Republic of Iraq

Ministry of Higher Education

And Scientific Research

University of Babylon

College of Engineering

Time-Series Deep-Learning Classifier for Human Activity Recognition Based On Smartphone Built-in Sensors

A thesis

Submitted to the College of Engineering\University of

Babylon as Partial Fulfillment of the Requirements for the

Degree of Master in Engineering/Electrical Engineering/Electronics

By

Ghada Qanbar Ali

Supervised

Asst.Prof. Dr. Hilal Al-Libawy

2022A.D 1443A.H

جمهورية العراق

وزارة التعليم العالي والبحث العلمي

جامعة بابل / كلية الهندسة

قسم الهندسة الكهربائية

مصنف التعلم العميق من السلسلة الزمنية للتعرف على النشاط البشري استنادًا إلى أجهزة الاستشعار المدمجة في الهاتف الذكي

قـدمت هـذا الرسالة إلى قسم الهندسة الكهربائية – كلية الهندسة – جامعة بابل كجـزء من متطلبات الحصول على درجــة الماجستير في الهندسة /الهندسة الكهربائية /الكترونيك

قـــدمت من قبـل

غاده قنبر علي

بأشراف

أ.م.د. هلال الياوي

1443 هـ 2022م

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

وَمَآ أُغۡنِي عَنكُم مِّنَ ٱللَّهِ مِن شَيۡءٍۖ إِنِ ٱلۡحُكۡمُ إِلَّا لِلَّهِۖ عَلَيۡهِ تَوَكَّلۡتُۖ وَعَلَيۡهِ فَلۡيَتَوَكَّلِ ٱلۡمُتَوَكِّلُونَ ٦٧

صدق الله العلي العظيم

سوره يوسف الايه (67)

I dedicate this work:

To "ALLAH" Almighty who opens my mind before my eyes to guide me, enlighten my heart, and show me clearly the right way so I follow it, and show me clearly the wrong way so I avoid it.

To the person who is the source of inspiration, wisdom, knowledge, understanding, mercy, peace, and the person who encourages people to be patient, our

"Prophet Mohammad".

To the one who dosed the cup empty to give me a drop of love. To the one who harvested the thorns from my path to pave the path of knowledge for me, to the person who is always praying for me and encouraging me.

"My father and my mother"

ACKNOWLEDGEMENTS

I want to pay special thanks to my thesis advisor Dr. Hilal Al-Libawy for his guidance, answers, and patience every time that I needed him during my work, I have always been able to find the clear path to the answer. I want to acknowledge my appreciation to him for all his effort and support in helping me complete my thesis my great gratitude goes to my family, who has supported and encouraged me for all these years of spending time and effort on my research.

الخلاصة

تعتبر رعاية المسنين والتمريض والمساعدة الصحية من المجالات الرئيسية التي يمكن أن يكون التعرف على النشاط البشري (HAR) أداة مفيدة للغاية كتكنولوجيا مساعدة. ارتفع دور HAR مؤخرًا بسبب انتشار الأجهزة الإلكترونية مثل الهواتف الذكية والساعات الذكية وكاميرات الفيديو. يجب معالجة العديد من المشكلات لتحسين النظام وتحسين طريقة تفاعل البشر مع الهواتف الذكية. لا تزال الحاجة إلى خوارزمية أكثر دقة وموثوقية مجال بحث مفتوح.

في هذه الأطروحة ، تم تصميم مصنف HAR وتنفيذه باستخدام نهج التعلم العميق. يتم إجراء التجارب مع عدد كبير من المتطوعين في بيئات حقيقية لتقديم مجموعة بيانات موثوقة لبناء المصنف. يتم جمع البيانات باستخدام هاتف ذكي يعمل بنظام Android يحتوي على نوعين من أجهزة الاستشعار (مقياس التسارع والجيروسكوب) للتعرف على خمسة أنشطة بشرية تشمل المشي والجري وصعود السلالم والنزول على السلالم وأخيرًا الوقوف والجلوس على الكراسي في تجربة بيئة حقيقية مع مساعدة 50 متطوعا. تم إنشاؤه باستخدام خوارزميات الشبكة العصبية التلافيفية ثنائية الأبعاد (CNN) عن طريق تحويل السلاسل الزمنية إلى صور باستخدام التحويل المويجي المستمر. تم تقديم مجموعة من التجارب للتحقق من صحة هذا العمل ، وتم توضيح فعالية الإطار المقترح. تتم مقارنة نتائج النموذج المقترح بأحدث الأبحاث. وجاءت المقارنة لصالح نتائج هذا العمل ، حيث أظهرت أن أعلى دقة بلغت 99.6٪.

Chapter One

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Abstract

Eldercare, nursing, and health assistance are the main fields that Human Activity Recognition (HAR) can very helpful tool as assistive technology. The role of HAR has risen recently due to the proliferation of electronic devices such as smartphones, smart watches, and video cameras. Several issues must be addressed to improve the system and improve the way humans interact with smartphones. The need for a more accurate and more reliable algorithm is still an open research area.

In this thesis, the HAR classifier is designed and implemented using a deep learning approach. Experiments with a large number of volunteers are conducted in real environments to offer a reliable dataset for classifier building. Data is collected using an Android smartphone that contains two types of sensors (accelerometer and gyroscope) to identify five human activities including walking, running, walking up stairs, walking down stairs, and finally standing and sitting on chairs in a real environment experiment with the help of 50 volunteers. It is built using two-dimensional Convolutional Neural Network (CNN) algorithms by converting time series into images using continuous wavelet transformation (CWT). A set of experiments were presented to validate this work, and the effectiveness of the proposed framework was demonstrated. The results of the suggested model are compared to the state-of-the-art research. The comparison was in favor of the results of this work, showing that the highest accuracy was 99.6 %.

List of Contents

Abstract I

List of Contents II

List of tables V

List of figures VI

List of abbreviation VIII

Chapter One 1

General Introduction 1

1.1 Introduction 1

1.2 Human activity recognition using smartphone sensors 1

1.3 Motivation 5

1.4 Literature Review 5

1.5 Problem statement 10

1.6 Contribution of Research 10

1.7 Thesis Organization 11

Chapter Two 12

Theoretical Background 12

2.1 Introduction 12

2.2 Human Activity recognition based on the inertial measurement unit 12

2.3 Experiment reliability and validity 14

2.4 Deep learning 15

2.4.1 Recurrent Neural Network (RNN) 20

2.4.1.1 Long short term memory (LSTM) 22

2.4.1.2 Gate Recurrent Unit (GRU) 25

2.4.2 Convolutional Neural Network (CNN) 25
2.4.2.1 Architecture of a convolutional neural network 27
2.4.2.2 Human activity performance evaluation and classification algorithms 32
2.5 Time series 35
2.6 Inertial Measurement Unit (IMU) 37
2.6.1 Accelerometer 41
2.6.2 Gyroscope 41
2.7 Body Mass Index (BMI) 42
2.8 Wavelet 44
Chapter Three 47
Proposed Framework 47
3.1 Introduction 47
3.2 The proposed framework 47
3.2.1 Experiment setup and data collection 50
3.2.2 Data organization 55
3.2.3 Converting time series to 2D images using CWT 57
3.2.4 Convolution neural network architecture 61
Chapter Four 63
Results and Discussion 63
4.1 Introduction 63
4.2 Real Environment Dataset 63
4.3 Image generation using CWT 64
4.4 CNN model 66
4.5 Evaluation of different types of wavelet families and image color map 69
4.6 HAR classification results 74

4.7 Results of a 30 volunteers' sub-dataset 82

4.7 Discussion 88

4.8 Comparisons with previous work 89

Chapter Five 91

Conclusions and Future works 91

5.1 Conclusions 91

5.2 Future works 92

References 93

List of tables

No	Table name	Page
1.1	Example of using HAR	3
2.1	Lab environment vs. real-life environment	15
2.2	Comparison of Extract features manually and deep learned features	19
2.3	Body Mass Index and Status of Nutrition	44
3.1	Some deep learning challenges for human activity recognition	49
3.2	The characteristics of the subjects	55
4.1	Dataset characteristics	64
4.2	The hyper parameters used in the experiment	66
4.3	CNN architecture	68
4.4	Comparing three types of wavelet family using two types of color map	73
4.5	Evaluation metrics of proposed algorithm (for the accelerometer 50 volunteers)	77
4.6	Evaluation metrics of proposed algorithm (for gyroscope 50 volunteers)	80
4.7	Evaluation metrics of proposed algorithm (for both 50 volunteers)	82
4.8	Evaluation metrics of proposed algorithm (for accelerometer 30 volunteers)	84
4.9	Evaluation metrics of the proposed algorithm (for gyroscope 30 volunteers)	86
4.10	Evaluation metrics of proposed algorithm (for both 30 volunteers)	88
4.11	Comparison of our method with some state-of-art studies	90

List of figures

No	Figure name	Page
1.1	Possible Application domain by using HAR	2
2.1	Relationship between AI, ML and DL	16
2.2	Deep learning vs. traditional machine learning	18
2.3	Examples of deep learning applications	19
2.4	Deep Learning models for time series	20
2.5	Simple recurrent neural network	21
2.6	Illustration of LSTM	24
2.7	The architecture of the convolutional neural network	27
2.8	Convolution Operation	29
2.9	Max pooling and Average pooling operation	31
2.10	Time series for recognizing human activity using the smartphone	37
2.11	The coordinate system and the process of conversion	39
2.12	Sensor modality	40
2.13	Three axes of the Accelerometer and Gyroscope in smartphones	42
3.1	Proposed system	48
3.2	A screenshot of a smartphone application Kinetics Sensor Pro	51
3.3	Activities performed during data collection in real world conditions	54
3.4	Data sorting stages	56
3.5	Conversion of all activity from three-dimensional to one-dimension using Signal Vector Magnitude	59
3.6	Stages of converting time series to images using CWT	60
3.7	CNN workflow	62
4.1	Three different types of wavelet family for running activity	65
4.2	CNN architecture for proposed system	67
4.3	Confusion matrix Amor wavelet for the accelerometer (30 volunteers)	70
4.4	Confusion matrix Morse wavelet for the accelerometer (30 volunteers)	71
4.5	Confusion matrix Bump wavelet for the accelerometer (30 volunteers)	72

4.6	Relation between accuracy and epochs (above) and relation between loss and epochs (below)	75
4.7	Confusion matrix for the accelerometer (50 volunteers)	76
4.8	Relation between accuracy and epochs (above) and relation between loss and epochs (below)	78
4.9	Confusion matrix for gyroscope (50 volunteers)	79
4.10	Confusion matrix for accelerometer and gyroscope (50 volunteers)	81
4.11	Confusion matrix for the accelerometer (30 volunteers)	83
4.12	Confusion matrix for gyroscope (30 volunteers)	85
4.13	Confusion matrix for accelerometer and gyroscope (30 volunteers)	87

List of abbreviation

Abbreviation	Definition
AI	Artificial Intelligence
BMI	Body Mass Index
CNN	Convolutional Neural Network
CWT	Continuous Wavelet Transform
DL	Deep Learning
FC	Fully Connected
FN	False Negative
FP	False Positive
GRU	Gate Recurrent Unit
HAR	Human Activity Recognition
IMU	Inertial Measurement Unit
LSTM	Long Short-Term Memory
MEMS	Microelectro mechanical Systems
ML	Machine Learning
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
SGD	Stochastic Gradient Descent
STFT	Short Time Fourier Transform
TN	True Negative
TP	True Positive
UCI HAR	University of California, Irvine Human Activity Recognition
WISDM	Wireless Sensor Data Mining

Chapter One

General Introduction

1.1 Introduction

Learning about the activity of humans is a significant area of research computing ubiquitous, the analysis of human behavior, as well as interaction of human-computer. Recognizing daily activities is critical to maintaining a healthy lifestyle and rehabilitating patients. This chapter provides basic information for the identification of human activity using smartphones, which includes a diverse range of applications, the statement of motivation, and the importance of this study. Previous studies in the recognition of human activity broadly can be categorized based on the diversity of devices, patterns sensors, and data used to reveal the details of the activity. Finally, the research contribution and the outline of the remaining chapters of the thesis are discussed.

1.2 Human activity recognition using smartphone sensors

Recognition of human activity is considered an effective research area that is gaining more interest in recent years due to its growing role in many human-related sectors because it has the potential to automatically provide useful knowledge and context about a person's actions based on sensor input. Its significance is demonstrated by the wide range of fields in which it is used, including mobile computing[1], computing of context-aware[2], sports [3], the health sector especially with elderly people, and motion restricted patients which helps detect and diagnose critical conditions [4, 5], elderly care[6], and living with assisted of ambient [7]. In recent years, there has been great progress in identifying human activity using various machine learning approaches. However, traditional methods of feature extraction are the most challenging in the feature selection process. Deep learning (DL) is a promising approach in the human activity recognition research area and has overcome the feature selection problem. Because of the increased sensor devices' availability at a low cost and power, particularly those incorporated into mobile devices, and the advancement of techniques for data processing. There has been a surge in interest in the field; the aging population is a

source of interest increases in applications for health care and elderly care. HAR aims to learn about activities of human in both controlled and uncontrolled environments. Recognition of human activity with video datasets has made significant contributions it does involve some risks, of which personal data is more vulnerable; In addition, its processing is heavier. Some of the applications are summarized in the figure (1.1) with examples of applications as shown in the table (1.1).

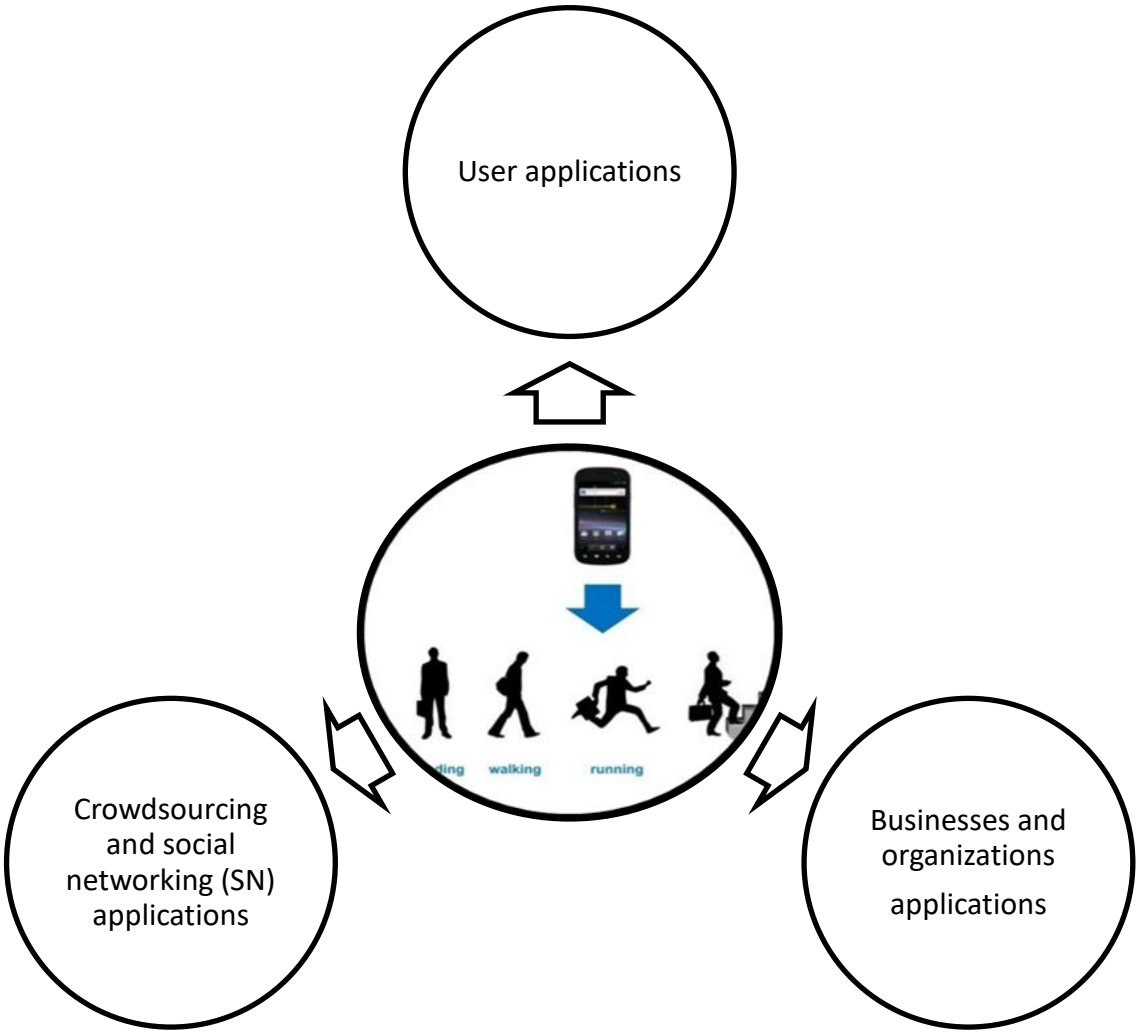


Figure (1.1): Possible Application domain by using HAR

Table (1.1): Example of using HAR

HAR application		An example
Applications of users	Tracking of physical fitness	Applications provide a history of online activity
	Monitoring of health	Instead of a single session, evaluate patients over time
	Detecting falls	Recognize falls to take appropriate action
	Context-sensitive behavior	Disable phone calls while jogging
	Automation at home/work	Smart homes expect the needs of their inhabitants
	Systems that self-manage	Turn off WiFi while jogging to save battery life
Businesses' and organizations' applications	Advertising that is targeted	Display relevant advertisements to users
	Research platform	Create a platform for gathering activity data
	Corporate management &accounting	Keep track of employee time and make sure it is spent wisely
Crowdsourcing and Social Networking (SN) applications	Conventional SNs	Inform your friends and followers about your activities
	SNs based on activity	People are connected according to their activity profiles
	Detection of place and event	Determine popular exercise and recreation areas

A mechanism for discovering postural and motor activities and body movements is provided by the framework for recognized human activities. In previous studies, different devices, sensor patterns, and data used to detect large numbers of activity information can be classified accordingly. HAR can be classified into two main types [8]: video and sensor-based on HAR. the first type analyzes images or videos that contain human movements from the camera which are among the surrounding sensors, which are environmental sensors or video cameras placed in the environment at specific points [9]. Sensor-based HAR is focused on smart sensor motion data are integrated sensors that are integrated with clothing or other medical devices or integrated into personal devices like smartphones or smart watches for example accelerometers, gyroscopes, sound sensors, Bluetooth [10-13] …. etc. On the other hand, sensor data utilization generated with smartphones as well as other wearable technology has dominated the human motion analysis research landscape, monitoring, and detection of activity because of their obvious advantages over other sensing methods[4]. Smartphones are ubiquitous, cheap to install, and easy to use. Mobile phones evolved into an integral part of our everyday lives and are easy to carry anywhere. In the age of smartphones, multi-sensor systems embedded are exploding that allow the researchers are collect signals of human physiological to monitor daily life activity, human motion analysis has become an essential part of our daily lives. Smartphones have access to a variety of sensors, including Bluetooth, accelerometers, Wi-Fi, gyroscopes, magnetometers, cellular radio sensors, and microphones …so on, which can be used to deduce activity information. Deep learning allows the models for computational with multiple layers of processing for learning data representation at various abstraction levels. These methods have significantly improved the most recent advances in speech recognition technology, optical object recognition, detection of an object, as well as numerous other fields for example genomics and drug discovery. Convolutional deep networks have made significant advances in video, image, audio processing, and speech, while recursive networks have facilitated sequential data for example speech and text.

1.3 Motivation

The motivation behind the proposed model is:

- Increasing people's need for health, fitness, sports injury detection, elderly care, and rehabilitation, under real-life conditions.
- Due to the development of modern technology for smartphones and the increase in sensors, the motivation was to facilitate work in this field.
- Deep learning is the advancement of machine learning technology, which includes multi-layer convolutional neural networks capable of automatic learning. These deep network methods have achieved great success in many difficult problems, as well as high accuracy results.

1.4 Literature Review

During the past decade, significant research work has been carried out in the field of HAR to better understand and quantify patterns of human behavior for various applications, live assisted environments, healthcare, the Internet of things, security, location, and navigation. Many sensors have been used to recognize the activity of humans such as smartphones, wearable sensors, etc. A large amount of data has been collected as a result of the rapid development of the wireless sensor network to determine human activities using various types of sensors. Machine learning algorithms require manual extraction of some distinguishing features from large amounts of sensory data. Unlike deep learning which extracts features automatically

(Bayat, Pomplun et al. 2014) [14] Explains how to use acceleration data generated by the user's cell phone to distinguish between various types of human physical activities. The system has gone through training and testing as part of an experiment with many people in real-life circumstances. It was found that employing the average odds as a method of fusion can achieve a rate of the overall accuracy of 91.15%. There are four subjects including two males and two women aged 29 - 33 years volunteered to be a part of this research study. These people, each with a cell phone, completed all six activities: slow walking, running, brisk walking, aerobic dance, stairs up, and stairs down.

(Ronaoo and Cho 2015) [15] Use the ConvNets for classifying activities centered on time series data are collected from smartphone sensors and to evaluate the various architectures. Experiments demonstrate that increasing the number of convolutional layers improves performance, however, with each additional layer, the derived features' complexity decreases. All of the experiences made use of the UCI repository's publicly available HAR smartphone dataset. Contain the data from accelerometer and gyroscope sensors for 30 people performing 6 various activities it is as follows: walking, walking upstairs, lying, walking downstairs, sitting, and standing. The accuracy was approximately 90.0% for three layers.

(Zebin, Scully, et al. 2016) [16] The research focused to use approaches of deep learning for human activity recognition where the inputs are signals of time series. Using CNN architecture for feature learning automation from the initial input to a human activity recognition task collected data from 12 healthy volunteers undertaking six common daily life activities such as, walking, walking upstairs, downstairs, sitting, standing, and lying down. The results of the experiments revealed that CNNs show a marginally better than MLP and SVM regarding classification accuracy of 97.01%, but achieved significant acceleration in terms of computational overhead CNN Training.

(Ronao, Charissa Ann, et al. 2016) [17] A deep convolutional neural network (convnet) is proposed to perform efficient and effective HAR using smartphone sensors by exploiting the inherent characteristics of activities and 1D time-series signals. Accelerometer and gyroscope triaxial sensor data were collected from 30 volunteer subjects who performed six different activities while the smartphone was in their pockets Convnets also achieved an almost perfect classification of moving activities, especially very similar ones which were previously perceived to be very difficult to classify. Dataset was collected from 30 volunteer subjects, achieving an overall performance of 94.79% on the test set with raw sensor data.

(Alsheikh, Selim, et al. 2016) [18] Focus on the use of 3-axis accelerometers Use three data sets WISDM Actitracker Dataset. The data samples belong to 29 users with 6 distinct activities of humans like walking, climbing stairs, jogging, standing, and sitting. The samples

of acceleration are collected using android mobile phones. Daphnet freeze for gait data set uses this data set to show how deep activity recognition models can be used in healthcare applications, for10 users. Skoda Checkpoint Dataset, The accuracy was obtained by (98.23%), (91.5%), and (89.38%) respectively.

(Lee, Yoon, et al. 2017) [19] A convolutional neural network in one dimension (1D) based method has been proposed to recognize the activity of humans using a 3-axis accelerometer collected data from the smartphones of users. The three activities were to be recorded by five graduate students i.e. running, walking, and remaining still, utilizing the accelerometer sensor on a smartphone. The acceleration data for x, y, and z are converted to signal vector size equation data and used as input into a 1D convolutional neural network learning algorithm. CNN-based ternary recognition of activity performance showed an accuracy of 92.71%.

(Nwankpa, Ijomah, et al. 2018) [20] Provide a survey on architectures of deep learning that use (AFs) which is an acronym for activation functions to perform various calculations between any given deep learning architecture's hidden and output layers. AFs have the potential to improve the learning of patterns in data and thus automate the process of features to detect and justify their use in hidden layers of neural networks. The benefit of this study is that it collects the majority of AFs used in deep learning and identifies current application trends, as well as the use of these functions in practical deployments of deep learning. This compilation will assist in making informed decisions when selecting the most suitable and appropriate activation function for any given application.

(Ignatov 2018) [21] The proposed approach of deep learning is to learn about human activity online using statistical features in convolutional neural networks that preserve the global feature of the time series of the accelerometer.

(Ku Abd. Ku Nurhanim, et al. 2018) [22] In this paper, classification of human daily activities using Ensemble Methods based on data acquired from smartphone inertial sensors with UCI dataset involving about 30 subjects with six different activities walking, walking upstairs, walking downstairs, sitting, standing, and lying. Five types of ensemble classifiers

utilized are Bagging, Adaboost, Rotation forest, Ensembles of nested dichotomies (END), and Random subspace. These ensemble classifiers employed a Support vector machine (SVM) and Random forest (RF) as the base learners of the ensemble classifiers. SVM produced a better accuracy rate with 99.22% compared to RF with 97.91% based on a random subspace ensemble classifier.

(Wang, Chen, et al. 2019) [8] Survey of recent advances in sensor-based activity recognition using deep learning approaches. In comparison to conventional pattern recognition techniques, deep learning can reduce reliance on man-made feature extraction and learn high-level representations of sensor data automatically to improve performance. Showing highlighted recent advances in three crucial categories starting from sensor method, application, and deep model. Also provided detailed insights into current work and suggested significant future research challenges.

(Zhou, Yang, et al. 2019) [23] Suggest a method for identifying smartphone-based activities using a convolutional neural network. Experiment results indicate that the proposed technique achieves an accuracy of approximately 98% in determining nine different types of activities, such as walk, still, upstairs, up the elevator, down the elevator, up the escalator, down the escalator, and downstairs, and turning. The dataset is collected from the participation of ten people and includes data from accelerometers, magnetometers, and gyroscopes.

(Wan, Qi et al. 2020) [24] Designs an inertial accelerometer-based architecture for HAR smartphones. Compare the benefits and drawbacks of five algorithms, including CNN, in recognizing human behavior in the UCI HAR dataset it is as follows walking, walking upstairs, sitting, walking downstairs, standing, and finally laying. The second dataset was the Pamap2 dataset which record 18 physical daily activities for 9 subjects (8 male and 1 female). The collected data include sensors such as accelerometers, gyroscopes; temperature, magnetometers, heart rate, etc. getting an accuracy of 92.71% and 91% sequentially.

(Shiranthika, Chamani, et al. 2020) [25] Presented an approach to predict human activities provided the CNN model and an LSTM model with 99.593% accuracy and 84.71% accuracy respectively for 6 daily life activities with the WISDM dataset. The use of Conv2D layers for CNN, Dropout regularization, and using perfect model hyper parameters in the networks of the two models has made them fast and robust in terms of speed and accuracy.

(Cruciani, Vafeiadis et al. 2020) [26] CNN was used as a HAR feature learning method. Two applications of CNN were taken into account: first inertial measurement unit and second HAR based on audio. On the UCI-HAR dataset, the balanced accuracy for the IMU data had a value of 91.98% and 67.51% on the extrasensory dataset collected in the real world. Regarding the audio data, The DCASE 2017 dataset had a balanced accuracy of 92.30 %, and 35.24% on the dataset extrasensory.

(Abid, Mariem, et al. 2021) [27] The purpose of the study is to investigate a human activity recognition method of accrued decision accuracy and speed of execution to be applicable in healthcare. This method classifies wearable sensor acceleration time series data of human movement the data acquired from 44 subjects wearing a single waist-worn accelerometer on a smart textile, and engaged in a variety of 10 activities, yielded an average recognition rate of 90%.

(Mekruksavanich, Sakorn et al. 2021) [28] In this work, the generic HAR framework for smartphone sensor data is proposed based on Long Short-Term Memory (LSTM) networks for time-series domains. Four baseline LSTM networks are comparatively studied to analyze the impact of using different kinds of smartphone sensor data. A hybrid LSTM network called 4-layer CNN LSTM is proposed to improve recognition performance with a high accuracy rate of 99.39%. The HAR method is evaluated on a public smartphone-based dataset of UCI-HAR.

The sensor-based approach is much better suited to environmental demands. Wearable sensors had been explored as well as proven to be feasible in the beginning stages of research on sensor-based [29]. However, sensor wearables have advantages in terms of privacy and operating area. For the first case, users are less hesitant to use each site if there is no photo or

video capture. Secondly, since the user is always carrying these sensors, it is ubiquitous and their coverage of locations is almost unrestricted. It is also portable and does not necessitate the use of fixed equipment. Wearable sensors, on the other hand, have introduced new challenges: conserving the life of the battery and reducing intrusion while gathering reliable information derived from a limited sensor. Sometimes the sensors are inconvenient for the average user (for example if they are clamped too tightly, wired, or if they must be constantly reset after wearing) and are not capable of providing a solution for the long term to monitoring activity without regularly recharging them. People will be required to keep the sensors on all of the time, which is inconvenient. With the growing popularity of smartphones outfitted utilizing several sensors, it is an appealing activity recognition solution [30].

1.5 Problem statement

- Obtaining high accuracy is a challenge for researchers to HAR classifier to provide better application performance. The application of various classification methods results in varying degrees of accuracy.
- The need for a large dataset to provide a reliable result.
- The results of laboratory experiments do not always correspond to the results of real-world experiments. The problem of determining physical activity performed by an individual based on movement tracking within a given environment is known as the recognition of human activity.

1.6 Contribution of Research

- Performance Metrics: Our main focus is on HAR based on sensors (accelerometer and gyroscope). Then data processing and analysis using a convolutional neural network, converting time series into images using continuous wavelet to get high accuracy of 99.6%
- Large data set: The system was trained and tested via a data set created with the help of 50 volunteers through five activities starting walking, running, walking upstairs, walking downstairs, and finally standing, sitting on chairs. The process uses the

smartphone's accelerometer and gyroscope sensor available in all smartphones, making the frame ubiquitous.

- Real environment dataset: An experiment set up in the real environment to identify detailed activities. The dataset is collected from many volunteers regardless of locations and volunteer ability.

1.7 Thesis Organization

This thesis is arranged as follows:

- Chapter One: This chapter provides a general introduction to the recognition of human activity (HAR) in deep learning, a literature survey of related works, and the contribution of the study.
- Chapter Two: Background theory on deep learning, CNN, and RNN is presented how these methods relate to time series and how they have been successfully applied to human activity (HAR). Also, provide formal and mathematical descriptions of each approach. The tools and software used in this study are described.
- Chapter Three: provide a full description of all our methods including many of the basic classifiers and proposed network architectures of multilayer CNNs. It started with the collection of volunteers and the procedures used to collect the data set that will be used in the experiment, then present the data pre-processing and select it to be ready for the classification process.
- Chapter Four: The experimental results are detailed and discussed in this chapter. The performance of the model is evaluated and the results are compared with previous studies in the same field.
- Chapter Five: This chapter includes conclusions and recommendations for future work

Chapter Two

Theoretical Background

2.1 Introduction

A recent development technique is to try, whenever possible, to make all of our environment's devices smart. One of the objectives of developing smart systems is to learn about human activities daily using smartphones to encourage a safe and people who are concerned about their health should live a healthy lifestyle. This chapter is separated into two parts. Started with the first section explains a method for automating features from a smartphone dataset for a HAR mission using an algorithm of deep learning the convolution neural networks. The proposed method builds a new architecture for CNN deep learning to deal with data of time series. It uses convolution and aggregation to identify prominent patterns in sensor signals at various time scales. The second section is about the sensors in a smartphone used to collect data from people to build this system.

2.2 Human Activity recognition based on the inertial measurement unit

Recent advances in the recognition of human activity have enabled a myriad of applications such as healthcare [31], smart homes[32], and manufacturing improvement [33]. Mobile phones have grown in popularity an indispensable part of our everyday lives. This is because of the increased sophistication and integration of high-performance sensors, large storage capacity, as well as continuous connectivity between mobile devices. In daily activities, people are constantly interacting with compact smartphones. As a result; there has been an increase in research into extracting knowledge from data obtained by sensors installed in mobile devices [34]. Powerful sensors are built into mobile devices, such as GPS, accelerometers, gyroscopes, magnetometers, ambient light… so on. Researchers can collect sensor data more easily with smartphones. Modern techniques of deep learning are used to discriminate and identify the activities of humans according to the data collected. A smartphone can assist in resolving the problem of keeping a detailed record of the activities daily of the user. Advances in deep learning and feature selection methods when combined

with a variety of sensors, could push the limits of deeper levels of human activity recognition. Although there are still many challenges in this field, such as smartphone orientation [35] and location, user demographics…etc. The majority of smartphone users tend to hold the phone in their hands. Activity recognition is required for humanity because it uses data to record people's behaviors that enables computing systems to analyze, monitor, and assist their daily lives. Because of the privacy concerns associated with putting cameras in our private spaces, the systems sensor-based have taken precedence in applications for monitoring activities' daily lives. Furthermore, with the increasing proliferation of sensors as a result of smart device proliferation and the Internet of Things; the sensors can be integrated into smart devices like watches and phones. Sensors record human movement information without interruption and interference. Recognizing human activity is seen as a very critical problem currently, filled with significant research challenges. It has a broad range of applications, including medical care, public health, and personal monitoring. Technical facilities based on inertial measurement units are used to capture HAR sensor-based raw data such as gyroscopes and accelerometers. An accelerometer is a device that measures an object's perceived physical acceleration. It's been used in a variety of applications in medicine, science, industry, and engineering including measuring machine vibrations, high-speed vehicle acceleration, and all matters HAR. The most common sensor used to read the body movement signals is the accelerometer [36]. Its working principle typically consists of a displaced seismic mass concerning the acceleration it experiences, the displacement then is converted into an electrical signal that can be measured. This phenomenon has been used in the advancement of micro-electromechanical (MEMS) sensors for systems; their technology enables the fabrication of nanoscale devices made of semiconductors. It has an advantage over other sensing technologies in that it can be mass-produced at a low cost. Most MEMS accelerometers function as capacitive sensors consisting of a beam with a cantilever with a mass of proof whosoever's deflection is related to the sensor's experience with acceleration [37]. The direction and magnitude of acceleration can be quantified as a vector quantity as determined by an orthogonal arrangement of the sensors in the three dimensions spatially. This can also be constructed on a single chip, which is becoming more common to find 3-axis accelerometers in many electronic devices. A gyroscope is a sensor that measuring direction. It has been used in a variety of applications such as systems of inertial navigation, air vehicles

to increase stability (e.g. in quadcopters), It has also recently been introduced into electronic devices (for example game consoles, smartphones) to improve gaming experience and user interfaces. This sensor has occurred used in various applications of HAR, including activity detection (e.g. walking, climbing stairs). Gyroscopes were also produced using MEMS techniques. Gyroscopes sensors can measure direction indirectly, calculate angular velocity instead which can be combined in time to get direction. However, it is first required to have an initial corner position referenced to achieve this. These sensors are also very susceptible to noise which can cause measurement deviation. In addition to the advancement of technologies, nowadays; participants are required to carry a smartphone to collect data. Almost all the data collected from these smart devices can be used to learn about human activities based on deep learning.

2.3 Experiment reliability and validity

Data collection from various sources is the foundation of scientific research, and it is used to reach new conclusions or to corroborate previous findings. There are numerous data collection techniques and methods available; there is a distinction between these methods in terms of how they are carried out as well as the accuracy, validity, and quantity of collecting of information. Some methods of data collection are possibly more appropriate in certain situations but may not be the best, and thus, if used, may produce inaccurate or useful research results. The distinction is based on whether the research is carried out inside or outside the laboratory, as it has a significant impact on the validity of the data collected, and analyzed, as well as the main results of the problem of research. Field research, which is conducted in an environment of reality or nature, as well as controlled laboratory research, are the two types of research. Field research has a significant advantage in its applicability to real-world situations because it represents the wide range of situations and environments that people encounter in their daily lives [38]. However, the results imply that field experience is better suited for larger-scale research [39, 40]. Table (2.1) shows the difference between the laboratory environment and the real-life environment[41].

Table (2.1): Lab environment versus real-life environment

Laboratory environment	• collaborative work • time and place restrictions • limited data
Real-life environment	• lack of collaboration • no time and place restrictions • Unlimited data

Big data provides numerous opportunities in a variety of fields. However, it raises several difficult issues in terms of processing information, because of its characteristics of high volume, wide variety, quickness, and honesty [42]. Bigger data leads to more accurate predictive models[43]. It performs significantly better rather than individual classifiers[44]. It appears that (small) samples of experimental were a failure in comparison to large samples [45]. Deep learning has become increasingly important for solutions to big data analytics.

2.4 Deep learning

For many years, adopted machine learning approaches, which is an application of artificial intelligence (Artificial Intelligence is defined in short as the ability to learn and adapt like a human being) that includes algorithms that analyze data, and learn from that data, so the algorithms can learn without being explicitly programmed. It largely focuses on manually designed features, from fundamental statistics to broad methods for reducing dimensions to various domain-specific metrics. Features are input to most deep learning (DL) classifiers; it is a subset of machine learning (ML) that builds algorithms in layers to create an (Artificial Neural Network) that can learn from a huge amount of data and make intelligent decisions on its own, as shown in figure (2.1). It is based on patterns of information processing discovered within the human brain. Deep learning does not necessitate the use of any human-created rules to function; instead, it makes extensive use of data to assign specific labels to the given input. DL is built with multiple layers (Artificial Neural Networks algorithms); each provides an interpretation of unique data fed to the layer to complete the classification task.

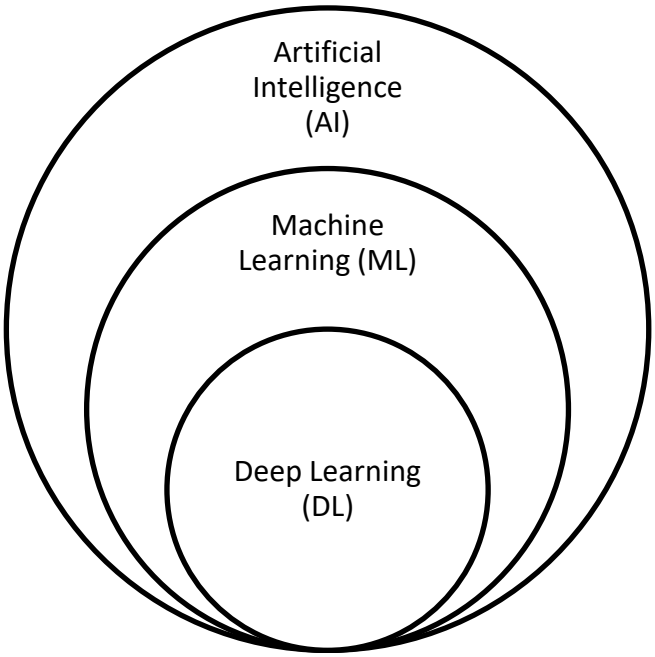


Figure (2.1): Relationship between AI, ML, and DL

Using traditional ML techniques necessitates a series of steps, including operations, starting from preprocessing, and extraction of feature, selection of prudent feature, learning, and finally classification. Furthermore, the selection of features has a significant influence on the performance of techniques machine learning. Feature selection that is biased may result in wrong class discrimination. Deep learning can automate the learning in terms of feature sets for a variety of tasks, conversely, conventional machine learning methods [46, 47]. Deep learning allows for the completion of feature extraction and classification in one shot figure (2.2). Generally, two approaches to extracting features are derived from raw data sensors. The first is to handcraft features based on expertise in this domain [48] Deep networks, on the other hand, are capable of automatically learning features. [17, 49, 50] . In HAR applications, features hand-crafted have had a great deal of success [49, 51]. Deep learning methods have been developed and used in HAR over the years to automatically learn HAR features [52, 53]. Table (2.2) shows the difference between manually extracted features and deep features. Despite these feature generation methods having been successfully used in a variety of settings, they typically involve a time-consuming design phase and necessitate the engineer developing knowledge concerning the types of data patterns that capture pertinent information. Machine learning systems can only fill specific roles and are incapable of exploiting subtle

patterns, but these patterns are important, which is why researchers are looking for inspiration in neural networks. Neural networks (artificial), which were inspired in part by biological networks, have been used for many years and have proven to be effective in a broad range of machine learning problems [54]. These networks are made up of several interconnected simple computational units known as artificial neurons. Networks with shallow depths, usually consisting of two layers of densely artificial neurons linked, have success but some types of problems may necessitate manual feature engineering, such as general object recognition in computer vision. The deep networks, which are distinguished by the use of multiple layers, have also existed in a wide range of contexts [46, 55]. Similar to the way biological network information processing, deep networks can learn from a feature hierarchy in which each layer represents a level of features. However, the use of deep networks in practice has proven to be challenging because they contain large parameters number as a result, optimizing weights for deep networks is difficult [56]. Many advances to deep networks have made them more efficient and practical use. The researchers note biological neural networks frequently contain multi-layered, specific, sparse, and repetitive connection patterns. By representing the connection for artificial neural network patterns, typically in a very simplified form, free parameters number can be minimized in addition to encouraging the network to learn strong general representations. Also, the availability of a lot of high-quality datasets for deep networks with multiple layers and free parameters allows the learning of complex representations that generalize well. Currently, various DL applications are being used all over the world. The applications of DL include healthcare, analysis of social network, processing of speech and audio (such as enhancement and recognition), methods for visual data processing (like computer vision and analysis of multimedia data), and natural language processing (sentence classification and translation), and many more applications as shown in figure (2.3) [57-59] . This application can be classified to five of categories: classification, detection, localization, registration, and segmentation; despite the fact that each of these tasks has a specific goal. Finally, improvements have been made in optimization algorithms and computers, especially traceable processors for training large networks on large data sets [60]. Currently, deep networks maintain standard performance on several standardized data sets and are generally considered to be among the most recent machine learning tasks [46].

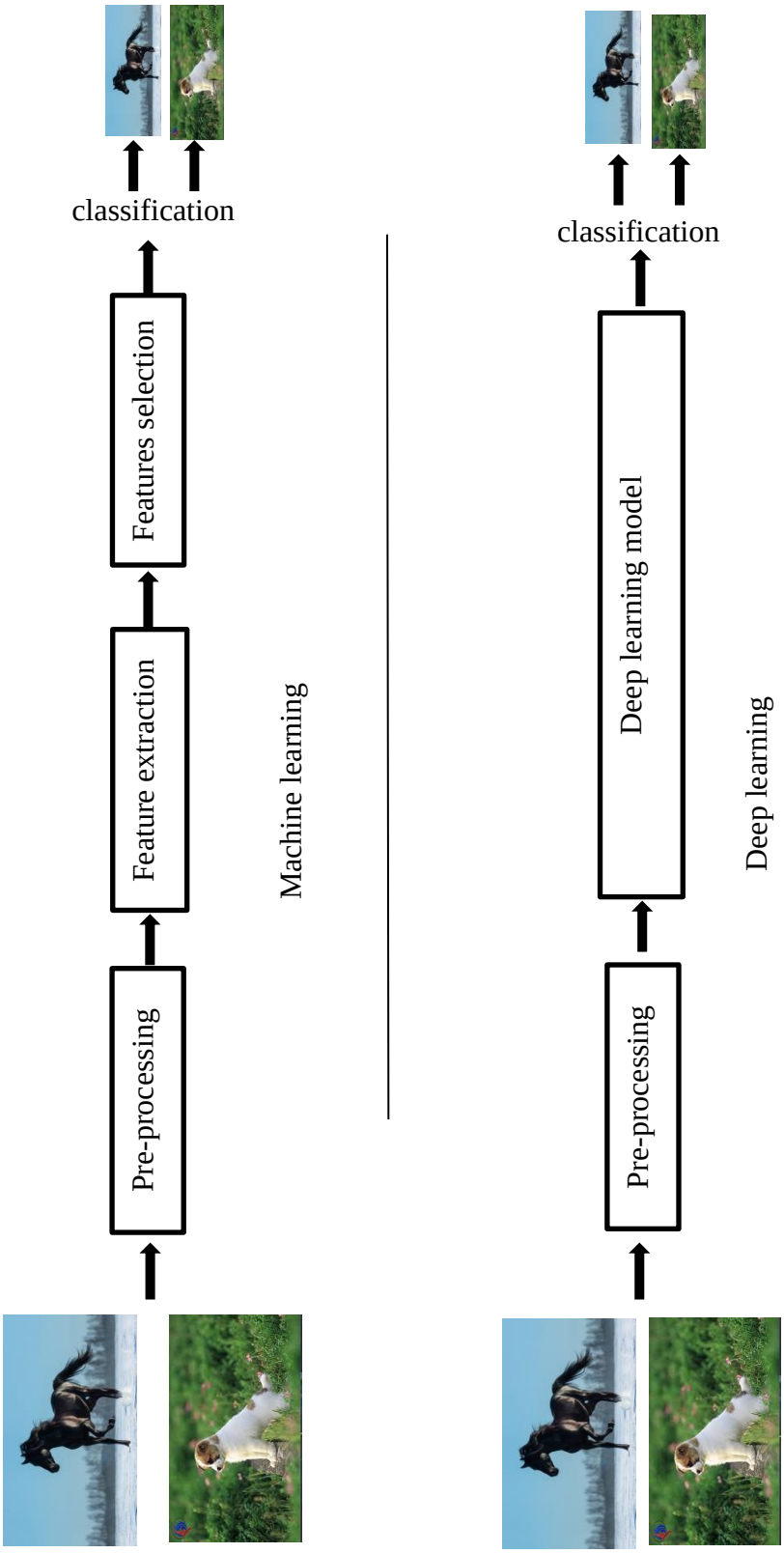


Figure (2.2): Deep learning vs. traditional machine learning

Table (2.2): Comparison of extract features manually and deep features

Type of feature	Advantages	Disadvantages
Extract features manually	1- The physical meanings of the features are easy to understand. 2- Extraction is efficient and simple to implement. 3- Work well for many ranges of activity recognition issues.	1-Required domain knowledge. 2-Specific sensor type. 3-More feature selection is required.
Deep features	1-No domain knowledge is required. 2-Learning features from raw data automatically. 3-More robust and generalized features.	1-A large number of computing resources. 2-It is difficult to adjust parameters 3-Less is interpretable for the learned features.

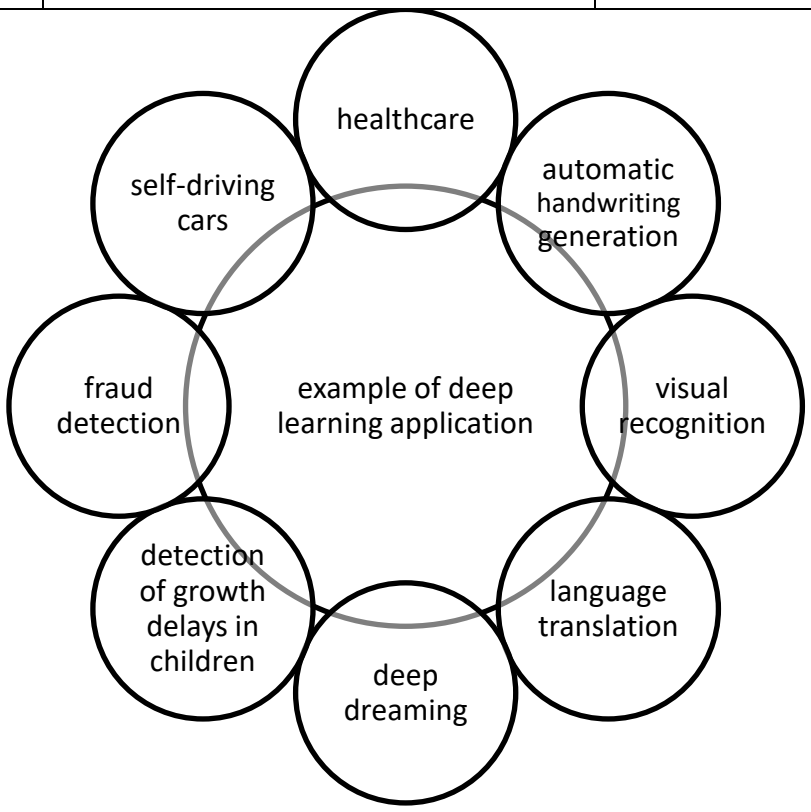


Figure (2.3): Examples of deep learning applications

The most important algorithms for deep learning that deal with time series are convolution neural networks and recurrent neural networks as described in figure (2.4). For better results with time series, designed a model of deep learning for the neural network using a convolution neural network (CNN) algorithm.

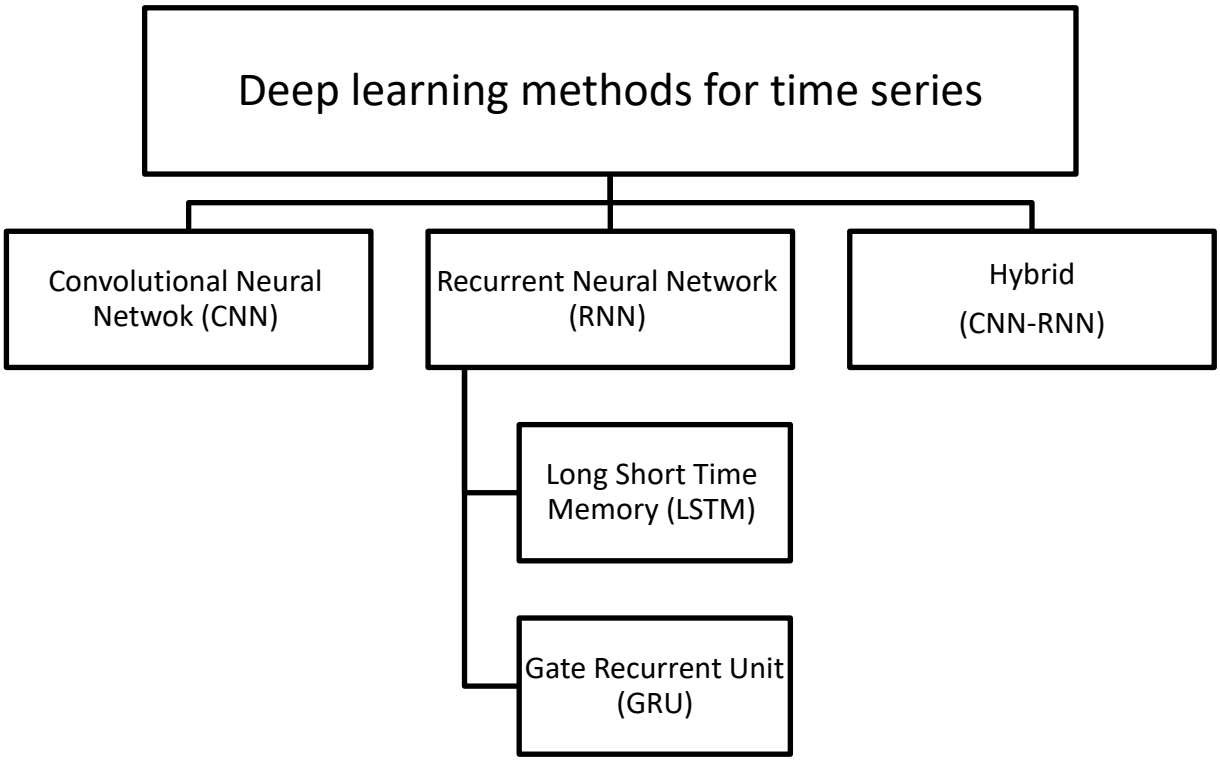


Figure (2.4): Deep Learning models for time series

2.4.1 Recurrent Neural Network (RNN)

Sequential data modeling like raw sensor data or time series, a recurrent neural network was developed. The RNN integrates a time layer for sequential information capture and then learns changes that are complex by using the current cell's hidden module. The information available to the network can cause hidden unit cells to change, this information has been updated regularly to reflect the network's current state. The current hidden state is computed by the RNN by considering the next hidden state to be an activation of the previous

hidden state. However, it is hard to train the model and suffers from fading or bursting gradients which limit its application for modeling long-term activity sequences and time dependencies in the data of the sensor [61]. Recognizing human activity with recurrent neural networks is a classic time-series classification problem with complex kinematics that varies over time. In traditional RNNs, where the network is trained via inverse propagation through time, there is a problem known as exploding and fading gradients [62] which causes the gradient to explode. This occurs when large error gradients begin to accumulate, which leads to training large changes in the model of the neural network. This prevents a model from being trained using the available data and leads to instability of the trained model. Gradient fades occur when the loss function gradients become very small (close to zero), as the initial layers' weights and biases are not effectively updated with training sessions, the network becomes increasingly difficult to train. LSTM is designed to overcome these problems RNNs and LSTMs are extremely similar with the distinction being that LSTMs' hidden layers contain memory blocks made up of cells rather than nonlinear units, which are capable of storing information for extended periods. In other words, conventional RNN cells have one inner layer running on the current state (h_{t-1}) and input (X_{t-1}) as shown in figure (2.5), while an LSTM cell has three layers.

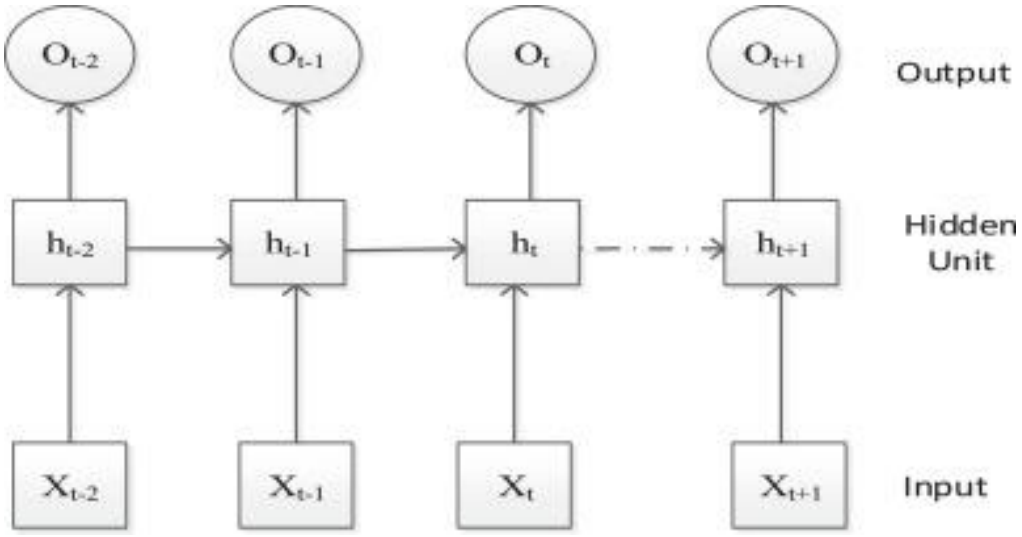


Figure (2.5): Simple recurrent neural network [4]

2.4.1.1 Long short term memory (LSTM)

Long short-term memory is a type of RNN that can learn long-term dependencies, capable of learning and remembering over long sequences of input data. Hence, LSTM is commonly used for time series analysis problems. LSTM does not use activation functions within its redundant components. The values stored there are not modified. LSTM also does not have the problem of gradient fading during training [63]. Long short- term memory is useful for dealing with a variety of problems, including handwriting and speech recognition, because it is capable of learning from inputs when significant events are separated by a long time. The units of LSTM are made up of cells as well as three gates: forgotten, input, and output. The LSTM memory is represented by the cell, which is used to remember values over arbitrary time intervals. The (gate) of an LSTM is a network structure that is distinct from the input of the vector and a set of outputs ranging from 0 to 1. If the output value is 0, there is no information permitted to pass. When the value of the output is 1, all information is permitted to be passed. The forgotten gate's f_t output is calculated as in the equation (2.1), where $W^*$and U^* weight matrices respectively, the parameters of which will be learned during training of model.

$$f_t = sigmoid(W_t^* x_t + U_t^* h_{t-1} + b_t) \quad (2.1)$$

The model is capable of learning by considering preceding output to collect specific information j_t based on the current time stride. The update gate, like the forget gate, determines which data that added to the cell state starting with a current time step

$$i_t = sigmoid(\, W_i x_e + u_i h_{t-1} + b_i) \quad (2.2)$$

$$j_t = \tanh\big(\, W_j x_t + U_i h_{t-1} + b_j\big) \quad (2.3)$$

Lastly, the updated state can be computed using a previous state, the output of the forget gate, and an update gate output.

$$c_t = f_t \odot c_{t-1} + i_t \odot j_t \quad (2.4)$$

an element-wise vector product is denoted by the operation⊙. The cell output h_t the current time step is computed using the updated cell states c_t

$o_t = \text{sigmoid}(W_o x_t + U_o h_{t-1} + b_o)$ (2.5)

$h_t = o_t \odot \tanh c_t$ (2.6)

The function of Sigmoid and Tanh is defined as follows

- Sigmoid function:

  The logistic function, also known as the sigmoid function, is one of the oldest functions of activation currently in use. It is common to associate its input value in the range [0, 1]. In the early days of artificial neural networks, it was a popular choice. The sigmoid function is primarily employed in classification tasks. The expression for sigmoid is as follows:

  $\sigma(x) = \frac{1}{1+e^{-x}}$ (2.7)

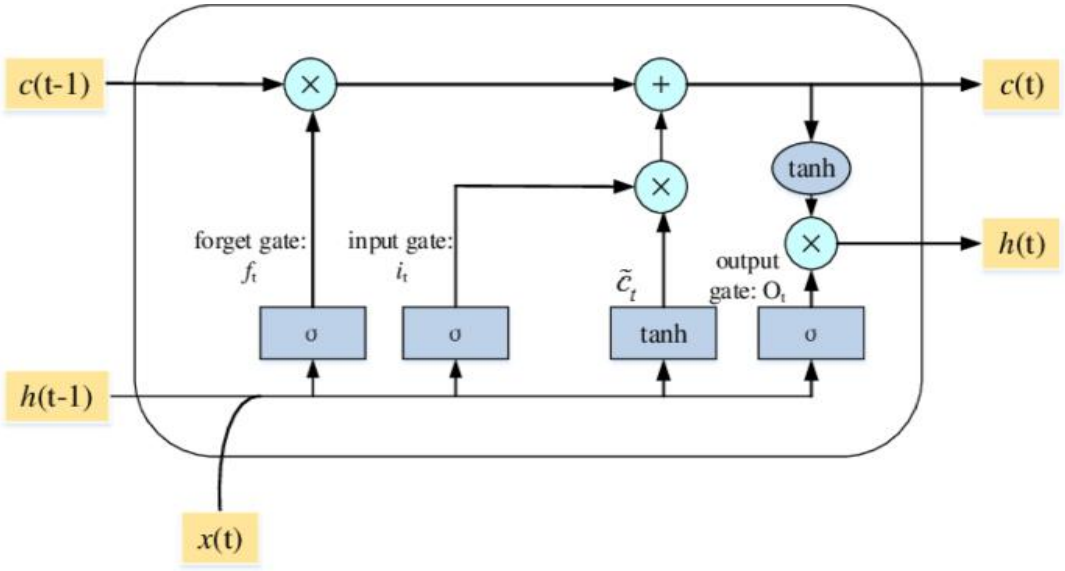
- Hyperbolic tangent (Tanh):

  The hyperbolic form is similar to the sigmoid function but the product of the values and their location is always in the range [-1, 1]. The following is an expression for tanh:

  $f(x) = \frac{(e^{x}-e^{-x})}{(e^{x}+e^{-x})}$ (2.8)

Many recent publications have demonstrated that LSTMs perform well and are relatively simple to train. LSTM has become the standard architecture for tasks requiring the processing of sequential data with temporal information. The LSTM is the architecture of artificial recurrent neural networks (RNN) which is mainly used in the deep learning field. Recurrent neural networks generally have a short memory. Derived from the feed-forward neural network, the recurrent neural network has the internal state (memory) to save the import data; it not only allows LSTM to have feedback connection but also makes LSTM process sequences of data possible. The LSTM layer is composed of a cell state, a hidden state, and three gates. The three gates are the forget gate, the input gate, and finally the output gate. The cell state and hidden state are where the LSTM layer uses memory to store the information from the previous time step. In the schematic diagram figure (2.6), the cell state is the upper horizontal line that goes through layers and the hidden state is the lower one. The information

stored in the cell state will be managed by the three gates and then passed to the next LSTM layer. The forget gate is in charge of removing the information derived from the cell state, the information that is less important or is no longer needed for the learning process, used the Sigmoid function as our recurrent activation function to preserve the Non-Linearity from the values. The input gate is accountable for adding information to the cell state. The input gate can be described as two parts. The first part does similar work to the forget part, which regulates the values that should be added into the cell state from the hidden state and the input value of the current state. The second part creates a vector that contains the potential values that can be inserted into the cell state; this part uses the tanh function as the activation function. The output gate is responsible for exporting the result of the current LSTM layer and sending the hidden state value to the next LSTM layer. After applying the activation function to the cell state and hidden state vectors, the output result will be the product of the two-state vectors.



Figure(2.6): Illustration of LSTM [64]

2.4.1.2 Gate Recurrent Unit (GRU)

A recurrent gated unit is analogous to LSTM memory [65]. The main difference is the method of applying the gates. The Reset gate and Update gate are only two gates for GRU instead of three for LSTM. A reset gate controls how new inputs and prior memory are combined as well as controls how much previous memory is left behind. The previous memory increment that will be discarded is represented by the smaller value of the reset gate. The update gate controls how much of the previous memory is retained. The previous information that will be fetched is the most valuable aspect of the update gate. There are a few differences between LSTM and GRU. For starters, LSTM has control over memory exposure, whereas GRU has uncontrollable memory. The GRU then lacks an output gate, as does the LSTM. The update gates replace the input and forget gates in LSTM; As a result, the reset gate is used for the previously hidden state directly. Because the GRU possesses fewer parameters than the LSTM, it will run faster and less data is required. If the amount of data is large, LSTM may produce a better outcome [66].

2.4.2 Convolutional Neural Network (CNN)

The convolutional neural network is a type of feed-forward neural network inspired by studies on the visual cortex of mammalians. CNNs are used for a variety of tasks such as recognition of handwriting, tracking motion in videos, recognizing objects in images, and so on, and are considered to be the most effective for image and video processing [46]. Because of the success of the CNN architecture in these various fields, researchers have started to use it for time series analysis [67]. The hidden layer (In this case, it is also referred to as the convolutional layer) is made up of several neuronal groups. All of the neurons in a group share their weights. In general, each group is composed of as numerous neurons as are required to completely cover the image. As a result, it's almost as if each neuron's collection in the layer of hidden computed an image convolution using their weights. Figure (2.7) depicts the convolutional neural network's structure. A CNN is made up of several different layer types which are piled on top of one another. There isn't any way to stack the different layers; it varies depending on the designer. The classification of an object is a simple example that demonstrates the fundamentals of CNNs, However, it can also be applied to other types of

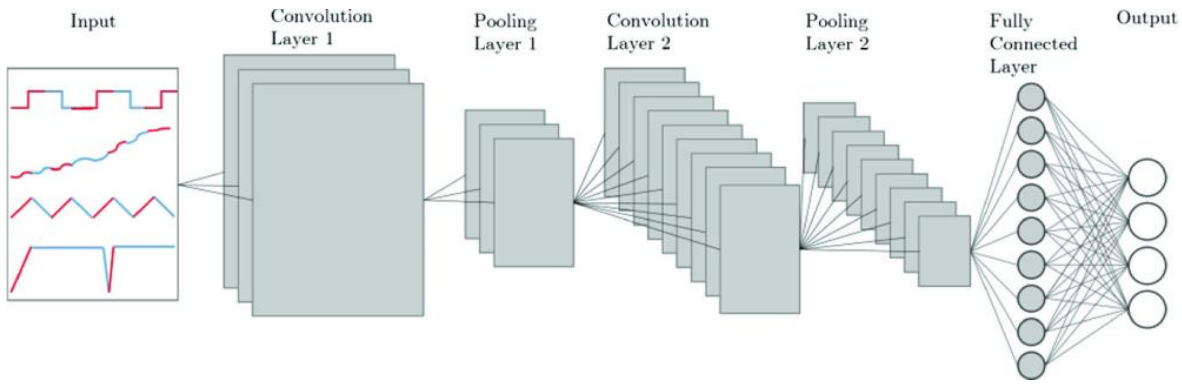
data, such as sound or text [68]. In the following subsections, various types of layers like convolutional, pooling, fully connected layers, and so on [69] will be discussed. The convolution neural network belongs to the Deep Learning field, the original purpose of CNN is to improve image classification but the method that CNN provides is not limited to the image recognition area. Nowadays CNN has been widely employed in different fields and time-series data is one of the kinds for many years, automatic learning methods are heavily based on handmade features ranges from basic statistics to public techniques to reduce dimensions into different scales. Although the methods of generating features have been successfully used in many places, usually including an arduous design phase, it requires the engineer to create knowledge about types of patterns in data taken by relevant information. This has led to specialized educational systems that are only able to fill narrow and unable roles, but important, deep learning tends to overcome these restrictions. Instead of designing the deep learning features automatically via the network. Also, the profound neural network can extract deep class high-level representation, making them more convenient to learn about complex tasks. A deep neural artificial (ConvNets or CNN) network usually has several lays of the neural network, each having more than one layer of neurons. CNN is specialized in the extraction of signal features and has demonstrated promising image classification, voice recognition, and text analysis outcomes far superior to other classifiers [8]. The outcomes may be further developed exactly by increasing the numbering of hidden neurons and layers of convolution [70] many important network layers have different roles in CNN like convolutional, pooling, and fully connected layers. Using time-series data in experiments in this paper the convolutional layer in extracting data and in extracting functions from subsequent convolutions are more complex and efficient. The number of features can be reduced by using a pooling layer and the maximum pooling is used in the experiment. The intermediate layer of the neural network acts as each layer of convolutional and single pooling layer. Before the output, the layer is placed in the final fully-connected layer (Fc), which collects preceding layer results for the score calculation for each last class. The role of the output layer produces classification outcomes based on the output of fully connected layers[24]. A performance for convolution is calculated as:

$$C_i^{l,j} = \sigma\left(b_j^l + \sum_{m=1}^{M} W_m^{l,j} X_{i+m-1}^{l-1,j}\right) \quad (2.9)$$

Where σ function of activation, the term b points to feature map bias, the symbol l indicates an index of layer, M is the size of kernel/filter, and W refers to feature map weight.

2.4.2.1 Architecture of a convolutional neural network

A convolution neural network consists of several convolution and sub-sampling layers. Each layer contains several feature maps, which are related to previous layer maps via a collection of small receptive fields. Input data, as expected is split into several feature maps. The outcome convolution defines input data features like lines, curves, corners, etc. The occurrence of a feature is mirrored on the feature map by the location of the corresponding receptive field, defining the feature map. The subsampling layer comes after each convolution layer. A CNN consists of three different parts, and the format of the input layer is the organization of the shape of the input data. Any data passed to the CNN model requires that it fit the shape required by the input layer, so data pre-processing is necessary. In CNN, hidden layers are the place to extract features containing convolutional layers (necessary) and pooling layers (optional) the convolution layer extracts features when the input does not fit the filter size exactly and uses the activation function to remove non-linear data. it requires a fully connected layer to receive all the feature vectors from the hidden layers, and another layer fully connected with a softmax activation function to shrink the feature vectors, so CNN will be able to do classification Data Preparation Conventional algorithms particularly deep learning models on neural networks of convolution (CNNs) can be taught to (learn) patterns and be adapted to a wide range of problems and tasks, such as denoising super-resolution, and segmentation. Model training, whether deep or not, is heavily reliant on data.



Figure(2.7): The architecture of the convolutional neural network [66]

❖ Input layer:

The raw input data are stored in the input layer. It's a 3D input of width and height of the image with depth of the color channels, usually 3 for RGB

❖ Convolutions layer:

The basis of the CNN is the layer of convolution, which uses smaller filters or kernels than the image input. The wrapping is performed with a portion of the input and the core, which finally covers the whole input of the image. A feature map, also known as an activation map in this process is the output; the activation map is stacked to produce a 3D tensor. As filters are trained, they will be able to identify patterns and edges more deeply on the network. The function ReLU is widely employed CNNs. All negative inputs are taken from this layer and set to zero. The ReLU layer does not contain hyperparameters, which are the designer's parameters specified; every ANN has a primary structure that is made up of convolutions. The convolution operator has the following parameters:

- Filter size
- Padding
- Stride
- Activation function

Every convolutional neuron used the operation of dot product with its inputs and matrix of weights. A kernel or filter is a term used to describe this operation. When it comes to 2D convolutions, the filter is slid across the input's width and height to generate a weighted sum. The several parameters including the padding, stride, and the number of feature maps choose the output's shape. The stride influences the step length as the filter slips over the input. The most common padding is zero padding, which adds a layer of 0s to an input's edges. This is frequently used to modify the output dimensions of input so that the dimensions in the output are the same. Finally, the number of feature maps or channels determines how many feature detectors can be learned within that layer. The output dimensions of a layer can be calculated using this equation:

$$\frac{input\ size + 2*padding - filter\ size}{stride} + 1 \qquad (2.10)$$

To illustrate the convolution process, features can be extracted from images using the first layer, which is a convolutional layer. Pixels are only associated with neighboring and nearby pixels; Convolution enables to maintain the relationship between the image's various components. Convolution filters the image using a smaller pixel filter to reduce image size without loss the relationship between the pixels[70]. Convolution can apply to a 7×7 image with a 3×3 filter in 1×1 step (1-pixel offset per step); end up with a 5×5 output as shown in figure (2.8)

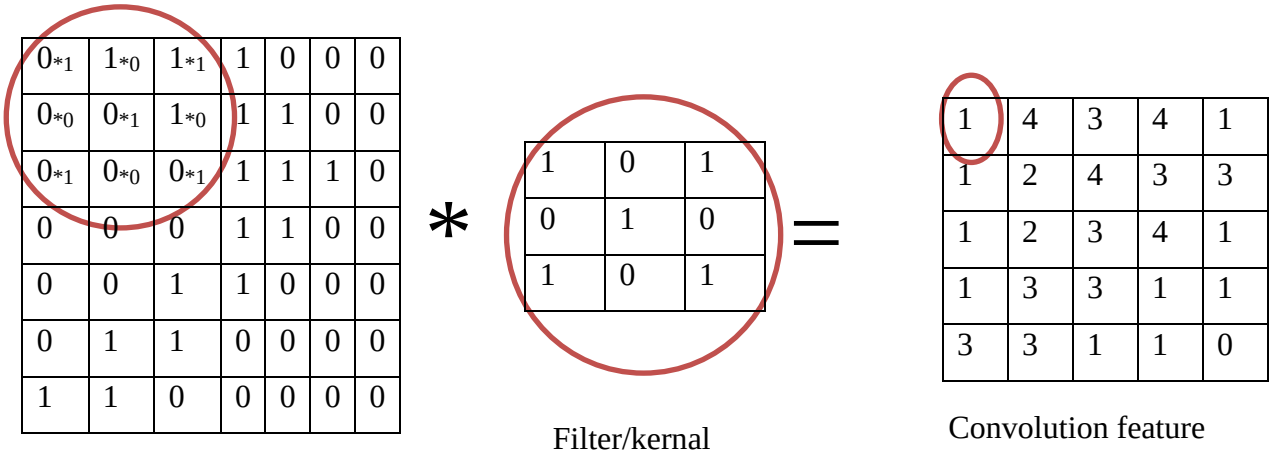


Figure (2.8): Convolution Operation

❖ Rectified linear units:

A linear function lacks the complex mappings required for deep neural networks' high dimensionality. ANNs rely heavily on activation functions to incorporate nonlinearity into linear functions. These functions of activation are commonly applied to each neural network layer's output. There are many different types of activation functions, for example, tanh and sigmoid. The sigmoid and tanh functions, on the other hand, suffer from the vanishing gradient problem. ReLU is an abbreviation for rectified linear units; has become a common activation function in the majority of neural networks. The ReLU function is a fast monotonic operation that avoids the vanishing gradient problem. There is one drawback to ReLU when a neuron receives a significant amount of negative bias, the

functional gradient becomes 0, or in other words, the rectified linear unit is a function of activation that searches its input arguments for positive if the input is positive, the value is returned if not, i.e. Zero is returned as the final output value when the value is less than zero. ReLU is the following expression:

$$y = x+ = \max(0, x) \tag{2.11}$$

- Batch normalization layer:

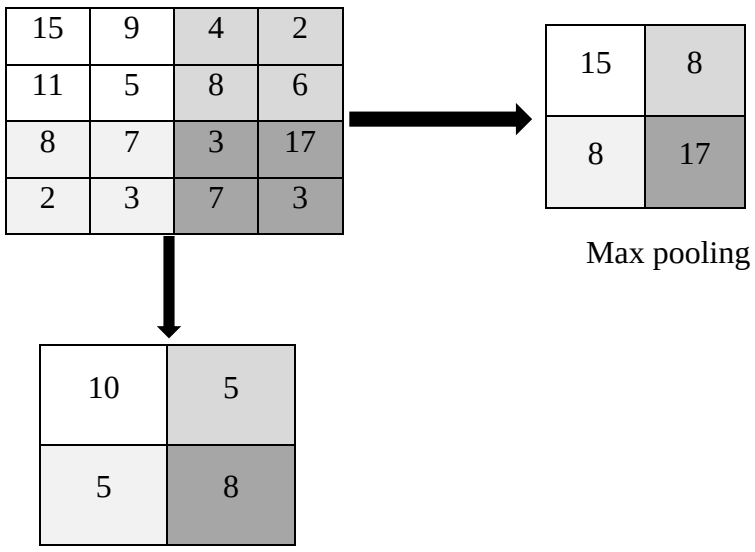
The batch normalization technique allows training of very deep networks in a very quick and efficient manner and solves the problem a concern about initializing the weights of the neural network. The benefit of normalization is to avoid model bursting while increasing the gradient with a large rate of learning. This layer must be inserted between the layer of convolution and the activation function. The operations of batch normalization are normalized to the input data with zero mean and unit variance within a training set of images. Deep neural network training is complicated since the distribution of the inputs of each layer changes that occur during training as the parameters of the preceding layers change. This slows down training by necessitating lower rates of learning and precise parameter initialization. The power of this method stems from incorporating normalization into the model architecture and performing normalization is carried out for each small training batch. The batch settlement enables much higher learning rates and is less constrained in terms of configuration. It also functions as a regulator, eliminating the need for leakage in some cases.

- Pooling layer:

A pooling layer is always accompanied by a convolutional layer; max-pooling is popular. Since its output pooling layer reduces data size, its pooling layer takes as many small features as possible the convolutional layer is followed by the pooling layer to convert the latter's output into a type of summary statistic of nearby outputs. The operation of max-pooling generates the maximum value among a collection of nearby inputs, as given by:

$$p_i^{l,j} = \max_{r \in R}\left(c_{i \times T + r}^{l,j}\right) \tag{2.12}$$

Where *R* is representing the pooling size, *T* is the stride of pooling, and $p_i^{l,j}$ is the output of the pooling layer. Deep neural network architecture can be built by stacking several convolutional and pooling layers. These layers function as a feature extractor in a hierarchy, which extracts discriminative and representations of information concerning the data pooling layer do not provide any learning in itself; they are simply mechanisms that introduce sparseness and translation invariance. Pooling layers are commonly added after each convolution layer to reduce the representation's spatial size. The number of parameters is reduced by this layer as a result, computational complexity is reduced. Shared layers help solve the problem of over-fitting. choose the size of the pool to reduce the number of parameters by specifying the maximum, average, or total values within these pixels [70]. Figure (2.9) shows the maximum pooling process and the average pooling process.



Figure(2.9): Max pooling and Average pooling operation

❖ Dropout layer:

The dropout layer is useful for avoiding data overload in the course of the training process and is being added after the layers of fully connected at the neural network's end.

❖ Soft-max layer:

A very simple softmax classifier is used at the topmost layer to recognize activities. Given activity feature values that have been flattened $p^l = [p_1, \dots p_I]$ where I pointing to the total number of units in the final pooling layer, the softmax classifier's output the activity class c:

$$P(c/p) = arg\max_{c \in C} \frac{exp(p^{L-1} w^L + b^L)}{\sum_{k=1}^{N_C} exp(p^{l-1} w_k)} \quad (2.13)$$

Where L is the last layer index, and N_C is the total number of activity classes. The final output is finally transferred to the soft-max layer for the calculation of the probability distribution over the predicted classes' also known as logistic regression multinomial, which is a mathematical function that normalizes the classes' predicted scores. The scores are the convolutional neural network's output values. As classifier outputs, they are placed in a vector format, with each position corresponding to a class, so that in an inference test, the score vector's maximum value should be assigned to the appropriate output class. After a classifier, the softmax function is used to compute the probabilities of each class. It takes the scores and compresses them into a value vector ranging from zero to one. As a result, the outputs of a softmax can be interpreted as a vector of label probabilities during training in a classification task. The goal of the softmax is to maximize the probability of the correct classification of class. As a result, the negative log-likelihood of the correct class is minimized.

2.4.2.2 Human activity performance evaluation and classification algorithms

Classification is an essential component of the process of identifying human activity. Training, testing, and the use of rating scales to assess the algorithm performance proposed are all part of the rating process. In human activity recognition, various classifiers have been

implemented over the years to classify the specifics during the training and testing of activity. The process of training extracts the attribute vectors fed to the classifiers via fully connected layers to obtain each discrete time step of sensor data that has a different probability distribution class. The extracted feature vectors' performance is evaluated using predefined metrics for evaluation and access to identification accuracy and complexity. The performance metrics like accuracy, recall, precision, and F-measure provide basic information for the ability to recognize feature vectors. Start by providing training for each of the DL methods then classification algorithm of inference and evaluation performance measures to recognize the human activity. The evaluation metrics used in deep learning tasks are critical in achieving an optimized classifier. They are used in a standard classification of data procedures in two stages: training and testing. It is used to improve the classification algorithm in the course of training. In other words, the evaluation metric is used to discriminate and choose the best solution, which can produce an extremely accurate prediction of upcoming evaluations in the case of a particular classifier. The evaluation metric is used to assess the effectiveness of the created classifier, for example, as an evaluator, TN and TP are defined as the number of negative and positive instances successfully classified, respectively. Furthermore, FN and FP denote the number of positive and negative instances that were misclassified, respectively. The following are some of the most widely used evaluation metrics.

- Recognition rate: The ratio of correctly predicted classes concerning the total number of samples is computed. Also called accuracy.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{2.14}$$

- Sensitivity: Calculates the percentage of positive patterns that are correctly classified. Also called recall.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{2.15}$$

- Specificity: Calculates the percentage of Patterns of negativity that are correctly classified.

$$\text{Specificity} = \frac{TN}{FP+TN} \tag{2.16}$$

- Precision: Calculates all positive patterns which are correctly predicted by all In Predicted patterns in a positive class.

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{2.17}$$

- F1-Score: The harmonic average of the recall and precision rates is computed.

$$\text{F1-score} = \frac{2*\text{precision}*\text{recall}}{\text{precision+recall}} \tag{2.18}$$

2.4.2.3 Training Algorithms

Early works were trained using deep neural networks with high ratio optimization as adjustments are made to the weights and biases because of the large number of parameters. Researchers of deep learning adopt strategies when it comes to proving the efficacy of their methods. The main goal of training DL algorithms is to identify network parameters to minimize refactoring errors between input and output [71]. Networks discover how to extract salient features from data of sensor through fine-tuning and pre-training then passed to SoftMax to highlight details of the activity. Therefore, several regularization methods are proposed that the learning algorithm be modified to reduce generation errors by using hyperparameters to control network behavior. According to [72], the hyperparameters involve learning rate values, weight decrement, initial weight values, momentum, and mechanism for weight updating, activation functions, optimization procedures, network depth, small batch sizes, and training epochs...etc. In the recognition of human activity based on deep learning, various studies identify variable values for the hyper-parameters that depend on the network Moreover the sensor data size for training.

- Stochastic Gradient Descent

The traditional random gradient descent (SGD) method is used to recursively adjust the network weights, based on the training data[73]. Stochastic gradient descent (SGD) is the most common variation and implementation of gradient descent. In gradient descent, updates are applied after running a small set of several samples that can converge towards the global minimum faster which gives us the network error values. Random gradient descent is used for training on mini-batches examples of sensor training

- Learning rate.

This is the most important hyperparameter because it can greatly change the accuracy of the model. In general, a small learning rate takes a very long time to converge, while a large learning rate may perform well at the beginning, but does not converge as a result of overshooting [74]. The use of learning rates varied across research (e.g. 0.01, 0.0001, 0.001, etc.)

- Epoch

In a neural network, a period presents a complete learning process for a training data set. The number of epochs is being increased; the learning process for the same training data set is also repeated. The result shows that by adding the number of epochs, the prediction accuracy is increased, as the neural network field loss measured the absolute error between the model output value and the actual value of the real-world data. So, accuracy is not the only factor in choosing the number of epochs; also needs to reduce the value of the loss to a minimum.

2.5 Time series

Time series are very common to present collected data such as economic indicators (stock analysis and forecasting), natural phenomena (temperature changes, earthquake records), medical metrics (electromyography, electroencephalography recordings), and control engineering data, among others The time series are viewed as a sequence of equally spaced

consecutive data where the time information may or may not be present. It is a series of data points indexed (or listed or graphed) in time order as shown in figure (2.10) time divided into discrete or continuous [75]. A time series is a sequence taken at successive equally spaced points in time. Techniques of deep learning have a significant and play an effective role in the resolution of problems of time series, this is demonstrated by their ability to deal with a variety of variables input and complex nonlinear relationships, and may not necessitate the use of a stationary or scaled as input of time series[67]. The time series methods of prediction are based on the concept that data contain intrinsic patterns that provide future-relevant information descriptions of the phenomenon under investigation. Typically, identifying these patterns is not easy, and one of time series processing's primary goals is to discover them, the circumstances under which the discovered patterns will reoccur, and what kinds of changes they might experience over time. The time series are distinguished from other types of data by several characteristics. To begin with, the data of time series includes a high dimensionality and a lot of noise, reducing dimensionality, filtering, or wavelet analysis are techniques for signal processing that can be used to reduce dimensionality and remove noise. The following property of time-series data is there is no guarantee that the information available is sufficient to comprehend the process. Lastly, data from time series are complex, multidimensional, and have distinct characteristics that distinguish time-series modeling and analysis difficult. Representation of the data of time series is critical to extracting pertinent information and reducing dimensionality. The right representation of time series is important to the success of any application.

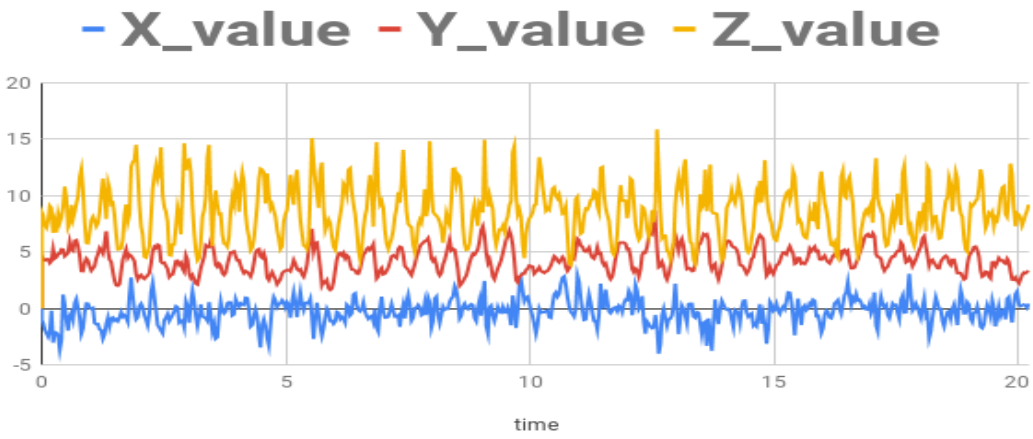


Figure (2.10): Time series for recognizing human activity using the smartphone

2.6 Inertial Measurement Unit (IMU)

A system that combines the output of several inertial sensors is known as an inertial measurement unit to provide a more comprehensive set of telemetry data than a single sensor could provide on its own. A standard low-cost Micro Electro Mechanical Sensors (MEMS) sensor[76] such as those discovered on a smartphone commonly uses a three-axis for each accelerometer, magnetometer, and gyroscope to provide true acceleration, orientation, and angular rate. These devices have been widely used over the past few years for tracking a moving body's orientation for an extended time. A common method for estimating orientation is inertial navigation, but it suffers from the accumulation of errors. The accumulating error in IMU measurements is well known because is a built-in hardware limitation. As a result, various attempts have been made to compensate for this error [77] using a much less complicated method, where a step is identified by looking at patterns in data from the accelerometer as well as calculating the distance traveled by multiplying the number of steps with average stride length. Because of estimation, such methods discard a large portion of the input data. Our method estimates the elevation angle and bank angle of the orientation over a persistent human movement using accelerometers and gyroscopes, which is new in the field. Since the estimated orientation is based on prior knowledge about human motion, it is prevented from growing unboundedly. The measurement model is designed to estimate the

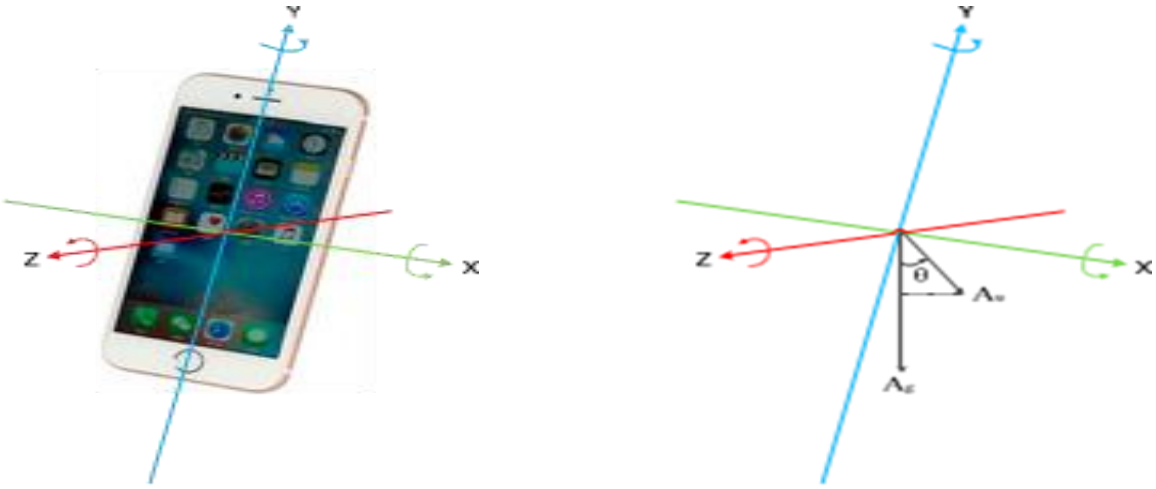
position for (t) seconds in the future. The knowledge of the estimated position for a few seconds further in the future provides feedback for orientation estimation during the periods when the accelerometer’s readings have significantly deviated from gravity. The Inertial Measurement Units (IMUs) are used to measure motion. Electronic stability control, inertial navigation systems (INS), and medical field applications related to movement analysis are just a few of the many fields in which IMU's have applications. Rapid advances in technology, functionality, cost, size, and real-time applications have made IMUs a popular choice for a variety of applications. Inertial measurement units (IMUs) can be worn on different body segments to record movement. Accelerometers and gyroscopes make up the IMUs. Gyroscopes are used to measure the angle change concerning a known initial angle by integrating the velocity over time. Accelerometers measure gravitational and translational acceleration. They are used to measure the speed of a moving object. As a result of transforming acceleration from sensor frame to earth frame, the gravitational vector can be determined. When computing the velocity of an object, the acceleration is first integrated, and then again when computing the position of that object. Inertial navigation is a well-known term for this technique. Gyroscope data and acceleration data can be used to determine the smartphone's state of motion and stability. It should be noted that as illustrated in figure (2.11 a) the data are proportional to the device's coordinate system, so they will change as soon as its orientation shifts. As a result, if users maintain their smartphones in various orientations, the acceleration will change, regardless of whether they are performing the same task in the same manner, posing a major problem for activity detection. It is converted into vertical and horizontal directions to avoid the effects of device orientation. The smartphone's acceleration is changed from triaxial to horizontal and vertical. In contrast to[35], which relies on magnetic field sensor data. They no longer need to filter the data in three axes, as they have the ability now to obtain separate gravity and user acceleration [78]. Our planet's gravitational field is oriented toward the center so that the user's acceleration can be broken down into two components. One component is oriented toward the earths, while the other component is perpendicular to it. Using projection, the angular relationship between gravity and the user's acceleration can first be determined.

$$\cos \emptyset = \frac{A_{gx}A_{ux}+A_{gy}A_{uy}+A_{gz}A_{uz}}{\sqrt{A_{gx}^2+A_{gy}^2+A_{gz}^2}\cdot\sqrt{A_{ux}^2+A_{uy}^2+A_{uz}^2}} \quad (2.19)$$

$$A_v=\sqrt{A_{ux}^2+A_{uy}^2+A_{uz}^2}\cdot\cos \emptyset \quad (2.20)$$

$$A_h = \sqrt{A_{ux}^2+A_{uy}^2+A_{uz}^2}\cdot \sin \emptyset \quad (2.21)$$

Where $\emptyset$ gravity-user acceleration angleA_{gx}, A_{gy} and A_{gz} represent the gravity in the X, Y, and Z axes. A_{ux} , A_{uy} and A_{uz} respectively, pointing to the user's acceleration for three axes. A_v And A_h accelerations of users in both the vertical and horizontal directions this is illustrated in figure (2.11 b). Users' accelerations for vertical and horizontal accelerations were converted through coordinate transformation. Smartphone orientation does not affect these accelerations because they only relate to the user's activities.



(a) A smartphone's coordinate system (b) Acceleration conversion process

Figure(2.11): The coordinate system and the process of conversion [78]

The performance of the activity recognition system critically depends on the sensor method used. Sensing methods can be classified into four strategies [19] as shown in figure (2.12). The data was collected by a smartphone, so will focus on two sensors, the accelerometer, and the gyroscope

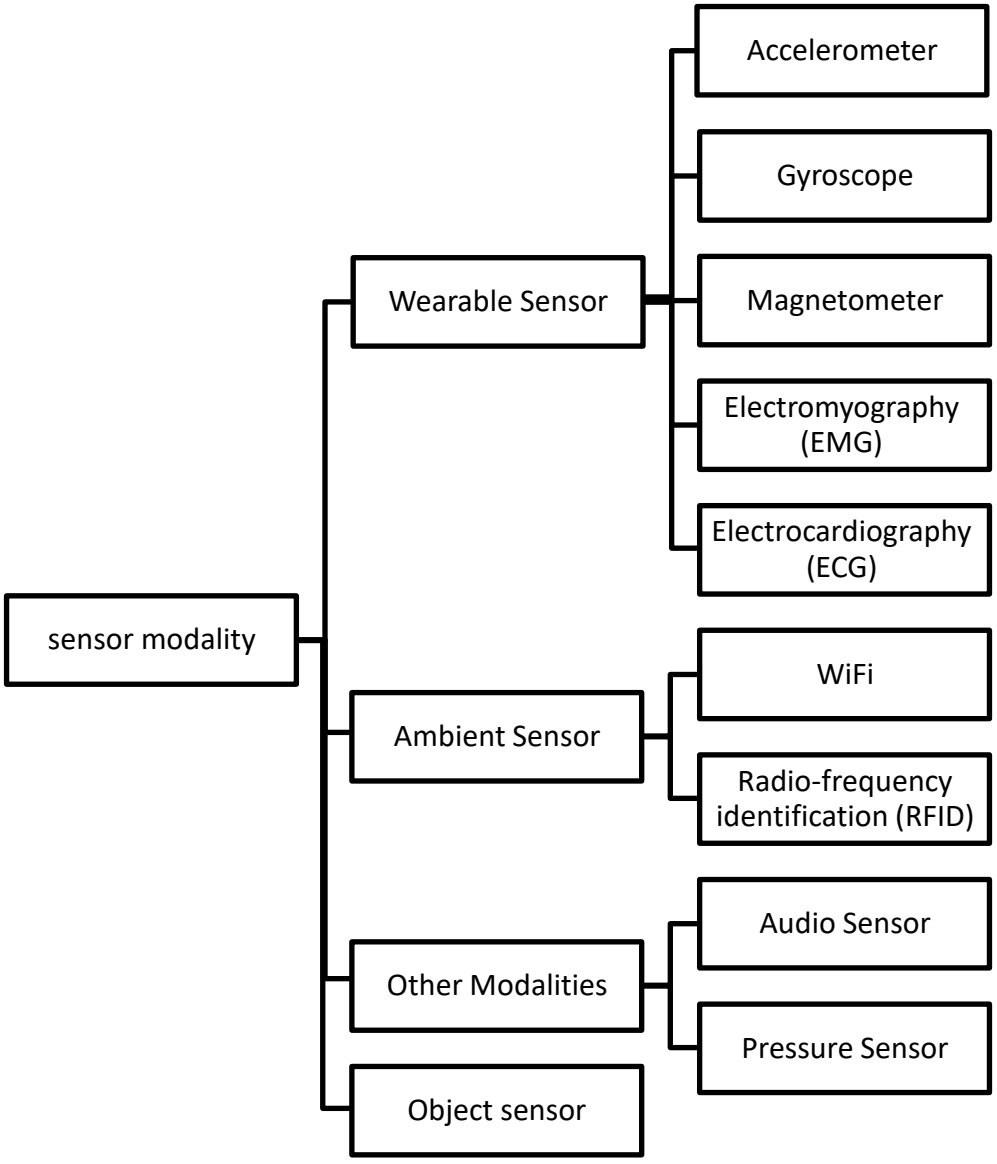


Figure (2.12): Sensor modality

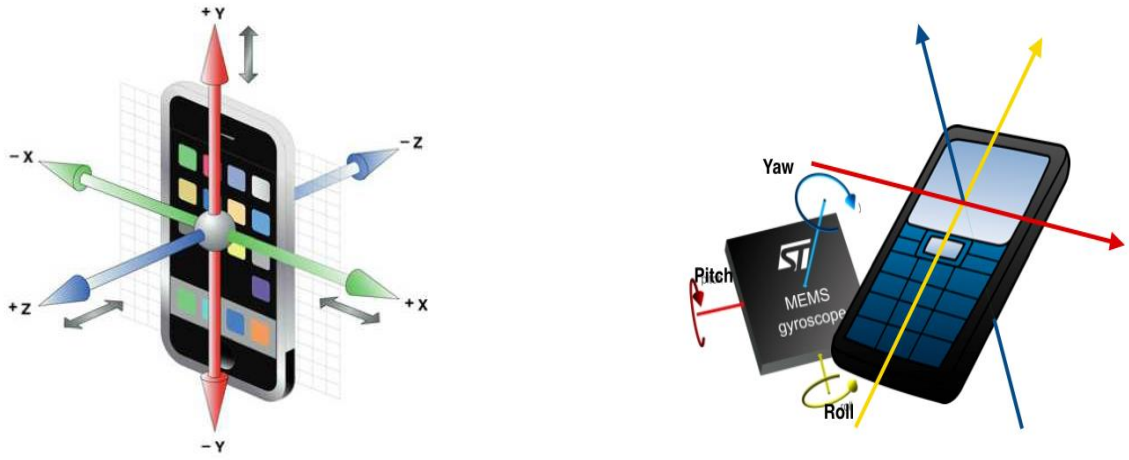
2.6.1 Accelerometer

An accelerometer is a device that measures an object's perceived physical acceleration. It was employed in many implementations for science, industry, engineering, and medicine, like measuring machinery vibrations, high-speed vehicle acceleration, and relocating loads on bridges. One of the most widely used sensors for reading movement signals of the body is the accelerometer, which is important to HAR [79]. The magnitude, as well as the direction of acceleration, can be quantified as a quantity of vector by the orthogonal arrangement of the spatial three dimensions of sensors. It can as well as is constructed on a single chip and it is now widely used to find 3-axis accelerometers in many product devices electronic. This is the case with the smartphones exploited in this study. The smartphone's accelerometer sensors detect acceleration events. The reading includes three axes; the raw data stream from the accelerometer is the acceleration of each axis in units. A set of vectors is used to represent the raw data (x_i,y_i,z_i) where i= 1, 2, 3… and timestamp can also be returned with the readings for the three axes. Most current accelerometers function as a user interface for configuring the frequency of sampling so that the user can experiment to find the best sampling rate. The accelerometer has been used extensively in smartphone sensor-based recognition of activity. Its popularity is because it directly measures the state of the physiological movement of the subject. As an example, if the user switches from walking to running, this will be reflected as a signal reading the acceleration along the vertical axis - there will be a sudden change in amplitude. Moreover, accelerometer data can indicate the movement pattern during a certain period, which is useful for recognizing complex activity. The measuring unit is metered over a second squared $\left(\frac{m}{s^2}\right)$ or (g forces).

2.6.2 Gyroscope

A sensor that using to measure direction is a gyroscope [80]. Many applications have been used like systems of inertial navigation, and aircraft to increase stability. Newly it has made its way into electronic devices (smartphones and game consoles, for example) to enhance gameplay and user interfaces. While HAR, the sensors are used in various applications such as detecting activities like walking, walking upstairs...etc. The gyroscope measures the rotation rate of the phone by detecting the rotation and yaw movements of the

smartphones along the three-axis (x_i,y_i,z_i) where i= 1, 2, 3… The rate rotation is represented by the raw data stream from the gyroscope sensor in rad/sec (radians per second). The axes of the accelerometer and gyroscope are shown in figure (2.13 a, b)



(a) Accelerometer axes on smartphones (b) Three axes of the gyroscope on smartphones

Figure (2.13): Three axes of the Accelerometer and Gyroscope in smartphones. [81]

2.7 Body Mass Index (BMI)

Obesity is one of the most important health problems spread all over the world, which has a close association with a higher disease risk such as heart disease, diabetes, blood vessels, and stroke. BMI simple and dependable indicator based on height and weight that is employed in identifying and classifying adults as being normal weight, obese, underweight, or overweight (obese). BMI is calculated according to its traditional definition using the height (in centimeters) and weight (in kilograms) of the body as in the equation (2.22). The BMI for adults is one of the categories according to the World Health Organization [82], which are listed in the table (2.3). The traditional method for obtaining a BMI is common to carefully calculate the height and weight of the body. Some obese or pre-obese patients might delay

medical treatment because of concerns about the underestimation of physicians and health care workers and for personal reasons [83]. Users can obtain their BMI by using online applications on smartphone terminals. These apps, however, required personal data of the body such as age, height, gender, and weight, which are usually extremely sensitive subjects. Sensor-rich smartphones are popular all over the world. These sensors are built-in, such as the touch screen, camera, motion sensors and microphone, which were initially employed to improve the phone, now using in a wide range of applications of sense. a wide range of human traits and behaviors can be inferred, As a result, some research has begun developing methods for automatic measurement of the mass index the body for long-range remote monitoring using smartphones is widely used. Several current studies have focused on learning based on the human face[84-86] and speech cues[87]. However, the above sensors are also sensitive to the environment. The interaction of hundreds of muscles and the body's joints results from human movement, sensors of motion can pick up on it and translate it into distinct patterns associated with its features. Practically, it is better to get physical attributes from sensors that detect movement, for the reasons listed below:

- First of all, Sensors for movement are usually considered insensitive to personal privacy and accepted by users.
- Various sensors that are sensitive environmentally, like motion sensors, cameras, and microphones always loosen the constraints of an environment.
- Sensors of motion are widely common in comparison to other sensors, not just on smartphones but as well in else smart devices, like smart bracelets.

As a result of the low signal-to-noise ratio of motion sensor data, it may be difficult to accurately predict BMI. For conventional feature-based approaches, motion sensor data is also multidimensional and has a special temporal-spatial structure [88]**.**

$$BMI = \frac{mass}{height^2} \tag{2.22}$$

Table (2.3): Body Mass Index and Status of Nutrition

Body Mass Index (BMI)	Status of Nutrition
18.5>	Underweight
18.5–24.9	Weight is normal
25.0–29.9	Obesity onset
30.0–34.9	Obesity Classification I
35.0–39.9	Obesity Classification II
40 <	Obesity Classification III

2.8 Wavelet

A wave is a perturbation of periodic oscillations that propagates across space and time, usually as energy is transmitted. The word wavelet means (small wave) and is given in the early 1980s by Morlet and Grossmann as little because the wavelet functions have a finite length. When time localization of the spectral components is required, a transformation is required that gives the temporal frequency representation of the signal. Wavelet transformations have advantages over traditional Fourier transforms to represent functions with sharp discontinuities and peaks, and to precisely deconstruct and reconstruct finite, signals which are non-periodic and/or stationary. In the last decade, several solutions were developed to solve this problem and better represent a signal in the domain of time and frequency. This deficiency can be overcome by wavelet function representation and wavelet transfer. Therefore, for signal analysis, which can gather all information, wavelet transfers are a mathematical approach [89] widely used for signal processing applications. It can decompose special patterns hidden in the data block. The wavelet transform can display functions simultaneously and demonstrate their local properties in the time-frequency domain. Signals are measured in the time domain and so get a time-amplitude representation of the signal. The time-amplitude representation gives us only limited information about the signal whereas can get a deeper understanding of the signal if observe the signal in the frequency domain (frequency-amplitude representation). Traditionally Fourier Transform (FT) had been

the most popular technique to analyze various frequencies present in a signal. Using FT [90] were able to know what frequencies were present in a measured signal. But there was a major drawback in this technique, FT could not tell at what time a particular frequency is present [91] using Short Time Fourier Transform (STFT) in which the signal is not analyzed completely at once. In STFT a small window is defined and takes the FT of the part of the signal within the window. This window is then shifted and the procedure is repeated till the complete signal is covered. The use of a window helps in localizing the signal in time and helps solve the issue of knowing the frequency component in time to some extent. The problem with STFT is the resolution because once choose the window size; it is fixed during the entire analysis. A fixed window does not resolve all the frequency components within a signal. With Wavelet Transform, the resolution is much better in extracting frequency information concerning time. Because of the Heisenberg Uncertainty Principle, one cannot know the exact frequency at a given moment in time but can know the range of frequencies present at a specific time. As human activity signals are inconsistent and vary over time in response to different environmental influences, in addition to differing adaptability and nature from one person to another. Therefore, it is more effective to replace the short-term Fourier transform (STFT) with a Wavelet Transform (WT) which is a more powerful tool for extracting features in the frequency domain. Replacing STFT with WT addresses the unstable nature of a signal. For better visualizing the transformation will use a scalogram, a tool that builds and displays the 2D spectrum for Continuous Wavelet Transform (CWT). A scalogram takes the absolute value of the CWT coefficients of a signal and plots it. Therefore, to solve the problem of the resolution and to extract accurate frequency information this research has applied Continuous Wavelet Transform (CWT). The CWT is defined in Equation as follows:

$$C(\tau, s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} f(t)\, \Psi^* \left(\frac{t-\tau}{s}\right) dt \qquad (2.23)$$

Where Ψ is the transforming function denotes to mother wavelet, s pointing to scale, τ is the translation, and ***** represents the operation of complex conjugation. Translation corresponds to time information whereas scale is for frequency. But scale is inversely proportional to frequency. As mentioned earlier, scale is the inverse of frequency and this implies that higher scale corresponds to lower frequencies and lower scale values correspond to higher

frequencies. Transition provides time information for the beginning and end of an instant. And CWT amplitude shows the strength of the signal. Regarding the issue of prediction by time series and neural networks, the CNN route has been proposed along with the use of WT to extract frequency features. The proposed model is trained with the proportion of the data; Moreover, it solves the problem of over-fitting

Chapter Three

Proposed Framework

3.1 Introduction

In this chapter, the details of creating the proposed system and more general explanations about it are explained. It started with the collection of volunteers and the procedures used to collect the data set that will be used in the experiment, then present the data pre-processing and select it to be ready for the classification process. HAR algorithms still face many challenges. The most important challenges faced were presented. Deep learning, especially convolutional neural networks (convNets), has attracted a great deal of interest in recent years for its success in the image and speech fields due to its powerful feature extraction mechanism. Activities were classified using convolutional neural networks using time series data collected from smartphone sensors.

3.2 The proposed framework

A typical activity definition framework often follows stages including as shown in figure (3.1) data collection, data pre-processing, and classification. Multiple sensors collect data related to human usage and behavior. Several classification techniques are applied to capture activity. Sensors and smart wearable devices are used to collect data. One of the most important issues in data collection is the selection of sensors and measurement of characteristics that play a significant role in the performance of an activity recognition system. If sensors are not specified, recognition performance may have a negative impact. To recognize the activity, inertial sensors such as accelerometers and gyroscopes are used. The accelerometer measures the non-gravity velocity of a smart device while the gyroscope detects a change in direction or angular velocity.

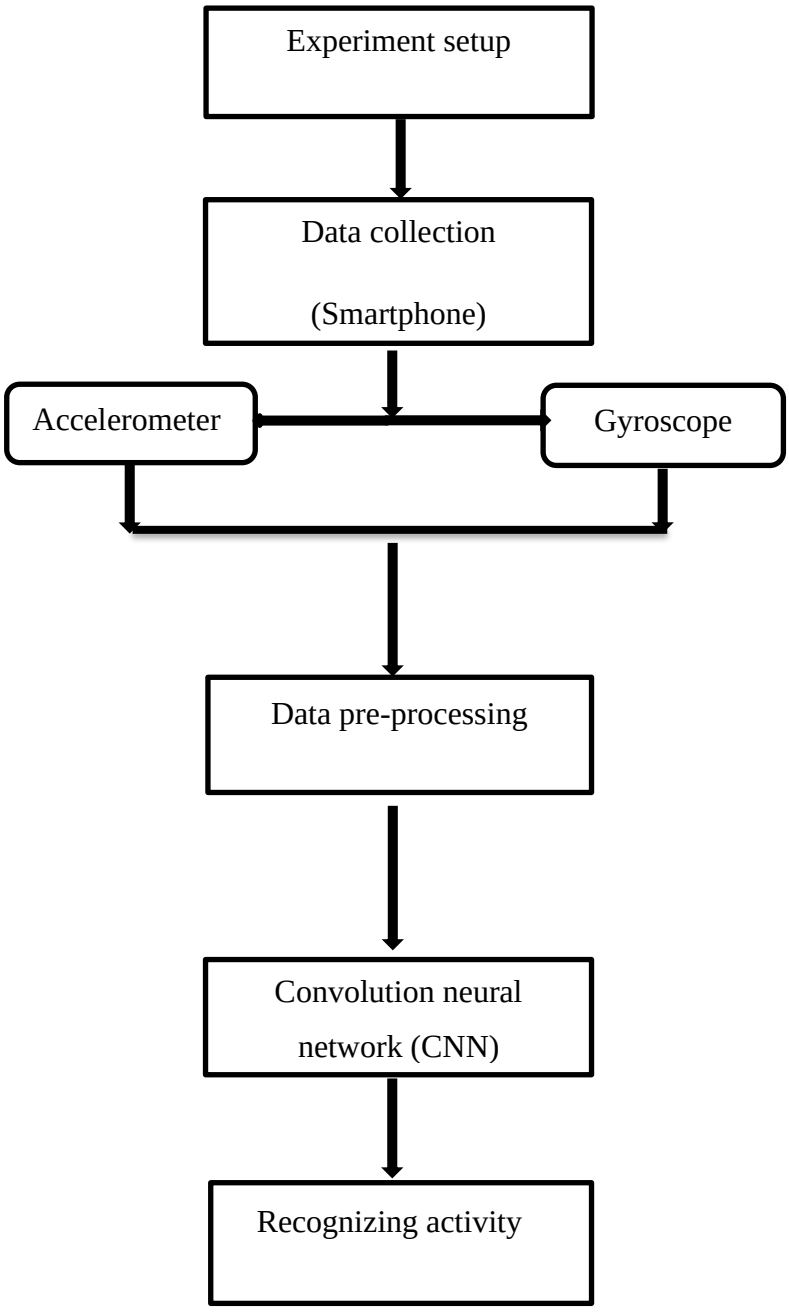


Figure (3.1): Proposed system

Recognizing human activity has become a popular topic previously decade due to its importance in the study of many fields, including healthcare, interactive games, sports, and general-purpose monitoring systems. HAR aims to learn about human activities in controlled and uncontrolled environments. HAR algorithms continue to face multiple challenges, including.

- The variety and complications of daily activities
- Intra-subject and cross-subject variation of the same activity
- The trade-off between specificity and performance
- Mobile and computational expertise
- Data are hard to explain [92]

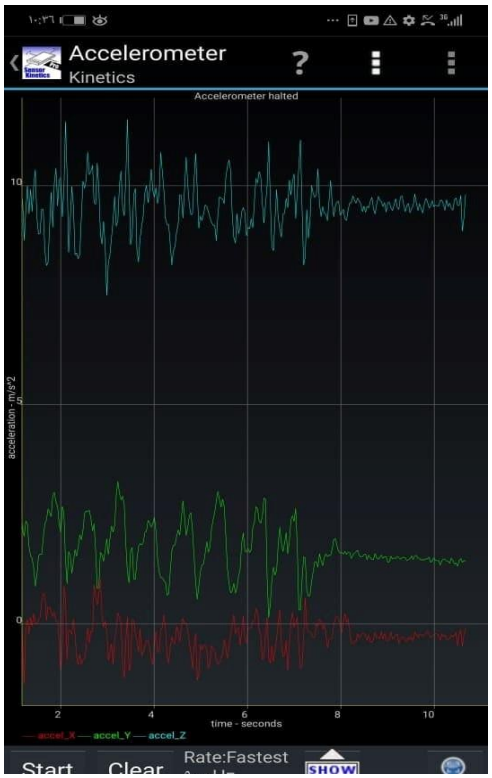
Many types of research have addressed the challenges [3, 77]. Many challenges require identifying sensor-based activity when collecting data and applying it in real-life conditions. The most important challenges that were faced in our study can be summarized in the table (3.1). Each of the challenges is explained in the following sections, each topic as appropriate for the challenge (indicated by the word in bold black).

Table (3.1): Some deep learning challenges for human activity recognition

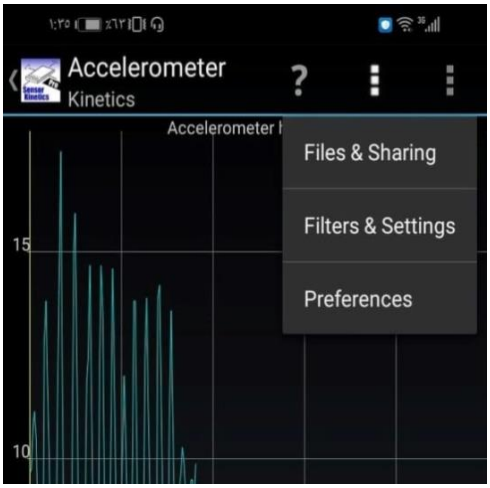
HAR challenges	Class Imbalance	Data Level
		Algorithmic Level
	computational cost	Layer Reduction
		Network Optimization
	Privacy	Transformation
		Perturbation
	Differentiation	User
		Sensor
		Time
	Interpretability	Attentive Selection
		Feature Visualization

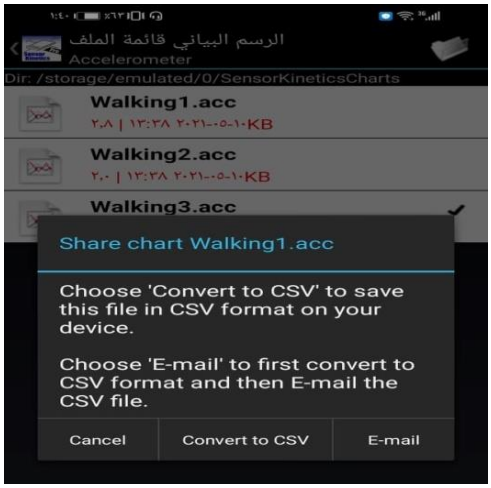
3.2.1 Experiment setup and data collection

The first step in the proposed framework was to capture the raw data of the sensors. An Android-based mobile app already available from the Google Play Store was used. This application detects the sensors that the phone has and allows them to be tested. Allows the user to set filters to the data flow, save raw data sensor the user's phone, display the data as format tabular directly on the user's phone, and export it to a computer. Kinetics Sensor Pro [93] gives the user a detailed and comprehensive view of the common sensors' operations. Figure (3.2, a) shows X, Y, and Z graphs on the Sensor Kinetics Pro app for accelerometer and gyroscope. Sensors include an accelerometer, gyroscope, and magnetometer. The derived 3D sensors are the gravity sensor, linear acceleration sensor, and gyro sensor. The last group of sensors is the numerical sensor group consisting of the ambient temperature sensor, the proximity sensor, the light sensor, the pressure sensor, and the relative humidity sensor. This study used the accelerometer and gyroscope sensors since they were the most common sensors found in smartphones and their combination covered the required range of motion. Sensor Kinetics Pro allows the user to monitor each sensor individually as the chart, save and share. This option can be selected by pressing the button in the upper right corner to open a menu where one has the option to save data. Clicking on the Files & Sharing option allows the user to save the file and data on the smartphone as shown in figure (3.2, b). After collecting the data, the application allows the user to save the data. It saves motion data points as comma-separated value (CSV) files. Then share it with others via email Clicking on the file brings up the options screen as shown in figure (3.2, c) to send the file, press the share button. To be sent to an address via email.



(a) Graphs X, Y, and Z for accelerometer and gyroscope (walking activity for example)





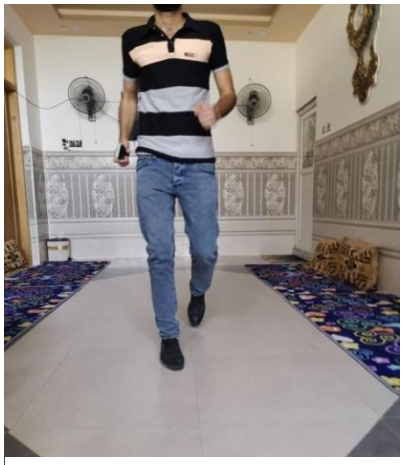
(b) Files and Sharing option (c) sharing file

Figure (3.2): A screenshot of a smartphone application Kinetics Sensor Pro

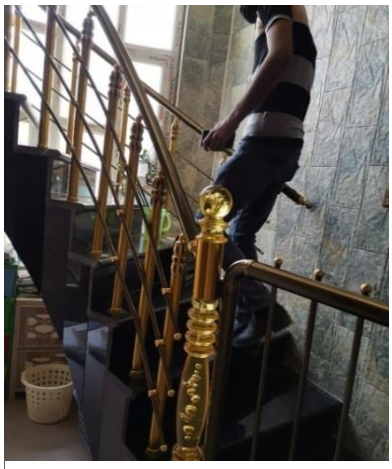
Data information was collected from smartphone accelerometers and gyroscope sensors. Various types of smart mobile devices have been used, such as (Huawei Mate 20, Galaxy J7 Pro, and Xiaomi Redmi 5). Where 50 volunteers were in the age group from (18 to 49 years) (number of males 31 and number of females 19) with weights between (53-90) range of physical activities including walking, running, walking upstairs, walking downstairs, and finally standing and sitting on chairs events under real-world situations as shown in the figure (3.3) displays an example of the performed activities during the collection of experimental data. Start in sequence from the top left walking, running, walking upstairs followed by the activities walking downstairs, sitting, and standing on the chair. Data obtained from the real world is the first and essential material for tasks requiring recognition following identifying types of sensors. While gathering data can be a time-consuming and labor-intensive task, the researchers may encounter a string of issues when collecting data from the real world, like annotation, how to use sensors, environment for experiment, intrusion, chronological order, volunteer capacity ...etc. Data from the real world for a particular task should be included as many target populations as much as possible of different ages, gender, height, weight, and conditions of health. Whereas because of the cost of time and will of people, the volunteers' number assigned to collect data is very limited, the data collection protocol also affects recognition performance; there are factors such as the number of participants and activities, and performance of activities in a normal or restricted manner, regulated or uncontrolled environment…etc. faced many difficulties in different locations, such as the different stairs to conduct the climbing and descending activity, the different yards in which volunteers running, the yards or places where walking was carried out, and the different chairs in which the sit-stand activity was conducted. Data is collected by researchers for their research. They can also make use of the public data sets made available by HAR to evaluate or compare their proposed methods with comparison to other research on the same data sets. The mean and the collected data's standard deviation were computed. As shown in table (3.2). Sensor time-series information from the phone and gyroscope accelerometer is shown as shown in figure (3.2). Learning evaluation and training techniques require a large number of samples data annotated. However, data collection and annotation of sensory activity is a cumbersome process and consumes time. Also, it is difficult to get data for some activities that are emerging or unexpected (For example. accidentally falling), which results in a new challenge known as a

class imbalance. To make the experience different volunteers from different segments of society gathered despite the strong opposition to conducting this experiment, whether for unknown reasons, health reasons, or starting the experiment and not completing it until the end. In this case, the volunteer experiment was neglected for not completing it. The five activities and the recognition of human activity using its features also require a substantial amount called sensor data used for training to achieve an accurate performance of recognition. To collect a substantial amount of data in real life, a lot of time is consuming. On the contrary, it is easy to get untagged data by making use of smart homes, the Internet of Things (IoT)... etc. Another factor that must be taken care of is the system's ability to recognize the human activity. Efforts should be devoted to ensuring that the system is usable by a large number of people because the identification for activity is extremely close to the everyday life of a human being, which can be two-fold. First, the system must be useful so that it is suitable for mobile devices and can give an immediate response. Therefore, it is necessary to consider the **computational cost**. Second, since the system of recognition deals with the lives of users on an ongoing basis, **privacy** is one of the most important issues that must be dealt with, including the risks related to the disclosure of personal information to make the system possible in private places, and then the data was collected by a volunteer by holding the smartphone in hand in the position the subject used to hold their phone or in the way he/she feels comfortable.

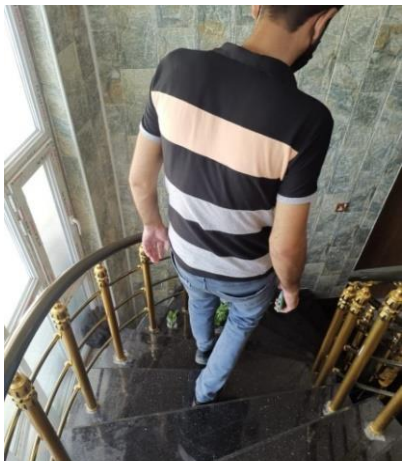
(a) Walking



(b) Running



(c) Walking upstairs



(d) Walking downstairs

(e) Standing

(f) Sitting

Figure (3.3): Activities performed during data collection in real-world conditions

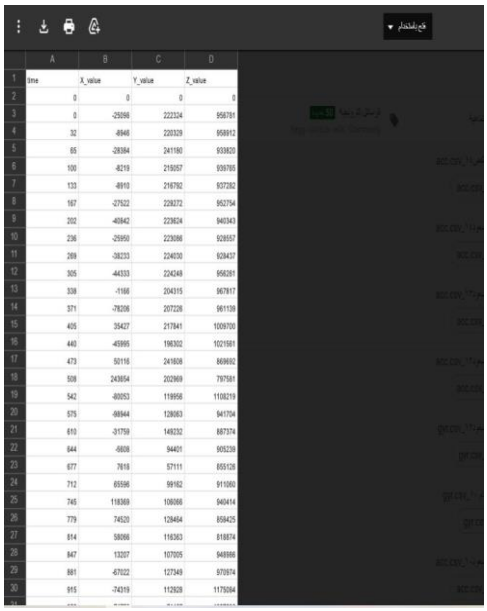
Table (3.2): The characteristics of the subjects

Subject	Total	Female	Male	Parameter
	50	19	31	
Age	18-49	19-46	18-49	Max-Min
	30.78±9.698	33.526±7.77	29.096±10.473	Mean±Std
Height	140-190	140-170	155-190	Max-Min
	166.24±11.631	155.894±6.748	172.58±9.182	Mean±Std
Weight	50-90	50-81	60-90	Max-Min
	70.88±9.356	66.947±8.8661	73.29±8.944	Mean±Std
BMI	18.559-35.555	21.359-35.555	18.559-35.338	Max-Min
	26.076±4.282	28.057±4.215	24.798±4.009	Mean±Std

3.2.2 Data organization

There are three factors on which recognition of human activity depends: sensors, time, and users. First, with the various sensors, whether they are on human bodies or in environments. Sensors greatly affect the data that the activities stimulate. Second, concepts of activity change over time. Assuming that the patterns of users' activity will remain unchanged for an extended time is impracticable. Moreover, new activities are more likely to arise when they are used. Third, activity patterns depend on the person. Different users may have different activity patterns. All three factors lead to a **differentiation** of training and test data and must be urgently mitigated. Raw time series sensor data is recorded by the accelerometer and gyroscope of a smartphone, each consisting of three X, Y, and Z axes. All experiment data was captured on Android phones, and after converting to (.CSV) file, uploaded to Google Drive folder as shown in figure (3.4) shows stages of data arrangement, and downloaded Excel file to the laptop. Sensor data is stored in separate subdirectories. Each volunteer is arranged with its profile, accelerometer, and gyroscope data for each device stored in two subdirectories. There is a file for each topic within each subdirectory, and all volunteers were collected in one file after each volunteer was numbered. In the end, a folder was obtained consisting of 500 Excel files, merging accelerometer and gyroscope data into one folder

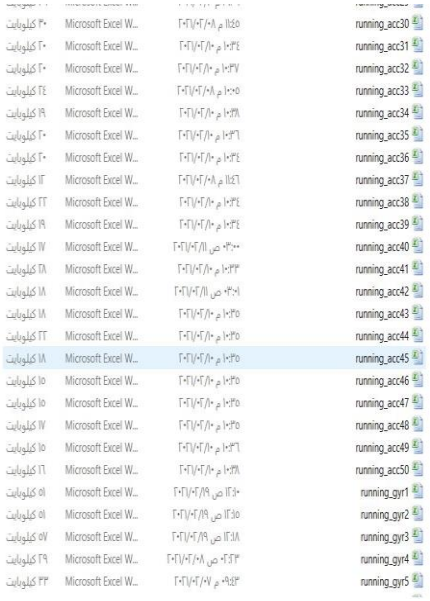
containing all activities for 50 volunteers including walking, running, walking upstairs, walking downstairs, and finally standing and sitting on chairs.



(a) Google Drive

(b) Create a special folder for each volunteer

(c) The folder includes activities for each volunteer



(d) Folder includes all activities

Figure (3.4): Data sorting stages

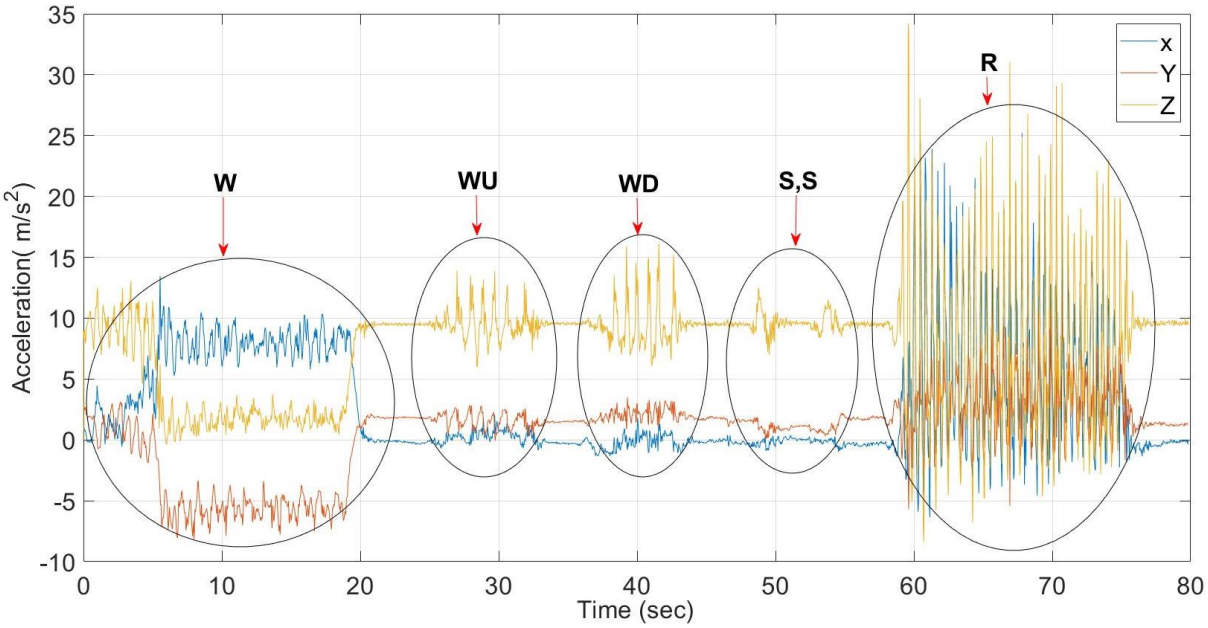
3.2.3 Converting time series to 2D images using CWT

Sensory accuracy data is complex and unreadable, unlike images or text. Sensory data inevitably includes a lot of noise information due to inherent defects in the sensors or abnormal heights on the smartphone, due to an unintended change in the direction of the sensor, a certain change in the position of the device or its fall, or noise caused by dynamic movement of humans. Therefore, it must have the reliability of identification solutions, the ability to **interpret** perceptual data, as well as the ability to understand which portion of the data facilitates identification and which portion degrades. Other challenges are problems of diversity within a class and similarity between classes. In this instance, Different people may perform the same activities in different ways, or it appears that the various activities follow a similar pattern. Filtering systems remove signal noise from the accelerometer and preserve intermediate frequency signal components. An important process for validating the validity of the filtered and normalized database is data transformation. The square root of a value is another non-linear process used for statistical analysis. The Signal Vector Magnitude (SVMag) method is used to convert data from three dimensions (the accelerometer and gyroscope sensors are three dimensions(3D) (Ax, Ay, Az) (Gx, Gy, Gz) respectively) into one dimension(1D) as shown in figure (3.5) this facilitates the process of converting the time series a series of temporally indexed data points (inserted or graphed). Particularly often, a time series is a sequence captured at successive equidistant time points on images. Apply the following equation [94-96] for conversion.

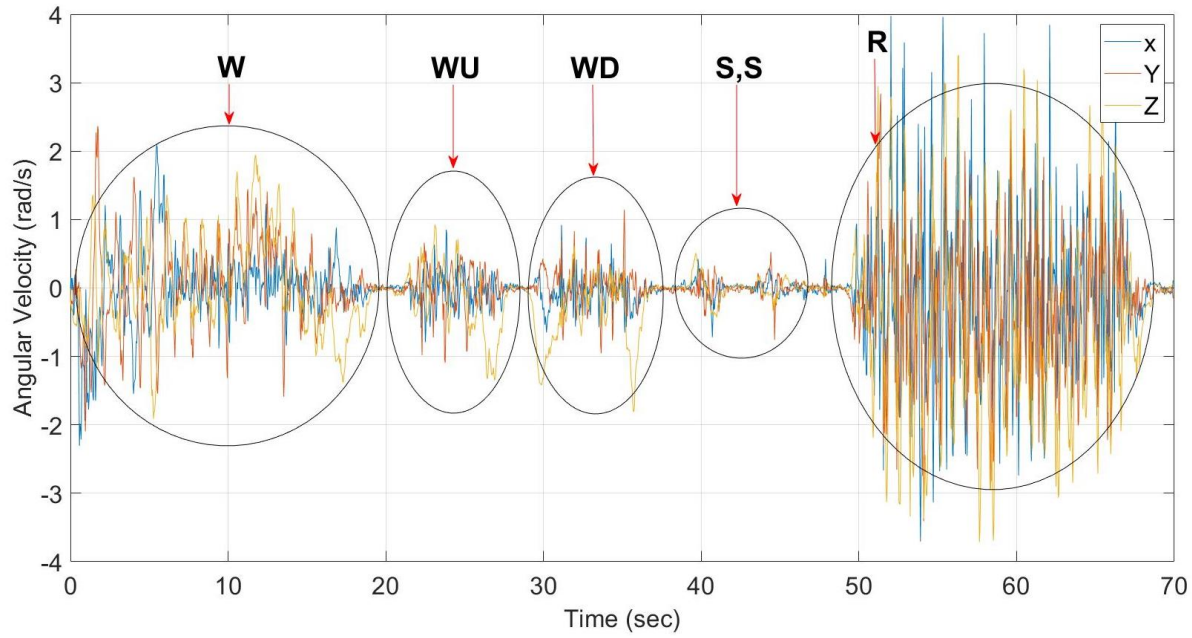
$$\text{SVMag for Accelerometer} = \sqrt{A_x^2 + A_y^2 + A_z^2} \tag{3.1}$$

$$\text{SVMag for Gyroscope} = \sqrt{G_x^2 + G_y^2 + G_z^2} \tag{3.2}$$

Color maps are an extremely important way for users' to earn insights into data visualization. Users can obtain more effectively and efficiently extract information from data by using a good selection of color maps. Figure (3.6) shows the stages of transformation of time series into images.

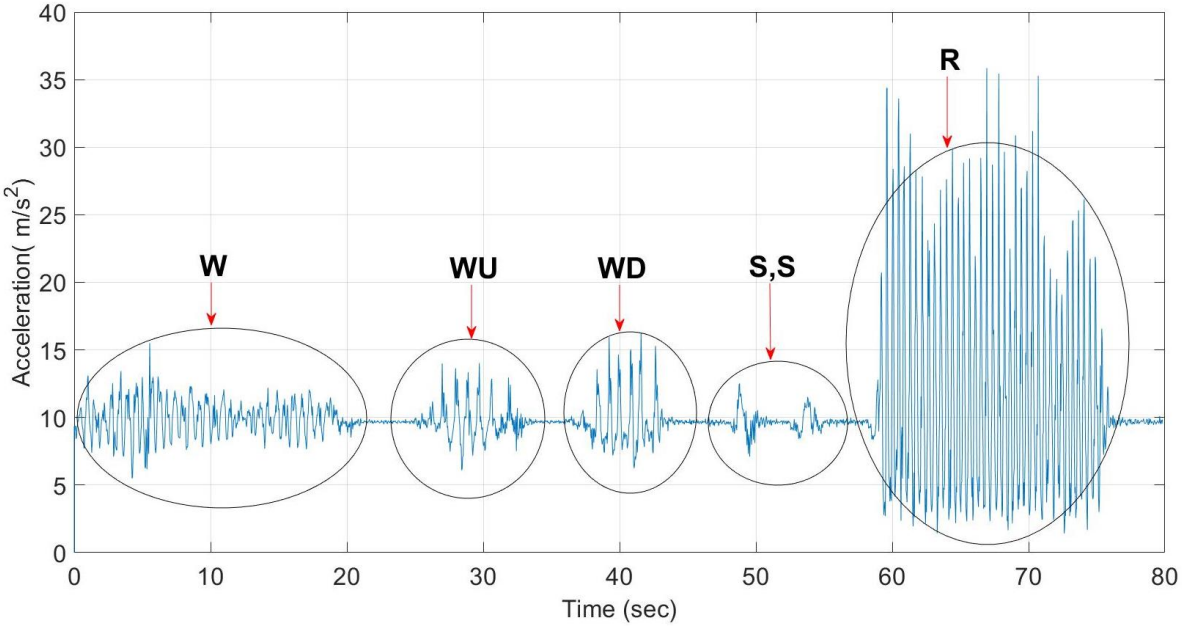


(a) Accelerometer for all activity (three-dimension)

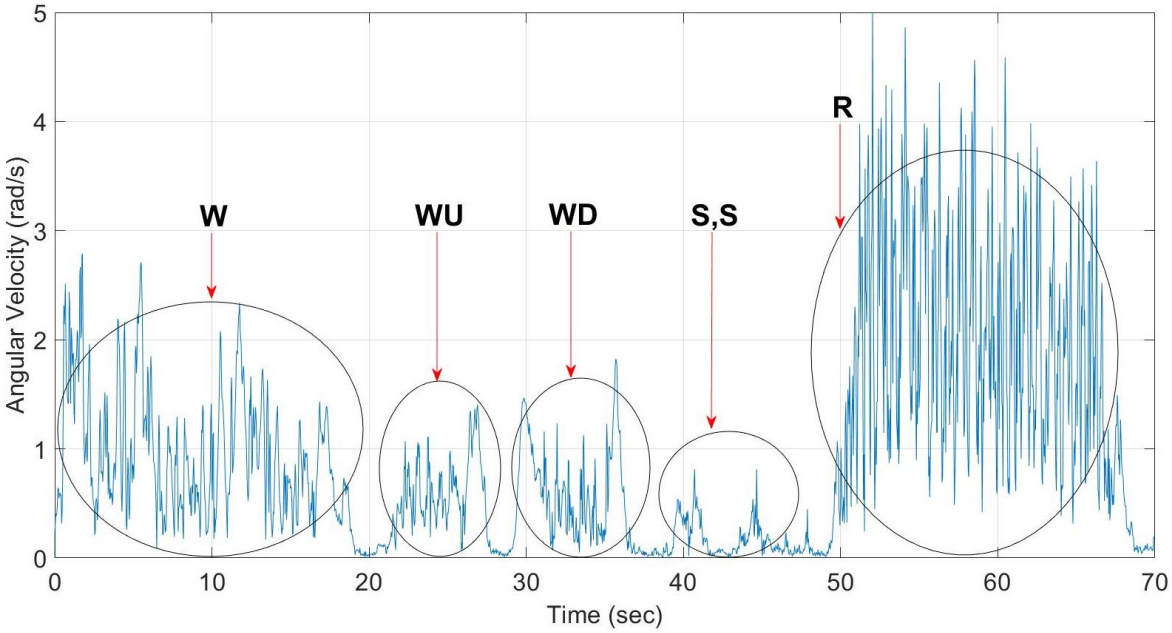


(b) Gyroscope for all activity (three-dimension)

(W=Walking, WU=Walking upstairs, WD=Walking downstairs, S, S=Standing, Sitting, R=Running)

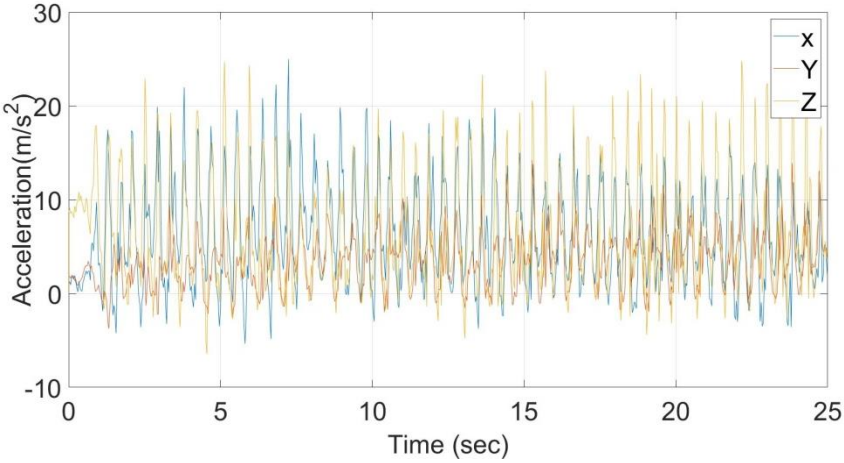


(c) Accelerometer for all activity (one- dimension)

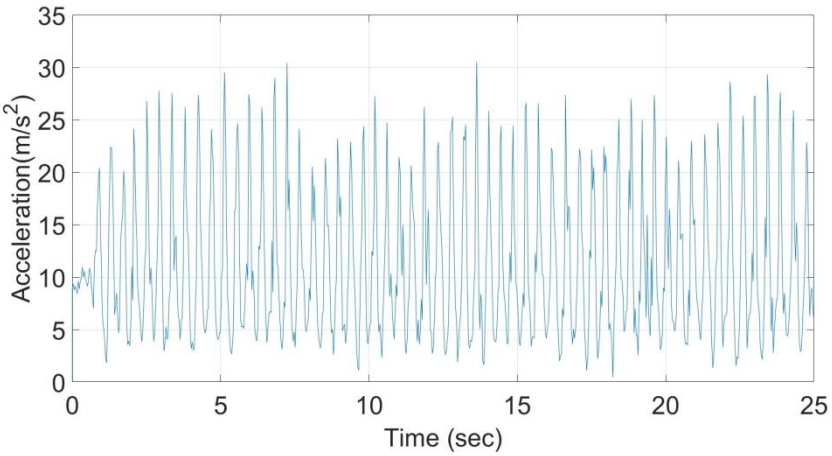


(d) Gyroscope for all activity (one- dimension)

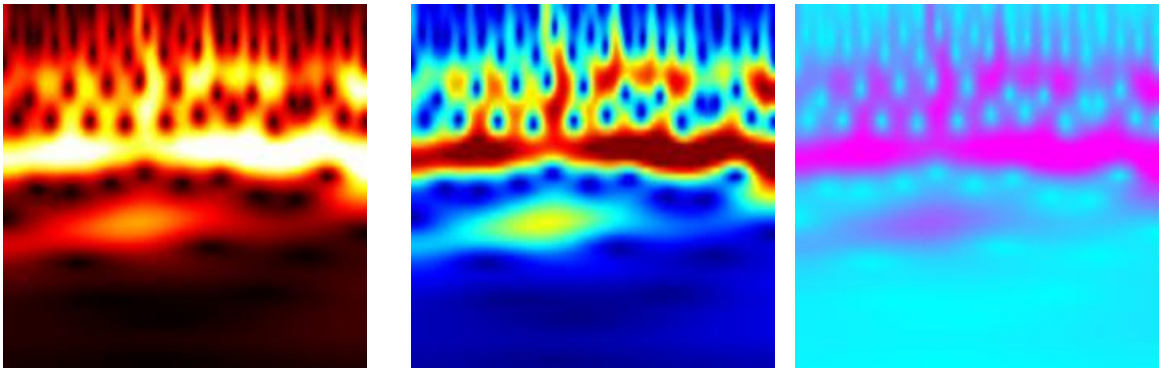
Figure (3.5): Conversion of all activity from three-dimensional to one-dimension using Signal Vector Magnitude



(a) Time series for a running activity



(b) Signal vector magnitude for running activity



(c) Image running activity using three different types of the color map (hot, jet, and cool)

Figure (3.6): Stages of converting time series to images using CWT

3.2.4 Convolution neural network architecture

Over the past years, several CNN builds have been introduced. The architecture of the model is an important factor in enhancing the various applications' performance. The CNN architecture has undergone several modifications. Structural refactoring, regularization, and optimizations of parameters are the modifications in the CNN. The primary reason for considering CNN is the feature of weight sharing; this reduces the number of the parameters for trainable networks which contributes to the network's improvement generalization and avoidance of over-fitting. The data was collected from a smartphone, and converting this data into color images using a continuous wavelet. Then splitting the data into training, validation, and testing as a CNN training input to classify activities. Figure (3.7) shows the general workflow. The CNN as explained in the previous chapter consists of three different parts, the input layer organizes the input data into a CNN model after it is converted into images, so the pre-processing of the data is necessary for a CNN, and the convolution layer extracts the features and uses the activation function to remove Non-linear data. Nonlinearity is a principle of a neural network. The last part is the output, it requires a fully connected layer to receive all feature vectors from the hidden layers and another fully connected layer with a softmax activation function to shrink the feature vectors, and then to prevent over-fitting, added a dropout layer to remove the extreme situation. The CNN architecture model consists of five convolutions layers (C1, C2, C3, C4and C5), five batch normalization layers (BN1, BN2, BN3, BN4andBN5), five rectified linear units, five max-pooling layers, one dropout layer, and one fully connected layer.

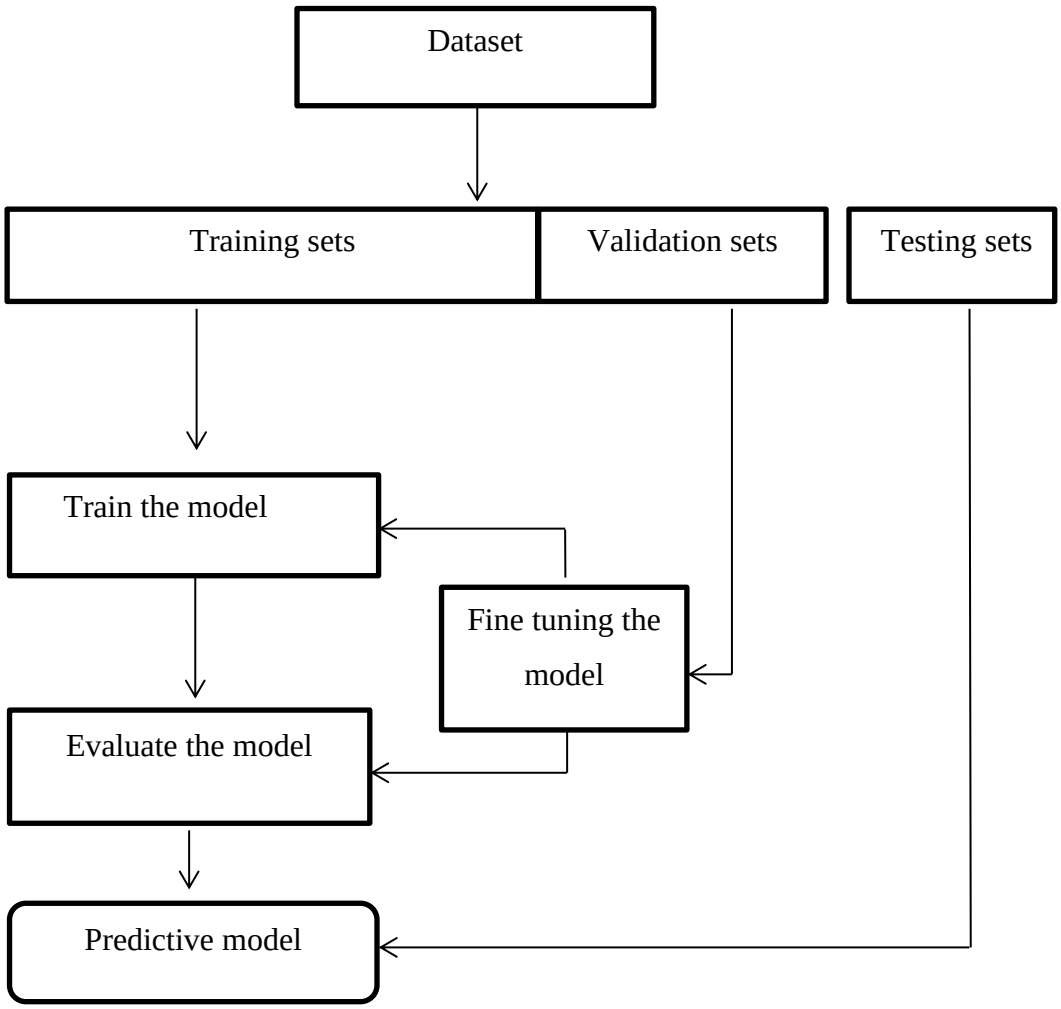


Figure (3.7): CNN workflow

Chapter Four

Results and Discussion

4.1 Introduction

This chapter presents in detail the results of the proposed system. A CNN classifier is selected to classify five human activities as a part of the HAR system based on smartphone inertial sensors. The results of the inducted experiment are organized in a time series dataset including accelerometer and gyroscope data. Three families of wavelets were tested and two types of color maps were chosen to select the best combination of them in terms of representing human activities when converting time series into images. Then, the hyperparameter of the proposed model was chosen and the structure was optimized to design a better CNN architecture. Finally, the results of classifier performance metrics such as accuracy are presented and fair comparisons with previous work are conducted.

4.2 Real Environment Dataset

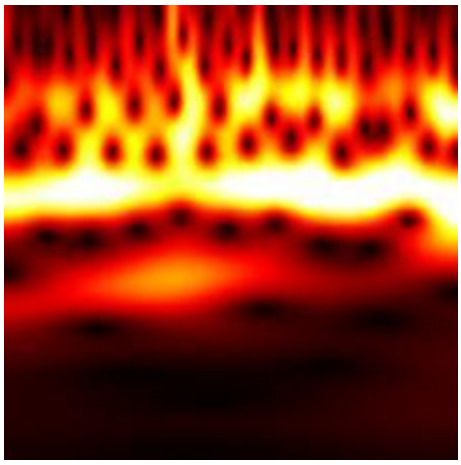
For the development of the HAR model, a data set was collected with the five activities (running, walking, walking upstairs, walking downstairs, and standing, sitting on chairs) the main purpose of this data set was to train and evaluate the performance of the proposed HAR model. This data set was collected using smartphone accelerometers and gyroscopes with the help of 50 volunteers in real-life conditions. Details of data collection, number, and number of images are shown in table (4.1)

Table (4.1): Dataset characteristics

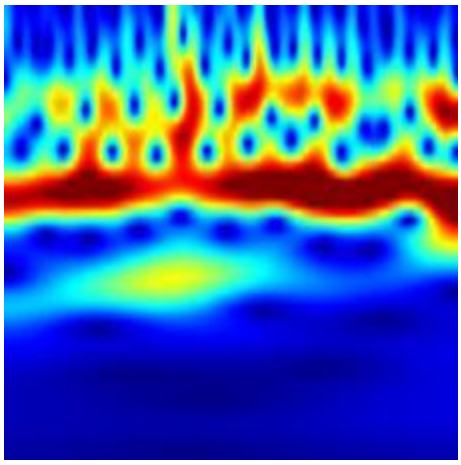
Collected data	Time series X, Y, Z, Number of activities = 5, Sample rate=30	
	Total number of volunteers = 50	
	Accelerometer or Gyroscope excel sheet = 50	
	Total excel sheet for each activity = 100	
	The number of excel sheets 5 activities for 50 volunteers is= 500	
Data preparation	Number of images for 50 volunteer Accelerometer or Gyroscope = 7500 Number of images for 50 volunteers all =15000	
Activity ID	Running Standing, Sitting on chairs Walking Walking upstairs Walking downstairs	**R** **S, S** **W** **WU** **WD**

4.3 Image generation using CWT

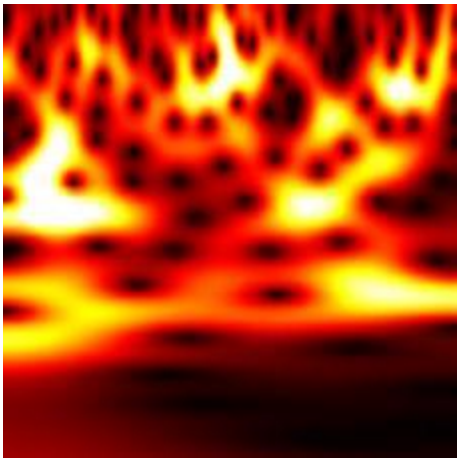
The continuous wavelet transformation displays the signal's scale-dependent structure as it changes over time. As a result, CWT displays the signal's frequency versus time behavior. In this thesis, the results of time series analysis will be used to convert them into images for training using a convolutional neural network. Three types of continuous wavelet families Amor, Morse, and Bump were used by applying two types of color maps hot and jet as shown in figure (4.1). The Amor of the wavelet family with the hot color map is distinguished by its clear representation of the movement of the activity; it is good to be used for classifying pictures.



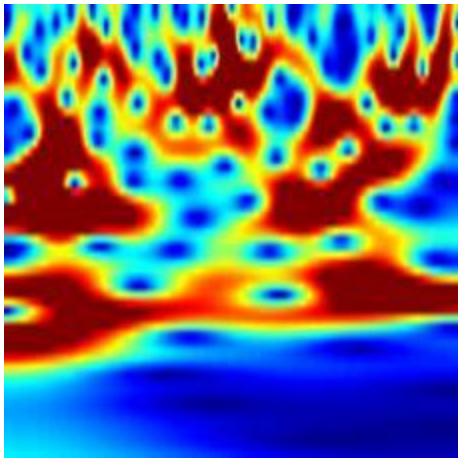
(a)Amor wavelet hot colormap



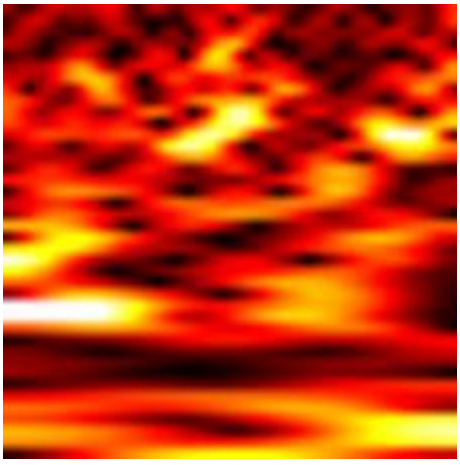
(b) Amor wavelet jet colormap



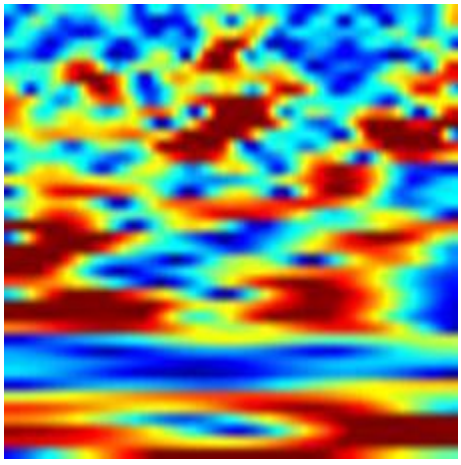
(c)Morse wavelet hot colormap



(d)Morse wavelet jet colormap



(e)Bump wavelet hot colormap



(f) Bump wavelet jet colormap

Figure (4.1): Three different types of wavelet family for running activity

4.4 CNN model

The hyperparameters as mentioned in table (4.2) were mini-batch size data, the learning rate, and the epoch number. Batch size refers to the number of trials that CNN will train or test the network on in one period. The number of periods thus indicates the number of times the entire training data set passes through the CNN at once. 20 epochs remain constant because changing them to a shorter number would cause the test accuracy to be reduced by a greater or greater extent than the epochs in which accuracy is established and the time to completion of training increases. The learning rate is the final parameter and is a numerical value that the algorithms multiply in the gradient to determine the next point. If the learning rate is too small, it will take a long time to learn, but if it is too large, a lot of data will be bypassed and lost. Thus there is an optimal learning rate value that must be found for each set of data

Table (4.2): The hyperparameters used in the experiment

Splitting data	Training sets = 0.6
	Validation sets = 0.1
	Testing sets = 0.3
Training Options	The number of layers = 25
	The learning rate = 0.001
	The mini-batch size = 30
	The validation frequency = 30
	The max epochs = 20
	Dropout =50%

CNN models are generated using a combination of three major layers convolutional, pooling, and fully connected layers as shown in fig (4.2). The first layer convolutional is a basic building block for CNN models, which tries to learn to map features that are spatially activated via lots of filters via inputs while sharing parameters. The pooling layer is mainly used to reduce the size of feature representations, and the number of parameters learned in the convolutional layer. The fully connected layer connects to all the outputs from its previous layer. Table (4.3) shows the CNN architecture.

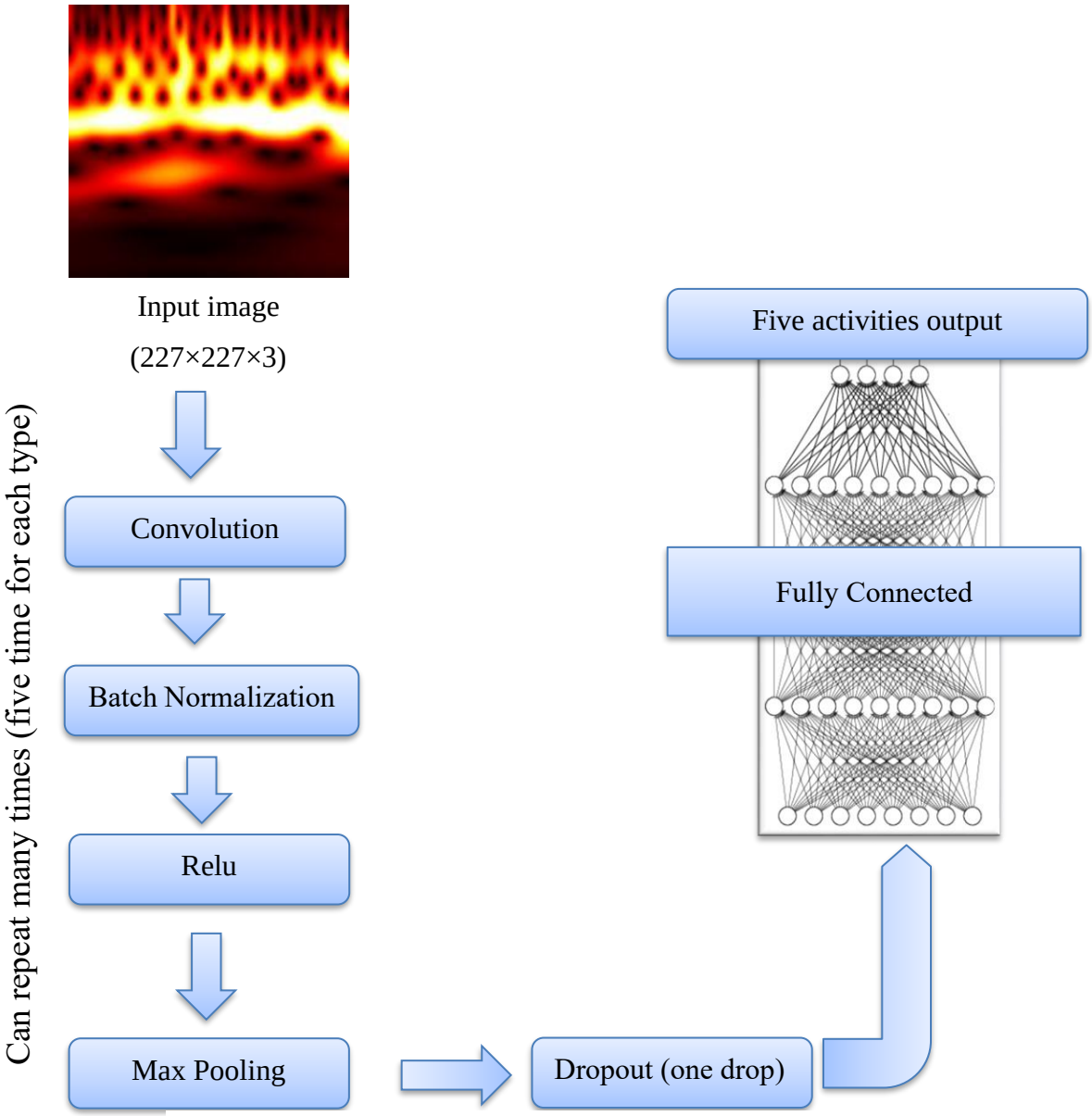


Figure (4.2): CNN architecture for proposed system

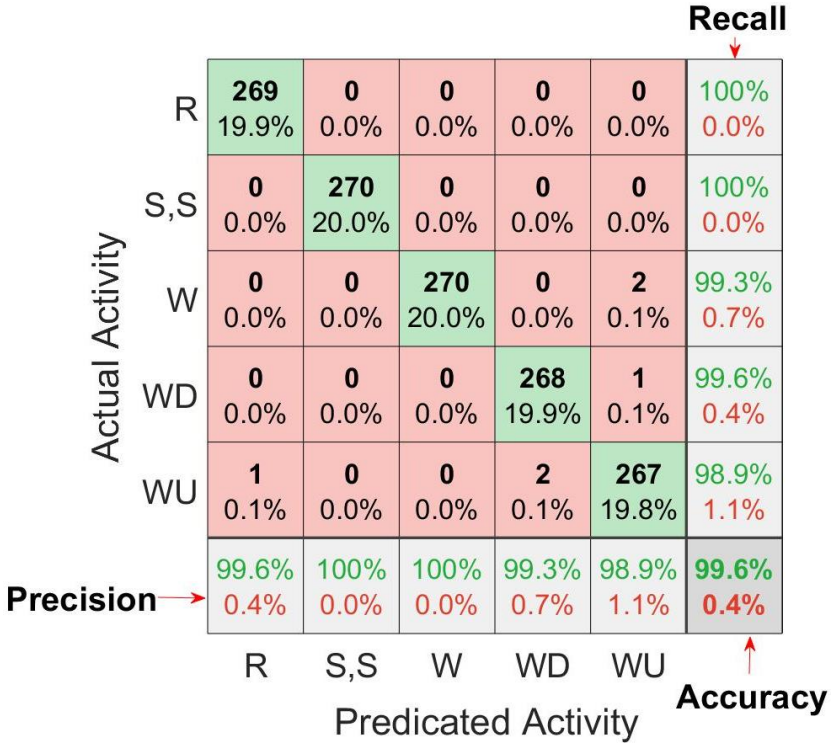
Table (4.3): CNN architecture

layer	Type/ layer name	Size	Other parameters
1	Image input	227×227×3	
2	Convolution (conv_1)	227×227×8	FS=[3,3] NF=8 S=[1,1]
3	Batch Normalization (BN_1)	227×227×8	
4	Relu (relu_1)	227×227×8	
5	Max Pooling (maxpool_1)	113×113×8	PS=[2,2] , S=[2,2]
6	Convolution (conv_2)	113×113×16	FS=[3,3], NF=16 , S=[1,1]
7	Batch Normalization (BN_2)	113×113×16	
8	Relu (relu_2)	113×113×16	
9	Max Pooling (maxpool_2)	55×55×16	PS=[5,5] , S=[2,2]
10	Convolution (conv_3)	55×55×32	FS=[3,3] ,NF=32 , S=[1,1]
11	Batch Normalization (BN_3)	55×55×32	
12	Relu (relu_3)	55×55×32	
13	Max Pooling (maxpool_3)	25×25×32	PS=[6,6] , S=[2,2]
14	Convolution (conv_4)	25×25×64	FS=[3,3], NF=64 , S=[1,1]
15	Batch Normalization (BN_4)	25×25×64	
16	Relu (relu_4)	25×25×64	
17	Max Pooling (maxpool_4)	9×9×64	PS=[8,8] , S=[2,2]
18	Convolution (conv_5)	9×9×64	FS=[3,3] ,NF=64 , S=[1,1]
19	Batch Normalization (BN_5)	9×9×64	
20	Relu (relu_5)	9×9×64	
21	Max Pooling (maxpool_5)	3×3×64	PS=[5,5] , S=[2,2]
22	Dropout (drop1)	3×3×64	
23	Fully Connected (fc)	1×1×5	
24	Softmax (softmax)	1×1×5	
25	Classification Output (classOutput)		

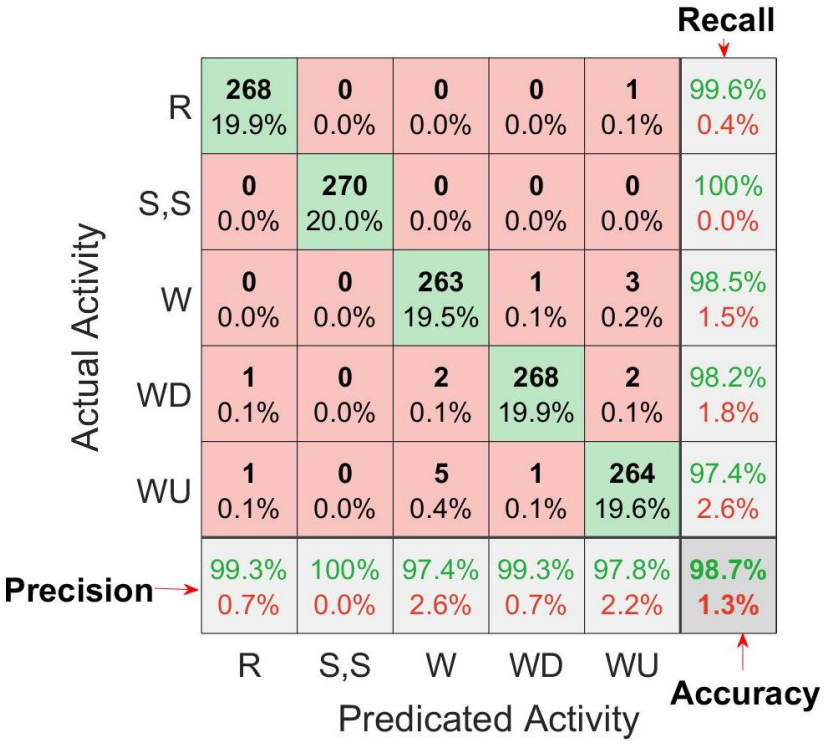
(FS=filter size,NF= numper of filter,S= stride,PS=poolsize)

4.5 Evaluation of different types of wavelet families and image color map

The first classifications made to an image data set obtained from the application of three wavelet species Amor, Morse and Bump were applied to 30 volunteers. Color mapping can help to get an accurate visual understanding of the images and thus These groups were first analyzed using two types of colored maps (hot and jet) to distinguish which types of the family are better compared to the representation of movement and high accuracy. First, Amor was applied to classify the activities with two types of the color map, which are hot and jet, and the accuracy was equal to (99.6%) in the case of hot and (98.7%) in the case of jet, as shown in the confusion matrix Fig(4.3). Secondly, Morse was applied to classify activities, and the accuracy was (98.4%) in the case of hot and (99.0%) in the case of jet, as shown in the confusion matrix of fig (4.4). Thirdly, the Bump application to classify the activities, and the accuracy was (96.8%) in the case of hot and (97.1%) in the case of the jet as shown in the confusion matrix of fig (4.5). conclude from the above that Amor using hot got the highest results inaccuracy as well as the best representation of the image. Table (4.4) shows a comparison between the Wavelet families using two types of the color map. The algorithm evaluation metrics were calculated, such as accuracy, recall, and precision based on the results obtained.

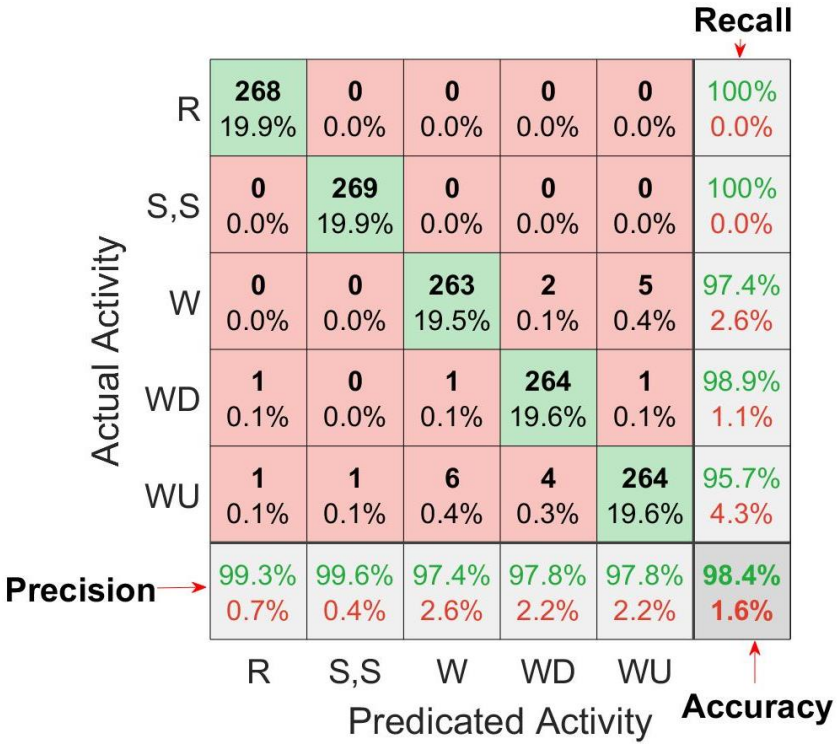


(a) Confusion matrix for testing (Amor /hot)

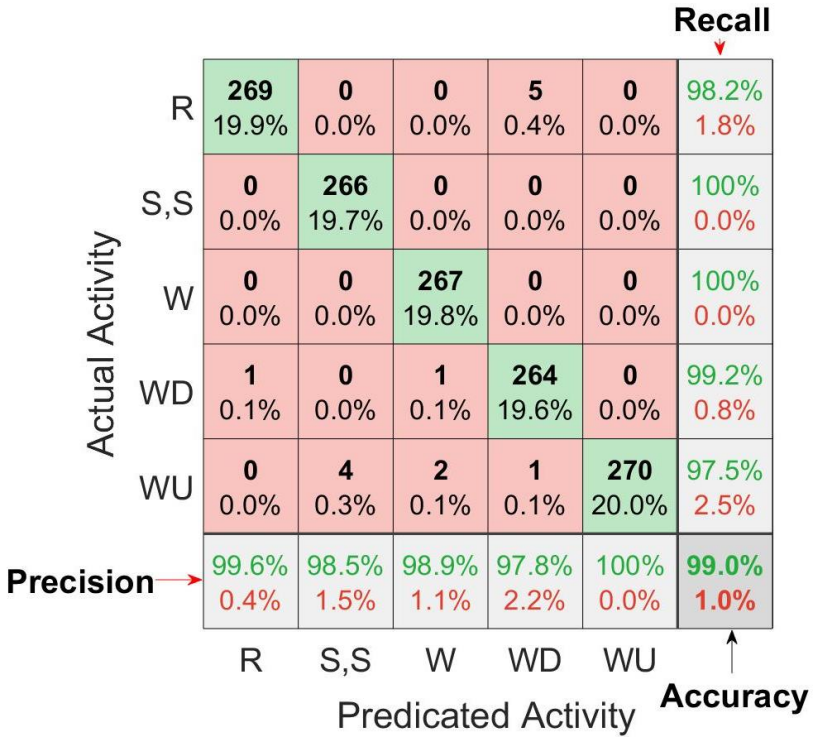


(b) Confusion matrix for testing (Amor/ jet)

Figure (4.3): Confusion matrix Amor wavelet for the accelerometer (30 volunteers)

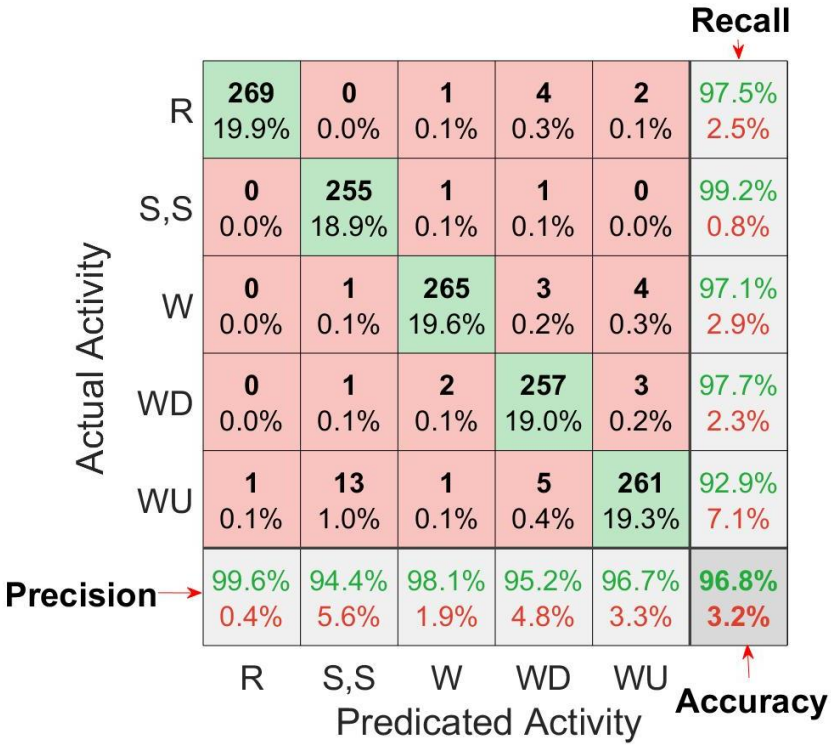


(a) Confusion matrix for testing (Morse /hot)

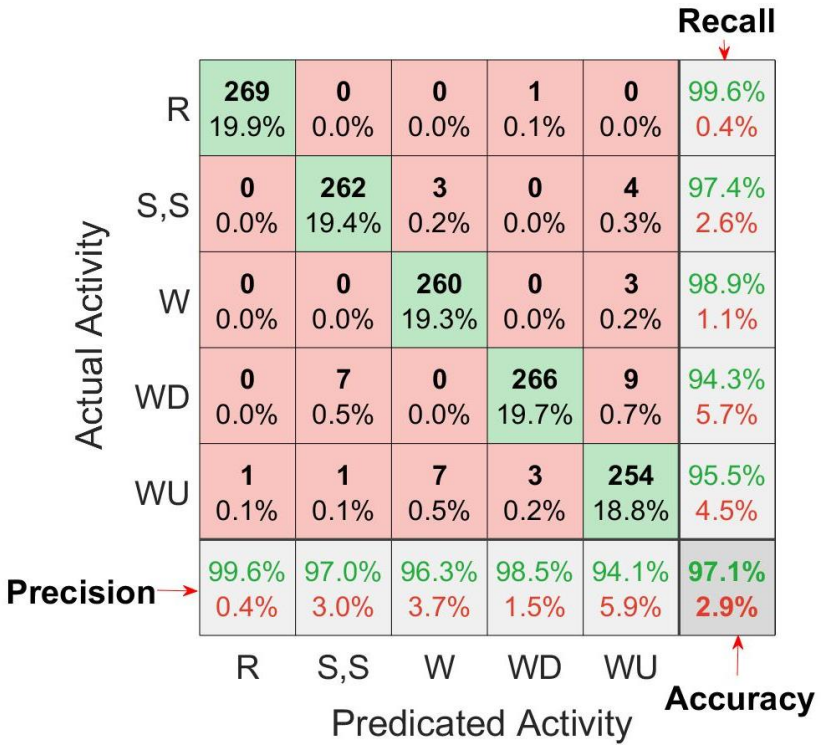


(b) Confusion matrix for testing (Morse/ jet)

Figure (4.4): Confusion matrix Morse wavelet for the accelerometer (30 volunteers)



(a) Confusion matrix for testing (Bump /hot)



(b) Confusion matrix for testing (Bump/ jet)

Figure (4.5): Confusion matrix Bump wavelet for the accelerometer (30 volunteers)

Table (4.4): Comparing three types of wavelet families using two types of color map

	hot				jet			
	activity	recall	precision	accuracy	activity	recall	precision	accuracy
Amor	R	100%	99.6%	**99.6%**	R	99.6%	99.3%	98.7%
	S,S	100%	100%		S,S	100%	100%	
	W	99.3%	100%		W	98.5%	97.4%	
	WD	99.6%	99.3%		WD	98.2%	99.3%	
	WU	98.9%	98.9%		WU	97.4%	97.8%	
	Average	99.56%	99.56%		Average	98.74%	98.76%	
Morse	R	100%	99.3%	98.4%	R	98.2%	99.6%	99.0%
	S,S	100%	99.6%		S,S	100%	98.5%	
	W	97.4%	97.4%		W	100%	98.9%	
	WD	98.9%	97.8%		WD	99.2%	97.8%	
	WU	95.7%	97.8%		WU	97.5%	100%	
	Average	98.4%	98.38%		Average	98.98%	98.96%	
Bump	R	97.5%	99.6%	96.8%	R	99.6%	99.6%	97.1%
	S,S	99.2%	94.4%		S,S	97.4%	97.0%	
	W	97.1%	98.1%		W	98.9%	96.3%	
	WD	97.7%	95.2%		WD	94.3%	98.5%	
	WU	92.9%	96.7%		WU	95.5%	94.1%	
	Average	96.88%	96.8%		Average	97.14%	97.1%	

4.6 HAR classification results

In this section, the results obtained from data collection were calculated by analyzing the capability of the accelerometer and gyroscope when used separately and then combined. All our experiments are conducted in MATLAB 2019 environment on a normal PC with CPU 2.70 GHz and 16GB memory to improve neural network training.

To train and evaluate the proposed model HAR data set containing movement data of 50 volunteers performing five activities. Classification accuracy is displayed for existing classifiers, when the dataset contains 50 volunteers only the data collected from the accelerometer sensor, converting the time series by CWT Amor wavelet type to images using hot colormap type the figure (4.6) showing the training progress relationship of epoch with accuracy on the one hand, and loss on the other. The case-based group classifier provides improved results. The overall classification accuracy test is 99.1% as shown by the confusion matrix in figure (4.7).

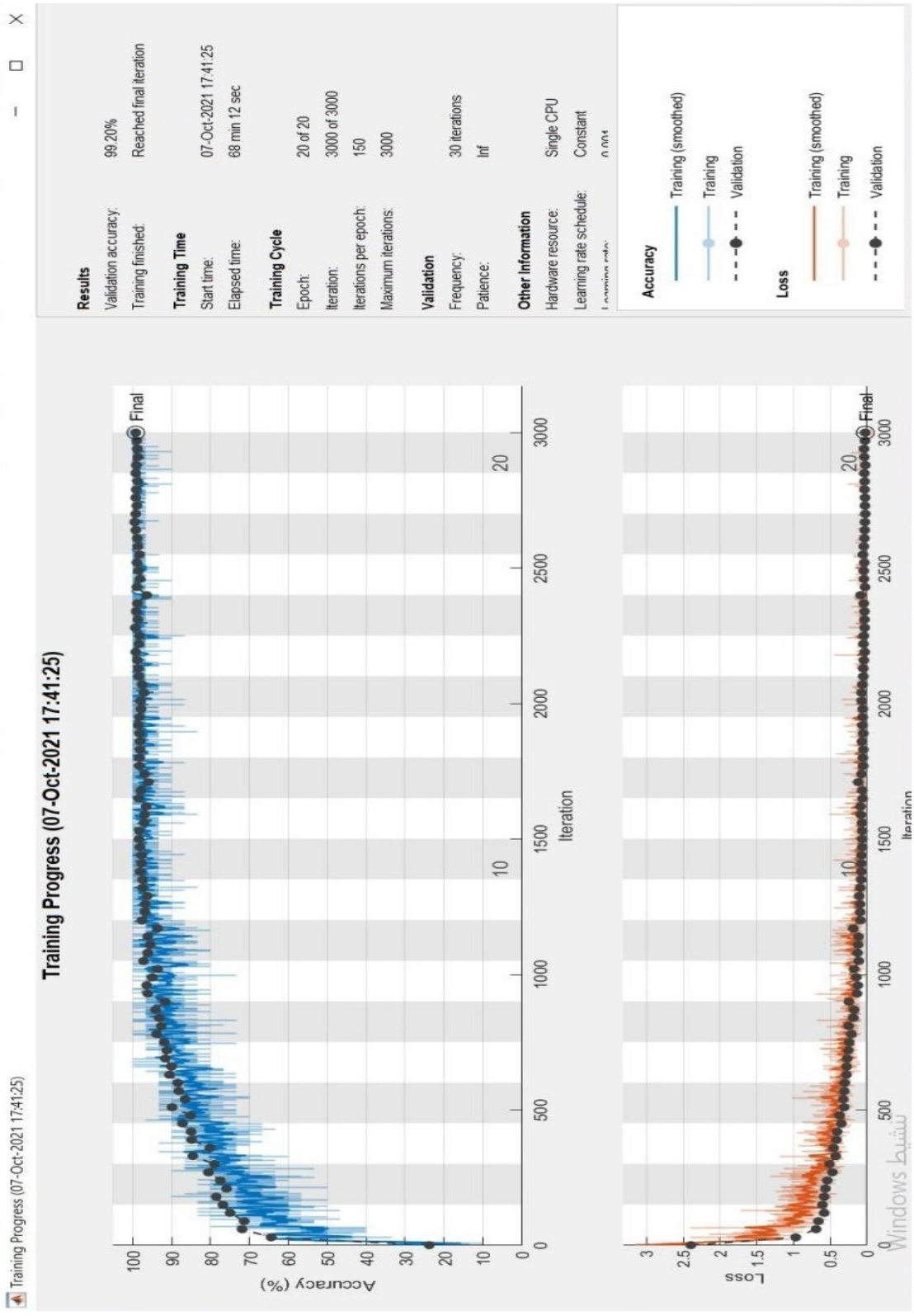
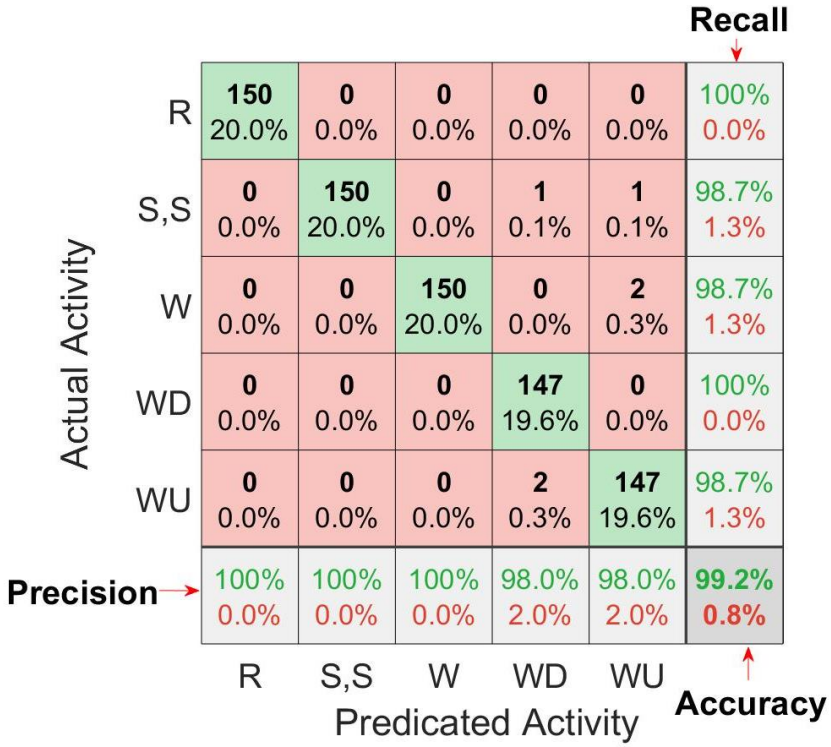
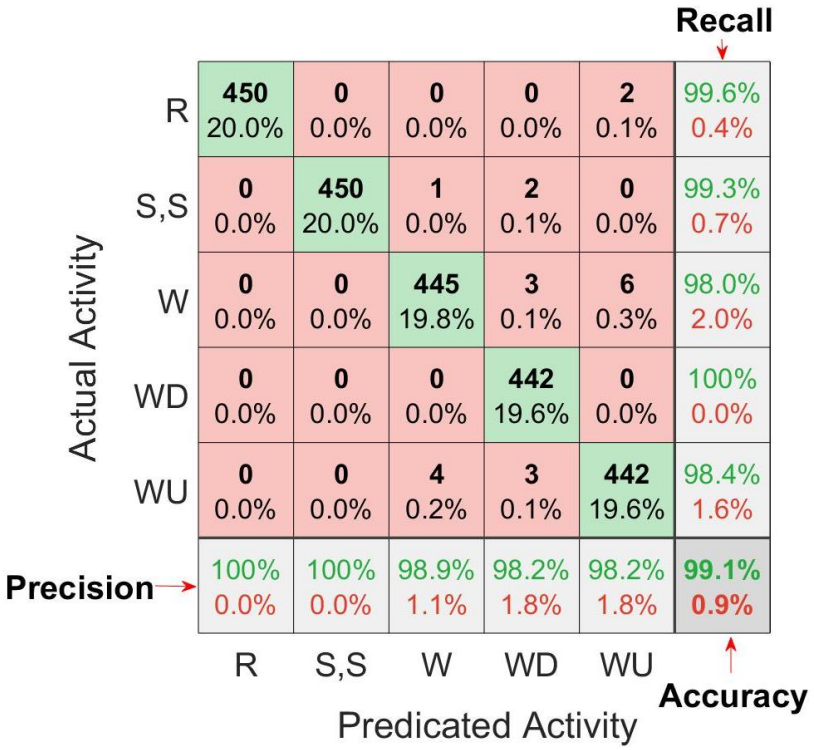


Figure (4.6): Relation between accuracy and epochs (above) and relation between loss and epochs (below)



(a) Confusion matrix for training



(b) Confusion matrix for testing

Figure (4.7): Confusion matrix for the accelerometer (50 volunteers)

The performance-specific classification model is being evaluated using a collection of metrics that show mathematically, how reliable the HAR model is. The most widely employed rating scales in HAR literature based on smartphones are accuracy, precision, sensitivity, and F-score as shown in Table (4.5).

Table (4.5): Evaluation metrics of proposed algorithm (for the accelerometer 50 volunteers)

Measure	R	S,S	W	WD	WU
Sensitivity, Recall	99.6%	99.3%	98%	100%	98.4%
Precision	100%	100%	98.9%	98.2%	98.2%
F-score	99.8%	99.6%	98.4%	99.1%	98.3%
Accuracy	99.1%				

In the same way, training data was collected from the gyroscope sensor. Figure (4.8) shows the relationship of epoch to accuracy and loss. The overall classification accuracy is 96.5% as shown by the confusion matrix in figure (4.9). The performance measures of the proposed algorithms were calculated as shown in table (4.6).

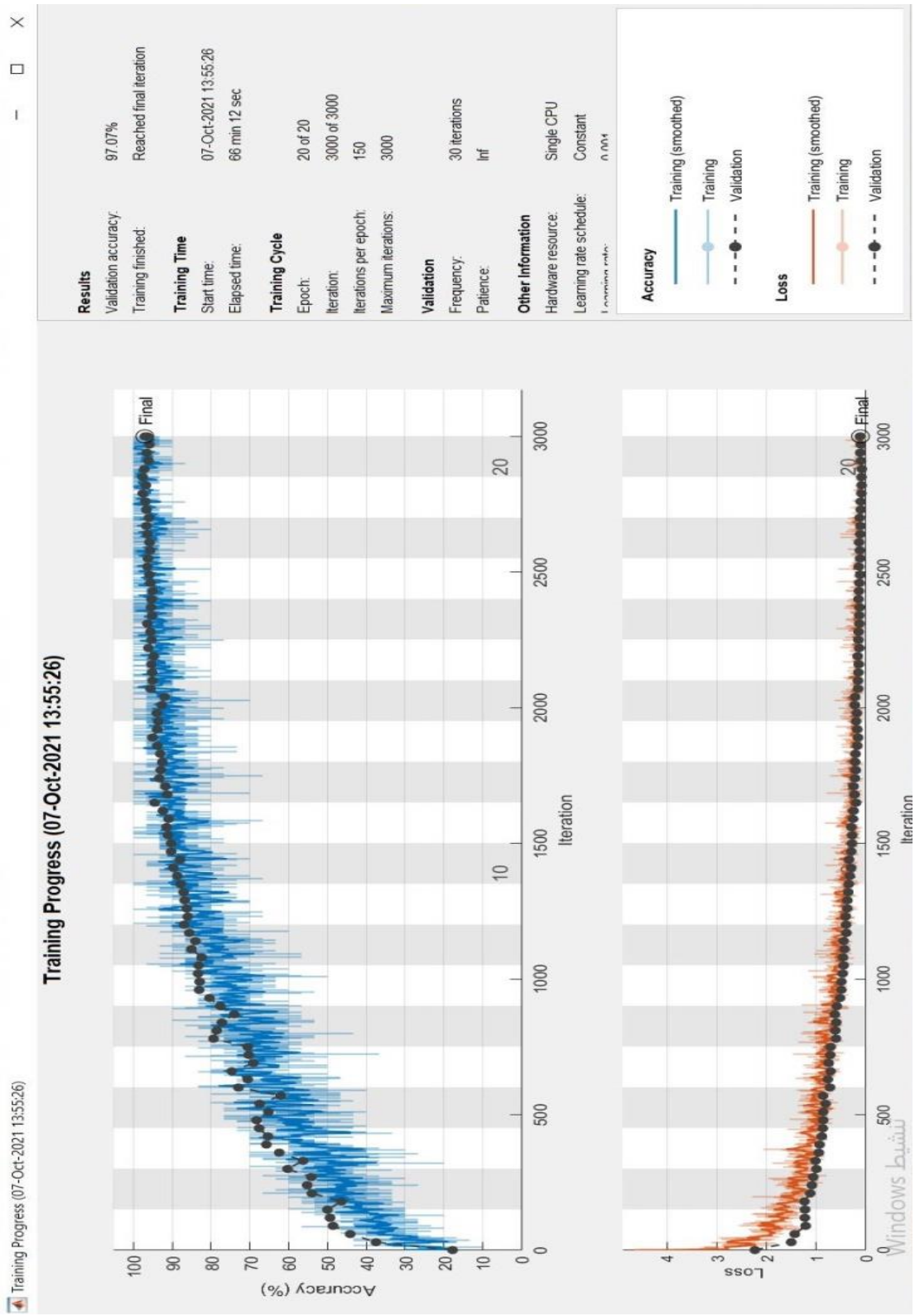
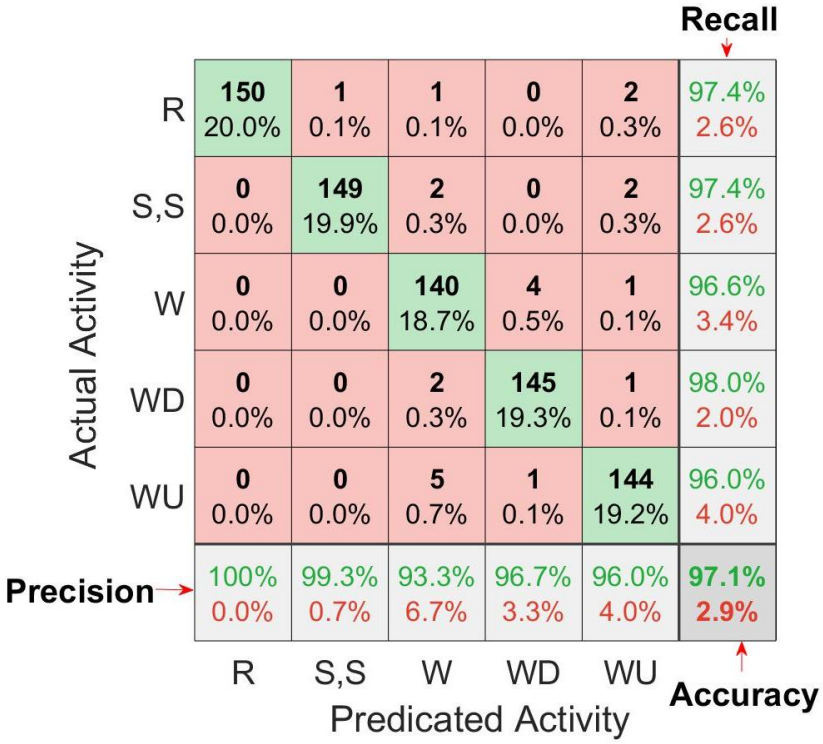
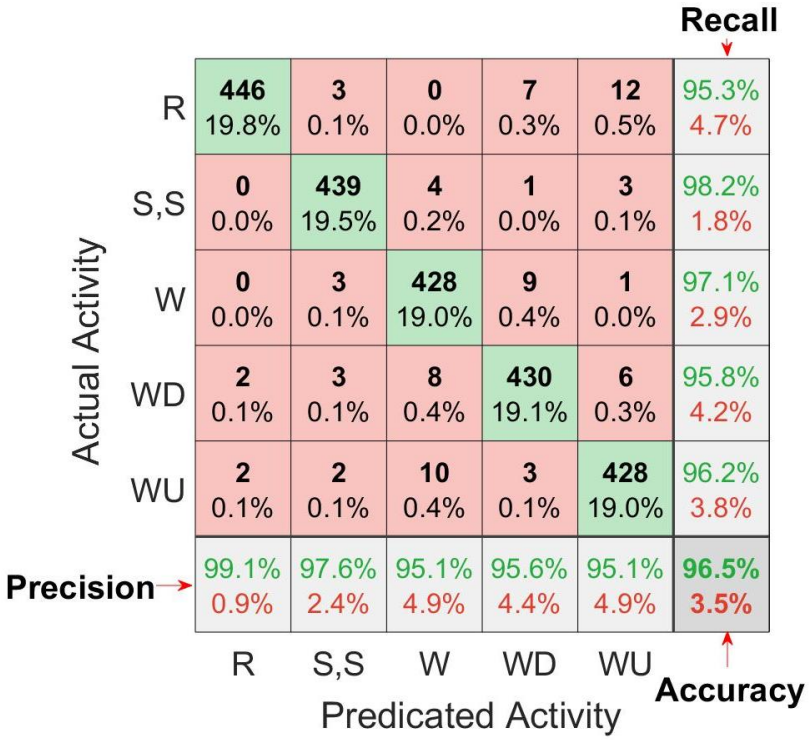


Figure (4.8): Relation between accuracy and epochs (above) and relation between loss and epochs (below)



(a) Confusion matrix for training



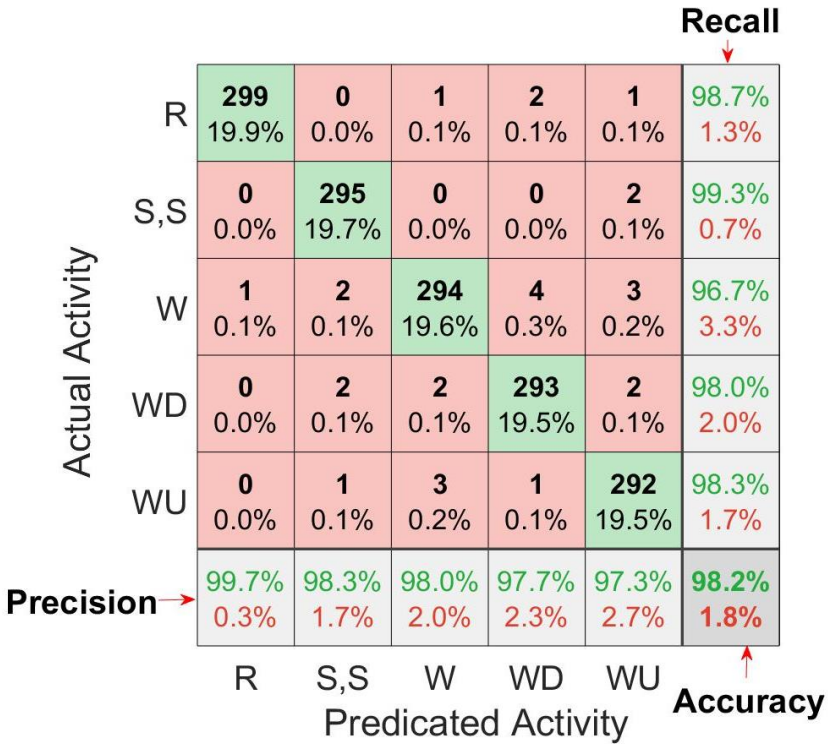
(b) Confusion matrix for testing

Figure (4.9): Confusion matrix for gyroscope (50 volunteers)

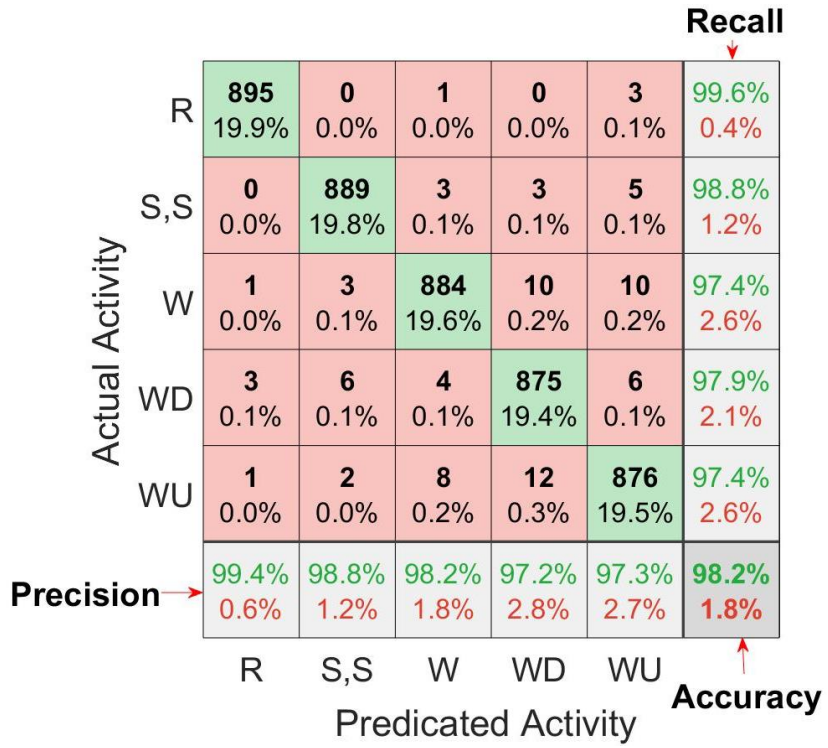
Table (4.6): Evaluation metrics of proposed algorithm (for gyroscope 50 volunteers)

Measure	R	S,S	W	WD	WU
Sensitivity, Recall	95.3%	98.2%	97.1%	95.8%	96.2%
Precision	99.1%	97.6%	95.1%	95.6%	98.1%
F-score	97.2%	97.9%	96.1%	95.7%	97.1%
Accuracy	96.5%				

Experiments are conducted to identify the activity of the smartphone using a group classifier using the data collected from both the accelerometer and gyroscope, and the results are reported for the general experimental procedure with an accuracy of 98.2% as shown in the confounding matrix figure (4.10). The rating scales were calculated to measure the performance of proposed algorithms as shown in table (4.7). From the results obtained, the accelerometer shows a high accuracy compared to the gyroscope sensor or when both accelerometer and gyroscope sensors are combined.



(a) Confusion matrix for training



(b) Confusion matrix for testing

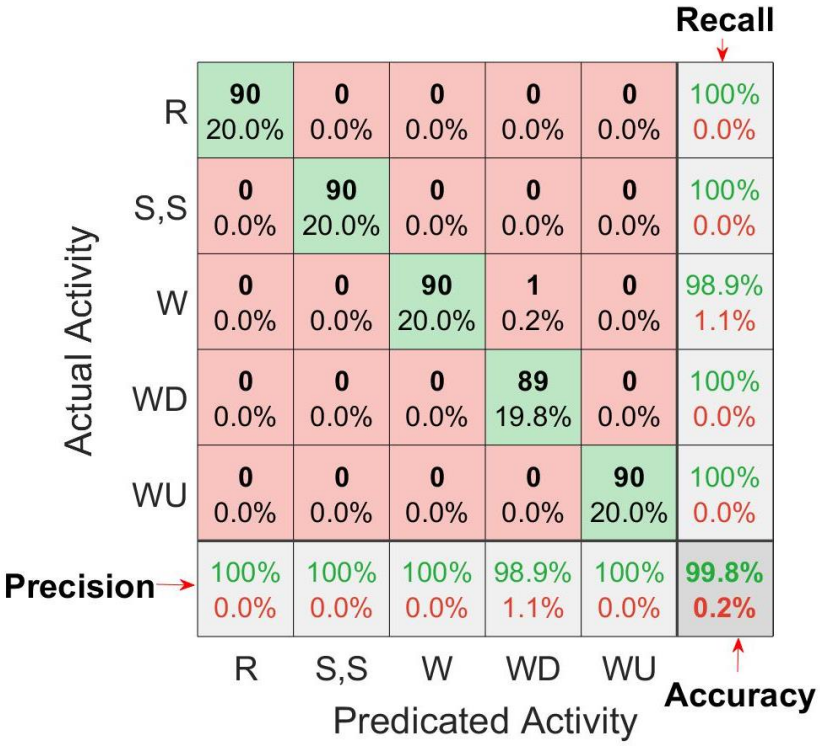
Figure (4.10): Confusion matrix for accelerometer and gyroscope (50 volunteers)

Table (4.7): Evaluation metrics of proposed algorithm (for both 50 volunteers)

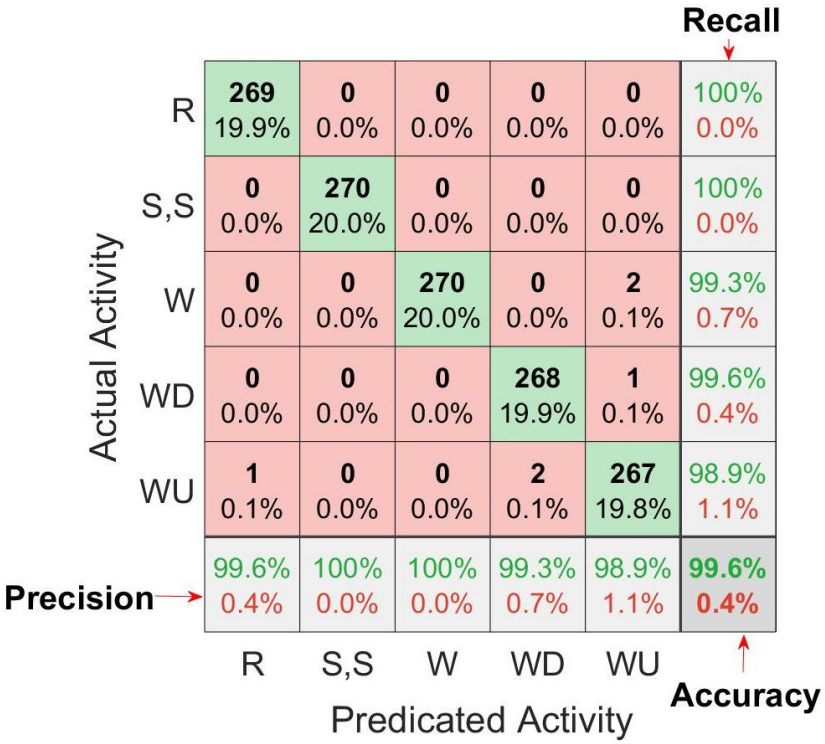
Measure	R	S,S	W	WD	WU
Sensitivity, Recall	99.6%	98.8%	97.4%	97.9%	97.4%
Precision	99.4%	98.8%	98.2%	97.2%	97.3%
F-score	99.5%	98.8%	97.8%	97.5%	97.3%
Accuracy	98.2%				

4.7 Results of a 30 volunteers' sub-dataset

To compare with the data set that was suggested in the previous research that used the most popular general data set such as UCI HAR, in which the number of volunteers was 30 volunteers, selected 30 volunteers out of 50 volunteers. The number of images that were calculated for 30 volunteers with the accelerometer or gyroscope is 4500. The number of images for the accelerometer and gyroscope when combined is 9000. Using the data collected from the accelerometer, the accuracy of up to 99.6% is obtained as shown in the confounding matrix fig (4.11). The evaluation metrics were calculated to measure the performances of the proposed algorithms are as shown in table (4.8)



(a) Confusion matrix for training



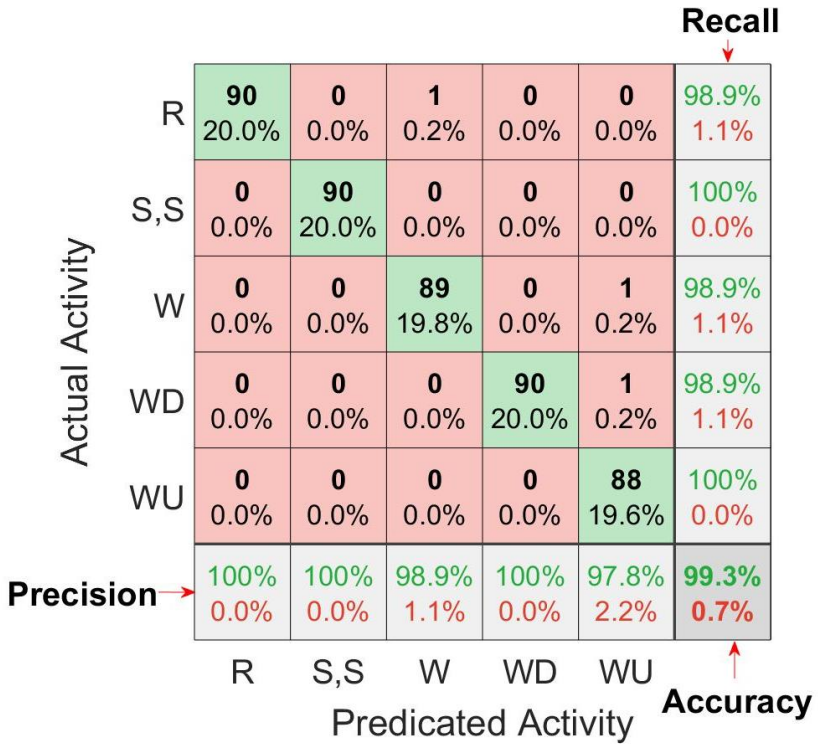
(b) Confusion matrix for testing

Figure (4.11): Confusion matrix for the accelerometer (30 volunteers)

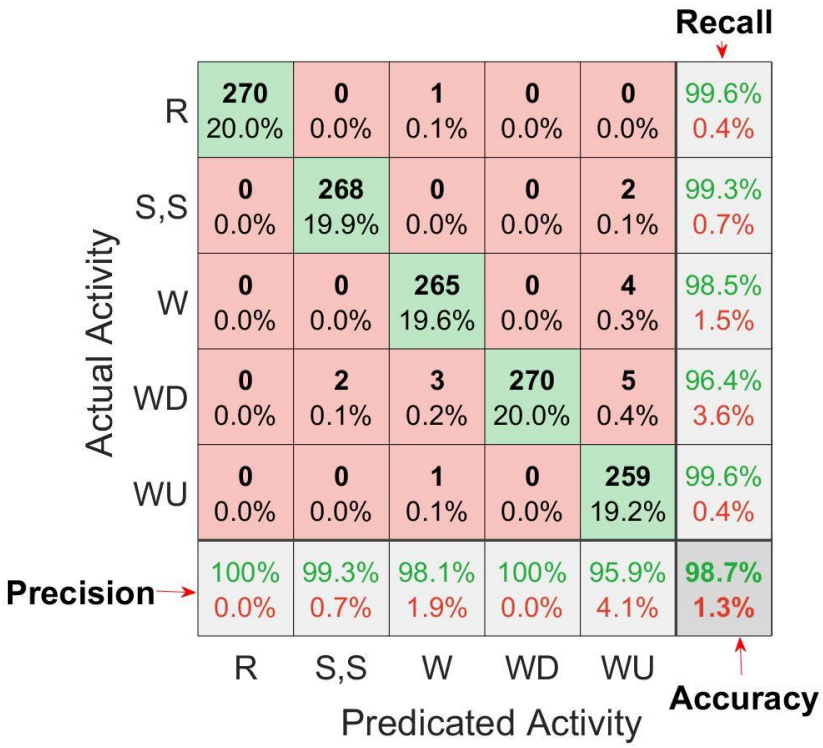
Table (4.8): Evaluation metrics of proposed algorithm (for accelerometer 30 volunteers)

Measure	R	S,S	W	WD	WU
Sensitivity, Recall	98.9%	100%	100%	100%	98.9%
Precision	100%	98.9%	100%	98.9%	100%
F-score	99.4%	99.4%	100%	99.4%	99.4%
Accuracy	99.6%				

With the same steps, the accuracy of the gyroscope sensor was calculated accurately up to 98.7%, as shown in figure (4.12). The rating scales were calculated to measure the performance of proposed algorithms as shown in table (4.9).



(a) Confusion matrix for training



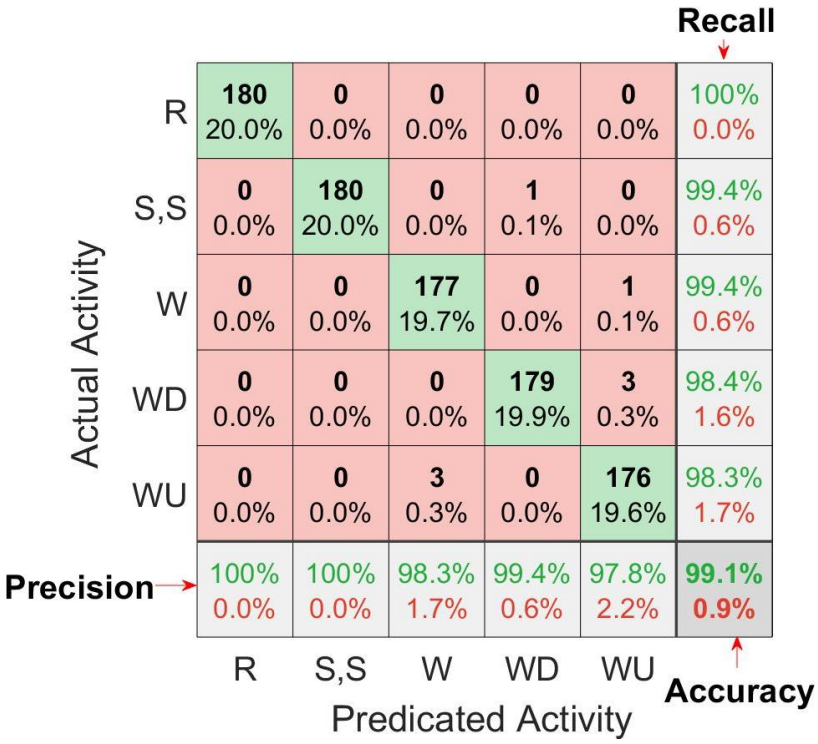
(b) Confusion matrix for testing

Figure (4.12): Confusion matrix for gyroscope (30 volunteers)

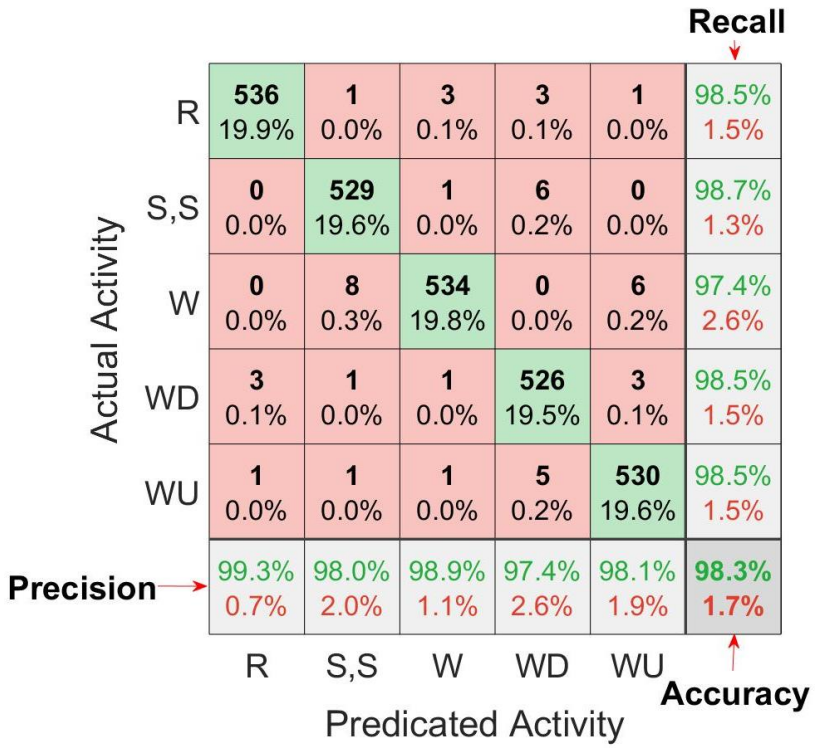
Table (4.9): Evaluation metrics of the proposed algorithm (for gyroscope 30 volunteers)

Measure	R	S, S	W	WD	WU
Sensitivity, Recall	99.6%	99.3%	98.5%	96.4%	99.6%
Precision	100%	99.3%	98.1%	100%	95.9%
F-score	99.8%	99.3%	98.3%	98.2%	97.7%
Accuracy	98.7%				

98.3% The accuracy was obtained when accelerometer and gyroscope sensors were combined, as shown by the confusion matrix fig(4.13). Rating scales were calculated to measure the performance of the proposed algorithms as shown in table (4.10).



(a) Confusion matrix for training



(b) Confusion matrix for testing

Figure (4.13): Confusion matrix for accelerometer and gyroscope (30 volunteers)

Table (4.10): Evaluation metrics of proposed algorithm (for both 30 volunteers)

Measure	R	S,S	W	WD	WU
Sensitivity, Recall	98.5%	98.7%	97.4%	98.5%	98.5%
Precision	99.3%	98%	98.9%	97.4%	98.1%
F-score	98.9%	98.3%	98.1%	97.9%	98.3%
Accuracy	98.3%				

4.7 Discussion

This chapter has aimed to demonstrate the fundamental theoretical concepts and applicability of CNNs for deep neural network construction. The study of the theoretical and practical applications of CNNs has shown that the main advantage of using CNNs concerning being fully connected is the reduced number of parameters that must be learned. When the parameter's number is reduced, during the training process, there is less noise. The reason for this is that the number of parameters is determined by the kernel width. The number of parameters in the model increases as the kernel width increases. In contrast, CNN normally needs thousands of labeled data. When the weight decay parameters are reduced, the model parameters grow larger. To avoid an increase in the number of iterations required to converge, the dropout rate should be reduced. Should be tuned learning rate optimally, low or high rates will optimization problems and a decrease in the effective capability of the network. The model's representation capacity grows as the number of hidden units Increases. When zero padding is used before convolution, the representation size is kept large. According to the experimental results; a 3-axis accelerometer consistently gets better performance than using a 3-axis gyroscope. Testing cases for 20 epochs, the results are excellent for a simple model

with limited CPU training. While some images are difficult to identify, our model can properly classify them. The accuracy of the test indicates that the model is well trained to predict. As the number of data increases, the size of the exercise affects the increased accuracy. The more data in the training set, the smaller the effect of training errors and test errors. When evaluating an activity performance recognition system, can achieve high levels of accuracy. Therefore, if identified correctly, the activity can be classified as topical true, true negative (TP), false positive (FP), or false negative (FN) when incorrectly classified. also noted that our results compared to other research are the best since the experiment has been applied in real-world conditions, and efforts should be made to make the system acceptable to many users, because human identification activity is very close to the daily life of human beings, and therefore the system must be useful to enable it to respond immediately for mobile devices.

4.8 Comparisons with previous work

In this section, to evaluate the proposed model, compare our approach with related work in HAR using a smartphone that implemented the data set with various methodologies. The CNN methodology proposed in this thesis gives the highest accuracy for HAR of all previous experiments. Comparing results registered for the proposed model with the related works as shown in the table (4.11). As for the relevant work, the data was collected using the smartphone using different sensors, whether it was an accelerometer, gyroscope, or other sensors, as well as for the data, either their data set or ready-made data was used with different types of deep learning approaches (CNN, LSTM or hybrid methods). While the data of our study were collected from two types of sensors built into the smartphone; an Accelerometer and a gyroscope

Table (4.11): Comparison of our method with some state-of-art studies

Reference	Dataset	Subjects	Activities	Methods	Overall Accuracy %
[15] (2015)	UCI HAR	30	6	CNN	90%
[16] (2016)	Their own data set	12	6	CNN	97.01%
[17] (2016)	-	30	6	CNN	94.79%
[18] (2016)	WISDM Actitracker	29	6	CNN	98.23%
[19] (2017)	Their own data set	5	3	CNN	92.71%
[23] (2019)	Their own data set	10	9	CNN	98%
[24] (2020)	UCI HAR	30	6	CNN	92.71%
[25] (2020)	WISDM dataset	36	6	CNN /LSTM	99.593% / 84.71%
[26] (2020)	UCI HAR	30	6	CNN	91.98%
[27] (2021)	Their own data set	44	10	CNN	90%
[28] (2021)	UCI HAR	30	6	CNN-LSTM	99.39%
Our experiment dataset		**50**	**5**	**CNN**	**99.1%**
		30		**CNN**	**99.6%**

Chapter Five

Conclusions and Future works

5.1 Conclusions

This research presented a human activity recognition system and emphasized the design issues such as selection of sensors, challenges, etc. Through the results that have been obtained in this work and from the proposed practical system, the conclusions can be summarized as follows:

- The primary project's goal is to create a system capable of obtaining information on human activities and use a case recognition model.
- A new method has been proposed for the classification of human activity recognition based on the transformation of time series into images using continuous wavelet transformation. The internal structures of the input time series can be discovered and extracted automatically by CNN to generate deep classification features. Instead of using features created by humans.
- In terms of accuracy of classification and tolerance of noise, the results of the experiments show that the CNN method performs better than competing baseline methods.
- The system was trained and tested via a large dataset created with the help of 50 volunteers through five activities. The process uses the smartphone's accelerometer and gyroscope sensor available in all smartphones, making the frame ubiquitous.

5.2 Future works

The study described in this thesis can be expanded in a variety of significant ways:

- For HAR systems to reach their full potential, more human activities are required. Activities recognized in existing systems have been simple and atomic, which could be a part of more complex composite behaviors. Recognition of composite activities can enrich context-awareness.
- Expanding on the work carried out on deep learning algorithms, one dimensional and two-dimensional convolutional neural networks, hybrids of convolutional networks and LSTMs should be further studied to determine their suitability to solve the problem of human activity recognition from raw signal data.
- Existing HAR systems are mainly focused on individual activities but could be extended further towards recognizing patterns and activity trends for a group of people with the use of social networks. Finally, recognition systems that could predict actions before they take place by the user could be a revolutionary development in certain applications.

References

[1] F. Wang, W. Gong, J. Liu, K. J. I. T. o. N. S. Wu, and Engineering, "Channel selective activity recognition with WiFi: A deep learning approach exploring wideband information," vol. 7, no. 1, pp. 181-192, 2018.

[2] Z. Gao, D. Liu, K. Huang, and Y. J. R. S. Huang, "Context-aware human activity and smartphone position-mining with motion sensors," vol. 11, no. 21, p. 2531, 2019.

[3] Y.-L. Hsu, H.-C. Chang, and Y.-J. J. I. A. Chiu, "Wearable sport activity classification based on deep convolutional neural network," vol. 7, pp. 170199-170212, 2019.

[4] H. F. Nweke, Y. W. Teh, M. A. Al-Garadi, and U. R. J. E. S. w. A. Alo, "Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: State of the art and research challenges," vol. 105, pp. 233-261, 2018.

[5] F. Demrozi, R. Bacchin, S. Tamburin, M. Cristani, G. J. I. j. o. b. Pravadelli, and h. informatics, "Toward a Wearable System for Predicting Freezing of Gait in People Affected by Parkinson's Disease," vol. 24, no. 9, pp. 2444-2451, 2019.

[6] X. Chen, H. Xue, M. Kim, C. Wang, and H. Y. Youn, "Detection of falls with smartphone using machine learning technique," in *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2019, pp. 611-616: IEEE.

[7] M. Li, M. O'Grady, X. Gu, M. A. Alawlaqi, and G. J. I. t. o. e. t. i. c. O'Hare, "Time-bounded activity recognition for ambient assisted living," 2018.

[8] J. Wang, Y. Chen, S. Hao, X. Peng, and L. J. P. R. L. Hu, "Deep learning for sensor-based activity recognition: A survey," vol. 119, pp. 3-11, 2019.

[9] Y. Liu, L. Nie, L. Liu, and D. S. J. N. Rosenblum, "From action to activity: sensor-based activity recognition," vol. 181, pp. 108-115, 2016.

[10] E. Kanjo, E. M. Younis, and C. S. J. I. F. Ang, "Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection," vol. 49, pp. 46-56, 2019.

[11] L. Liu, Y. Peng, M. Liu, and Z. J. K.-B. S. Huang, "Sensor-based human activity recognition system with a multilayered model using time series shapelets," vol. 90, pp. 138-152, 2015.

[12] N. Neverova *et al.*, "Learning human identity from motion patterns," vol. 4, pp. 1810-1820, 2016.

[13] L. Onofri, P. Soda, M. Pechenizkiy, and G. J. E. S. w. A. Iannello, "A survey on using domain and contextual knowledge for human activity recognition in video streams," vol. 63, pp. 97-111, 2016.

[14] A. Bayat, M. Pomplun, and D. A. J. P. C. S. Tran, "A study on human activity recognition using accelerometer data from smartphones," vol. 34, pp. 450-457, 2014.

[15] C. A. Ronaoo and S.-B. J. 한. 학. Cho, "Evaluation of deep convolutional neural network architectures for human activity recognition with smartphone sensors," pp. 858-860, 2015.

[16] T. Zebin, P. J. Scully, and K. B. Ozanyan, "Human activity recognition with inertial sensors using a deep learning approach," in *2016 IEEE SENSORS*, 2016, pp. 1-3: IEEE.

[17] C. A. Ronao and S.-B. J. E. s. w. a. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," vol. 59, pp. 235-244, 2016.

[18] M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin, and H.-P. Tan, "Deep activity recognition models with triaxial accelerometers," in *Workshops at the Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

[19] S.-M. Lee, S. M. Yoon, and H. Cho, "Human activity recognition from accelerometer data using Convolutional Neural Network," in *2017 ieee international conference on big data and smart computing (bigcomp)*, 2017, pp. 131-134: IEEE.

[20] C. Nwankpa, W. Ijomah, A. Gachagan, and S. J. a. p. a. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," 2018.

[21] A. J. A. S. C. Ignatov, "Real-time human activity recognition from accelerometer data using Convolutional Neural Networks," vol. 62, pp. 915-922, 2018.

[22] K. N. Ku Abd. Rahim, I. Elamvazuthi, L. I. Izhar, and G. J. S. Capi, "Classification of human daily activities using ensemble methods based on smartphone inertial sensors," vol. 18, no. 12, p. 4132, 2018.

[23] B. Zhou, J. Yang, and Q. J. S. Li, "Smartphone-based activity recognition for indoor localization using a convolutional neural network," vol. 19, no. 3, p. 621, 2019.

[24] S. Wan, L. Qi, X. Xu, C. Tong, Z. J. M. N. Gu, and Applications, "Deep learning models for real-time human activity recognition with smartphones," vol. 25, no. 2, pp. 743-755, 2020.

[25] C. Shiranthika, N. Premakumara, H.-L. Chiu, H. Samani, C. Shyalika, and C.-Y. Yang, "Human Activity Recognition Using CNN & LSTM," in *2020 5th International Conference on Information Technology Research (ICITR)*, 2020, pp. 1-6: IEEE.

[26] F. Cruciani *et al.*, "Feature learning for human activity recognition using convolutional neural networks," vol. 2, no. 1, pp. 18-32, 2020.

[27] M. Abid *et al.*, "Physical Activity Recognition Based on a Parallel Approach for an Ensemble of Machine Learning and Deep Learning Classifiers," vol. 21, no. 14, p. 4713, 2021.

[28] S. Mekruksavanich and A. J. S. Jitpattanakul, "Lstm networks using smartphone data for sensor-based human activity recognition in smart homes," vol. 21, no. 5, p. 1636, 2021.

[29] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. J. S. Amirat, "Physical human activity recognition using wearable sensors," vol. 15, no. 12, pp. 31314-31338, 2015.

[30] A. Wang, G. Chen, J. Yang, S. Zhao, and C.-Y. J. I. S. J. Chang, "A comparative study on human activity recognition using inertial sensors in a smartphone," vol. 16, no. 11, pp. 4566-4578, 2016.

[31] X. Li, Y. Zhang, M. Li, I. Marsic, J. Yang, and R. S. Burd, "Deep neural network for RFID-based activity recognition," in *Proceedings of the Eighth Wireless of the Students, by the Students, and for the Students Workshop*, 2016, pp. 24-26.

[32] S. Ishimaru, K. Hoshika, K. Kunze, K. Kise, and A. Dengel, "Towards reading trackers in the wild: detecting reading activities by EOG glasses and deep neural networks," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, 2017, pp. 704-711.

[33] R. Grzeszick, J. M. Lenk, F. M. Rueda, G. A. Fink, S. Feldhorst, and M. ten Hompel, "Deep neural network based human activity recognition for the order picking process,"

in *Proceedings of the 4th international Workshop on Sensor-based Activity Recognition and Interaction*, 2017, pp. 1-6.

[34] A. J. Perez, M. A. Labrador, and S. J. J. I. N. Barbeau, "G-sense: a scalable architecture for global sensing and monitoring," vol. 24, no. 4, pp. 57-64, 2010.

[35] Y. E. Ustev, O. Durmaz Incel, and C. Ersoy, "User, device and orientation independent human activity recognition on mobile phones: Challenges and a proposal," in *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, 2013, pp. 1427-1436.

[36] J. L. R. Ortiz, "Smartphone-based human activity recognition," 2015.

[37] C.-C. Yang and Y.-L. J. S. Hsu, "A review of accelerometry-based wearable motion detectors for physical activity monitoring," vol. 10, no. 8, pp. 7772-7788, 2010.

[38] H. A. J. A. o. C. Aziz and B. Research, "Comparison between field research and controlled laboratory research," vol. 1, no. 2, pp. 101-104, 2017.

[39] X. Sun and A. J. A. i. H.-C. I. May, "A comparison of field-based and lab-based experiments to evaluate user experience of personalised mobile devices," vol. 2013, 2013.

[40] L. Zeuwts *et al.*, "Is gaze behaviour in a laboratory context similar to that in real-life? A study in bicyclists," vol. 43, pp. 131-140, 2016.

[41] H. Jaya, S. Haryoko, L. Taris, and P. Ida, "Use of Remote Lab for Online and Real time Practicum At Vocational School in Indonesia," 2020.

[42] Q. Zhang, L. T. Yang, Z. Chen, and P. J. I. F. Li, "A survey on deep learning for big data," vol. 42, pp. 146-157, 2018.

[43] E. Junqué de Fortuny, D. Martens, and F. J. B. D. Provost, "Predictive modeling with big data: is bigger really better?," vol. 1, no. 4, pp. 215-226, 2013.

[44] M. Abid *et al.*, "Physical Activity Recognition Based on a Parallel Approach for an Ensemble of Machine Learning and Deep Learning Classifiers," 2021.

[45] G. Koppe, A. Meyer-Lindenberg, and D. J. N. Durstewitz, "Deep learning for small and big data in psychiatry," vol. 46, no. 1, pp. 176-190, 2021.

[46] Y. LeCun, Y. Bengio, and G. J. G. S. G. S. C. R. C. R. Hinton, "Deep learning. nature 521 (7553), 436-444," 2015.

[47] A. Shrestha and A. J. I. A. Mahmood, "Review of deep learning algorithms and architectures," vol. 7, pp. 53040-53065, 2019.

[48] P. Vepakomma, D. De, S. K. Das, and S. Bhansali, "A-Wristocracy: Deep learning on wrist-worn sensing for recognition of user complex activities," in *2015 IEEE 12th International conference on wearable and implantable body sensor networks (BSN)*, 2015, pp. 1-6: IEEE.

[49] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. J. F. G. C. S. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," vol. 81, pp. 307-313, 2018.

[50] J. Suto, S. Oniga, P. P. J. I. J. o. C. C. Sitar, and Control, "Feature analysis to human activity recognition," vol. 12, no. 1, pp. 116-130, 2016.

[51] J. Morales, D. J. B. Akopian, and B. Engineering, "Physical activity recognition by smartphones, a survey," vol. 37, no. 3, pp. 388-400, 2017.

[52] N. Hammerla, J. Fisher, P. Andras, L. Rochester, R. Walker, and T. Plötz, "PD disease state assessment in naturalistic environments using deep learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2015, vol. 29, no. 1.

[53] S. Sani, S. Massie, N. Wiratunga, and K. Cooper, "Learning deep and shallow features for human activity recognition," in *International conference on knowledge science, engineering and management*, 2017, pp. 469-482: Springer.

[54] S. S. Haykin, "Neural networks and learning machines/Simon Haykin," ed: New York: Prentice Hall, 2009.

[55] J. J. N. n. Schmidhuber, "Deep learning in neural networks: An overview," vol. 61, pp. 85-117, 2015.

[56] Y. Bengio, P. Simard, and P. J. I. t. o. n. n. Frasconi, "Learning long-term dependencies with gradient descent is difficult," vol. 5, no. 2, pp. 157-166, 1994.

[57] X. Wu, D. Sahoo, and S. C. J. N. Hoi, "Recent advances in deep learning for object detection," vol. 396, pp. 39-64, 2020.

[58] G. Yolcu, I. Oztel, S. Kazan, C. Oz, F. J. J. o. a. i. Bunyak, and h. computing, "Deep learning-based face analysis system for monitoring customer interest," vol. 11, no. 1, pp. 237-248, 2020.

[59] S. Kuutti, R. Bowden, Y. Jin, P. Barber, and S. J. I. T. o. I. T. S. Fallah, "A survey of deep learning applications to autonomous vehicle control," vol. 22, no. 2, pp. 712-733, 2020.

[60] J. Martens and I. Sutskever, "Training deep and recurrent networks with hessian-free optimization," in *Neural networks: Tricks of the trade*: Springer, 2012, pp. 479-535.

[61] Y. Guan, T. J. P. o. t. A. o. I. Plötz, Mobile, Wearable, and U. Technologies, "Ensembles of deep lstm learners for activity recognition using wearables," vol. 1, no. 2, pp. 1-28, 2017.

[62] S. W. Pienaar and R. Malekian, "Human activity recognition using LSTM-RNN deep neural network architecture," in *2019 IEEE 2nd wireless africa conference (WAC)*, 2019, pp. 1-5: IEEE.

[63] C. Jobanputra, J. Bavishi, and N. J. P. C. S. Doshi, "Human activity recognition: A survey," vol. 155, pp. 698-703, 2019.

[64] A. Ferrari, D. Micucci, M. Mobilio, and P. J. J. o. R. I. E. Napoletano, "Trends in human activity recognition using smartphones," pp. 1-25, 2021.

[65] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014.

[66] A. Mahmoud and A. Mohammed, "A survey on deep learning for time-series forecasting," in *Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges*: Springer, 2021, pp. 365-392.

[67] J. C. B. J. a. p. a. Gamboa, "Deep learning for time-series analysis," 2017.

[68] Z. Chen and D. J. a. p. a. Yi, "The game imitation: Deep supervised convolutional networks for quick video game AI," 2017.

[69] S. Øyen, "Forecasting multivariate time series data using neural networks," NTNU, 2018.

[70] M. A. Hossain, M. S. A. J. G. J. o. C. S. Sajib, and Technology, "Classification of image using convolutional neural network (CNN)," 2019.

[71] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. J. P. R. Leckie, "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning," vol. 58, pp. 121-134, 2016.

[72] G. E. Hinton, "A practical guide to training restricted Boltzmann machines," in *Neural networks: Tricks of the trade*: Springer, 2012, pp. 599-619.

[73] S. J. a. p. a. Ruder, "An overview of gradient descent optimization algorithms," 2016.

[74] B. Almaslukh, A. M. Artoli, and J. J. S. Al-Muhtadi, "A robust deep learning approach for position-independent smartphone-based human activity recognition," vol. 18, no. 11, p. 3726, 2018.

[75] W. Palma, *Time series analysis*. John Wiley & Sons, 2016.

[76] V. Gikas and H. J. S. Perakis, "Rigorous performance evaluation of smartphone GNSS/IMU sensors for ITS applications," vol. 16, no. 8, p. 1240, 2016.

[77] V. Renaudin, M. Susi, and G. J. S. Lachapelle, "Step length estimation using handheld inertial sensors," vol. 12, no. 7, pp. 8507-8525, 2012.

[78] R. Yang and B. J. I. Wang, "PACP: A position-independent activity recognition method using smartphone sensors," vol. 7, no. 4, p. 72, 2016.

[79] A. Mannini and A. M. J. S. Sabatini, "Machine learning methods for classifying human physical activity from on-body accelerometers," vol. 10, no. 2, pp. 1154-1175, 2010.

[80] O. J. Woodman, "An introduction to inertial navigation," University of Cambridge, Computer Laboratory2007.

[81] X. Su, H. Tong, P. J. T. s. Ji, and technology, "Activity recognition with smartphone sensors," vol. 19, no. 3, pp. 235-249, 2014.

[82] Y. Yao, L. Song, and J. J. S. Ye, "Motion-To-BMI: Using Motion Sensors to Predict the Body Mass Index of Smartphone Users," vol. 20, no. 4, p. 1134, 2020.

[83] C. J. Billington, L. H. Epstein, N. J. Goodwin, and R. L. J. A. f. p. Leibel, "Medical care for obese patients: advice for health care professionals," vol. 65, no. 1, p. 81, 2002.

[84] E. Kocabey *et al.*, "Face-to-BMI: Using computer vision to infer body mass index on social media," in *Proceedings of the International AAAI Conference on Web and Social Media*, 2017, vol. 11, no. 1.

[85] E. Kocabey, F. Ofli, J. Marin, A. Torralba, and I. Weber, "Using computer vision to study the effects of bmi on online popularity and weight-based homophily," in *International Conference on Social Informatics*, 2018, pp. 129-138: Springer.

[86] L. F. Polania, G. M. Fung, and D. Wang, "Ordinal regression using noisy pairwise comparisons for body mass index range estimation," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019, pp. 782-790: IEEE.

[87] C. Berkai, M. Hariharan, S. Yaacob, and M. I. Omar, "Estimation of BMI status via speech signals using short-term cepstral features," in *2015 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2015, pp. 195-199: IEEE.

[88] Q. Riaz, A. Vögele, B. Krüger, and A. J. S. Weber, "One small step for a man: Estimation of gender, age and height from recordings of one step by a single inertial sensor," vol. 15, no. 12, pp. 31999-32019, 2015.

[89] T. Furuya, H. Kondo, and F. N. Kondo, "Convolutional Neural Networks on Time Series for Smartphone Application Activations Using Wavelet Transform," in *2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI)*, 2019, pp. 521-527: IEEE.

[90] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi, *Wavelets and their Applications*. John Wiley & Sons, 2013.

[91] R. Polikar, "The wavelet tutorial: the engineer's ultimate guide to wavelet analysis," 2002.

[92] F. Demrozi, G. Pravadelli, A. Bihorac, and P. J. I. A. Rashidi, "Human activity recognition using inertial, physiological and environmental sensors: a comprehensive survey," 2020.

[93] (2017,June 13). *Sensor Kinetics Pro.* Available: https://play.google.com/store/apps/details?id=com.innoventions.sensorkineticspro&hl=en

[94] J. Saha, C. Chowdhury, I. Roy Chowdhury, S. Biswas, and N. J. I. Aslam, "An ensemble of condition based classifiers for device independent detailed human activity recognition using smartphones," vol. 9, no. 4, p. 94, 2018.

[95] J. R. Kwapisz, G. M. Weiss, and S. A. J. A. S. E. N. Moore, "Activity recognition using cell phone accelerometers," vol. 12, no. 2, pp. 74-82, 2011.

[96] M. Shoaib, H. Scholten, and P. J. Havinga, "Towards physical activity recognition using smartphone sensors," in *2013 IEEE 10th international conference on ubiquitous*

intelligence and computing and 2013 IEEE 10th international conference on autonomic and trusted computing, 2013, pp. 80-87: IEEE.