DEEP LEARNING BASED INDOOR HUMAN ACTIVITIES RECOGNITION USING CHANNEL STATE INFORMATION

KHOO BEE SZE

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DEEP LEARNING BASED INDOOR HUMAN ACTIVITIES RECOGNITION USING CHANNEL STATE INFORMATION

KHOO BEE SZE

This report is submitted in partial fulfilment of the requirements for the degree of Bachelor of Electronic Engineering with Honours

Faculty of Electronic and Computer Engineering
Universiti Teknikal Malaysia Melaka

2020

DECLARATION

I declare that this report entitled "Deep Learning Based Indoor Human Activities Recognition Using Channel State Information" is the result of my own work except for quotes as cited in the references.

Signature :

Author : KHOO BEE SZE

Date : 30/6/2020

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Bachelor of Electronic Engineering with Honours.

Signature : .....

Supervisor Name : PM Dr. Wong Yan Chiew

Date : 30/06/2020

DEDICATION

I would like to dedicate this thesis to myself.

ABSTRACT

Do you ever believe that the human activities can be recognized by only using Wi-Fi signals? In this project, a Wi-Fi based activity recognition using deep neural networks is proposed to recognize the indoor human activities. Compared to the traditional human activities recognition approaches, which employed the used of the camera and the sensors, the proposed method is unobstructive, respect to the individual's privacy and works without affected by the lighting condition. Moreover, the complexity of feature extraction processing is simplified due to the powerful inference of deep neural networks. In this project, the standard Unidirectional Long Short Term Memory (Uni-LSTM) had been developed for human activity recognition. Besides, more complex architectures such as Bidirectional LSTM (Bi-LSTM) and cascaded (Cas-LSTM) has also been investigated. An accuracy of 98.33% had been achieved by using self-collected dataset using the proposed Bi-LSTM model. All of the experiment results including the effects of human activities on the Wi-Fi signal and the performance of the proposed networks have been analyzed and evaluated in the project thesis.

ABSTRAK

Adakah anda percaya bahawa aktiviti manusia dapat dikenali dengan hanya menggunakan isyarat Wi-Fi? Dalam projek ini, pengecaman aktiviti berasaskan Wi-Fi menggunakan rangkaian neural mendalam dicadangkan untuk mengenali aktiviti manusia dalam ruangan. Berbanding dengan pendekatan pengiktirafan aktiviti manusia tradisional, yang menggunakan kamera dan sensor, kaedah yang dicadangkan adalah bebas daripada sekatan, menghormati privasi individu dan berfungsi tanpa dipengaruhi oleh keadaan pencahayaan. Lebih-lebih lagi, kerumitan pemprosesan pengekstrakan ciri dipermudah kerana kesimpulan kuat dari rangkaian saraf dalam. Dalam projek ini, Memori Jangka Panjang Panjang Unidirectional standard (Uni-LSTM) telah dikembangkan untuk pengiktirafan aktiviti manusia. Selain itu, seni bina yang lebih kompleks seperti Bidirectional LSTM (Bi-LSTM) dan lata (Cas-LSTM) juga telah disiasat. Ketepatan 98.33% telah dicapai dengan menggunakan set data yang dikumpulkan sendiri menggunakan model Bi-LSTM yang dicadangkan. Semua hasil eksperimen termasuk kesan aktiviti manusia pada isyarat Wi-Fi dan prestasi rangkaian yang dicadangkan telah dianalisis dan dinilai dalam tesis projek.

ACKNOWLEDGEMENTS

First and foremost, I am deeply gratitude to my supervisor, Prof. Madya Dr. Wong Yan Chiew for providing me the opportunity to learn and explore new things and knowledge which cannot be learn in lecture slides. Dr. Wong had assisted me on my project throughout the semesters and had given me advices and hints to help me complete my project.

Besides that, I am glad and grateful to the people that had helped me throughout this project so that I can complete the project in time. Upon completing this project, I have learnt managing time properly is crucial. Because completing this project is time-consuming. I have to plan and implement this planning to have produce desired result.

Furthermore, I would also like to take this opportunity to express my gratitude to all the lecturers of Faculty of Electronic and Computer Engineering who had provide me comments and advices. Last but not least, I would like to express my gratitude to one and all who directly or indirectly lend their hand for completing this project and thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

CSI	:	Channel State Information
RNN	:	Recurrent Neural Network
LSTM	:	Long Short Term Memory

CHAPTER 1

INTRODUCTION

1.1 Overview

This thesis proposes the implementation of deep learning based indoor human activity recognition by using channel state information (CSI). This chapter will present the project background, problem statement, objectives, the scope of the project, project significant and chapter review.

1.2 Project Background

Human activity recognition (HAR) is one of the more recent research topics that recently gained on popularity and focus of both academic and commercial researchers. Since human activity monitoring has a broad range of applications like homecare systems, prisoner monitoring, physical therapy and rehabilitation, public security, military uses and others, the motivation to create a reliable human activity recognition system is considerable.

In prior work, the recognition system required the supervised person to be located in a smart environment, equipped with devices such as sensors. Then, the sensor data is processed locally on the wearable device or transmitted to a server for feature extraction and supervised learning algorithms for classification. This type of monitoring is known as active monitoring. At the same time, some systems employed multiple view cameras or speakers which often suffer from some issue such as privacy concerns and lighting issue as well.

However this issue can be overcome as with the fast advances in mobile devices and communication technologies, various machine and devices are capable of interacting with each other within a network which is the Internet of Things (IoT). Wi-Fi technology is everywhere as it seems like almost all of the electronic devices can connect to the wireless network connection. Wi-Fi techniques are being used in our day to day life. They invisibly fill the air with a spectrum of radio frequency (RF) signals. Since the Wi-Fi signals are unobtrusive, respects to the individual's privacy and works without affected by the lighting condition. Therefore, the Channel State Information (CSI), which consists of sufficient and stable information regarding the characteristic of Wi-Fi is employed to classify the human activity through deep learning approach. The CSI signals can be obtained by using one access point and one mobile device equipped with an Intel Wi-Fi link 5300 NIC.

1.3 Problem Statement

Indoor human activity recognition played an essential role in the smart home, smart health and smart security [1]. For instance, an accident may occur when children have activities inside the home [2]. Moreover, elderly people may have the potential to fall when they try to sit down or stand up [1]. There are multiple approaches for human

activities recognition which employing wearable sensor measuring the acceleration [3], radars [4] or camera [5]. These traditional activity recognition methods obtain good performances and are widely used in application such as child care, smart home, intrusion detection, etc. Nevertheless, wearable based solutions result in inconvenience as elderly people often complain about having to carry a sensor around is not convenience. Nonetheless, the installation of a camera in a private area had raised the privacy issue [6] and limited by the requirement of light.

Thus, it is of ultimate interest to design a passive, non-invasive and ambient human activity recognition solution that does not present privacy concern. Therefore, wireless signals have been utilized to track human motion and recognize activities performed by a human.

1.4 Objective

1. To investigate the effect of interference (human movement) on Wi-Fi CSI data.
2. To design a deep neural network model to classify human activity by using Wi-Fi CSI data.
3. To analyze the performance of the proposed model.

1.5 Scope of Project

This project focuses on the classification of human activities through deep learning approach which is using Recurrent Neural Network (RNN), LSTM model. The performance of the designed deep learning model is evaluated by using two datasets. The first dataset is collected at an indoor living room. The activity samples included in the first dataset are walking, falling, running, and when the surrounding area didn't involve any activity. The second dataset [7] is downloaded from online which

collected by Z. Wang et al. in an indoor office area. There are seven activities included in the dataset which are walking, running, sitting down, stand up, walking, bed and falling. Both of the datasets are collected under a relatively stable environment. Only a single person is performing the activity during the data collection.

1.6 Project Significant

The priority purpose of this project is to propose a solution to indoor human activity recognition by using Channel State Information (CSI) which consists of sufficient and stable information regarding the characteristic of Wi-Fi. Indoor Human activity recognition by using CSI Wi-Fi signals can be applied in application such as child care, smart home, and intrusion detection. This approach is more convenient as users do not require to take any devices for recognition.

Besides, this approach is suitable to monitor the safety of elders, especially for those elders who live alone as this system reduces the privacy issue and will not be affected by the lighting issue. The CSI can be collected by equipment which can regularly found in homes today, namely Wi-Fi access point (APs).

1.7 Thesis Outline

Chapter 2 discusses the literature review on the general framework for deep learning in RF sensing, Recurrent Neural Network (RNN) and Long Short Term Memory (LSTM). Besides, the properties of Wi-Fi such as Channel State Information (CSI) and (Received Signal Strength) had been discussed in detail. Moreover, the related work on Wi-Fi based human activity recognition system had been included in this chapter as well. In Chapter 3, the design model, methodology, resources of

training and testing dataset are explained. In Chapter 4, experiments result are presented. In Chapter 5, this work is concluded with recommendations for future work

CHAPTER 2

BACKGROUND STUDY

2.1 Introduction

This chapter will be started with the general deep learning framework for RF sensing, and follow by Wi-Fi properties and related work on human activity recognition using Wi-Fi in its subsections.

2.2 General Deep Learning Framework for RF Sensing

As shown in Figure 2.1, various types of RF signals such as Wi-Fi, RFID, UWB and acoustics signals can be used as an input to deep learning algorithms. Since deep learning algorithms have an excellent capability to represent data and then extract features from the data, therefore, comparing to traditional shallow machine learning techniques such as SVM and KNN feature extraction is not necessary in this

framework. However for RF signals, preprocessing which is a data preparation step is still required before applying the deep learning algorithms in the applications.

Pre-processing is an essential step to remove randomness errors from other factors such as the packet boundary detection (PBD) error, the sampling frequency offset (SFO) and central frequency offset (CFO). In the pre-processing step, parameters such as phase difference between two antennas should be implemented to prepare the data for training. Generally, the different architectures of deep learning have different input in pre-processing step. For example, images can be constructed from the calibrated phases or amplitudes of signals as an input in the data pre-processing step [8] when Convolutional Neural Network (CNN) is employed. The signal can be divided into small time series in pre-processing step before it is used as input data in the LSTM architecture. Signal can be directly used for the deep learning framework. When auto-encoder is exploited, signals can be directly leveraged for the proposed deep learning framework.

The RF sensing deep learning framework consists of two stages which are an offline stage for training and an online stage for prediction. Training data obtained from the pre-processing step is used to train the deep learning model in the offline phase. The deep network models exhibit different potentials for different types of applications. For example, CNN is useful in image classification and pattern recognition while LSTM achieves more exceptional performance at processing variable-length input sequence. The results are predicted in the online stage by feeding the data into the well trained deep network. The output result can be used directly for the recognition process. Transfer learning also can be used in the deep learning framework to update the training weights with small measurement dataset.

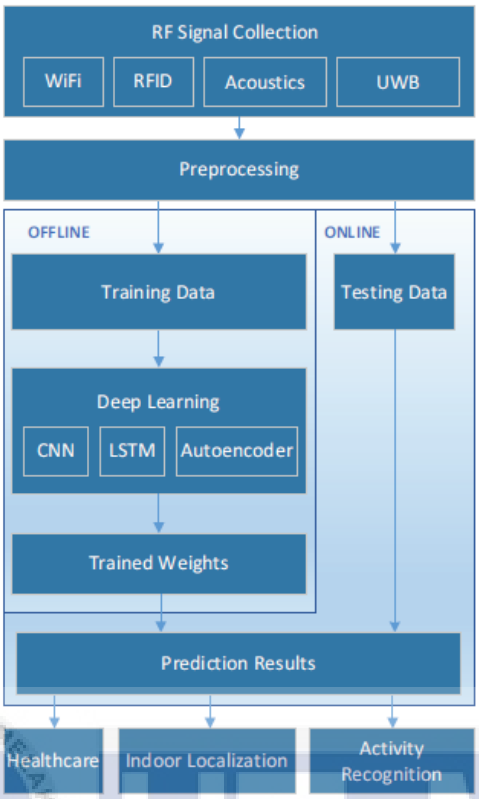


Figure 2.1: A general deep learning framework for RF sensing [9]

2.2.1 Recurrent Neural Networks (RNN)

Neural network forms the base of deep learning, which is a subfield of machine learning where the algorithms are inspired by the structure of the human brain. Neural network takes in the data and trains themselves to recognize the patterns in the data and then predicts the outputs for a new set of similar data. Neural network is made up of layers of neurons, which are the core processing units of the network.

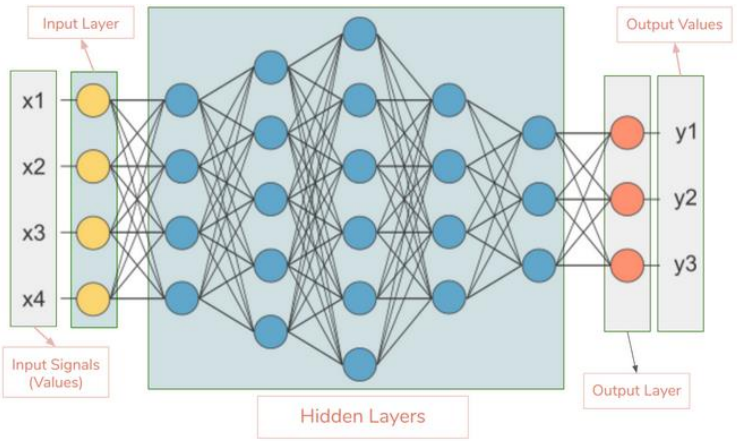


Figure 2.2: The layout of neural network.

Figure 2.2 shows the basic layout of neural network. The input layer is used to receive the input, while the outputs layer is used to predict the final output. The hidden layers which perform most of the computations required in the network are existing between the input and output layers.

Neurons of one layer are connected to the neurons of the next layer through the channel. Each of these channels is assigned a numerical value to know as weight. The input is multiplied to the corresponding weights, and their sum is sent as input to the neurons in the hidden layer. Each of these neurons is associated with a numerical value called the bias, which is then added to the input sum. This valued is then passed through a threshold function called the activation function. The result of the activation function determines if the particular neurons will get activated or not. An activated transmits data to the neurons of the next layer over the channel. In this channel the data is propagated through the network. This is called the forward propagation.

In the output layer, the neuron with the highest value or known as probability will determine the output. During the training process along with the input, the weight will be adjusted by comparing the predicted output with the input to realize the error in prediction. The process where the information is transferred backward through the network is known as back propagation. The training cycle with forward propagation and backward propagation is iteratively performed until the weight are assigned such that the network can predict accurately.

Recurrent Neural Network is a generalization of feed forward neural network that has internal memory. RNN is recurrent in nature as it performs the same function for every input of data while the output of the current input depends on the past one

computation. After producing the output, it is copied and sent back into the recurrent network. For making a decision, it considers the current input and the output that it has learned from the previous input.

RNNs can use their internal state or known as memory to process sequences of inputs. This makes them applicable to tasks such as unsegmented, connected handwriting recognition or speech recognition. In other neural networks, all the inputs are independent of each other. But in RNN, all the inputs are related to each other.

Figure 2.3 shows an unrolled recurrent neural network [10]. First, it takes the X(0) from the sequence of input and then it outputs h(0) which together with X(1) is the input for the next step. So, the h(0) and X(1) is the input for the next step. Similarly, h(1) from the next is the input with X(2) for the next step and so on. This way, it keeps remembering the context while training. The formula for the current state is defined as:

$$h_t = \mathrm{f}(h_{t-1}, x_t) \tag{1}$$

which can be expressed as:

$$h_t = \tanh(w_{hh} h_{t-1} + w_{hx} x_t) \tag{2}$$

where W is weight, h is the single hidden vector, w_{hh} is the weight at previous hidden state, w_{hx} is the weight at current input state, tanh is the activation function, that implements a non-linearity that squashes the activations to the range [-1.1] . Where output is expressed as:

$$y_t = W_{hy} h_t \tag{3}$$

where y_t is the output state. W_{hy} is the weight at the output state.

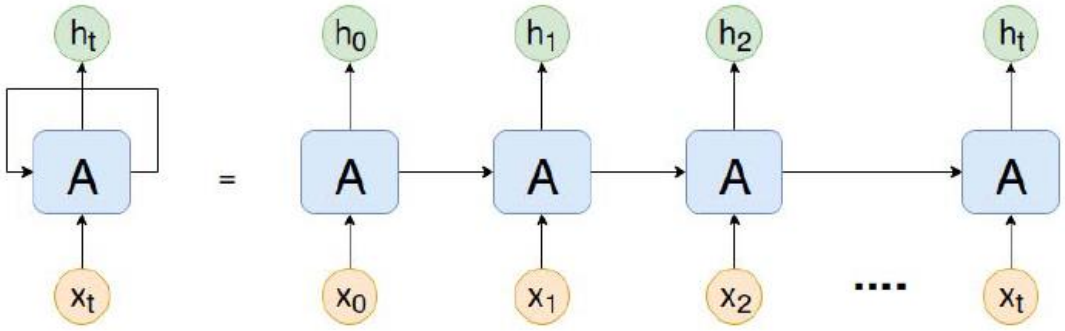


Figure 2.3: An Unrolled Recurrent Neural Network [10]

Although RNN can model a sequence of data so that each sample can be assumed to be dependent on previous ones, however it experiences the gradient vanishing and exploding problems [11]. They have a limited capacity to learn 'long-term dependencies'. For example, to make a forecast at time step t = 100, we may want some information of what we observed at time step t = 10. However, because of the way RNNs are designed, the information obtained at time step t = 10 is almost completely lost once we get to t = 100. To deal with such a shortcoming, a more sophisticated recurrent neural network called a Long Short Term Memory (LSTM) Network was developed.

2.2.2 Long Short Term Memory (LSTM)

Long Short-Term Memory (LSTM) [12] networks are a modified version of recurrent neural networks, which makes it easier to remember past data in memory. The vanishing gradient problem of RNN is resolved here. LSTM is well-suited to classify, process and predict time series given time lags of unknown duration. It trains the model by using back-propagation.

In addition to the recurrent component h_t, the model also includes a long term memory component c_t which is manipulated at each time step through various 'gates'. The training algorithm is identical to the RNNs except that an LSTM has more parameters involved. As shown in Figure 2.4, there are three gates present in an LSTM network which are forget gate, input gate and output gate. An input gate is used to decide whether a new value could flow into the memory or not while a forget gate is used to determine whether it should remain in the memory. Then whether the value in memory could be used to compute the output unit is determined by the output gate. These gates ensure that gradient-based optimization methods could be used to train the LSTM.

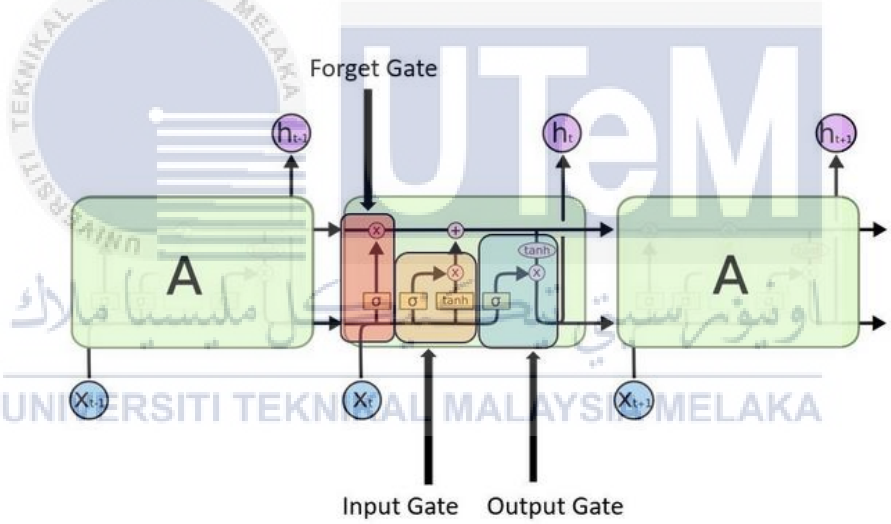


Figure 2.4: The architecture of long short-term memory (LSTM)

2.3 Wi-Fi Properties

Wi-Fi is a technology that provides network connectivity for devices through radio waves based on the 802.11 family of standards [13]. When wireless transmitter receives data from the internet, it converts the data into a radio signal that can be received by Wi-Fi enable devices. The data is then exchanged between the transmitter and the device. The basic specification for Wi-Fi which allowing two megabytes per

second of data transfer wirelessly between devices had sparked a development of routers to comply with IEEE802.11. The ubiquitous characteristic of Wi-Fi in indoor and outdoor environment has made it become an ideal candidate for RF sensing to capture the changes in the environment. As compared to the traditional sensor, Wi-Fi is capable of monitoring an area which is much larger and crowed [14][15].

Since Wi-Fi signal is sensitive to environment dynamics, it has now been widely as a 'sensor' for various sensing tasks, such as health monitoring, gesture recognition, contextual information acquisition, and authentication as showed in Figure 2.5. Computer vision seems to be perfect in the existing health monitoring system. However, it raises privacy issues, and it works under line-of-sight (LOS) only [16]. Therefore, these limitations can be overcome by utilizing using Wi-Fi sensing. The multi-path effect of Wi-Fi which is typically used for data communication can now use to perform sensing tasks under NLOS scenarios. Studies [17][18] had shown that fall will cause a sudden change in the Channel State Information (CSI) values, which is one of the fine-grain information of Wi-Fi. By utilizing the amplitude and the phase differences across the antennas, fall can be detected at an accuracy at over 87% [19]. Figure 2.5 shows the overview of applications based on Wi-Fi sensing.

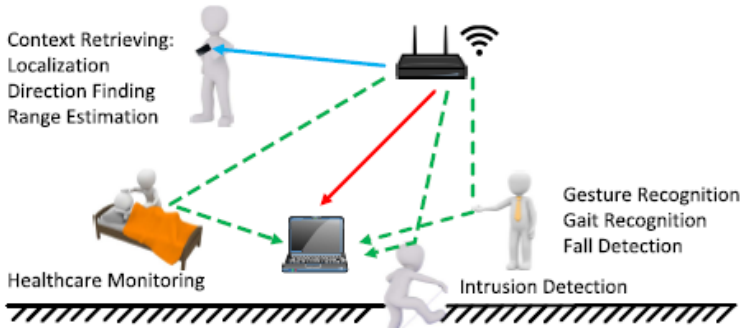


Figure 2.5: An Overview of Applications based on Wi-Fi Sensing [17]

The attributes of Wi-Fi signal which could be used to extract information about the human activities which are RSSI and CSI and will be explained in the following section.

2.3.1 Received Signal Strength Indicator (RSSI)

Received Signal Strength Indicator (RSSI) is the measurement of taking the signal power into a known channel fading model to estimate the distance between the transmitter and receiver in device-free sensing. In short, RSSI contains the MAC layer information of Wi-Fi which represents the strength of the received signal compared to the transmitted signal. The RSSI can be expressed as followed:

$$P(r) = P_0 - 10_n log_{10} \frac{r}{r_0} \quad (4)$$

Where P (r) is the received signal power (dB) measured at a distance r; P_0 is the received signal power measured at a reference distance r_0 and n denotes the path loss exponent. It can be seen from (4) that the signal power decreases as the propagation distance increase. Therefore, it can be used to estimate the distance between the receiver and the transmitter.

The RSSI is affected by the surrounding environment when a signal propagates for the transmitter; it gets reflected by different surfaces before reaching the receiver. This phenomenon is known as multi-path which the signal reached at the receiving antenna by two or more paths. RSSI is given as the superposition of all these multi-path signals received. The properties such as delay, attenuation and phase shift are different among the multi-path signals [13]. Figure 2.6 illustrates an example of the multi-paths and the resulted signal received at the receiver.

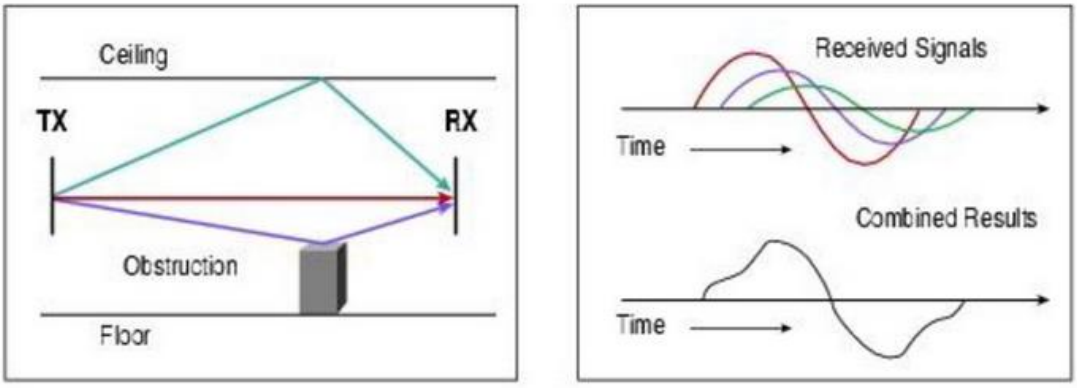


Figure 2.6: Multi-paths in Wi-Fi signals [13]

The complex baseband signal voltage measured at the receiver at a specific time is denoted as

$$V = \sum_{i=1}^{N} ||V_i|| e^{-j\theta_i} \tag{5}$$

Where V_i and θ_i are amplitude and phase of the i^{th} multi-path component and N is the total number of components. The RSSI could be given as the power of the received signal as:

$$V = 10 \ \ log_2 ||V||^2 \tag{6}$$

From (6), each multi-path component contributes to the RSSI value. The changes in the RSSI values cannot be modelled for the moving bodies as a slight change in environment may cause multi-path components to add up in constructive or destructive manners.

2.3.1.1 RSSI-Based Human Activity Recognition System

Scholz et al. [20] presented a device-free and device-bound activity recognition using 802.15.4 RSSI. The proposed system can classify activities such as sitting, lying, standing and walking. The data is collected in two different scenarios. The first

scenario (device-bond) required user to perform activity while carrying along with a wireless node; while the second scenario (device-free) required a user to perform activity within a wireless sensor networks without carrying a node. An average accuracy rate of 89%, had been achieved in both scenarios by using kNN (k-Nearest Neighbors) classifier.

From the examination of raw data, they found that activity 'walking' has a more substantial impact on the signal for device-bond compared to device-free. Besides, the changes in signal magnitude level are affected by the size of an object being in the wireless sensor network for device-free. While for device-bond the parameter is affected by the motion of the transceiver itself. Besides, the relationship on the impact of the subject on the proposed recognition system had been tested by applying the trained data from one subject to another subject and vice-versa. The results highlighted that device-bond activity recognition is strongly affected by the intense activity signal from the motion of the body. In contrast, device-free activity recognition solely depends on the link strength between the infrastructures.

Gu [21] presented a Wi-Fi assisted human activity recognition method by applying the data mining techniques to abstract footprints of different activities on the radio signal strength (RSS) data. The proposed system had achieved an accuracy of 75% by using a single feature and the common k-NN classifier. Moreover, the average recognition ratio can be achieved at 75% to 92.58% by using the proposed fusion algorithm as shown in Figure 2.7.

The impact of different environments on the RSS data had been studied in this paper by experimenting in indoor and outdoor environment. Due to more substantial interference caused by uncontrollable factors such as passing-by objects and other Wi-

Fi APs in outdoor environment, the system had a more excellent performance for the indoor environment.

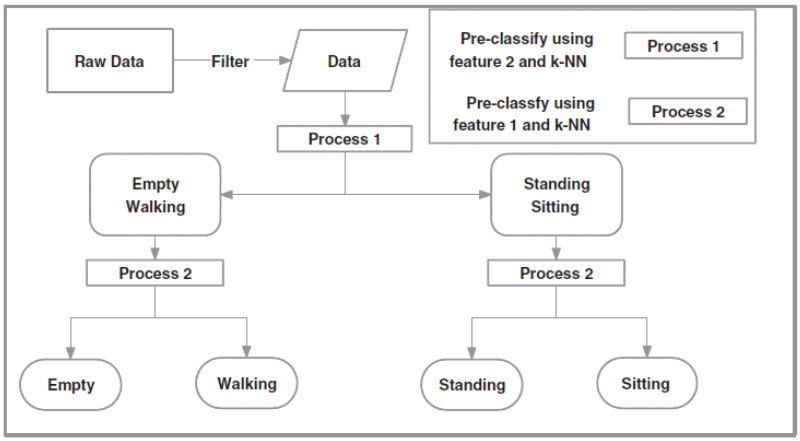


Figure 2.7: The fusion algorithm [21]

Abdelnasser et al. [22] proposed a gesture recognition system termed Wi-Gest with Wi-Fi RSS measurements. Wi-Gest is a system that able to sense in-air hand gestures around the user's mobile device by leveraging the relationship between the changes in RSS and the human movement. The system can control the media player action without modification on the standard Wi-Fi equipment and without involving training for gesture recognition. An accuracy of 87.5% had been achieved in Wi-Gest by using a single AP.

Figure 2.8 shows the overview layers proposed in Wi-Gest system which cover primitives, gesture families, and application actions layers. In primitive's layer, the basic changes in the raw RSSI will be detected, which are further used to define the speeds and magnitudes of the gestures. The gesture family layer represents a set of gestures that consists of the same sequence of primitives, which give the same effect on the signal strength. This layer allows the developers to choose their own gestures for their applications. The number of consecutive up-down gestures and the speed of

the repetition can be extracted from this layer. The gestures performed are mapped to its action in the application action layers. The system allows users to control the same function by performing different gestures.

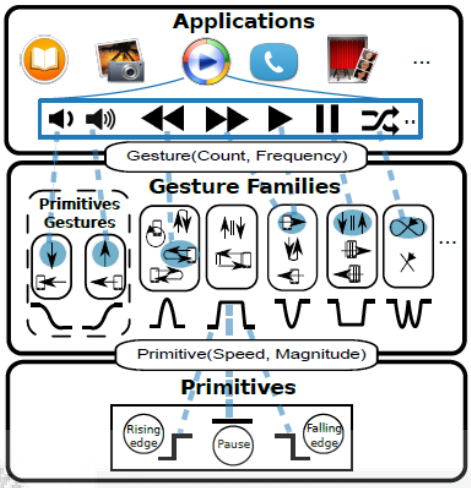


Figure 2.8: Layer overview of Wi-Gest system [22]

Figure 2.9 shows the processing flow of the Wi-Gest system. The basic primitives are extracted from the raw signal during the stages of primitive extraction. Discrete Wavelet Transform (DWT) is selected to reduce the noise contained in the raw data as it can achieve fine-grained multi-scale analysis [23]. The wavelet-based de-noising method [24] had been applied to enhance the performance of this system as it does not make any particular assumptions about the nature of the signal and allows discontinuities in the signal.

In gesture identification stage, different gestures and their attributes which include frequency and count, are extracted by segmentation and matching. The purpose of segmentation is used to determine whether there is a gesture generating by the user. The string pattern is extracted and compared with the gestures templates to find the best match after the gesture boundary is determined. Action mapping is the process

where the developer can determine which actions are mapped to which gesture family by using the count and the frequency attributes.

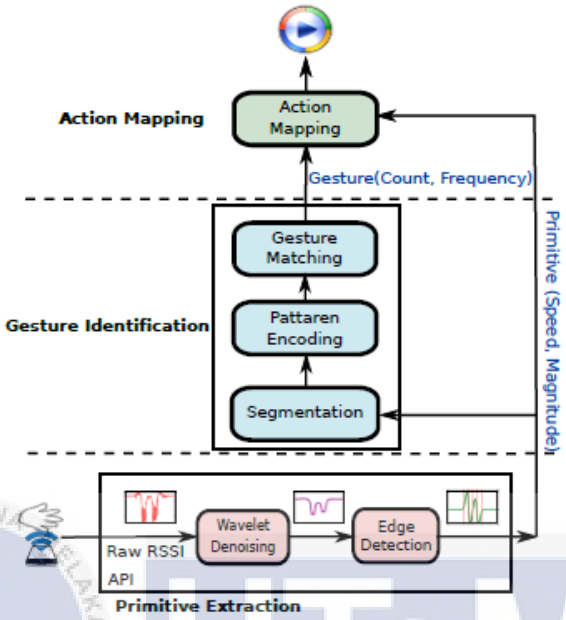


Figure 2.9: Logic flow of Wi-Gest system [22]

Table 2.1: Results of the proposed approach for RSSI-Based Human Recognition

Literature	Classifier	Accuracy
Scholz et al. [20]	kNN	89.00%
Gu [21]	kNN + fusion	75.00 % to 92.58%
Abdelnasser et al[22]	Primitives	87.50%

2.3.2 Channel State Information (CSI)

In modern digital wireless communication systems, OFDM is widely used (e.g., in Wi-Fi standards such as IEEE 802.11a/g/n) to combat frequency selective fading in multi-path propagation environments. As shown in Figure 2.10, at the OFDM transmitter, data is encoded and mapped into multiple orthogonal subcarriers and then transmitted over the subcarriers. With inverse Fast Fourier Transform (IFFT), the

subcarriers are converted from the frequency domain to the time domain. To reduce the inter-symbol interference (ISI), the cyclic prefix is added in the time domain. Then, in-phase and quadrature (I-Q) modulation is used for transmission in the multi-path channel. The digital data is converted into analog data with the Digital to Analog Converter (DAC). Finally, the analog signal is up-converted and amplified by the high power amplifier (HPA). At the OFDM receiver, the signal is down-converted to the baseband. The Automatic Gain Controller (AGC) can compensate for the signal amplitude attenuation. The inverse process of that at the transmitter is implemented for recovering the data at the receiver. Figure 2.10 shows the block diagram of an OFDM transceiver.

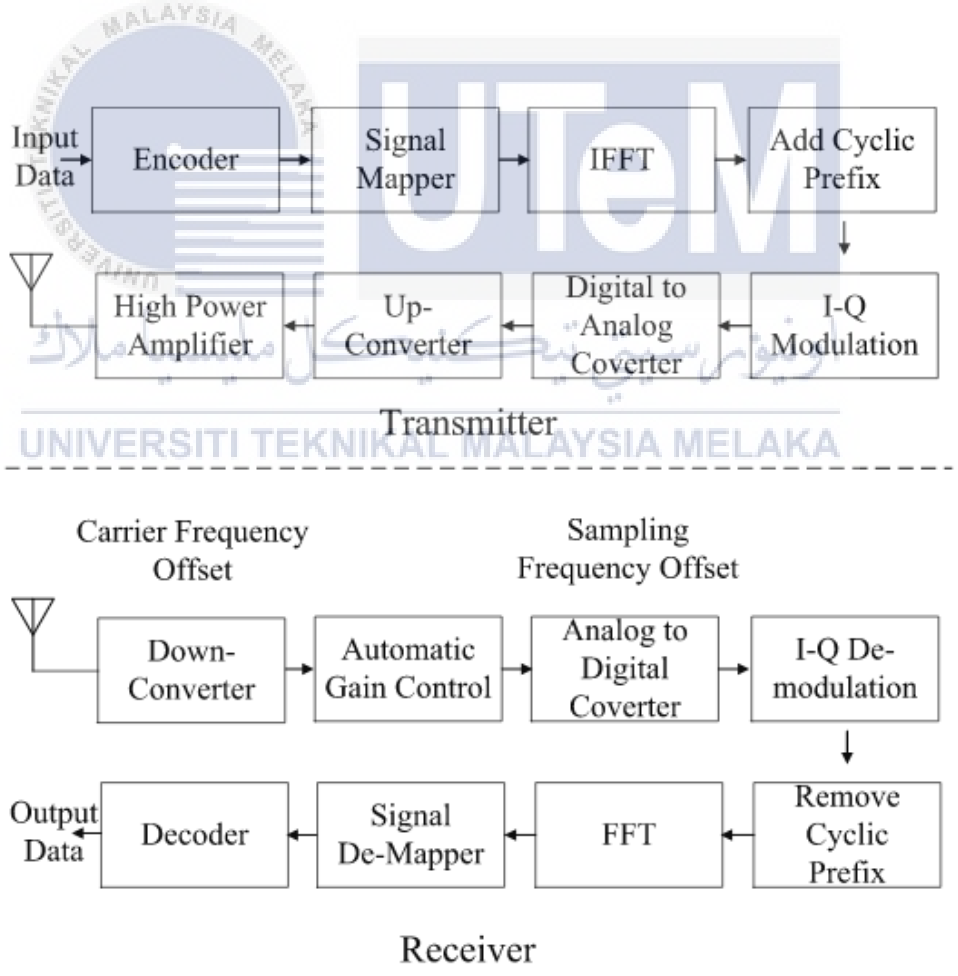


Figure 2.10: The block diagram of an OFDM transceiver [26]

CSI is considered the new trending metric in Wi-Fi –based sensing technology. By modifying the device driver of off-the-shelf NICs, i.e., Intel's IWL 5300, CSI as fine-grained PHY information can be obtained by using the CSI-Tool presented by Halperin et al. [27], which represents the subcarrier-level channel measurements. In addition, CSI describes the channel properties experienced by the packet. For example, a wireless signal in propagation may undergo considerable impairments due to shadowing, multi-path propagation, and distortion, which are reflected in the CSI.

CSI is the collection of information that describes how wireless signals propagate from the transmitter to the receiver. To precisely define the CSI, some background knowledge about Multiple Input Multiple Output (MIMO) technology is needed. Figure 2.11 shows a MIMO equivalent model. The received signal of the j antenna can be defined as:

$$y_i(t) = \sum_{i=1}^{n_t} h_{i,j}(t) * x_i(t) + n_j(t), i = 1,2, \dots n; j = 1,2, \dots, n_r, \quad (7)$$

where H (i, j) is the channel fading factor between the transmitted antenna i and received antenna j. X_i is the transmitted signals of the antenna i. Y_i is the receive signal of the antenna j. Considering the narrowband flat fading channel, Equation can be simplified as:

$$y_j(t) = \sum_{i=1}^{n_t} h_{i,j}\, x_i(t) + n_j(t) \quad (8)$$

which can be expressed as:

$$y(t) = Hx(t) + n_j(t), \quad (9)$$

where MIMO system transmit matrix X(t), MIMO system receive matrix y(t), channel additive white Gaussian noise matrix n(t), and channel fading factor matrix H are represented as:

$$H = \begin{bmatrix} h_{1,1} & \cdots & h_{1,n_t} \\ \vdots & \ddots & \vdots \\ h_{n_r,1} & \cdots & h_{n_r,n_t} \end{bmatrix} \tag{10}$$

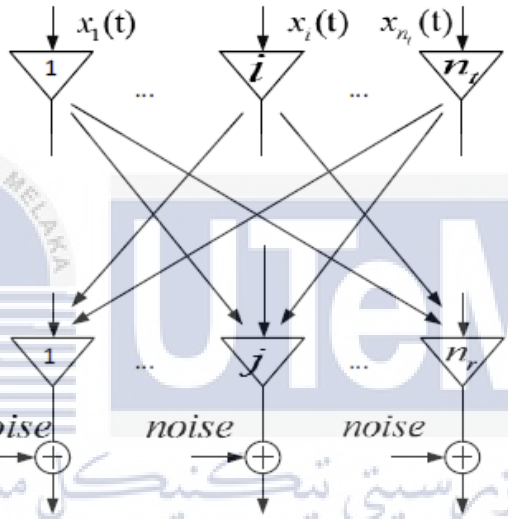


Figure 2.11: Multiple Input Multiple Output (MIMO) equivalent model. [28]

The CSI uses the channel fading factor matrix H, that is defined as Equation (5). Each element in the matrix is complex represented [29] as Equation (6):

$$H_{i,j}(f_k) = H_{i,j}(f_k)e^{j<H_{i,j}(f_k)}, \tag{11}$$

where f_k is the central frequency of the OFDM subcarrier that is defined in the 802.11n protocol,$H_{i,j}(f_k)$ is the amplitude, and $< H_{i,j}(f_k)$ represents the phase shift information.

2.3.2.1 CSI-Based Human Activity Recognition System

In E-eyes [30], a device-free location-oriented activities identification by extracting the channel state information (CSI) from the IEEE802.11n had been proposed by Yan Wang. The proposed system can classify activities into in-place activities and walking activities. The main idea in E-eyes is to match the CSI pattern against the activities profile. An accuracy of 96% true positive rate had been achieved in this system.

The system started with collecting the time-series amplitude measurement in each subcarrier on a link as shown in Figure 2.12. The collected data is then pre-processed to remove outliers by employing the use of low pass-filter. Dynamic exponential smoothing filter (DESF) is applied to smooth the factor dynamically based on the previous sample. This filter not only removes high free frequency but also preserved the features caused by human activities. The data also filtered by removing all the CSI measurements with MCS index less than 263 as those data are unstable even through the environment does not change.

A moving variance thresholding technique is applied to differentiate the activities between in-place activities or walking activities. This technique is selected as the CSI amplitude will changes significantly during walking activities while changes slightly for an in-place activity. The start and end of an activity also can be determined by the moving variance. Next, the activities are identified by calculating the similarity between the CSI segment and the pre-constructed activity profiles. Multiple-Dimensional Dynamic Time Warping (MD-DTW) technique is employed for walking activities to align a trace with larger CSI changes to the profile and correcting the difference in speed. While for in-placed activity, the Earth Move Distance (EMD) is used to qualify the similarity of two CSI histogram distributions.

In E-eyes, the activities profile is constructed by continuously monitoring the home environment. Then, the clustering-based method is applied to identify multiple similar instances of an activity without a matching profile. Then the resulted clusters can be used to construct known activity profiles. This technique enables users to detect and update the activity profile if there are any significant changes in the environment such as the Wi-Fi device had been moved from one place to another place.

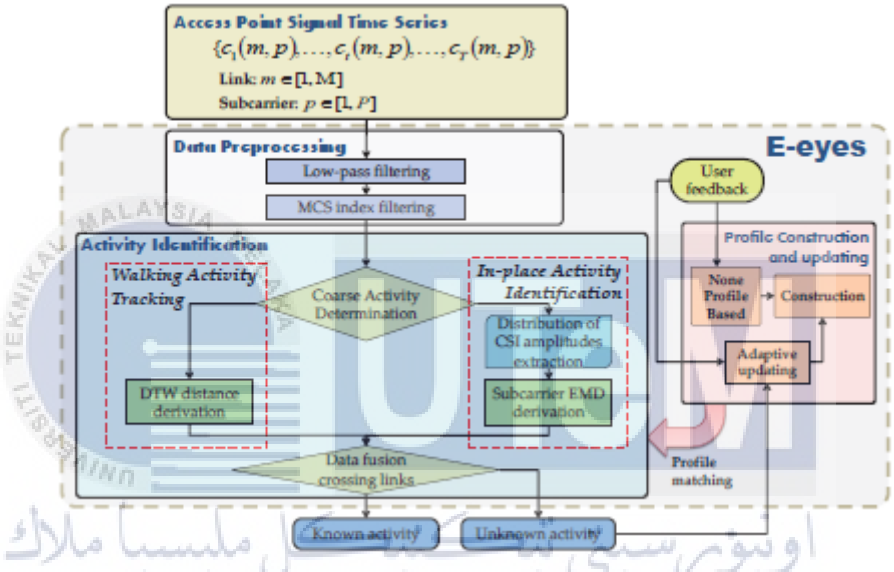


Figure 2.12: System flow of activities recognition using CSI in E-eyes [30]

Wenjun Jiang presented an EI [25], a deep learning-based device-free activity recognition framework that can remove the specific information of environment and subject contained in the activity data. The proposed system is able to extract the independent features of environment or subject shared by the data collected on different subjects under different environments. An accuracy of 75% had been achieved in this system in recognizing activities such as wiping the whiteboard, walking, moving a suitcase, rotating the chair, sitting, standing up and sitting down.

Figure 2.13 shows the system framework in EI which consists of three major components, which are data collection, data pre-processing and deep learning model. The system is started with collecting the CSI data for each activity in a different environment. The data collected is separated into label data and unlabeled data. The CSI data had been interpolated to obtain a uniform sampling period and is normalized to have a mean of zero and a standard deviation of one. Hampel filter is applied to remove the outliers and reduce the sample of the CSI measurements. Finally, the transformed signal is split into short segments to train the activity recognition model. A deep learning model that incorporates an adversarial network is employed to predict the unlabeled data.

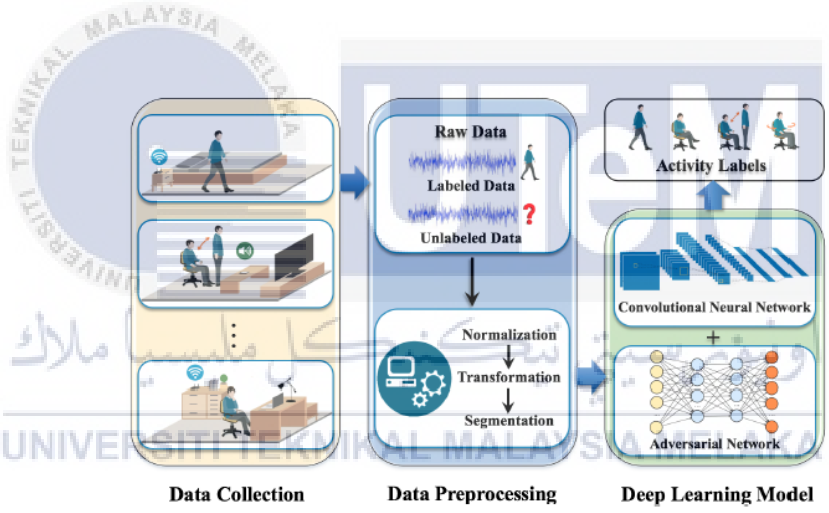


Figure 2.13: System framework of EI [25]

Figure 2.14 shows the core of EI, which consists of three main components which are feature extractor, recognizer, feature activity, and domain discriminator. CNN is employed to extract activity features. 2D kernels are used as the filters in each layer of CNNs, then the mean and variance of the data at each layer is normalized by a batch norm layer. A rectified linear unit (ReLU) and a max-pooling layer are used to introduce the non-linearity and reduce the size of the representation respectively. From

the output of the feature extractor, a fully connected layer, which followed by an activation function, is used to introduce non-linearity.

A domain discriminator is used to recognize the environment where the activities is recorded so that the environment independent activity features can be produced from the feature extractor. In order to prevent over fitting due to insufficient data, confidence constraints and smoothing constraints are designed so that the unsmooth latent space of deep neural networks and the over confidence problem can be solved. Balance constrains is designed to deal with the condition where the same label is assigned to the data samples corresponding to multiple similar but different activities by the model.

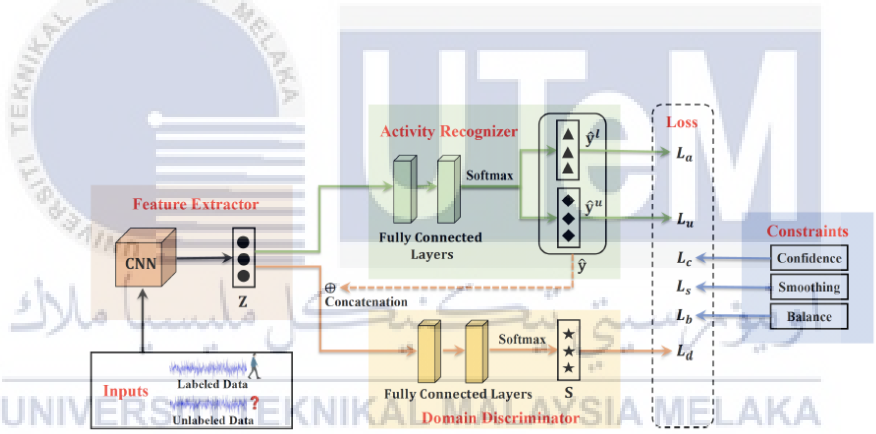


Figure 2.14: Model Overview in EI [25]

A CARM [31], CSI based human activity recognition and monitoring system had been proposed by Wei Wang. CARM consists of two Commercial Off-The-Shelf (COTS) Wi-Fi devices, as shown in Figure 2.15. Router is used to continuously sending signals while the laptop is used to receive the signals continuously. The system consists of 2 models, which are CSI-speed model and CSI-activity model. The correlation between CSI value dynamics and a specific human activity is built by combining two of the models. A set of signal processing techniques, such as PCA

based de-noising and DWT based feature extraction, for human activity recognition based on the CSI-speed model and the CSI-activity model had been proposed in this paper. CRAM had achieved an average activity recognition accuracy of 96%.

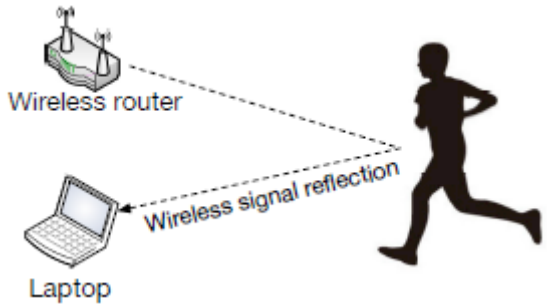
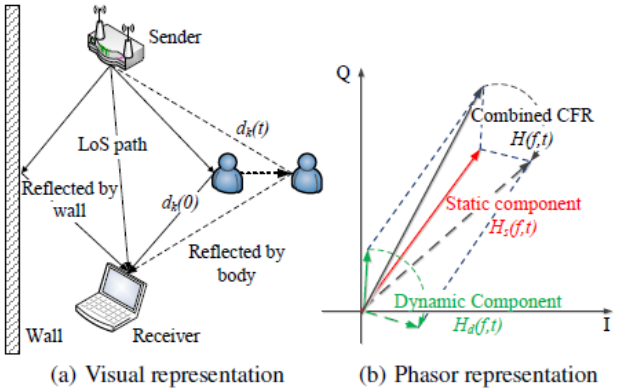


Figure 2.15: CARM system [31]

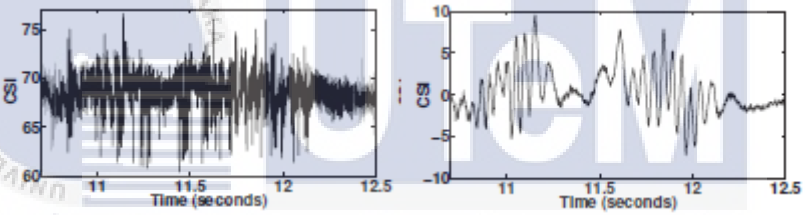
The CSI-speed model is designed by relating the variations in CSI power to the movement speeds as the CFR power varies according to lengths of multi-path as show in Figure 2.16. In the CSI-speed model, the Hilbert Transform is used to calculate the phase change of the waveform. The CSI-activity model which quantifies the correlation between the movement speeds of different human body parts and a specific human activity is built by using the energy profile of different frequencies. Hidden Markov Model (HMM) is used to build CSI activity models that consist of multiple movement states. Information from all training samples can be capture by this model even there is high within-class variance.



(a) Visual representation (b) Phasor representation

Figure 2.16 : Multi-paths caused by human movements [31]

The HMM model is built by firstly collecting the CSI data and the noise is removed by Principle components analysis (PCA) which is a technique in identifying correlations and patterns in a data set so that it can be transformed into a data set of significantly lower dimension without loss of any important information. Figure 2.17 shows the time-series of CSI values after being filter by PCA. Instead of Short-Time Fourier Transform (STFT), discrete wavelet transform (DWT) is applied to extract human movement features to decompose the PCA components. This is because DWT has better tradeoffs in time and frequency resolutions which can capture both high and low speed movements.



(a) Original CSI Stream **(b) PCA based denoising**

Figure 2.17: Denoising the time-series of CSI values [31]

Wi-Motion [32], a Wi-Fi based human activities recognition system had proposed by Heju Li. Two respective classifiers are constructed by adopting the amplitude and phase information extracted from the CSI sequence and the outcome is combined by using a combination strategy based on posterior probability. A mean accuracy of 98.4% had been achieved in Wi-Motion in recognizing six human activities.

In the amplitude data preprocessing stage, the method weighted moving average (WMA) is used to remove the noise from the CSI data. In order to reduce the dimensionality of the filtered data is then reduced by applying PCA algorithm in order

to reduce the complexity of the system. In classification stages, discrete wavelet transform (DWT) is applied to analyze signals on multiple frequency scales and extract the features.

In the phase data pre-processing, a linear transformation is executed to mitigate the effects of random noise. The algorithm which proposed by Wang et al [33] is performed to calibrate the phase data so that it becomes more stable. The CSI phase difference is simplified by singular value decomposition (SVD) in classification stages.

Table 2.2: Results of the proposed approach for CSI-Based Human Recognition

Literature	Classifier	Accuracy
E-eyes [30]	EMD + MD-DTW	96.0%
EI [25]	CNN	75.0%
CARM [31]	HMM	96.0%
Wi-Motion [32]	DTW + SDV	98.4%

2.3.3 Comparison between RSSI and CSI

Channel state information is the physical layer information of the Wi-Fi signal. For OFDM Wi-Fi signals, CSI gives the channel response of the system. In contrast with the RSSI information, CSI provides information on different channels instead of a cumulative signal strength indication. Hence frequency selective fading could be observed as a loss of information from particular channels, and so CSI is more informative than RSSI. Figure 2.18 shows the analogy between the RSSI and CSI for better understanding.

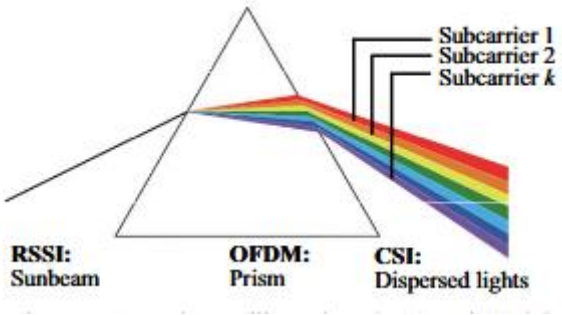


Figure 2.18: An analogous representation of CSI and RSSI [34]

Since this research focuses more on human activity recognition by using Wi-Fi signals, therefore there is a need to evaluate the characteristic for both Wi-Fi attributes before deciding which of them will be used as fingerprint in this research.

Device–Free Passive Localization by using LTE Signal as fingerprinting by Pecoraro [34], had proved that performance for CSI-based is better than RSSI-based. This is because the received signal strength is never consistent as it actually gets affected by many external factors, including temperature humidity and air pressure, as the signal which the router is emitting is never consistently strong. In contrast to RSSI, the CSI data comes closer to the physical layer of route and the signal is relatively consistent. An experiment had been conducted by Li [35] to show the superiority of CSI over RSSI as shown in Figure 2.19. Therefore, the Wi-Fi CSI data is chosen to be used in this research as although the environment noise may distort some streams, but the information in CSI is still more than RSSI. The summary for the comparison between the RSSI and CSI is shown in Table 2.3 as well.

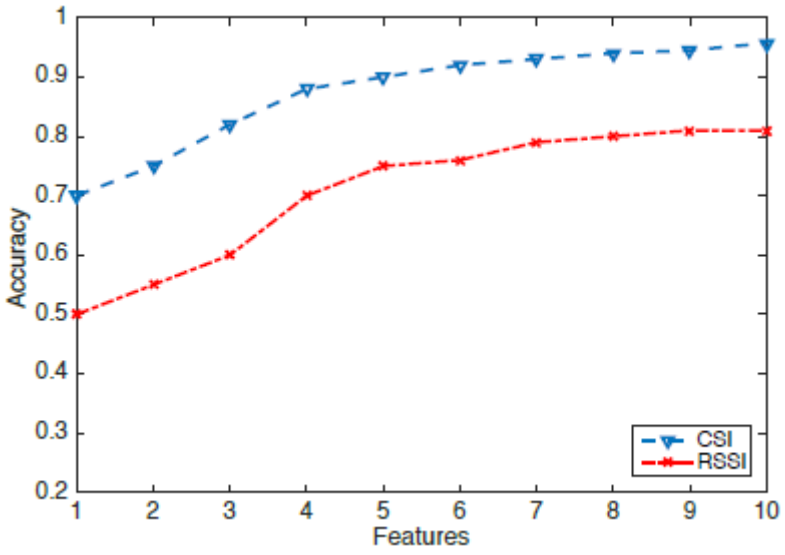


Figure 2.19 : Channel State Information (CSI) vs Received Signal Strength Indicator (RSSI) in an LOS scenario [35]

Table 2.3: Comparison between Receives Signal Strength Indicator (RSSI) and Channel State Information (CSI)

Metric	RSSI	CSI
Network Layer	MAC layer	Physical Layer
Time resolution	Packet Size	Multipath signal cluster scale
Frequency resolution	No	High
Temporal stability	Low	High
Measurement Band	RF Band	Base Band
Gradularity	Coarse-grained (per packet)	Fine-grained (per subcarrier)
Universality	Almost all Wi-Fi devices	Some Wi-Fi devices

2.4 Related Work

The state of the art methods that use Wi-Fi to track human activities may be categorized as coarse-grain, fine-grain or super-fine grain detection. Applications that use coarse grain detection only detect the presence of human motion such as instruction detection, where the presence of a moving human can be detected. For fine-grain detection, activity such as running, walking, and falling can be classified and detected. Last but not least, super-fine grain detection focuses on movement of a specific body part, such as the movement of the chest in monitoring heart-beat rate. In this section, the approaches to achieve activity detection for these three granularity levels will be summarized.

2.4.1 Coarse-Grain Activity Detection

A novel scheme for device-free passive detection of moving humans with dynamic speed (PADS) had been proposed by Kun Qian et [36]. By implementing the space diversity offered by multi-antenna systems in MIMO, the researcher used the amplitude and phase information obtained from the CSI data and achieved an almost 99% true detection with no false detection in detecting the dynamic speed of human movement. To achieve this, firstly the raw data is pre-processed which outlier filtering is applied to sift out the skewed amplitude information. For phase information, the phase differences between antenna pairs are measured and unwrapped to remove the substantial random noise while retaining its sensitivity towards any changes in environment. After the data is pre-processing, the respective three maximum significant Eigen values of correlation matrix of CSI are used for feature extraction so that it is extendable to various situations without a precise configuration of the environment. Next, the presence of human moving is estimated by implementing the

Support Vector Machine (SVM), a supervised machine learning algorithm that mostly used in classification problems. Figure 2.20 shows the architecture overview of PADS.

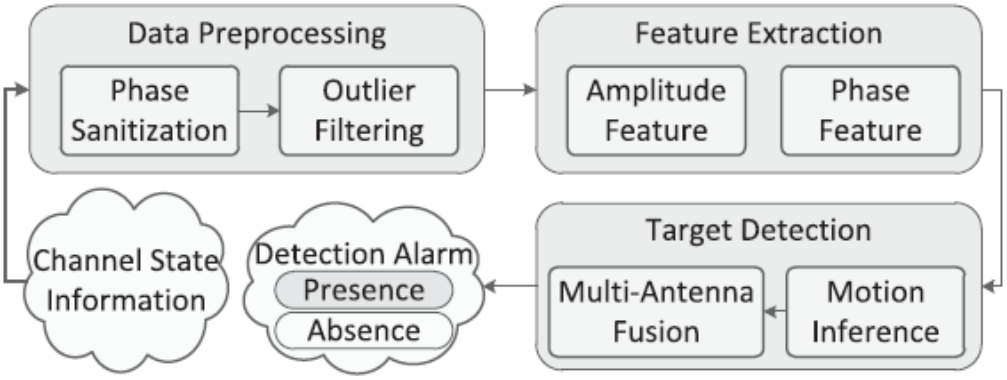


Figure 2.20: Architecture overview of PADS [36]

By employed a single Wi-Fi link, a count estimator [37] that had been proposed by Ibrahim to estimate the human count in specific area The proposed system had achieved a human counting accuracy of 100% within a maximum of 2 persons within the indoor area where a single Wi-Fi link whose transmitter and receiver are behind the walls. The key idea is to focus on the temporal link-blocking pattern as a discriminating feature that is more sensitive to wireless channel noise than the signal power so that higher accuracy can be achieved. Firstly, the RSS data for single person is obtained from the receiver and is used to generate the training set for all of the other count classes as well. Instead of using the typical way which the data is prepared through collecting the real data for different counting of persons inside the area, the training set for multiple-user classes is generated by using the training data synthesizer module.

Since the system is designed by implementing the uses of deep learning, which is known as long short term memory (LSTM), algorithms which required large data to train a model, therefore a module namely training classes balancer is used to

synthesize the data. During the training the blockage pattern for each human count class is being characterized. Then, the trained module is implemented in the Count Estimator module to predict the result after the LoS Blockage Detector module processed the RSS value received. Figure 2.21 shows the architecture overview of CROSSCOUNT [37].

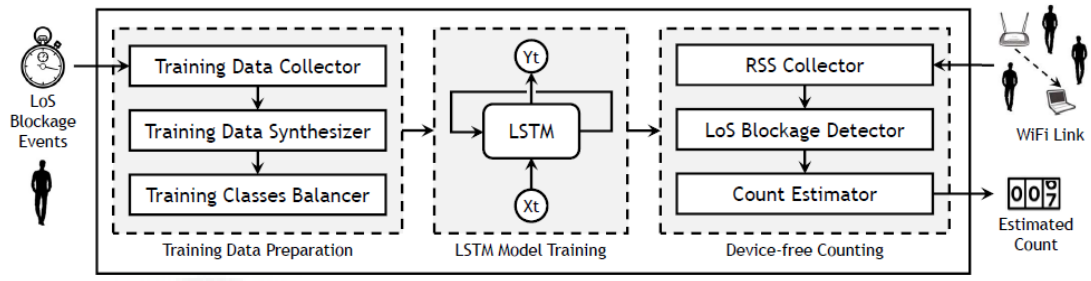


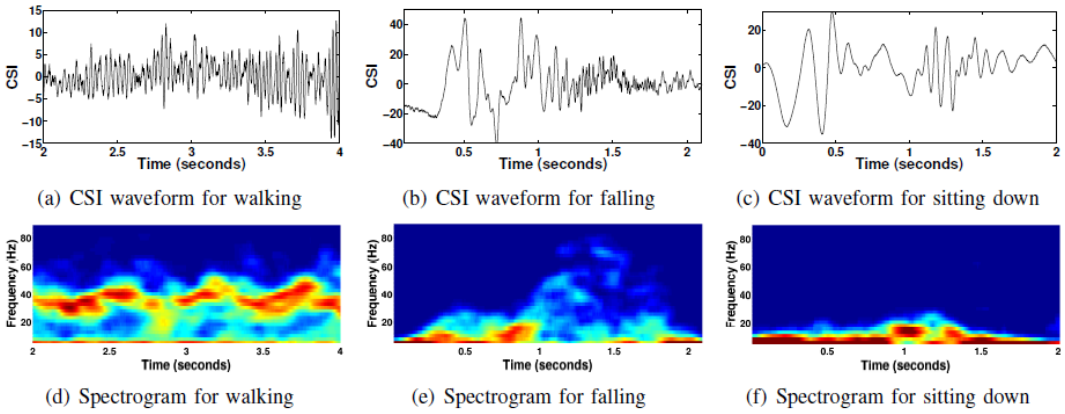
Figure 2.21: Architecture overview of CROSSCOUNT [37]

Avinash Kalyanaraman et all [38] proposes a radar-based system to track people based on body shapes known as FormaTrack. The idea behind this system is to scan the body with a radar sensor and to analyse the reflection profile based on the amount of energy that reflects back from each part of the body. This system can achieve a tracking accuracy at over 99% with 2 people in a home with 5 or more rooms. To achieve this, firstly a radio transmitter is placed just above the door frame to detect people which the range for the radio system used is only about 1 meter. When there is people pass by, the electromagnetic energy pulse generated from the pulse radar will get reflected. Therefore, the energy reflected when people pass by supposed to be higher than when no one pass through. Then, a 3-frame moving target indication (MTI) filter is used to remove all the unwanted static clutter. Next, the total power in each radar frame is computed as the sum of the absolute reflected power at each distance in the frame and is filtered via a two-stage discrete FIR filter to remove impulse noise. A crossing is detected when the total radar frame power exceeds

threshold in the first filter. While the second filter is used to increase the system accuracy by filter out those false positives situation due to the transient device noise which caused high radar frame power.

2.4.2 Fine-Grain Activity Detection

An activity recognition system had been proposed by Wei Wang et al [39] by modelling the correlation of CSI dynamics with human activities. CSI speed model which quantifies the relation between human speed of movement and CFR power of all the channels and CSI activity model which relates speed of movement and type of action is introduced in this system. The concept behind the CSI speed model is that CFR power changes according to the change of path length when the lengths of multi-path change due to human motion. The higher frequency channel and lower frequency channel will get affected easily by movements in high speed and low speed respectively. Figure 2.22 shows activity such as walking has higher frequency components continuously compare to sitting and falling. While a high frequency component for very small duration only is consisted in activity falling. The proposed system achieves a cross validation accuracy of 96.5% for 7 different activities. The accuracy is achieved by training the Hidden Markov Model (HMM) module through de-noised CSI data. Then the transition from lower speed of movement to higher is modelled by HMM. The dimensionality and noisy components is removed through Principle Component Analysis (PCA) as well.



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

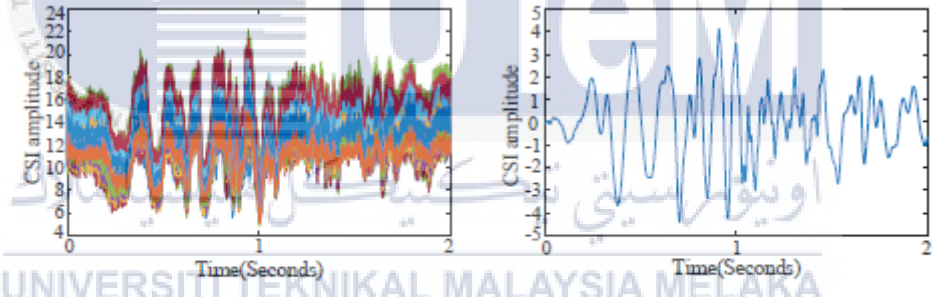
(d) Spectrogram for walking (e) Spectrogram for falling (f) Spectrogram for sitting down

Figure 2.22: Activity-speed model in CARM [39]

On the contrary to CARM, Siamak Yousefi et all [40] explores deep learning methods instead of machine learning techniques. It uses Long Short Term Memory (LSTM) under recurrent neural network (RNN) for training and classification of the activities and compares it with HMM and Random Forest method. The accuracy of LSTM is higher than the other two the experimental results could even distinguish well between similar activities like sitting, standing and lying down. The performance is evaluated by using Tensor Flow in Python. The raw CSI amplitude is used as the input feature vector for the LSTM module. In contrast with traditional approached where PCA and STFT is required to extract the feature, the features can be obtained directly through LSTM and no pre-processing steps are necessary to prepare the data for training.

A transfer-learning based device free fall detection system known as TL-Fall [41] had been proposed by Zhang. In contrast to the existing commercial Wi-Fi based fall detection systems, TL-Fall model can be trained with only a few labelled data when working in the new environment. This system is able to achieved fall with 86.83% sensitivity.

The concept behind this system is that whenever there is presence of body movement, the velocity in multi-path wills changes and it is independent from the number of multi-path changes. To achieve this, firstly the amplitude of CSI data is filtered and smoothed through band-pass filter and weight moving average method. Similar to CARM, TL-Fall uses the second principal component and applied dynamic time wrapping (DWT) to extract feature as the first component contain much higher noise. Figure 2.23 shows the extraction signal from Principal Component. In contrast to activity recognition through deep learning LSTM, signal processing is required before passing through the data for training and classification by using traditional machine learning SVM approach. Figure 2.24 shows the overview system of the TL-Fall.



(a) Original CSI Stream (b) The Second Principal Component

Figure 2.23: Principal Component Extraction [41]

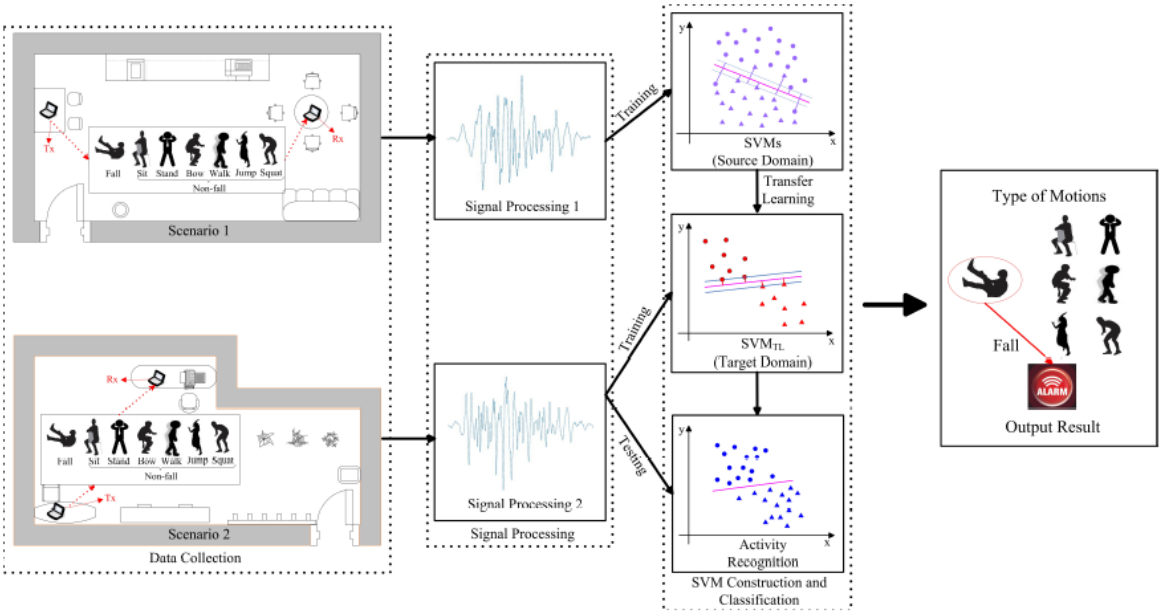


Figure 2.24: The System Overview of TL-Fall. [41]

2.4.3 Super-Fine-Grain Activity Detection

WiHear [42] which proposed by Guanhua Wang enables Wi-Fi signals to hear people talk and transmit the talking information to the detector without deploying any devices. The concept of this system is through the Mouth Motion Profile which can be generated by analysing the multi-path reflection of fine-grained radio from mouth movements. The system is able to detect at an accuracy at 91% for single individual speaking with less than 6 words and accuracy with up to 74% for a maximum of 3 people talking simultaneously. Firstly, the out-band interference of the raw signals is filtered to constructed the mouth motion profile by using discrete wavelet packet decomposition. Then machine learning is applied to recognize the pronunciations and the result is classified and undergoes error correction process to increase the performance of the system. The system framework of Wi-Hear is shown as in Figure 2.25.

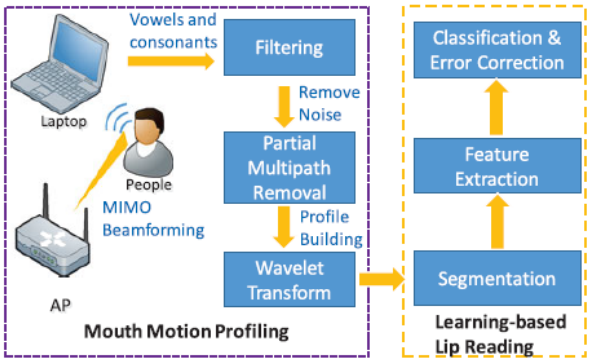


Figure 2.25: Framework of Wi-Hear [42]

Wang had proposed a multi-user respiration detection system using Wi-Fi CSI namely as TinySense [43]. The system had achieved an accuracy with over 95% of the respiratory data for the presence of single person only, while achieve an accuracy at 88% when there is some others present around the user. In order to detect multiple respirations of persons at a time, multiple TX-RX antennas pair is employed to capture the changes in Wi-Fi CSI. Then the data whose time-of-arrival (TOA) is bigger than a truncation threshold is filter to reduce the effect of surrounding people and subcarrier that are not significantly affected by multi-path had been selected before the signal curve is smooth by wavelet filter and mean filter. In the respiration detection step, peak-valley detection algorithm is used to detect the peaks of the wave, where which peak represents the occurrence of one respiration. Figure 2.26 shows the overview system architecture of the TinySense.

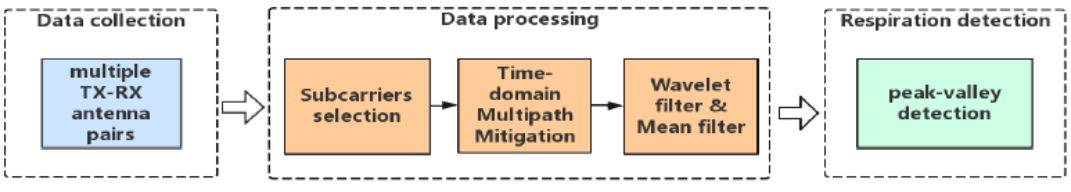


Figure 2.26: System architecture of TinySense [43]

Figure 2.27 shows the WiKey [44] system which had been proposed by Karman Ali. WiKey is the first system which is able to recognise keystrokes by using CSI Wi-Fi Signals by using one transmitter and one receiver only. This system had achieve an accuracy of more than 97.5% and 93.5% in detecting and recognizing keystrokes in a continuously typed sentence respectively. To achieve this, firstly the high frequency noises of the CSI data is removed through low pass filter. Then PCA is applied to extract the signals that caused by movement of hand with acceptable noise. Then an algorithms is applied to identify the starting and ending points of keystroke in order to reduce the detection errors. Then, the shape feature of keystrokes is extracted by DWT as DWT can reduce the number of samples in CSI which help to reduce the computation cost. Finally, k-Nearest Neighbour (kNN) classifier is used to classify for each key stroke

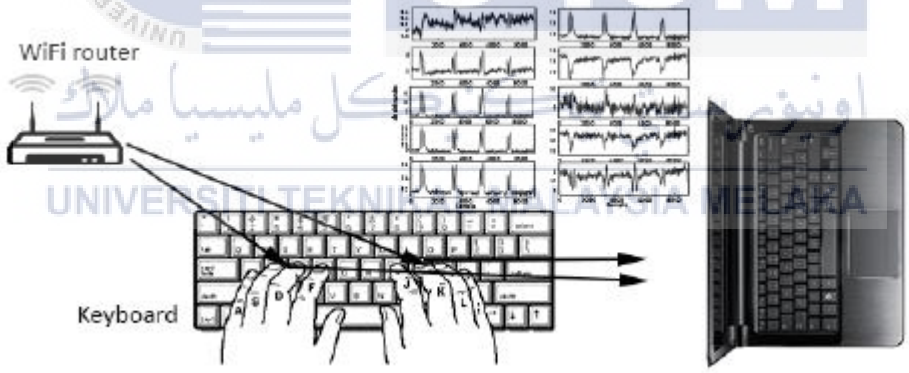


Figure 2.27: Detection Process of WiKey system [44]

2.5 Chapter Summary

An overview of general deep learning framework for RF sensing had been discussed and the network structure of RNN and LSTM had been discussed in detailed. Besides, the Wi-Fi properties such as RSSI and CSI had been compared and tabulated. The results of the proposed approach for human activities recognition by using RSSI and CSI is tabulated as well. Moreover, the related works on human activities

recognition by using Wi-Fi signal and its performance have been discussed as well in this chapter.

CHAPTER 3

METHODOLOGY

3.1 Introduction

This chapter will present the methodology of implementation a deep learning based namely LSTM network to recognize human activities by employing the Wi-Fi CSI data. Despite of the flow charts, the method selection, project management and environment and sustainability of the system will be discussed in details.

3.2 Flowchart

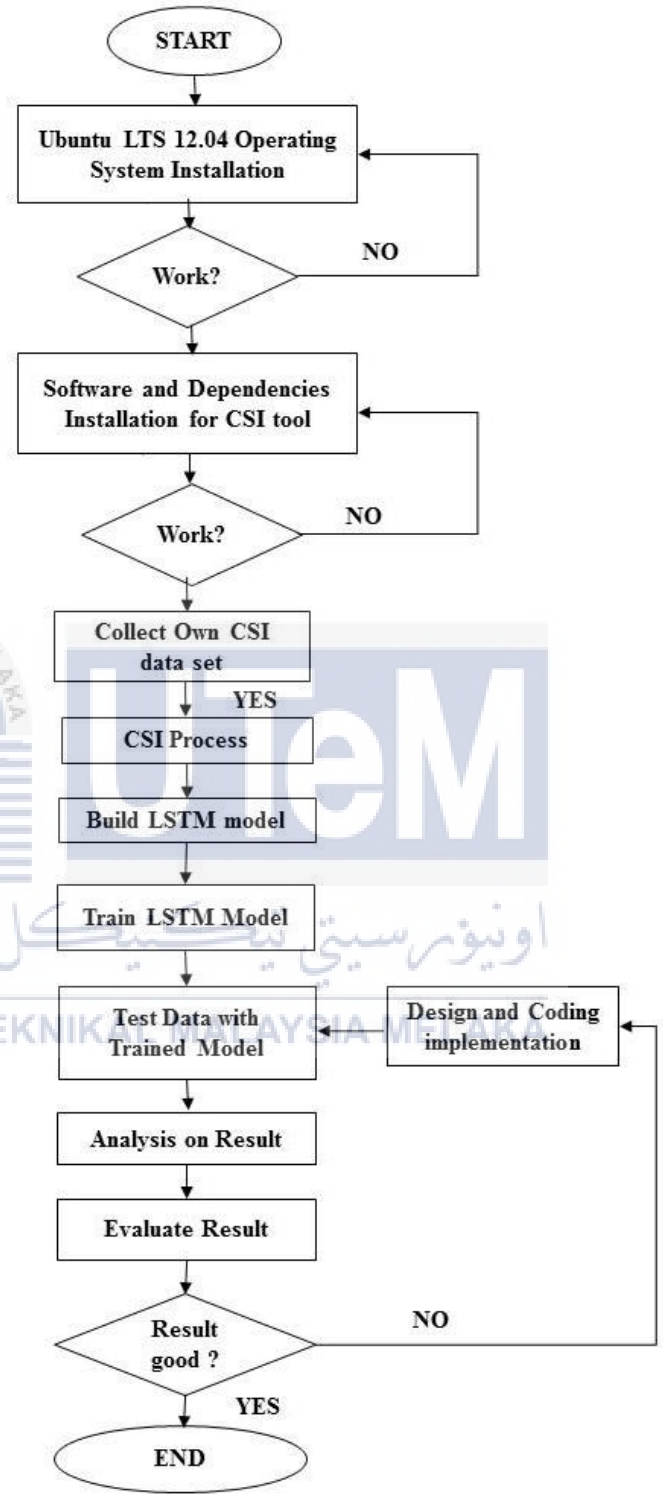


Figure 3.1: Flow Chart of This Project

3.3 Data Preparation

There are two Wi-Fi CSI datasets used in this project. First dataset is a self-collected dataset. The second data set is obtained online [40]. Table 3.1 shows the summary of human activity datasets involved in this project.

Table 3.1: Summary of Human Activity Wi-Fi CSI dataset

Dataset	Class of Activities	Number of subjects involved	Locations	Duration for each activity
Self-collected	4	1	Indoor Living Room	10 seconds
Online	7	6	Indoor Office	20 seconds

3.3.1 Self-collected Dataset

In this project, Linux 802.11n CSI Tool [27] is used to collect the Wi-Fi CSI data for different human activities. Since the CSI Tool only work on Linux operating systems that are based on an upstream Linux kernel version between 3.2 (e.g. Ubuntu 12.04) and 4.2 (e.g. Ubuntu 14.04.4). Therefore, Ubuntu 12.04 LTS with the 2.6.36 kernel is selected to install the CSI tool. Ubuntu is an operating system that includes a lot of free and open source applications and uses Linux as its core.

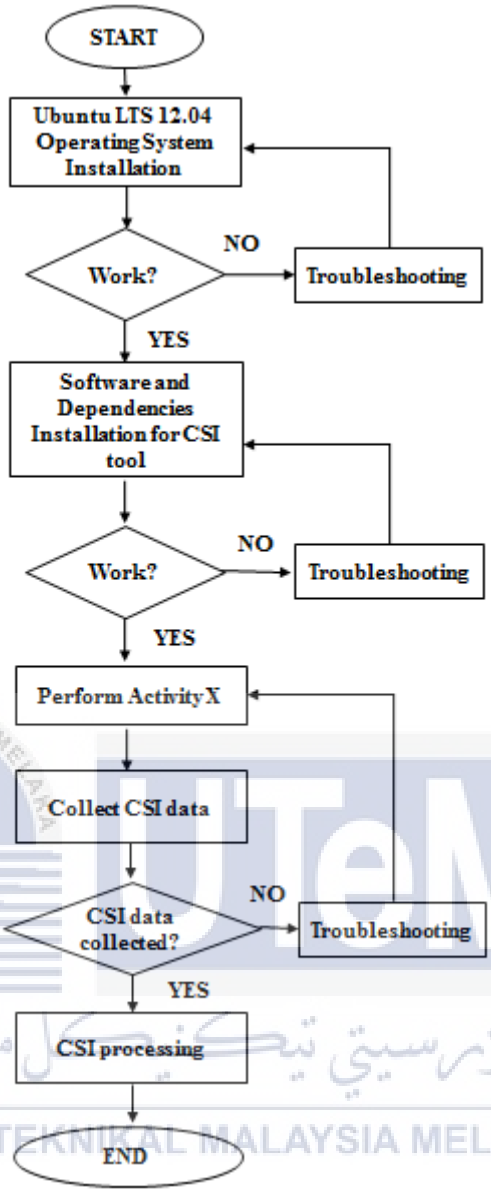


Figure 3.2: Flowchart of Self-collected Dataset

Figure 3.2 shows the flowchart to prepare the self-collected dataset. To install the CSI tool, firstly the Linux developer header is installed in the Ubuntu OS. The kernel headers contain the C header files for the Linux kernel, which offers the various function and structure definitions required when compiling any code that interfaces with the kernel, such as kernel modules or device drivers and some user programs. It is very crucial to note that the kernel headers package installed should match with the currently installed kernel version on the system.

Next, the modified wireless driver is built and installed so that the CSI Tool Linux source tree can be obtained. By modifying the driver, access point functionality for controlling both ends of the link can be accessed. Modified firmware is installed in order to obtain the supplementary material for the CSI tool. After the firmware is installed, log_to_file, is built so that obtained CSI via the driver can be written into a file for further processing.

Now, the CSI data can be collected by unloaded the driver and reload it with CSI logging enable mode by connecting it to an 802.11n access point. After connected, CSI data is enable to logging to a file by running the IP address of the connected access point at a different terminal. The IP address of the connected access point can be checked via ipconfig commands.

After the CSI Tool is successfully installed, the Wi-Fi CSI data are ready to be collected. In this project, we are interested in how human activity can be recognized through CSI. Therefore, a real-time data processing and visualization plugin for the Linux 802.11n CSI Tool by Lu et al. [45] is implemented to observe the effect of the changes in surrounding before the CSI data for different activities are collected. This is to double confirm that the CSI Tool installed was working properly. The changes of CSI for various environment conditions had been recorded in chapter 4.

The CSI data for each human activity is collected in a living room as shown in Figure 3.3. During the data collection, the router model with TP-Link TD-W8961N modem router is acting as a transmitter and Dell Latitude D630 Intel Core 2 Duo 2GB laptop which equipped with Intel 5300 Wi-Fi mini card and Ubuntu 12.04 with the Linux kernel version 3.2 installed acts as a receiver.

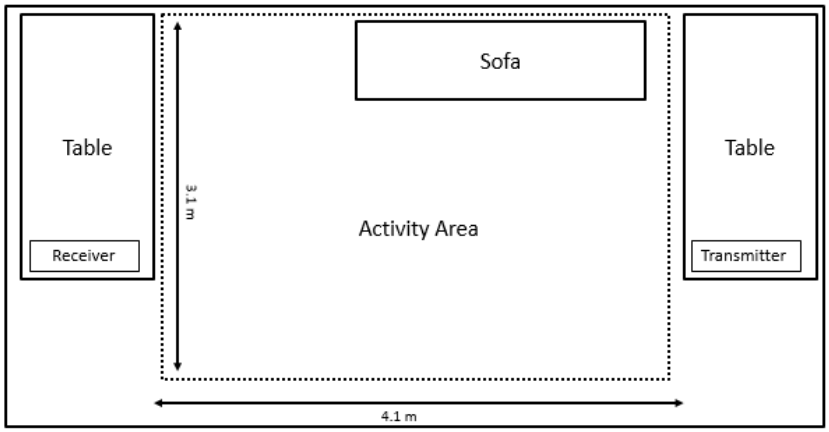


Figure 3.3: Layout of the living room

In the experiment, one antenna in the transmitter, and three antennas in the receiver are used. The CSI data are performed in the 2.4 GHz frequency band with 20 MHz channel bandwidth. By using the ping command, 1000 packets are sent from the transmitter to the receiver in 1 second during the CSI data collection, as suggested in [40]. The packet is ping continually before the activity start until the end of the activity. The volunteer remains stationary at the beginning and end time. Activities data such as walking, running and falling and no activity is collected from one volunteer. Thirty samples are collected for each of the activities. As shown in Figure 3.4, a .dat file is created to collect the CSI information.

(a)

(b)

Figure 3.4: Self collected Data (a) CSI file created (b) command to write the CSI data into the csi.dat file.

3.3.2 Online Dataset

This dataset is collected in an indoor office area with sampling rate of 1 kHz where the Tx and Rx are located 3 m apart in LOS condition [40]. This dataset includes six persons, seven activities, denoted as "Fall, Walk, Run, Sit down, Stand up, Pick Up and Bed" and 78 trials for each one. During the data collection, a person starts moving and doing an activity within a period of 20 s in LOS condition. In the beginning and at the end of the time period, the person remains stationary.

3.4 Deep Learning Implementation

Figure 3.5 shows the flowchart of workflow for designing LSTM model for human activities classification. Firstly, the input and output parameter need to be defined based on type of designed deep learning network. As discussed in Chapter 2.1, different type of input is required by different deep learning network. Therefore, the input data is in sequence as LSTM network is used to classify human activities. Then, the training parameter is randomly select in training the proposed LSTM network to classify the human activities. The training parameter is keep adjusting until the proposed network achieve at higher accuracy.

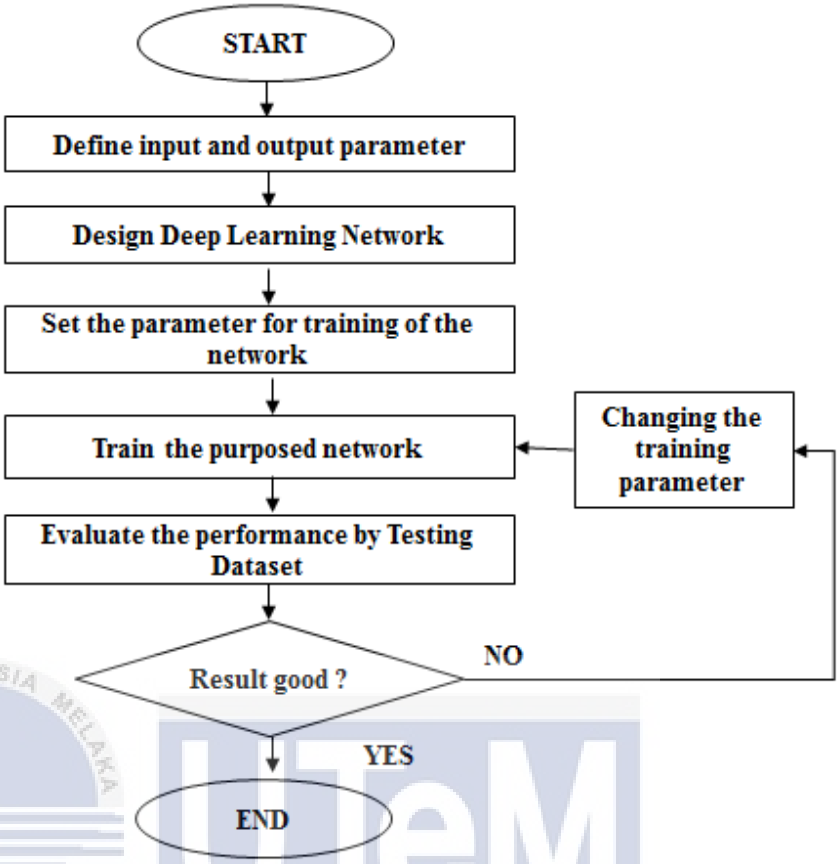


Figure 3.5: Flow Chart for Workflow of Designing LSTM model

3.4.1 Training and Testing Subset

Table 3.2: Training and Testing Subset

Dataset	Training Subset	Testing Subset	Total
Self-collected	60	60	120
Online (Reduced)	105	105	210
Online	546	546	1092

The CSI data will fetch into the proposed networks separately to be used as a training and testing subsets. The training subset is a set of dataset utilized for learning to satisfy the parameter of the classifier. Each input CSI data have their respective ground truth label for supervised learning. Testing subset is a set of dataset utilized to evaluate the performance of a fully trained classifier. No same input CSI data should

be as the part for both training and testing subsets. Both datasets have been split in the ratio of 5:5 as the training set and the testing set, as shown in Table 3.2.

3.4.2 Proposed LSTM network

As shown in Figure 3.6, the proposed LSTM network implemented the use of an RNN architecture with one hidden layer of LSTM cells in classifying the time-series CSI data. The core components of an LSTM network are a sequence input layer and an LSTM layer. To predict the activity labels, the network ends with a fully connected layer, a softmax layer, and a classification output layer.

In this project, we not only developing and testing the standard unidirectional LSTM, but also a more complex architecture such as bidirectional LSTM and cascaded LSTM as shown in Figure 3.7 for human behavior recognition.

The use of an RNN architecture with one hidden layer of LSTM cells is proposed, in order to classify the action time series. The input layer of this RNN comprises an embedded vector that contains the sequence of the CSI amplitude. Then, n LSTM cells are fully connected to these inputs and have recurrent connections with all the LSTM cells. A dense output layer performs the classification task.

Figure 3.7(a) shows the LSTM single cells over time. The single cell layer is presented at time t, where Xt and Yt are the input and output states, respectively. Data from each subcarriers represent the input, while the different activities for each dataset represent the output.

As shown in Figure 3.7(b), the Bi-LSTM, which extract patterns from the past and the future, includes two parallel LSTM tracks defined by forward and backward loops. Compared to the standard Uni-LSTM, Bi-LSTM has better performance in modeling

time dependency. The forward track (green arrow) reads the input data Xt from left to right, whereas the backward track (red arrow) reads the input data from right to left. The output prediction is the weighted sum of the prediction score, resulting from the forward and backward tracks.

Figure 3.7(c) shows the cascaded LSTM, which implements the use of Uni-LSTM and Bi-LSTM. The feature vector is firstly fetched into the Bi-LSTM and then, feed into the Uni-LSTM. The cascade architecture is inspired by [46]. According to [46], the cascaded LSTM able to improve the speed of the gradient computation on each replica.

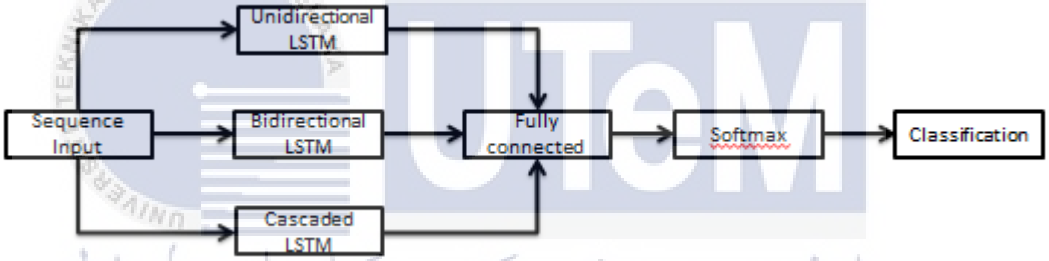
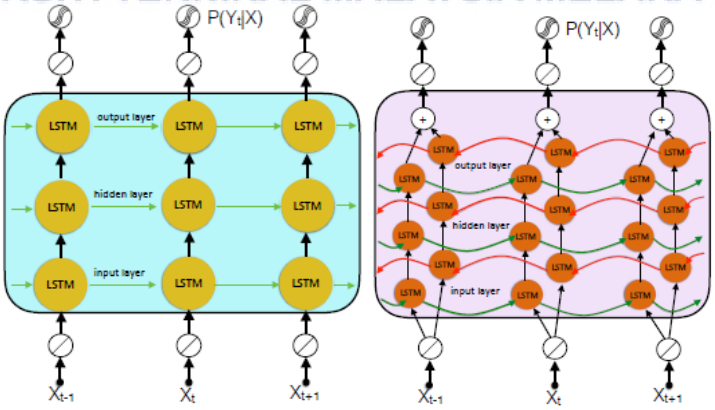
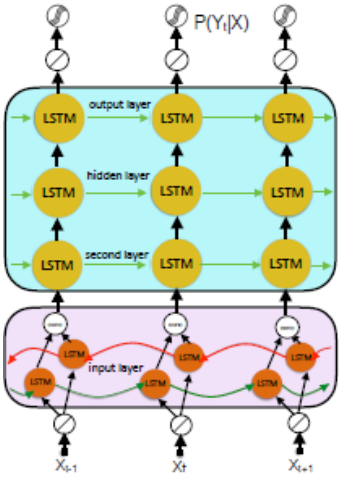


Figure 3.6: The architecture of the proposed LSTM networks



(a) (b)

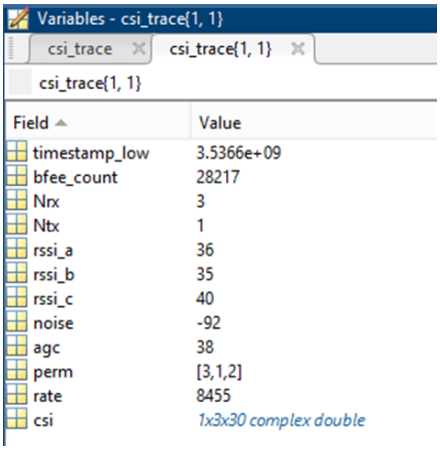


(c)

Figure 3.7: The architecture of (a) Unidirectional LSTM (b) Bidirectional LSTM (c) Cascaded LSTM

3.4.3 Training LSTM

Before starting to do training for proposed LSTM model, it is essential to extract the CSI information from each of the packet. The CSI amplitude for each of the sample is imported into the Matlab for further processing. Then the data will be separated into testing dataset and training dataset.



(a)

```
>> csi_trace = read_bf_file('fall/tyy_fall_01.dat');
>> csi_entry = csi_trace{1}
```

(b)

Figure 3.8: (a) Example of CSI packet (b) function to extract the CSI info from packet

There are few parameters needed to be adjusted before starting to train. Firstly, CUDA that acts as a parallel computing platform and programming model that utilizes GPU for general-purpose computing is enabled during the training of LSTM model. Besides, the training parameters in the training algorithm such as batch size, learning rate, number of the epoch are varied during the training process and the performance of the trained model is evaluated by testing dataset. The performance for each of the model trained with different parameters will be tabulated and analyzed in chapter IV.

The input file required to train an LSTM network is shown in Figure 3.9 (a). The 'XTrain.mat' contains the CSI amplitude for human activities used for training while 'YTrain.mat' contains the respective label for each of the activities in 'XTrain.mat'. Figure 3.9 (b) shows the options which defined the parameter used during training. Figure 3.9 (c) shows the command to generate the trained network. The layers defined using the proposed LSTM network, which consists of different layers, as shown in Figure 3.6.


```
load('XTrain.mat')
load('YTrain.mat')
load('layers.mat')
load('options.mat')
```

(a)

```
options = trainingOptions('adam', ...
    'MaxEpochs',70, ...
    'MiniBatchSize', 64, ...
    'InitialLearnRate', 0.001, ...
    'SequenceLength', 500, ...
    'GradientThreshold', 1, ...
    'ExecutionEnvironment',"auto",...
    'plots','training-progress', ...
    'Verbose',true);
```

(b)

```
net = trainNetwork(XTrain,YTrain,layers,options) % training network
```

(c)

Figure 3.9: LSTM Training (a) File required (b) training parameter (c) command to train the network

3.4.4 Evaluate the proposed LSTM models

After the training for each of the model is done, the trained module is used to predict the activity contained in the testing dataset. The metrics that implemented to evaluate the proposed module is named as confusion metric. Confusion metric is a specific table layout that allows visualization of the performance of an algorithm. The accuracy for each of the module is calculated and tabulated in chapter 4 as well.

Figure 3.9 shows the command to calculate the accuracy of the trained module. 'testPred' is the activity predicted by the trained module while 'YTest' is the actual label of the activity.

```
testPred = classify(net,XTest,'SequenceLength',1000);
LSTMAccuracy = sum(testPred == YTest)/numel(YTest)*100
```

Figure 3.10: Command to calculate the Test accuracy

3.5 Method Selection

According to Figure 3.10, the amplitude of Wi-Fi CSI from 90 channels is used as an input to train the LSTM model. Before the data is fetched into the LSTM model,

the CSI data will be under pre-processed stage. During the preprocessed stage, the CSI information will be extracted from each of the packet. Then the data is filtered as only amplitude of CSI is used as an input data in this project. Finally, the human activity can be classified by using the trained LSTM model.

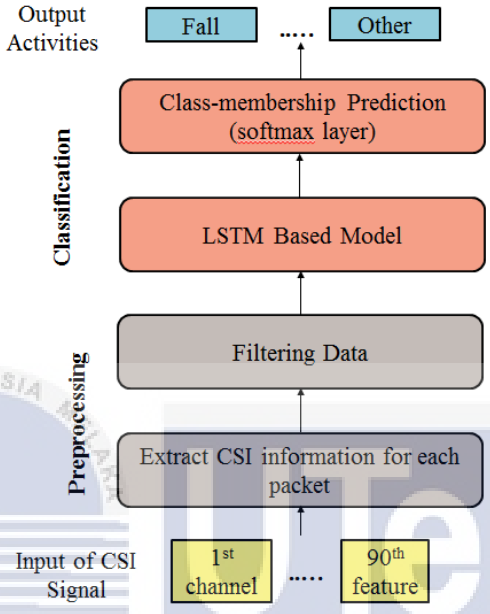


Figure 3.11: Work Flow of Proposed HAR approach

3.6 Project Management

The financial budget of this project is within the financial budget since all the materials, tools and developer kits are sponsored by supervisor and FKEKK UTeM such as GPU and computer. Hence, there will be zero cost for this project.

There are three objectives will be achieved in this project. Firstly is to investigate data. Secondly, to design a deep learning neural network model to classify the human activities by using Wi-Fi CSI data. The third objective is to analyze the performance of the proposed model. The steps and results of this project will be discussed in Chapter III and Chapter IV respectively.

3.7 Environment and Sustainability

The priority benefit of the proposed LSTM network is that it has the capability of automatic learning of spatial-temporal information from the input CSI data without the need of handcrafted features or kernel fusion approaches [12]. Besides, human activities can be recognized by using Wi-Fi CSI data which does not raise up the privacy issue and free of lighting issues. Moreover, the Wi-Fi CSI data can be collected without required users to carry any hardware thus indirectly reducing the product waste.

3.8 Project Summary

In this project, there are two Wi-Fi CSI datasets used to evaluate the proposed LSTM models. The trained LSTM module is generated by using three different types of LSTM cell to classifier the human activities by using the training datasets. Finally, the performance of the proposed LSTM module is evaluated by classifying the human activities in the testing set by using the trained module. The flowchart of this project and significant steps that needed to be achieved have discussed respectively in this chapter.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Introduction

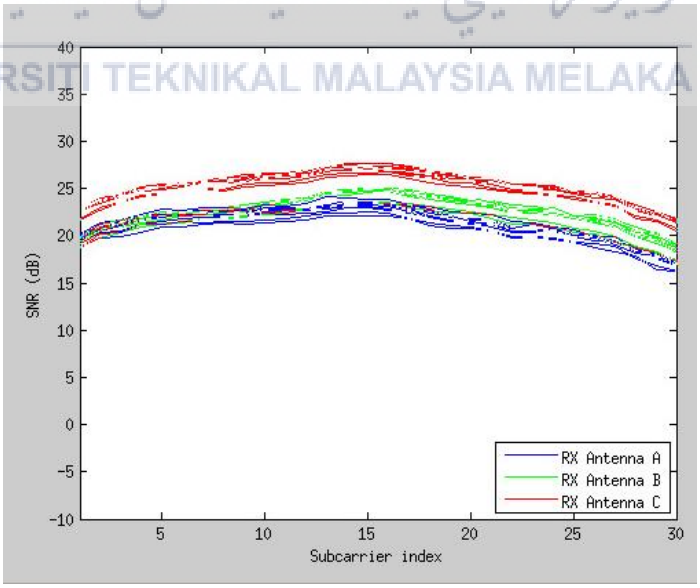
This chapter started by presenting the visualization of CSI data in different types of human activity followed by the proposed LSTM modules under subsection 4.2 and subsection 4.3, respectively. Subsection 4.2 discusses the effect of CSI on different human activities. The proposed unidirectional LSTM, bi-directional LSTM, and cascaded LSTM for human activity recognition and the results achieved for human activity recognition by using different datasets had been shown under subsection 4.3.

4.2 Visualization of Channel State Information on Human Activities

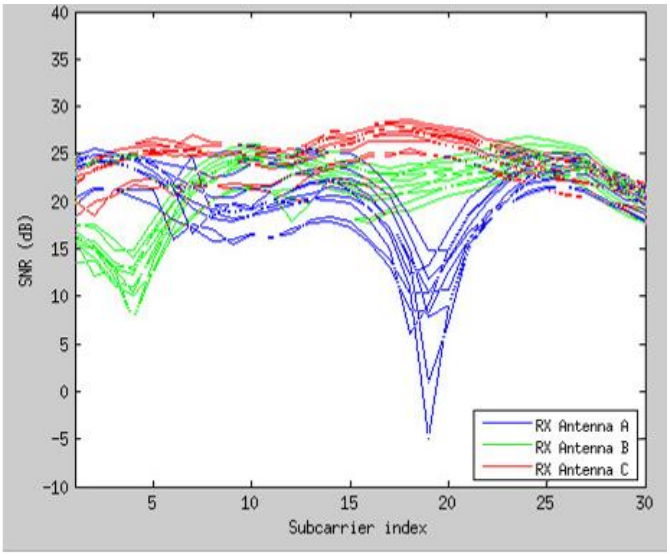
4.2.1 Real time processing CSI signal visualization

In this project, we are interested in how human activity can be recognized through CSI. Therefore, a real-time data processing and visualization plugin for the Linux 802.11n CSI Tool by [45] is implemented to observe the effect of the changes in surrounding before the CSI data for different activities are collected.

A simple experiment is conducted by placing the receiver (TP-Link Router) and transmitter (Dell Laptop) as the access point (AP) with 2m apart. To verify that CSI variance correlates with multi-path effect and surrounding movement, the CSI data is recorded in two conditions, with interferences and without interference. The CSI data is recorded by the AP which including Signal-to-Noise Ratio (SNR) values across each antenna and subcarrier. Figure 4.1 (a) shows the data when there is no activity or movement within the specific area. The SNR values for CSI signal when no activity in Figure 4.1 (a) are very high for each antenna and subcarrier and with little variance observed. Further, the signal curves for each antenna exists a strong similarity in terms of curve shape due to the strong and stable connection between the wireless networks. Figure 4.1 (b) shows the CSI signal when there is interference (body movement) is the surrounding. From Figure 4.1 (b), a high degree of variance between each frame is clearly seen. Therefore, it had been proved that CSI variance relates closely to interference. Hence, it is possible to recognize human activity by using CSI data.



(a)



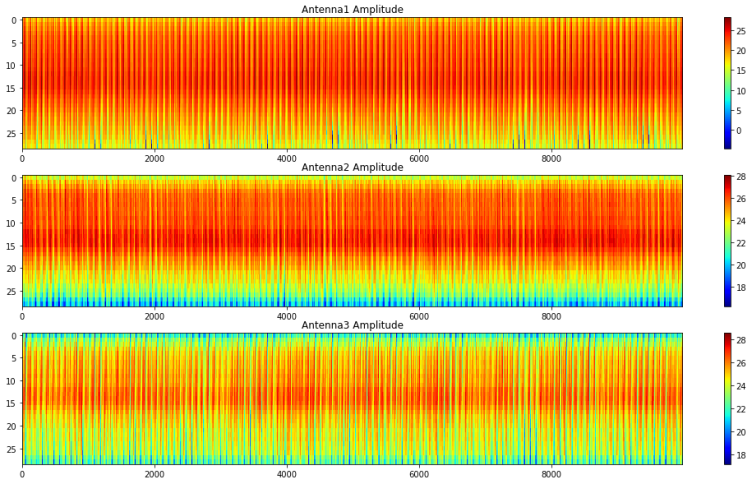
(b)

Figure 4.1: CSI signal visualization (a) without interference (b) with interference

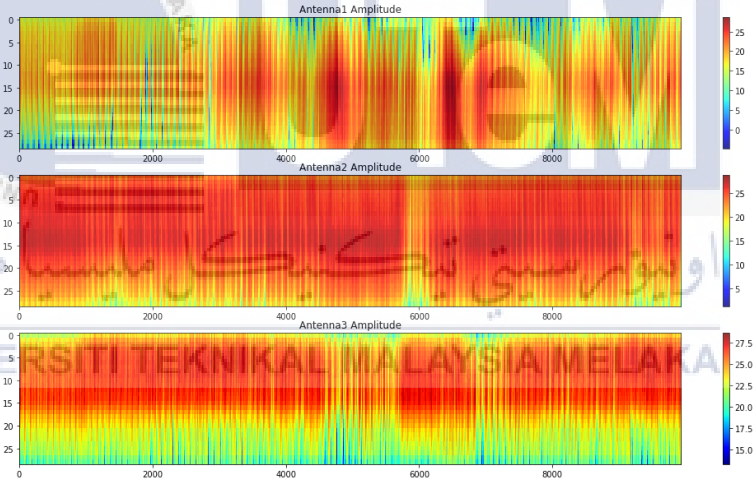
4.2.2 CSI Amplitude versus the number of packets

The effect of human activity on Wi-Fi CSI amplitude is visualized by implementing the method suggested in CARM [39]. Figure 4.2 (a) shows the visualization of CSI amplitude without any activity in the interested area at each of the received antenna which its amplitude is relatively constant during the transmission of 10000 packets. Figure 4.2 (b) and (c) visualize the CSI amplitude for the activity that involved radical movements that are walking and running, respectively. As observed, the amplitude of CSI undergoes a huge changes compare to Figure 4.2(a) (data conducted when there is no activity). This shows that the CSI amplitude can be affected by human activity. Figure 4.2 (d) visualized the CSI data collected for falling action. From observation, when, there is are sudden rise in CSI amplitude during the transmission from 4000th packet to 6000th packets, which indicates the sudden fall of the volunteer, especially for the CSI signal received at the second antenna. This means that, the CSI value received at each antenna is different, and may have different sensitivity. Therefore, by implementing the CSI amplitude value received at all of the three antennas, an activity

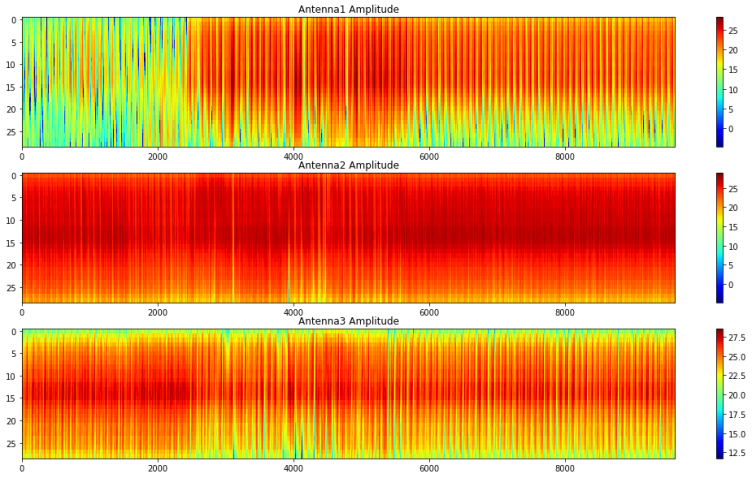
can be recognized as each of the activities will have a different effect on the CSI amplitude.



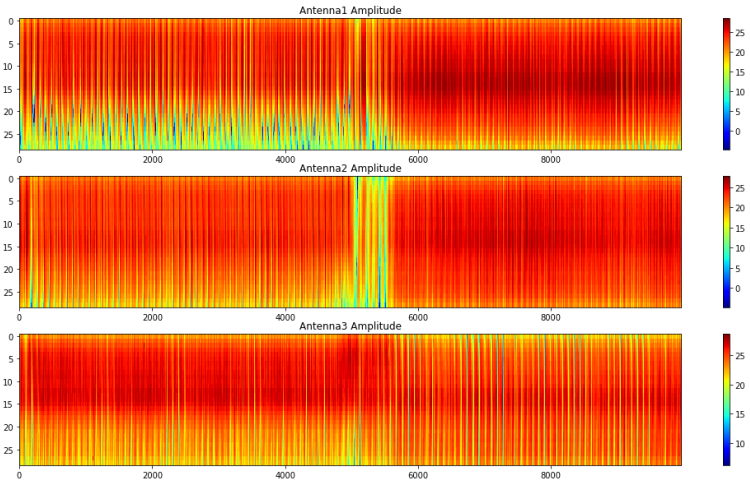
(a)



(b)



(c)



(d)

Figure 4.2: Visualization of CSI amplitude for (a) no activities (b) walking (c) running (d) falling

4.2.3 Effect of Human Activities on CSI Amplitude

In this project, CSI value is the most concerning value as it contains most of the information as it is needed as input data to classify activity. Inside the packet, the CSI information is in the form of complex values in a matrix of the order Tx × Rx × 30, where Tx is the number of antenna used to transmit and Rx is the number of antenna used to receive the packet.

In this project, one antenna is used to transmit and three antennas is used to receive. Therefore, a total of 90 channels or features can be used an input vector to developed the deep learning network. To investigate the effect of human activity on CSI amplitude, the average amplitude of CSI across 30 subcarrier for each activity by using self-collected dataset is plotted as shown in Figure 4.3. From observations, the CSI value received at each antenna is different and have their own characteristic for each of the activity. Therefore, the deep learning algorithm is implemented for human

activities recognition as it has the ability to execute features that correlate and combine them to enable faster learning without being explicitly told to do so.

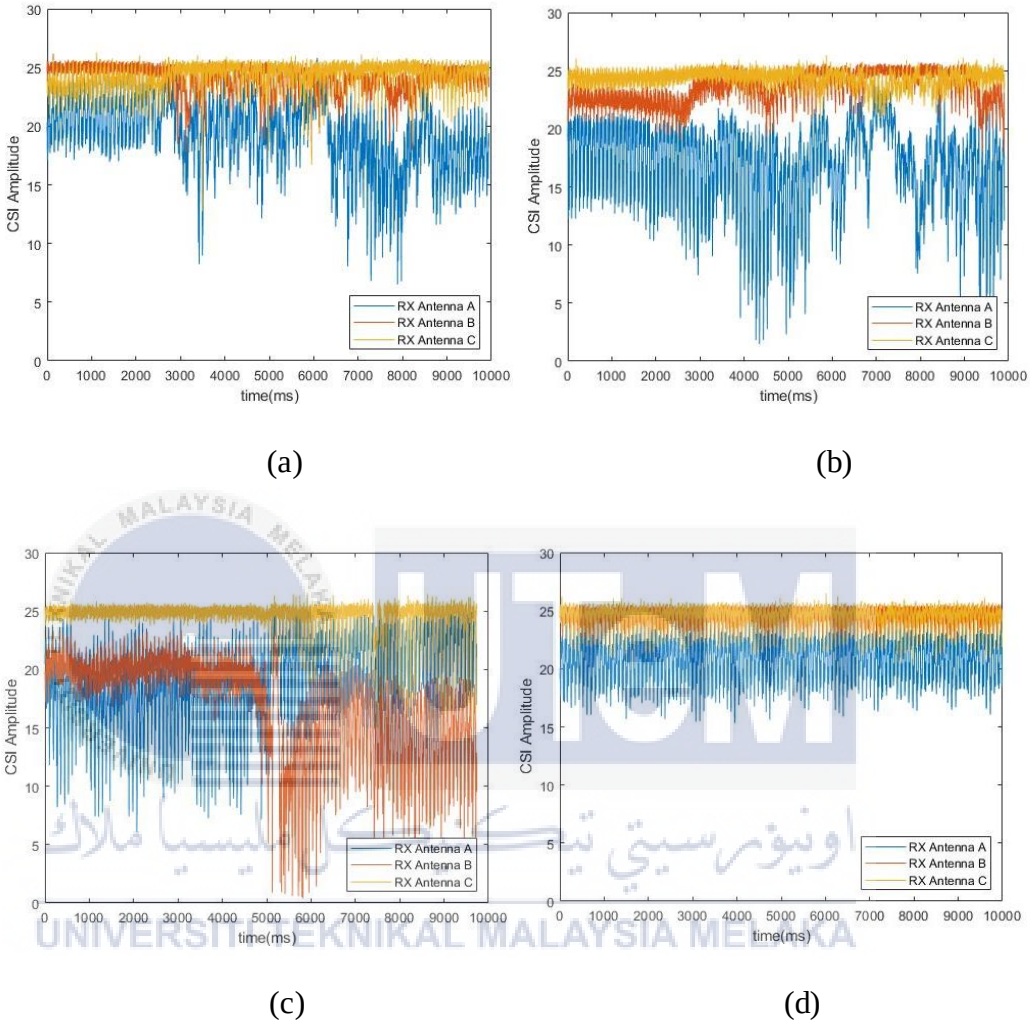


Figure 4.3: The Average CSI Amplitude for Human Activity (a) No activities (b) Fall (c) Walk (d) Run

4.3 Proposed LSTM module

In this project, the standard Unidirectional LSTM had been developed for human activity recognition. Besides, more complex architectures such as bidirectional LSTM and cascaded LSTM has also been investigated. The architecture of different types of LSTM modules had been discussed in detailed in Chapter 3. All of the proposed LSTM modules consist of 5 layers. The first layer is sequence input layer, followed by LSTM layers, fully connected layers, softmax layers and classification layers. The

proposed LSTM network implemented the use of an RNN architecture with one hidden layer of LSTM cells (second layer) in classifying the time-series CSI data.

During the classification process, firstly the CSI input data is fetched into the sequence input layer. The example of the input data is shown in Figure 4.4. The value for each row represents the CSI amplitude collected at different time steps for each subcarrier. In this project, there are 30 subcarriers collected at each received antenna. Since there are three received antenna is used, thus, 90 subcarriers or 90 features for each activity are employed to develop the LSTM module.

Then, the sequence input will fetch into the LSTM layer which consists of LSTM Block. The LSTM contains four interacting layers which are forget gate layers, input gate layers, cell state layers and output gate layers in each module as shown in Figure 4.5. The sequential data will be processed in this layer and the information is passing on as it propagates forward.

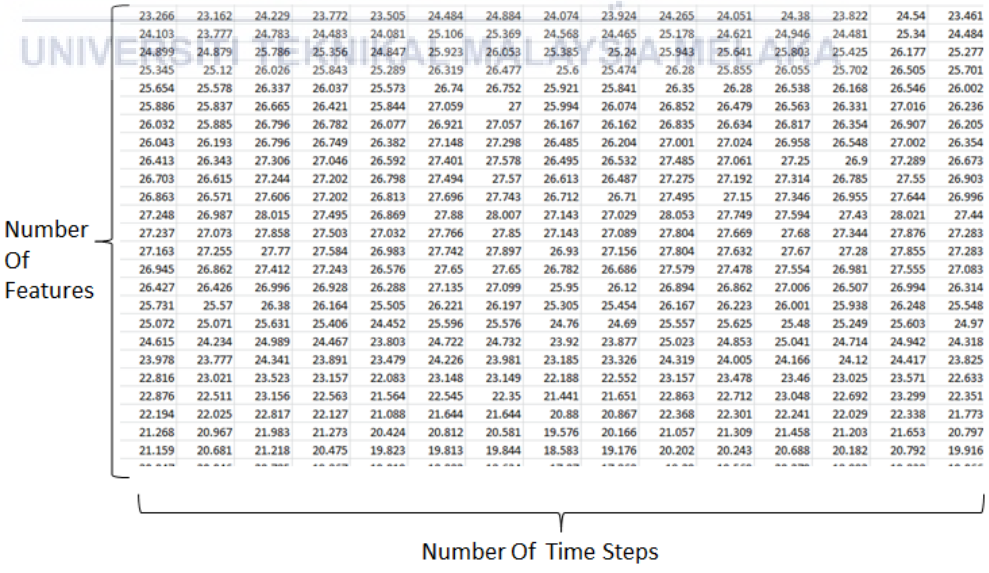


Figure 4.4: Example of the Data Fetched into Input Sequence Layer

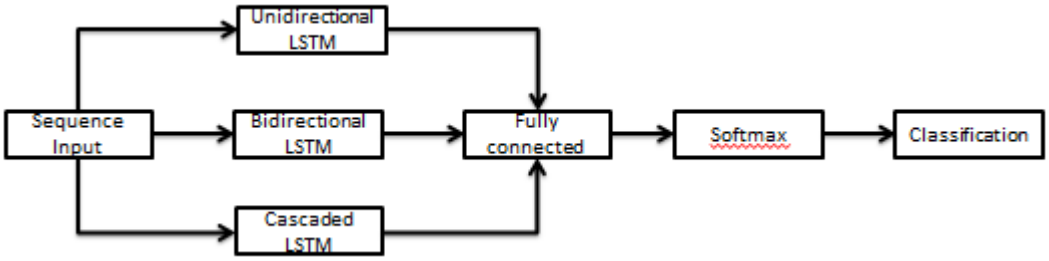


Figure 4.5: The architectures of proposed LSTM module

According to Figure 4.6, there are three different gates used to regulate the data in a LSTM cell which are forget gate, input gate and output gate. The data from the previous hidden state and current input will firstly be passed through the forget gate in the LSTM cell. The gate contains of sigmoid function which output values range between 0 and 1. This gates decides which information is allowed on the cell state, and learns what information is relevant to keep or forget during training. The closer to 0 means to forget, while closer to 1 means the data is decided to keep.

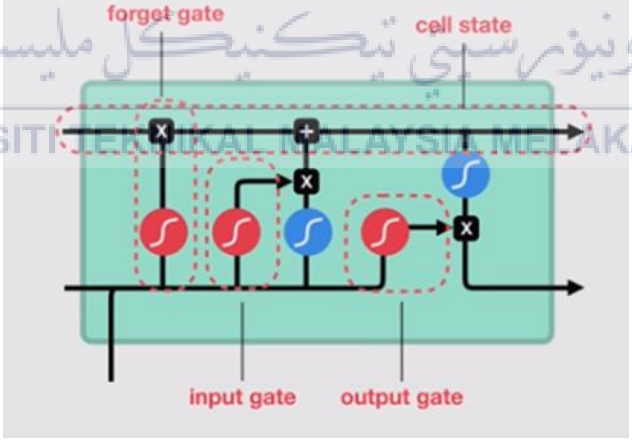


Figure 4.6: The architecture of the LSTM Block

The cell state is the memory of the network, as it contained the information from earlier time steps to the last time step. To update the cell state, we have the input gate, first we passed the previous hidden state and the current input to a sigmoid function that decide which values will be updates by transforming values to be between 0 (least

important) and 1(most important). The hidden state and current input are also passed into the Tanh function to squeeze values between -1 and 1, which help to regulate the network. Then the Tanh output is multiplied with the sigmoid output. The sigmoid output will decide which information is important to keep from the Tanh output. The cell state is then multiplied by the forget vector. The values in a cell state are dropped if it multiplies by values near 0. Then a pointwise addition is performed by using the output from the input gate and the previous cell state. Now, the cells state had updated to new values.

The next hidden state is decided by the output gate. The information of previous inputs is contained in a hidden state thus, the hidden state is also used for prediction. The previous hidden state and the current input are fetched into the sigmoid function. Then the newly modifies cell state is passed into the Tanh function. The Tanh function output is multiplied with the sigmoid output to decide the information carried by the hidden state. The output is the hidden state. The new cell state and a new hidden state are then carried over to the next time step.

To predict the activity labels, the network ends with a fully connected layer, a softmax layer, and a classification output layer. A fully connected layer is used to multiply the input by a weight matrix and then adds a bias vector. Bias vector is added so that the activation function is allowed shifted to the left or right, which may be critical to fit the prediction with the data better.

The input is then passed through a softmax layer before entering the classification layer. In softmax layer, the input data is applied with softmax function which outputs a vector that represents the probability distributions of the label activity. The classification layer computes the cross entropy loss for multi-class classification

problems with mutually exclusive classes. The layer infers the number of classes from the output size of the previous layer.

4.3.1 Training Long Short Term Memory (LSTM) module

In training process, the parameters utilized in the training algorithm for both neural networks are very paramount important in order to regulate the precision of recognition output. Both neural networks are trained separately with their respective datasets. Since all trained modules for human activity recognition using CSI amplitude depend on the input data collected from the surrounding environment, therefore training is required to generate an LSTM module to evaluate the performance in a new environment.

There are several parameters generated during training, as shown in Figure 4.7. The training process will keep updating the weights of the LSTM model for each iteration. Parameters of 'Epoch' represents the number of times for an entire dataset is passed forward and backward through the neural network. Since the self-collected dataset does not contain large samples, therefore updating the weights with single epoch is not enough, as shown in Figure 4.12. The accuracy is only 25% by using one epoch. Therefore, multiple numbers of epochs is needed during training to achieve better performance. Figure 4.12 shows that when the number of epoch increases, the number of times weight is changed in the neural network increases. However, a higher number of epoch does not necessarily can achieve better accuracy, as it depends on the diversity of the dataset.

'Mini-batch Accuracy' corresponds to the accuracy of the particular mini-batch at the given iteration. During training by ADAM optimizer, the algorithm groups the full dataset into disjoint mini-batches. As the entire dataset is considered too large to pass

into the neural net at once, therefore the dataset is divided into number of batches. The number of batches refers to the number of parts divided from the training dataset. Before the module weights updated, the batch size, which refers to the total number of training examples presented in a single batch, is used to estimate of the error gradient. There are advantages and disadvantages as well for different batch sizes. By utilizing the parallelism of GPUs, larger batch size can be trained speedily. However, training with a larger batch size often leads to poor generalization. On the other hand, better performance can be achieved by using a smaller batch size as the module can start learning before have to go through all the data.

The 'Mini-batch Loss' is the quantitative measure of deviation or difference between the predicted output and the actual output in anticipation of the mini-batch. It indicates the measure of mistakes made by the network in predicting the output.

Parameter of 'Base Learning Rate' indicates the amount that the weights are updated during training. It is a configurable parameter and often in the range between 0.0 and 1.0. The backpropagation of error estimates the amount of error for which the weights of a node in the network are responsible during the training. Instead of updating the weight with the full amount, it is scaled by the learning rate. Generally, a large learning rate allows the model to learn faster. However it may cause the weight to explore high results in the numerical overflow. A lower learning rate may allow the model to learn a more optimal or even globally optimal set of weights but may take significantly longer to train. Therefore, the range of values to consider for the learning rate is less than 1.0 and greater than 10^-6.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:40	25.00%	1.7081	0.0010
3	50	00:02:34	63.33%	1.0615	0.0010
5	100	00:03:01	83.33%	0.7294	0.0010
8	150	00:03:25	83.33%	0.7054	0.0010
10	200	00:03:45	85.00%	0.5865	0.0010
13	250	00:04:05	81.67%	0.5736	0.0010
15	300	00:04:25	85.00%	0.4659	0.0010
18	350	00:04:44	85.00%	0.4612	0.0010
20	400	00:05:03	90.00%	0.3686	0.0010
23	450	00:05:22	83.33%	0.4873	0.0010
25	500	00:05:41	86.67%	0.3431	0.0010
28	550	00:06:01	85.00%	0.4284	0.0010
30	600	00:06:21	93.33%	0.2591	0.0010
33	650	00:06:40	86.67%	0.3627	0.0010
35	700	00:06:59	93.33%	0.2405	0.0010
38	750	00:07:18	88.33%	0.3127	0.0010
40	800	00:07:37	96.67%	0.1891	0.0010
43	850	00:07:58	86.67%	0.3528	0.0010
45	900	00:08:16	91.67%	0.2579	0.0010
48	950	00:08:35	83.33%	0.4151	0.0010
50	1000	00:08:54	93.33%	0.2373	0.0010
53	1050	00:09:14	90.00%	0.2760	0.0010
55	1100	00:09:33	95.00%	0.1771	0.0010
58	1150	00:09:52	85.00%	0.3588	0.0010
60	1200	00:10:11	96.67%	0.2140	0.0010
63	1250	00:10:30	88.33%	0.3171	0.0010
65	1300	00:10:49	93.33%	0.2179	0.0010
68	1350	00:11:09	91.67%	0.2923	0.0010
70	1400	00:11:28	91.67%	0.1887	0.0010

Figure 4.7: Output Parameters Produces During Training of LSTM.

4.3.2 Performance Evaluation of LSTM models

The proposed module had been evaluated by predicting the human activity of testing datasets as shown in Table 3.1. There are three different datasets used in this project. The first dataset is a total of 30 samples for each of the activity collected at a standard living room as discussed in Chapter 3 at sampling rate of 1k Hz. The second dataset chosen, which consists of 78 samples for each of the activity, is collected at 1k Hz as well so to evaluate the performance of the LSTM module. In contrast to the second dataset, the third dataset is a reduced version of second dataset which 30 samples are randomly chosen from the second dataset. The third dataset is used to observed performance of the LSTM module by using the same number of samples as the first dataset.

The training and testing accuracy of Unidirectional LSTM for different numbers of units (i.e., n = 32, n = 64, and n = 128) for self-collected dataset is shown in Figure 4.8. The lowest generalization error in the testing set was achieved with n = 64.

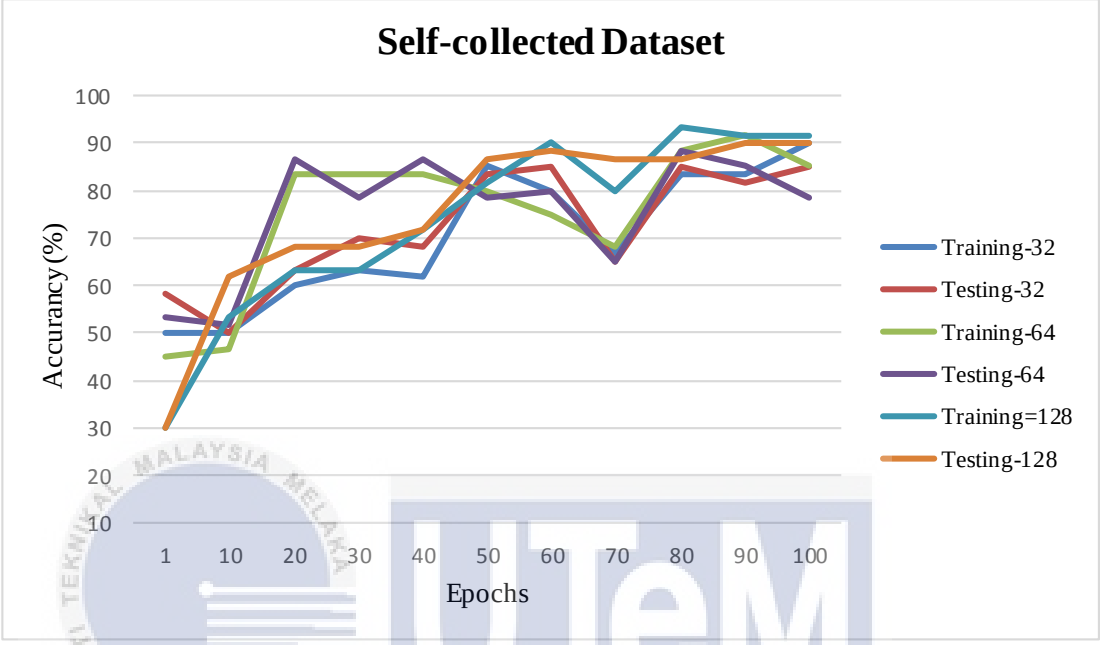


Figure 4.8: Training and validation accuracy with number of units, n.

Figure 4.9 shows the testing accuracy for all the LSTM-based models using different datasets. For self-collected dataset, the accuracy achieved by each of the modules is very similar. However, the highest accuracy is achieved by using Bi-LSTM and Casc-LSTM. This is because the inputs data will be run in two directions, one from the past to the future and another from future to the past, which the information from both past and future is preserved. The confusion matrices for the self-collected dataset, obtained by each different LSTM model is shown in Figure 4.10.

As observed in Figure 4.9, the overall performance of the result by using self-collected dataset is better than the online dataset. This is because the complexity level of the online dataset is higher than the self-collected dataset. For instances, the number of activities and the subjects involve in the online dataset is higher than the self-

collected dataset. Therefore, the variation of the online dataset is higher as the input data (CSI amplitude), are not the same for different people. Moreover, the size of each of the samples in online dataset is larger than the self-collected dataset as showed in Table 3.1.

The differences between the online-reduced dataset and online-default dataset is the number of samples. The total number of samples is 210 and 1092, in online reduced-dataset and online default-dataset, respectively. From Figure 4.7, the accuracy achieved by Bi-LSTM module is the highest for the online-reduce dataset. This is because that the Bi-LSTM module is able to handle data with dependence on long range by utilized the two data information, both from the front and the back [47].

However, the highest accuracy achieved using online default-dataset is obtained by Cas-LSTM module. Studies [48][49] had shown that the Cas-LSTM algorithm are able to overcome the limitations of traditional prediction models related to input data, such as the need for substantial prior knowledge, the satisfaction of certain distribution characteristics, mutual independence between factors, etc. The main feature of this algorithm is the adoption of the cascade LSTM structure, which use the perception range characteristics of different LSTM structures as data features and extracts and fuses the features of different input factors to acquire deeper and more complete features of each factor. Therefore, it outperformed the proposed Uni-LSTM and Bi-LSTM using default dataset, which contains the largest samples among the three datasets.

The confusion matrices for the second dataset and third dataset obtained by each different LSTM model Figure 4.12 and Figure 4.13, respectively.

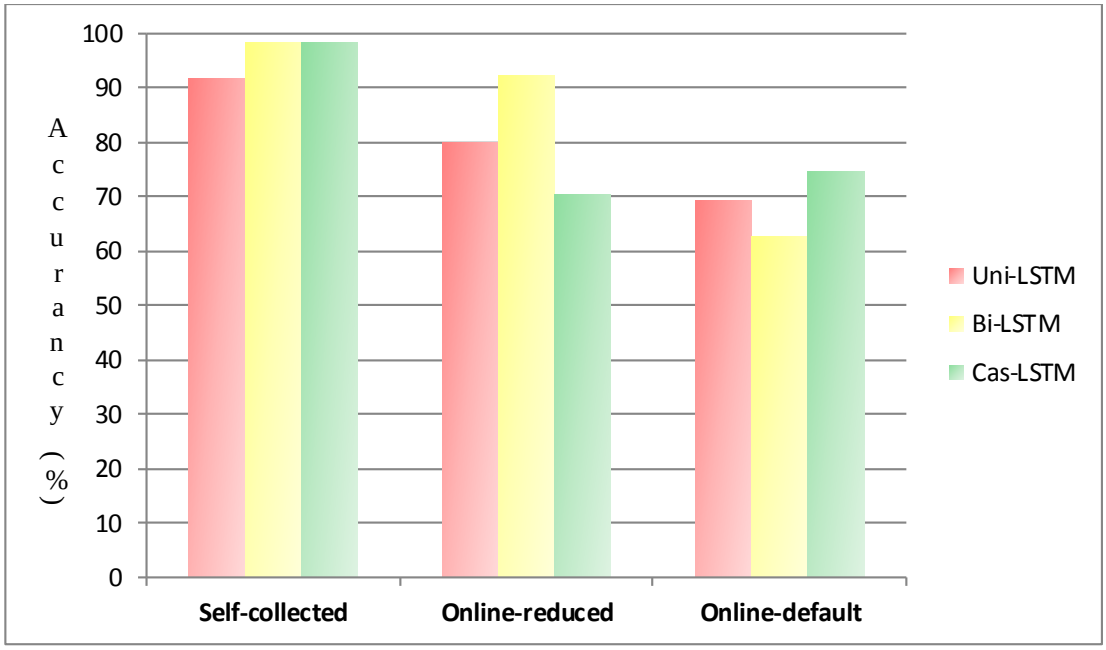
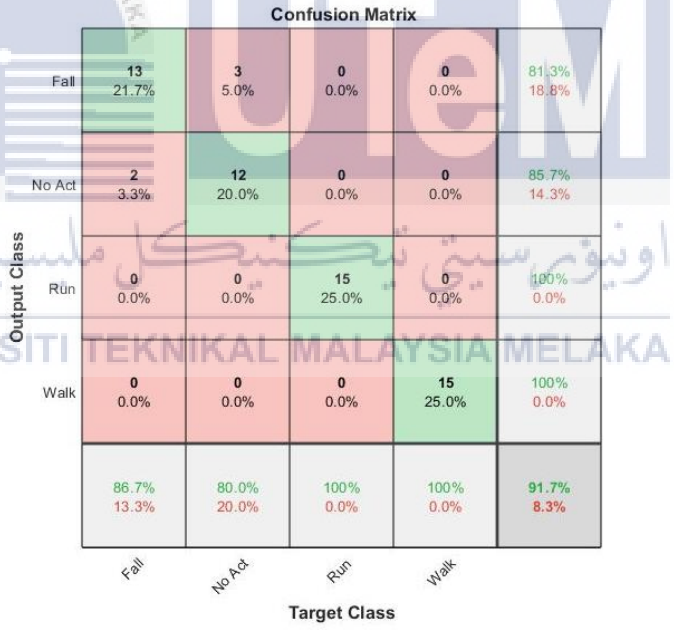
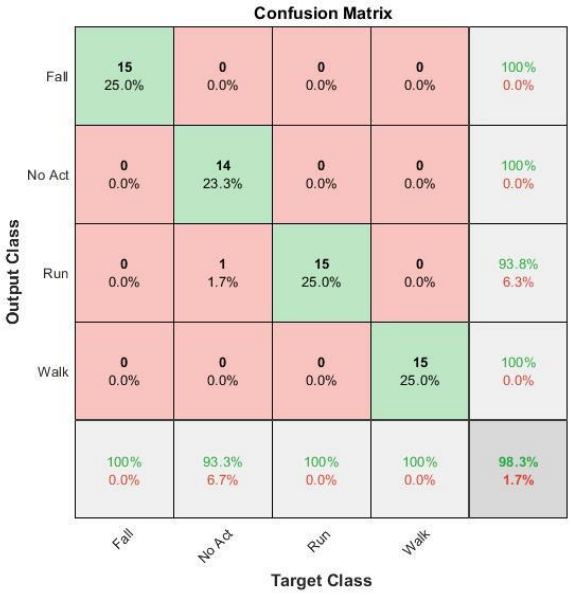


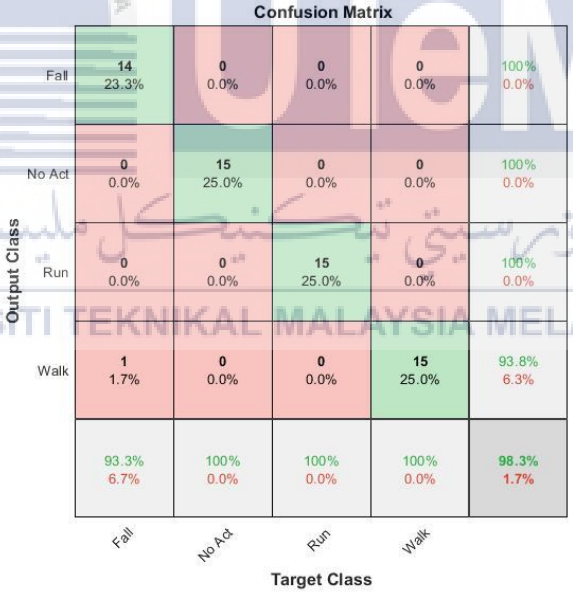
Figure 4.9: The Accuracy for all the LSTM-based models.



(a)

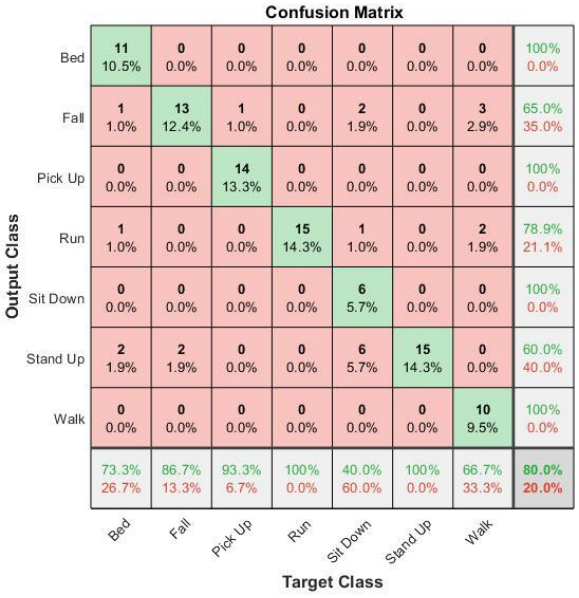


(b)

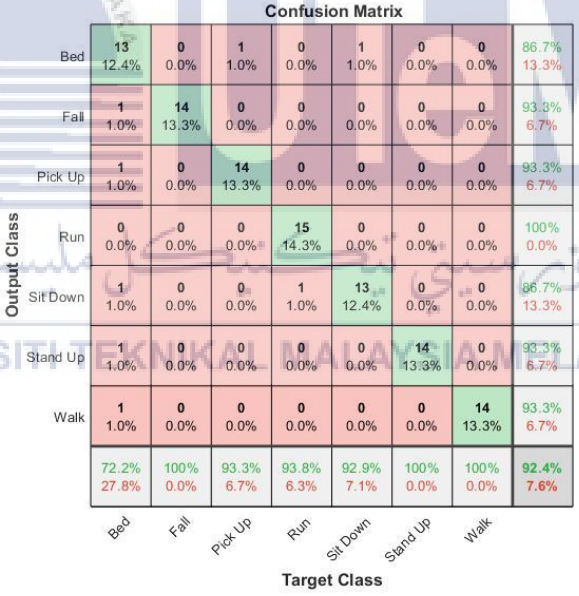


(c)

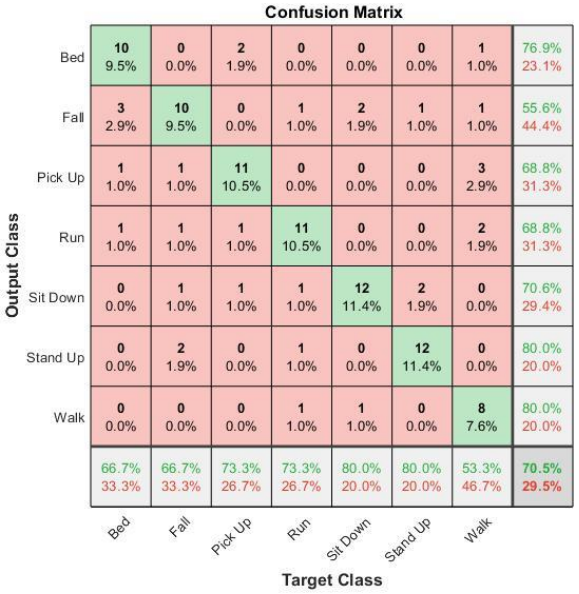
Figure 4.10: Confusion metric of the self-collected dataset for (a) Uni-LSTM module (b) Bi-LSTM module (c) Cas-LSTM module



(a)

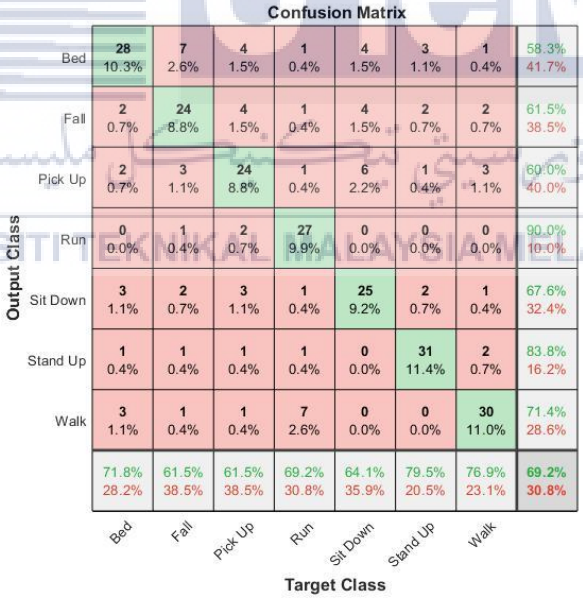


(b)

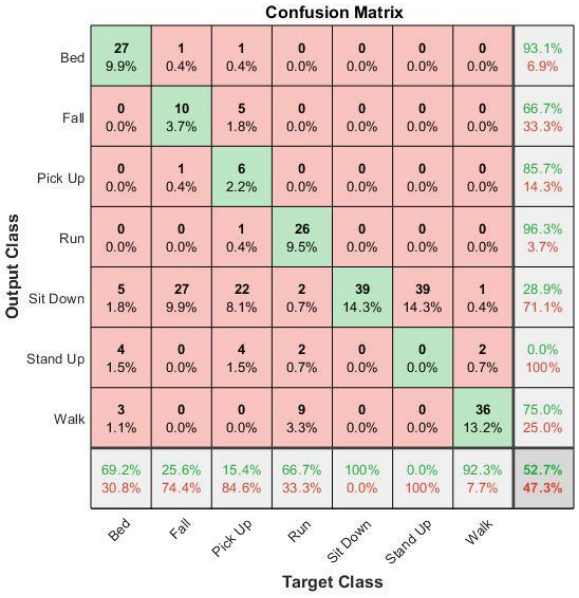


(c)

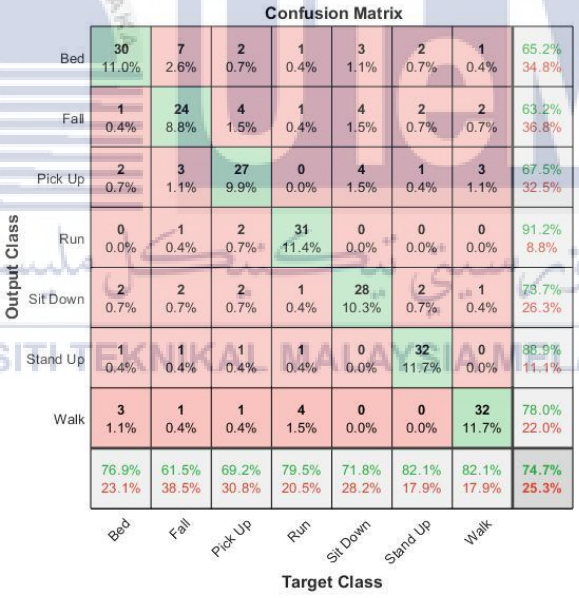
Figure 4.11: Confusion metric of the online reduced dataset for (a) Uni-LSTM module (b) Bi-LSTM module (c) Cas-LSTM module



(a)



(b)



(c)

Figure 4.12: Confusion metric of the online default dataset for (a) Uni-LSTM module (b) Bi-LSTM module (c) Cas-LSTM module

4.3.3 Benchmarking Table on Wi-Fi-based Human Activities Recognition

Table 4.1 shows the accuracy comparison of self-collected dataset with previous works. By using the self-collected dataset, a highest accuracy had been achieved by using Bi-LSTM approach. This is mainly due to several reasons. The main reason that

leads to the high accuracy is due to the uniformity of the self-collected dataset. For instances, the datasets employed in previous works are collected from different persons and tested in different environment, while the self-collected dataset is collected from the same single person and is tested in the same environment. Therefore, it can the assumed that the complexity level of the self-collected dataset is much lower than others datasets used in previous works. Besides, instead of course-grain (RSSI), the self-collected dataset contains the fine-grain properties (CSI) of the Wi-Fi. Moreover, compared to the previous work's method which employed machine learning techniques and signal processing techniques, Bi-LSTM is proposed to classifier human activities. By implemented Bi-LSTM, the features can be extracted automatically without the needs to pre-processing the data. Moreover, Bi-LSTM run the inputs data in two ways, one from past to future and one from future to past, which preserve the information from both past and future.

Table 4.1 : Accuracy Comparison with Previous Works

Dataset	Data Properties	Methods	Accuracy
[20]	RSSI	kNN	89.00%
[21]	RSSI	kNN + fusion	75.00% to 92.58%
[22]	RSSI	Primitives	87.50%
[30]	CSI	EMD + MD-DTW	96.00%
[25]	CSI	CNN	75.00%
[31]	CSI	HMM	96.00%
[32]	CSI	DTW + SDV	98.40%
Self-collected	CSI	Bi-LSTM	98.33%

4.4 Chapter Summary

In the first part of this chapter, a real-time data processing and visualization plugin for the Linux 802.11n CSI Tool by [45] is implemented to observe the effect of the changes in surrounding before the CSI data for different activities are collected. In the second part of this chapter, the effect of human activities on the Wi-Fi CSI signals is investigated. The changes of the amplitude of the CSI data is plotted for each of the activities. The third part of the chapter presents the deep neural network proposed, starting from the standard LSTM formulation, a more complex LSTM based models such as Bi-LSTM and Cas-LSTM models is explored in order to improve the generalization performance of HAR prediction. Then, the performance of the proposed networks have been analyzed and evaluated.

CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Introduction

This chapter will summarize the results and discuss the conclusion of this project. The recommendation for the future development and work will also be discussed in order to enhance the quality of the project.

5.2 Conclusion

Activity monitoring has gradually become important in home environments, especially to elder care, child safety, and patient monitoring. Traditional approaches for human activities recognition which employing wearable sensor or camera obtain good performances and are widely used in application. Nonetheless, these approaches had raised the privacy issue and are limited by the requirement of light. As a result, a deep learning based indoor human activities recognition by using Wi-Fi signals had been implemented. Since Traditional power features like Received Signal Strength Indicator (RSSI) fail to provide sufficient distinction and robustness in complex indoor environments because RSSI is the superimposition of multipath signals with fast-

changing phases. Recent advances in the wireless technologies give alternative solutions for the above problems in a new way. In 802.11 a/g/n standards, channel information can be partially obtained from Orthogonal Frequency Division Multiplexing (OFDM) receivers in the format of Channel State Information (CSI).

Therefore, this project focus more on software development in designing a deep neural network for human activities recognition by using CSI data. Network topology implemented in this project are Uni-LSTM, Bi-LSTM and Cas-LSTM. All the behavior of LSTM are analyzed with some quantitative results which are presented in Chapter 4. By referring to the result predicted, it can be observed that the human activities recognition is not 100% accurate as it will still output wrong predicted result. For self-collected dataset and online reduced-dataset, which has smaller number of samples, Bi-LSTM-based modeling offers better predictions than regular Uni-LSTM-based models. Because of the model using bidirectional will run the inputs data in two ways, one from past to future and one from future to past, which preserve the information from both past and future. However, Cas-LSTM offers a better prediction for the online default-dataset, which contains the largest samples among the three datasets. This is due to the cascade LSTM structure, which use the perception range characteristics of different LSTM structures as data features and extracts and fuses the features of different input factors to acquire deeper and more complete features of each factor.

5.3 Recommendation

In this project, the rate of transmission is set at 1 kHz in order for the CSI to show noticeable changes due to the movement. The effect on the rate of transmission should be investigate to enhance the efficiency of the model as increasing the frame rate will

increases the number of samples, and hence the training time increases. Increasing the frame rate may also not help further after some point because human movement speed is limited in indoor areas.

Apart from that, the CSI characteristics are not the same for different environments and different people. Therefore, the CSI information due to the stationary objects can be obtained by using technique such as PCA. By discarding the first PCA, the information due to the mobile target is mainly captured. Due to similarity in the change of signal reflections, similar features can be obtained for different environments and people by using this technique, which reduce the complexity of the dataset.

REFERENCES

[1] B. Jokanovic, M. Amin, and F. Ahmad, "Radar fall motion detection using deep learning," *2016 IEEE Radar Conf. RadarConf 2016*, 2016.

[2] J. Goto, T. Kidokoro, T. Ogura, and S. Suzuki, "Activity recognition system for watching over infant children," *Proc. - IEEE Int. Work. Robot Hum. Interact. Commun.*, pp. 473–477, 2013.

[3] A. S. Abdull Sukor, A. Zakaria, and N. Abdul Rahim, "Activity recognition using accelerometer sensor and machine learning classifiers," *Proc. - 2018 IEEE 14th Int. Colloq. Signal Process. its Appl. CSPA 2018*, no. March, pp. 233–238, 2018.

[4] S. Zhu, J. Xu, H. Guo, Q. Liu, S. Wu, and H. Wang, "Indoor human activity recognition based on ambient radar with signal processing and machine learning," *IEEE Int. Conf. Commun.*, vol. 2018-May, pp. 1–6, 2018.

[5] K. T. Song and W. J. Chen, "Human activity recognition using a mobile camera," *URAI 2011 - 2011 8th Int. Conf. Ubiquitous Robot. Ambient Intell.*, pp. 3–8, 2011.

[6] Y. Nishida, S. Murakami, T. Hori, and H. Mizoguchi, "Minimally privacy-violative human location sensor by ultrasonic radar embedded on ceiling,"

Proc. IEEE Sensors, vol. 1, pp. 433–436, 2004.

[7] Z. Wang *et al.*, "A Survey on Human Behavior Recognition Using Channel State Information," *IEEE Access*, vol. 7, no. October, pp. 155986–156024, 2019.

[8] X. Wang, X. Wang, and S. Mao, "ResLoc: Deep residual sharing learning for indoor localization with CSI tensors," *IEEE Int. Symp. Pers. Indoor Mob. Radio Commun. PIMRC*, vol. 2017-Octob, pp. 1–6, 2018.

[9] X. Wang, X. Wang, and S. Mao, "RF Sensing in the Internet of Things: A General Deep Learning Framework," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 62–67, 2018.

[10] S. K. Saksena, B. Navaneethkrishnan, S. Hegde, P. Raja, and R. M. Vishwanath, "Towards Behavioural Cloning for Autonomous Driving," *Proc. - 3rd IEEE Int. Conf. Robot. Comput. IRC 2019*, no. June, pp. 560–567, 2019.

[11] J. H. Han, "Comparing Models for Time Series Analysis," 2018.

[12] S. Hochreiter, "Long Short-Term Memory," vol. 1780, pp. 1735–1780, 1997.

[13] M. X. Gong, B. Hart, and S. Mao, "Advanced Wireless LAN Technologies," *GetMobile Mob. Comput. Commun.*, vol. 18, no. 4, pp. 48–52, 2015.

[14] C. Gioia, F. Sermi, D. Tarchi, and M. Vespe, "On cleaning strategies for WiFi positioning to monitor dynamic crowds," *Appl. Geomatics*, vol. 11, no. 4, pp. 381–399, 2019.

[15] C. Chilipirea, A. C. Petre, C. Dobre, and M. Van Steen, "Presumably simple:

Monitoring crowds using WiFi," *Proc. - IEEE Int. Conf. Mob. Data Manag.*, vol. 2016-July, no. November 2019, pp. 220–225, 2016.

[16] Z. Hussain, M. Sheng, and W. E. Zhang, "Different Approaches for Human Activity Recognition: A Survey," pp. 1–28, 2019.

[17] H. Jiang, C. Cai, X. Ma, Y. Yang, and J. Liu, "Smart Home Based on WiFi Sensing: A Survey," *IEEE Access*, vol. 6, no. c, pp. 13317–13325, 2018.

[18] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-Free Fall Detection by Wireless Networks," *IEEE Trans. Mob. Comput.*, vol. 16, no. 2, pp. 581–594, 2017.

[19] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "RT-Fall: A Real-Time and Contactless Fall Detection System with Commodity WiFi Devices," *IEEE Trans. Mob. Comput.*, vol. 16, no. 2, pp. 511–526, 2017.

[20] M. Scholz, T. Riedel, M. Hock, and M. Beigl, "Device-free and device-bound activity recognition using radio signal strength," *ACM Int. Conf. Proceeding Ser.*, pp. 100–107, 2013.

[21] Y. Gu, L. Quan, and F. Ren, "WiFi-assisted human activity recognition," *Proceedings, APWiMob 2014 IEEE Asia Pacific Conf. Wirel. Mob. 2014*, pp. 60–65, 2014.

[22] H. Abdelnasser, M. Youssef, and K. A. Harras, "WiGest: A ubiquitous WiFi-based gesture recognition system," *Proc. - IEEE INFOCOM*, vol. 26, pp. 1472–1480, 2015.

[23] F. Jaais and A. F. A. Rahman, *Trabeculectomy - A review and 2 year follow up*, vol. 59, no. 3. 2004.

[24] S. Sardy, P. Tseng, and A. G. Bruce, "Robust wavelet denoising," *IEEE Trans. Signal Process.*, vol. 49, no. 6, pp. 1146–1152, 2001.

[25] W. Jiang *et al.*, "Towards environment independent device free human activity recognition," *Proc. Annu. Int. Conf. Mob. Comput. Networking, MOBICOM*, pp. 289–304, 2018.

[26] X. Wang, "RF Sensing for Internet of Things: When Machine Learning Meets Channel State Information," 2018.

[27] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *Comput. Commun. Rev.*, vol. 41, no. 1, p. 53, 2011.

[28] M. A. A. Al-Qaness *et al.*, "Channel state information from pure communication to sense and track human motion: A survey," *Sensors (Switzerland)*, vol. 19, no. 15, pp. 1–27, 2019.

[29] J. Sanguino and R. W. Roberts, "Wireless Communications Adapter For A Hearing Assistance Device," *J. Acoust. Soc. Am.*, vol. 129, no. 4, p. 2354, 2011.

[30] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," *Proc. Annu. Int. Conf. Mob. Comput. Networking, MOBICOM*, pp. 617–628, 2014.

[31] A. Q. Mohammed, "Indoor micro-activity recognition method using ubiquitous WiFi devices," *Proc. 5th IEEE Conf. Ubiquitous Positioning, Indoor Navig. Locat. Serv. UPINLBS 2018*, pp. 1–7, 2018.

[32] F. Adib and D. Katabi, "See through walls with WiFi!," *Comput. Commun. Rev.*, vol. 43, no. 4, pp. 75–86, 2013.

[33] X. Wang, S. Member, L. Gao, S. Mao, and S. Member, "IEEE INTERNET OF THINGS JOURNAL 1 CSI Phase Fingerprinting for Indoor Localization with a Deep Learning Approach," pp. 1–11.

[34] G. Pecoraro, E. Cianca, S. Di Domenico, and M. De Sanctis, "LTE Signal Fingerprinting Device-Free Passive Localization Robust to Environment Changes," *6th Glob. Wirel. Summit, GWS 2018*, pp. 114–118, 2018.

[35] F. Li, M. A. A. Al-Qaness, Y. Zhang, B. Zhao, and X. Luan, "A robust and device-free system for the recognition and classification of elderly activities," *Sensors (Switzerland)*, vol. 16, no. 12, pp. 1–12, 2016.

[36] K. Qian, C. Wu, Z. Yang, Y. Liu, and Z. Zhou, "PADS: Passive detection of moving targets with dynamic speed using PHY layer information," *Proc. Int. Conf. Parallel Distrib. Syst. - ICPADS*, vol. 2015-April, pp. 1–8, 2014.

[37] O. T. Ibrahim, W. Gomaa, and M. Youssef, "CrossCount: A deep learning system for device-free human counting using WiFi," *IEEE Sens. J.*, vol. 19, no. 21, pp. 9921–9928, 2019.

[38] A. Kalyanaraman, "FormaTrack : Tracking People based on Body Shape," vol. 1, no. 3, pp. 1–21, 2017.

[39] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Device-Free Human Activity Recognition Using Commercial WiFi Devices," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 5, pp. 1118–1131, 2017.

[40] S. Yousefi, H. Narui, S. Dayal, S. Ermon, and S. Valaee, "A Survey of Human Activity Recognition Using WiFi CSI," pp. 1–8, 2017.

[41] L. Zhang, Z. Wang, and L. Yang, "Commercial Wi-Fi Based Fall Detection with Environment Influence Mitigation," *Annu. IEEE Commun. Soc. Conf. Sensor, Mesh Ad Hoc Commun. Networks Work.*, vol. 2019-June, pp. 1–9, 2019.

[42] G. Wang, Y. Zou, Z. Zhou, K. Wu, and L. M. Ni, "We Can Hear You with Wi-Fi!," *IEEE Trans. Mob. Comput.*, vol. 15, no. 11, pp. 2907–2920, 2016.

[43] P. Wang, B. Guo, T. Xin, Z. Wang, and Z. Yu, "TinySense: Multi-user respiration detection using Wi-Fi CSI signals," *2017 IEEE 19th Int. Conf. e-Health Networking, Appl. Serv. Heal. 2017*, vol. 2017-Decem, pp. 1–6, 2017.

[44] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," *Proc. Annu. Int. Conf. Mob. Comput. Networking, MOBICOM*, vol. 2015-Septe, pp. 90–102, 2015.

[45] B. Lu, Z. Zeng, L. Wang, B. Peck, D. Qiao, and M. Segal, "Confining Wi-Fi coverage: A crowdsourced method using physical layer information," *2016 13th Annu. IEEE Int. Conf. Sensing, Commun. Networking, SECON 2016*, vol. 2, 2016.

[46] Y. Wu *et al.*, "Google ' s Neural Machine Translation System : Bridging the

Gap between Human and Machine Translation," pp. 1–23.

[47] Y. Yu, C. Hu, X. Si, J. Zheng, and J. Zhang, "Averaged Bi-LSTM networks for RUL prognostics with non-life-cycle labeled dataset," *Neurocomputing*, 2020.

[48] L. Zhu *et al.*, "Landslide susceptibility prediction modeling based on remote sensing and a novel deep learning algorithm of a cascade-parallel recurrent neural network," *Sensors (Switzerland)*, vol. 20, no. 6, 2020.

[49] D. C. Nguyen, G. Bailly, and F. Elisei, "Comparing cascaded LSTM architectures for generating head motion from speech in task-oriented dialogs," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 10903 LNCS, pp. 164–175, 2018.