UKRAINIAN CATHOLIC UNIVERSITY

BACHELOR THESIS

Human Activity Recognition based on WiFi CSI data

Author:
Andrii ZHURAVCHAK

Supervisor:
PhD. Oleg KAPSHII

A thesis submitted in fulfillment of the requirements
for the degree of Bachelor of Science

in the

Department of Computer Sciences
Faculty of Applied Sciences

Lviv 2020

Declaration of Authorship

I, Andrii ZHURAVCHAK, declare that this thesis titled, "Human Activity Recognition based on WiFi CSI data" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

UKRAINIAN CATHOLIC UNIVERSITY

Faculty of Applied Sciences

Bachelor of Science

Human Activity Recognition based on WiFi CSI data

by Andrii ZHURAVCHAK

Abstract

Using Wi-Fi Channel State Information (CSI) is a novel way of sensing and human activity recognition (HAR). Such a system can be used to ensure safety and security without any violence of privacy versus a vision-based approach.

The main goal of this thesis is to explore current methods and systems that use Wi-Fi CSI, conduct experiments to analyze how different hardware configurations affect data and possibility to detect human activity, collect datasets and build a classification model for the HAR task.

Performed eight experiments and dataset was was collected in three different rooms. Built and trained – an InceptionTime and LSTM-based classification model. We show a full pipeline of building a Wi-Fi CSI-based system. The results show the 61% human activity classification accuracy (80-82% for particular activities) by using simple-commodity Wi-Fi routers. Other systems show better results, but we concluded that their dataset is not representative enough, so we propose a more realistic data-collection approach for the Wi-Fi CSI-based HAR problem. Also, we show the problems of our methods and describe possible ways to deal with them.

Source code and dataset [1] are publicly available and can be used for future work in other studies.

[1] Github repository: github.com/Retsediv/WIFI_CSI_based_HAR

Acknowledgements

First of all, I express my gratitude to my supervisor Oleh Kapshii from the Ukraine Cypress Semiconductor company for his guiding me throughout this thesis, sharing his knowledge, help, and fun meetings we had.

Also, I thank the Ukrainian Catholic University and UCU Applied Science Faculty for the opportunity to study in a great university and high-quality academic program, which played an essential role in shaping my professional interests as well as personal formation.

My special thanks are extended to the ASR group members of the Ukraine Cypress Semiconductor company.

Contents

List of Figures

List of Abbreviations

HAR	**H**uman **A**ctivity **R**ecognition
ANN	**A**rtificial **N**eural **N**etwork
DNN	**D**eep **N**eural **N**etwork
RNN	**R**ecurent **N**eural **N**etwork
GMM	**G**aussian **M**ixture **M**odel
CNN	**C**onvolutional **N**eural **N**etwork
CSI	**C**hannel **S**tate **I**nformation
LIDAR	**Li**ght **D**etection **A**nd **R**anging
OFDM	**O**rthogonal **F**requency **D**ivision **M**ultiplexing
MIMO	**M**ultiple **I**nput **M**ultiple **O**utput
RSSI	**R**eceived **S**ignal **S**trength **I**ndicator
BLE	**B**luetooth **L**ow **E**nergy
AP	**A**ccess **P**oint
Tx	**T**ransmitter
Rx	**R**eceiver

Chapter 1

Introduction

1.1 Motivation

Modern people are using various sensors even if they do not understand that. Proximity sensors control garage light, LIDAR is actively being developed for self-driving car [1] [2], cameras watch you in the airport or capacitive sensors as buttons in your smartphone [3]. We would not be humans if we could not feel and sense. At the same time, the technologies are growing and we are looking for new ways to sense the surrounding world and get more data from it.

In this desire, the industry is rethinking the use of already familiar devices and technologies. The last decade two essential things have been rapidly developing this work would not be possible without:

- Wireless communication and Wi-Fi Channel State Information that provide information about the Wi-Fi radio waves propagation.
- Machine Learning to process that information and make use of it.

Today, there is a Wi-Fi router nearly in every house. So, it becomes a perfect sensor as we can re-use the existing tool with no production cost.

This thesis shows how a combination of Wi-Fi Channel State Information and Machine Learning can be used for Human Activity Recognition(HAR). By using such a system, we can determine if a person drops down with a heart attack, enhance security systems by detecting people present in the total darkness or even check whether a person is breathing and decide whether to notify relatives and call the emergency.

1.2 Problem framing

1.2.1 Human Activity Recognition

Human Activity Recognition has developed tremendously in recent years. Usually, people face it in medical institutions, where patients have special beacons for their location tracking, accelerometers, and gyroscopes to detect movement. Also, it is well combined with different IoT solutions, where activity recognition is used for efficient energy consumption in a smart home. In such a system, a person usually has to wear a device equipped with some sort of motion sensors. It continuously collects data and process it either locally or sends it to a server that performs the feature selection and then some supervised machine learning algorithm is used for activity classification.

1.2.2 Wi-Fi RSSI and CSI

A usual home Wi-Fi router is generally used only for communication. Wi-Fi works using radio waves, and when people are in the Wi-Fi range, signals propagate differently. This effect can be empirically observed by Wi-Fi channel measurements. Based on this fact, these Wi-Fi channel changes can be used for human activities classification. Primarily, RSSI (Received Signal Strength Indicator) and CSI (Channel State Information) are used for understanding the environment and based on it, finding better configuration for Wi-Fi communication. But, they can be used for other purposes as well:

- The RSSI has been actively used for active localization based on the Wi-Fi fingerprinting technique [4] or as a metric for mobile devices passive tracking [5]. However, when a person is not directly between the Access Point (AP) and Wi-Fi device, the RSSI can not accurately capture real changes in signal and human movements. This is because the RSSI is not a stable metric, even when there is no dynamic change in the environment. [6]
- CSI is based on Multiple-Input Multiple-Output (MIMO) that provides high throughput of the wireless data traffic and Orthogonal Frequency-Division Multiplexing (OFDM). It provides information for each transmitter and receiver antenna pair at each carrier frequency. CSI represents how wireless signals propagate from the transmitter to the receiver at certain carrier frequencies along multiple paths. For example, CSI amplitude variations in the time domain have different patterns for different humans, activities, gestures, etc., that can be used for HAR. [7]

1.2.3 Using Wi-Fi CSI for HAR

The RSSI suffers from dramatic performance degradation in complex situations due to multipath fading and temporal dynamics, so, it does not fit well for the HAR problem and we will focus on CSI only. However, using CSI in HAR systems has its pitfalls:

1. The performance hugely depends on the environment and the more complex it is, the worst accuracy we will get.
2. CSI does not perform well in new environments.
3. There are a few datasets publicly available ready for this particular task. Yet, it is hard to use an ML model trained on a dataset not collected by your Wi-Fi device due to hardware differences: different number of antennas, sensitivity, frequency etc.

1.3 Bachelor thesis goals

1. Explore the existing approaches on how to use Wi-Fi CSI for HAR.
2. Analyse the Wi-Fi CSI and how it depends on the system configuration and environment.
3. Collect a dataset for HAR based on CSI data.
4. Develop an ML model for the HAR problem based on CSI data from commodity Wi-Fi routers.

1.4 Thesis structure

The rest of the thesis is organized as follows:

Chapter 2. Related Works
Analyzes previous works and research conducted in the area of our problem.

Chapter 3. Background information
Specifies the WiFi CSI and machine learning background for this project:

- how WiFi radio waves propagate
- how to capture that propagation
- what deep learning is
- how recurrent networks work.

Chapter 4. Experiments and data retrieval
Describes experiments with WiFi CSI and dataset collection approach.

Chapter 5. Proposed solution
Describes our solution in detail.

Chapter 6. Results and summary
Summarizes the achieved results and their comparison and shares our ideas for future work.

Chapter 2

Related Works

2.1 Human Activity Recognition overview

HAR can be defined as a method of recognizing a human activity based on the information received from various sensors. It has always been a challenging problem that requires a solution. However, we see it as a trendy research topic due to the variety of sensors, rapid development of machine learning, data analysis and Internet of Things (IoT). [8]

HAR is mainly used for eldercare and healthcare as an assisting technology in a combination with other technologies like IoT. Along with that, there is a huge perspective of usage HAR for gaming, security, and enhancing human-computer interaction. [9]

2.1.1 Classification by activity type

However, HAR is not only about detecting a human activity at a particular moment. For example, it also includes Gestures Recognition [7], which gives you the ability, for example, to control your devices only by moving your hands in a specific way. [8] presents an overview of the HAR techniques (Fig. 2.1) divided into the three main categories and into ten sub-categories.

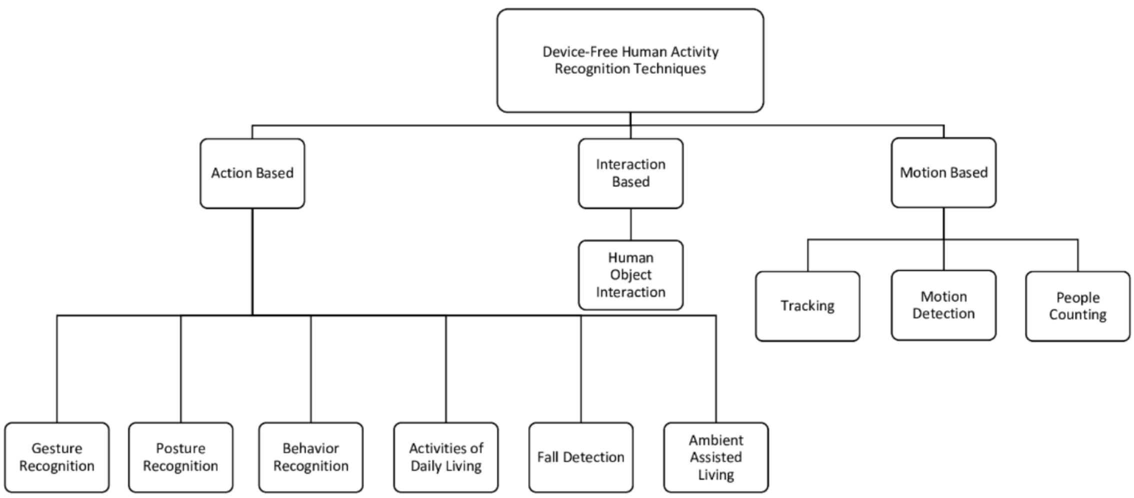


FIGURE 2.1: Overview and classification of human activity recognition techniques from [8]

HAR could be classified in another way - by the used instrument, which is used for collecting data and its further analysis. Different surveys describing and summarizing the work done in the activity recognition field can be divided into three categories.

2.1.2 Classification by instruments and methods

[10] presented a comprehensive survey showing various aspects of the sensor-based activity recognition. They divide the current approaches into:

- Vision-based – video cameras and computer vision techniques are commonly used. This is a data-driven approach that uses the generative and discriminative modelling.
- Sensor-based – emerging sensor network technologies that generate time-series data of system parameters are used. This is a knowledge-driven - logic, ontology and mining-based approaches.

In these approaches, the sensor can be used by the actor as a wearable device(e.g. GPS, BLE beacon, accelerometer) or as an object that constitutes the activity environment - namely "dense sensing".

Sensor-based

Accelerometer sensors are the most used wearable sensors for HAR. [10] They are extremely effective in classifying repetitive body motions, such as running, sitting, standing, walking etc. and widely used in deep learning-based HAR [11] [12]. Gyroscopes and magnetometers are frequently used along with an accelerometer, as they have been seen as an effective combination for recognition activities of daily life and sports. [13] They can be even used without extracting statistical or frequency data from the movement data, as the original signal can be directly used as an input for the network.

Vision-based

Vision-based HAR has long been a research focus due to the importance of the problem it can solve: surveillance, robot learning, and anti-terrorist security. Researchers are trying to combine different sources of information, such as a video camera, sound, and infra-red [10]. [14] present a survey of existing research with the vision-based approach for HAR. It divides it into two approaches:

- Unimodal methods use data from single modality and are classified as stochastic, rule-based, space-time based, and shape-based methods.
- Multimodal approaches use data from different sources and are classified as behavioural, effective, and social-networking methods.

2.1.3 Machine learning techniques

Different datasets are used for each of the methods wherein data is collected by different means such as sensors, images, accelerometer, gyroscopes, and the placement of these devices at various locations. The results obtained by each technique and the type of a dataset are then compared. The Machine learning techniques like decision trees, K-nearest neighbours, support vector machines, hidden Markov models are the most commonly used classical methods for HAR as well as deep neural networks (DNN) techniques like artificial neural networks (ANN) together with its subclasses - convolutional neural networks (CNN) and recurrent neural networks (RNN).

2.2 Wi-Fi-based HAR

While most HAR applications are based on using either wearable devices or video camera data, researchers are trying to find new ways of solving this problem to:

1. Remove any wearable devices in the system to facilitate the usage.
2. Remove video camera usage to avoid people's privacy violation.

The advantages of using Wi-Fi radiowaves information approach:

1. No privacy violation.
2. The ability to "see" in the darkness and bend obstacles.
3. No need in a special HW infrastructure, as Wi-Fi is a widely-spread technology.

Wi-Fi Channel State Information is commonly used as one of the most accurate and full data source about the WiFi communication channel so that we focus only on it.

People tracking and localization

Widar [15], IndoTrack [16] and Chronos [17] systems are developed for indoor localization based on WiFi CSI data. These systems use regular routers and report an average accuracy of 20-40 cm. [15] proposes switching from CSI to PLCR(Path Length Change Rate) to extract the change rate of each signal propagation path, from which we can directly estimate both user's velocity and location. This way, they enable precise tracking of user's moving velocity and location using COTS (commercially available off-the-shelf) Wi-Fi devices.

[16] proposes Doppler-MUSIC method to estimate the speed of a moving human's reflection path-length change. The Doppler-MUSIC method addresses the varying intervals issue of CSI samples, and the random CSI phase offsets problem due to the unsynchronized Wi-Fi transmitter and receivers. It includes a novel method to measure the subtle Doppler shift introduced by human movements accurately. This technique looks very promising from the data interpreting side and can be used together with the ML approach to get better results All the systems use classical Machine Learning methods, with no deep learning and show pretty solid results. However, mostly, all tests were performed in a laboratory environment and results may not be highly reliable. More testing is required.

People detection and pose estimation

[18] shows that people detection problem is generally solved. They propose PADS system which detect humans moving with a dynamic speed using CSI data. PADS accurately alarms human movements by 99% on average with almost no false-negative errors. Moreover, it consistently achieves excellent performance in cases of walking humans with dynamic speeds and outperforms existing approaches.

CSI measurements contain significant phase noise random and occasionally amplitude outliers, so is planned to be removed by the phase sanitization module. Meanwhile, an outlier filter based on the Hampel identifier algorithm is applied to eliminate the outlier observation in the amplitude sequences of CSI. PADS proposes to exploit the amplitude and phase information of CFR (Channel Frequency Response) simultaneously. To avoid the influence of diverse transmitting power in specific scenarios, they devise a novel feature using the respective three maximum eigenvalues of the correlation matrices of amplitude and phase information over a particular time window. The features are then fed to an inference model to alarm user's movements.

The classification algorithm is based on SVM (Support Vector Machine), and along with great results, they did extensive experiments showing that this problem is solved and the system is robust.

Intrusion detection

Robust Device-Free Intrusion Detection (RDFID) system [19] leveraging fine-grained CSI and aims to detect possible intrusions. Their key idea is to deal with the limitations state-of-the-art human-detection techniques have: can only detect a human walking with a regular pattern, require on-site calibration, fail in detecting intruders in security systems.

The noises in the signals are removed by Principle Component Analysis (PCA) and a low pass filter. They extract a robust feature of the frequency domain utilizing Continuous Wavelet Transform (CWT) from all subcarriers. RDFID captures the changes from the whole wireless channel, and a threshold is obtained self-adaptively, which is calibration-free in different environments and can be deployed in smart home scenarios.

Gaussian Mixture Model (GMM) was chosen as a baseline model based on the fact that the distribution of the wavelet variance when there is human motion different from that of static scenario. It has two Gaussian models: for static and human motion model.

They achieved high intrusion detection accuracy with simple hardware. However, based on their analysis, the relative location of the intruder and transceivers can profoundly affect the system detection accuracy. Also, the system was tested only with one intruder, while in real life, there may exist more.

CSI is a rich source of information with its problems dedicated firstly to the environment, where the system is located. Based on that, researchers are trying to extract the most useful information and features from it, illuminate noise and environment influence. Already, there are a lot of systems showing the perspective of using the CSI in HAR area. However, this is a new approach which needs more testing, increasing reliability and robustness.

Chapter 3

Background Information and Theory

3.1 Wi-Fi Channel State Information

Commonly, we use Wi-Fi for communication purposes, but it could also be used for solving the HAR problem. Human movement within the range of a Wi-Fi network affect the multipath propagation [20]. Why this happens: a Wi-Fi router works with radio waves that propagate through its location environment, and when a person is on the way of the radio wave, it get round this person and its characteristics will change. Moreover, the Wi-Fi channel also consists of a signal reflected by static objects in some environments, such as furniture or others. As the Channel State Information represents the signal propagation effect in the channel [21], these additional reflection caused by different activities are observed as in Fig. 3.1.

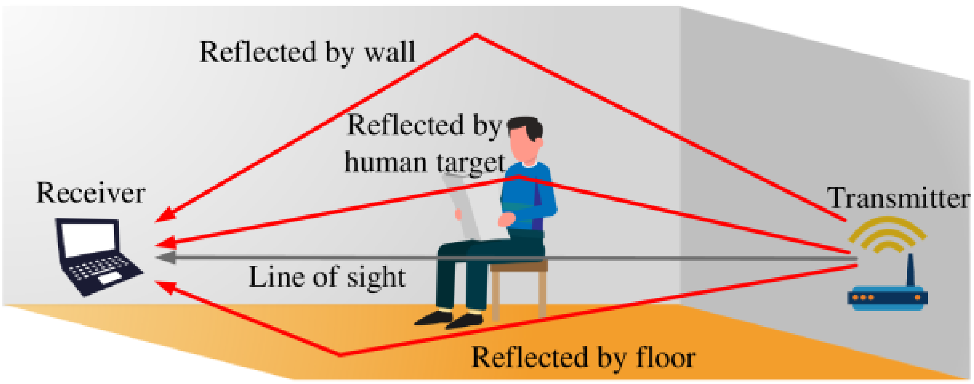


FIGURE 3.1: Wi-Fi radio waves propagation in a room from [22]

More precisely, CSI represents the change of the signal from the transmitter (denote it as x) to the receiver (denote it as y) and the Wi-Fi channel in the frequency domain can be represented as:

$$y = Hx + n$$

where H is a complex matrix consisting of CSI values and n is the channel noise. The CSI is estimated for each Orthogonal Frequency Division Multiplexing (OFDM) subcarrier links [23]. OFDM splits the total frequency spectrum into 56 or 114 frequency subcarriers for a channel bandwidth of 20 and 40 MHz respectively. The CSI for each subcarrier is:

$$h = |h|e^{j\theta}$$

where $|h|$ represents the amplitude and θ the phase. To measure the CSI, the transmitter sends Long Training Symbol (LTS), which contains pre-defined information for each subcarrier. When the receiver receives the LTS, it estimates the CSI having the difference between the original and received LTS. However, in real-world systems, the CSI is affected by a multi-path channel, receiver/transmit processing, hardware and software errors. [7]

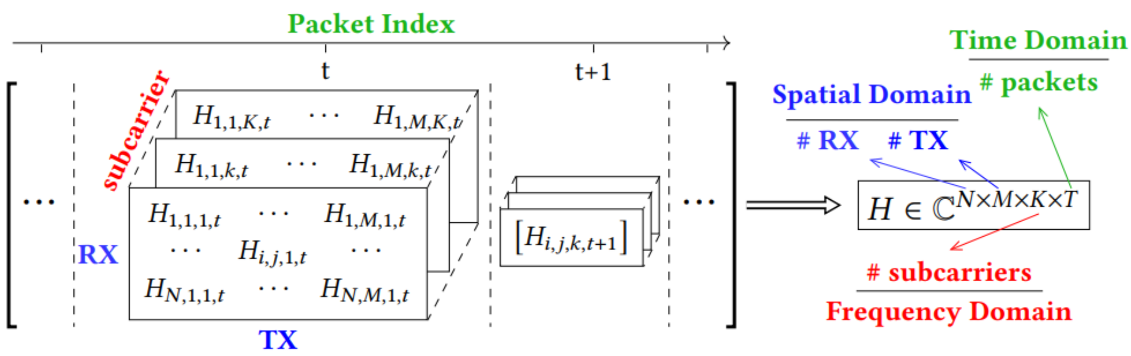


FIGURE 3.2: 4D CSI tensor is a time series of CSI matrices of MIMO-OFDM channels

In our research, the CSI matrix consist of 114 complex H matrices for 5 GHz frequency band or 56 for 2.4 GHz of dimension $N_{Tx}xN_{Rx}$. 56 and 114 are defined by the number of subcarriers the CSI extraction tool can handle (on 2.4 and 5 GHz). Its structure is shown in Fig. 3.2.

3.2 Neural Networks

Neural Network is the mathematical model inspired by a study of the biological structure and organization of biologic neural networks in the human brain. More precisely, it is a sequence of neurons, connected via synapses. Based on such architecture, the model can analyze, remember or even create new information as it works with the human brain, which consists of millions of neurons which exchange information based on the electrical impulses.

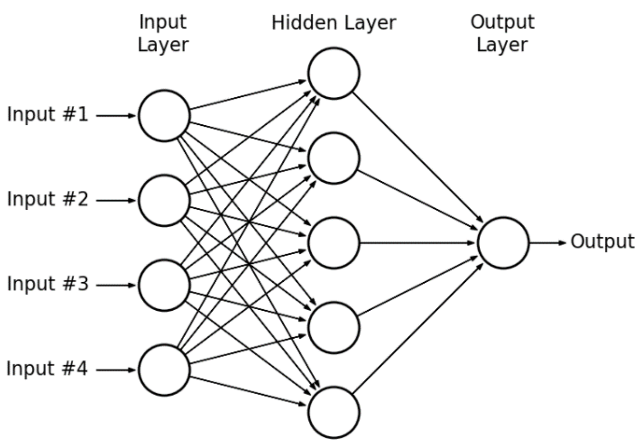


FIGURE 3.3: Multilayer perceptron architecture

One of the most straightforward NN architectures is the multilayer perceptron (shown in Fig. 3.3). It consists of three layers: input, hidden, and output. Each circle represents a neuron, each line – the connection between the neurons. It differs from other architectures because its neurons from two adjacent layers are fully interconnected. Each neuron has an activation function, which maps the weighted inputs to the output. The weights are a crucial factor, which NNs are trying to learn.

The universal approximation theorem [24] states that "a standard multilayer feedforward network with a locally bounded piecewise continuous activation function can approximate any continuous function to any degree of accuracy if and only if the network's activation function is not a polynomial". From this perspective, NNs are very powerful as they can approximate any model even with the most straightforward architecture as the multilayer perceptron with a single hidden layer it. However, the more difficult problem is, the more data and computational power it requires.

The training consists of two steps:

1. The trained data is forwarded through the network. The output and error between the network prediction and the target values from the dataset are calculated.
2. This error backpropagates through the network and weight are updated concerning the error gradients [25].

This cycle is repeated many times until the network is stable and outputs accurate results.

3.2.1 Recurrent Neural Networks

The standard multilayer perceptron approach would not work well in sequence data. Let's imagine, that we want to summarize the text.

- Firstly, the inputs and outputs will not have the same length, as the article and summary lengths can vary greatly.
- Secondly, the studied functions will not be distributed at different positions of the text, which will lead to a decrease in performance.

This is why Recurrent Neural Networks (RNNs) are very popular for serial data such as sentences or music.

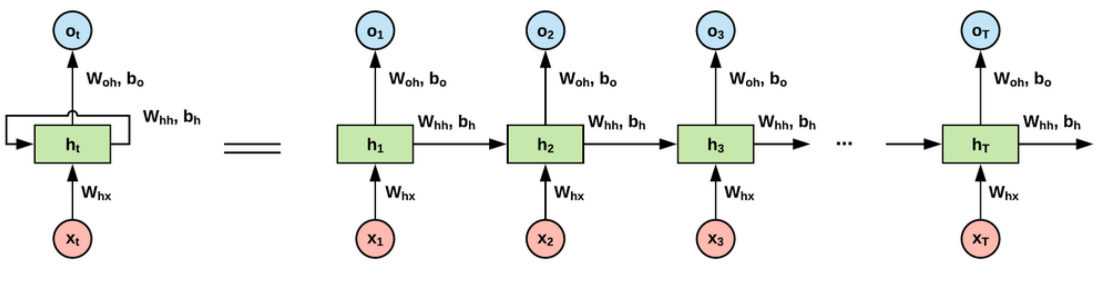


FIGURE 3.4: RNN architecture (compressed and unrolled schemes)

The basic RNN cell is shown in Fig. 3.4. As the multilayer perceptron it has input cell (x), hidden (h) and output (o). This structure allows the RNN to associate information from previous steps with the current step [25]. However, if the RNN has many layers (deep), it will be prone to a vanishing gradient.

Bidirectional RNNs

In certain types of tasks, it is essential to make predictions taking into account both past and future context. Those tasks include speech and handwriting recognition. Bidirectional Recursive Neural Networks [26] were invented to deal precisely with this problem. The idea is to feed data to the RNN network simultaneously in a positive and negative time direction.

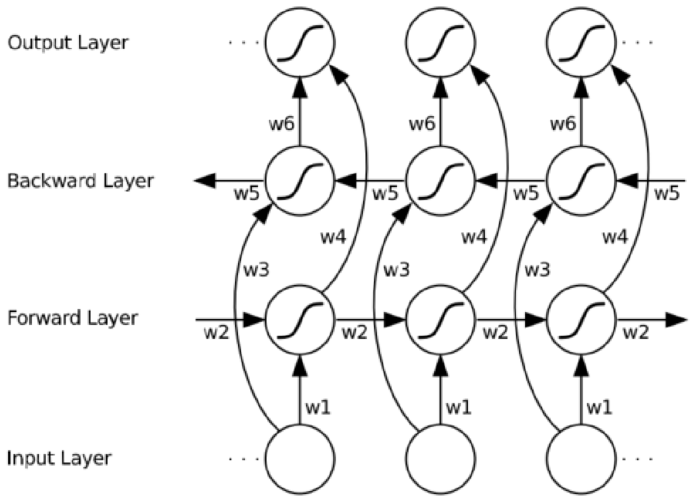


FIGURE 3.5: Bidirectional RNN architecture

Gated RNNs

Human thoughts have a beneficial property – persistence. It means that we do not start thinking from zero reading a new article or paragraph, you understand each word based on the previous words no matter how far they were before. However, vanilla RRNs cannot handle it, and when we are interested in understanding long-term dependencies, it becomes unfeasible due to the vanishing gradient problem [27].

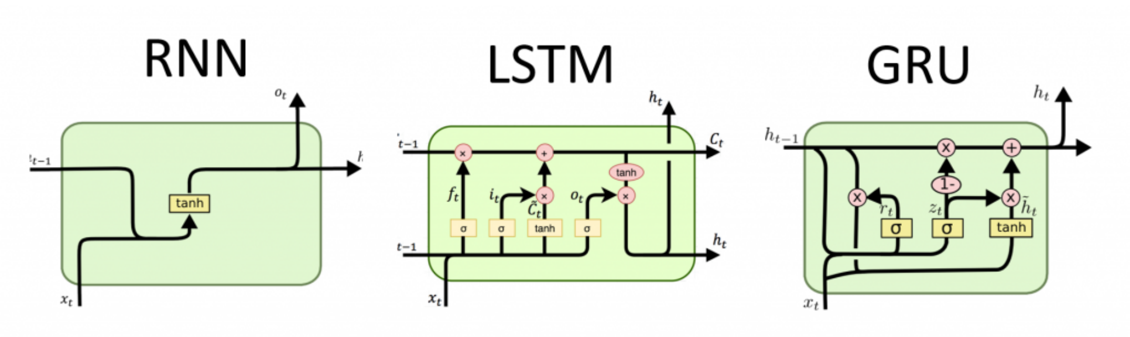


FIGURE 3.6: RNNs architectures (Vanilla RNN, LSTM and GRU)

Long short-term memory (LSTM) is a special kind of architecture of recurrent neural networks, capable of learning long-term dependencies [28]. LSTMs are designed specially to avoid the vanishing gradient problem and handle long-term dependencies. The key to LSTMs is the cell state. The cell state tries to keep the information untouched along with gates responsible for deleting or modifying the information, which is going through the call state.

Another popular approach to solve this problem is Gated Recurrent Units (GRU) [29]. Its "forgetting" and entry gates are combined into one "update gate". Besides, the state of the cell is combined with the hidden state. The resulting model is simpler than a standard LSTM, and its popularity is steadily increasing. LSTM and GRU architectures along with "vanilla" RNN are shown in Fig. 3.6.

Chapter 4

Experiments and data retrieval

4.1 Hardware setup

We have found three options to get CSI data from the router: Linux 802.11n CSI Tool [30] Atheros CSI Tool [31], and Nexmon Channel State Information Extractor [32]. All of them are custom modified firmware and open-source Linux wireless drivers to be installed on Wi-Fi devices instead of the vendor-provided.

Linux 802.11n CSI Tool is the oldest one and is already used in more than 500 publications. However, it supports only Intel 5300 Network Interface Controllers (NICs), which are quite old and hard to purchase. Also, using that NIC is not representative as it could differ much from the modern ones.

Nexmon CSI Extractor allows extracting CSI of OFDM-modulated Wi-Fi frames on a per-frame basis with up to 80 MHz bandwidth on the Broadcom Wi-Fi Chips, which are used on smartphones (Nexus 5, Nexus 6P and others), Raspberry Pi B3+/B4 or Asus RT-AC86U router. It improves continually, but due to the lack of supported devices and information, we decided not to use it

Atheros CSI Tool is an open-source 802.11n measurement and experimentation tool. It enables the extraction of detailed PHY wireless communication information from the Atheros Wi-Fi NICs, including the CSI, the received packet payload, and other additional information (the data rate, the timestamp, the RSSI of each antenna and others). There is a big community around it, which tested different Atheros NICs and did projects to get started easier. Based on that, we decided to use it in this research.

Two TP-link TL-WDR4300 routers were used as the receiver and the transmitter. Custom Atheros CSI Tool firmware was built and installed there to enable PHY (Physical) layer information extraction. They work in AP-Client Mode, so the transmitter and receiver are configured to work in Access Point (AP) mode and Client mode respectively. To create AP, we used OpenWRT (linux operating system targeting embedded devices) interface, which gives the ability to configure operating frequency, mode, channel bandwidth, and other parameters. After that, the client router has to associate with the created AP via an OpenWRT interface in the same way as we connect to the Wi-Fi network via phone. After that, they become in the same network and can freely communicate.

FIGURE 4.1: TP-link TL-WDR4300 router

The next step was to build and deploy the system, where the transmitter sends data to the receiver, which handles it, computes CSI and logs it for further analysis.

It was built based on [32] and [33]. As far as the Wi-Fi router does not have enough space to save a big amount of data, we upgraded the system, so it works as follows:

1. The transmitter sends data to the receiver.
2. The receiver gets the data and computes CSI.
3. The receiver sends raw CSI data to the user laptop.
4. The user's laptop handles data, stores it and visualizes on the fly.
5. Simultaneously with the previous point, the camera makes a photo of the room, where the experiments are done (was added for dataset collection and better labelling).

4.2 Experiments

The Wi-Fi network can be configured in different ways. In particular, it is possible to choose the frequency, channel, bandwidth, number of antennas, which affect the Wi-Fi channel and its throughput, bandwidth, speed and coverage. Based on the environment and goals, people configure it in a way to satisfy their needs. Our goal was to test the hypothesis, that Wi-Fi configuration, routers location and environment affect the CSI data and the ability to recognize human activity using that information. As well, we tested the possibility to detect human gestures (arm and palm moving) using CSI data because in that case, the signal change will be dramatically smaller due to the smaller objects moving. Finally, based on the conducted experiments we recommended choosing better network configuration and provide a summary of our observations.

The experiments are grouped into two categories:

1. *Activities* - To check how different Wi-Fi configurations affect the possibility to classify activities (Experiment 1-5). There are lot of commonly used activities, but it is hard to analyse all of them. So, we focused only on the two of them: walking and absence of activity. We considered configuring the following system parameters: channels, bandwidth, base frequency (2.4 GHz, 5 GHz), and different antennas quantity (Experiment 1-4). The goal of these experiments was to test different system parameters configurations, understand how they influence the system itself and data we receive as well as the possibility to detect human activity based on the data. As a result, recommendations on the system configuration for HAR were given. Along with that, the possibility of detecting human activity in the room without a routers was tested (Experiment 5).
2. *Gestures* - To test the possibility to classify or at least to find some pattern of Wi-Fi gestures among CSI data (Experiment 6). Gestures are a form of non-verbal communication in which visible bodily actions communicate a particular message. As well as activities, there are lots of commonly-used gestures, so we focused only on two: a moving(horizontally and vertically) palm and arm. The goal is to understand whether there is a pattern of gesture on the data we collect during the experiment.

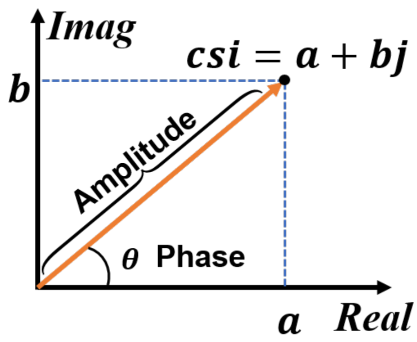


FIGURE 4.2: Single CSI complex value

CSI consists of complex numbers in the form of, for example $csi = a + bj$, as shown in Fig. 4.2. Each such an entry represents the amplitude and phase change, that happened by original signal sending from the transmitter to receiver. In an environment with no human activity, still the signal changes due to its reflections from walls, furniture and other objects. Moreover, in such an environment, this shift will be constant. However, this data is useless in our task because we cannot separate the change in phase and amplitude from chair and human. We have some resulting change value, which consists of all reflections, obstacles avoidance and human activity. Due to that fact, isolated amplitude and phase shift samples are not informative for us because they do not contain information about previous states and we cannot understand whether there was any change in the system (environment) or not. Based on that, we are interested in the data variations in time, taking into account the previous situation and searching for the patterns in data.

We are interested in patterns where the amplitude and phase in time change. For example, during walking, the amplitude and phase are changing often and rapidly, as the person changes their pose and location in a room. So, in this case, we are interested in fast and changeable variations. This is the pattern we are looking for. However, during sitting, the person does not move much, and CSI data is quite stable. So, in this case, the pattern we would like to look for is how the CSI values have changed from the previous activity to the current one – the transition between two activities.

We visualize CSI data to be able to see the data change in time and look for the pattern through it. An example of such a graph, is shown in Fig. 4.3. It consists of 4 sub-charts showing the CSI values for different system setup (Wi-Fi channel in that case). Let's look at the graph in the top-right corner. Its *y-axis* corresponds to the subcarriers, the *x-axis* is a timeline and represents each CSI packet number. The colour of each point on that graph is the value of the amplitude corresponding to the colour bar attached to the right corner. In this example, the walking area is labelled on the *x-axis* and we can clearly see the pattern between "walking" and "no activity" data. If there are no labels, it means that no activity was performed at that time. It's worth to mention that such a graph represents CSI data for a single antenna pair. Also, at the bottom of Fig. 4.4b the single carrier charts are shown. The only difference with the previous is that the single subcarrier is visualized by using a simple line chart.

The system configuration information is attached as a table to each experiment, where the number next to Rx (Receiver) and Tx (Transmitter) means antennas quantity used by transmitter and receiver correspondingly.

The goal of the experiments from "Activities" group is the selection of the router configuration, including (a) channels, (b) bandwidth, (c) frequency, (d) antenna number, and (e) locations of transmitter and receiver (same room / different rooms). The methodology for the selection is comparing the amplitude and phase measured over different channels, bandwidths etc. for the same human motion pattern (walking), plotting results on a graph and choose the configuration that yields more contrasted view so that the activity pattern can be easier observed. The same methodology is used for the experiments from "Gestures" group, but its goal is to understand the possibility to detect gestures based on Wi-Fi CSI data.

Experiment 1: Wi-Fi channels

System configuration	
Bandwidth	40 MHz
Channels	3, 7, 10, 11
Frequency	2.4 GHz
Antennas	2Rx vs 2Tx

WiFi channels are smaller bands within WiFi frequency bands used by your wireless network to send and receive data. Depending on the frequency band the router is using, there is a specific number of channels to use: 11 channels in the 2.4 GHz frequency, 45 channels in the 5 GHz. It is worth noting that, channel numbers in 5 GHz frequency band are not numbered successively, and the biggest channel which is currently possible to use has the number of 173.

While setting up a WiFi network, it is recommended choosing 1, 6, or 11 channel. Those channels yield better WiFi performance than others because they are non-overlapping with others. Based on this fact, we assume, that the activity pattern can be better visible on one channel and worse on another. It was tested from the activity recognition perspective on channels 3, 7, 10, 11 (Fig. 4.3).

The area with the walking activity is labelled at the bottom of each graph. As we can see from the results, the pattern of the walking area during the experiment on channel 11 has a bigger difference with the non-activity area (yields more contrasted view) compared to other channels. This can be exactly due to the fact, that channel 11 does not overlap with other channels.

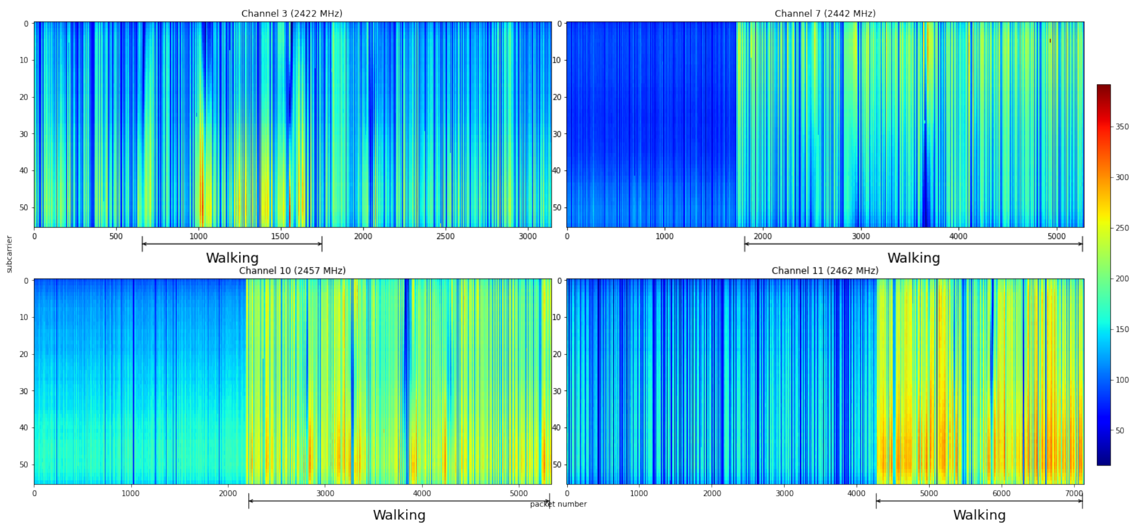


FIGURE 4.3: WiFi 3, 7, 10 and 11 channels comparison (amplitude)

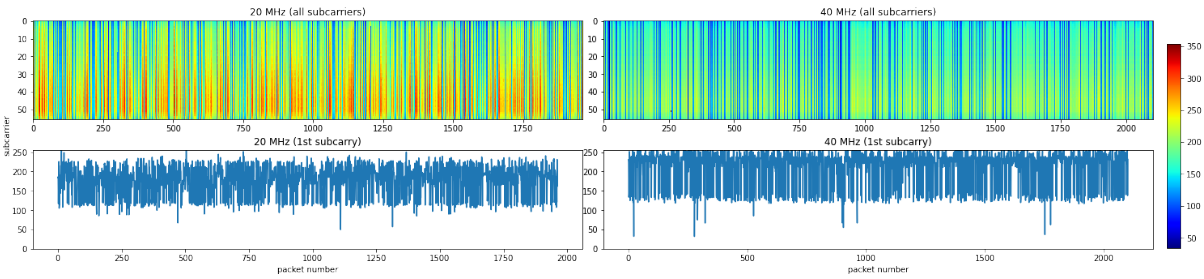
Experiment 2: Bandwidth 20 and 40 MHz

System configuration	
Bandwidths	20, 40 MHz
Channel	11
Frequency	2.4 GHz
Antennas	2Rx vs 2Tx

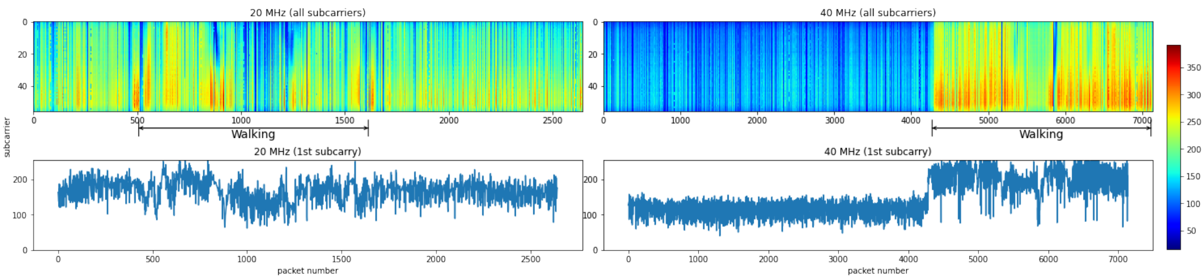
Working with IEEE 802.110n (a wireless-networking standard that uses multiple antennas to increase data rates), there is the possibility of using signal bandwidths of either 20 MHz, 40 or 80 MHz (the latter frequency is not supported by our hardware). A higher bandwidth corresponds to a higher data throughput. However, it reduces the number of channels that can be used. It was tested from the activity recognition perspective on bandwidths 20 and 40 MHz (shown in Fig. 4.4a and Fig. 4.4b).

As we see from the results in Fig. 4.4b, the walking area during the experiment with a bandwidth 40 MHz is much better highlighted and the difference between the walking and "silence" areas is bigger than at 20 MHz.

However, 40 MHz bandwidth adds more noise to the system (Fig. 4.4a). It is clearly seen in the bottom graphs, where a single first subcarry is visualized, the variance of the amplitude is bigger with 40 MHz than at 20 MHz.



(A) (no activity)



(B) (walking)

FIGURE 4.4: WiFi 20 and 40 MHz bandwidth comparison (amplitude)

Experiment 3: Frequency 2.4 and 5 GHz

System configuration	
Bandwidths	40 MHz
Channel	11
Frequency	2.4, 5 GHz
Antennas	2Rx vs 2Tx

There is a possibility to use a Wi-Fi network with the 0.9, 2.4, 3.6 and 5 GHz frequencies. However, only 2.4 and 5 GHz are the most commonly used and spread, so we focus only on them. In general, 2.4 GHz outshines 5 GHz as to the Wi-Fi coverage, but 5 GHz gives a much higher speed.

The experiment results are shown in Fig. 4.5. Walking areas pattern can be observed roughly the same in both frequencies (showed in the bottom graphs). However, the variation at the amplitude of 5 GHz is less than at 2.4 GHz (showed in the top graphs) and results in less data noise.

The reason could be the absence of another Wi-Fi network on 5 GHz frequency during the experiment. The 5 GHz band is not widely used yet, and there is no influence from neighbours' nodes.

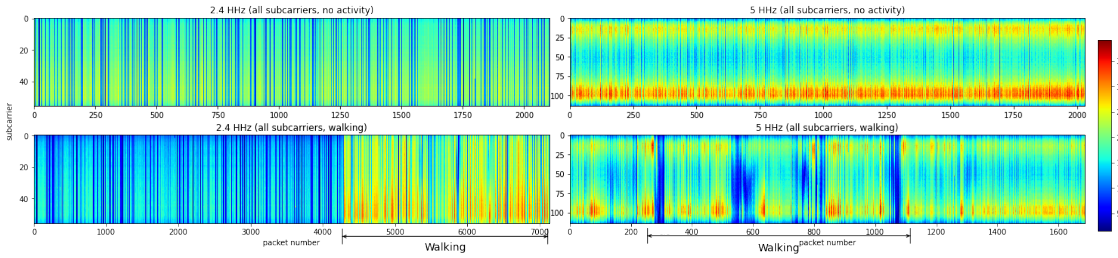


FIGURE 4.5: 2.4 vs 5 GHz WiFi frequency comparison

Experiment 4: Different antennas quantity effect on HAR not in the LoS

Used in this research TP-link TL-WRD4300 has three antennas (all three can work at the 5 GHz frequency, but only two at 2.4 GHz). Different antennas configurations (number of enabled antennas on the transmitter and receiver routers) were tested to understand how their number influences the ability to detect a walking activity. In this experiment, routers were located diagonally in the room. The person was walking next to one of the walls (not in the routers line of sight) (Fig. 4.6).

System configuration	
Bandwidths	40 MHz
Channel	60
Frequency	5 GHz
Antennas	1Rx x 1Tx, 2Rx x 2Tx, 2Rx x 1Tx, 1Rx x 2Tx, 2Rx x 3Tx

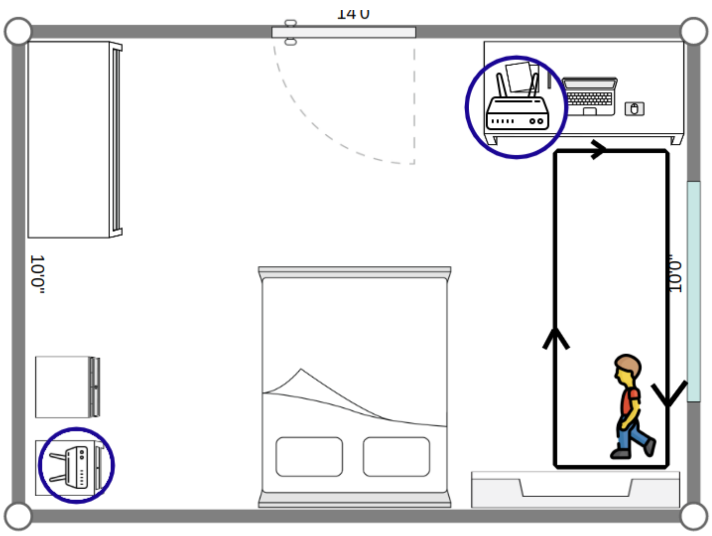


FIGURE 4.6: A person is walking next to the wall, not in the line of the routers sight

The results are shown in Fig. 4.7. As we can see, there is no actual possibility to detect an activity with the 1Rx x 1Tx configuration, when a person is not on the sight line. Also, 1Rx x 2Tx and 2Rx x 1Tx configurations give worse results compared to the 3Rx x 2T and 2Rx x 2Tx. Clearly, there is a change in the amplitude in the beginning and the end of the walking, however, the change itself takes very short time and during the activity, no changes are observed. However, with the 3Rx x 2T and 2Rx x 2Tx configurations, the activity pattern is better visible and the change of the amplitude during the activity can be observed. So, based on that, we can state that we need at least two receiver and two transmitter antennas to detect "walking" activity when the person is not on the sight line.

Along with that, we have discovered problems with Atheros CSI Tool:

1. Having one transmitter and two receiver antennas configuration, the system reports that there are two transmitters antennas and returns noise on that antenna as CSI (shown of Fig. 4.8 (A). The amplitude is shown on the top graph, the phase – on the bottom one. The subcarriers are shown along x-axis, the amplitude/phase value along the y-axis. The lines are coloured by the antenna pairs: the first Tx antenna to first Rx – yellow, the first Tx antenna to second Rx – red, the second (not actually existed in this experiment) Tx antenna to Rx antennas – light-blue and blue.

2. Having three transmitter and two receiver antennas, the system incorrectly estimates the CSI data, which always gets a sinusoidal shape (both amplitude and phase) (shown in Fig. 4.8 (B). The colouring is the same as in the previous graphs.

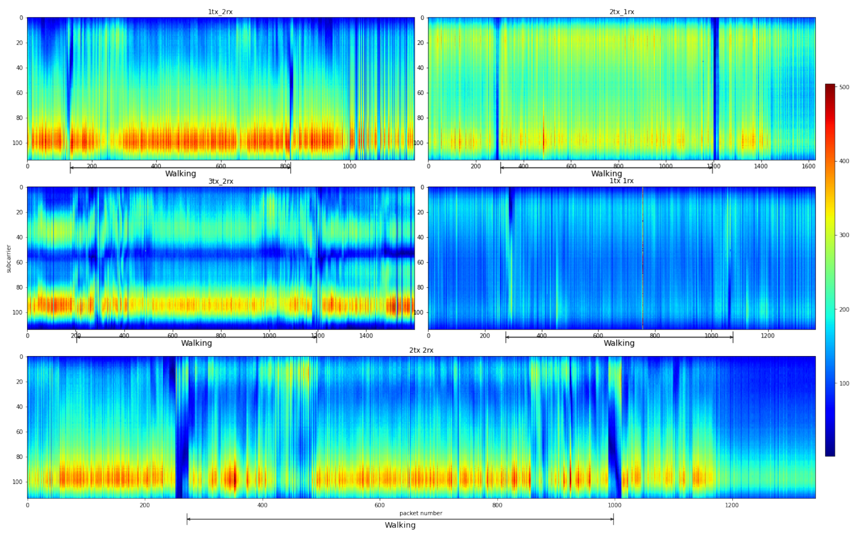


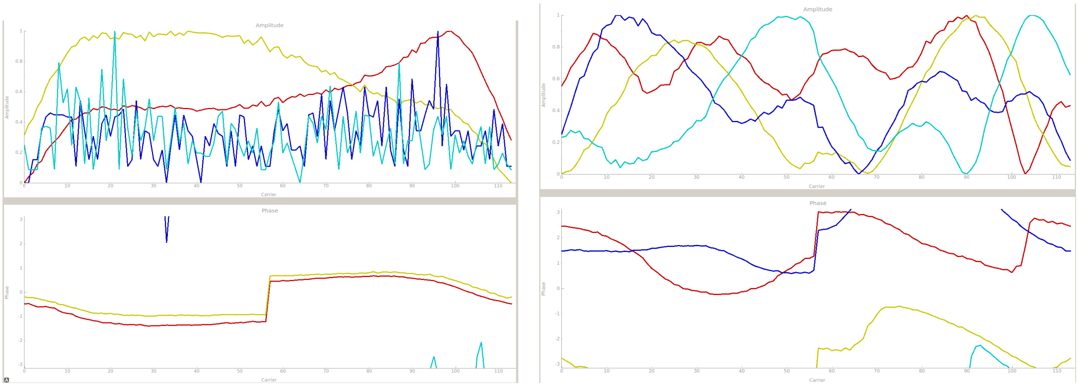
FIGURE 4.7: Different antenna configuration comparison (1x1, 1x2, 2x1, 2x2, 3x2)

Experiment 5: Detecting activity in another room with no Wi-Fi device

A. Routers are in different rooms

In this experiment, the transmitter and receiver, routers were located in different(adjoining) rooms. After that, "walking" activity was done in both rooms. The results are shown in Fig. 4.9, there is

System configuration	
Bandwidths	40 MHz
Channel	11
Frequency	2.4 GHz
Antennas	2Rx vs 2Tx



(A) 1 transmitter antenna and 2 receiver antennas (B) 3 transmitter antennas and 2 receiver antennas

FIGURE 4.8: Problems found with CSI calculation while using a different numbers of antennas on transmitter and receiver

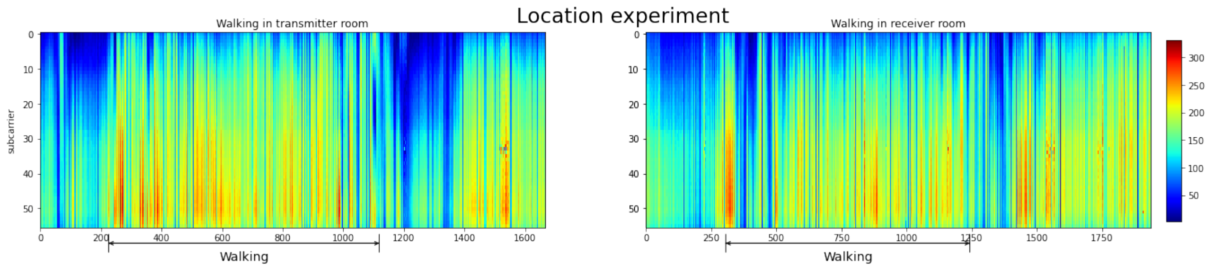


FIGURE 4.9: Both routers are in the different rooms (amplitude)

no difference in CSI data in both rooms and activity can be detected, which was expected.

B. Routers are in the same room

In this experiment, the transmitter and receiver, routers were located in the room. After that, "walking" activity was first done in the room, where both routers are located, and then in another room. The results are shown in Fig. 4.10. Human activity can be easily detected in a room with routers, but it is unfeasible when a person is in another room (at least by using simple commodity routers)

System configuration	
Bandwidths	40 MHz
Channel	60
Frequency	5 GHz
Antennas	2Rx vs 2Tx

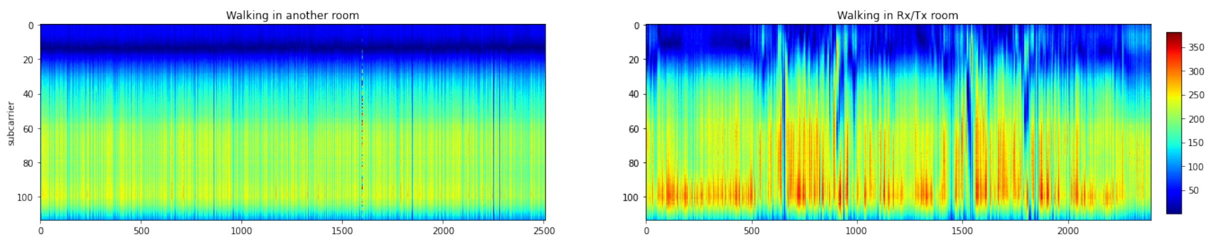


FIGURE 4.10: Both routers are in same room (amplitude)

Experiment 6: WiFi gestures

A gesture is a form of non-verbal communication in which visible bodily actions communicate particular messages — for example, okay symbol or clapping hands to switch the light in the room. The possibility of detecting gestures based on WiFi CSI data was tested. In these experiments, routers were located in 1 meter between them and gestures were performed in their line of sight.

A. Arm gesture

These experiments performed an arm moving gesture: 5 times – a vertically moving arm, a little pause, and 5 times horizontally. Based on the results (shown in Fig. 4.11), there is a specific pattern of data when the arm was moving vertically and horizontally, so we can distinguish them.

System configuration (for both)	
Bandwidths	40 MHz
Channel	60
Frequency	5 GHz
Antennas	2Rx x 2Tx

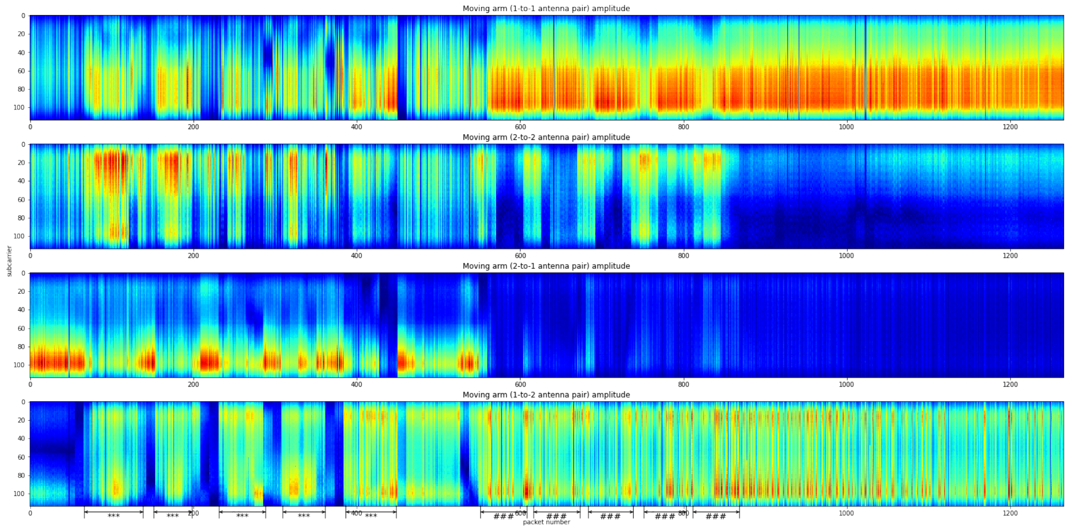


FIGURE 4.11: Arm gesture detection

B. Palm gesture

This experiments conducted palm moving gestures: 5 times – a vertically moving palm, little pause and 5 times horizontally. Based on the results (shown in Fig. 4.12), there is some data change throughout the time, however, it is hard to find the pattern of data to distinguish between the horizontal and vertical palm moving and detect the exact time period they were done. It may be due to the moving palm, which is quite small (compare to arm) and does not affect CSI data much.

4.2.1 Summary

Based on the conducted experiments, we can state that the proper hardware configuration is crucial for the HAR system based on Wi-Fi CSI data. The conclusions and recommendations based on each experiment are the following:

1. The same as with a Wi-Fi network set up for common needs, for better HAR, it is better to choose a not overlapping channel. For example, for 2.4 GHz, it could be 11, 1 or 6. During our experiment, the pattern of the walking area on channel 11 yielded a more contrasted pattern compared to the other channels.

2. The experiment with the 40 MHz bandwidth shows better results regarding the pattern visibility then 20 MHz. However, the variance of amplitude and noise by using 40 MHz is bigger than on 20 Mhz.

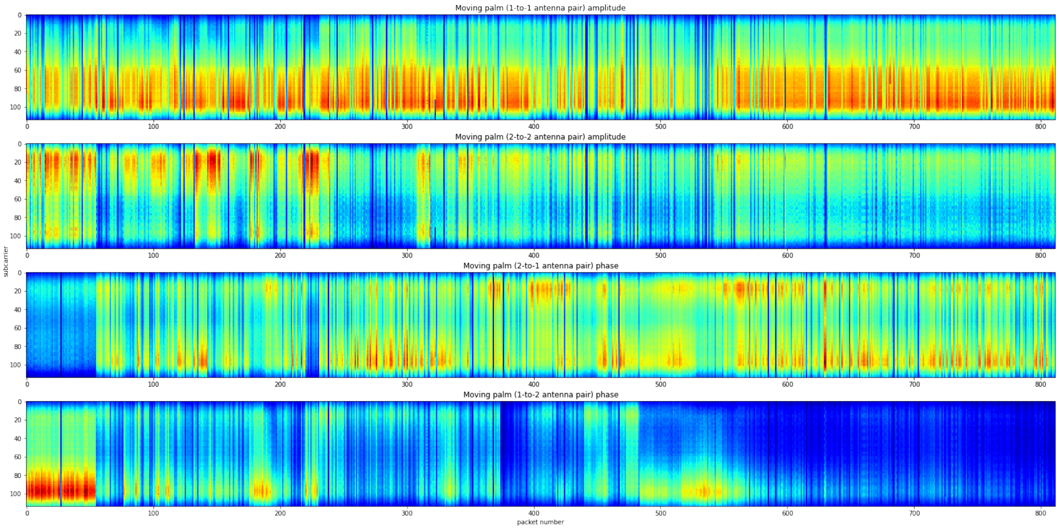


FIGURE 4.12: Palm gesture detection

3. 2.4 and 5 GHz Wi-Fi networks are the most popular now and they were used in the experiment. From the results, we cannot say that one frequency gives better results than others, as the activity patterns are comparable. However, the amplitude at 5 GHz is less than at 2.4 GHz and results in less data noise.

4. Regarding the antennas quantity – the more, the better. From the results, while a person is not on the sight line, it is almost impossible to detect any activities and movements in the 1Tx 1Rx configuration. We can state, that at least two receiver and transmitter antennas are required to detect human activity when they are not on the sight line. Also, several problems with Atheros CSI-Tool were discovered regarding the incorrect data processing in a case with different antenna numbers in the transmitter and receiver.

5. As well as we can access the Internet in different rooms, we can detect walking activity in another room if there is a receiver (client). However, the experiment shows, that there is no possibility to detect any movements in a room with no device (Access Point or client).

6. The experiment with Wi-Fi gestures shows that in an environment close to the lab, it is possible to detect and recognise them. We see a pattern in data during an arm gesture, however not with a palm. Probably, it is due to the size of an object – Wi-Fi still can detect an arm, but a palm is too small for it. Yet, this topic needs more investigation.

To sum it up:

At least two antennas on the receiver and transmitter are required to detect movements not only on the sight line, the 5 GHz band is recommended for now as it is not affected from neighbours' nodes. A bigger bandwidth is better as it gives more sensitivity even at the cost of increased noise, but that can be eliminated later. The channel has to be chosen considering the frequency band you are using and overlapping with other channels. If you are not in a room, with an AP or client router, it is not possible to detect your activity and movements. Recognizing gestures using the Wi-Fi CSI appears perspective and has to be further investigated.

4.3 Data retrieval

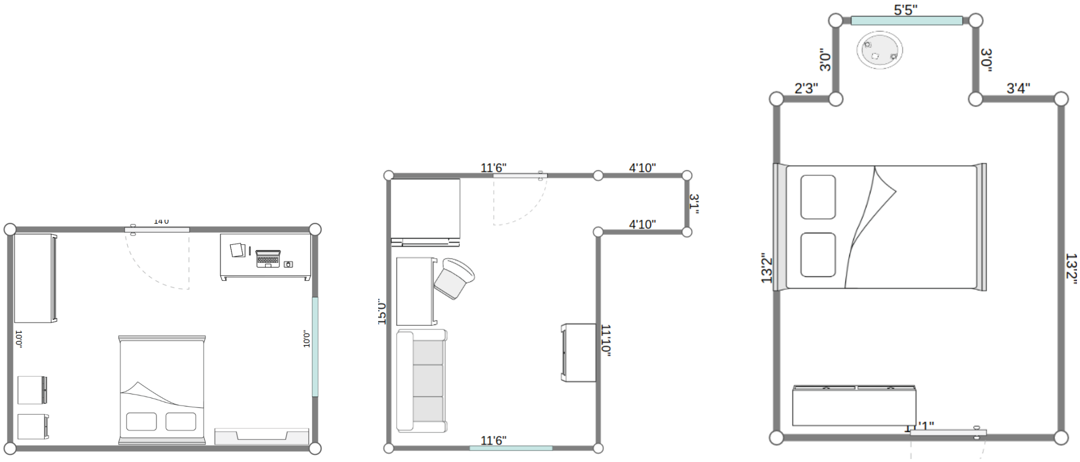
To build and train a model for the HAR, a dataset is required. We found two publicly available datasets: [34] and [35]. Both of them were collected using Linux 802.11n CSI Tool with three antennas on the transmitter and receiver routers.

However, we decided not to use them because of their data collection approach. They collect data by sequences(fixed time period), during which only one activity can be performed. Those sequences may be quite large in time, but the problem is that they contain a single action during them, so, transitions between different actions cannot be observed as well as quick and frequent change of actions during small periods of time. We state that it is a poor representation of real-world data and human's behavior. Also, their hardware and environmental setup are completely different from ours.

Activities	walking, sitting, standing, lying, getting up, getting down, no activity
Size	1.2 Gb (no images), 9.1 Gb (with images)
Labels	activity, person bounding box coordinates
# of people involved	1
# of rooms used	3
WiFi router	TP-Link TL-WDR4300
Bandwidth	40 MHz
Channel	60
Frequency	5 GHz
Antennas	2Rx x 2Tx
of subcarriers	114

TABLE 4.1: Collected dataset information

Based on that, we collected a new dataset. The workflow is described in 4.1. The activities performed: walking, sitting, standing, lying, getting up, getting down, and absence of activity as there is no person in the room. Totally, three different rooms are used, their plans are shown in Fig. 4.13. Each CSI packet is labelled with an image, activity and bounding box of the person located on the image, which may be interesting for further work. Data for each activity is shown in Fig. 4.14, the length distribution of each activity in 4.15. The description of the collected dataset is shown in Table 4.1.



(A) (B) (C)

FIGURE 4.13: Plans of rooms, the dataset was collected in

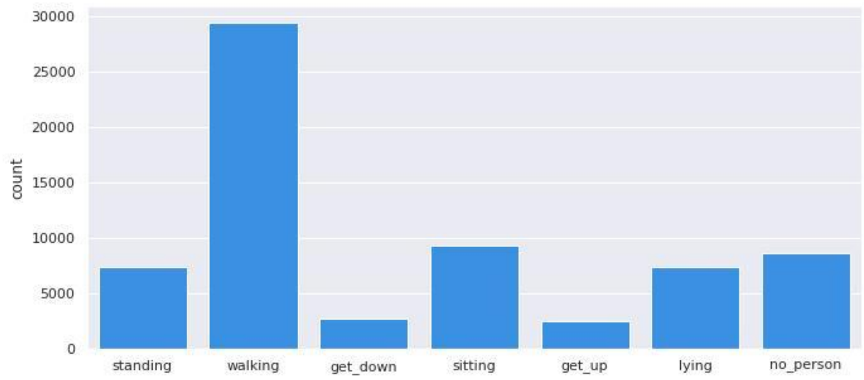


FIGURE 4.14: Collected dataset: number of CSI packets to correspond to each activity

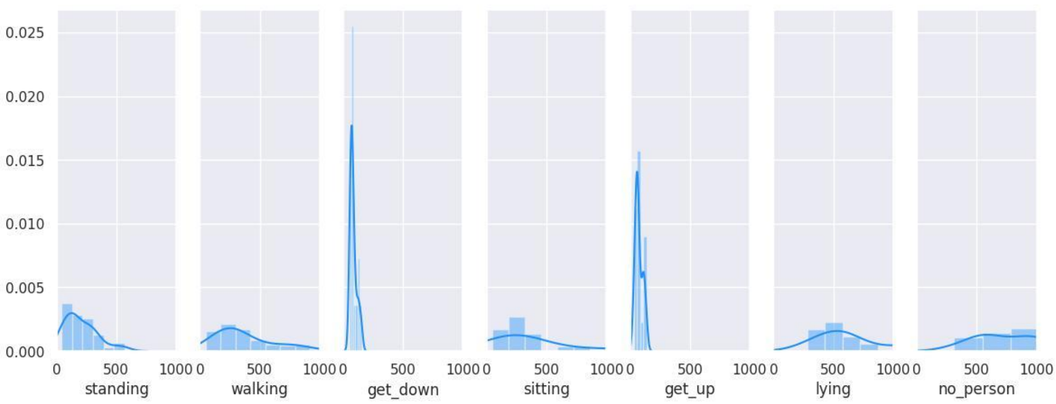


FIGURE 4.15: Collected dataset: distribution of lengths for each activity

Chapter 5

Developed solution

5.1 Data preprocessing

Even with the best possible hardware and configuration, CSI data has much noise and raw phase information cannot be directly used. This is caused firstly by the environment and reflected radio waves, secondly, possible hardware and software errors. That is why the data preprocessing is an essential part of building a stable and accurate HAR model based on WiFi CSI data. A full workflow from the routers to model prediction and preprocessing is shown in Fig. 5.1.

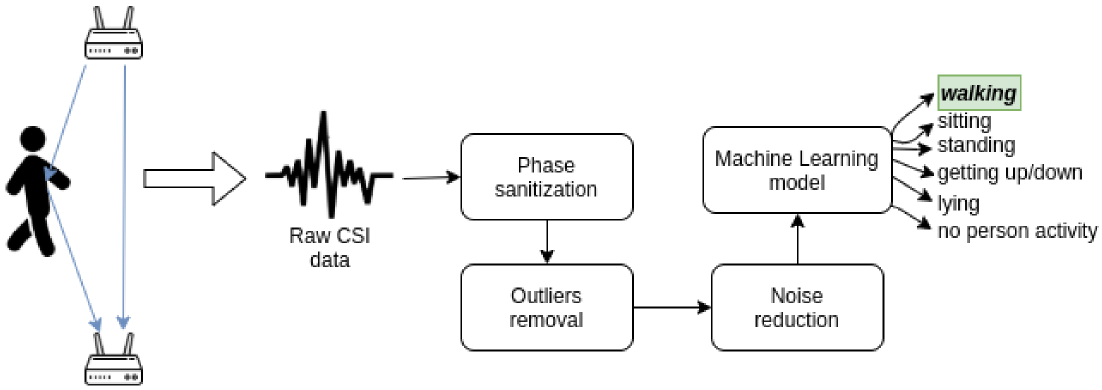


FIGURE 5.1: System workflow from Wi-Fi router to Machine Learning classification model

5.1.1 Phase sanitization

In contrast to amplitude, raw phase information cannot be used. The problem is that it is affected by the carrier frequency offset (CFO) and sampling frequency offset (SFO). The CFO arises when the transmitter and receiver do not precisely synchronize their timing and phases before transmitting a packet. For example, if the receiver and transmitter clocks difference is 50s, then at the 5 GHz Wi-Fi band, this leads to a phase change of 8π. And based on that, the phase changes due to the human moving commonly less than 0.5π, it cannot be observed from phase change [7]. The SFO is caused by analogue to digital converter, and it also varies by subcarrier, so each of them has a different error in the end.

As we do not know both CFO and SFO, the raw phase information is not useful. But the linear transformation described in [36] aims to fix it.

The results of the phase sanitization technique are shown in Fig. 5.2. As we can see, raw phase information in the left graphs does not give us any information about walking activity which done during its collection. However, after sanitizing, the phase data became less noisy, and we see that during the walking, the phase changed.

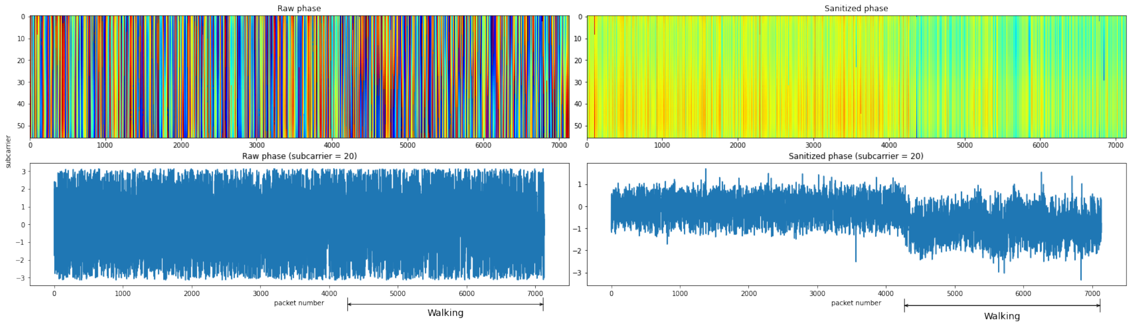


FIGURE 5.2: Phase sanitization algorithm results: raw phase and sanitized

5.1.2 Outliers removal technique

Amplitudes and phases data contains noises caused by the transition rate and power adaptations, thermal noise etc. As a result, it introduces outliers to the signal that are not caused by human actions. Hampel Identifier algorithm [21] was to eliminate this problem. Its results are shown in Fig. 5.3

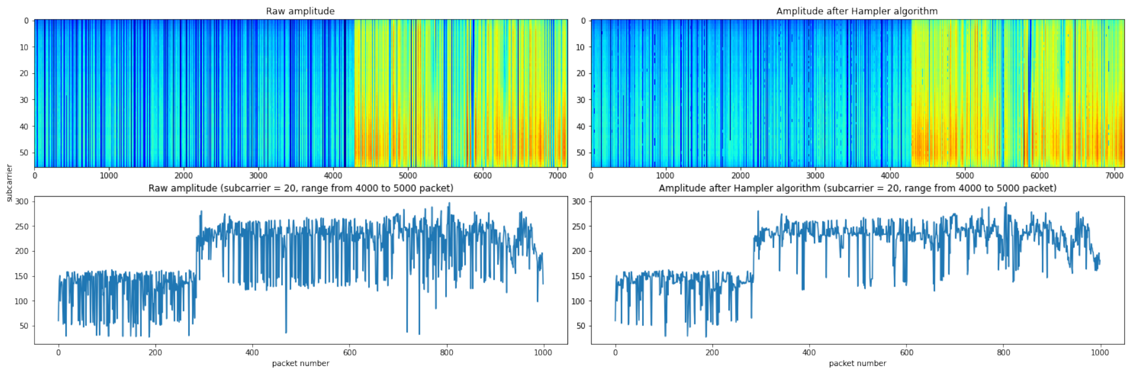


FIGURE 5.3: Hampel outlier removal algorithm results ($k = 5$, $t_0 = 2$)

5.1.3 Noise reduction based on Discrete Wavelet Transform

Still, there is much noise in CSI data left and based on [21] the Discrete Wavelet Transform based noise reduction algorithm was implemented. The Pywt library [37] was used for this purpose. The results of this method are shown in Fig. 5.4.

5.2 Model

It is hard to interpret CSI data by using classical methods like SVM, decision tree, fuzzy rule-based classifiers, therefore we have chosen to use the neural network approach instead. The idea behind it is that must be capable of learning complex, long-term dependencies and features needed for our problem.

The dataset used for training and validation of the network is described in 4.3. As it's shown in Table 4.1, two antennas for the receiver and transmitter routers were

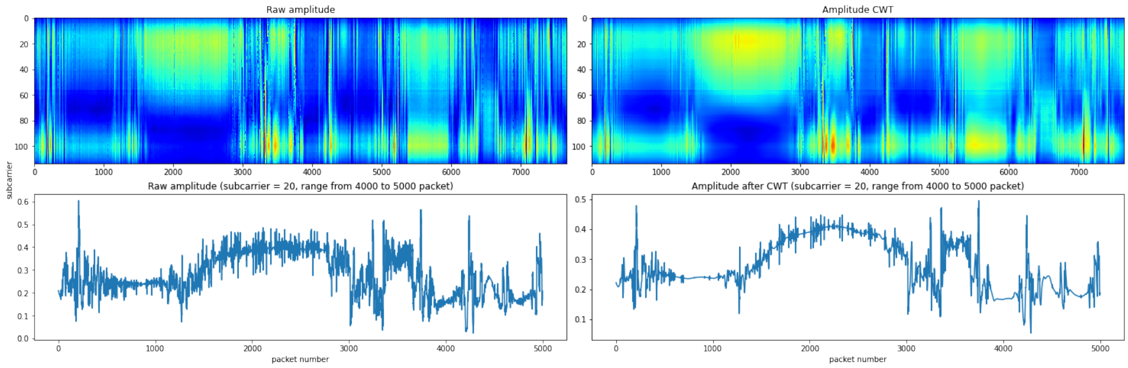


FIGURE 5.4: Discrete Walewet Transform for noise removal results (threshold = 20%)

used, so totally there are four antenna pairs. Each antenna pair has CSI data (phase and amplitude) for each of 114 subcarriers. So, there are 4 * 114 * 2 = 912 features for each CSI packet we receive. Those features are further used for model training.

Along with that, we need to consider that people do not immediately change their activity, and this transition can take some time. Also, due to the environment characteristics and reflected waves, the model has to be able to learn patterns throughout time. The data is fed to the model as sequences with some step between them. The target is to detect the activity performed at the end of that sequence.

Based on that, two methods were tested:

InceptionTime model [38] proposes an architecture based on deep one-dimensional Convolutional Neural Network aimed at Time series classification problem. They show promising results in their architecture accuracy as well as in training time, which is smaller compared to similar architectures.

InceptionTime model is an ensemble of Inception networks followed by a Global Average Pooling layer and a Dense layer with a softmax activation function. The model is constituted by the following layers:

- A bottleneck layer to reduce the dimensionality of the input sequence
- Max Pooling layer of size 3, through which the input is also passed
- The output of the bottleneck is fed to three one-dimensional convolutional layers of kernel size 10, 20 and 40.
- The output of the Max Pooling layer is fed to a single one-dimensional convolutional layer of kernel size 1.
- Finally, the outputs of all four convolutional layers are concatenated in the last layer.

So, the InceptionTime architecture was used for Wi-Fi CSI-based HAR. Its input is the sequence of CSI data of length 1024. That sequence label is the activity class performed at the end of that sequence. The number of input channels for the first convolutional layer is equal to the (number of subcarriers) * (number of antenna pairs) * 2 = 1024, where "2" stands for phase and amplitude The dataset is divided by sequences with the step of 8 time points. It means that the second sequence is shifted by eight packages in time compared to the first one.

# of blocks	3
# of input channels	1024
# of output channels	912
# of bottleneckchannels	12
kernel size	15
use residuals	True
# of predicted classes	7
Learning rate	0.00147
Batch size	10
# of training epochs	50

TABLE 5.1: Hyperparameters used for InceptionTime model

Cross Entropy Loss was used as a loss function. The model was built based on the codebase [38] provides. The training was performed on the Google Colab service with Nvidia K80s GPU. Training for 50 epochs took approximately 10 hours.

After training the model after 50 epochs, its accuracy on validation data was 38.2%; however, after training the data, it reaches 99.2%. We see overfitting here, and the main reason is probably the insufficient amount of data, as we tried different validation and training subsets and it did not give a better result. The model hyperparameters used in this model are shown in 5.1.

LSTM-based model RNN works well with sequence data, while its LSTM improvement helps with long term dependencies. Based on that, this approach was tested with the model architecture as follows:

- A single bidirectional LSTM layer of the hidden dimension of size 256
- Dense layers of sizes 512, 256, 128, 7, and the activation function ReLU. The last dense layer is the system output.

Model's input is the sequence of CSI data of length 1024. That sequence label is the activity performed at the end of that sequence. Using that, we tried to simulate a realtime HAR system. The dataset is divided by sequences with the step of 16 time points. It means that the second sequence is shifted by eight packages in time compared to the first one.

Cross Entropy Loss was used as a loss function. The model was built using Python PyTorch framework. The training was performed on Google Colab service with Nvidia K80s GPU. Training for 60 epochs took approximately 7 hours.

Table 5.2 summarises the model and hyperparameters used for training of this model. After the model training, its accuracy was 61.1%, precision 59% and recall 55% on validation data, 87.8% accuracy on testing data. The resulting confusion matrix is shown in Fig. 5.5. This model also suffers from overfitting, but not as much as the previous one. Also, the accuracy is much higher. Based on that, we can state that LSTM-based approach works better in this specific case when there is no much data available for training.

The system accuracy is worse compared to similar systems [35] [34] [21], but we suppose that this is due to the difference between our dataset collection approach because the models' architectures are quite similar.

These models look promising and we presume that larger amount of data used for training can significantly improve their accuracy; however, collection of respective data sets requires substantial additional effort and if out of the scope of present

Cell type	LSTM
# of layers	1
RNN hidden size	256
Dense layers for classification	512 -> 256 -> 128 -> 7
Dense layers activation	ReLU
Sequence size	1024
Data step	16
Dropout	0.2
Learning rate	0.00147
Batch size	16
# of training epochs	60

TABLE 5.2: Hyperparameters used for LSTM-based model

bachelor research. Also, even after data preprocessing, we still have just amplitude and phase data, which is hard to transform into valuable information about human movements and activity. This problem can be solved by the introduction of more high-level features that can capture special special characteristics about the signal we are receiving.

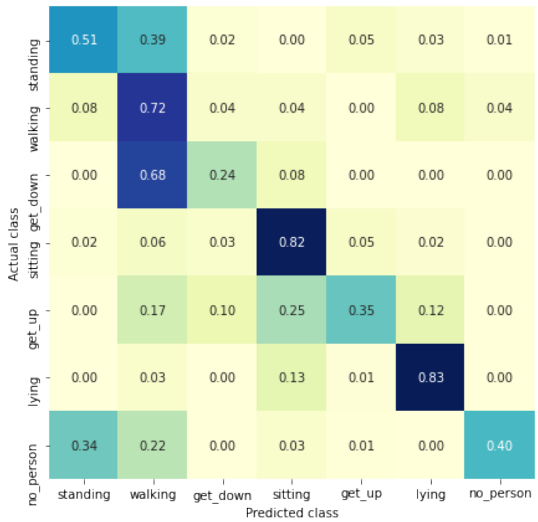


FIGURE 5.5: Confusion matrix of LSTM-based model results

Chapter 6

Results and summary

6.1 Results discussion

From our analysis, the existing studies in this field show that there are a lot of systems indicating the perspective of using CSI in the HAR area. For example, people detection and fall recognition problems are shown as mostly solved. Nevertheless, some issues remain open:

- Those systems need more testing
- Those systems often need a better data collection approach to represent real life not labs
- The accuracy and robustness must be increased to use them publicly.
- The problem with using CSI and already collected datasets in different environments is not solved yet.

Based on research goals, which includes: analysis of existing approaches (described above) and how the Wi-Fi CSI depends on the system configuration, dataset collection and HAR model development, we achieved the following results:

1. Unlike with configuration Wi-Fi network for everyday usage, there are no recommendations on which system configuration to use to get the best Wi-Fi CSI-based HAR system accuracy. We conducted experiments to understand how the Wi-Fi configuration, routers location and environment affect the CSI data and the ability to recognize the human activity. Based on the received results, we recommended a particular network configuration and provided a summary of our observations. Also, we have shown that Wi-Fi gestures are possible to detect and recognize, but it needs more investigation to understand all limitations of such a system in real life.
2. Previous studies use a simplified data collection approach, that does not fully represent real-world data and human's behaviour. Based on that, we proposed new data collection approach, where during the small amount of time, different activities can be performed with no strict order. The collected dataset is publicly available and can be used for further work.
3. We proposed the data preprocessing workflow, which includes: phase sanitization, outliers remove and data denoising before feeding it into the model. InceptionTime and LSTM-based models were trained and described. The final model gives 61% classification accuracy for 7 activities (83% for lying recognition, 82% for sitting). Although the success is limited due to the limited volume of the dataset, these initial results look promising, and further research

in this direction could be conducted with larger data sets. Also, we tried the model architecture very close to what similar works(which got better results) in this field are using, so we can state that data collection approach is crucial and the more it's close to the real world, the harder the task is. Furthermore, replacement of the pre-processed amplitude and phase shift data with a more advanced feature sets may yield better results as well as other ideas described in 6.2.

6.2 Future work

We see possible areas for improvements in:

- Extend the dataset for other locations, people and hardware variations
- Look into other NN architectures as well as try to use Transfer Learning to get the ability to use data collected by other tools and hardware
- Use origin complex numbers CSI is providing instead of phase and amplitude, as they can better save the relation between them especially on different subcarriers
- Features extractions to get the most from the data: PCA, Power Spectral Density, signal skewness, Median Absolute Deviation, Autospectrum etc.
- Data augmentation and using synthetic data to extend the current dataset
- Atheros CSI Tool: improve the system for ensuring correctness when there is different antennas number on transmitter and receiver (see 4.2)

Bibliography

[1] Claudine Badue et al. *Self-Driving Cars: A Survey*. 2019. arXiv: 1901.04407 [cs.RO].

[2] R. W. Wolcott and R. M. Eustice. “Visual localization within LIDAR maps for automated urban driving”. In: *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2014, pp. 176–183.

[3] Tobias Grosse-Puppendahl et al. “Finding Common Ground: A Survey of Capacitive Sensing in Human-Computer Interaction”. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI ’17. Denver, Colorado, USA: Association for Computing Machinery, 2017, 3293–3315. ISBN: 9781450346559. DOI: 10.1145/3025453.3025808. URL: https://doi.org/10.1145/3025453.3025808.

[4] A. Tahat et al. “A Look at the Recent Wireless Positioning Techniques With a Focus on Algorithms for Moving Receivers”. In: *IEEE Access* 4 (2016), pp. 6652–6680.

[5] J. Wilson and N. Patwari. “Radio Tomographic Imaging with Wireless Networks”. In: *IEEE Transactions on Mobile Computing* 9.5 (2010), pp. 621–632.

[6] Fadel Adib et al. “3D Tracking via Body Radio Reflections”. In: *Usenix NSDI* (Dec. 2013).

[7] Yongsen Ma, Gang Zhou, and Shuangquan Wang. “WiFi Sensing with Channel State Information: A Survey”. In: *ACM Comput. Surv.* 52.3 (June 2019). ISSN: 0360-0300. DOI: 10.1145/3310194. URL: https://doi.org/10.1145/3310194.

[8] Charmi Jobanputra, Jatna Bavishi, and Nishant Doshi. “Human Activity Recognition: A Survey”. In: *Procedia Computer Science* 155 (2019). The 16th International Conference on Mobile Systems and Pervasive Computing (MobiSPC 2019),The 14th International Conference on Future Networks and Communications (FNC-2019),The 9th International Conference on Sustainable Energy Information Technology, pp. 698 –703. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2019.08.100. URL: http://www.sciencedirect.com/science/article/pii/S1877050919310166.

[9] Zawar Hussain, Quan Sheng, and Wei Emma Zhang. “Different Approaches for Human Activity Recognition: A Survey”. In: (June 2019).

[10] Liming Chen et al. “Sensor-Based Activity Recognition”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42 (Nov. 2012), pp. 790–808. DOI: 10.1109/TSMCC.2012.2198883.

[11] Y. Chen and Y. Xue. “A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer”. In: *2015 IEEE International Conference on Systems, Man, and Cybernetics*. 2015, pp. 1488–1492.

[12] Jian Bo Yang et al. "Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition". In: *Proceedings of the 24th International Conference on Artificial Intelligence*. IJCAI'15. Buenos Aires, Argentina: AAAI Press, 2015, 3995–4001. ISBN: 9781577357384.

[13] Jindong Wang et al. "Deep learning for sensor-based activity recognition: A survey". In: *Pattern Recognition Letters* 119 (2019). Deep Learning for Pattern Recognition, pp. 3 –11. ISSN: 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2018.02.010. URL: http://www.sciencedirect.com/science/article/pii/S016786551830045X.

[14] Michalis Vrigkas, Christophoros Nikou, and Ioannis Kakadiaris. "A Review of Human Activity Recognition Methods". In: *Frontiers in Robotics and Artificial Intelligence* 2 (Nov. 2015). DOI: 10.3389/frobt.2015.00028.

[15] Zheng Yang Yunhao Liu Kyle Jamieson Kun Qian Chenshu Wu. "Widar: Decimeter-Level Passive Tracking via Velocity Monitoring with Commodity Wi-Fi". In: *Mobihoc '17: Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing* (July 2017). URL: http://dx.doi.org/10.1145/3084041.3084067.

[16] Qin Lv Jie Xiong Shengjie Li Yue Zhang Hong Mei Xiang Li Daqing Zhang. "IndoTrack: Device-Free Indoor Human Tracking with Commodity Wi-Fi". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (Sept. 2017). URL: https://dl.acm.org/doi/10.1145/3130940.

[17] Deepak Vasisht, Swarun Kumar, and Dina Katabi. "Decimeter-Level Localization with a Single WiFi Access Point". In: *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*. Santa Clara, CA: USENIX Association, Mar. 2016, pp. 165–178. ISBN: 978-1-931971-29-4. URL: https://www.usenix.org/conference/nsdi16/technical-sessions/presentation/vasisht.

[18] Zheng Yang Kun Qian Chenshu Wu and Yunhao Liu. In: *ACM Transactions on Embedded Computing Systems (TECS)* (Jan. 2018). URL: https://dl.acm.org/doi/10.1145/3157677.

[19] J. Lv et al. "Robust WLAN-Based Indoor Intrusion Detection Using PHY Layer Information". In: *IEEE Access* 6 (2018), pp. 30117–30127. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2017.2785444.

[20] H. Wang et al. "RT-Fall: A Real-Time and Contactless Fall Detection System with Commodity WiFi Devices". In: *IEEE Transactions on Mobile Computing* 16.2 (2017), pp. 511–526.

[21] Tahmid Chowdhury. "Using Wi-Fi Channel State Information (CSI) for Human Activity Recognition and Fall Detection the thesis can also be accessed at: https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0365967". PhD thesis. Apr. 2018. DOI: 10.14288/1.0365967.

[22] Youwei Zeng et al. "FarSense: Pushing the Range Limit of WiFi-Based Respiration Sensing with CSI Ratio of Two Antennas". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3.3 (Sept. 2019). DOI: 10.1145/3351279. URL: https://doi.org/10.1145/3351279.

[23] "IEEE Standard for Information technology—Telecommunications and information exchange between systems Local and metropolitan area networks—Specific requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications". In: *IEEE Std 802.11-2016 (Revision of IEEE Std 802.11-2012)* (2016), pp. 1–3534.

[24] Moshe Leshno et al. "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function". In: *Neural Networks* 6.6 (1993), pp. 861 –867. ISSN: 0893-6080. DOI: https://doi.org/10.1016/S0893-6080(05)80131-5. URL: http://www.sciencedirect.com/science/article/pii/S0893608005801315.

[25] Jeffrey Heaton. "Ian Goodfellow, Yoshua Bengio, and Aaron Courville: Deep learning: The MIT Press, 2016, 800 pp, ISBN: 0262035618". In: *Genetic Programming and Evolvable Machines* 19 (Oct. 2017). DOI: 10.1007/s10710-017-9314-z.

[26] M. Schuster and K. K. Paliwal. "Bidirectional recurrent neural networks". In: *IEEE Transactions on Signal Processing* 45.11 (1997), pp. 2673–2681.

[27] Y. Bengio, Patrice Simard, and Paolo Frasconi. "Learning long-term dependencies with gradient descent is difficult". In: *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council* 5 (Feb. 1994), pp. 157–66. DOI: 10.1109/72.279181.

[28] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Comput.* 9.8 (Nov. 1997), 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: https://doi.org/10.1162/neco.1997.9.8.1735.

[29] Kyunghyun Cho et al. "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation". In: *CoRR* abs/1406.1078 (2014). arXiv: 1406.1078. URL: http://arxiv.org/abs/1406.1078.

[30] Daniel Halperin et al. "Tool Release: Gathering 802.11n Traces with Channel State Information". In: *SIGCOMM Comput. Commun. Rev.* 41.1 (Jan. 2011), p. 53. ISSN: 0146-4833. DOI: 10.1145/1925861.1925870. URL: https://doi.org/10.1145/1925861.1925870.

[31] Yaxiong Xie, Zhenjiang Li, and Mo Li. "Precise Power Delay Profiling with Commodity WiFi". In: *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. MobiCom '15. Paris, France: ACM, 2015, 53–64. ISBN: 978-1-4503-3619-2. DOI: 10.1145/2789168.2790124. URL: http://doi.acm.org/10.1145/2789168.2790124.

[32] NovelSense. *Atheros-CSI-Tool-UserSpace-APP*. https://github.com/NovelSense/Atheros-CSI-Tool-UserSpace-APP. 2017.

[33] Yaxiong Xie. *The user-space applications for Atheros-CSI-TOOL*. https://github.com/xieyaxiongfly/Atheros-CSI-Tool-UserSpace-APP. 2018.

[34] Jialin Liu Wei Zhou Linlin Guo Lei Wang and Bingxian Lu. "HuAc: Human Activity Recognition Using Crowdsourced WiFi Signals and Skeleton Data". In: *Wireless Communications and Mobile Computing* (2018).

[35] S. Yousefi et al. "A Survey on Behavior Recognition Using WiFi Channel State Information". In: *IEEE Communications Magazine* 55.10 (2017), pp. 98–104.

[36] Souvik Sen et al. “You Are Facing the Mona Lisa: Spot Localization Using PHY Layer Information”. In: *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*. MobiSys ’12. Low Wood Bay, Lake District, UK: Association for Computing Machinery, 2012, 183–196. ISBN: 9781450313018. DOI: 10.1145/2307636.2307654. URL: https://doi.org/10.1145/2307636.2307654.

[37] Gregory Lee et al. “PyWavelets: A Python package for wavelet analysis”. In: *Journal of Open Source Software* 4.36 (2019), p. 1237. DOI: 10.21105/joss.01237. URL: https://doi.org/10.21105/joss.01237.

[38] Hassan Ismail Fawaz et al. “InceptionTime: Finding AlexNet for Time Series Classification”. In: (Sept. 2019).