Wi-Fi Based Supervised Machine Learning Approach to Detect Objects Activities

by

Samia Salam Suzana
17301075
Israt Jahan Prome
16101102
Ahnaf Hossain Siddique
16101119
Zakia Rahman
16101089
Farzana Nasrin
17101090

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
January 2021

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Zakia Rahman
16101089

Farzana Nasrin

Farzana Nasrin
17101090

Israt Jahan

Israt Jahan Prome
16101102

Samia Salam Suzana
17301075

Ahnaf Hossain Siddique
16101119

Approval

The thesis/project titled "Wi-Fi Based Supervised Machine Learning Approach to Detect Objects Activities" submitted by

1. Samia Salam Suzana (17301075)
2. Israt Jahan Prome (16101102)
3. Ahnaf Hossain Siddique (16101119)
4. Zakia Rahman (16101089)
5. Farzana Nasrin (17101090)

Of Fall, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January 11, 2021.

Examining Committee:

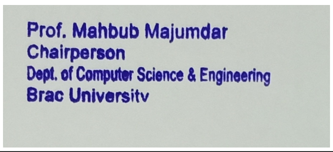
Supervisor: (Member)

Esfar E Alam Turzo
Associate Lecturer
CSE Department
Brac University

Program Coordinator: (Member)

Dr. Md. Golam Rabiul Alam
Associate Professor
CSE Department
Brac University

Head of Department: (Chair)



Mahbubul Alam Majumdar, PhD
Professor and Chairperson
Department of Computer Science and Engineering
Brac University

Ethics Statement

Since we have aimed our research on the Wi-Fi based supervised learning to detect movements, we had to collect many research papers and projects from other authors. In order to take the help and guide from the research works, we made sure to take their permission through electronic mail (e-mail) to continue our research further based on their papers. And also we vow to provide all our research work to build the system in the better and utmost usage of mankind.

Abstract

Wireless signal based movement detection technologies are emerging effectively. This type of technology is quite popular compared to other ones because it relieves people from the hassle of wearing sensors and coming in close contact with the detection model. It is possible to detect details of movement pattern of the subject if it comes within the range of wireless signal. In this paper, we have presented a supervised learning approach based on Wi-Fi to detect movements of object. A supervised learning is a machine learning approach where a model is trained with input and desired output beforehand. For our research, we have used a dataset which contains Channel State Information (CSI) data for walk, run, standup, seat etc. separately. In this paper we have worked on three kinds of data such as walk, run and standup. We have trained our model with these CSI data and later applied time series, Augmented Dickey-Fuller Test (ADF) and different machine learning algorithms such as Decision tree, Liner Regression model, Random Forest etc. on those data. Lastly, a rigorous comparison is made between trained and test data in order to validate the accuracy of our result.

Keywords: Machine Learning, Supervised Learning, W-Fi, Channel State Information (CSI), Movement Detection.

Dedication

This research paper is dedicated to the peoples suffering in lack of surety and uncertainty. In contemplation of being in a safe zone and secured around known movements of everyone is a comfort and amiable. Our system will serve the purpose of certifying the security concerns of everyone in beneficial to their safety and certainty.

Acknowledgement

First of all, we want to express our gratitude towards Almighty Allah who has kept us safe during this pandemic. As a result, we could complete our research. Besides, we want to appreciate the effort of our department and all the faculty members for introducing online learning platform and successfully running the semesters. We want to thank our supervisor specially for providing us continuous feedback, correcting our mistakes and guiding us properly throughout the whole time. Lastly, we want to thank our parents and family members whose prayers and good wishes helped us to continue our research in this extra ordinary situation.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Ensuring security is becoming more concerning day by day. With the rise of technology, criminal activities are also increasing. The more the world is evolving rapidly with the present and upcoming developed technologies, innovations and researches, the more security issues are rising. To prevent these issues many people are working to improve the security system which will protect people as well as the environment from attack or any kind of harm. Likewise, our main goal was also to find the accurate ways to prevent these issues. We tried to present the supervised learning approach which is based on Wi-Fi to detect movements of an object for the security purposes. In order to achieve our goal, we researched many procedures which will help us to implement further in building the system which will be able to identify any object movement using the wireless based signals. These signals will not only be efficient but also accurate. For example, when we are at home we are untold about our outer environment. We do not know who is passing by near our houses or apartments. Not only in our houses but also we are always unspecified about the movements of living beings or objects around us while we are at home or at work. This not only brings out our concerns but also makes us vulnerable. For another scenario, whenever working parents will be outside of their houses leaving their children behind there can happen a lot of mishap regarding this security issue. Any kind of subject can break into the house and cause harm. This torment can happen to anyone and it can happen anywhere. Outside the bank there should be protection in order for the security and safety of peoples valuable wealth and goods. Similarly in an open environment like parks, lakes, picnic spots or any other, the security must be widened regarding preventing any kind of harmful events or situations. Sometimes, it is seen that due to lack of effective and accurate recognition or even unknown whereabouts, criminals and law offenders are walking in broad daylight under the nose of the law enforcement authority. In many countries in the world, People have been using many systems to ensure their security and safety. One kind of security is the detection of the movements of the living things which will notify anybody using the system about the movements of anyone near them or near their house premises or offices. We have gone through many studies that researched this particular topic about detecting object movement using Wi-Fi based sensors. In our research we tried to come about a Wi-Fi based supervised learning approach to detect movement successfully so that it can alert about the unknown entrance and identify the objects movement. Users can be always alerted and prevent any attack or vicious activity. In this paper, an approach towards the detection using Wi-Fi which can detect movements and then using the Channel State Information (CSI) to process the data with the time series is made. We have worked on our system's accuracy and efficiency as well as researched our limitations which we will implement in our future work.

1.1 Problem Statement

In this era of 2021, many people are still suffering from security due to lack of unauthorized access of living beings as well as objects. Financial crises are also at peak due to pandemic issues as well as economic crises. In order to use the technology to detect movement using wireless based signals, it is not always possible to set up the system. People who have financial difficulties, will not be able to use the system and that is why, as a reason there will be breach in their security control and privacy violation. When illegal access by any object or human presence will remain undetected then the owner of the house or people in the area of the environment will not stay protected and may be attacked. The person's vulnerability will increase so as many crimes may also rise which will go unnoticed. It will then not only be a silly issue or a sudden occurrence rather it will cause huge mayhem if it continues to proceed further. People will feel unsafe with their life as well as their expensive goods and wealth. This also refers to the issue of environmental limitation. Due to the intrusion without the detecting systems the living environment will not sustain and will fail to provide the security of the people but also it will not be able to identify the attacks.

1.2 Aim of Study

To reduce this kind of activity is our main aim of this research. In our thesis paper, we tried to present the supervised learning approach which is based on Wi-Fi to detect movements of an object for security purposes. It is such a machine learning approach where a model is trained with input and desired output beforehand. Not only it will contain the kinds of data such as walk but also it will contain the data with run and stand up. We tried to train our model with the using of CSI data and then processing it further with the time series. These procedures will help us to implement the system which will be able to identify the object movement using wireless based signals accurately and efficiently. It will thus provide and ensure security and protection. As we are heading to a digital new era, our system will get ready in order to provide the services in every sector.

1.3 Research Methodology

In our research we tried to develop a Wi-Fi based supervised learning approach to detect movements successfully so that it can be used for security purposes. We will attempt to make a system that will be able to detect any type of doubtful person appearing in a system by scanning the data and then matching or analyzing the data in the future. The system then shall be able to distinguish the subjects presence of who are suspicious and who are not and give an alert through the security system.

Since we have developed the study regarding Wi-Fi based supervised learning, we tried to extract the data from our dataset which contained the data of the running file, walking file and the stand up file. It also contained the ideal running and walking value from which we then took three dataset of each category and after visualizing it by Python, we compared the train and test visualization graph. Later on we used time series and the Augmented Dickey-Fuller (ADF) test due to having a timestamp in our data set and to find out if the data is stationary or non stationary. Later on, we tested the data to determine how accurately the model can predict the result, and thus we have reached a conclusion that our supervised approach is successful.

1.4 Thesis Outline

This research paper authorizes UN constructing this system which would be beneficial in the future. It emphasizes to assure The supervised learning approach based on Wi-Fi to detect movements of an object. The following steps for our research are described below.

Firstly, in the introduction part of Chapter 1, it states the reason, aim, and the motivation which brought a huge importance regarding this topic. It not only focuses on technological development but mainly focuses on the security of everybody. The reason behind selecting this topic, the limitations, the aim to research on this topic, and the methodology which are used in this paper are briefly described.

Secondly, in Chapter 2, we have discussed the similar papers we have found while doing our research. Most authors have had the same aim and motivation like us and were succeeded. Many of them also had their limitations like us and worked on the solutions regarding those problems. Few papers having the solution to those limitations were able to shed light on many different technologies which are more efficient, beneficial and cost effective. In this chapter we have discussed those matters.

Thirdly, in Chapter 3, we have discussed about data collection, approach and extraction, human movement impact on remote channel, highlight extraction process in details. Description of Channel State Information, data plotting, PCA component analysis along with graphs of amplitude, spectrogram, PCA component and magnitude spectrum are included here. We have discussed about our present work flow and initial workflow in this section. For graph generation we have selected three data sets for each type of ideal walking, running and stand up.

Furthermore, in Chapter 4, we have stated machine learning, supervised learning, the usage of algorithms such as, decision tree algorithm, random forest and linear regression algorithm. Then we have described the time series which we used due to having a timestamp in our data set and also the usage of the Augmented Dickey-Fuller test which we used to find out if the data is stationary or non-stationary. After that, we have calculated test-statistic, p-value and critical value at different percentages. Then we have compared the p-value obtained from train data of walk, run and stand up data with the test data. Therefore, we have compared the graph obtained from the Principal component analysis to detect if the graphs obtained from train and test data are similar or not. Based on their similarities we then work on their categories of which serves the same as the graph.

Lastly, in Chapter 5, we have concluded the whole research paper and have discussed our future work plan and the limitations we have faced while doing this research.

1.5 Literature Review

Wi-Fi based activity detection is a new technology of supervised learning recognition system. This process gathers human movement and extracts, which authenticate the specific identity of a person in motion. Here [5], we came to know that this technique involves 2 stages: Gathering information from human locomotion and secondly extraction of data for further process.

In [24], we came to know about that, currently developed Wi-Fi based recognition systems passes by two critical processes gathering data and extracting the features of those data. Including these as Wi-Fi devices contain much noise it creates difficulty to detect the signal.

In [18], we can learn about developing a Wireless Fidelity (Wi-Fi) based recognition system. As each human being has unique body shape, movement gesture, walking style and some continuous habit of a specific movement, it is possible to examine the changes on the Wi-Fi spectrum. This uniqueness will lead us to get different result for different individuals which will help us to identify by examining these using Channel State Information (CSI).

In [14], we can learn about using the Wi-Fi signals to profile human movements using Channel State Information (CSI). Specifically we can use signal processing technique. To extract features from spectrograms that best characterize the ambulating pattern, By performing autocorrelation on the torso reflection to abstract false signals in spectrograms. By calculating that on a pre-set of data with 2,800 gait instances gathered by 50 human beings in a space of 50 square meter. Experimental findings show that WifiU achieves 79.28%, 89.52%, and 93.05% respectively top-1, top-2, and top-3 accuracies in perception.

In [43], we can learn about extracting the unique features from the system by analyzing the Channel State Information (CSI). They characterized by several differences like hand, feet etc. to make their system more accurate and efficient. These patterns are also reliable as they are highly repetitive. We also can learn about the usages and implementation of passive reception of Wi-FI signals. There we can also learn about the signal separation equations and methods depending on the signals. Also in similar matter in [8] the author did a same experiment and showed us the correlation between the velocity of movement of human limbs and the frequency components.

Chapter 2

Related Work

Detection of subject presence technology is preceding effectively through various recognition systems. Many have used the detection using Wi-Fi and then finding with time series as well as Many used Channel State Information (CSI) dataset for their research purpose. We have found many related papers regarding this topic. In one paper it was discovered that, [23] in recent years it's been a rapid development of object activity recognition using Wi-Fi, given its advantages of low cost, ubiquitous deployment, privacy protection, etc. Under many circumstances it is really useful if there is any presence of humans or any living being detections anywhere in the environment. Analyzing many factors they tried to focus on the effective movement during travel through generating Wi-Fi signals and specific signal patterns in their subject body. They also tried to focus on the low-cost budget which can be used almost in every country regarding their usage and security. This is one of the important factors which is similar to ours. We tried to analyze our research work which could be implemented in our future work. The main purpose of our research is also to focus on the security issue, privacy protection, and implementing it all in a low budget.

In another paper it is stated that, [9] behavior identification in smart environments is really important using interaction based sensing. They tried to address the multiple activities which will mostly concentrate on the single resident case considering the domain specific needs with multiple purposes. In order to initiate their plan they applied machine learning which would give the accurate and efficient result. They mainly emphasized on detection along with the diverse characteristics of the human and other beings, their movements and interaction with the environments and their nature of differences of activities. Not all humans or living beings have the same pattern of movement or reciprocity. Similarly, their pattern of activities becomes pretentious due to external factors. In another paper, [15] the authors focused on the subject detection using Wi-Fi signals which can be uniquely done and initiated a system, known to be Wi-Fi-ID which analyzes the channel state identification to identify the signals of the movement of an object without making any error. They also proposed to ensure that the system can be conducted in a small area, for example in home or in a small office by [15] demonstrating the whole system with an accuracy of 93% to 77% from a group of 2 to 6 living subjects, respectively.

Likewise, in another paper, it is mentioned that many problems might arise while detecting and recognizing the movement due to the customized wear, emission of energy and the external factors accurately. It will be highly challenging for the researchers to carry on the system procedure whereas [32] Vision based object activity analysis which is based by computer vision and machine learning has its limitations that are unable to detect behind the wall surfaces or in the dark places and also it is unable to detect the sensation of the subject with cameras everywhere. Recent studies in this sector [32] gave new solutions as it has been proven that the movement of the objects' physiques will affect the Channel State Information (CSI) of wireless signals in the indoor environments. This gives the advantage to work in dark areas as well as in the areas where the sight gets an optical resistance. Another shows that, [17] since PIR sensors experience low range and require a number of sensors to detect the movement accurately, technologies such as radar, ultra wideband SDR-based solutions are the best option but they are too expensive. It has optical resistances also and the audio record can be a concern in the light of privacy issues. Also, some sensors require the pedestrian or object to carry it along them which can be highly comfortable and may cause malware function. To remove all these complications, they examined that the Channel State Information(CSI) is highly efficient to identify basic human movements, due to [17] invisibly filling in the air with a spectrum of radio frequency signals. Using the system, when a person or an object is in motion, the signal they propagate will be captured and shown on the Wi-Fi spectrum. The changes will be visible on the spectrum and therefore will be uniquely examined by the Channel State Information (CSI).

Chapter 3

Data collection and Analysis

3.1 Data Collection

There are mainly two ways to collect data and create a dataset, Primary collection and secondary collection. Primary data collection is collecting data directly from the field while researching. Researchers examine and collect data from the scratch. Secondary data collection is collecting data from researched studies, tests from primary data to continue their research. In our research, we have used the method of secondary data collection to collect our data. Our purpose was to detect object movement using Wi-Fi based supervised learning. We searched the internet for related topic, found many works related to this. Finally, we found a related dataset where they worked on human behavior recognition using Wi-Fi. Although, we used a secondary dataset, at the same time we have studied and done our research by ourselves. We continued our research, extracting and understanding the data, data analysis, techniques, feature and Doppler extraction, CSI, graphs and comparison.

First of all, we have collected our dataset and used Linux OS to pull out the data from our dataset. It was compressed in tar.gz format. We have used following commands to do it:

gunzip -v Dataset.tar.gz
The command uncompressed the dataset from gz to tar file.
tar -tvf Dataset.tar
tar -xvf Dataset.tar
These commands extracted the data from the dataset tar file. Our data was in CSV format.

3.2 Approach and Extraction

3.2.1 Wi-Fi Frequency

Received signal strength (RSS) is used to measure the signal of the desired area. It is easy to use but it cannot show the exact change of object movement because it is not very stable even in a stable situation [7]. To have a good Wi-Fi signal in an active localized area, the RSS technique is implemented here. [12]. In case of different objects like mobile also can be tracked down in a large manner by using RSS [3]. The RSS indicates the measurement of the signal power between the access point (AP) and the router. When an individual comes between this, the signal will be sent and RSS will be measured.

3.2.2 Configuring Wi-Fi

Since, the RSS cannot show the exact measure of change, to get more information Wi-Fi systems can be modified. It will extract and gather more information from the Wi-Fi signal. A modified Wi-Fi system that can be used to track passively 3D in WiTrack is The USRP software radio system [15]. Frequency modulated carrier wave (FMCW) is a technique that is used to detect human movement. It can take the measure happening in the Doppler shift in OFDM signals. Doppler shift measures the distance between the receiver and the object, USRP system is used to have it from OFDM signals due to human movement. When the object is closed to the receiver it causes a positive Doppler shift. And, when the object is moving in distance from the receiver it causes a negative Doppler shift. For example, if the object is making movement at 0.73 m/s at a 6.1GHz module, the approximate Doppler shift will be 17.1Hz [6]. These tiny change or shifts cannot be detected. We have to focus on the smaller Doppler shifts as it is really necessary.

3.2.3 Extraction of Doppler Shift

After extricating the Doppler data, WiSee calculates the repeat time Doppler profile. It takes the FFT models in a range of an enormous part of a single second. This system is generally called STFT, stands for short time Fourier transform. It has been used in various places too [2], [11]. Human movement fundamentally has a range of motion of 0.27m/s-4m/s. 5 GHz of the Doppler shifts is some place in the scope of 8Hz and 134Hz, hence the FFT yield in this repeat range is considered in WiSee.

3.2.3.1 Segments

The subsequent stage is to fragment the STFT information to recognize various examples. For example, with a positive and a negative Doppler move, a movement may involve one area at Doppler developments in positively or negatively, or might be two parts. Distinguishing a bit relies upon the energy ID over a little term. Starting of the part can be found when the energy level is 3dB higher than the disturbance level. When the level is under 3dB, by then the segment has been ended.

3.2.4 Human Movement Impact on Remote Channel

Improvement of the individuals and articles change the multipath typical for the distant channel and from now on it will have a substitute ampleness and stage. For one subcarrier and all of the radio wires, the CSI sufficiency is connected to an individual. Walking and plunking down Between a Wi-Fi transmitter and recipient the sit-down and walking is spoken in fig 3.1. The individual is fixed for the underlying 400 bundles yet then beginnings walk or sit-down. As seen, when the individual isn't moving, the CSI amplitudes for all gathering contraptions are for the most part consistent, regardless, the CSIs begins advancing unquestionably when the individual begin to move. In the examination the sitting period is lesser than the walking period, since when the individual plunks down he remains fixed. The got stage, is uncommonly turned on account of the SFO and the CFO, as referred to before. Regardless, using the stage disinfection technique, the effect of goofs in stage can be abstained from.

3.2.5 Highlight Extraction

One approach to remove highlights from a sign is to change it to another area, for example, recurrence space. The Fast Fourier Transform (FFT), which is a productive execution of DFT-discrete Fourier transform that could be utilized for the reason. Then, a particular number of CSI tests choses a window size. Then, afterward by sliding the window, the First Fourier Transform is prosecuted on each section. The procedure is called short time Fourier transform (STFT), which can identify the recurrence changes of a sign after some time. The STFT was utilized in radar for location of the development of the middle and legs in [2]. Various exercises appears for CSI information gathered at 1KHz rate for the spectrogram of the CSI. The exercises that include extreme developments show high steam in high frequencies, for running and walking in the spectrogram. [13], [11], [10] shows, removing highlights from CSI was done by utilizing DWT, as a component of time. It provides elevated time goal to exercises with the high frequencies, also high recurrence goal for exercises with the low rates. Every DWT degree speaks to a recurrence range. In the range, higher recurrence data are low at level and lower frequencies have more significant level. Benefits of DWT to STFT which is referenced in [11] may be:

- A pleasant compromise as expected and recurrence space can be provided by DWT,
- DWT diminishes the size of the information too, so it gets reasonable for AI calculations.

DWT of a 12level in CARM is utilized to disintegrate the 5 main parts (subsequent to eliminating the main head segment). At that point the five estimations of the DWT are found in the middle value of it. For each 200ms, CARM removes a 27 dimensional component vector which includes three arrangements.

3.2.6 Wi-Fi channel state information

3.2.6.1 CSI of Wi-Fi System

The far off devices with IEEE 802.11n/ac standards are using diverse information different yield (MIMO) structure. By using MIMO advancement, it is possible to fabricate the diversity gain, array gain, and also multiplexing gain, in the co-channel lessening block [1]. In IEEE 802.11 has changed OFDM. It is the place where the information transmission is split between different even subcarriers. On account of the little information move limit, the obscuring that each subcarrier faces is shown as level obscuring. Thus, the restricted scale obscuring things of the channel can be updated using OFDM. The channel cross section for subcarrier Greetings, which includes complex characteristics, can be surveyed by parceling the yield signal with an alluded to progression of data in any case called pilot. The channel lattice is generally called the CSI. Since, it shows how the channel is affecting the data picture to arrive at the recipient. Here in OFDM structures, each and every subcarrier faces a tight band obscuring channel. There will be assortment in the saw channel components by gaining the CSI for each subcarrier. Compared to RSS, it is the essential piece of elbowroom of using CSI, in that the movements are found the center estimation of out over all the Wi-Fi bandwidth and along these lines can't get the variation at explicit frequencies. Also in few business network interface cards (NICs).

3.2.7 Deficiency and Errors of Wi-Fi Systems

The plenitude of CSI is usually a strong estimation using for incorporate extraction, more for game plan. Despite the way that transmission power can be changed with it, and also the change of transmission rate. As will be analyzed, the burst clatter can be lessened by applying filtering strategies [11]. Nevertheless, rather than sufficiency, the time of Wi-Fi structure is affected by a couple of wellsprings of misstep, for instance, carrier repeat balance and testing repeat balance (SFO). Differentiation in central frequencies, nonattendance of synchronization between the beneficiary tickers and the transmitter. A 5 GHz Wi-Fi band of the CFO for a time range of 50 MS can be as broad as 80 KHz, provoking stage change of 8. Therefore, the stage changes as a result of the improvement of the body, which is all around more unobtrusive than 0.5ms, isn't discernible from stage change. The other wellspring of slip-up, SFO, is delivered by the recipient easy to cutting edge ADC converter. Moreover, the SFO is moving by subcarrier list, thusly, every subcarrier finds a substitute error. Dark CFO and SFO causes, applying the unrefined stage knowledge which might not be important. In any case, an immediate change is proposed in [4], with the ultimate objective that the CFO and SFO can be wiped out from the changed stage. This cycle is generally called stage cleansing. In LOS condition, the subcarrier list versus CSI abundance, CSI stage and cleansed CSI stage are graphed for a circumstance where the gatherer and the Wi-Fi transmitter are arranged in the district of each other. It is seen, the CSI adequacy is for the most part consistent yet shapes a couple of packs as referred to in [4]. The unrefined stage developed with subcarrier list, as the SFO makes with subcarrier list. After stage cleansing, the distinction in stage as a result of SFO will be diminished.

3.2.8 De-noising of CSI

As the CSI is tumultuous and might be it cannot show specific features for different movements. Thusly, this is essential for the first filter through the racket and subsequently separate a couple of features using AI techniques to be used for portrayal. There are different systems for isolating the racket [11]. In any case, in light of the presence of burst and drive uproars at CSI. It has high bandwidths, the lower channel can't create a sleek CSI stream [11]. Also, it is shown that there can be better methodologies hence, for instance, de-noising of PCA. This is a method for lessening of a tremendous estimation structure implying the likelihood that most of the facts and knowledge about the sign is concentrated upon a part of the features. The essential head part is overboard to diminish the interference, and accompanying 5 ones are used for incorporate eviction. The facts from the dynamic reflection which is coming from the versatile object isn't lost by taking out the primary head section, since this is similarly trapped in another head parts. A couple of features are isolated from it, after PCA demising of CSI data, so it will in general be applied for course of action. The component extraction will be inspected underneath.

3.2.9 Discovery of Human Presence

FreeSense is a novel device, free human presence in indoor area structure, which was proposed by Xin et al. It realizes presence affirmation of human just as shows the recognizing consideration for improvements. The structure can subsequently recognize inside customers by applying the estimation of the MUSIC, AoA, Fresnel zones model, and WI-HD model, have amazing foe impedance ability to contradict disturbance, for instance, multi-way effect and device qualification. At first, the makers use the MUSIC figuring to survey the stage contrast between some radio wired waveforms. Additionally, AoA, WI-HD model and the Fresnel zones model, are improvised to evaluate the identifying granularity and the incorporation size. This investigations show that FreeSense gets a typical 0.54% wrong certain rate and also an ordinary 1.50% sham opposed rate. Also, FreeSense has a typical validity of 1.37 m of the consideration evaluation approach.

3.3 Data Plotting and Analyzing

In our dataset there are around 2200 CSV files of the total data. From those data, we plotted and analyzed the data to prove our test and methods. We did our coding in python and matlab, successfully visualized the plotting in various graphs and tested the comparison to the perfection.

Here, we will be showing our plotting visualization. There are mainly five kinds of information- walking, sitting, fall, run, standup. We have plotted total of 168 files to analyze our data. However, we are visualizing and comparing three Run data plotting in this paper. We have generated four plots per data file. Those are: CSI Amplitude, PCA component, Short-Time Fourier Transform (Spectrogram STFT) and Magnitude spectrum.

3.3.1 CSI Amplitude

As CSI is very noisy it cannot show characteristics for different movement. So, it is important to remove the noise to extract the data from the signal frequency. There are many ways to remove the noise. Howsoever, the existence of huge noise in CSI with high wave frequency, normal filter cannot provide a sleek CSI result. There is more efficient way such as Principal component Analysis (PCA). It could be a way for spatial property reduction for a large data system. Here, most of the data getting from the signal is targeted to change. CARM system, the primary major component is used to cut back the noisy frequency. Thus, the next can be used to extract the features. PCA de-noised the CSI data, and then some date are extracted from it so that it can be used for the classification part.

Figure 3.1-3.3: Amplitude: CSI amplitude plotting of three run files, run.20, run.19 and run.18.

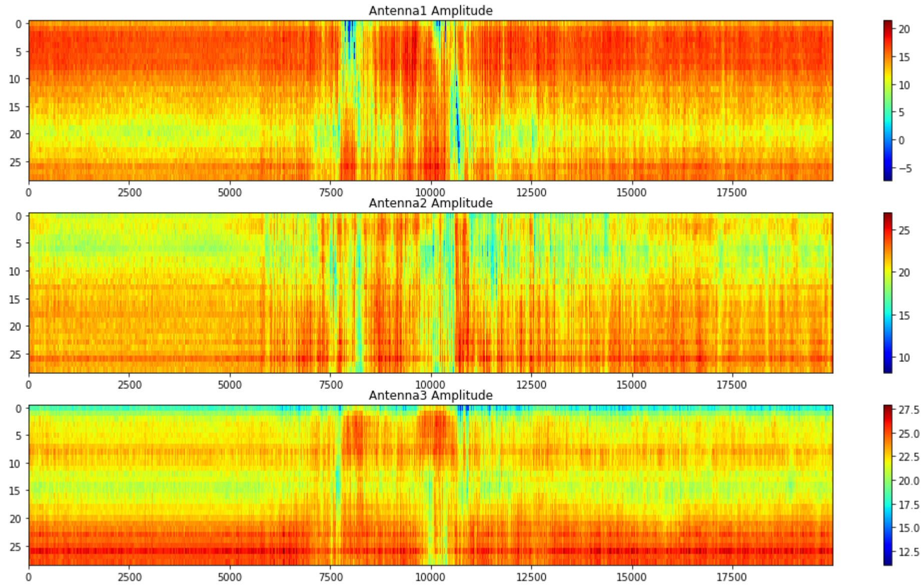


Figure 3.1: run.20 CSI Amplitude

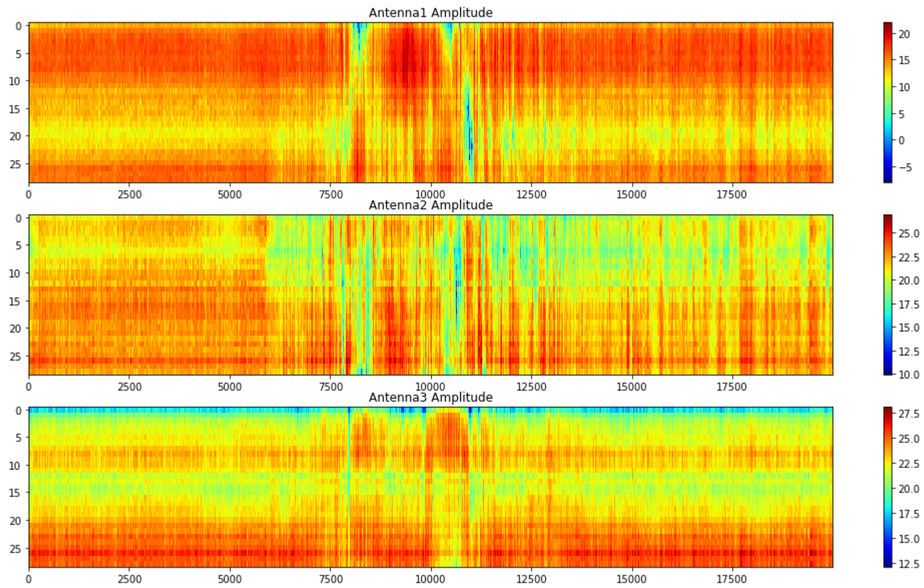


Figure 3.2: run.19 CSI Amplitude

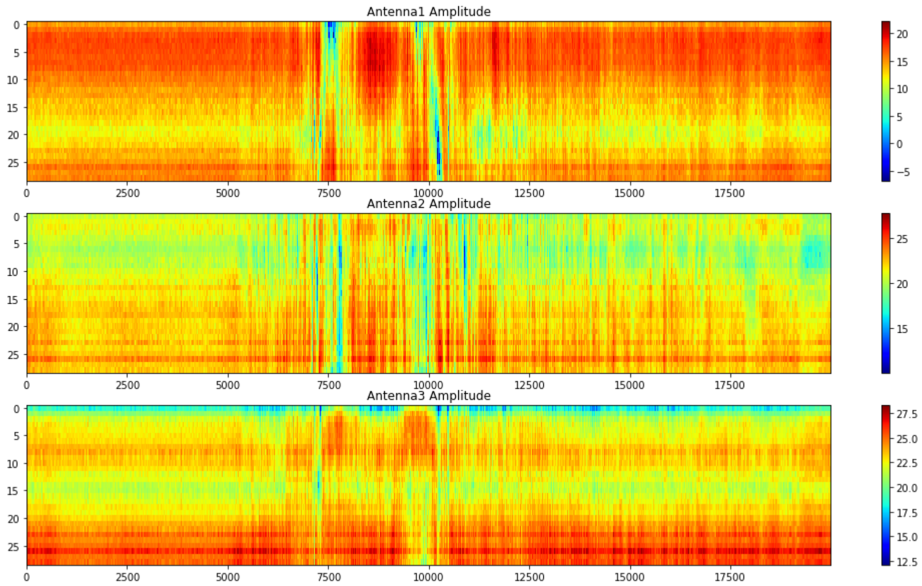


Figure 3.3: run.18 CSI Amplitude

3.3.2 PCA Component

PCA is applied on CSI amplitude. STFT is used to extract options within per 100 MS frequency. As we have a tendency to solely use the primary frequencies. Most of the power of movements is getting in low level frequencies. HMM is applied upon features of extraction which mistreatment STFT, We used the MATLAB program which helped training of HMM, with higher computation time needed for training. Evaluating the LSTM by Tensor flow using Python. There were three antennas. The CSI characteristics are not identical for individuals.

Figure 3.4-3.6: PCA component test figures of 3 run files.

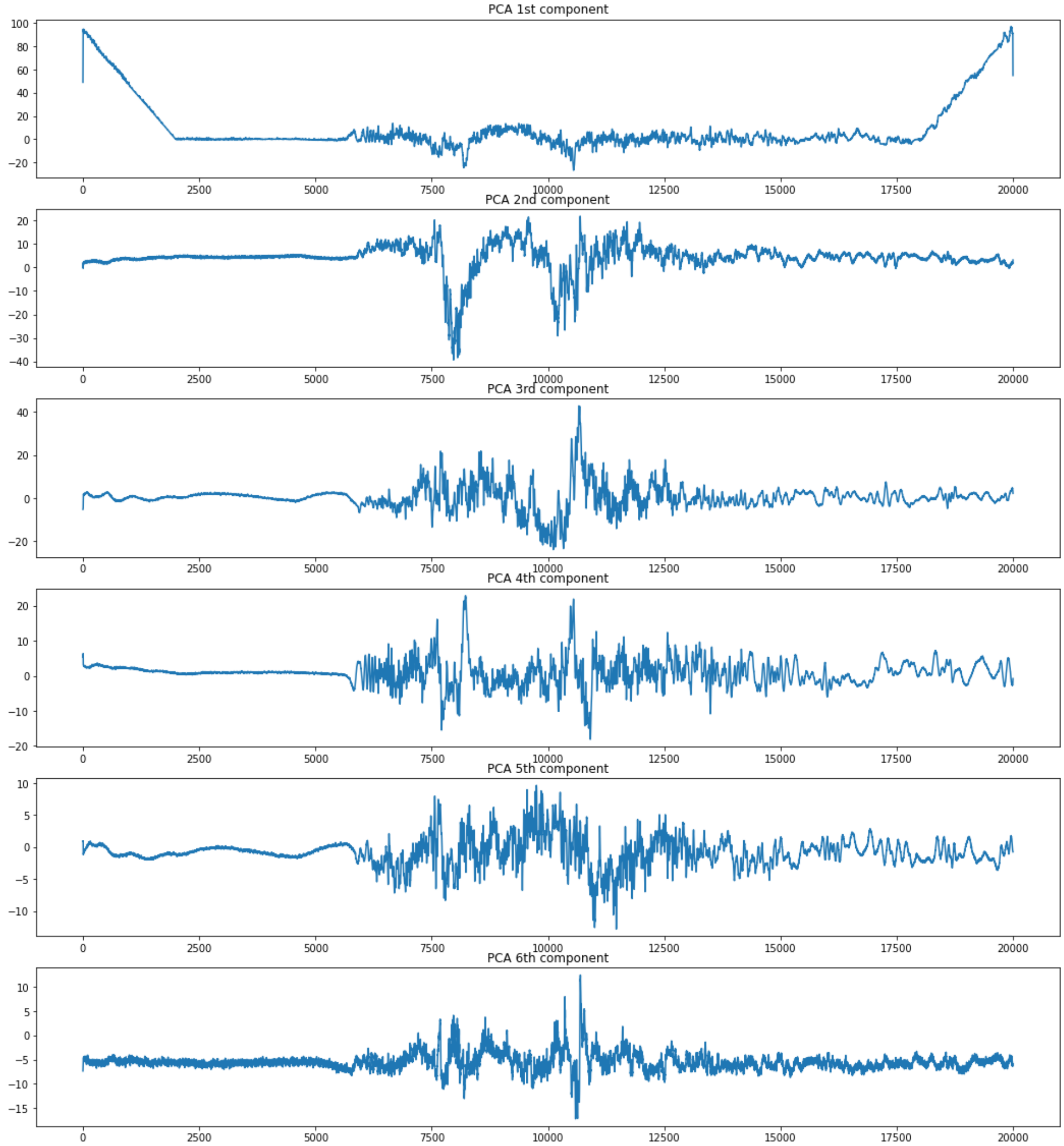


Figure 3.4: run.20 test

The upper data graph was plotted for run file 20. Similarly, here every first figure is shown for run file 20, and the second and third figures are for serially run file 19 and 18. We are comparing 18 and 19 files' plot with run 20 to test if data are validate for the test.

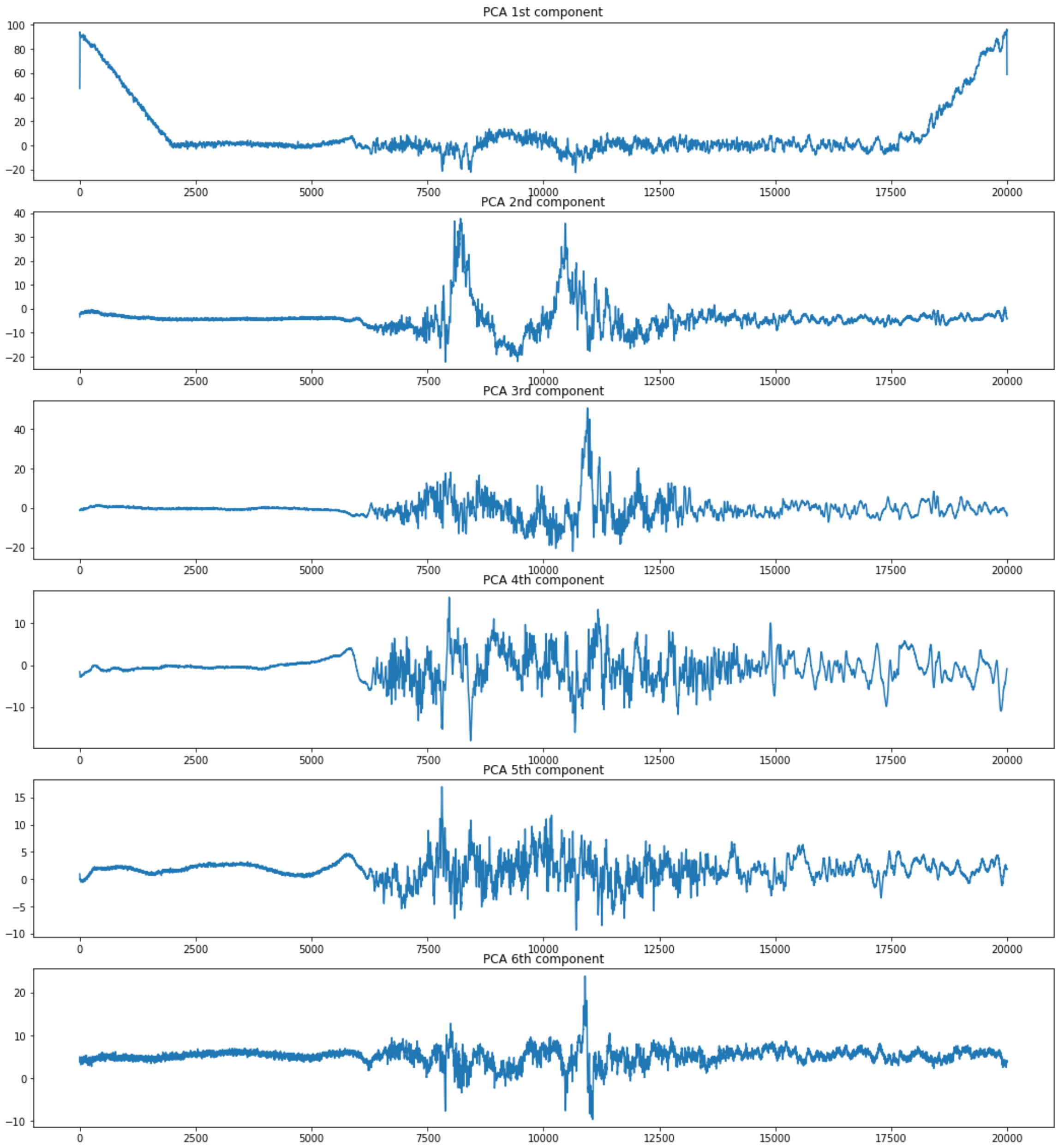


Figure 3.5: run.19 test

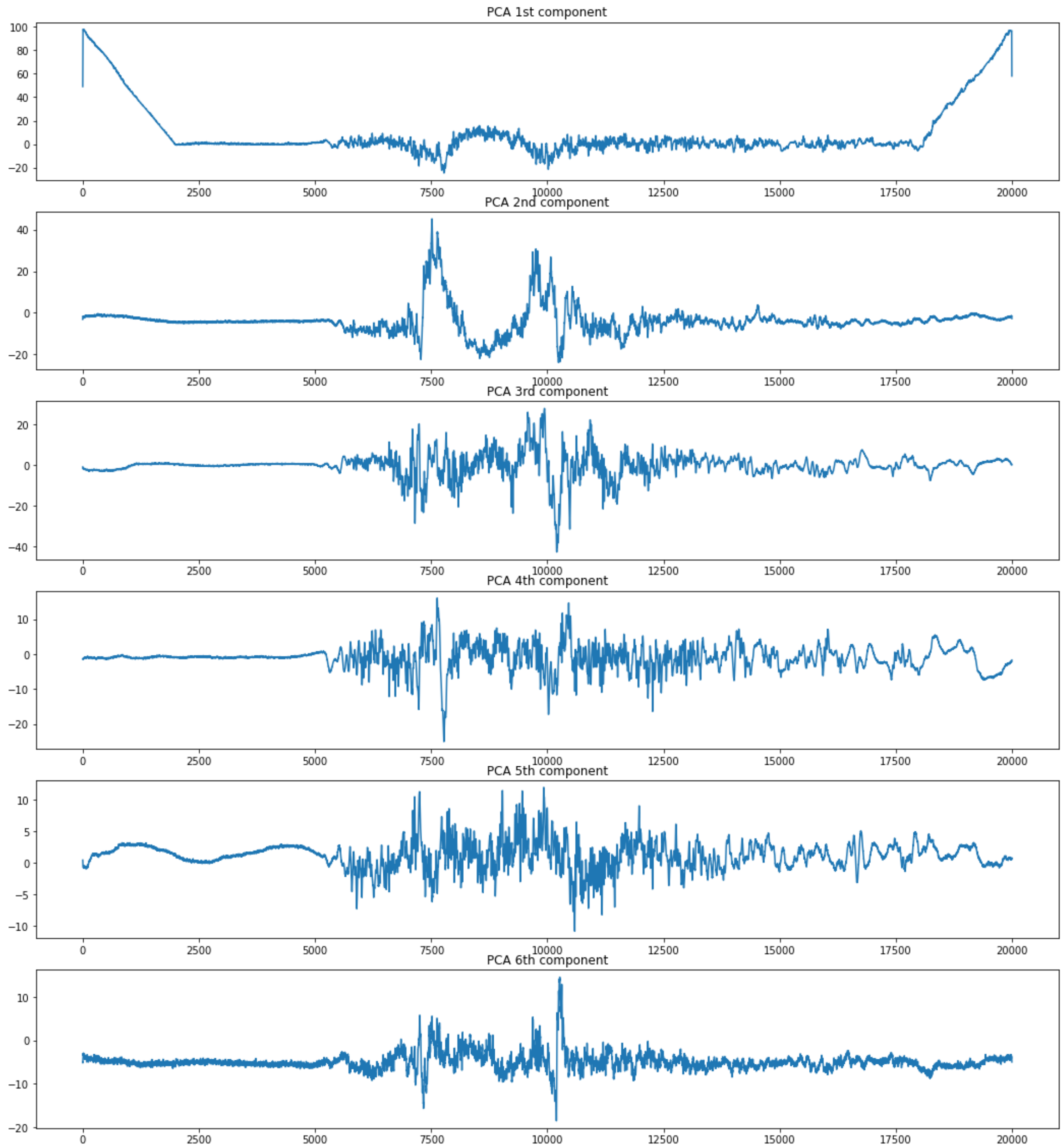


Figure 3.6: run.18 test

In these figures, we can see the positive and negative energy of the frequency. Which stands that data were collected from running activity, as when the object is steady there is less noise and it shows positive. When the movement increases, the noise also increases and the energy is negative.

3.3.3 Spectrogram STFT

There is a way to extract options from a signal which can be done by reworking it and take it to the variant domain. Fast Fourier transform system (FFT), that gives an implementation in an economic way of DFT. It can be useful to do it. At present, various CSI data samples are selected. Then, Fast Fourier Transform can be applied to the segments of the window. The process is called STFT or short-time

Fourier transform, it can notice the change of frequency of the signal from time to time. It was utilized in the radar which detects the movements of objects. CSI STFT spectrogram shows various activities of CSI data. Discovering, it shows the positive and negative result for movement frequency in high or low power.

Figure 3.7-3.9: The STFT spectrogram is plotted as frequency vs. time. It is detecting the object movement and capturing the change of frequency from time to time.

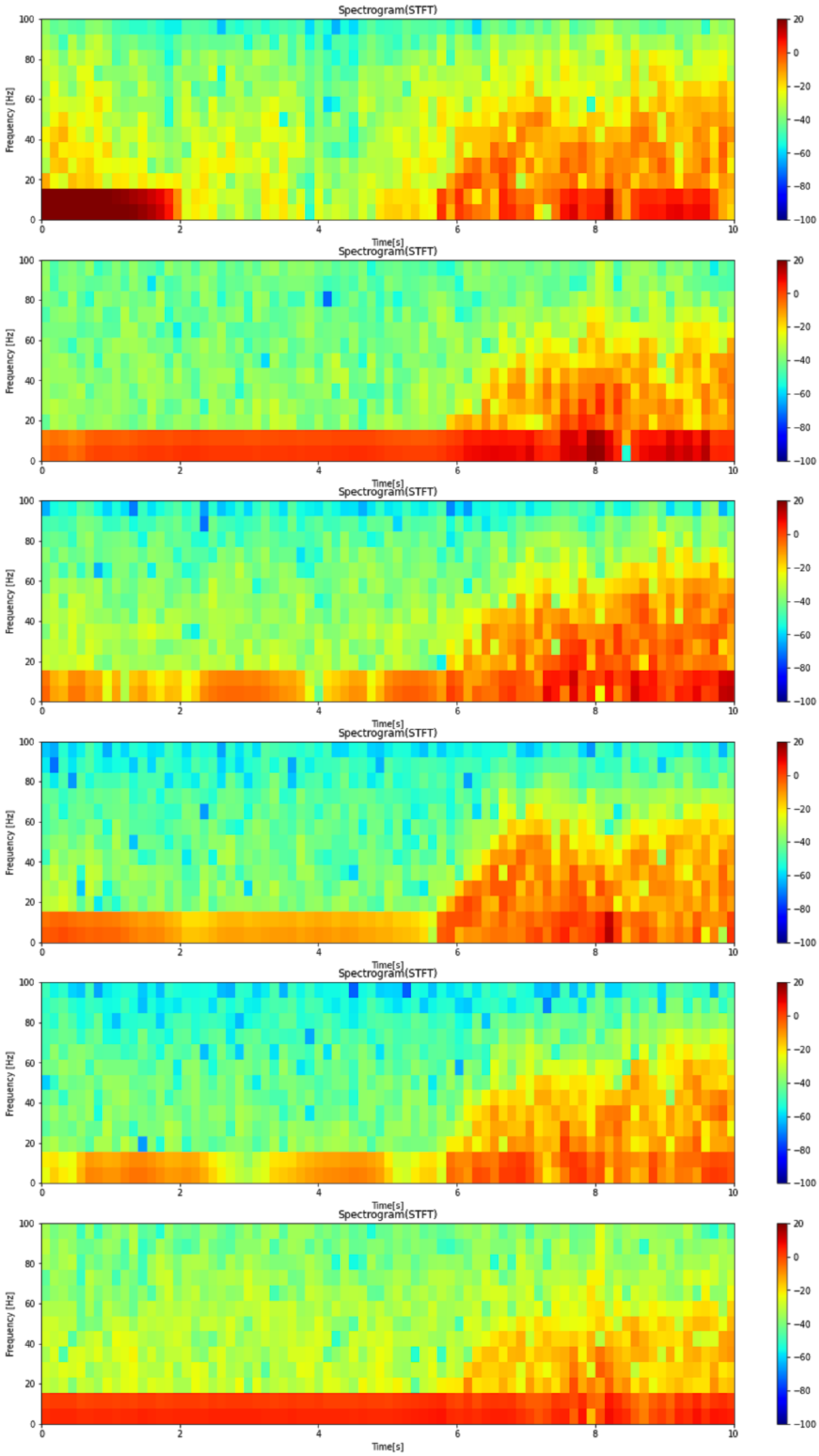


Figure 3.7: Frequency changed over time for run.20

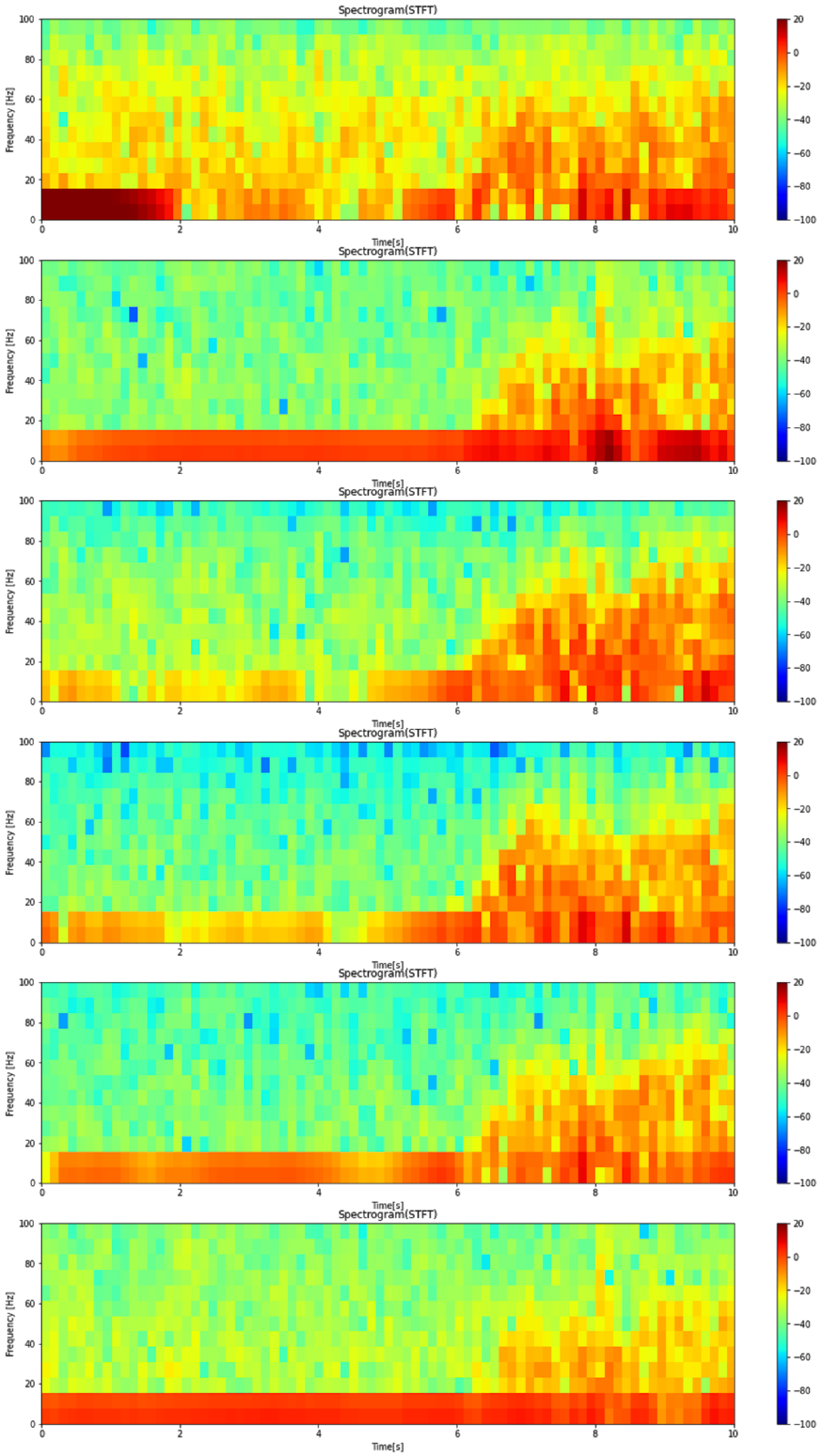


Figure 3.8: Frequency changed over time for run.19

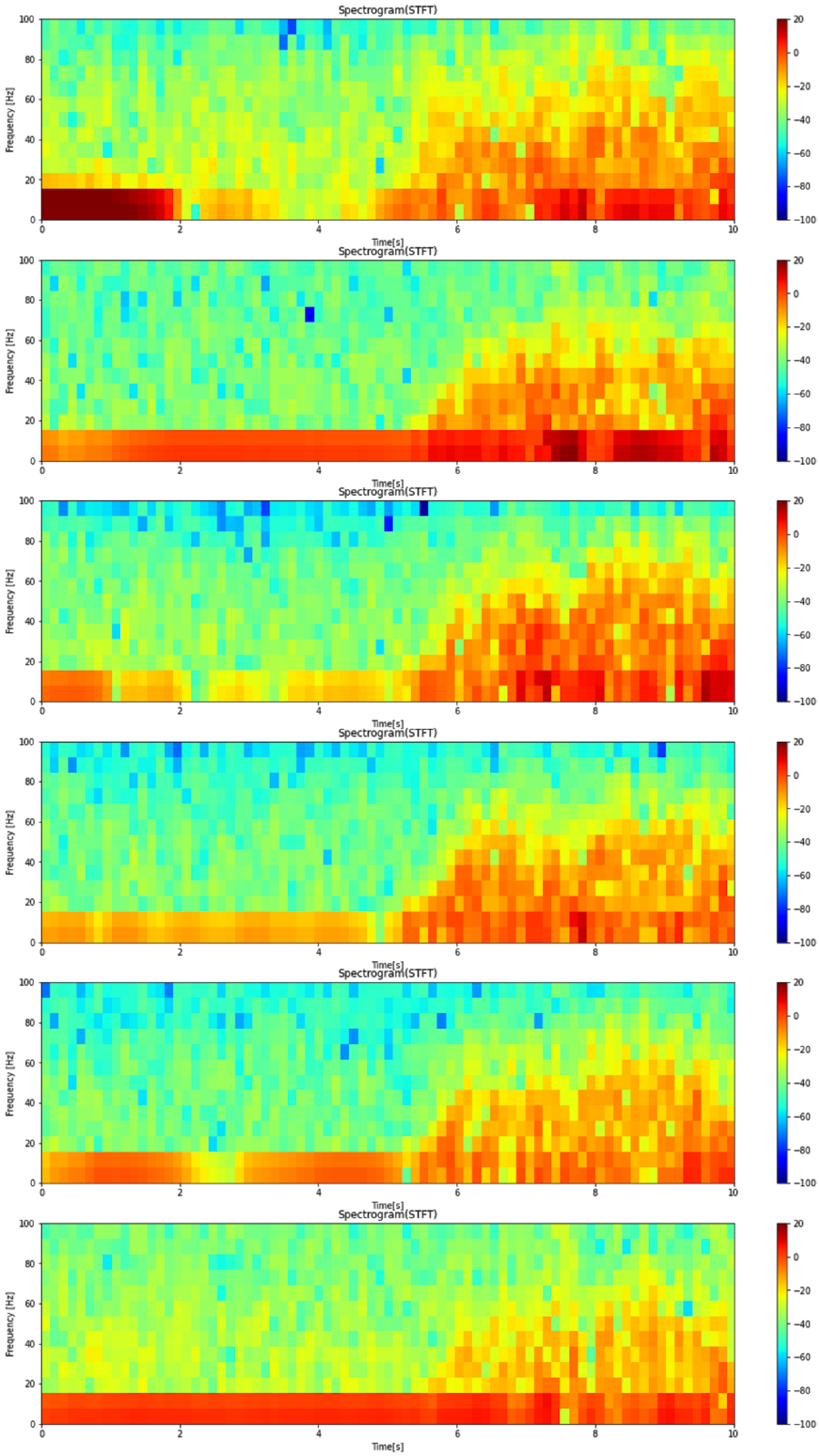


Figure 3.9: Frequency changed over time for run.18

3.3.4 Magnitude spectrum

Figure 3.10-3.12: It is a graph for plotting magnitude vs. frequency.

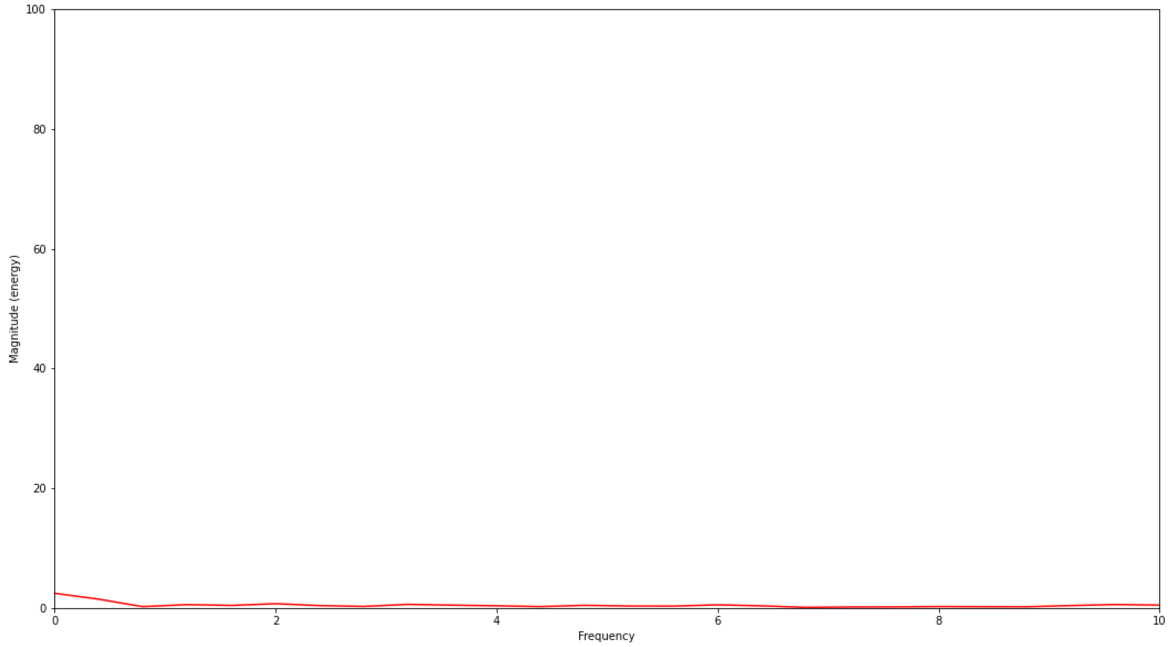


Figure 3.10: Showing the change for run.20

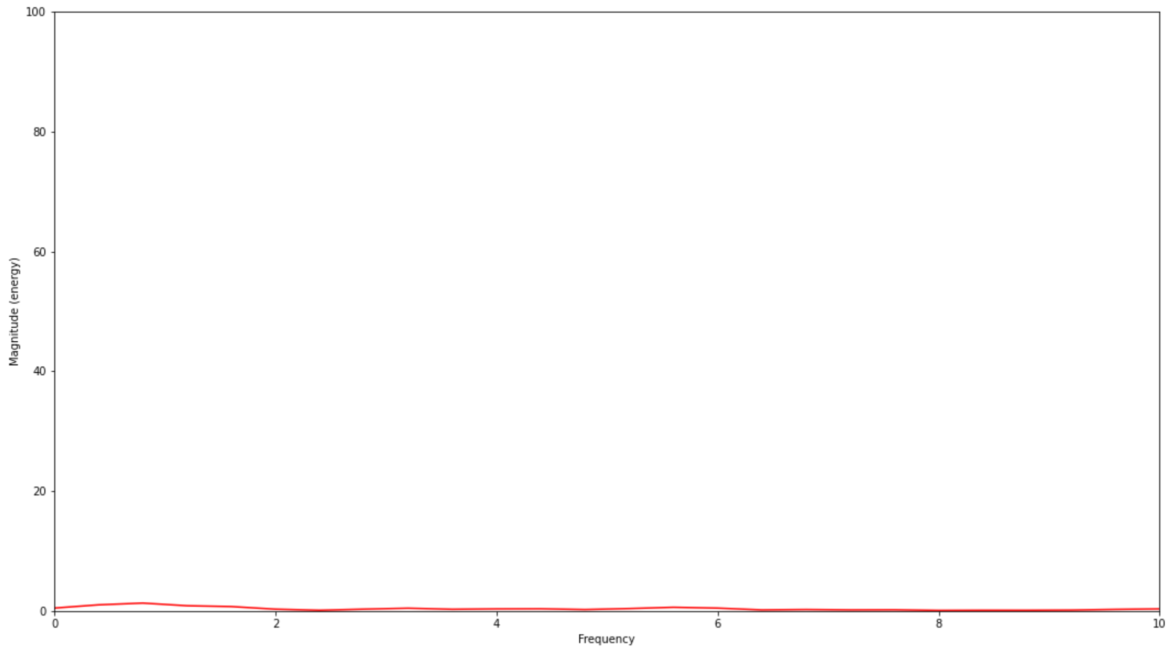


Figure 3.11: Showing the change for run.19

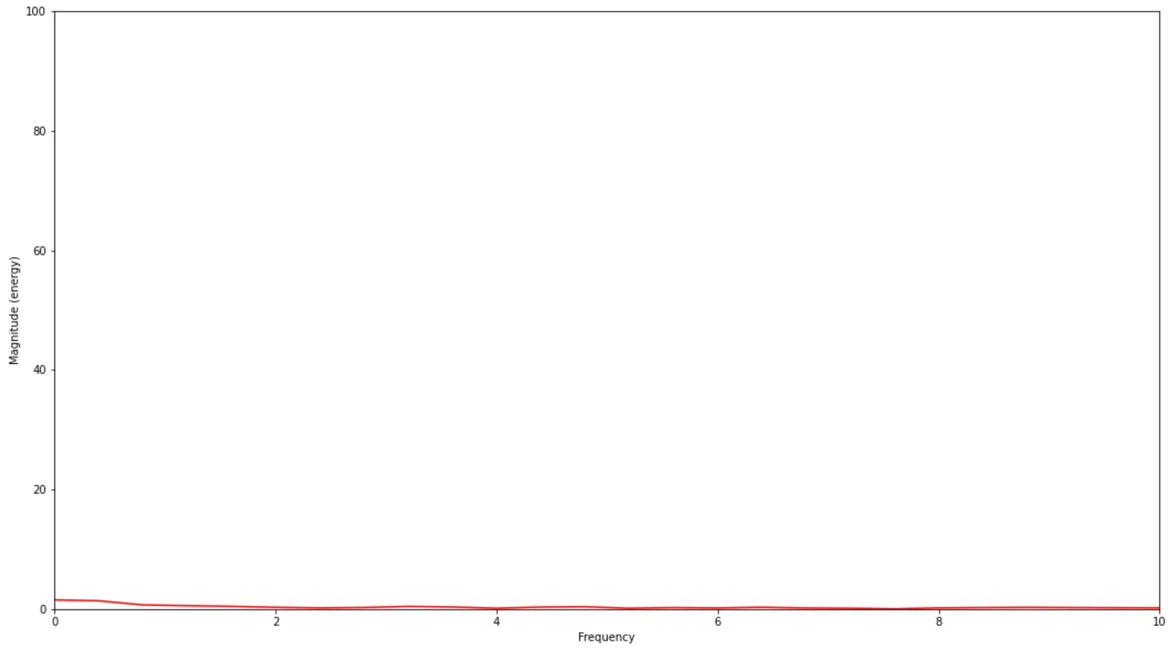


Figure 3.12: Showing the change for run.18

We can see the similarities between all the plots and figures. Which testify that, our test is close to accurate to detect movement of object using Wi-Fi frequency.

3.3.5 Data flow for present

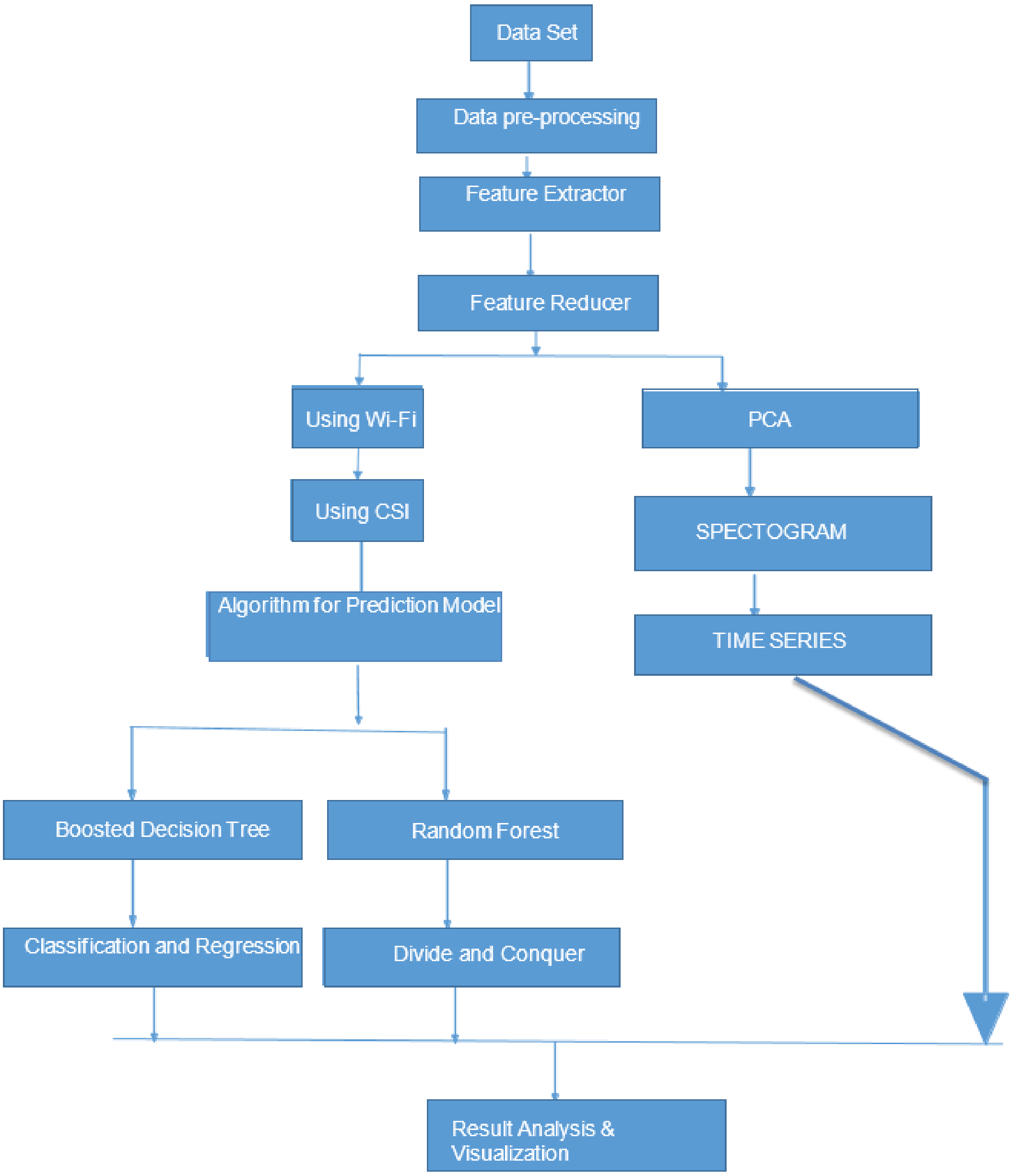


Figure 3.13: Data Flow

DATA FLOW WORKS FOR PRESENT

Here, we can see a practical study of object movement detection and presence detection using Wireless Fidelity (Wi-Fi) and Time series analysis which are under supervised learning. We have used a secondary dataset where Wi-Fi devices has been used to capture fine-grained movements to recognize object's or human's activity. As every person and objects are different, their walking datasets will be also different. So it will create distinctive pattern or variation which will be received by Wi-Fi through Channel State Information (CSI). To profile human or objects movement or presence using CSI, we use signal processing techniques to generate spectrograms from CSI measurements and prediction algorithm models. The CSI data can only illustrate the signal changing patterns, where the signal reflections of different objects parts are mixed together. When a subject is walking, its body parts (such as legs, arms and torso) have different moving speeds, and the signals reflected by different subject's parts have different energy. In using times Series analysis, first we preset some data as ideal input. We trained the data. After training we take out test data and analyze it with principle component analysis and spectrogram through time series analysis. This is how we can detect the objects presence.

Some important steps including pre-processing, feature extraction and feature reducer are described below:

1. DATA SET:
For datasets we can collect data in two process.
i.Primary data collection and
ii.Secondary data collection.

Primary Data:
Primary data is a real time data that is collected by first hand process. It is always specific to the researchers needs and collected through survey, experiments, observations, personal interview etc. It's a long term process and a bit costly too [27].

Secondary data:
Secondary data is a past data that is already collected by someone earlier. In this way we can find the data quickly and easily. It's also cost effective as it is gathered by government publications, books, journals, articles, websites etc. [27].

As this is very crucial time for us for corona virus, we done our thesis mostly dependent on secondary dataset.

2. DATA PRE-PROCESSING:

Doing preprocessing for datasets, we can find out the best and improved datasets for us that can help us to fulfill our aim. In real life world data is noisy, inconsistent, incomplete, error full etc.

By preprocessing, we can extract the clean and original data. There are some steps for preprocessing.

i. Importing Libraries
ii. Importing datasets.
iii. Missing value checking,
iv. Categorical value observation
v. Data set splitting as training and testing
vi. Scalling feature.

By following those steps we can do preprocessing for machine learning [26]. It is really helpful for big size image, bulk amount of datasets etc.

3. FEATURE EXTRACTOR:

By identifying important feature for our datasets, we can extract those particular featured data. We can use it for clustering, classification, detection, recognition etc. [30] We use pattern recognition in our paper for identifying objects presence.

4. FEATURE REDUCER:

To reduce aspects and variety of feature, we can reduce feature by this reducer. We can only take our essential parts. This will help us to do our work efficiently and reduce wastes of our time also. We also used feature extractor for reducing our datasets feature. Like walking, running etc.

After doing feature reduction, we tried to implement CSI using Wi-Fi for detecting objects presence through generating different type of pattern. Here 3 algorithms can be used. 1.Boosted decision tree, 2.Random forest algorithm and 3.Linear regression algorithm. Here with supervised learning, we train our expected datasets and test with our test datasets. We can compare the differences between this two over time as well as can detect the presence of a subject also.

3.4 Initial planning

Detection of human presence technology is proceeding effectively through various recognition systems i.e. face recognition, fingerprints detection, gait recognition, iris, hand geometry etc. we were working on detecting human presence by gait recognition using Wi-Fi and pyro electric sensors. Biometric gait recognition system works on identifying a person by their walking style. Gait is unique and differs from person to person. In addition, if we can use this presented system to identify human presence it will be very efficient in security advancement. It can be used to secure home, office, hotel, bank, and other organizations by identifying unauthorized people and access. We will attempt to handle this security issue examining Radio Frequency (RF) signals on the basis of changes of the Wi-Fi spectrum. Moreover, pyro electric sensors will be used to recognize gait. As every individual's walking style, body shape is unique in relation to each other, each individual will affect various changes when they come to the Wi-Fi range and we will see unique frequency. Thusly, it is possible to recognize an individual remarkably by experimenting these frequency using Channel State Information (CSI) data. We will use image processing technology which is a part of artificial intelligence (AI) to access the data. Furthermore, Pyroelectric Infrared (PIR) Sensors will be used to collect data of human movement and detect unwanted human presence for security purpose. PIR sensors will also help to detect the direction and speed of movement.

In our research, we were working on implementing biometric gait recognition with Wi-Fi and Pyroelectric Infrared Sensors approach for security system, along with various algorithms, signals, visions, calculations in order to ensure the safety. Crimes are daily issues for now–a-days. We will attempt to make a system that will be able to detect any type of doubtful person appearing in the systems by scanning the data and then matching or analyzing the data. It should be able to distinguish the human presence of who are suspicious and who are not, and alert through the security system.

Furthermore, we were focused on implementing a hybrid technology combining the Wi-Fi signals and Pyro electric infrared sensors (PIR). Camera visions are the most common technique in this method. But, as we wanted to work our system at lowlight and identify individuals matching data, we decided to take advantage by a combined system. We can be able to detect suspicious human's physical behavior within few seconds that can make us alert about the situation. Our objective was to find or establish a system for security purpose which will serve the law enforcement authority, private and public goods and help them to protect system environment.

If we wanted to see in the simplest term, biometric gait recognition means identifying a person's appearance by collecting the related data over necessary technologies. For this, we are using Wi-Fi signals to collect data. Wi-Fi device and resources can be used in low cost, and availability is very high as now almost everywhere every person is using Wi-Fi for the internet. There is router everywhere.

A router can send signals continuously, which we can receive through other devices like computer, laptop or mobile. Wi-Fi signals gives us unique gait information and reflect every human body differently. This helps us to recognize the different gait patterns and detect the presence. As Wi-Fi signals collect unique biometrics information, it gives us human recognition uniquely.

This technology can be used in office or at home and many other places. For example, let us see an office scenario. In an office, gait data of every employee can be collected. When a person enters the office Wi-Fi signals compare the new data with the old collected data. If it does not match, that means the person is not from the employees of office. Now it will alert the security about the unknown entrance immediately and identify the person's movement. We can also use it in public places like shopping malls or in shops individually. For example, if we can record one person's common and normal movements position such as walking style, hand's position style etc. So if any thief or unwanted person entered in that shop, who's activity like walking style or hands position is not matched with our pre-defined pattern, then it will give us an alert. So that we can easily suspect that specific person and can notice his/her activity. User can also be always alerted and prevent any attack or vicious activity. It will improve the security systems in an efficient way. Though using Wi-Fi signals is really efficient. But sometimes there can be many problems to run it perfectly. For example, electricity is must to run a router. If the electricity is gone, the imposter can take benefit from it and achieve their purpose. To prevent these issues, we have to search for alternatives and apply them to work it effectively.

Beside this, personal identification and authentication of an individual is important, because of this, biometric technologies across the world developing rapidly. Biometric recognition identifies physical and behavioral traits which are always present with a person and there is no way of faking it. Thus it plays an important role in identification. There are different types of biometric systems, such as DNA matching, face, iris, retina, fingerprint, hand geometry, finger geometry recognition etc. But all these systems have some drawbacks as well. For instance, most of these systems are not capable of identifying people wearing sunglasses, masks, helmets, gloves etc.

Besides, a person undergoing such recognition procedures has to be in close contact with the system and mentally prepared beforehand. In order to overcome such difficulties biometric gait recognition technique is used. Gait biometrics record stride patterns, walking speed, posture combined with characteristic motion etc. with the help of video imaging then the mapped data is transformed into a mathematical equation. It emphasizes on the gait cycle of human which is considered unique for every individual. Gait biometrics can quickly identify people from a distance even under low light where other biometrics fail to do so. In a crowd surveillance images can be obtained performing gait recognition without the cooperation and awareness of the person under observation. Due to all these features gait biometrics differs from the rest of the biometric recognition techniques. With the increasing rate of crimes around the world the demand for biometric gait recognition is also emerging. Different organizations like Federal Bureau of Investigations (FBI) are using biometrics in criminal investigations and so on [35]. Sometimes it is seen that due to lack of efficient recognition techniques criminals and law offenders are walking in broad daylight under the nose of law enforcement authority. Biometric gait recognition provides a solution to this problem to a large extent. According to the Critical review of the use and scientific basis of forensic gait analysis, forensic gait analysis has been used as supportive evidence in criminal cases in the United Kingdom for more than 15 years and in Denmark for more than 10 years. In the Netherlands, gait analysis has been performed rarely in the past 20 years However, two recent criminal cases renewed interest in the topic in the Netherlands [19] Since gait biometrics has become quite popular over the recent years, researches has been already conducted to identify criminals. In this paper, we were concerned about detecting presence and movement of human before conducting a biometric gait recognition. Presence of human can be detected using different methods such as Wi-Fi signals, infrared sensors etc. But individually these type of methods do not give satisfactory results. For this reason, we have come up with an idea of combining these two methods in order to get a better result. In this paper, an approach towards biometric gait recognition to identify vicious person and movement using Wi-Fi signals and Pyro Electric Infrared (PIR) Sensors has been introduced. The hardware's required for the system are personal computer, router, Pyro electric Infrared (PIR) Sensors and the software is MATLAB. In our system, we will use CASIA gait database that has been provided by The Institute of Automation, Chinese Academy of Sciences (CASIA) in order to promote research [33]. Here Vision based Pyro Electric Infrared (PIR) Sensors are used which can detect presence along with distance of the subject from the sensor, direction and speed of movement, basic gaits etc. Wi-Fi emits Radio Frequency (RF) signals continuously and if a person arrives within the range, he/she propagates through the signals. As a result, upon observation changes in the spectrum can be found. After examining the propagation with Channel State Information (CSI) using signal processing techniques basic human gaits can be identified. A Boosted decision tree is used for taking decision and splitting the dataset. For error detection Random Forest Algorithm has been applied with formulas. Then Silhouette Algorithm is used to detect motion. Three steps are involved in the development of Image-Processing Algorithm such as pre-processing, feature extraction and feature reducer. Pre-processing enhance features of image, feature extraction is used to reduce dimensionality and feature reduction is used to work with essentials reducing aspect. Finally, we tried to visualize and analyze our result from the algorithms.

3.5 Initial Work Flow

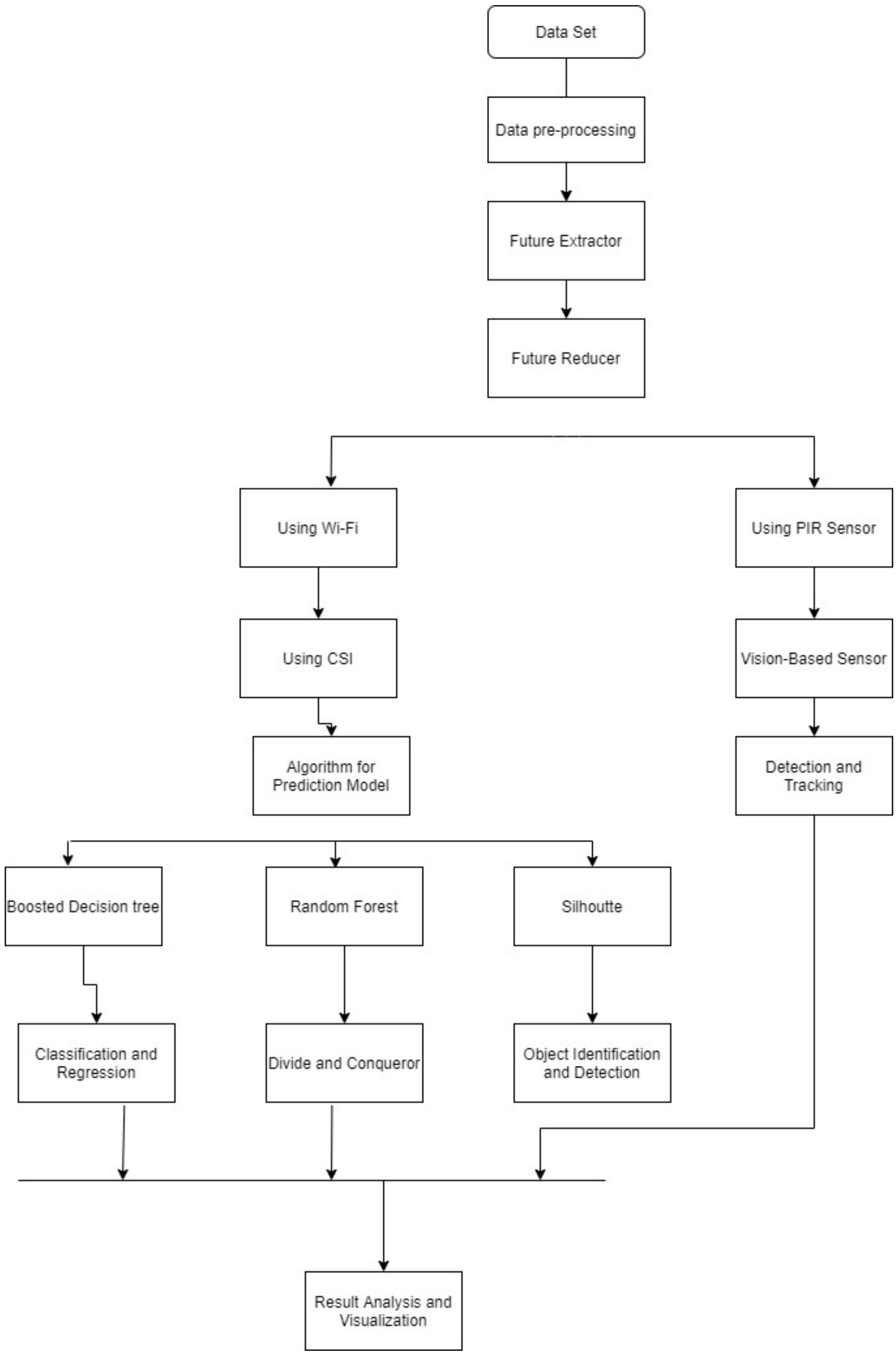


Figure 3.14: Initial Work Flow

3.6 Initial Work Flow Explanation

Here, we can see a pragmatic study of human movement detection and identification using Wireless Fidelity (Wi-Fi) and the Pyroelectric Infrared (PIR) Sensors.

We have developed a set of data where we use commercial Wi-Fi devices to capture fine-grained gait patterns to recognize humans. As every person is different, their walking pattern will be also different. So it will create distinctive pattern or variation which will be received by Wi-Fi through Channel State Information (CSI). To profile human movement using CSI, we use signal processing techniques to generate spectrograms from CSI measurements and prediction algorithm models. The CSI data can only illustrate the signal changing patterns, where the signal reflections of different body parts are mixed together. When a subject is walking, its body parts (such as legs, arms and torso) have different moving speeds, and the signals reflected by different body parts have different energy. The Boosted Decision tree which is used for predictive algorithms takes decision at each point and splits the dataset. When we use a huge balk number of datasets with a lot of features, this algorithm can find the best features among them by splitting. It works very efficiently. [20]. we used the Random Forest algorithm, which is used with certain formulas in a sequential way to decide whether the malware is benign or not. It can be used for both classification and regression problem and very easy to use it [21]. To be able to have a more reliable and fast algorithm for background estimation based human motion detection and analysis, silhouette algorithm is used. The method combines periodic motion estimation with static symmetry analysis of the silhouettes of a person in each frame of the sequence [22]. In using PIR sensors, which is the Pyroelectric Infrared Sensor we use the vision-based sensor. The vision-based sensors mainly consider position, speed, direction, shape and size (i.e., the number of pixels in cameras) as the principal context for identifying the users and understanding their activities. [31] The steps involved in the development of the image-processing algorithm are preprocessing, feature extraction and feature reducer. The aim of preprocessing data is to find out the best data's or improved data that would help us to reach to our goals. This procedure is helpful when it's about large image sizes. Feature representation can be reduced for quickly complete tasks like image matching and retrieval. Similarly, the feature reducer is mainly done to reduce the aspect and to work with the essentials. This helps with technicality and time. Then the results from these algorithms are then analyzed and visualized.

Chapter 4

Model selection and Result analysis

4.1 Machine learning

Artificial Intelligence is one of the most important aspects of our life. We are using it in our daily life. In AI, machine learning is most significant algorithm that we used for problem solving. A set of instructions that we use to solve problems is called algorithm [38]. Machine learning can automatically learn and improve by itself without being explicitly commanded [16]. We can make our decision or can predict things depending upon machine learnings algorithm's trained data or sample data [38], [39]. In many field it used widely. In our paper, we also use machine learning algorithm for training and testing our datasets.

4.2 Supervised learning

We can categorize machine learning as 1.Supervised learning, 2.Unsupervised learning, 3.Semi-supervised learning and 4.Reinforcement learning [38], [39], [41]. Among them, supervised learning is the subcategory of machine learning and artificial intelligence [29].

With supervised learning, we can train a dataset with algorithm. After that we can take a new dataset as input and supervised algorithm then analyze it and give us correct result or such an output that helps us to predict results for future. For example, if we take a basket of fruit, we have to train the machine by all fruits shape, size, colors etc. Then if we give an apple as input, the machine will be able to identify the apple correctly by supervised algorithm [28]. Two types of algorithm we can use here.1. classification and 2 regressions. In our thesis, We mainly focus on supervised learning algorithm to find our expected outcomes. We first set datasets for input as ideal from different sources. Then we test it one by one through our sample input dataset. This is how we can detect the presence of an object in our work.

4.3 Decision Tree Algorithm

Decision tree algorithm works under supervised learning. It works as a tree. It makes trained datasets tree with leaf node and attributes. Then through some calculation it can match with test datasets. So in our project we also can use it. We can first make a decision tree according to its category. For example, if one object or person are walking, running there will be variation in their pattern. So according to their variations characteristics, we can make a decision tree. We can use Boolean functions also for discrete attributes. Statistical methods can be used for ordering.

Decision tree works as total of product which is also known as disjunctive normal. [20]. Here attributes finding is the most challenging tasks. For this, decision tree algorithm uses attribute selection measure.

For attribute measure selection,
1.Information Gain and
2.Gini index is used usually [20]
This is how we can use decision tree in our code.

4.4 Random Forest Algorithm

This also falls under supervised learning. It works on bagging method [21]. The bagging model represent the combination of model which is learned. This can increase results overall. Here we also make decision tree but multiply. We make multiple decision trees. After making tress we join those tree and can find more reliable and right answer. We can use it for classifying our datasets. Then we can generate multiple tress. This is easy and more accurate process than decision tree.

4.5 Linear Regression Algorithm

Linear Regression algorithm is a machine learning and statistical approach widely used for prediction based analysis. It is mainly based on supervised learning. The algorithm can predict continuous values such as weather, age, salary, price etc. In this approach the model is simple which makes it easier to understand than other machine learning approaches. [37] The algorithm is used to find the rate of change of dependent variable with respect to independent variable. The equation for linear regression is shown below:

$$y = a_0 + a_1 x + \varepsilon \tag{4.1}$$

Here,
y= Dependent variable
x= Independent variable
a_0= Intercept
a_1= Coefficient
ε = Error

In our model we have timestamp as dependent variable. We have determined the rate of change of time with respect to amplitude and phase. Since we have more than one independent variable so it is Multiple Linear Regression. The best fit line is determined by minimizing the error between actual and predicted value.

Also, we have determined the actual performance of our model with the help of cost function. [36] Mean Squared Root (MSE) cost function is the average squared error between actual and predicted values. The equation is given below:

$$MSE = \sum_{i=1}^{n} (y_i - (W_1 x_1 + W_2 x_2 + W_3 x_3))^2 \tag{4.2}$$

Here,
n= Number of observation considered
y_i= Actual value
$W_1 x_1 + W_2 x_2 + W_3 x_3$= Predicted value

We have determined gradient of cost function in order to reduce MSE with the help of Gradient descent. One of the property of linear regression model is to modifying line coefficients by minimizing cost function. [36] This is also done with the help of Gradient descent by selecting values randomly and modify it in order to get minimum cost function. Formula of Gradient descent using chain rule is given below:

$$f'(W_1) = -x_1(y - (W_1x_1 + W_2x_2 + W_3x_3)) \tag{4.3}$$
$$f'(W_2) = -x_2(y - (W_1x_1 + W_2x_2 + W_3x_3)) \tag{4.4}$$
$$f'(W_3) = -x_3(y - (W_1x_1 + W_2x_2 + W_3x_3)) \tag{4.5}$$

Again, [15] the Gradient descent is improved by using matrix multiplication. It has been done to avoid repetitive data. The formula is given below:

$$gradient = -X(targets - predictions) \tag{4.6}$$

Here,
X= Feature matrix

Finally, we have determined how good the model fits with the help of R-squared method. [37] We have received high value for R- square which means our predicted value and actual value has very small difference representing a good model. Formula for R-squared method is given below:

$$R - squared = \frac{Explainedvariation}{Totalvariation} \tag{4.7}$$

4.6 Time Series

Time series analysis is an important part of machine learning to understand the present behavior of a model and to predict how it will behave in future. Time series states that collection of data is done at a definite interval of time. It is helpful to understand various pattern in a model, training a model, predicting future patterns from past patterns etc. [42] Here data is categorized into three types:

- *Times series data:* Data points based on time in sequential manner.
- *Cross-sectional data:* Collection of data of single or multiple variable at same time.
- *Pooled data:* Time series data and cross-sectional data.

In our dataset we have three types of data such as timestamp, amplitude and phase. For this reason we have used time series to observe the changes in data over time. For instance, for one of our walking data we find out the specific date, month, and year of data collection in date-time format. It is observed that all of the data were collected on the same date after a specific period of time.

Since, our dataset is huge (19994 rows and 185 columns) the values for first eight rows after time series application are given below:

Timestamp	Amplitude	Amplitude.1	Amplitude.2	Amplitude.3	Amplitude.4	Amplitude.5	Amplitude.6	Amplitude.7	Amplitude.8	Amplitude.9	Amplitude.10
1970-01-01 00:00:00.000000011	11.435	13.449	13.752	14.721	14.721	14.550	14.226	14.948	15.225	14.226	15.225
1970-01-01 00:00:00.000000011	11.435	13.449	13.752	14.721	14.721	14.550	14.226	14.948	15.225	14.226	15.225
1970-01-01 00:00:00.000000011	12.398	12.398	14.471	15.608	14.718	15.145	15.063	15.960	15.776	13.786	15.408
1970-01-01 00:00:00.000000011	12.398	12.398	14.471	15.608	14.718	15.145	15.063	15.960	15.776	13.786	15.408
1970-01-01 00:00:00.000000011	12.398	12.398	14.471	15.608	14.718	15.145	15.063	15.960	15.776	13.786	15.408
1970-01-01 00:00:00.000000011	10.950	13.267	14.799	14.236	14.464	14.943	15.245	15.375	16.107	14.799	15.952
1970-01-01 00:00:00.000000011	10.950	13.267	14.799	14.236	14.464	14.943	15.245	15.375	16.107	14.799	15.952
1970-01-01 00:00:00.000000011	10.950	13.267	14.799	14.236	14.464	14.943	15.245	15.375	16.107	14.799	15.952

8 rows × 185 columns

Figure 4.1: Time series application on walking data (row: 1-8, column: 1-11)

Phase.79	Phase.80	Phase.81	Phase.82	Phase.83	Phase.84	Phase.85	Phase.86	Phase.87	Phase.88	Phase.89	Hour	Minute	Second	Microsecond	Nanosecond
13.056	13.020	13.073	13.211	13.188	13.221	13.165	12.949	12.740	12.534	12.754	0	0	0	0	11
13.056	13.020	13.073	13.211	13.188	13.221	13.165	12.949	12.740	12.534	12.754	0	0	0	0	11
14.070	14.087	14.183	14.289	14.288	14.318	14.223	14.054	13.817	13.646	13.907	0	0	0	0	11
14.070	14.087	14.183	14.289	14.288	14.318	14.223	14.054	13.817	13.646	13.907	0	0	0	0	11
14.070	14.087	14.183	14.289	14.288	14.318	14.223	14.054	13.817	13.646	13.907	0	0	0	0	11
14.328	14.333	14.371	14.472	14.435	14.463	14.431	14.239	14.022	13.798	14.061	0	0	0	0	11
14.328	14.333	14.371	14.472	14.435	14.463	14.431	14.239	14.022	13.798	14.061	0	0	0	0	11
14.328	14.333	14.371	14.472	14.435	14.463	14.431	14.239	14.022	13.798	14.061	0	0	0	0	11

Figure 4.2: Time series application on walking data (row: 1-8, column: 170-185)

From this it is seen that data has been collected after 11 nanosecond each time which validates the concept of time series.

4.7 ADF (Augmented Dickey-Fuller) Test

Time series analysis depicts two types of results, stationary and non-stationary. A stationary series does not change over time and external factors cannot alter its pattern. Whereas a non-stationary series is affected by external factors and changes over time. Moreover, stationarity represents normal distribution of series and the mean and variance are constant. Augmented Dickey-Fuller Test is used to understand the stationary and non-stationary nature of a time series. ADF checks whether a unit root is present in a time series or not. [34] ADF is the augmented version of Dickey-Fuller Test which has higher regression process. The equation of ADF is shown below:

$$\Delta yt = \alpha + \beta t + \gamma yt - 1 + \delta 1 \Delta yt - 1 + \delta 2 \Delta yt - 2 + \dots \tag{4.8}$$

ADF test is done with the help of statsmodel package.[40] From the test we receive:

- *Test statistic value*
- *p-value*
- *Critical value at different percentages*
- *Lags used in observation*

Whether a series is stationary or not is decided based on p-value. [25] ADF consists of null and alternate hypothesis.

- *Null hypothesis:* Unit root is present in series and it is non-stationary
- *Alternate hypothesis:* Unit root is not present in series and it is stationary

Mean ad Standard Deviation is calculated to show that the results are constant and they are not changing over time. So, we can say unit root is absent here and data is stationary. Mean and Standard deviation of walk data is given below:

```
# Mean and Standard Deviation Calculation
mean=indexedDataset.rolling(window=365).mean()
std=indexedDataset.rolling(window=365).std()
print(mean,std)
```

```
                               Amplitude  Amplitude.1  ...  Microsecond  Nanosecond
Timestamp                                              ...
1970-01-01 00:00:00.000000016        NaN          NaN  ...          NaN         NaN
1970-01-01 00:00:00.000000016        NaN          NaN  ...          NaN         NaN
1970-01-01 00:00:00.000000016        NaN          NaN  ...          NaN         NaN
1970-01-01 00:00:00.000000016        NaN          NaN  ...          NaN         NaN
1970-01-01 00:00:00.000000016        NaN          NaN  ...          NaN         NaN
...                                  ...          ...  ...          ...         ...
1970-01-01 00:00:00.000000036   9.662206    11.049666  ...          0.0        36.0
1970-01-01 00:00:00.000000036   9.665552    11.051113  ...          0.0        36.0
1970-01-01 00:00:00.000000036   9.666695    11.054655  ...          0.0        36.0
1970-01-01 00:00:00.000000036   9.669201    11.050862  ...          0.0        36.0
1970-01-01 00:00:00.000000036   9.667814    11.046938  ...          0.0        36.0

[19995 rows x 185 columns]                               Amplitude  Amplitude.1  ...  Microsecond  Nanosecond
Timestamp                                              ...
1970-01-01 00:00:00.000000016        NaN          NaN  ...          NaN         NaN
1970-01-01 00:00:00.000000016        NaN          NaN  ...          NaN         NaN
1970-01-01 00:00:00.000000016        NaN          NaN  ...          NaN         NaN
1970-01-01 00:00:00.000000016        NaN          NaN  ...          NaN         NaN
1970-01-01 00:00:00.000000016        NaN          NaN  ...          NaN         NaN
...                                  ...          ...  ...          ...         ...
1970-01-01 00:00:00.000000036   1.036382     0.802488  ...          0.0     0.00001
1970-01-01 00:00:00.000000036   1.036189     0.804686  ...          0.0     0.00001
1970-01-01 00:00:00.000000036   1.036565     0.808353  ...          0.0     0.00001
1970-01-01 00:00:00.000000036   1.035967     0.811827  ...          0.0     0.00001
1970-01-01 00:00:00.000000036   1.037253     0.815503  ...          0.0     0.00001

[19995 rows x 185 columns]
```

Figure 4.3: Mean and Standard Deviation of walk data no 18 (train)

ADF Test is performed on three types of dataset which include walk, run and standup data of a subject. The results of ADF test are given below:

	Walk.18.train	Walk.19.train	Walk.20.test
Test Statistic	-5.406939	-6.057467e+00	-5.141050
p-value	0.000003	1.235270e-07	0.000012
No. of Observations Used	46	44	46
No. of Lags Used	19948	19949	19948
Critical Value (1%)	-3.430678	-3.430678e+00	-3.430678
Critical Value (5%)	-2.861685	-2.861685e+00	-2.861685
Critical Value (10%)	-2.566847	-2.566847e+00	-2.566847

Table 4.1: Results of ADF test on walk data

	Run.18.train	Run.19.train	Run.20.test
Test Statistic	-4.633695	-6.169251e+00	-5.476064
p-value	0.000112	6.868921e-08	0.000002
No. of Observations Used	46	37	42
No. of Lags Used	19950	19959	19954
Critical Value (1%)	-3.430678	-3.430678e+00	-3.430678
Critical Value (5%)	-2.861685	-2.861685e+00	-2.861685
Critical Value (10%)	-2.566847	-2.566847e+00	-2.566847

Table 4.2: Results of ADF test on run data

	Standup.18.train	Standup.19.train	Standup.20.test
Test Statistic	-3.521198	-3.876179	-4.490680
p-value	0.018790	0.002217	0.000204
No. of Observations Used	37	24	43
No. of Lags Used	19956	19969	19951
Critical Value (1%)	-3.430678	-3.430678	-3.430678
Critical Value (5%)	-2.861685	-2.861685	-2.861685
Critical Value (10%)	-2.566847	-2.566847	-2.566847

Table 4.3: Results of ADF test on standup data

The results show p-value <= 0.5 in all cases, which is alternate hypothesis. So, there is no unit root in data and the series or data is stationary. Furthermore, critical values at different percentages are also greater than Test Statistic values which is acceptable.

4.8 Results and Analysis

Using supervised learning approach we have trained our data first. Then we tested the data to determine how accurately the model can predict the result. If our train and test data depicts similar result (approximately) we have reached a conclusion that our supervised approach is successful.

In order to do so, we have selected three datasets for each type such as walk, run and standup. Among these three datasets of each type we have selected two for training and one for testing purpose. Each of these selected dataset has 185 columns and 19994 rows approximately. The columns contain timestamp, amplitude and phase data of on object while walking, running and standup.

Firstly, we have compared the p-value obtained from Augmented Dickey-Fuller Test (ADF) for two training and testing datasets. From our observation, we can see that p-values for each dataset is less than 0.5 which means data is stationary and there is no null point in our dataset. In Figure 4.5 we can see that p-value of run.train.19 and run.train.20 overlaps which means they are equal (almost) so we can predict they will give similar result. Comparison between p-values of walk, run and standup data is given below:

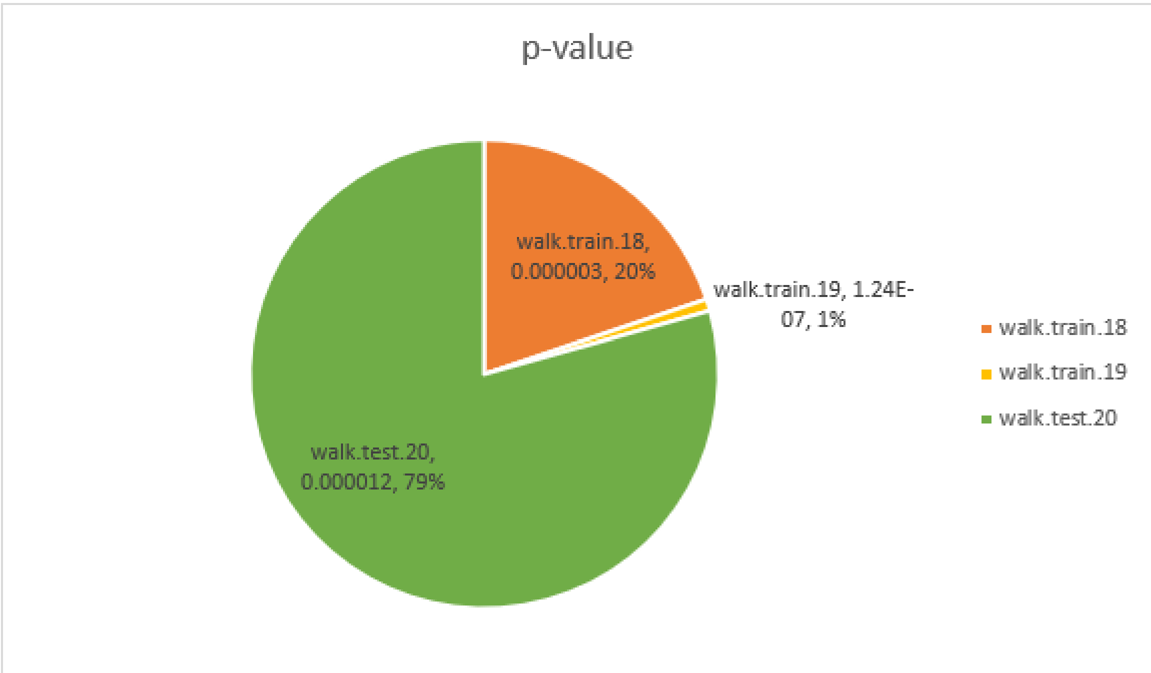


Figure 4.4: p-value comparison for walking data (train and test)

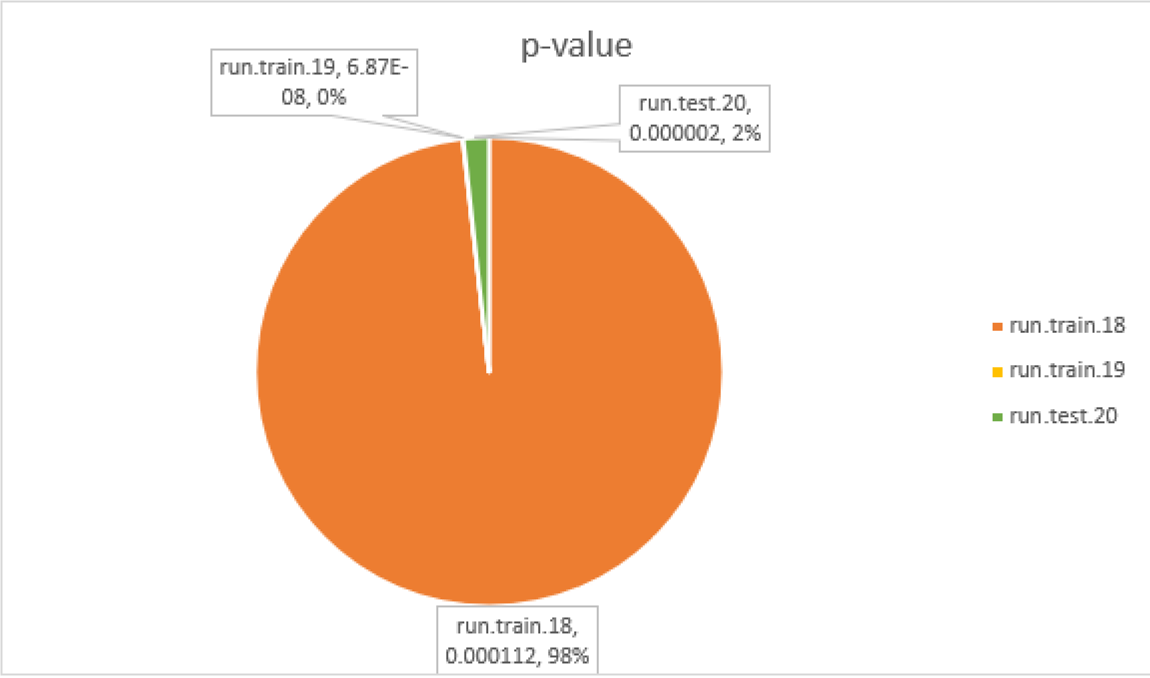


Figure 4.5: p-value comparison for running data (train and test)

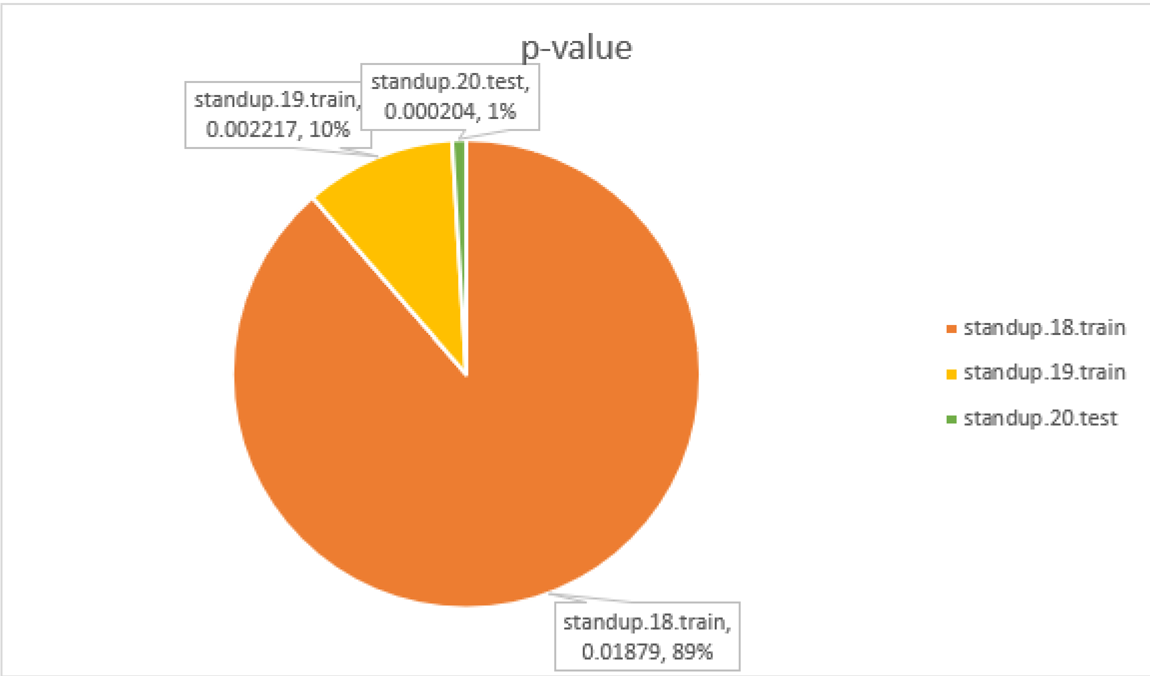


Figure 4.6: p-value comparison for standup data (train and test)

Furthermore, we have compared the graph obtained from Principal Component Analysis to detect if the graphs obtained from train and test data are similar are not. If they are similar to some extent we have decided that the graphs are of same categories i.e. either of walk, run or standup data.

Figure 4.7, 4.8 and 4.9 shows walking features for first and second Principle Component of a subject under observation. We have used walking data no 18 and 19 for training our model. Later we have compared outputs from training sets with that of walking data no 20 which is a testing set. From our observation we can see outputs of three datasets exhibit similar pattern or feature. So we can say these are outputs from data which belongs to similar categories of movements i.e. all three outputs are of walking data. We have followed similar approach for run and standup data. Figures for different kind of movements are listed below:

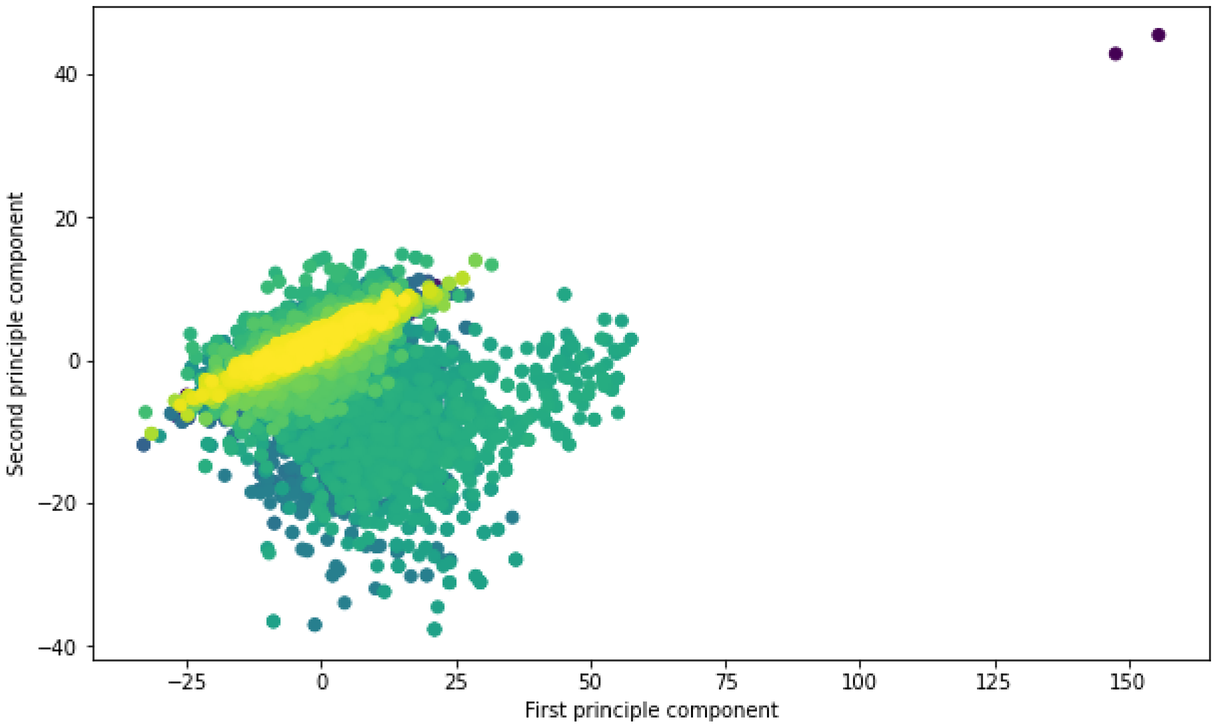


Figure 4.7: PCA of first and second component for walking data no 18 (train)

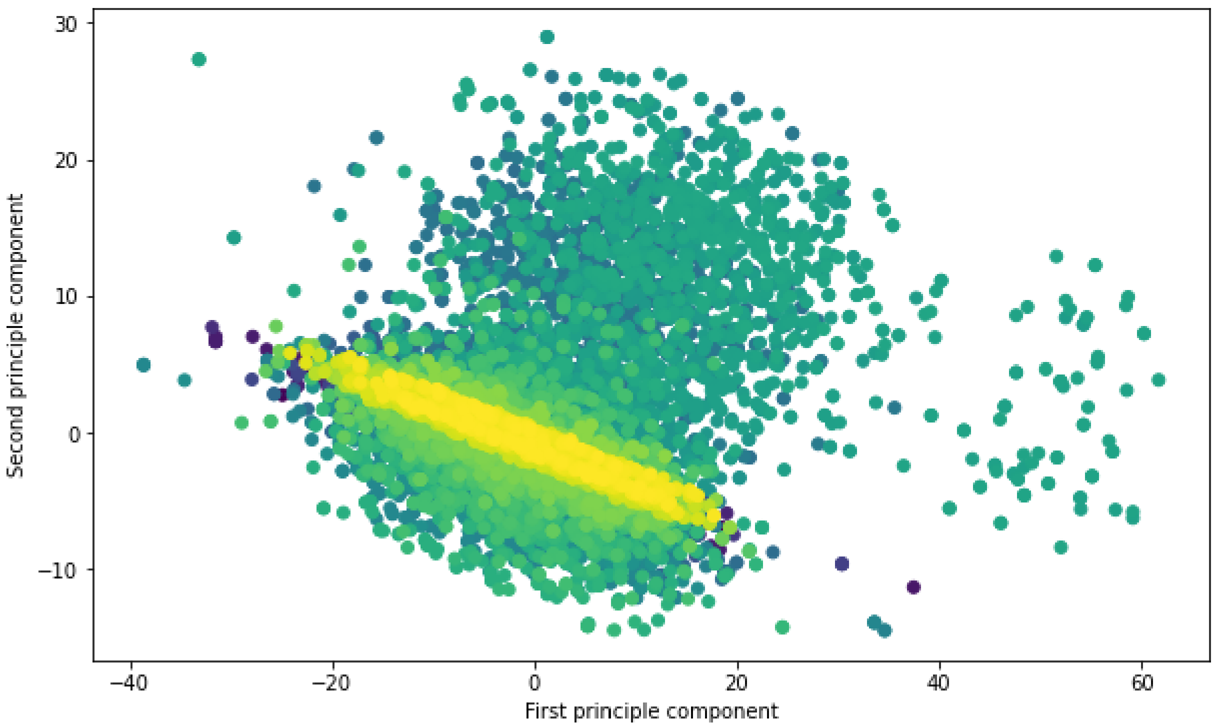


Figure 4.8: PCA of first and second component for walking data no 19 (train)

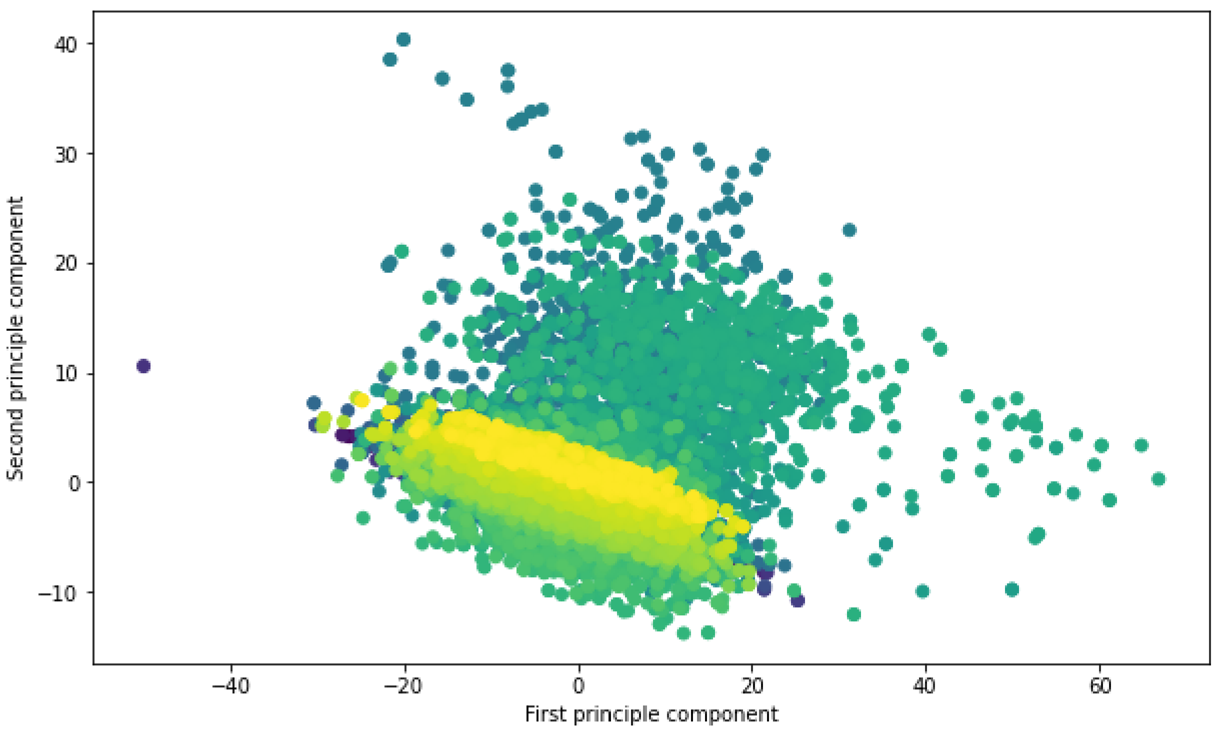


Figure 4.9: PCA of first and second component for walking data no 20 (test)

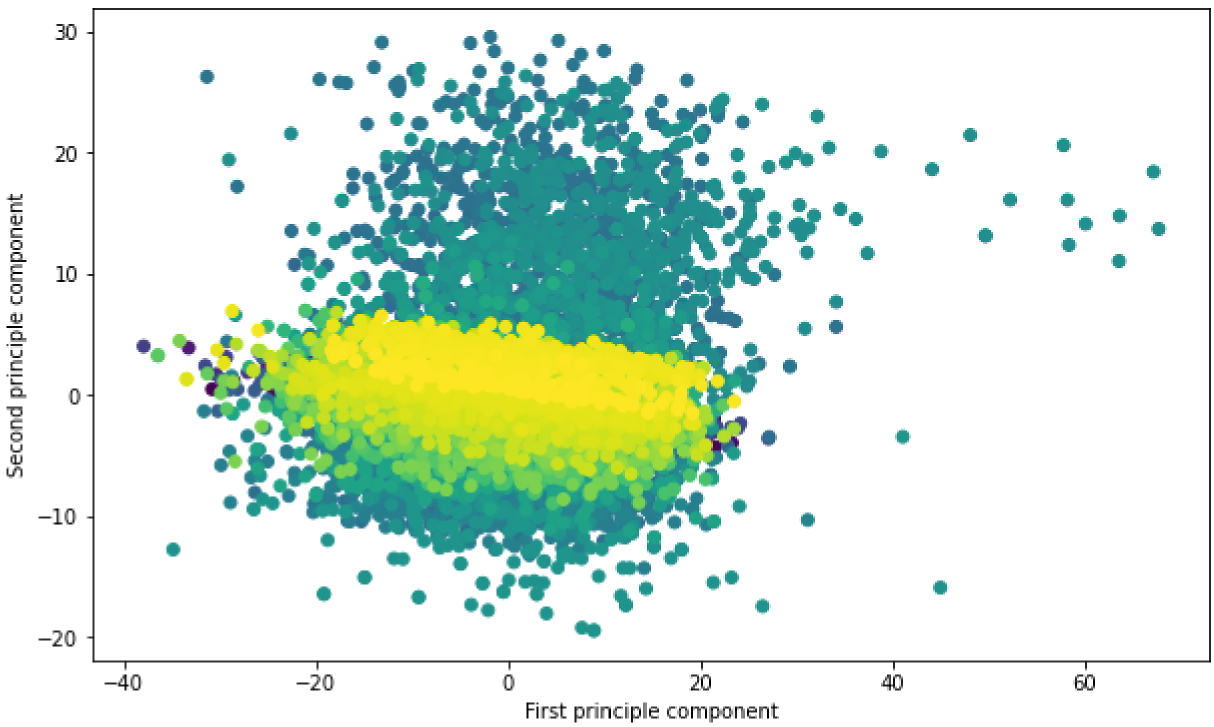


Figure 4.10: PCA of first and second component for running data no 18 (train)

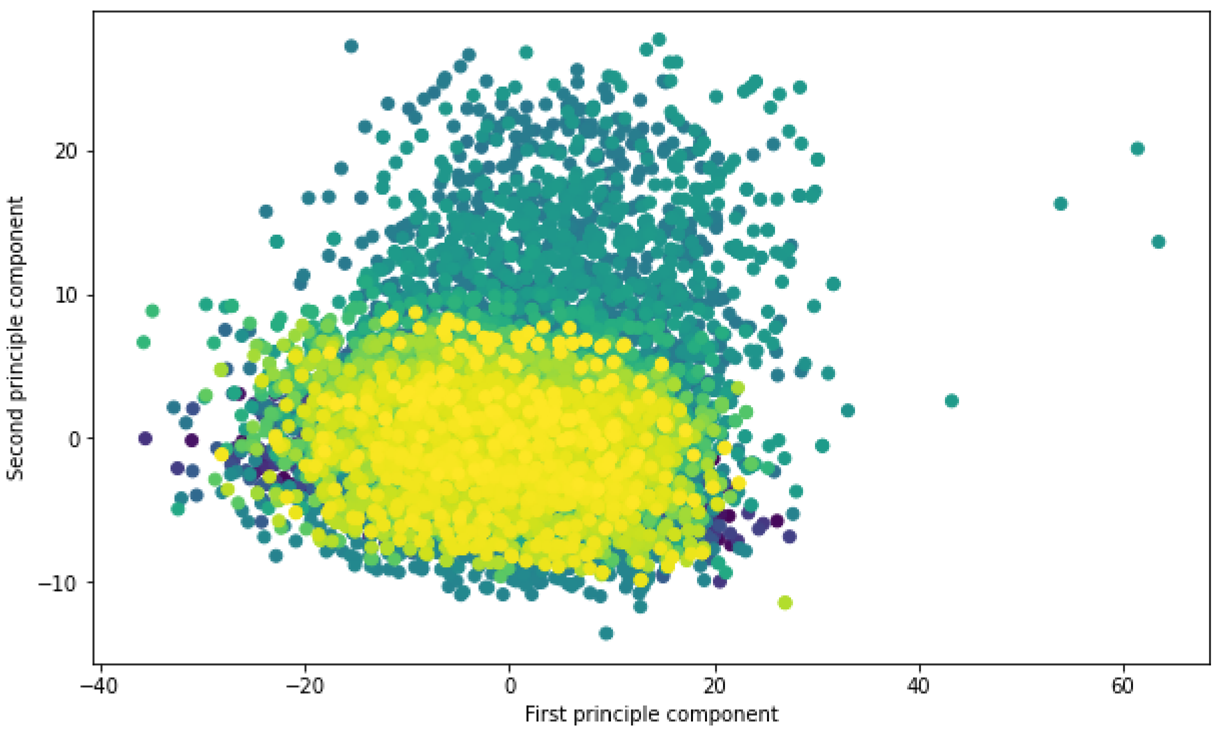


Figure 4.11: PCA of first and second component for running data no 19 (train)

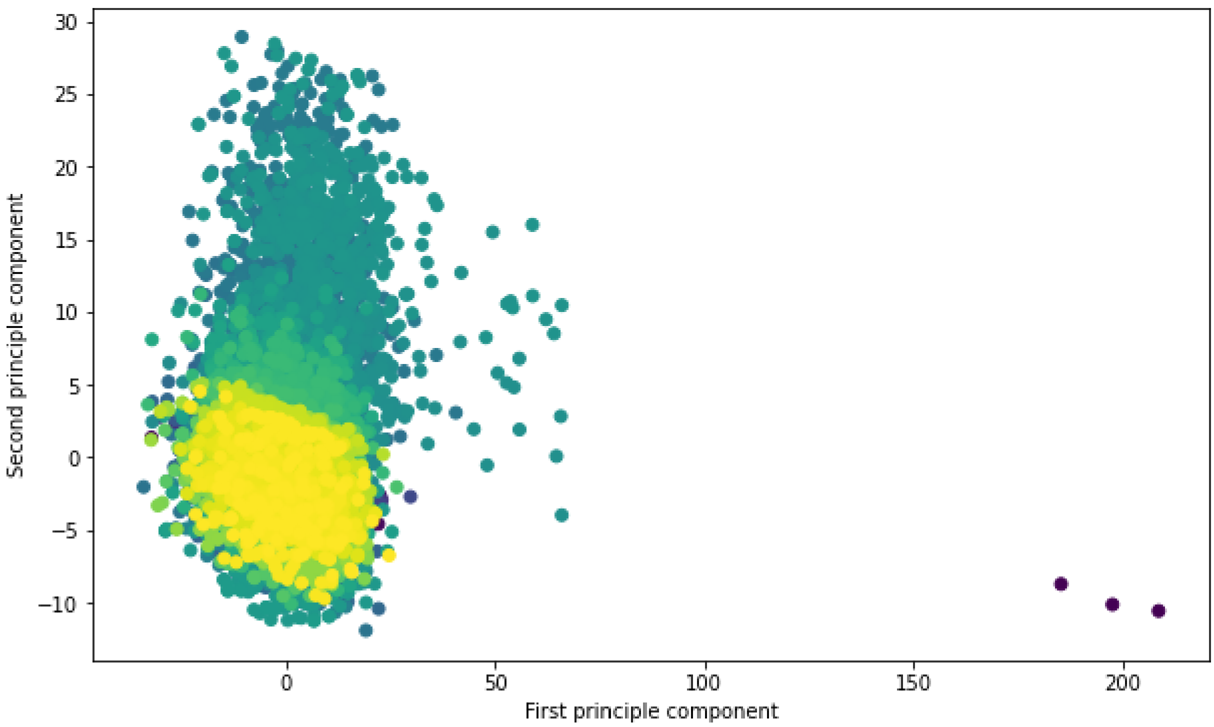


Figure 4.12: PCA of first and second component for running data no 20 (test)

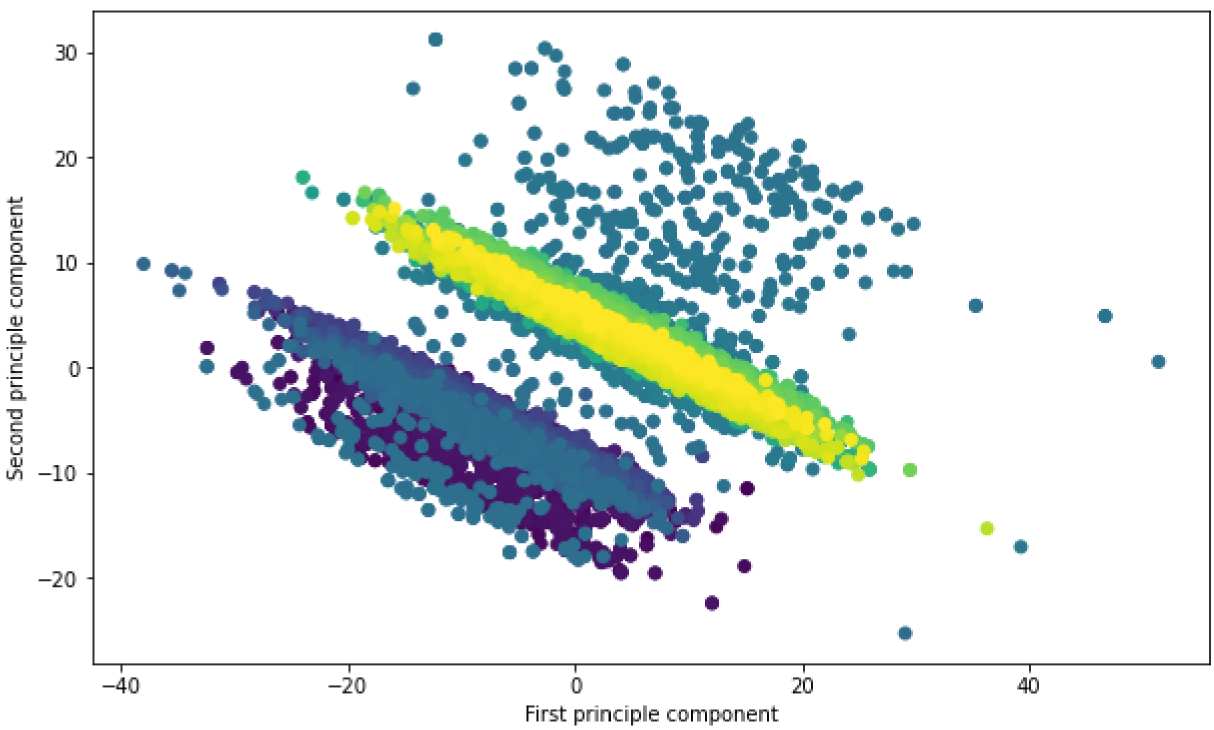


Figure 4.13: PCA of first and second component for standup data no 18 (train)

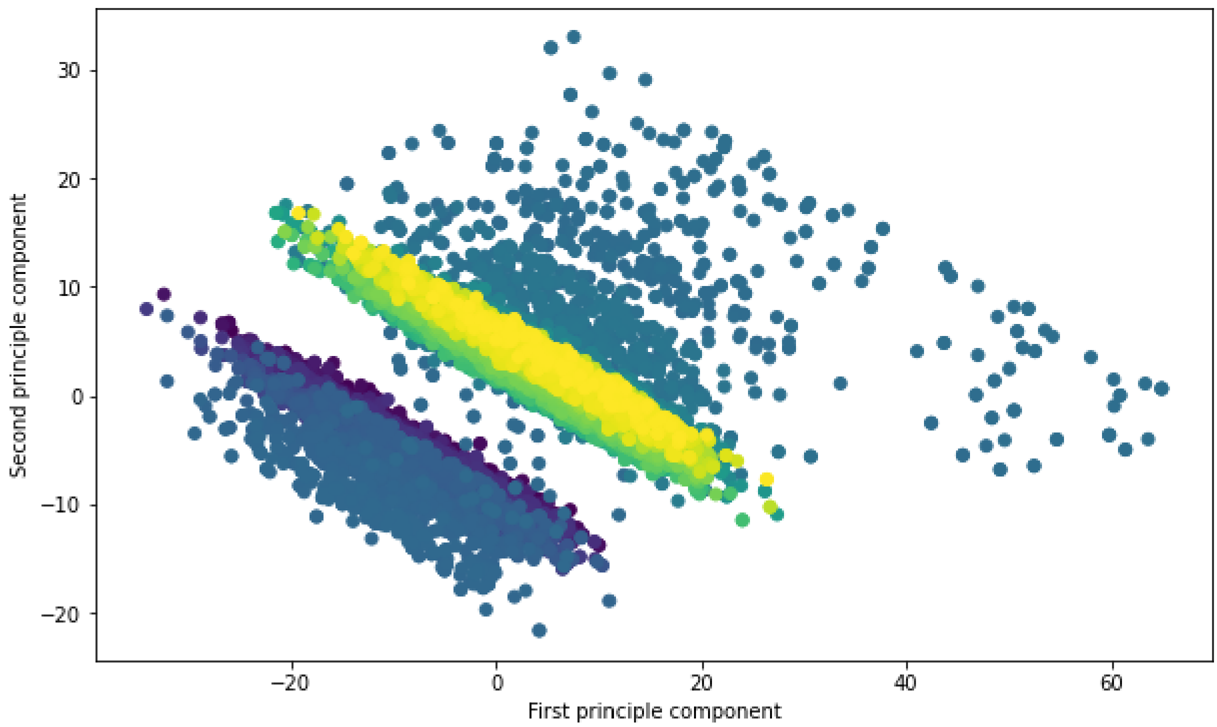


Figure 4.14: PCA of first and second component for standup data no 19 (train)

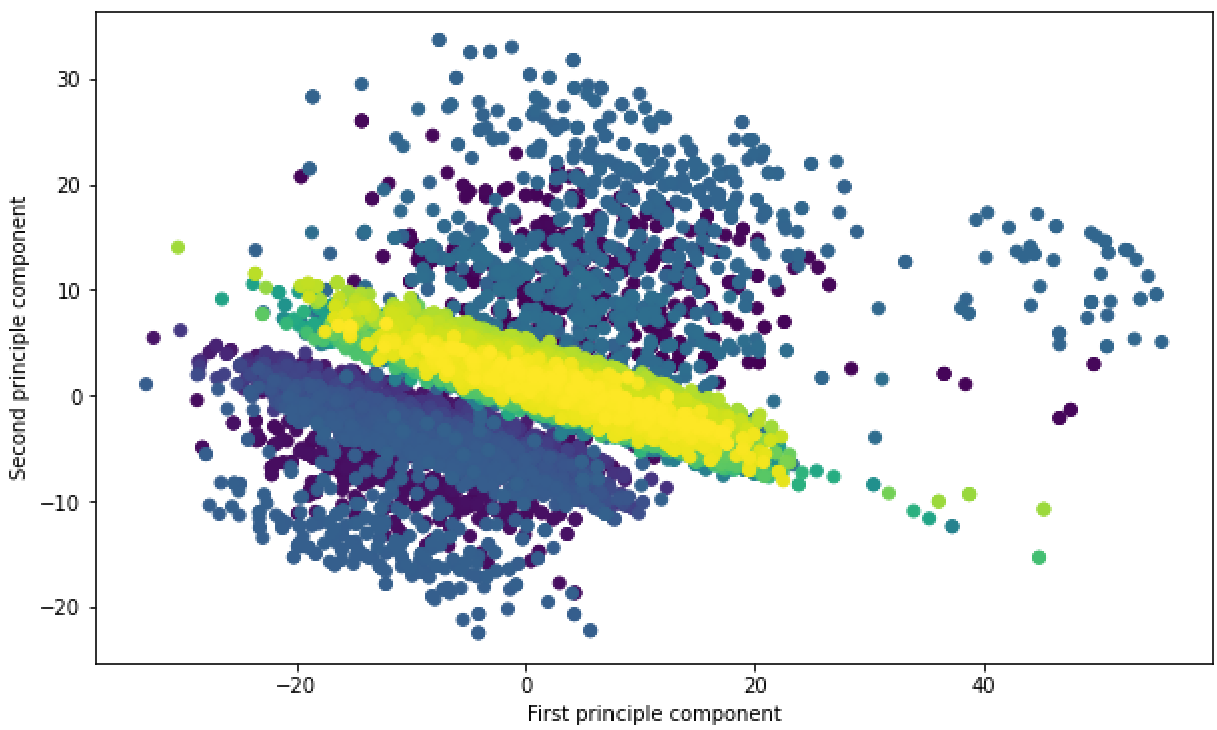


Figure 4.15: PCA of first and second component for standup data no 20 (test)

From the above figures we can see that for walking data we are getting similar types of patterns for both trained and test data. Features for running trained and test data are also similar. The patterns originated from standup data is unique and match has been found between trained and test ones. In conclusion, we can say that our supervised model is able to distinguish a particular type of movement from other types of movement successfully.

Chapter 5

Conclusion and Future Work

5.1 Future Work

The results obtained from our supervised learning approach will be followed and we wish to develop a gait based biometric system to detect gait patterns. After successful completion of a gait based model we will make comparison between the results of our Wi-Fi based approach and gait based approach. [23] Biometrics are physical and behavioral characteristics of human which are used to validate the identity of an individual to give access to a system, data or device. Gait refers to the study of identifying living beings by the unique gait patterns generated due to movement of different body parts. By combining these two methods we would like to develop a model that can identify a person and if that person is authorized to access a particular system it will then give permission. [9] We can use database created by others or we can develop a model of our own for collecting new data. As we know gait based technologies are based on image processing so the hardware requirements are computer and high resolution camera for collecting videos or series of images. We can use Python for programming and coding part of our system.

After collection of data the process of gait recognition begins. [9] We extract silhouette from our data, calculate the gait cycles using algorithms, the silhouette is added and calculation is done to get result. Lastly gait recognition is done by matching results and similarity score computation. In order to increase the efficiency and accuracy of our system we can use algorithms such as Boosted decision tree for taking decision and splitting the dataset, Random Forest Algorithm for error detection.

We also use Image Processing algorithms for improving quality of data. Three steps are involved in the development of Image Processing such as pre-processing, feature extraction and feature reducer. Pre-processing enhance features of image, feature extraction is used to reduce dimensionality and feature reduction is used to work with essentials reducing aspect. All algorithms are implemented in ideal conditions and on particular database while calculating the results. Finally, we try to visualize and analyze our result from the algorithms.

For determining the accuracy of our model we have to make comparison between images. [9] Two or more images can be compared using root mean square value. If the obtained values are between a given range we say match has been found.

Lastly we make a final comparison between our results obtained from Wi-Fi based and gait based approach and take decision accordingly.

5.2 Conclusion

The sole purpose of this study is to develop a Wi-Fi based supervised learning approach to detect movements successfully so that it can be used for security purposes. With the increasing rate of crimes around the world, demand for movement detection based security systems are also increasing. Our supervised approach is able to detect movements of subject from a distance even under low light where other approaches fail to do so. In a crowd surveillance CSI can be obtained through this process without the cooperation and awareness of the subject under observation. As a result, this approach can be implemented in banks, hospitals, offices and even residents to ensure proper security without much hassle.

Our future prospect for this study is to convert our proposed approach into a biometric system. Biometrics identify people from their physical and behavioral traits, as these are always present with a person and there is no way of faking it. Different organizations like Federal Bureau of Investigations (FBI) are using biometrics in criminal investigations and so on. Sometimes it is seen that due to lack of efficient recognition techniques criminals and law offenders are walking in broad daylight under the nose of law enforcement authority. Biometric gait recognition provides a solution to this problem to a large extent. [19] According to the Critical review of the use and scientific basis of forensic gait analysis, "forensic gait analysis has been used as supportive evidence in criminal cases in the United Kingdom for more than 15 years and in Denmark for more than 10 years. In the Netherlands, gait analysis has been performed rarely in the past 20 years. However, two recent criminal cases renewed interest in the topic in the Netherlands.

By using our current supervised approach we can only detect particular activities and movements of a subject successfully such as walking, standing, running, fall, seat etc. In future we would like to incorporate suspicious movements and activities detection in our approach. For example, a repetitive walking pattern can be listed as suspicious activity. If such kind of patterns are detected security and law enforcement authority can take actions. Furthermore, comparison of detected movements can be made with the database which contains movements obtained from criminals and vicious people so that we can identify them. This will be possible if we can upgrade from our current approach to a biometric gait recognition system. Such developed biometric system can be used in entrance of important buildings. If the movement or gait detected from a person trying to enter the building matches with that of a criminal that will mean the person has a criminal record and access will be denied. Furthermore, the biometric system can also alert security guards and other people if a suspicious person is present around them. The biometric system can ensure safety of residents and offices by allowing access only to the people whose gait patterns are recorded in database rejecting unauthorized access. Lastly, there can be drawbacks in gait detection due to difficulty in gait cycle calculation, effect of bad weather, shoes and outfits variation etc. We have to come up with ideas to reduce such effects and make our system more precise so that it can serve the purpose of ensuring security without any hindrance.

References

[1] X. Yang, "Ieee 802.11 n: Enhancements for higher throughput in wireless lans," *IEEE Wireless Communications*, 2005.

[2] P. V. Dorp and F. Groen, "Feature-based human motion parameter estimation with radar," *IET Radar, Sonar Navigation*, 2008.

[3] J. Wilson and N. Patwari, "Radio tomographic imaging with wireless networks," *IEEE Trans. on Mobile Computing*, 2010.

[4] S. Sen, "You are facing the monalisa: Spot localization using phy layer information," *Proc. of the 10th Int. Conf. on Mobile systems, applications, and services.ACM*, 2012.

[5] N. A. Makhdoom, "Human gait recognition and classification using similarity index for various conditions," *IOP Conference Series: Materials Science and Engineering*, 2013. [Online]. Available: https://arxiv.org/pdf/1811.07162.pdf?fbclid=IwAR1DWGDnScuCxh8YxcPiNJDy50ClDIu4ZEc9190nb3TVBMktP_jrn3pCHAI.

[6] Q. Pu, "Whole-home gesture recognition using wireless signals," *Proc. of the 19th annual Int. Conf. on Mobile Computing and Networking. ACM*, 2013.

[7] F. Adib, "3d tracking via body radio reflections," *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, 2014.

[8] O. Politi, "Human motion detection in daily activity tasks using wearable sensors," *Proc. of the 22nd European Signal Processing Conference (EUSIPCO). IEEE*, 2014.

[9] H. Alemdar, "Human activity recognition with wireless sensor networks using machine learning," 2015.

[10] K. Ali, "Keystroke recognition using wifi signals," *Proc. of the 21st Annual Int. Conf. on Mobile Computing and Networking. ACM*, 2015.

[11] W. Wang, "Understanding and modelling of wifi signal based human activity recognition," *Proc. of the 21st Annual Int. Conf. on Mobile Computing and Networking. ACM*, 2015.

[12] A. Tahat, “A look at the recent wireless positioning techniques with a focus on algorithms for moving receivers,” *IEEE Access*, 2016.

[13] G. Wang, “We can hear you with wi-fi!” *IEEE Trans. on Mobile Computing*, 2016.

[14] W. Wang, “Gait recognition using wifi signals,” 2016. [Online]. Available: https://dl.acm.org/doi/10.1145/2971648.2971670.

[15] J. Zhang, “Wifi-id: Human identification using wifi signal,” *International Conference on Distributed Computing in Sensor Systems (DCOSS)*, 2016.

[16] “Artificial intelligence and machine learning: Policy paper,” *Internet Society*, 2017. [Online]. Available: https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/?gclid=CjwKCAiAudD_BRBXEiwAudakX4YTXec2JW3Cwksoxd939nkNW%5C%5CfCXaAfE7XcJtGeByKDgrZq0BFhlVRoCEXQQAvD_BwE.

[17] S. T. Md. Nafiul Alam Nipu, “Human identification using wifi signal,” 2017.

[18] M. N. A. Nipu, “Human identification using wifi signal,” 2017. [Online]. Available: http://dspace.bracu.ac.bd/xmlui/bitstream/handle/10361/9505/13201006%5C%2C13201061_CSE.pdf?sequence=1%5C&isAllowed=y%5C&fbclid=IwAR1NI1WdtJjO-G9aQRM1LIIrtXqeIOf0wOWZgXuRVRgKr-uU-7sbbC8kYaA.

[19] N. M. van Mastrigt, “Critical review of the use and scientific basis of forensic gait analysis,” *NCBI*, 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6201773.

[20] “Decision tree introduction with example,” *GeeksforGeeks*, 2019. [Online]. Available: https://www.geeksforgeeks.org/decision-tree-introduction-example.

[21] N. Donges, “A complete guide to the random forest algorithm,” *builtin*, 2019. [Online]. Available: https://builtin.com/data-science/random-forest-algorithm.

[22] “Silhouette algorithm to determine the optimal value of k,” *GeeksforGeeks*, 2019. [Online]. Available: https://www.geeksforgeeks.org/silhouette-algorithm-to-determine-the-optimal-value-of-k.

[23] F. Wang, “On spatial diversity in wifi-based human activity recognition: A deep learning-based approach,” *IEEE INTERNET OF THINGS JOURNAL*, vol. 6, no. 2, 2019. [Online]. Available: https://www2.cs.sfu.ca/~jcliu/Papers/OnSpatialDiversity.pdf.

[24] Y. Xu, “Attention-based walking gait and direction recognition in wi-fi networks,” *arxiv*, 2019. [Online]. Available: https://arxiv.org/pdf/1811.07162.pdf?fbclid=IwAR1DWGDnScuCxh8YxcPiNJDy50ClDIu4ZEc9190nb3TVBMktP_jrn3pCHAI.

[25] J. Brownlee, “How to check if time series data is stationary with python,” *Mchine learning mastery*, 2020. [Online]. Available: https://machinelearningmastery.com/time-series-data-stationary-python.

[26] Mohit, “Data preprocessing: 6 necessary steps for data scientists,” *hackernoon*, 2020. [Online]. Available: https://hackernoon.com/what-steps-should-one-take-while-doing-data-preprocessing-502c993e1caa.

[27] S. S, “Difference between primary and secondary data,” *keydifferences*, 2020. [Online]. Available: https://keydifferences.com/difference-between-primary-and-secondary-data.html.

[28] “Supervised and unsupervised learning,” *GeeksforGeeks*, 2020. [Online]. Available: https://www.geeksforgeeks.org/supervised-unsupervised-learning.

[29] “Supervised learning,” *IBM Cloud Education*, 2020. [Online]. Available: https://www.ibm.com/cloud/learn/supervised-learning.

[30] Wamidh, “Feature extraction methods: A review,” *Journal of Physics: Conference Series*, 2020. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1591/1/012028.

[31] lady ada, “Pir motion sensor,” *Adafruit*, [Online]. Available: https://learn.adafruit.com/pir-passive-infrared-proximity-motion-sensor.

[32] M. Aljumaily, “A survey on wifi channel state information (csi) utilization in human activity recognition,”

[33] “Casia gait database,” *CBSR*, [Online]. Available: http://www.cbsr.ia.ac.cn/english/Gait%5C%20Databases.asp.

[34] “Dickey-fuller and augmented dickey-fuller tests,” *nwfsc-timeseries*, [Online]. Available: https://nwfsc-timeseries.github.io/atsa-labs/sec-boxjenkins-aug-dickey-fuller.html.

[35] “Fingerprints and other biometrics,” *SERVICES*, [Online]. Available: https://www.fbi.gov/services/cjis/fingerprints-and-other-biometrics.

[36] “Linear regression,” *Linear Regression*, [Online]. Available: https://ml-cheatsheet.readthedocs.io/en/latest/linear_regression.html.

[37] “Linear regression in machine learning,” *Javatpoint*, [Online]. Available: https://www.javatpoint.com/linear-regression-in-machine-learning.

[38] “Machine learning,” *Machine learning*, [Online]. Available: https://en.wikipedia.org/wiki/Machine_learning.

[39] "Machine learning definition," [Online]. Available: https://www.google.com/search?safe=active%5C&sxsrf=ALeKk00ciNnWgoheWkZqxERDTgsDJto2cQ%5C%3A1609863309043%5C&ei=jZD0X6yiAq_G4-EPxsyW-Ac%5C&q=machine+learning+definition%5C&oq=mechine+learning+defi%5C&gs_lcp=CgZwc3ktYWIQARgAMgoIIxCwAhDJAxAnMgQIABAN.

[40] S. Prabhakaran, "Augmented dickey fuller test (adf test) – must read guide," *Let's DataScience*, [Online]. Available: https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test.

[41] M. Rouse, "Machine learning," *Search Enterprise AI*, [Online]. Available: https://searchenterpriseai.techtarget.com/definition/machine-learning-ML.

[42] "Time series analysis," *Statistics Solutions*, [Online]. Available: https://www.statisticssolutions.com/time-series-analysis.

[43] J. Zhang, "Wifi-id: Human identification using wifi signal," [Online]. Available: https://www.cse.unsw.edu.au/~wenh/zhang_dcoss16.pdf.