Non-Invasive Human Activity Recognition using Radar dataset

Author

AQSA RAHIM

Reg. Number

00000318251

Supervisor

DR SAJID GUL KHUWAJA

Co-Supervisor

Dr. MUHAMMAD USMAN AKRAM

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

April 2021

Non-Invasive Human Activity Recognition using Radar dataset

Author

AQSA RAHIM

Reg. Number

00000318251

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Software Engineering

Supervisor:

DR SAJID GUL KHUWAJA

Co-Supervisor

Dr. MUHAMMAD USMAN AKRAM

Thesis Supervisor's Signature:

____

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD

April 2021

DECLARATION

I certify that this research work titled "Non-Invasive Human Activity Recognition using Radar dataset*"* is my work. The work has not been presented elsewhere for assessment. The material that has been used from other sources has been properly acknowledged/referred to.

Signature of Student

AQSA RAHIM

MS – 19 – CSE

PLAGIARISM REPORT (TURNITIN REPORT)

This thesis has been checked for Plagiarism. Turnitin report endorsed by Supervisor is attached.

Signature of Student

AQSA RAHIM

00000318251

Signature of Supervisor

ACKNOWLEDGEMENT

I am thankful to my Creator Allah Subhana-Watala to have guided me throughout this work at every step and for every new thought indeed. Whosoever helped me throughout my thesis, whether my parents or any other individual was Your will, so indeed none be worthy of praise but You.
I would also like to express special thanks to my supervisor for his help throughout my thesis. Finally, I would like to express my gratitude to all the individuals who have rendered valuable assistance to my thesis.

ABSTRACT

Human Activity Recognition has enabled state of art applications in medical healthcare, surveillance systems, digital entertainment and various other sectors. Therefore, prediction of such kind of movements remained an interesting aspect in the field of research. Wearable sensor and vision- based systems have been utilized for the detection of Activities of Daily Life (ADL), however suffer from various limitations including intrusiveness, lighting conditions and privacy issues. This study proposes a transfer learned, deep learning model for Frequency Modulated Continuous Wave (FMCW) radar-based system operating at a frequency of 5.8 GHz with 400 MHz bandwidth for classification of human activities. Radars are highly sensitive to human body movements and can capture small variations as Doppler shift. I have designed and demonstrated a generalized deep learning classification system for the detection of ADL irrespective of the geographical location and the ages of the subjects based on data from FMCW radar-based system. I have utilized experimental public data from 99 participants consisting of 1453 micro-Doppler signatures, at nine different locations including five retirement homes and four laboratories. This study propose micro-Doppler images normalized for speed profile of individuals to obtain age group independent feature maps of human activities and engineer a deep convolutional neural network architecture with a high classification accuracy, sensitivity and specificity of 99.1%, 99.2% and 100%, respectively.

Key Words: Activity Detection, Radars, Data classification, Data preprocessing, Feature extraction, CNN model

TABLE OF CONTENT

DECLARATION iii

PLAGIARISM REPORT (TURNITIN REPORT) iv

COPYRIGHT STATEMENT v

ACKNOWLEDGEMENT vi

LIST OF FIGURES xi

LIST OF TABLES xii

CHAPTER 1. INTRODUCTION 13

1.1. MOTIVATION 13

1.2. PROBLEM STATEMENT 14

1.3. AIMS AND OBJECTIVES 15

1.4. STRUCTURE OF THESIS 15

CHAPTER 2. ACTIVITY RECOGNITION SYSTEM 16

2.1. HUMAN ACTIVITY RECOGNITION SYSTEM 16

2.2. HUMAN ACTIVITY RECOGNITION SYSTEM 18

2.2.1. DAILY LIFE MONITORING 18

2.2.2. PERSONAL BIO-METRIC SIGNATURE 19

2.2.3. ELDERLY AND YOUTH CARE 20

2.2.4. LOCALIZATION 20

2.2.5. INDUSTRY MANUFACTURING, ASSISTING 21

2.3. HAR FEATURES 21

2.4. TECHNIQUES FOR HAR 22

2.4.1. IMAGING SENSOR 22

2.4.2. WEARABLE SENSOR 22

2.4.3. ACCELEROMETERS 23

2.4.4. GYROSCOPE 24
2.4.5. RADAR SYSTEMS..25
2.5. RADARS..25
2.6. BASIC FEATURES OF RADARS 27
2.6.1. Doppler radar:..27
2.6.2. FMCW radar:..27
2.6.3. Interferometry radar:..27
2.6.4. UWB radar: 28
2.7. DEEP LEARNING APPROACHES FOR HUMAN ACTIVITY RECOGNITION IN RADAR 28
2.7.1. Deep Learning Approaches in 3D Radar Echo 28
CHAPTER 3. LITERATURE REVIEW..31
3.1. TECHNOLOGIES FOR HUMAN ACTIVITY RECOGNITION..32
3.1.1. Surveillance Cameras...32
3.1.2. Cameras..33
3.1.3. Wi-Fi 33
3.2. Sensors 34
3.2.1. Accelerometer:..34
3.2.2. Motion Sensor:..34
3.2.3. Proximity Sensor: 34
3.3. RFID..34
3.4. Related Work..35
3.4.1. Radio Frequency Based..35
3.4.2. Sensor Based 37
3.4.3. Wearable Device Based 37

3.4.4. Vision Based 38
3.5. RESEARCH GAP 40
CHAPTER 4. METHODOLOGY 42
4.1. FMCW 43
4.2. MICRO-DOPPLER SIGNATURE 44
4.3. DATA COLLECTION 44
4.4. DATA PRE-PROCESSING 48
4.5. DATA CLASSIFICATION 49
4.5.1. Transfer Learning: 50
4.5.2. Neural Networks 51
4.5.3. Convolutional Neural Networks 52
4.6. MICRO-DOPPLER SIGNATURE 55
4.7. SUMMARY 56
CHAPTER 5. RESULTS AND DISCUSSION 57
5.1. Performance Metrics for Classification 57
5.2. Classification Results 57
5.2.1. Classification using Machine Learning 57
5.2.2. Classification using Convolutional Neural Network 60
CHAPTER 6. CONCLUSION & FUTURE WORKs 64
REFERENCES 65

LIST OF FIGURES

Figure 1: Multiple Activity Detection System [10] 17
Figure 2: Level of Activity 18
Figure 3: Different types of Activity 19
Figure 4: Biometric Signature 20
Figure 5: Industry Manufacturing Assisting 21
Figure 6: Imager 22
Figure 7: Wearable Sensor [18] 23
Figure 8: MEMS S 24
Figure 9: Subject Against Different Activites 24
Figure 10: Different Axes 25
Figure 11: Traditional Machine Learning (ML) techniques. [22] 26
Figure 12: Doppler Radar 27
Figure 13: FMCW Radar 27
Figure 14: Interferometry radar 27
Figure 15: 2D Radar Echo 29
Figure 16: Some technologies used for activity recognition: (a) Surveillance camera (b) Depth camera (c) Wi-Fi (d) Accelerometer (e) Magnetometer (f) Motion sensor (g) Proximity sensor .32
Figure 17: General process of human activity recognition 33
Figure 18: Ultra High Frequencies 33
Figure 19: Proposed Framework 42
Figure 20: Collection of Data 45
Figure 21: Six different Activities 46
Figure 22: Scalograms collected for six activities 49
Figure 23: Transfer Learning Architecture Model 51
Figure 24: Architecture of CNN [56] 52
Figure 25: Graphical Representation of model 54
Figure 26: Objective function model 58
Figure 27: Test accuracy for 30 iterations with combined datasets 60
Figure 28: Normalized Confusion Matrix of (a) CNN and (b) Transfer Learned AlexNet. 62

LIST OF TABLES

Table 1: Comparison of Already Existing Work 39
Table 2: Advantages and Disadvantages of different techniques 41
Table 3: C-band Ancortek Radar Parameters 43
Table 4: Details of the activity's dataset [46]. 46
Table 5: Average information about data subjects [46]. 46
Table 6: Performance Metrics 57
Table 7: Confusion Matrix for KNN 59
Table 8: Confusion Matrix for SVM 60
Table 9: Comparison of the state-of-the-art techniques with the proposed system. 62

CHAPTER 1. INTRODUCTION

Human activity recognition (HAR) remained a well know part of research in computer vision and image processing domain. HAR has enabled state of art applications in medical healthcare, surveillance, digital entertainment and various other sectors [1]. Therefore, prediction of such kind of movements remained an interesting aspect in the field of research. Human activities are envisioned by introducing many sensor-based techniques such as gyroscope, accelerometer etc, having their own benefits and drawbacks.

In a research field Human activity recognition (HAR) is a challenging and highly active topic. Its objectives are to identify several activities performed by an individual or a group of people that are based on sensors and observed data and also includes knowledge due to which the observed activities are carried out [2]. Moreover, an activity is recognized no matter in what environment it is taking place or who is the one performing the activity.

1.1. MOTIVATION

The number of publications has been increasing rapidly, specifically in the field of HAR since last decades. To determine the particular activity and to achieve specific goals in this domain, many researchers have proposed application domains. Many types of devices and sensors are needed such as body inertia sensors, video sensors and environment activities sensors to identify the various actions of human activities. Similarly, some other sensors are required to record or sense the human action. HAR systems have used many other sensors but the effect of outdoor activities and environment on them have limited their usage such as GPS receiver that is only useful for the outdoor environment [3].

Normally, HAR process consists of multiple steps including information collection from raw sensor data on human behavior to find conclusion as well about the currently performed actions. Following are the steps: (i) to remove noise, redundancy, to deal with incompleteness and performing data aggregation and normalization pre-processing is done. (ii) Segmentation involves the most important data segments. (iii) In feature selection process, the essential attributes of features are extracted like statistical moments and temporal and spatial information. (iv) Similarly, to improve the quality some features are lessened as well as to reduce the computational efforts that is necessary for classification in dimensionality reduction process. (v) The given activity is determined through a core machine learning classification process.

The essential objectives of HAR process is to successfully notice and evaluate the human actions and understand current events. Contextual data is retrieved and processed by HAR systems such as temporal, spatial, environmental etc to know the human actions using visual and non-visual sensory data. HAR concepts are explored and developed in various applications domain. These are divided into four categories; for smart homes, Active and Assisted Living (AAL) systems, and tele immersion (TI) applications, healthcare monitoring applications moreover, for indoor and outdoor activities there is a monitoring surveillance system [4]. Observing and analyzing human activities by a human operator was a traditional way. For example, monitoring patient health or in security and surveillance processes. However, this task becomes more cost intensive and much challenging for the operators with rapid growth of technical monitoring devices and camera views. Practically, personnel deployment for these tasks can become financially difficult for the scenario of home care. Therefore, the proficiency and efficacy of the observation and analysis process can be enhanced by replacing the human operator through HAR support in these fields. For example, HAR systems uses sensory devices to maintain the record of a patient health and informs health workers in any serious situation.

1.2. PROBLEM STATEMENT

Human Activity recognition is one of those topics which have large number of applications in different fields including health care, military systems, security related problems, smart home environment and monitoring driving activities. The problem is that the data collected for one certain field is not applicable for the other as well because of the difference of location and the subjects whose data was collected. To address this issue, I have proposed a solution which is independent of geographical location and the age of the participants. The dataset which i have used contains 1453 micro-Doppler signatures, collected at varios different locations including five old care homes and four laboratories, using Frequency Modulated Continuous Wave (FMCW) radar sensor. The preprocessing techniques will be caried out to achieve the good classification accuracy and deep learning algorithm (CNN) will be used for the classification purpose.

1.3. AIMS AND OBJECTIVES

The major objectives of the research are as follow:

- Develop algorithm for human activity recognition
- Preprocessing of the local dataset
- Classify data into six different activities
- Improve response time of system
- Offer low-cost, easy deployable and non-invasive activity detection
- Develop a flexible and scalable wireless sensing system
- work properly in the new environment independent of the location and age group

1.4. STRUCTURE OF THESIS

The structure of the thesis is as follows:

Chapter 2: covers the introduction to the Human Activity Recognition System.

Chapter 3: a review of the literature and the significant work done by researchers in the past few years for Human Activity Recognition.

Chapter 4: consists of the proposed methodology in detail.

Chapter 5: All the experimental results are discussed in detail with all desired figures and tables.

Chapter 6: concludes the thesis and reveals the future scope of this research

CHAPTER 2. ACTIVITY RECOGNITION SYSTEM

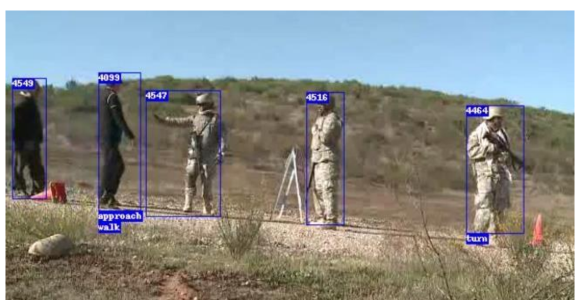
Activity is something that you do or something that is going on. Activities performed by the people for their living, profit motive, entertainment, mental peace, are known as human activities. For example: leisure, entertainment, industry, recreation, war, and exercise. Human activities are categorized into:

- gestures
- atomic actions
- human-to-object or human-to-human interactions
- group actions
- behaviors
- events

2.1. HUMAN ACTIVITY RECOGNITION SYSTEM

Human activity identification, or HAR, is a wide field of research that focuses on recognizing a person's individual behavior or behavior based on sensor data [5]. Indoors, common behaviors such as walking, chatting, standing, and sitting are examples of movements. They may also be more oriented tasks, such as those carried out in a kitchen or on a manufacturing floor. Cam, radar, or other wireless methods may be used to record sensor data from afar. Data may also be collected explicitly on the topic, such as by the use of custom hardware or mobile phones with accelerometers and gyroscopes [6]. The aim of HAR is to identify behaviors based on a set of observations of subjects' behavior and environmental conditions. Activity detection systems are a wide area of research and development, with a recent emphasis on sophisticated machine learning algorithms, hardware design advances, and lowering control costs while increasing protection. Smart home technologies, health care management applications, monitoring and surveillance services for indoor and outdoor sports, and tale -immersion applications are all examples of ambient assisted living (AAL) systems [7]. The technologies are categorized under these groups depending on the methodology used to recognize human activity, such as visual, non-visual, and multimodal sensor technology. It's a difficult challenge to classify time series. It entails predicting a person's movement based on sensor data, and it historically entails deep domain knowledge and signal processing methods to correctly engineer functionality from raw data in order to match a machine learning model. Human-to-human contact and intimate relationships are aided by HAR [8]. It is difficult to extract because it contains details about a person's identity, appearance, and psychological condition. It's a fictitious device that represents some kind of human behavior. The structures are notional in that they are mental models rather than representations of concrete real-world behavior. There are many examples in the world of groups of human activities that are linked to create a whole (system). For example, the actions that make British Rail a human activity mechanism are linked to constructed physical structures such as the railway network, with its platforms, tracks, and so on. One of the major topics of research in the scientific fields of computer vision and machine learning is the human capacity to perceive another person's activities [9]. Many implementations, such as video monitoring devices, human-computer interaction, and robots for human behavior characterization, now include a multiple activity detection system as a result of study as shown in Figure 1.

Two key questions emerge from different classification techniques: “What action?” and “How do I know?” (i.e., the issue of recognition) and “Where in the video?” (i.e., the issue of localization). When trying to identify human actions, one must first assess a person's kinetic states in order for the algorithm to recognize the action effectively. Human movements like "walking" and "running"



occur spontaneously in everyday life and are relatively easy to identify. More abstract tasks, such as "peeling an onion," are, on the other hand, more difficult to recognize. Complex tasks can be broken down into simpler activities that are more easily recognized.

Figure 1: Multiple Activity Detection System [10]

Object identification in a scene will usually aid in better understanding human behaviors by providing helpful knowledge about the current event [11]. The aim of human activity identification is to look at events in video or static photographs. Human activity recognition systems are motivated by this fact, and their goal is to correctly identify input data into its underlying activity group.

There are three major components of most activity detection systems are listed below and shown in Figure 2:

- A low-level sensing module that collects sensitive information about activities in real time using microphones, accelerometers, light sensors, and other sensors
- A feature processing and selection module that converts raw sensor data into features that can distinguish between activities
- A classification module that uses the features to infer what task an individual or a group of people is doing, such as driving, cooking.

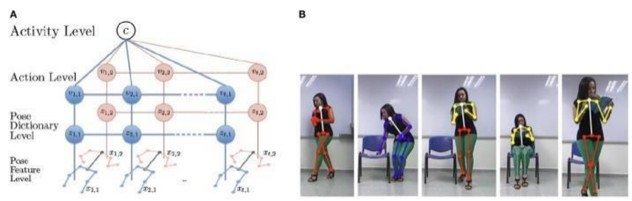


Figure 2: Level of Activity

2.2. HUMAN ACTIVITY RECOGNITION SYSTEM

Many interesting technologies depend on activity recognition as a key component [12]. Action detection applications can be categorized based on their intended beneficial subjects:

- Exercise tracking, wellness management, fall prevention, behavior-based context recognition, home and work automation, and self-managing systems are all applicable to end users.
- Third-party applications such as targeted ads, data processing tools for testing, business administration, and accounting [13].
- Crowd and community applications, such as social networking and activity-based crowdsourcing. Examine some sample applications in this section.

2.2.1. DAILY LIFE MONITORING

Applications for real life tracking are typically designed to serve as a handy reference for action recording or to aid in fitness and healthier lifestyles. These devices include embedded sensors such as an accelerometer, gyroscope, and GPS that measure people's steps, stairs climbed, calories burned, hours slept, distance travelled, and sleep quality, among other things. Users will study data monitoring and visualization in reports using an online service. These instruments are more advanced than mobile phone sensors because their sensors are developed primarily for movement identification and monitoring. The disadvantage is that they are far more costly. In recent years, smartphone apps that use behavior detection strategies have emerged as a viable option. These apps typically perform the same functions as the above-mentioned specialized instruments. Users' motion records, such as jogging routes, steps taken, and sleep time, are tracked as shown in Figure 3. They will be able to provide the user with a description of his or her lifestyle and report on sleeping efficiency by mining the logged data.

Figure 3: Different types of Activity

2.2.2. PERSONAL BIO-METRIC SIGNATURE

The motion pattern of a subject is normally exclusive and one-of-a-kind. When people lift their hands, for example, it is almost impossible for two people's hands to have the same gesture patterns. Because of the distinctions between motion-related bones and muscles on human bodies, even in a good imitation, variations still remain. Sensors like accelerometers can detect these variations. Human biometric signatures of patterns in motion/gestures can be solved using behavior detection techniques. Pattern detection techniques are used in these applications to extract special motion patterns, which are then saved in a database. Because of the widespread use of mobile devices, it is both simple and possible. On the other hand, the motion signature could be used for malicious purposes. People may, for example, use the studied patterns to decipher users' behavior, such as tapping on a smart phone keyboard or other espionage activities as shown in Figure 4.

Figure 4: Biometric Signature

2.2.3. ELDERLY AND YOUTH CARE

Because of the ageing of the baby boomer generation, there is greater need for elderly treatment (both physically and mentally). The development of emerging technology and software for elderly treatment is a main focus of ongoing studies in human activity tracking. These applications may be used to help avoid damage, such as detecting unsafe conditions in older adults. A mobile phone architecture is being designed with the aim of detecting users' falls. Elders could benefit from behavior detection and track sensors in a constructive manner, such as life routine reminders (e.g., taking medicine) and living activity tracking for a remote robotic assist. Another area that gains from behavior identification studies is youth care. Monitoring infants' sleeping patterns and anticipating their requests for food or other items are two examples of applications. Children with Autism Spectrum Disorder (ASD) are also detected using activity recognition techniques [14].

2.2.4. LOCALIZATION

Mobile phone activity recognition may aid context perception and, as a result, be used in localization. One explanation for using smartphone sensors instead of GPS for location is that GPS signals are usually slow inside buildings and underground [15]. Action detection techniques combined with mobile sensors, on the other hand, may aid in locating the location. Furthermore, GPS localization is a two-dimensional positioning system that does not take into account a user's altitude. Mobile phone activity recognition strategies may be able to fill this void. A common method is used for floor localization without the use of facilities. A third justification to use smartphone sensors for localization is that GPS precision degrades within cities surrounded by tall buildings. In this case, GPS-based localization may make the distinction between a movie theatre and a restaurant, which may be only a few feet apart. By augmenting the roles with people's actual activity type, activity recognition-related apps may reduce these types of errors.

2.2.5. INDUSTRY MANUFACTURING, ASSISTING

Staff may benefit from behavior detection strategies in their everyday jobs. Wearable sensors are used in this project. Wearable computing is a form of body extension that enables a worker to carry out exceptional tasks. Smart cameras that can recognize people's movements in the filming area, robot assistance in car manufacturing, and other applications focused on movement recognition as shown in Figure 5.

Figure 5: Industry Manufacturing Assisting

2.3. HAR FEATURES

Both HAR domain tasks necessitate the accurate recognition of human behaviors from sensor data, which necessitates the proper categorization and description of features extracted from sensor data. Following that, we'll go over some of the functionality that are found in the HAR domain.

Although certain features may be derived from physical activity signals, more features do not always mean better classification precision since the features could be redundant or not class-specific:

- Time domain features (applied to the amplitude and time dimensions of a signal, such as mean, median, difference, standard deviation, minimum, limit, and root mean square) are often used in many functional HAR systems since they are less computationally expensive and therefore can be derived in real time.
- To differentiate between various human operations, frequency-domain features necessitate a higher computational cost. As a result, they may not be appropriate for real-time AAL applications [16].

- Physical characteristics are obtained from a basic understanding of how a particular human activity produces a particular sensor signal. Physical parameters of human motions are normally used to derive physical characteristics from different sensor axes.

2.4. TECHNIQUES FOR HAR

2.4.1. IMAGING SENSOR

An image sensor, also known as an imager as shown in Figure 6, is a sensor that detects and transmits data used to create an image. It accomplishes this by translating light waves' variable attenuation into signals, which are short bursts of current that carry information. Light or other electric radiation may be used as waves.

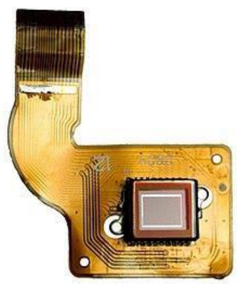


Figure 6: Imager

Image sensors are used in a variety of electronic imaging instruments, including digital cameras, camera modules, camera phones, optical mouse devices, medical imaging equipment, night vision equipment including thermal imaging devices, radar, and sonar, among others. Electronic and digital imaging are increasingly replacing chemical and analogue imaging as technology advances.

2.4.2. WEARABLE SENSOR

Wearable cameras are the most widely used for human behavior identification because they can directly and reliably capture body movements. Smartphones, watches, bands, and even clothing will all benefit from these sensors. With the miniature wearable sensors integrated, Wearable Technology (WT) provides new ways to actively track human behavior as shown in Figure 7. The analysis of human activity detection (HAR) using WT offers insight and interpretation of data gathered from the mining method, which aids in the improvement of wearable technology techniques and concepts [17]. Across sectors, it boosts performance, production, operation, and participation. Wearable sensor devices are gaining traction in the scientific community due to the use of highly miniaturized electronic components with low power consumption, making them suitable for indoor and outdoor applications of human activity recognition. These apps enable users to perform any physical exercise in a natural manner while delivering excellent outcomes in a variety of practical applications, including health recovery, respiratory and muscular activity assessment, athletics, and protection.

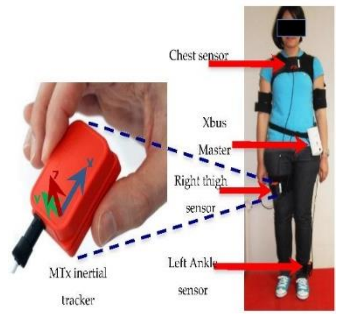


Figure 7: Wearable Sensor [18]

2.4.3. ACCELEROMETERS

In cell phones, accelerometers are used to detect the orientation or to feel the acceleration case. The linear acceleration of travel is measured by an accelerometer. Three axes of predetermined paths are used in the reading. The raw data stream from the accelerometer is the acceleration. The raw data is represented by a series of vectors: Acci =, (I = 1,2, 3...). An accelerometer can reliably determine a device's directional acceleration, but it can't fix the lateral orientation or tilt during that shift. The measurements from the three axes are paired with a time stamp. Many existing accelerometers have a user interface that allows the user to experiment with different sampling rates in order to sample the frequency. Accelerometers, both single and multi-axis, detect the amplitude and trajectory of angular, rotational, and gravitational acceleration [19]. They can be used to provide basic motion sensing capabilities. For example, in a fixed state location, a system with an accelerometer may sense rotation from vertical to horizontal. As a result, accelerometers are mainly used in consumer electronics for basic motion sensor applications such as switching a mobile device's screen from portrait to landscape orientation. Its success stems from the fact that it tests a subject's physiology motion status directly. For example, if a consumer switches from walking to jogging, the signal form of the accelerated reading along the vertical axis will change abruptly. Furthermore, acceleration data may suggest a motion pattern within a specified time span, which is useful in recognizing complex activities.

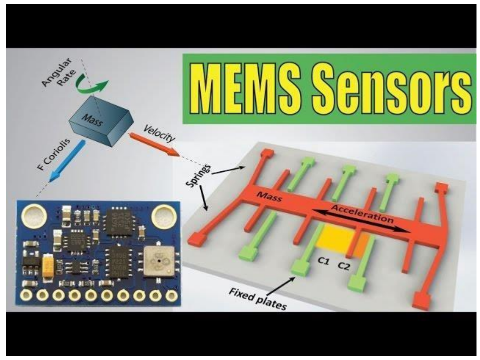


Figure 8: MEMS S

2.4.4. GYROSCOPE

The angular rate of rotational movement along one or more axis is calculated by gyroscopes. Unlike accelerometers, which can only monitor the fact that an object has shifted or is travelling in a certain direction, gyroscopes can precisely calculate dynamic motion in several dimensions, measuring the orientation and rotation of a moving object. Furthermore, unlike accelerometers and compasses, gyroscopes are unaffected by errors caused by external forces like gravity and magnetic fields as shown in Figure 9.

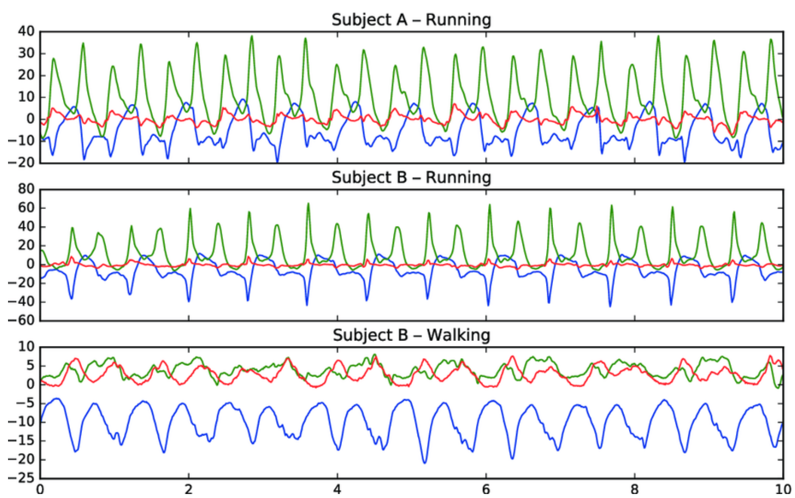


Figure 9: Subject Against Different Activites

As a result, gyroscopes significantly improve computer motion sensing capability and are used in sophisticated motion sensing applications in consumer electronics, such as absolute gesture and action tracking and simulation in video games. The tilt, pitch, and yaw movements of the Smart phones around the x, y, and z axes, respectively, are detected by the gyroscope, which determines the phone's rotation rate as shown in Figure 10. The axes, to be precise. The rate of rotation in rad/s (radian per second) along each of the three physical axes is the raw data stream from a gyroscope sensor: Rotation I = (i=1,2,3,...). The gyroscope is useful in navigation applications and some smart phone games that rely on rotation data. The gyroscope is used to assist the mobile orientation identification in activity recognition studies [20].

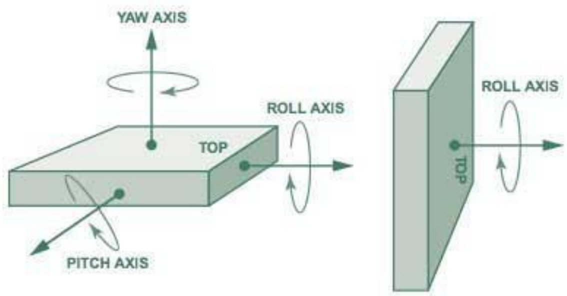


Figure 10: Different Axes

2.4.5. RADAR SYSTEMS

Radar has special characteristics as one of the sensors for human activity detection (HAR), such as privacy protection and contactless sensing. Human–computer contact, smart monitoring, and health evaluation have all benefited from radar-based HAR. The generalization capabilities of traditional machine learning approaches are constrained since they depend on heuristic hand-crafted feature extraction. Furthermore, manually scraping features is wasteful and time–consuming. Deep learning, which uses a hierarchical approach to automatically learn high-level functions, has outperformed HAR.

2.5. RADARS

Since radar is resistant to light and temperature, it can be used in rugged environments [21]. Radar has the potential to preserve visual privacy. Instead of catching the target's visual structure, the returned signals modulated by the target carry a wealth of time–varying activity range and velocity information. Humans can be detected by radar via walls. As a result, radar-based HAR can be used in a wider range of situations. Radar systems do not need the attachment of a sticker on the human body, making them more user-friendly. As a result, radar has been increasingly used to detect human activity in recent years. Radar-based HAR networks have traditionally used traditional machine learning (ML) techniques as shown in Figure 11. These standard algorithms are based on scientific foundations, making them understandable and logically optimizable. Their complexity is also smaller than that of deep learning models, resulting in a lighter computing load.

Radar is a kind of active sensing device that sends out radio waves and receives modulated signals from illuminated targets. In recent decades, it has mostly been used in remote sensing systems

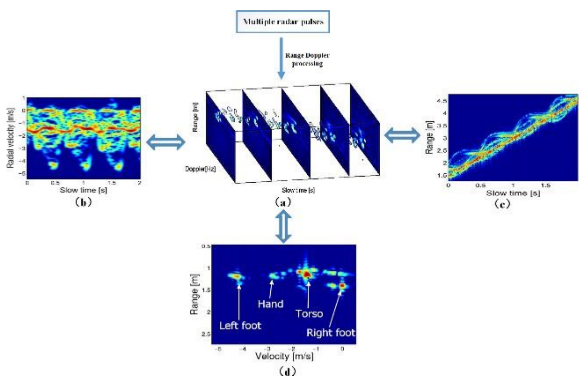


Figure 11: Traditional Machine Learning (ML) techniques. [22]

such as satellite remote sensing, air and terrestrial traffic control, and geophysical tracking. Furthermore, short-range radar for HAR activities has recently expanded. Since radar is insensitive to light and atmosphere, radar-based HAR approaches are more reliable than vision-based ones. Without any tag applied to the human body, they can sense human activity and behaviors directly. The speeds/Doppler frequencies of body parts differ with respect to the person's movement while he or she is driving. As a result, the distributions of these components are not continuous in time. The range, distance, and angle knowledge obtained by radar may be used to identify human activities. Radar is a viable human motion measuring technology because of its inherent advantages such as straightforward design, fast device integration, low cost, and penetration capabilities. Continuous-wave radar, ultra-wideband radar, and noise radar are among the radars used by HAR.

2.6. BASIC FEATURES OF RADARS

2.6.1. Doppler radar:

Doppler radar transmitting single-tone radio waves capable of acquiring target Doppler/radial velocity information as shown in Figure 12.

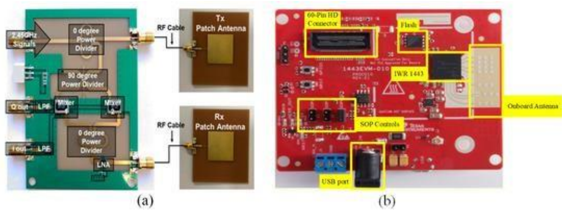


Figure 12: Doppler Radar

2.6.2. FMCW radar:

It provides target range and speed information at the same time, making it ideal for situations with many targets as shown in Figure 13.

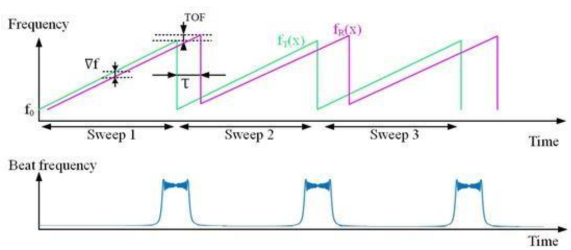


Figure 13: FMCW Radar

2.6.3. Interferometry radar:

It obtains the target's angular velocity independent of the target's travelling position by cross-correlating the output of two antennas as shown in Figure 14.

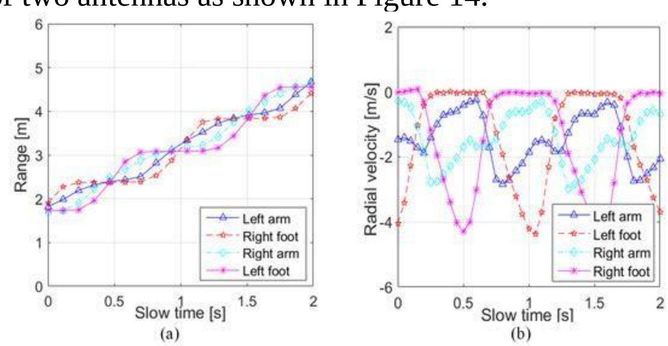


Figure 14: Interferometry radar

2.6.4. UWB radar:

It provides fine range precision, allowing the target's main scattering centers to be distinguished.

2.7. DEEP LEARNING APPROACHES FOR HUMAN ACTIVITY RECOGNITION IN RADAR

2.7.1. Deep Learning Approaches in 3D Radar Echo

Moving and micro-Doppler properties of targets are shown by range–Doppler frames. The 3D RD video series, which is made up of N time-sampled 2D range–Doppler frames, shows both spatial and temporal characteristics. Per RD frame contains range and Doppler information, while time information occurs within frames. The joint time–range–Doppler echoes comprise nearly half of the operation information that radar gets, compared to 1D and 2D echoes. It is necessary to develop models that can retrieve both temporal and spatial data. Since manually designing features from 3D echoes is challenging, DL approaches are more feasible and desirable for 3D echo-based HAR due to their ability to automatically remove deep features. Additionally, the introduction of GPU allows DL models to process 3D data rapidly and efficiently. DL approaches on 3D echoes are promising for HAR, despite the fact that there are few DL algorithms proposed for 3D radar echoes to date.

2.7.1.1. Deep Learning Approaches in 2D Radar Echo

2D radar echoes, also known as time–Doppler maps, time–range maps, and range–Doppler maps, hold enough information about human activity as shown in Figure 15. Since 2D echoes are typically viewed as images, CNN has become the most widely used standard for 2D echoes, alongside the line of computer vision. As a result, 2D echo-based HAR is often repurposed as an image classification task.

1. Doppler–time map
2. Doppler–range map
3. Doppler range map
4. Hybrid 2D maps

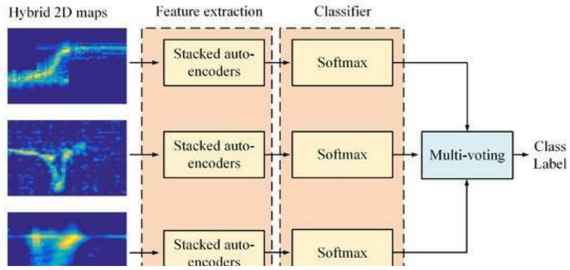


Figure 15: 2D Radar Echo

2.7.1.2. Deep Learning Approaches in 1D Radar Echo

1D radar echoes are basically time-series data, comparable to that collected from accelerometer and gyroscope sensors. As a result, several time series methods may be applied to 1D echo-based HAR. Because of the benefits of modelling sequential data, RNN is often used for 1D data. A. Graves et al., for example, suggested a speech recognition architecture that uses LSTM and the Connectionist Temporal Classification (CTC) algorithm to mark unsegment sequence data. This teaches us how to remember continuous events without having to manually annotate them beforehand. For the first time, A. Hamid et al. [23] applied 1D CNN to a hybrid NN-HMM paradigm for speech recognition and suggested partial weight sharing. DL methods have the ability to remove sequential features and produce successful classification results for 1D radar echoes, despite the fact that there are few DL-related studies for 1D radar echoes.

2.7.1.3. Perspectives for the Future

Despite the fact that radar-based HAR with DL algorithms has made significant strides, there is still a long way to go before it reaches maturity. It is critical for planned DL architectures to be capable of exploring activity information in radar echoes as much as possible as a method for feature extraction and activity detection.

2.7.1.4. Radar-based human activity recognition in real-world scenarios.

Most recent radar-based HAR methods are only valid in controllable settings, where a human subject performs a series of discrete and delegated tasks with minimal interference. Furthermore, the model's real-time computing power is not taken into account. However, some aspects must be closely considered before radar-based HAR can be used in real-world scenarios.

Light-weight deep model design

Training a deep learning model also necessitates a lot of computational power, so it's mostly done off-line with a small amount of data. In practice, behavior data is always received in a stream, necessitating extensive online and gradual learning. Traditional ML methods for real-world HAR are hindered by large feature engineering and hand-craft feature extraction, despite the fact that they are capable of processing and classifying data in real-time. As a result, for radar-based HAR, light-weight DL models must be created. Combining hand-crafted features with deep features and cooperating DL models with traditional ML algorithms are two concepts worth investigating [24].

Continuous activity segmentation and recognition:

In real-life situations, people work continuously and freely rather than only doing their given tasks. It is critical to accurately segment and recognize the events that are of concern. There has been a recent pattern of approaching segmentation and identification together. To understand continuous complex hand signals, a Connectionist Temporal Classification (CTC) algorithm was used in. CTC allows gesture recognition without the need for overt pre-segmentation and simultaneously tackles segmentation and recognition. More algorithms aimed at jointly segmenting and identifying a set of behaviors are required in future research.

Multi-target activity recognition:

It's worth looking at how to classify different targets' behaviors or how to distinguish a target from a squad. The use of DL approaches for multi-target human gait detection was investigated. In the presence of several targets, an FMCW radar was used to distinguish and identify several assigned hand signals. Such strategies, on the other hand, frequently operate in less alarming situations, such as where a human object is making movements and an individual is heading toward the radar at the same time. Applying DL models to learn high-level functionality is critical in situations where radar echoes are modulated by multiple moving targets. For multi-target behavior detection, more complex DL models should be created

CHAPTER 3. LITERATURE REVIEW

Over the past two decades, activity recognition (HAR) has become a hot subject in research because of its potential uses in fields like fitness, remote control, gaming, protection and surveillance, and human-computer interaction. The ability to recognize/detect current behavior based on input obtained from various sensors is known as activity recognition [25]. Cameras, wearable sensors, sensors mounted to everyday objects, and sensors deployed in the atmosphere are all examples of sensors. Logging everyday activities has become very common and realistic as technology has advanced and computer costs have decreased. People are keeping track of their everyday habits, such as cooking, dining, resting, watching television, or the number of steps they take. Various techniques have been used to document these events.

Because of the less price and innovation in technology using sensors, the majority of HAR study has moved to a sensor-based approach. Various sensors are used in the sensor-based approach to capture human activity as they conduct everyday life tasks. On the basis of sensor implementation, these solutions can be grouped into 3 main categories: I Wearable ii) Device based, and iii) Non-Invasive [26]. In a wearable, the user would wear the device when doing some operation. Although there has been a lot of work done on movement detection using wearable devices, the main issue with this technique is that wearing a tag is not always possible. In the case of the old or ill people, for example, they can fail to wear the tags or refuse to wear it. Sensors are added to everyday objects with solutions that use an object-tagged approach. Different behaviors are remembered based on a user's contact with these objects. This is a device-bound approach, which ensures that apps must only use unique objects (tagged-objects). This solution, like the wearable approach, will not always be possible since it requires consumers to use tagged-objects.

Researchers have been working on Non-Invasive solution in which consumers are not allowed to take any tags or devices with them for the past few years. The idea is to position sensors in the environment, and when a person participates in some activity, the data is collected by the device, which can then be used to detect the action or any movement. The gadget-free method is more convenient as it eliminates the need for the consumer to hold a device when engaging in some tasks. However, there are certain drawbacks to this strategy, such as intrusion from the climate. The data collected by the sensors can be disrupted by the environment, resulting in data noise.

As seen in Figure 1, human behavior identification is a complete mechanism that can be separated into four main types [26]. These steps are: I Selection of the Sensor and its deployment ii) Collection of the data from the devices, iii) Pre-Processing of the data and then selection of the features, and iv) Recognize human activities using machine learning algorithms

3.1. TECHNOLOGIES FOR HUMAN ACTIVITY RECOGNITION

Major study has been conducted in the field of behavior detection using technologies over the last decade. Accelerometers, motion sensors, biosensors, gyroscopes, friction sensors, contact sensors, and other sensors are among the most often utilized for behavior detection. Any of the sensors, such as RFID, are radio-based. This sensor can be found in a number of applications. They may be mounted to a variety of items, worn as sensors, or installed in the field. Many various types of inexpensive and compact sensors are now available that can detect and relay information over wireless networks. As seen in Figure 16, we include information on some of the technology used for human activity identification in this section.

Figure 16: TECHNOLOGIES FOR HUMAN ACTIVITY RECOGNITION

3.1.1. **Surveillance Cameras.**

Installing security cameras in the premises and monitoring human movements (as shown in Figure 17) is the most common and conventional method of action identification. Monitoring can be performed manually (by a human monitoring the videos and photographs collected by the cameras) or automatically. Various computer vision methods have been introduced to interpret camera data including videos and images and to identify events automatically.

3.1.2. Cameras

The problems with conventional cameras are their reliance on sun, which means they can't operate in low light. The creation of depth cameras like Kinect, which can operate in complete darkness, solved this problem. Kinect can produce a variety of data sources [27]. It can gather data about the human body and build a simulated skeleton in 3D. Since various behaviors of the body (skeleton) are linked to behaviors, activities may be identified using this information. Apart from the complicated computation, depth cameras are costly, which is a downside to using them for action detection.

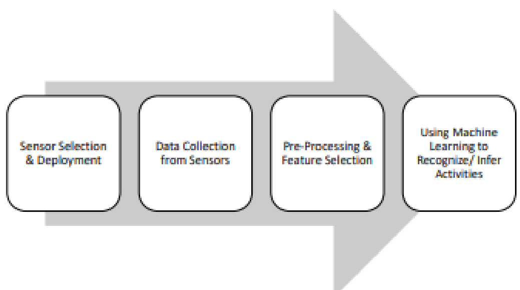


Figure 17: General process of human activity recognition

3.1.3. Wi-Fi

There has been a paradigm change in human activity identification studies in the last decade, from device based to non-invasive approaches. Researchers have begun to use Channel State Information (CSI) for behavior identification after studying the properties of wireless networks [28] (as shown in Figure 18). Many Wi- Fi-based applications for translation, monitoring, and fall detection have

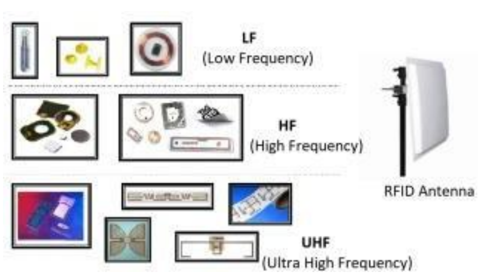


Figure 18: Ultra High Frequencies

been suggested. Wi-Fi has the advantage of being unobtrusive, and people do not need to take any computers with them.

3.2. Sensors

Major work has been performed in the area of technology in the twenty-first century, and several different types of sensors have been developed. These sensors are incredibly useful because they can detect the atmosphere and relay data wirelessly. The following are some of the sensors that are often used in activity detection studies.

3.2.1. Accelerometer:

An accelerometer is an electromechanical instrument that tests the acceleration of a moving item. It has the ability to track movement in different directions. The accelerometer is equipped with multi-axis (i.e., x, y, and z) sensors to do this. A multi-axis accelerometer can concurrently calculate acceleration in the x, y, and z directions. The accelerometer is commonly used in motion recognition, pose recognition, fall detection, monitoring, ambient assisted living, activities of daily living, and other uses.

3.2.2. Motion Sensor:

Motion sensors sense the movement or motion of a subject in a specific location. In the field of human behavior identification, motion sensors are often used in motion monitoring, tracking, and people counting.

3.2.3. Proximity Sensor:

A proximity sensor is an electronic device that senses the location of adjacent objects without having physical interaction. In motion recognition, proximity sensors are commonly used.

3.3. RFID

In the last ten years, the field of radio frequency identification has exploded. This technology, which was originally designed for the purpose of military, has advanced dramatically in recent years [30]. It's commonly used in supply chain monitoring and recording. The range of this technology was originally very narrow, but it has since been significantly extended [31]. The reader and tags are the two key components of RFID technology.

A reader is a computer that reads tags and collects data from them. The reader has a radio antenna that sends out radio waves. RFID tags receive and modulate these radio waves with their content, such as ID. Via an antenna with tag detail, the reader can catch these backscattered signals.

Tags are tiny chips that can be added to a variety of items. There are two kinds of tags: Active tag and the other is Passive tag. For active tags it has its own battery, on the other hand passive tags don't have one and focus on harvesting. Their electricity comes from the readers' radio waves. As compared to passive tags, active tags have a larger range.

RFID has been implemented in a number of fields due to its passive existence, low cost, and unobtrusiveness. In human behavior recognition science, RFID is now widely used. RFID technology is being used by researchers for posture recognition, gesture recognition, mapping, localization, and behavior recognition, among other things.

3.4. Related Work

Over the past decade, a significant amount of study has been undertaken in the field of human behavior identification. Many surveys exist that summarize studies in the field of activity detection. These studies look at various approaches to activity detection and can be divided into four major groups, which are mentioned below.

3.4.1. Radio Frequency Based

Scholz et al. [32] published a review of the literature on device-free radio-based behavior detection. The ongoing work in device-free radio-based localization (DFL) and device-free radio-based behavior detection is classified in this survey (DFAR). The authors cover a wide range of topics for DFL, including reliable presence detection, spatial coverage, adaptive machine learning, radio tomography, and mathematical modelling. Adaptive threshold-based DFAR, machine learning-based DFAR, and mathematical modeling-based DFAR are the three forms of DFAR in the literature. Wang and Zhou [33] presented a review of findings in the area of radio-based behavior detection. The current study is divided into four sections in this survey: I ZigBee radio-based, ii) Wi-Fi-based, iii) RFID-based, and iv) other radio-based (e.g., FM radio, microwave). The authors compare both of these approaches using metrics such as coverage, accuracy, operation types, and implementation costs. They also suggest some research topics for the future. This research is limited to a single device-free method focused on RFID. Cianca et al. [34] present a review of the work done in the field of human behavior detection using RF signals. Presence detection, fall detection, movement detection, gesture and pose recognition, people counting, personal characteristic identification, breathe and vital sign detection, and human-object interactions are among the sub-categories identified by the authors. This research focuses on device-free passive sensing approaches and classifies them according to signal characteristics (bandwidth, carrier frequency, and transmission mode), method of measurement on the obtained signal (directly generated CSI or raw data from an SDR platform), and signal descriptor used. This survey paper gives a thorough overview of the RF signal behavior recognition work. Amendola et al. [35] published a survey that summarized the use of RFID technologies in health-related Internet of Things (IoT) implementations. Environmental passive sensors, such as volatile compound sensors and temperature sensors, as well as body-centric tags, such as wearable tags and implantable tags, are discussed in this paper. This paper also discusses several RFID applications of human behavior analysis, including tracking, gesture detection, and remote control. In the field of RFID science, the authors have research guidance. This paper addresses the potential use of RFID technologies in various applications, but it does not go into depth regarding the research that has been performed in those fields. Ma et al. [28] presented a brief overview of behavior detection studies using a Wi-Fi-based approach. The paper provides a brief description of the core technologies in Wi-Fi-related work from the literature in order to devise a structure for a Wi-Fi-based behavior recognition scheme. Base signal collection, pre-processing, attribute extraction, and classification techniques are the main steps in this framework. There are three common types of base signals discussed: amplitude, phase, and phase difference. Outline elimination, irrelevant information removal, and redundancy removal are all part of the pre-processing phase. Space transformation and feature selection are part of the feature extraction step. Finally, two approaches are explored in the classification step: rule-based and machine learning-based. The literature on behavior identification is divided into two categories in this study: coarse-grained activities and fine-grained activities. This work only addresses Wi-Fi-based analysis, and there is no information about the given work. Instead of focusing on the steps involved in a Wi-Fi-based human activity recognition model, this survey focuses on the steps involved in a Wi-Fi-based human activity recognition model.

3.4.2. **Sensor Based**

Chen et al. [33] provided a comprehensive overview of sensor-based studies in the field of human behavior recognition. The current research efforts are divided into two sections in this survey I sensor-based vs. vision-based, and ii) data-driven vs. vision-based vs. a knowledge-driven approach. The first classification was based on the Sensor-based methods are the subject of the survey. Wearable sensors (such as accelerometers, GPS, and biosensors) and dense sensing are explored in detail. The literature in behavior identification is divided into data-driven and knowledge-driven categories in the second classification method comes into play.The writers talk about data-driven methods generative modelling and discriminative modelling techniques are developed further for knowledge-driven methods. There are three types of logic-based, ontology-based, and mining-based systems comes into play. The survey's primary priority is data-driven Techniques for recognizing activities. Wang et al. [36] conducted a survey that outlined the various deep learning techniques for human behavior detection using sensors. This paper organizes the behavior recognition literature by sensor modality, deep model, and application region. The literature is grouped into four categories based on modality: body-worn sensors, object sensors, environmental sensors, and hybrid sensors. Linked work is divided into three categories based on the deep model: discriminative deep architecture, generative deep architecture, and hybrid deep architecture. The related work is categorized as tasks of daily life, sleep, sports, and fitness in terms of the implementation field. This survey summarizes behavior recognition analysis, with a particular emphasis on the deep model used to process sensor data.

3.4.3. **Wearable Device Based**

Cornacchia et al. [37] conducted a thorough review and divided the current literature into two categories: global body motion behavior (e.g., walking, jumping, and running) and local contact activity (e.g., moving the extremities) (e.g., use of objects). The type of sensor used and the location of the sensor on the human body, such as waist mounted and chest mounted, are also classified in this article. The authors address various strategies that use sensors such as gyroscopes, accelerometers, magnetometers, portable cameras, and hybrid sensors (which combine several sensors). Since many approaches use a cell phone (built-in sensors) for movement detection, some surveys often rely on mobile phone-based solutions for HAR. A survey of this kind is presented. Shoaib et al. [38] have an overview of the research work cellular phones. The work of Lara

&Labarador [39] in behavior recognition using wearable sensors was described. This survey covers a wide range of topics in HAR architecture, including sensor and attribute selection, data collection and procedure, recognition efficiency, processing methods, and energy consumption. This study divides current work into three categories: supervised online, supervised off-line, and semi-supervised off-line. The survey's key focus is on human activity recognition technologies that capture data using wearable sensors.

3.4.4. **Vision Based**

Vrigkas et al. [40] conducted a review of recent studies on movement identification using vision-based methods and divided the literature into two categories: unimodal and multimodal approaches. Unimodal models use data from a single modality and can be categorized as stochastic, rule-based, space-time-based, or shape-based. Multimodal approaches draw on evidence from a variety of sources and are categorized into behavioral, effective, and social-networking approaches. This study focuses solely on vision-based movement detection methods. Herath et al. [41] presented a comprehensive review of the main studies in the field of action perception focused on vision. The overall work is divided into two sections in this survey: solutions based on representation and solutions based on deep neural networks. Holistic and local demonstrations, as well as aggregation processes, are two types of representation-based solutions. Multiple stream networks, temporal coherency networks, generative structures, and spatiotemporal networks are all examples of deep neural network solutions. This paper offers a thorough review of the work undertaken in the field of intervention identification. Many solutions for behavior detection using device-free RFID technology have been proposed recently as RFID technology has advanced. The specifics of these solutions were overlooked in previous surveys. There has been no prior survey that offers a systematic and informative overview of RFID-based device-free methods for behavior detection, to the best of our knowledge.

Table 1: Comparison of Already Existing Work

Categories	Paper	Main Focus	Future Directions	Comparison of techniques
RF-based	Scholz et al. [32]	Applicability of radio sensors in activity recognition	Yes	No
RF-based	Amendola et al [35].	Applications of RFID technologies in various fields	Yes	No
RF-based	Wang and Zhou et al. [36]	Use of radio signals for activity recognition	Yes	Yes
RF-based	Ma et al [28].	Wi-Fi based techniques	No	No
RF-based	Cianca et al. [34]	FM radio and Wifi based methods	No	No
Sensor based	Chen et al. [33]	Data centric activity recognition technique	Yes	Yes
Sensor based	Wang et al. [36]	Deep models for sensor-based approaches	Yes	Yes
Wearable device based	Lara & Labarador et al. [39]	Wearable sensor-based approaches	Yes	Yes
Wearable device based	Shoaib et al. [38]	Mobile phone-based techniques	Yes	Yes
Wearable device based	Cornacchina et al.	Wearable sensor-based approaches	No	Yes
Vision based	Vrigkas et al. [40]	Vision based approaches	Yes	Yes
Vision based	Herath et al. [41]	Vision based solutions	Yes	Yes

3.5. RESEARCH GAP

We provided an overview of the total work performed in various areas of behavior recognition, with an emphasis on device-free methods. Human behaviors have been recognized using a number of methods. We discovered that comparing these strategies is difficult for the reasons mentioned below.

We discovered that comparing these methods is problematic for a variety of purposes. Action identification includes a number of sub-areas. We conducted a literature review on both of these sub-areas and discussed the various approaches proposed in each. The main focus of the work addressed differs, with some focusing on one sub-area and others concentrating on another. As a result, comparing both of these methods is challenging. It's difficult to compare a gesture recognition technique to an ADL recognition technique, for example. In gesture recognition, processing time is crucial, and the solution must produce results in original, whereas in Activity of Daily Living, time is less of a concern, but the results are more important. We have provided a comparison with radars with some common factors.

Some methods, for example, rely on wearable electronics, while others do not; some rely on sensors connected to objects, while others rely on Wi-Fi. Furthermore, these implementations use a variety of classification techniques (machine learning tools).It's not straightforward to compare strategies that are based on entirely different methods. Every strategy has advantages and disadvantages, but comparing these methods to others is difficult. We did our utmost to provide the reader with a comprehensive contrast.

There is no one-size-fits-all approach to testing these methods. The experimental setups used in various solutions vary from one another. Some solutions, for example, use devices that are wearable to conduct campaigns in room, while others use devices to conduct these campaigns. Since precision and other variables are dependent on the experimental context, comparing solutions with different experimental configurations is difficult. The table below summarize the approaches and define advantages and limitations of different methods.

Table 2: Comparison of different techniques

Approach	Technology	Advantages	Disadvantages
Vision based	Surveillance Camera	High Accuracy	High cost, Complex computation, Privacy issue
Depth Sensor	Kinect	High Accuracy	High cost, Privacy issue
Wearable Sensor	Gloves, Smart Watch	Low cost	Constraint to wear the device
Object tagged	Accelerometer, Ultrasonic Sensor, Microphone	Low cost	Device-bound
RFID	Passive RFID tags Arrays	Low cost, Passive	Environmental Interface

CHAPTER 4. METHODOLOGY

The aim of the methodology part in the proposed study is to present the details and implementation of the Human Activity Recognition and the CNN model followed by data collection and data preprocessing steps. This is a supervised machine learning algorithm for classification. Here below we have the flow diagram of our proposed methodology as shown in Figure 19 which presents the flow of the steps which we will follow for the validation of our proposed methodology.

First of all, we need to collect the data to validate our proposed methodology. Data has more importance because it depends on what we are trying to do mean in which field and about what problem. It depends on the aim of the study about which about what we are going to solve. After the dataset collection, there is some pre-processing which is compulsory to do to bring the data into an understandable form. Because data gathered from different resources and can be in different formats and could have many noises. So, the aim of the pre-processing is to remove the noise and unwanted data. After the preprocessing, the last step of our flow is to classify the data using the

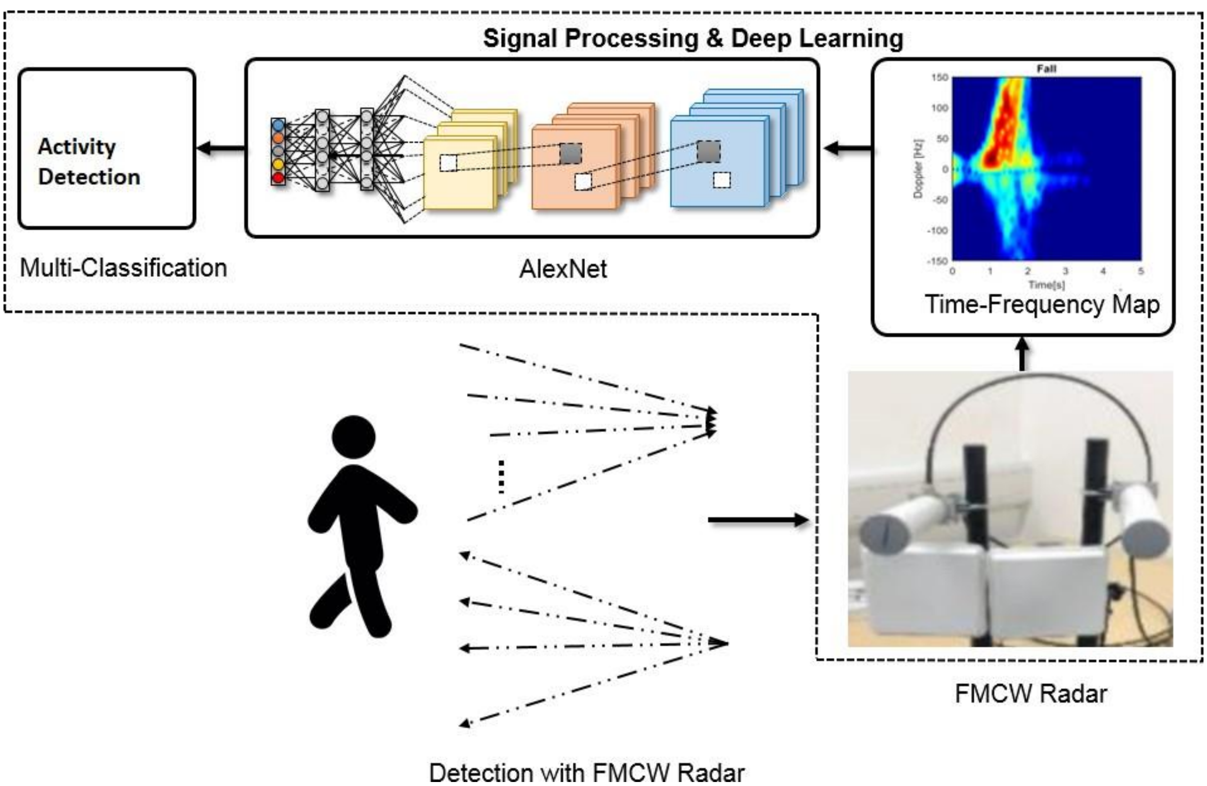


Figure 19: Proposed Framework

trained CNN model, and finally, we will compare their accuracy with the result which was collected from the generalized dataset.

4.1. FMCW

Electromagnetic waves are used by radar for object detection. Transceiver and Signal processing are the main components of a classic radar system. [42] RF signals are continuously disseminated by operated FMCW radar and waves are reflected to the receiving antenna by any object within the range of the system. In FMCW radar, the momentary transmission frequency changes linearly over the waveform, giving a broadly received arrangement for low-cost, brief to medium extend sensing application which also incorporates ADL [43]. FMCW radar provides benefits in terms of is its strength against outside narrow-band interference from other sources, less power at peak, and capability of recording micro-Doppler marks for recognition of targets [44]. Due to the above-mentioned desired qualities, we have used an FMCW radar sensor by Ancortek, the specifications of the radar are illustrated in Table 3.

Table 3: C-band Ancortek Radar Parameters

Radar Model	SDR 580AD
Waveform	FMCW
Operating frequency	5.6 to 6.0 GHz
Bandwidth	400 MHz
Sweep time	1 ms
Transmitting power	~20 dBm

Equation 1 gives the total number of signals that are transmitted by FMCW radar.

$$x(t) = \sum_{i=0}^{NF-1} \mathrm{x(i)\ (t - iT)} \tag{1}$$

In equation 1, T_F is the entire duration of a sweep, N_F tells the full number of transmitted sweeps. The transmitted signal with L number of chirps at the i^{th} sweep can be computed as:

$$x(t) = \sum_{i=0}^{L-1} \mathrm{xo\ (t - iT)} \tag{2}$$

In equation 2, FMCW chirp signal $x_0(t)$ is composed as follow:

$$\mathrm{x(t)} = \sum_{x}^{i=1} e^{ft+\frac{\mu}{2}t)2\pi j} \tag{3}$$

In the above equations, f_0 is the working or operating frequency, μ signifies alter in the instantaneous frequency of an FMCW signal.

4.2. MICRO-DOPPLER SIGNATURE

When the target moves [45] within the range of the radar, it actuates a recurrence tweak on the returned signal that creates sidebands proportional to the size of target's Doppler recurrence move, called the "micro-Doppler effect" [46]. Any moving individual found in K moves with frequency f and displacement D, has a relocation work given as, D (t) = $(2\pi f_v t)cos(\beta)cos(\alpha p)$. The mathematical expression for the received radar signal can be defined as:

$$s(t) = \rho\ (e^{2\pi ft+\frac{R(t)}{\alpha}t)j)} \tag{4}$$

In the above equation, f_0 is the frequency of the carrier, λ is the carrier's wavelength and ρ denotes the backscattering.

To summarize the equations, the micro-Doppler recurrence component is directly related to the speed of the displacement of the target's portion and the recurrence of the radar signal. Moreover, there's a reliance on the cosine of the perspective angles in azimuth and rise, in which the outspread component of the speed vector contributes to the micro-Doppler signature. With the presence of micro-Doppler marks, the extraction of such information from the reflected radar signal can play a crucial part in the activity detection framework. We will discuss the publicly available dataset that we utilised for our work in the next section.

4.3. DATA COLLECTION

The radar dataset utilised for this work was collected by University of Glasgow, United Kingdom (UK) as shown in Figure 20 and is publicly available at [46]. The data in [46] was obtained at three different organizations including the University of Glasgow, North Glasgow Housing Association Residential Center, and Age UK West Cumbria Daily center. The data was collected for 10 days at 9 different locations in the organizations. The dataset contains 4 locations from the University of Glasgow, 3 from North Glasgow Housing Association Residential Center and 2 from Age UK West Cumbria Daily center (Figure 20). The data consists of a total of 99 participants from the University of Glasgow and North Glasgow Housing Association Residential Center. The ages of the participants range from 21 to 98 years. The dataset contains 56 participants from Age UK West Cumbria Daily center with majority of them above 50.

Figure 20: Collection of Data

The experiments consist of six different daily routine activities as shown in Figure 21. These activities include standing, sitting, walk in the forward and backward direction, pick up objects, drinking, and falling. All the activities are repeated three times for every subject except for some people of very old age. The activities are for the duration of 5 seconds except the activity of walking which is for 10 seconds. The details of the dataset in [46] are given in Table 4. The average information about data subjects including average age, average height and male to female distribution is given in Table 5. The overall average age and height of all subjects is 42.6 years and 173.8 cm, respectively. The male to female distribution is 75.24% males to 24.76% females in entire dataset.

Figure 21: Six different Activities

Table 4: Details of the activity's dataset [46].

Location	ID	No. of obs.	Age Group	Environment
1	Dataset 1	360	Younger Participants	Lab Environment
2	Dataset 2	48	Younger Participants	Lab Environment
3	Dataset 3	162	Younger Participants	Lab Environment
4	Dataset 4	288	Younger Participants	Lab Environment
5	Dataset 5	141	Mature Participants	Retirement Home
6	Dataset 6	105	Mature Participants	Retirement Home
7	Dataset 7	60	Mature Participants	Retirement Home
8	Dataset 8	184	Mature Participants	Retirement Home
9	Dataset 9	105	Mature Participants	Retirement Home

The data in [46] is obtained from a deployable FMCW radar sensor [47] which is manufactured by Ancortek, the RF signal transmitted by it is of 400 MHz bandwidth at 1 kHz pulse repetition frequency and the transmitted power is nearly +20dBm. For transmission and reception of signals, two Yagis antennas have been utilised in [19], one as a transmitter and other as a receiver with gains equal to 17 dBi. The Universal Serial Bus (USB) is used to power the radar and its power consumption is limited by the USB standards. The C-band is used to operate the radar and the signal is centered at 5.8 GHz. In case of the realistic scenarios, when it is deployed in real-time scenarios the radar should always be connected with a laptop or desktop for the acquisition and processing of the data.

Table 5: Average information about data subjects [46].

Location	ID	Avg. Age (years)	Avg. Height (cm)	Male	Percen t. (%)	Female s	Percen t. (%)
1	Dataset 1	25.7	177.9	20	100.00	0	0
2	Dataset 2	25.0	178.5	4	100.00	0	0
3	Dataset 3	26.6	172.8	7	87.50	1	25.00
4	Dataset 4	27.3	174.4	15	93.75	1	6.25
5	Dataset 5	30.5	174.6	15	88.24	2	11.76
6	Dataset 6	57.6	161.4	8	80.00	2	20.00
7	Dataset 7	65.6	177.2	5	50.00	5	50.00
8	Dataset 8	67.5	-	4	30.77	9	69.23
9	Dataset 9	57.9	-	1	14.28	6	85.71

4.4. DATA PRE-PROCESSING

Regarding human activities, the FMCW radar offers information in both range and Doppler. In this study, we have focused primarily on the Doppler information from the dataset in [48], since this information is enough for the detection of ADLs. For collection of data, participants are asked to perform activities within a short range of 1 to 3 m and the radar sensor in [46] is placed on a wooden table in all nine locations. To maximize the strength of the received signals, two antennas are placed in a way that would allow keeping the participants' torso in the center of the beam. In order to characterize spectrograms and produce micro-Doppler signatures, the data collected using the FMCW radar sensor in [46] was processed using the Short-Term Fourier Transform (STFT). To get the range profiles, radar signal is stacked in matrix form and then applied Fast Fourier Transform (FFT) algorithm on it. To characterise their micro-Doppler signatures, STFT is applied to the range cells containing the target signatures, in this case, moving subjects. STFT applies a series of FFTs with small, overlapping intervals over the total length of the reported data; the squared absolute value of the complex result is the spectrogram and represents a plot of moving body parts velocities (measured by Doppler effect) as a function of time. To eliminate the contribution of static targets near 0 Hz such as furniture, walls, ceilings, and floors, a notch moving target indication filter is applied.

Figure 22 illustrates micro-Doppler signatures examples for six human activities such as standing walking back and forth, sitting, picking up an object, drinking, and a fall. The Doppler components' positive values correspond to the radar sensor movements, whereas any movement away from the radar produces negative Doppler values. In Figure 22, in the walking activity: the main contribution (in red) comes from the torso as the subject moves back and forth in front of the radar, hence resulting in alternate values between positive and negative.

An acceleration towards the ground is observed when examining the fall activity. The spectrograms for each activity were normalized in this figure as the distance varies between the radar sensor and subject. Figure 22 represents spectrograms for various subjects of different age groups, obtained at nine different locations such as four different rooms in the University of Glasgow, three at North Glasgow Housing Association Residential Centre, and two at Age UK West

Cumbria Daily Centre. Some of the participants could move quickly depending on their physical characteristics while few had limited walking capacity. For example, the spectrogram associated with Room 7 in Figure 5 was the slowest to move in backward and forward directions. Figure 22 was generated using radar dataset [46]

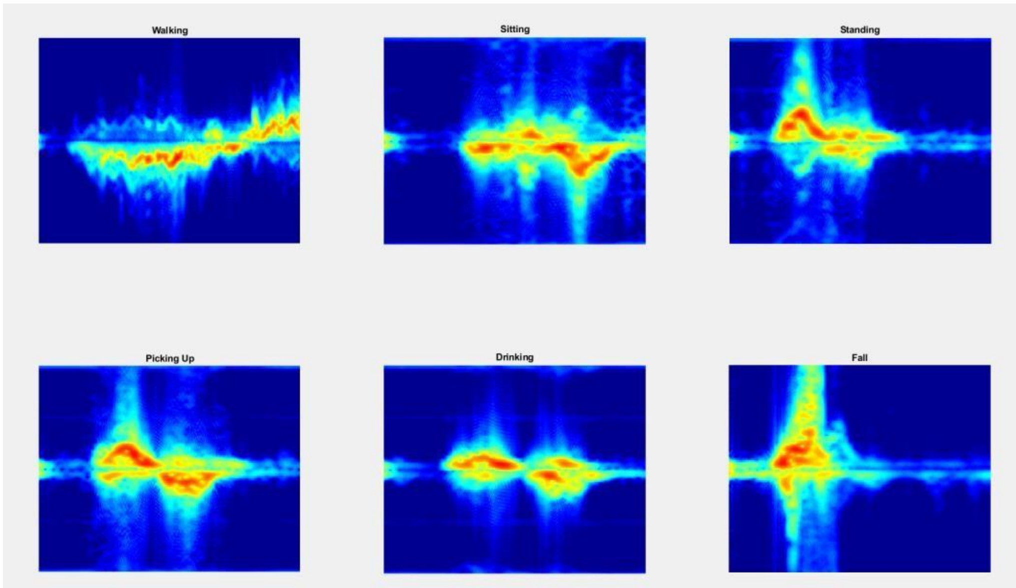


Figure 22: Scalograms collected for six activities

We extracted the average body velocity, as the spectrogram mass core. This can be a proxy for subjects' overall mobility because decreased mobility and the risk of falling are typically associated. In addition, this shows that our radar-based system is capable not only of measuring the velocity of different body parts but also of presenting the average motion velocity. We used the velocity profiles to normalize the spectrograms for age group independent features for classification.

4.5. DATA CLASSIFICATION

Deep learning [49] has increasingly gained attention due to rapid improvements in algorithms and higher computation power delivered by state-of-the-art graphics processing units. Deep neural networks utilise multiple layers of neurons. This increases the complexity of the non-linear input-output relationship and the overall size of the network. The inputs are combined linearly and plugged into an activation function to simulate neurons in the deep network. Formerly, hyperbolic or sigmoid functions were used as activation function, but there were limitations on the network size and appropriate training imposed by the vanishing gradient problem. During the backpropagation process, a gradient descent technique is used to train the neural network that helps in minimising the loss function of the network. However, as the input data flows through each layer step by step, the error rate gradually decreases leading to slow training with an increased number of layers. Nair et al. [50] has resolved this problem by using Rectified Linear Units (ReLU) as activation functions. The ReLU primarily has zero output for negative input values and positive values for positive inputs. When the network is arbitrarily initialised, ReLU enables a sparse representation of data. Furthermore, ReLU considers values between 0 and 1 to reduce the vanishing gradient problem substantially. ReLU activation functions have produced good results on huge datasets. An eight-layer architecture, AlexNet [51] won the visual recognition competition in 2012. Later, the same challenge was won by a 16-layer deep network [35] namely VGG-Net [52] and 152-layer network architecture, ResNet [53]. Currently, the research works on deep learning algorithms includes processing and classification of millions of images into thousands of classes. This allows the researchers to experiment with deep networks to classify datasets of RF signals within the radar and healthcare community. However, the availability of limited data sets[54] to apply deep network algorithms to RF signals classification is a big challenge. In the applications of radar systems for monitoring and surveillance purposes, the RF data collection is relatively difficult, costly to perform, and time-consuming. Therefore, to obtain millions of micro-Doppler signatures for human activity recognition is impractical. To mitigate the effects of smaller datasets in RF detection, transfer learning has been proposed, which utilises pre-trained deep networks and fine-tunes the deep network with a smaller radar dataset at the training stage. This approach seems to perform better in biases of the network and random initialisation of the weights.

4.5.1. **Transfer Learning:**

Transfer learning is a vital tool in machine learning and deep learning classification where insufficient data is available for training [55]. Figure 23 displays the architecture of Transfer learning. It predominantly provides the difference between target domains, tasks, and distributions of the training. This infers the accessible training and test datasets may utilize different distribution with the end goal that P(x) may follow different marking function P (y|x) which may have a diverse set of features for various classes. With regards to transfer learning, the datasets that utilize similar distributions having similar labels with similar capabilities are known as the source. The center thought of utilizing transfer learning is to learn machine learning algorithms for classification dependent on the target datasets that principally advantage from the existing datasets began from

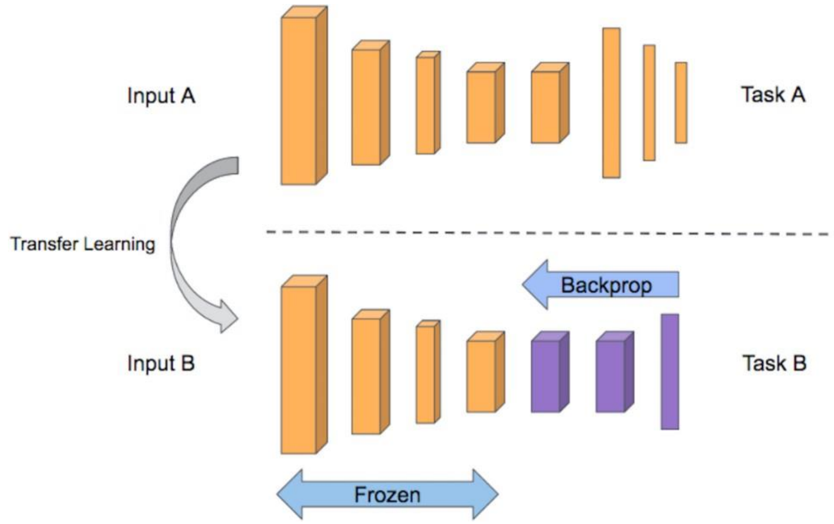


Figure 23: Transfer Learning Architecture Model.

different sources, for example, accessible data that comprises similar patterns yet don't need to be illustrative of the target datasets. We present one example of a transfer classifier that uses this same-and different-distribution training data for the neural network part, all followed by CNN. This means that It reuse the weights from a pre-trained network model into a new model, and then while training the model, either keep the weights fixed, fine-tuning them or adapt the weights entirely. It is pre-trained on a large related dataset and then transfer learns the knowledge from one trained model to the target model. [57] This prompts a critical positive effect on the issues where restricted datasets are accessible, such as in our situation and in numerous other automatic target classification problems based on radar systems.

4.5.2. **Neural Networks**

The architecture of Neural Networks is inspired by the functionality of the human brain. Brain takes the input, on that input it does some processing, and based on that it gives an output. The neural network exactly works this way. The individual components or units in the neural network are the neurons. Two steps are vital for the complete training of the network i.e. backward and forward propagation. In forward propagation, the network takes the image as an input; these images are stored in the form of numbers at the input layer. These numbers are used to define the intensity of the image's pixels. The hidden layer performs a mathematical operation on the input layer based on certain parameters and gives a final answer which is directed to the output layer and thus, final prediction is made. When the prediction is done, the output is compared with the actual value, and error is computed. Then we check that how much deviation is there from the actual answer and based on which weights and parameters are updated. The output is generated again using the updated weights and parameters. The process is repeated until the maximum accuracy is achieved.

In the neural network, the number of neurons is huge which makes the algorithm unmanageable. For instance, if there is an image of 224 x 224 x 1 dimensions than the required number of neurons will be 50176. This is a huge number which is very complex to manage and this makes the algorithm computationally inefficient. To overcome this problem, Convolutional Neural Network [58] was made. The main difference is that CNN extract the main features of the image and then reduce its dimensions without losing the useful features. In CNN, the image is represented by a three-dimensional matrix i.e. 224x224x3. The convolutional layer extracts the important features and reduces the dimensionality by applying the convolution. The dimensionality of 224x224x3 image is condensed to 1 x 1 x 1000 which states that now only 1000 neurons are required in the first layer of feed-forward prorogation making it less complex and manageable.

4.5.3. **Convolutional Neural Networks**

Deep convolutional Neural Network is a deep learning algorithm in which the image is taken as input and different weights and biases are assigned to them which is used for the classification and differentiates one image from the other. Convolutional Neural Network consists of a sequence of

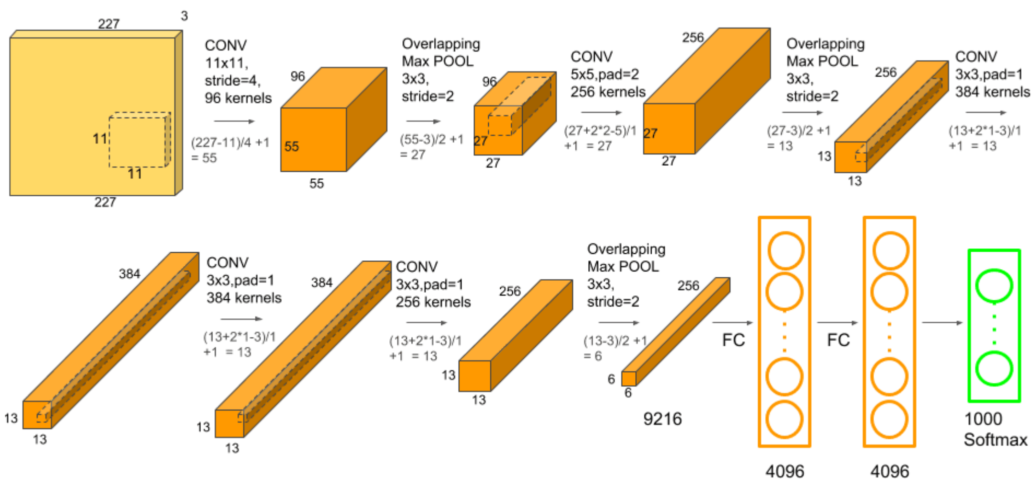


Figure 24: Architecture of CNN [56]

layers. [59]. Four main layers are used in the convolutional neural network architecture i.e. Input Layer, Rectified Linear Units layer, Convolutional Layer, Fully-Connected Layer, and a Pooling Layer. These layers are arranged in the form of a stack that forms a full convolutional neural network as shown in Figure 24. The number of layers can be increased as we go deep in the algorithm. Now we will discuss every single layer of Convolutional Neural Network in detail.

4.5.3.1. Input Layer

The input images are encoded into the color channels one of the most common is the red, green, blue (RGB) channel (W x H x D). When we have to transform the image in the pixel values than based on the intensity of these channels, we form three matrixes. Each matrix represents the intensity of the color at each point in the image. These three matrixes together form a tensor.

4.5.3.2. Convolutional Layer

Convolutional Layer is that layer where the features are extracted from the image, so it is also known as the feature extractor layer. The features are extracted by using the kernel convolution. It involves passing the kernel or filter through the image and in this way, image is transformed based on the values of the filter. The final output of this convolution is calculated using equation 5.

$$G[m,n] = (f * h)[m,n] = \sum\sum h[j,k] f[m-j, n-k] \tag{5}$$

Here, f shows input image, h shows the filter, and m, n represents dimensions of the final output matrix. If the image dimension is (n, n) and the filter dimension is (f, f) then the output dimension can be calculated using equation 6.

$$Dimension\ of\ Output = ((n-f+1),(n-f+1)) \tag{6}$$

The output of the convolutional layer is sent to the Rectified Linear Units layer (ReLU) which applies the activation function on the output. The activation function is given in equation 7. The images contain the non-linear features so this function is applied to decrease the non-linearity in the image. The function removes all the negative values from the output and keeps the positive values only.

$$f(x) = \begin{cases} x\ when\ x \geq 0 \\ 0\ otherwise \end{cases} \tag{7}$$

4.2.1. Pooling Layer

The size of the image is quite and by running our image through different filters the size of the image keeps on increasing which makes our algorithm inefficient. To overcome this problem, we will stack the pooling layer after every convolutional layer. It benefits in reducing the image size drastically by replacing the pixel with min, max, or average of the pixel. The main idea is that we will take a pixel and replace it with soma some function (min, max, average). The most common is the max function that is also called max-pooling. In the max-pooling layer, the size of the image is reduced drastically as we move the filter by multiple pixels in such a way that each pixel is seen by exactly one filter position. It uses two hyperparameters, first is the stride S, and second is their spatial extent F. The input is in the form of W x H x D and the output of this layer can be calculated using equation 8, 9, 10.

$$W_{new} = \frac{W - F}{S + 1} \tag{8}$$

$$H_{new} = \frac{H - F}{S + 1} \tag{9}$$

$$D_{new} = D \tag{10}$$

4.2.2. Fully Connected Layer

The output from the pooling layer is first transformed into a vector and then given as an input to the fully connected layer. A fully connected layer is nothing but a simple feedforward neural network. The graphical representation of ANN is shown in Figure 25.

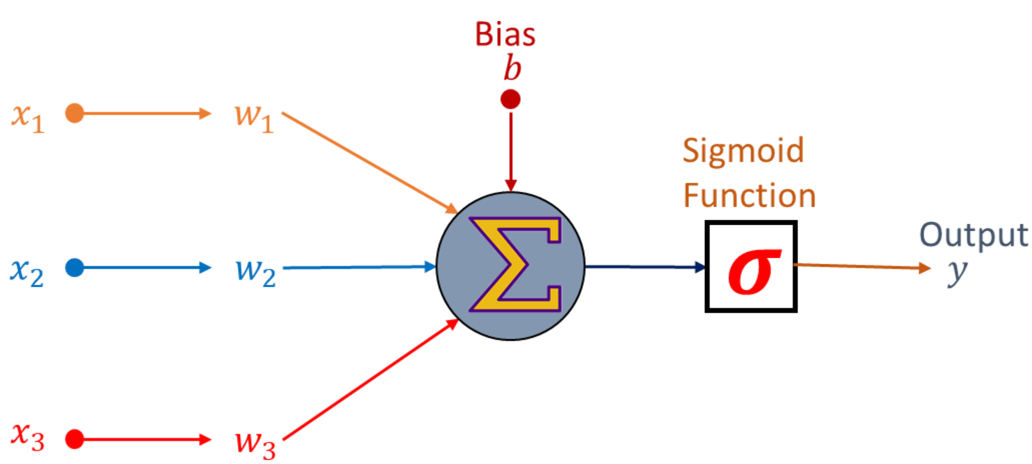


Figure 25: Graphical Representation of model

After getting the summation result, the sigmoid function is applied to the output of the summation. Equation 11 is used for calculating the sigmoid function.

$$X = \frac{1}{1 + e^{-x}} \tag{11}$$

In a fully connected layer, ReLu can also be used as the activation function. After passing from the fully connected layer the final layer used the softmax activation function which classifies input in the given classes based on the probability value. Softmax activation function. does the normalization to make sure that the sum of the output vector is equal to 1. This output vector tells that how much probability is there that the input will fall in the given classes.

4.6. MICRO-DOPPLER SIGNATURE

The CNN in our case takes 2D spectrograms as an input. The CNN layers are exploited to extract features from spectrograms and we do not further obtain hand-crafted features from the spectrograms. CNN is composed of various layers which are stacked upon each other. At the input layer, pixels are given as an input where the features are extracted and form a feature layer. Since we utilise the AlexNet architecture, so the input size 227 x 227 x 3 is dependent upon the architecture of the network. The inputs are convolved with pixels resulting in the convolutional layer output. We have used five convolutional layers followed by max pooling. In the first convolutional layer, 96 filters with size 11x11x3 are used. After the first two convolutional layers, a max-pooling layer is utilised to down sample the pixels and select the most relevant and useful features. The size of the pooling windows is 3x3 with a stride value of 2, which reduce the dimensions to 2x2. Next 3 convolutional layers i.e., third, fourth, and fifth are connected directly. The last convolutional layer is followed by a max-pooling layer and its output directly goes as in input into a series of two fully connected layers. The last fully connected layer feeds the output into a softmax classifier for multi-class classification. In fully connected layers, ReLU is used as the activation function to provide the normalization operation. Figure 8(b) gives the confusion matrix for the training dataset. We achieved a classification accuracy of 99.1% with only two samples misclassified.

4.7. SUMMARY

The proposed solution has been explained in detailed in the above paragraphs. We proposed frequency–time spectrogram feature maps for training of the deep learning model normalized on speed profiles of individuals to provide robust classification independent of age groups. Our engineered deep CNN has higher layers than existing state-of-the-art solutions and achieves the highest accuracy through transfer learning on the proposed feature maps for classification of activities including walking, sitting down, standing up, picking up objects, fall and drinking water. We have utilized last two full connected layers for transfer learning and achieved above 99% results for activity detection. Furthermore, we presented a robust generalized solution for detection of activities, which provides better generalization for geographical locations because of the diversity of dataset utilized for training.

CHAPTER 5. RESULTS AND DISCUSSION

This section discusses the results and analysis of data collected. The classification methods are divided into two main parts. Firstly, two conventional machine learning classifiers, namely Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) were used to classify activities. Secondly, transfer learning technique was used to extract features with one CNN (AlexNet) and train a machine learning/deep network model on the same features for classification.

5.1. Performance Metrics for Classification

In order to examine the performance of different classification algorithms, the performance metrics can be obtained from the true classification and misclassifications in a confusion matrix, as described in Table 6, as an example for a simple binary classification. The elements in diagonal show the correct classification for a specific class (class A in this case) and are termed as a True Positive and when the remaining classes which are not of interest are also correctly identified (Class B). The two terms namely False Positive and False Negative indicate ‘false alarm’ and ‘missed identification,’ respectively.

Table 6: Performance Metrics

True/Predicted	A	B
A	True Positive	False Negative
B	False Positive	True Negative

5.2. Classification Results

This section presents the classification results obtained using different methods. Each of the method is discussed in details as follows:

5.2.1. Classification using Machine Learning

Initially, we combined all available data sets (healthy individuals’ datasets from university lab environment and mature people datasets from residential and service care centers) and performed classification tasks to obtain some baseline results. Two conventional machine learning classifiers namely SVM and KNN were considered for classification of activities. The SVM uses features to produce a hyperplane margin that is based on the distribution of set of features for a particular class. This algorithm is already extensively used for human activity recognition in indoor settings and has been compared with other classifiers (Riazul Islam et al., 2015). The second classifier i.e. KNN is a non-parametric technique used for classification tasks. It compares the distance between an input test sample and the k nearest training samples in its features space, performing a majority vote between the closest neighboring points to assign the test sample to a specific class as shown in Figure 26. The training, validation, and testing processes were implemented using MATLAB. Different ways of calculating the distance between points or vectors in features spaces can be used, starting from the simplest Euclidean distance. The datasets obtained using FMCW radar for all nine locations and 1453 observations were divided into 70% (1017 observations) for training and 30% (436 observations) for testing per class. The undesired biases found in the results are minimized using this deterministic method that would be encountered in imbalance datasets between training and test classes. Initially, we have selected the optimum features such as mean,

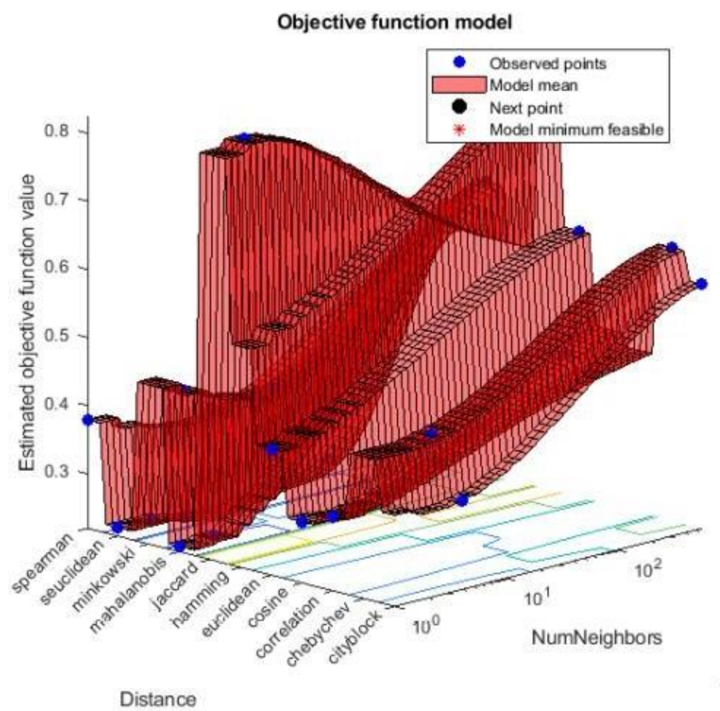


Figure 26: Objective function model

root-mean-square, median, skewness, variance, standard deviation, and kurtosis of Centroid and Bandwidth extracted from micro-Doppler signatures for each activity.

The KNN classification algorithm with optimized hyperparameters was obtained from estimated objective, distance and total number of neighbor's functions as in Fig. 26.

The distance function namely, ‘Mahalanobis’ and the 10 nearest neighbors were used as optimum hyperparameters, where the value of the estimated objective function was 0.1971 and the estimated function evaluation time was 0.056 seconds. We have used ‘holdout cross-validation’ that splits the data into training and test parts where there were no common data points between the two.
An example of a confusion matrix obtained from the 11th iteration for six human activities for all nine locations using KNN is shown in Table 4. A1, A2, A3, A4, A5, and A6 refers to walking, sitting down, standing up, pick up object. The test accuracy, in this case, is nearly 86%, where there was misclassification between picking up object from the ground and pick up glass from the table to drink water. These were similar activities; that is why the classifier was not able to fully discriminate between human actions. The average accuracy for all iterations was 81.18%, drink water and fall event, respectively. The comparable confusion matrix for the KNN is shown in Table 7.

Table 7: Confusion Matrix for KNN

Actual/Predicted	A1	A2	A3	A4	A5	A6
A1	47	0	0	0	1	0
A2	0	47	0	1	2	0
A3	0	0	45	1	3	2
A4	2	2	1	41	0	0
A5	0	2	4	7	43	1
A6	1	0	0	0	2	26

Figure 7 (orange) shows the percentage accuracy of KNN and SVM classification algorithms when training and testing were performed for 30 number of iterations. In this scenario, the training and test data (unseen to the classifier) were divided 30 times, and both procedures were performed using the same optimized hyperparameters.

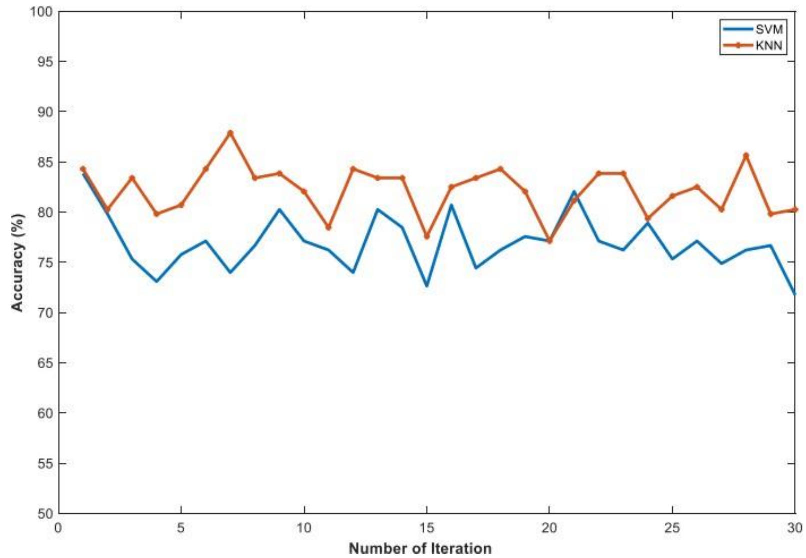


Figure 27: Test accuracy for 30 iterations with combined datasets

Table 8 describes the confusion matrix for SVM algorithm. In this case, radial basis kernel function was used as the features were linearly non- separable after they were mapped to a high dimensional feature space. The maximum accuracy obtained was nearly 80% for iteration number 4, and the average accuracy was 77.71% as in Fig. 7 (blue). The misclassification rate between activity 4 (picking up object) and activity 5 (drinking water) was higher than KNN classifier.

Table 8: Confusion Matrix for SVM

Actual/Predicted	A1	A2	A3	A4	A5	A6
A1	50	0	0	0	1	0
A2	0	47	0	1	0	0
A3	0	0	49	2	2	0
A4	0	1	1	25	9	0
A5	0	2	1	22	36	2

5.2.2. **Classification using Convolutional Neural Network**

This section discusses the results of our proposed technique. For data classification, we utilized the pre-trained AlexNet architecture and transfer learned it on frequency-time feature maps representing the Doppler shift for different activities. The frequency-time feature maps show the signal spectrogram in different colors. The blue color in spectrograms represents no movement and light green represents stationary background. The red color represents higher speed movements and yellow color represents moderate to low movement. The colors represent different movements and speeds and are utilized for training the deep learning network for classification of activities. Furthermore, we directly train a CNN with 5 convolutional layers, two fully connected layers and one SoftMax layer to compare with the transfer learned AlexNet and demonstrate the effectiveness of our proposed transfer learned model. A 5-fold cross validation technique is utilized in which the data is divided into 5 folds, 4 folds for training and 1-fold for testing. The process is repeated until all the data in 5 folds is used for testing. The final results are obtained by averaging over all the 5 folds. The multi-classification performance of 6 activities including falls is illustrated with the normalized confusion matrices in Figures 28(a) and 28(b). The overall multi-class accuracy of the proposed method is 98.4%, in comparison to a directly learned CNN which provides 79.5% multi-class accuracy. The proposed technique provides a fall classification accuracy of 99.1% with two fall samples misclassified as "standing up". This is due to the resemblance between the spectrograms of "standing up" and fall as shown in Figure 22. In comparison, the directly trained CNN provides a fall accuracy of 82.4% with misclassification of falls as sitting down, standing up and picking up activities. Furthermore, the proposed model achieves a high classification accuracy for "picking up object" and "drinking water" at 96% and 98.5%, respectively. The directly trained CNN provides relatively lower classification accuracy of 78.1% and 80.6% for "picking up object" and "drinking water", respectively. The spectrograms of both the activities "picking up object" and "drinking water" are nearly similar (Figure 22) and result in some misclassification between the two. The "walking and "drinking water" activities achieve a high classification accuracy of 100% and 99.7%, respectively as compared to the trained CNN, where the accuracies remain relatively low at 83.8% and 80.6% for "walking" and "drinking water", respectively. Our proposed system generates false positives for standing up, picking up and drinking activities. However, the system does not generate false positives for falls as illustrated by the confusion matrix in Figure 28(b). However, the system produces false negatives for falls by classifying 0.9% of the falls as standing-up activity. The inference latency for our network with 5 convolutional layers network is < 1ms.

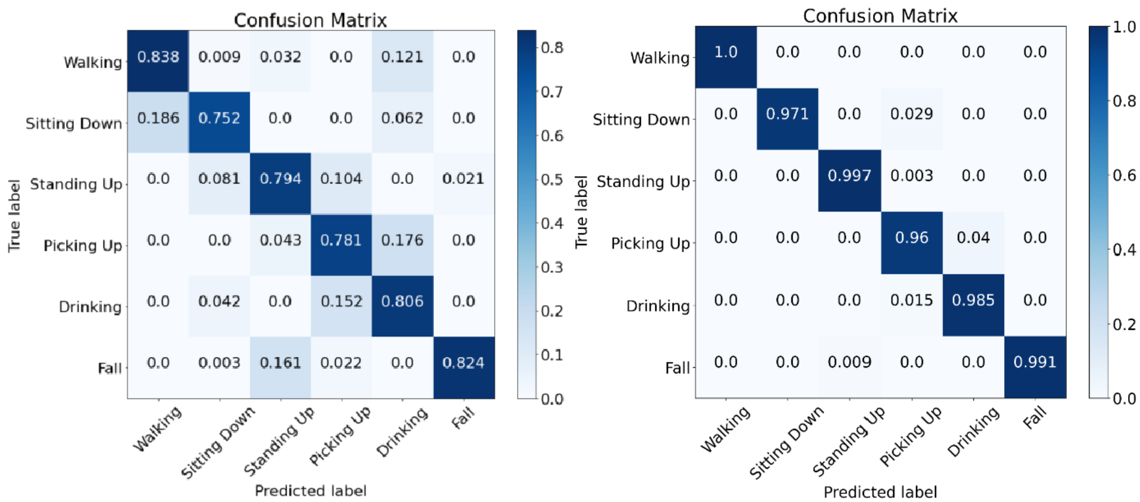


Figure 28: Normalized Confusion Matrix of (a) CNN and (b) Transfer Learned AlexNet.

Table 9, provides the comparison of state-of-the-art techniques with our proposed system and model. All the techniques use CNN except for [60] and [61], which utilize CNN with LSTM and recurrent auto encoder (RAE), respectively. Our proposed model utilizes a deeper CNN model as compared to the proposed CNN models and the RAE network. We achieve the highest classification accuracy of 99.1% for falls and an overall multiclass accuracy of 98.4%. The dataset sizes in [62] and [61] are much smaller at 201 and 50 (fall) samples. We have 6.4% higher accuracy and 8.3% higher specificity than [62]. While [61] has the second highest accuracy of 98%, the value is obtained for only 50 fall samples and the result suffers from low confidence and lacks reliability. Authors in [60] use overall 3060 samples, but utilize 600 samples for testing. We utilize bigger dataset for testing through 5-fold validation technique, which includes all the 1436 samples and have 6.1% higher accuracy than [60]. Furthermore, the problem is treated as a binary classification problem in [63], and provides lower accuracy than both our multi-classification and fall results.

Table 9: Comparison of the state-of-the-art techniques with the proposed system.

Reference	IEEE Trans C&S’19 [60]	ArXiv'20 [61]	IEEE Sensors'20 [62]	Sensors'20 [63]	Proposed Classification Model

Radar	Ultra Wideband	mmWave FMCW	FMCW	Ultra Wideband	FMCW
Frequency	5.9 - 10.3 GHz	77 GHz	2.425 GHz	4 GHz	5.8 GHz
Bandwidth	-	1.92 GHz	250 MHz	1.7 GHz	400 MHz
Dataset Size	201	50 (falls)	4000	3060	1453
Classification	Multiclass	Multiclass	Binary	Multiclass	Multiclass
Algorithm	CNN	RAE	CNN	CNN-ConvLSTM	Deep CNN
Sensitivity (%)	93.4	-	-	95.0	99.2
Specificity (%)	91.7	-	-	92.6	100
Accuracy (%)	92.7	98.0	95.0	93.0	99.1

CHAPTER 6. CONCLUSION & FUTURE WORKs

This paper presented a micro-Doppler radar-based system for multi-classification of human activities. We proposed frequency–time spectrogram feature maps for training of the deep learning model normalized on speed profiles of individuals to provide robust classification independent of age groups.

- Our engineered deep CNN has higher layers than existing state-of-the-art solutions and achieves the highest accuracy through transfer learning on the proposed feature maps for classification of activities including walking, sitting down, standing up, picking up objects, fall and drinking water. We have utilized last two full connected layers for transfer learning and achieved above 99% results for activity detection.
- We presented a robust generalized solution for detection of activities, which provides better generalization for geographical locations because of the diversity of dataset [46] utilized for training. For better generalization, we utilized a diverse publicly available dataset at
- [46] from nine different locations and 99 volunteer participants of different age groups.
- Our solution provides the highest accuracy, sensitivity and specificity for different activity at 99.1%, 99.2% and 100%, respectively than existing solutions with an overall multi-class accuracy of 98.4%.
- The proposed solution is deployable to new environments and has good generalization capability for robust and reliable classification independent of the age groups.

In future, we will extend our work to higher ranges by utilizing multiple radar sensors operating at different frequencies to avoid interference between respective sensors. We will also increase different activities along with different kinds of postures. We intend to perform multiclassification for different kinds of other activities as well.

REFERENCES

[1] J. Sunny, S. George and J. Kizhakkethottam, “Applications and Challenges of Human Activity Recognition using Sensors in a Smart Environment,” International Journal for Innovative Research in Science & Technology,vol. 2 , no. 04, 2015.

[2] A. Das Antar, M. Ahmed and M. A. R. Ahad, "Challenges in Sensor-based Human Activity Recognition and a Comparative Analysis of Benchmark Datasets: A Review," 2019 Joint 8th International Conference on Informatics, Electronics & Vision (ICIEV) and 2019 3rd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Spokane, WA, USA, 2019, pp. 134-139, doi: 10.1109/ICIEV.2019.8858508.

[3] G. Bhat, R. Deb, V. V. Chaurasia, H. Shill and U. Y. Ogras, "Online Human Activity Recognition using Low-Power Wearable Devices," 2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), San Diego, CA, USA, 2018, pp. 1-8, doi: 10.1145/3240765.3240833.

[4] S. Ranasinghe, F. Machot and H. Mayr, “A Review on Applications of Activity Recognition Systems with Regard to Performance and Evaluation.” International Journal of Distributed Sensor Networks, 2016.

[5] E. Bulbul, A. Cetin and I. A. Dogru, "Human Activity Recognition Using Smartphones," 2018 2nd International Symposium on Multidisciplinary Studies and InnovativeTechnologies (ISMSIT), Ankara, Turkey, pp. 1-6, 2018.

[6] A. K. Muhammad Masum, A. Barua, E. H. Bahadur, M. R. Alam, M. Akib Uz Zaman Chowdhury and M. S. Alam, "Human Activity Recognition Using Multiple Smartphone Sensors," 2018 International Conference on Innovations in Science, Engineering and Technology (ICISET), Chittagong, Bangladesh, pp. 468-473, 2018.

[7] P. Koleva, K. Tonchev, G. Balabanov, A. Manolova and V. Poulkov, "Challenges in designing and implementation of an effective Ambient Assisted Living system," 2015 12th International Conference on Telecommunication in Modern Satellite, Cable and Broadcasting Services (TELSIKS), Nis, Serbia, pp. 305-308,2015.

[8] F. Machot, Mohammed R. Elkobaisi and K. Kyamakya, “Zero-Shot Human Activity Recognition Using Non-Visual Sensors”, Sensors 2020, 20, 825.

[9] Michalis Vrigkas, Christophoros Nikou and Ioannis A. Kakadiaris,” A Review of Human

Activity Recognition Methods”, Frontiers in Robotics and AI,vol. 2,no. 28,2015

[10] https://www.cs.colostate.edu/~draper/MindsEye.php

[11] Vrigkas Michalis, Nikou Christophoros, Kakadiaris Ioannis A,” A Review of Human Activity Recognition Methods”, Frontiers in Robotics and AI,vol. 2,no. 28,2015.

[12] R. Bhardwaj, S. Kumar and S. C. Gupta, "Human activity recognition in real world," 2017 2nd International Conference on Telecommunication and Networks (TEL-NET), Noida, India, pp. 1-6,2017.

[13] A. R. Sanabria, T. W. Kelsey and J. Ye, "Representation Learning for Minority and Subtle Activities in a Smart Home Environment," 2019 IEEE International Conference on Pervasive Computing and Communications (PerCom, Kyoto, Japan, pp. 1-7,2019.

[14] A. Zunino et al., "Video Gesture Analysis for Autism Spectrum Disorder Detection," 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, pp. 3421-3426,2018.

[15] Z. Chen, C. Jiang, S. Xiang, J. Ding, M. Wu and X. Li, "Smartphone Sensor-Based Human Activity Recognition Using Feature Fusion and Maximum Full a Posteriori," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 7, pp. 3992-4001, 2020.

[16] M. Muaaz, A. Chelli, A. A. Abdelgawwad, A. C. Mallofré and M. Pätzold, "WiWeHAR: Multimodal Human Activity Recognition Using Wi-Fi and Wearable Sensing Modalities," in IEEE Access, vol. 8, pp. 164453-164470, 2020.

[17] N. S. Pathan, M. T. F. Talukdar, M. Quamruzzaman and S. A. Fattah, "A Machine Learning based Human Activity Recognition during Physical Exercise using Wavelet Packet Transform of PPG and Inertial Sensors data," 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, pp. 1-5,2019. [18]https://www.semanticscholar.org/paper/Physical-activity-recognition-using-inertial-%E2%80%94-A-of-Safi-Attal/0ad002077cb28cb8f1a9b09c931c83d00765bea0

[19] "IEEE Recommended Practice for Precision Centrifuge Testing of Linear Accelerometers," in IEEE Std 836-2001 , vol., no., pp.1-96, 2001.

[20] Ahmed, N., Rafiq, J. I., & Islam, M. R.,” Enhanced Human Activity Recognition Based on Smartphone Sensor Data Using Hybrid Feature Selection Model”, Sensors (Basel, Switzerland), 20(1), 2020.

[21] P. Knott, S. Stanko, H. Wilden, M. A. Gonzalez-Huici and J. Worms, "RADAR systems -

technology and challenges," 2017 18th International Radar Symposium (IRS), Prague, Czech Republic,pp. 1-4,2017.

[22] https://www.researchgate.net/figure/1D-2D-and-3D-radar-echoes-a-3D-time-range-Doppler- data-cube-b-2D-time-Doppler-map_fig7_332975833

[23] A. Hamid, Ossama & Mohamed, Abdel-rahman & Jiang, Hui & Deng, Li & Penn, Gerald & Yu, Dong, " Convolutional Neural Networks for Speech Recognition", Audio, Speech, and Language Processing, IEEE/ACM Transactions,2014.

[24] Nanni, Loris & Ghidoni, Stefano & Brahnam, Sheryl,"Handcrafted vs Non-Handcrafted Features for computer vision classification". Pattern Recognition. 71,2017.

[25] J. Yang, J. Lee, and J. Choi, "Activity recognition based on rfid object usage for smart mobile devices," Journal of Computer Science and Technology, vol. 26, no. 2, pp. 239–246, 2011.

[26] S. Wang and G. Zhou, "A review on radio based activity recognition," Digital Communications and Networks, vol. 1, no. 1, pp. 20–29, 2015.

[27] M. Scholz, S. Sigg, H. R. Schmidtke, and M. Beigl, "Challenges for device-free radio-based activity recognition," in Workshop on Context Systems, Design, Evaluation and Optimisation, 2011, Conference Proceedings.

[28] S. Amendola, R. Lodato, S. Manzari, C. Occhiuzzi, and G. Marrocco, "Rfid technology for iot-based personal healthcare in smart spaces," IEEE Internet of things journal, vol. 1, no. 2, pp. 144–152, 2014.

[29] J. Ma, H. Wang, D. Zhang, Y. Wang, and Y. Wang, "A survey on wi-fi based contactless activity recognition," in Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld). IEEE, 2016, pp. 1086–1091.

[30] E. Cianca, M. De Sanctis, and S. Di Domenico, "Radios as sensors," IEEE Internet of Things Journal, vol. 4, no. 2, pp. 363–373, 2017.

[31] M. Scholz, S. Sigg, H. R. Schmidtke, and M. Beigl, "Challenges for device-free radio-based activity recognition," in Workshop on Context Systems, Design, Evaluation and Optimisation, 2011, Conference Proceedings

[32] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu, "Sensorbased activity recognition,"

IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 42, no. 6, pp. 790–808, 2012.

[33] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, “Deep learning for sensor-based activity recognition: A survey,” arXiv preprint arXiv:1707.03502, 2017.

[34] O. D. Lara and M. A. Labrador, “A survey on human activity recognition using wearable sensors,” IEEE Communications Surveys and Tutorials, vol. 15, no. 3, pp. 1192–1209, 2013.

[35] M. Shoaib, S. Bosch, O. D. Incel, H. Scholten, and P. J. Havinga, “A survey of online activity recognition using mobile phones,” Sensors, vol. 15, no. 1, pp. 2059–2085, 2015.

[36] M. Vrigkas, C. Nikou, and I. A. Kakadiaris, “A review of human activity recognition methods,” Frontiers in Robotics and AI, vol. 2, p. 28, 2015.

[37] S. Herath, M. Harandi, and F. Porikli, “Going deeper into action recognition: A survey,” Image and vision computing, vol. 60, pp. 4–21, 2017.

[38] Y. Maret, D. Oberson, and M. Gavrilova, “Real-time embedded system for gesture recognition,” in IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2018, pp. 30–34.

[39] J. Landt, “The history of rfid,” IEEE potentials, vol. 24, no. 4, pp. 8–11, 2005.

[40] Y. Wu, D. C. Ranasinghe, Q. Z. Sheng, S. Zeadally, and J. Yu, “Rfid enabled traceability networks: a survey,” Distributed and Parallel Databases, vol. 29, no. 5-6, pp. 397–443, 2011.

[41] C.-H. Ko, “Accessibility of radio frequency identification technology in facilities maintenance.” Journal of Engineering, Project & Production Management, vol. 7, no. 1, 2017.

[42] S.A. Shah, A. Tahir, J. Ahmad, A. Zahid, H. Pervaiz, S.Y. Shah, A.M.A. Ashleibta, A. Hasanali, S. Khattak, and Q.H. Abbasi, “Sensor fusion for identification of freezing of gait episodes using Wi-Fi and radar imaging,” IEEE Sens. J., vol. 20, no. 23, pp. 14410-14422, 2020.

[43] K. Chetty, Q. Chen, M. Ritchie, and K. Woodbridge, “A low-cost through-the-wall FMCW radar for stand-off operation and activity detection,” Proc. SPIE, vol. 10188, 2017, Art. no. 1018808.

[44] D. Fan, A. Ren, N. Zhao, X. Yang, Z. Zhang, S.A. Shah, F. Hu, Q.H. Abbasi, “Breathing

rhythm analysis in body centric networks," IEEE Access, vol. 6, pp. 32507-32513, 2018.

[45] W. Taylor, Q.H. Abbasi, K. Dashtipour, S. Ansari, S.A. Shah, A. Khalid, and M.A. Imran, "A Review of the State of the Art in Non-Contact Sensing for COVID-19," Sensors, vol. 20, no. 19, p. 5665, 2020.

[46] F. Fioranelli, S.A. Shah, H. Li, A. Shrestha, S. Yang, and J. Le Kernec, "Radar signatures of human activities," 2019. analysis of radar micro-Doppler signatures from aircraft engine models," J. Electromagn. Waves Appl., vol 25, no. 8-9, 1069-1080, 2011.

[47] A. Tahir, J. Ahmad, S.A. Shah, and Q.H. Abbasi, "High Performance Big Data Graph Analytics Leveraging Near Memory System," In 2019 International Conference on Advances in the Emerging Computing Technologies (AECT), 2020, pp. 1-4.

[48] D. Haider, A. Ren, D. Fan, N. Zhao, X. Yang, S.A.K. Tanoli, Z. Zhang, F. Hu, S.A. Shah, and Q.H. Abbasi, "Utilizing a 5G spectrum for health care to detect the tremors and breathing activity for multiple sclerosis," Trans. Emerg. Telecommun. Technol., vol. 29, no. 10, Art no. e3454, 2018.

[49] J. Ahmad, A. Tahir, H. Larijani, F. Ahmed, S.A. Shah, A. Hall, and W. Buchanan, "Energy demand forecasting of buildings using random neural networks," J. Intell. Fuzzy Syst., vol. 38, no. 4, pp. 4753-4765, 2020.

[50] V. Nair, GE Hinton. "Rectified linear units improve restricted Boltzmann machines." In ICML, 2010.

[51] B. Monien, R. Preis, S. Schamberger, "Approximation algorithms for multilevel graph partitioning," in Gonzalez, T.F. (ed.) Handbook of Approximation Algorithms and Metaheuristics, chap. 60, pp. 60-1–60-15. Taylor & Francis, Abingdon, 2007.

[52] J.S. Khan, A. Tahir, J. Ahmad, S.A. Shah, Q.H. Abbasi, G. Russell, and W. Buchanan, "5G-FOG: Freezing of Gait Identification in Multi-class Softmax Neural Network Exploiting 5G Spectrum. In Science and Information Conference, Springer, Cham, 2020, pp. 26-36.

[53] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learn. Representations, 2015, pp. 521–534.

[54] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770- 778.

[55] F. Masood, J. Ahmad, S.A. Shah, S.S. Jamal, and I. Hussain, “A novel hybrid secure image encryption based on julia set of fractals and 3D Lorenz chaotic map,” Entropy, vol. 22, no. 3, p. 274, 2020.

[56] https://www.topbots.com/important-cnn-architectures/

[57] L. Shao, F. Zhu, and X. Li, “Transfer learning for visual categorization: A survey,” IEEE trans. neural netw. Learn. Syst., vol. 26, no. 5, pp. 1019-1034, 2014.

[58] J.A.F. Masood, S.A. Shah, S.S. Jamal, and I. Hussain, “A novel secure occupancy monitoring scheme based on multi-chaos mapping,” Symmetry, vol. 12, no. 3, p. 350, 2020.

[59] J. Ahmad, A. Tahir, H. Larijani, F. Ahmed, S.A. Shah, A. Hall, and W. Buchanan, “Energy demand forecasting of buildings using random neural networks,” J. Intell. Fuzzy Syst., vol. 38, no. 4, pp. 4753-4765, 2020.

[60] L. Ma, M. Liu, N. Wang, L. Wang, Y. Yang, and H. Wang, “Room-level fall detection based on ultra-wideband (UWB) monostatic radar and convolutional long short-term memory (LSTM),” Sensors, vol. 20, no. 4, p. 1105, 2020.

[61] F. Jin, A. Sengupta, and S. Cao, “mmFall: Fall Detection using 4D MmWave Radar and Variational Recurrent Autoencoder,” arXiv preprint arXiv:2003.02386.

[62] H. Sadreazami, M. Bolic, and S. Rajan, “Fall detection using standoff radar-based sensing and deep convolutional neural network,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 67, no. 1, pp. 197-201, 2019.

[63] A. Bhattacharya, and R. Vaughan, “Deep learning radar design for breathing and fall detection,” IEEE Sens. J., vol. 20, no. 9, pp. 5072-5085, 2020.