, L. Gao, S. Mao, and S. Pandey, "Csi-based fingerprinting for indoor localization: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, 2017.

APPENDIX A

ABLATION ON FILTERING TECHNIQUES

To evaluate the effectiveness of different filtering techniques for denoising and isolate movement related information from noisy outdoor CSI data. An ablation study was conducted with the goal of analyzing multiple filters such as Butterworth, Hampel and Kalman techniques to evaluate the performance across participants using a simple random forest model trained on amplitude and sanitized phase as the inputs to test to classify between $Empty$ and $Movement$ class.

TABLE A.1: COMPARISON OF FILTERING METHODS ON AVERAGE ACCURACY FROM ABLATION TESTS

Filtering Method	**Average Accuracy (%)**
No Filter Applied	82.2 ± 09.9
Butterworth Filtering	80.6 ± 07.1
Hampel Outlier Removal	85.0 ± 10.1
Kalman Adaptive Filtering	84.8 ± 10.4
Hampel + Kalman Filtering	86.4 ± 10.3

The combination of Hampel outlier removal and Kalman adaptive filtering yield the highest performance as seen in Table A.1 evident to the nature of CSI data requiring adaptive filtering. The results highlight that the inclusion of a filter increases the overall performance of the model by up to 4%. The results of this ablation testing also highlights the need to have robust features that can assist to isolate further noise in the CSI data as the highest accuracy of 86.4 with a standard deviation of 10.3 remains modest for a result from a binary classification.

Although several filters were considered as visualized in Fig. A.1, only the combined adaptive filtering of Hampel and Kalman were selected for preprocessing due to their simplicity of implementation and for the application of real-time processing.

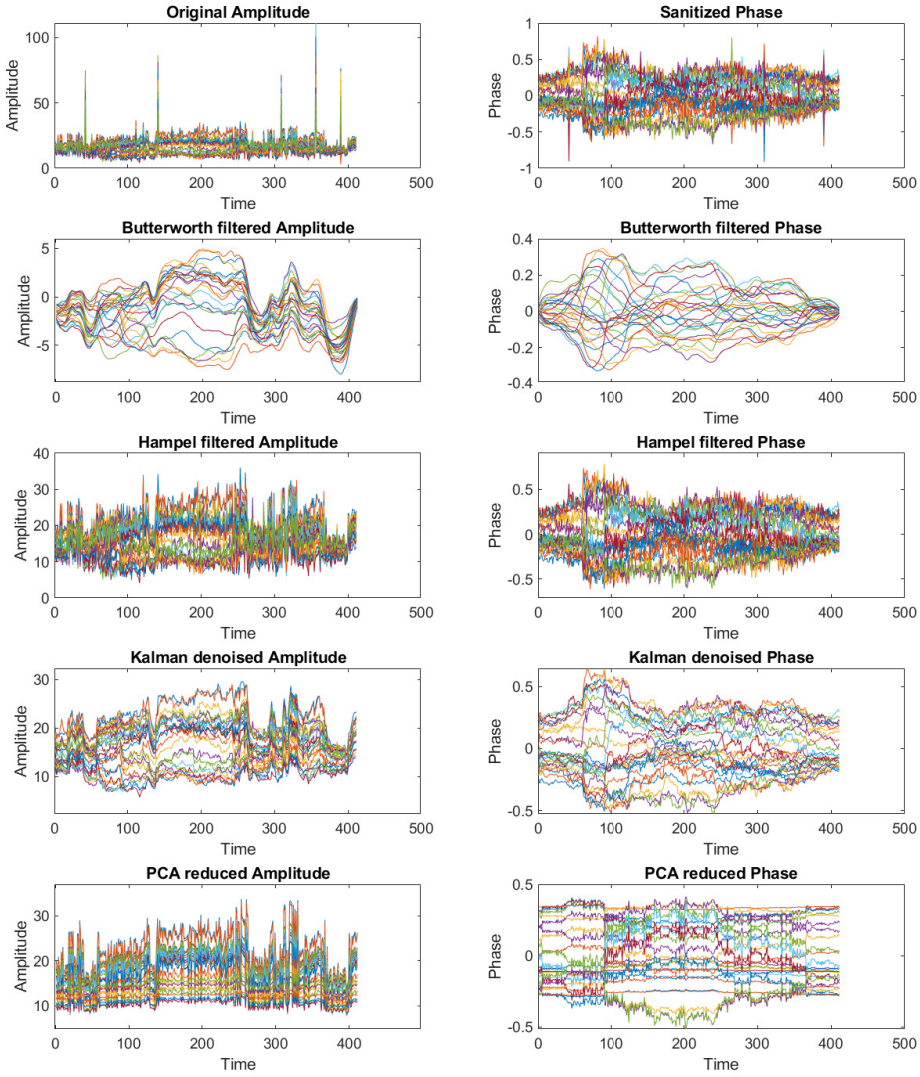


Fig. A.1: Visualization of different filtering methods applied to CSI data