UNIVERSITY OF BAHRAIN

College of Science



Department of Postgraduate Programs

Comparing Long Short-Term Memory and Graph Convolutional Network Models for Human Activity Recognition using WISDM Dataset

Thesis submitted in partial fulfilment of the requirements for the degree of M.Sc. In Big Data Science and Analytics

Prepared by:

Majeed Mohamed Hasan Ebrahim

202101787

Supervised by

Dr. Ahmed M. Zeki

Assistant Professor

University of Bahrain

Kingdom of Bahrain

April 2023

University of Bahrain

College of Science

Department of Postgraduate Programs

Comparing Long Short-Term Memory and Graph Convolutional Network Models for Human Activity Recognition using WISDM Dataset

Thesis submitted in partial fulfilment of the requirements for the degree of M.Sc. In Big Data Science and Analytics

Prepared by:

Majeed Mohamed Hasan Ebrahim

202101787

Supervised by:

Dr. Ahmed M. Zeki

Assistant Professor

University of Bahrain

Kingdom of Bahrain

April 2023

University of Bahrain
Deanship of Graduate Studies
And scientific Research

جامعة البحرين
عمادة الدراسات العليا والبحث العلمي

Thesis Discussion Committee Decision

Student Name: **MAJEED MOHAMED HASAN EBRAHIM**	Student ID: **202101787**
Department: Choose an item.	College: **Science**
Program: **M.Sc. in Big Data Science and Analytics**	
Thesis Title: Comparing Long Short-Term Memory (LSTM) and Graph Convolutional Network (GCN) Models for Human Activity Recognition using WISDM Dataset	
Supervisor Name: **Dr. Ahmed M. Zeki**	Student GPA: **3.78**

The Discussion Committee met on 29-May-23 In S41-1002. Upon discussing the student whose details are stated above in public, the discussion committee recommends the following:

The student is awarded a grade Choose an item. in the thesis subject.

☐ Without making any changes to the thesis.

☒ After making the modifications attached with this decision.

Discussion Committee Members Approval:

Name		Signature
1. Dr. Ahmed M. Zeki	(Advisor)	.
2. Dr. Riadh Ksantini	(Internal Examiner)	.
3. Dr. Naimul Khan	(External Examiner)	.

Approved at the University Council 6th meeting, 2013
In accordance with Decision No. 736/2013 dated 14/3/2013
Modified by University Council Decision No. 2043/2013 dated 20/11/2013.

Abstract

The development of technology in computer hardware has made observing and analyzing daily performed activities an easy task. Therefore, human activity recognition systems have benefited from the embedded sensors in smart devices in several fields, such as healthcare, security, and fitness. Furthermore, these sensors can generate sequenced data with temporal and spatial relationships. Therefore, Artificial Intelligence techniques, such as machine learning and deep learning, have been utilized for data processing and activity classification. While traditional machine learning methods have performed well, they require manual feature extraction from the raw data.

Meanwhile, deep learning methods can extract the required features respecting spatial and temporal dependencies. Among those methods, The Long Short-Term Memory (LSTM) method is proven efficient in dealing with time-series data. In addition, the Graph Convolutional Network (GCN) has the ability to extract spatiotemporal relations using the graph data structure. However, as per author knowledge, the previous studies have yet to compare the performance of both methods, so this research aims to compare both methods using the WISDM dataset. The dataset contains data from accelerometers and gyroscope sensors embedded in smartphones and smartwatches for 18 performed activities.

For each smart device, the data of two sensors have been combined to form two sub-datasets which are the preprocessed and prepared datasets for feeding models using the overlapped sliding windows. Hence, both methods have been applied to the resulting sub-datasets. As a result, for the smartphone sub-dataset, the GCN model outperformed the LSTM model by around 4%, scoring 94.4% for accuracy and F1-score. In addition, the GCN model performed better than some previous LSTM models considering a subset of activities. On the other hand, the smartwatch sub-dataset gave a slight advantage to the GCN model over the LSTM model by nearly 1%, as it scored 90.4% for accuracy and F1-score. However, the GCN method can be considered a rival to the LSTM method in such data type, which should be applied to other datasets in future works for generalization.

Dedication

This work is dedicated to the Masters of Kindness after God Almighty, who were of help to me in times of need. To all my little family, I appreciate their support and backing me in this important period of life with all its moments, especially the difficult ones. Thank you very much.

Acknowledgments

My sincere thanks go to my thesis advisor, Dr. Ahmed M. Zaki, for his support and guidance throughout the research project. I was able to shape my thoughts and develop my thesis with his valuable information, feedback, and encouragement.

Also, my grateful thanks go to the program coordinator, Dr. Sawsan Hilal, for her endless efforts and support since joining the program till the final stage.

Finally, I would like to extend my thanks to Program instructors, Professor Dr. Nabil Al Hewahi, Dr. Abdulla Al Qaddoumi, and Dr. Hesham Al-Ammal, for providing me with valuable tuition that added a lot to my knowledge.

Table of Contents

Defense Committee's Approval ..... A

Abstract ..... B

Dedication C

Acknowledgments ..... D

Table of Contents ..... E

List of Tables ..... H

List of Figures I

List of Abbreviations ..... K

Chapter One Introduction to the Study 2

1.0. Introduction 2

1.1. Research Problem 4

1.2. Research Questions 5

1.3. Research Objectives 5

1.4. Research Significance 6

1.5. Research Structure 6

Chapter Two Literature Review 8

2.0. Introduction 8

2.1. Human Activity Recognition 9

2.1.1. Definition 9

2.1.2. Importance of Human Activity Recognition 10

2.1.3. Application of Human Activity Recognition 11

2.1.4. Types of Human Activity Recognition Systems 13

2.1.5. Components of HAR Systems 14
2.2. Artificial Intelligence in HAR 16
2.2.1. The Impact of Artificial Intelligence 16
2.3. HAR using Machine Learning Algorithms 17
2.3.1. Application On Public Wearable Devices Datasets 18
2.3.2. Traditional ML On WISDM Dataset 19
2.4. HAR using Deep Learning Methods 21
2.4.1. Most Common DL Methods 21
2.4.2. HAR using LSTM Method 23
2.4.3. HAR using GCN Method 27
2.5. Conclusion 30
Chapter Three Methods and Procedures 32
3.0. Overview 32
3.1. Research Methodology 32
3.2. Research Dataset 34
3.2.1. Dataset overview 34
3.2.2. Dataset Preprocessing 36
3.3. Implemented Models 38
3.3.1. Long Short-Term Memory (LSTM) 38
3.3.2. Graph Convolutional Network (GCN) 40
3.4. Evaluation Metrics 44
Chapter Four Results and Discussion 47
4.0. Overview 47
4.1. Exploratory Data Analysis Results 47

4.1.1. Smartphone Sub-dataset EDA Results 47
4.1.2. Smartwatch Sub-dataset EDA Results 51
4.2. Evaluation of Smartphone Sub-dataset Models 54
4.2.1. Performance of LSTM Model 54
4.2.2. Performance of GCN Model.....59
4.2.3. Analysis and Comparison.....63
4.3. Evaluation of Smartwatch Sub-dataset Models.....67
4.3.1. Performance of LSTM Model 67
4.3.2. Performance of GCN Model.....72
4.3.3. Analysis and Comparison.....76
Chapter Five Conclusion and Recommendation 81
5.0. Overview.....81
5.1. Research Summary 81
5.2. Research Limitations 84
5.3. Recommendations and Future Work 84
References.....85
Appendices.....99
Appendix (A) 99
Appendix (B) 100
Appendix (C) 101
Appendix (D) 102
الملخص 103

List of Tables

Table (2.1): Summary of recent studies applied to the WISDM dataset using ML algorithms 20
Table (2.2): Summary of recent studies on the WISDM dataset using LSTM method 26
Table (3.1): List of Performed Activities (Classes) 35
Table (3.2): Description of Features for Raw Measurements 36
Table (3.3): LSTM Model Settings for Smartphone Sub-dataset 40
Table (3.4): LSTM Model Settings for Smartphone Sub-dataset 40
Table (3.5): GCN Model Settings for Smartphone Sub-dataset 43
Table (3.6): GCN Model Settings for Smartwatch Sub-dataset 44
Table (4.1): Performance of LSTM Model Across Each Class in Smartphone Sub-dataset 58
Table (4.2): Performance of GCN Model Across Each Class in Smartphone Sub-dataset 63
Table (4.3): Group-wise Performance Comparison for LSTM and GCN Models 65
Table (4.4): Class-wise Performance Comparison for LSTM and GCN Models 66
Table (4.5): Performance of LSTM Model Across Each Class in Smartwatch Sub-dataset 71
Table (4.6): Performance of GCN Model Across Each Class in Smartwatch Sub-dataset 75
Table (4.7): Group-wise Performance Comparison for LSTM and GCN Models 77
Table (4.8): Class-wise Performance Comparison for LSTM and GCN Models 78

List of Figures

Figure 2.1: Types of Human Activities 8
Figure 2.2: Approaches of human activity 14
Figure 2.3: Stages of HAR System 15
Figure 2.4: HAR – GCN Model Description 28
Figure 3.1: The Illustration of Research Methodology 33
Figure 3.2: LSTM Cell Architecture 39
Figure 3.3: Simple graph convolutional network design 41
Figure 3.4: Simple Graph Data structure 42
Figure 4.1: Smartphone's Accelerometer Measurements for Some Activities 48
Figure 4.2: Smartphone's Gyroscope Measurements for Some Activities 48
Figure 4.3: Correlation of Sensor's Measurements for Kicking Activity 49
Figure 4.4: The Spread of Sensor's Values in Smartphone Sub-dataset ..50
Figure 4.5: Distribution of Activities in Smartphone Sub-dataset 50
Figure 4.6: Smartwatch's Accelerometer Measurements for Some Activities 51
Figure 4.7: Smartwatch's Gyroscope Measurements for Some Activities 52
Figure 4.8: Correlation of Sensor's Measurements for Clapping Activity 52
Figure 4.9: The Spread of Sensor's Values in Smartwatch Sub-dataset ..53
Figure 4.10: Distribution of Activities in Smartwatch Sub-dataset 53
Figure 4.11: Effect of Layer Size on LSTM Model of Smartphone Sub-dataset 55
Figure 4.12: Effect of Batch Size on LSTM Model of Smartphone Sub-dataset 55

Figure 4.13: Result of Training LSTM Model Using Smartphone Sub-dataset....56
Figure 4.14: LSTM Model Performance using Smartphone Sub-dataset 57
Figure 4.15: Effect of Layer Size on GCN Model of Smartphone Sub-dataset....60
Figure 4.15: Effect of Batch Size on GCN Model of Smartphone Sub-dataset....60
Figure 4.17: Result of Training GCN Model Using Smartphone Sub-dataset....61
Figure 4.18: GCN Model Performance using Smartphone Sub-dataset ..61
Figure 4.19: Overall Comparison Between LSTM and GCN Models 64
Figure 4.20: Effect of Layer Size on LSTM Model of Smartwatch Sub-dataset....67
Figure 4.21: Effect of Batch Size on LSTM Model of Smartwatch Sub-dataset....68
Figure 4.22: Result of Training LSTM Model Using Smartwatch Sub-dataset....69
Figure 4.23: LSTM Model Performance using Smartwatch Sub-dataset 69
Figure 4.24: Effect of Layer Size on GCN Model of Smartwatch Sub-dataset....72
Figure 4.25: Effect of Batch Size on GCN Model of Smartwatch Sub-dataset....73
Figure 4.26: Result of Training GCN Model Using Smartwatch Sub-dataset....73
Figure 4.23: GCN Model Performance using Smartwatch Sub-dataset...74
Figure 4.28: Overall Comparison Between LSTM and GCN Models 77

List of Abbreviations

Abbreviation	Full Text
ADLs	Activities of Daily Living
AI	Artificial Intelligence
AUC	Area Under Curve (ROC curve)
HAR	Human Activity Recognition
CNN	Convolutional Neural Network
CRISP-DM	Cross Industry Standard Process for Data Mining
DL	Deep Learning
GCN	Graph Convolutional Network
GNN	Graph Neural Network
GRU	Gated Recurrent Unit
HCI	Human-Computer Interaction
KNN	K-Nearest Neighbors
LSTM	Long Short-Term Memory
ML	Machine Learning
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SVM	Support Vector Machines
WISDM	Wireless Sensor Data Mining dataset

Chapter One

Introduction to the Study

1.0. Introduction

1.1. Research Problem

1.2. Research Questions

1.3. Research Objectives

1.4. Research Significance

1.5. Research Structure

Chapter One
Introduction to the Study

1.0. Introduction

Human activities are actions people take that affect their bodies or the surrounding environment. Several aspects of people's lives, such as medical services, life assistance, well-being, and social connections, benefit from a greater knowledge of human behaviors. Using a series of observations and successfully identifying these activities could provide specialized help to improving plenty of daily life activities. Human activity recognition has received much attention recently, supported by significant technological advancements in sensing, processing, and analysis (Bian et al., 2022).

The data of human activity recognition systems usually come from sensors as a sequence, such as accelerometers and gyroscopes, which represent the activity instances in segments. The classification methods dealing with the data sequence are divided into two types: traditional Machine Llearning (ML) and Deep Learning (DL) methods. Although feature extraction and engineering are done manually in the traditional methods, it is still considered a shallow extraction. It is opposed to DL, which automatically pulls deeper and more complex characteristics from the original data. In addition, it can handle and separate the similarities of classes well (Li et al., 2023).

Among DL methods that proved successful in solving time sequence problems is the Long Short-Term Memory (LSTM), which takes care of long-term relationships. Also, it allows multiple concurrent input raw data sequences, where it can learn directly from it without feature engineering (Challa et al., 2021; Moradi et al., 2022). Recently, the Graph Convolutional Network (GCN) was also introduced to handle the prediction of time-series data, as it uses relationships between node characteristics and classes in a particular network to improve prediction performance (Mohamed et al., 2022).

While the LSTM technique has been developed based Recurrent Neural Network (RNN) method, the Convolutional Neural Networks (CNN) method was the base for GCN; therefore, each one has different power in DL classification (Mohamed et al., 2022). Therefore, the need to explore the abilities of each technique in Human Activity Recogintion (HAR) classification is present, in addition to comparing their performance.

The remainder of this chapter will discuss the following subjects of this research:

1. Research Problem
2. Research Questions
3. Research Objectives
4. Research Significance
5. Research Structure

1.1. Research Problem

The time series data has challenged ML algorithms to preserve the data sequence during model building. This action leads to improving the training performance of the model and then the prediction process at the end. Hence, DL offers some methods to deal with this data type, such as Neural Networks (NNs). One of these networks is the LSTM which has been used for a while with notable success in solving this due to its processing design. The GCN has also entered this domain by using the relationships between data elements with the time to make predictions (Mohamed et al., 2022).

The human activity recognition systems offer time series data using sensors like accelerometers and gyroscopes. However, while the LSTM method was used for recognizing human activity with various techniques, the GCN also was utilized for the same objective. Hence There was a missing part of knowledge that compares the performance of those methods for such classification problems.

The aim of this research is to provide a comparison between LSTM and GCN methods to recognize human activities by using accelerometers and gyroscopes data of smart devices. Furthermore, it will measure the performance of these methods to offer information that could help decide the proper method for processing this data type.

1.2. Research Questions

The conduction of this research was based on some questions, as the results should provide answers to the following questions:

Main Question: What is the difference in performance between the LSTM and GCN regarding the recognition and classification of human activities? Is there a preference for one of them?

Minor Question (1): Is there a difference in the performance between the two methods depending on the source device of the data?

Minor Question (2): Can the GCN method be a competitor to the LTSM method in processing such type of time-series data, at least?

1.3. Research Objectives

The main objectives expected to reach by carrying out this research are the following:

1. Developing a model to recognize the type of human activity based on data from smartphone and smartwatch sensors using DLmethods: LSTM and GCN.
2. Comparing the performance of the developed models using suitable metrics for each device's sub-datasets.

1.4. Research Significance

Although many of previous research has dealt with the topic of human activities recognition, especially using DL techniques, this research has some significance as follows:

1. It provides a comparison between LSTM and GCN methods in terms of processing time-series data for human activity recognition.
2. It explores the performance of the GCN method in such type of data.
3. The used data combination differs from other studies, as it combines the data of the two sensors with considering the complete list of attached activities.

1.5. Research Structure

The remainder of this research after this chapter will be as follows:

- Chapter Two: Literature Review
- Chapter Three: Research Methodology
- Chapter Four: Research Results
- Chapter Five: Research Discussion & Conclusion

Chapter Two

Literature Review

2.0. Introduction

2.1. Human Activity Recognition

2.2. Artificial Intelligence in HAR

2.3. HAR using Machine Learning Algorithms

2.4. HAR using Deep Learning Methods

2.5 Conclusion

Chapter Two

Literature Review

2.0. Introduction

Human daily activities and movement become essential to their lifestyle. They use their bodies to move, exercise and perform other necessary activities at school, work, or other places, as shown in Figure 2.1. The World Health Organization (WHO) considers such activities as energy-required activities due to body movement at any skill level. In addition, it leads to improved physical and mental health and quality of life and well-being (World Health Organization, 2022). Scientists have been interested in observing and studying these activities and researching their various aspects for a long time. Therefore, the term The Activities of Daily Living (ADLs) appeared by Sidney Katz in 1950 to describe the basic skills required by the individual to perform these activities by himself in life sides (Edemekong et al., 2022).

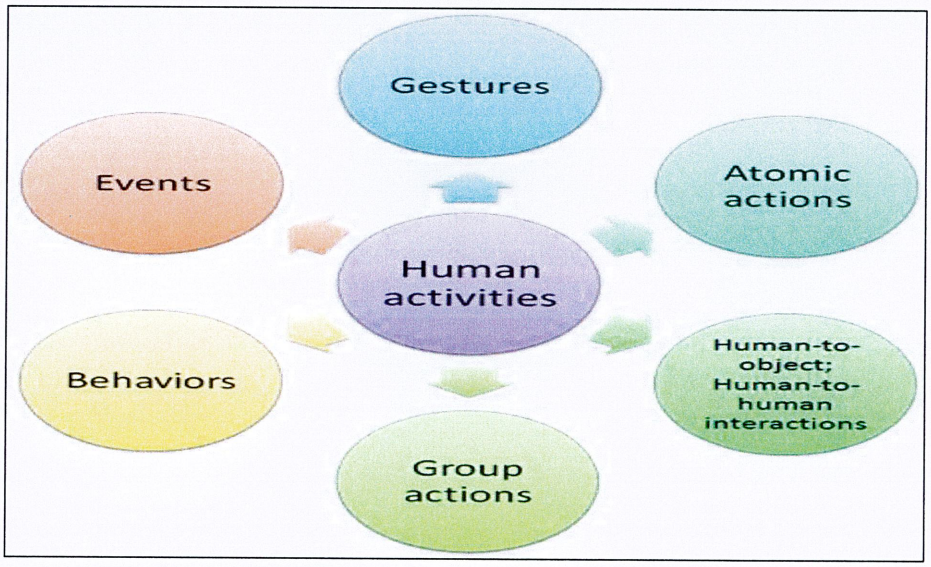


Figure 2.1: Types of Human Activities (Vrigkas et al., 2015)

Since then, this interest has been a task to observe, recognize and analyze the activities conducted by human power. However, it was not only a challenging task for the human operator, but it cost more. Due to that, the attempt to improve it and make it more efficient and effective was made by making it more automatic and systematic (Jaberi & Ravanmehr, 2022). Hence, the development of computer hardware and beyond technology, such as computer vision and image detection based on trailblazing algorithms, brought about the term Human Activity Recognition (Isakava, 2022). The historical emergence of this research field was not clear; it was dated back to the beginning of the 1990s as per this study (Dang et al., 2020), while it was traced back to late of the same decade as per Zheng et al. (2017).

2.1. Human Activity Recognition

2.1.1. Definition

There is no specific and agreed definition of this term, and the definitions vary among the researchers. However, as a general definition by Mondal et al. (2021), HAR can be described as the task of differentiating a physical activity performed by an individual capable of developing within a given domain.

According to Fra et al. (2022), it can be defined as the process of analyzing and categorizing signals related to human activities, which can be classified according to the equipment and sensors used to capture such signals data. This definition matches what Sarkar et al. (2021) came up with, as they focus on the raw sensor data classification task. However, while Yi and Hwang (2022) agree with this definition, they identify the data as a sequence. Although some researchers share the same idea, they

consider it a method of classifying a sequence of human activities by interpreting still photos, videos, and sensory data (Tina et al., 2021).

Using more specific expressions, other researchers stated that it is a classification task with a time series challenge that aims to identify people's activities using many observations from the human body and the surrounding environment (Shao et al., 2021). Also, Ramanujam et al. (2021) emphasize the time-series aspect, but they identify the data source from sensors built into smartphones and wearable devices. Some researchers have restricted the type of sensors to wearable ones same as the definition by Dhammi and Tewari (2021).

It can be concluded from the previous definitions that the primary purpose of this task is to classify human activities. However, the difference in the type of data is due to the difference in the purpose of each study.

2.1.2. Importance of Human Activity Recognition

Despite the novelty of this field, it has gained significant importance represented in the rapid development in the scientific research aspect. Which in turn produces a wide spread of applications and related products. The researchers mention several reasons, including:

- It has a wide range of applications because HAR's application has the potential enhancement by combining it with other technologies like wearable devices and the Internet of Things (IOT) (Ali et al., 2020; Suwannarat & Kurdthongmee, 2021).

- It has a vital role in extract an advanced information about human lifestyle due to the capability of observing daily occurred interactions. Which it gives the ability to learn more about person's identity, personality, and psychological state (Vrigkas et al., 2015; Athota & Sumathi, 2022).

- HAR can benefit from recent environmental advances, it can help in monitoring daily life by recognizing certain situations, such as abnormal and risky ones (Alexan et al., 2022).

- It is precious and practical for solving problems in various applications to bring more comfort to end-users and organizations to lessen employee workload (Bhambri et al., 2021).

As the aspects of significance for this system varied, the most significant importance remained focused on its scalability and practical embedding in various fields and applications.

2.1.3. Application of Human Activity Recognition

As a result of this importance, a solid ground has been provided to build on for many applications. Many previous studies have extensively reviewed the applications in various fields (Athota & Sumathi, 2022; Dhammi & Tewari, 2021; Host & Ivašić-Kos, 2022; Najeh et al., 2022; Tina et al., 2021; Zhang et al., 2022). Therefore, they can be mentioned as follows:

- **Security and Surveillance:** It can help the security staff discover suspicious behaviors and identity verification utilizing the gait characteristics. It is instrumental in monitoring the

region with relatively high accuracy without human supervision.

- **Health Care:** The medical services environment can be utilized for activity recognition. It can help better break down and understand the medical tests. In addition, it improves the reliability of diagnosing a patient's health, then giving the proper treatment and care. Also, it has been used for monitoring the daily living activities of elderlies, disabled persons, and children.
- **Sport and Fitness:** Applying a computer vision technique for monitoring player performance, such as detecting and tracking movement, identifying, and comparing performed activities, and doing automatic statistical analysis. It also contributes to detect sports injuries.
- **Human-Computer Interaction (HCI):** The communication abilities with electronics, computers, and robots have offered flexible and convenient methods for human interaction. Robots have been trained to follow and recognize the conducted actions by people. Then it is utilized for defense, rescue, or other similar purposes. While in the entertainment field, better interaction with hardware and software has been gained based on Augmented Reality (AR) and Virtual Reality (VR).
- **Smart Homes and Well-being:** This application is about caring for the daily activities of residents as part of their routine by using intelligent algorithms to track and recognize these activities. It also could provide information and guides about the significance of their practices.

2.1.4. Types of Human Activity Recognition Systems

The difference in the purpose of HAR is reflected in the design of data collection systems in addition to the processing algorithms and methods. Then it can be categorized according to that. Therefore, the number of methods varies according to the researchers' vision and field of research.

Several studies have agreed to divide the types of HAR systems into two main types; the first is vision-based, while the second method is the sensor-based method which includes all types of sensors (Gupta S., 2021; Ige & Noor, 2022; Islam et al., 2022; Vidya & Sasikumar, 2022).

In contrast, many other researchers divided the types into three categories as more details were added in the sensor-based branch appropriate to the field of study (Li et al., 2023; Wang M., 2022). For example, Athota and Sumathi (2022) categorized the strategies used in the mentioned systems into three main approaches based on data collection types, as shown in Figure. 2.2:

1- Video-based: It means computer vision-based systems, as the sensors are utilized for still and video imaging.
2- Sensor-based: It is further classified into three classes: wearable, object-tagged and dense sensing.
3- Time and Geography-based: Where human activities can be recognized using time and place information separately.

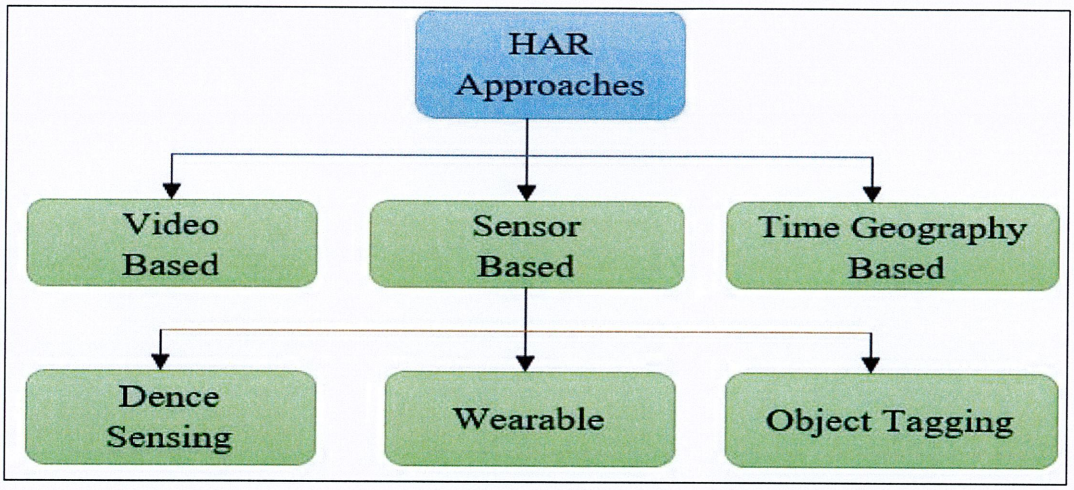


Figure 2.2: Approaches of human activity recognition (Athota & Sumathi, 2022)

Meanwhile Dang et al. (2020) classified the primary sensors into four classes as follows:

1. Wearable: like the Accelerometer, Gyroscope, Magnetometer, and Global Positioning System (GPS).
2. Object sensor: Including Radio-Frequency Identification (RFID) and WIFI.
3. Environmental: Sensors that measure the environmental elements such as atmospheric pressure, temperature, and sound level.
4. Hybrid: a mix of several types of sensors to increase activity recognition rate.

Finally, regardless the variation in system types, they share almost the same stages despite the purpose or usage of these systems.

2.1.5. Components of HAR Systems

Like other systems, the HAR system consists of several parts representing this process, each playing a key role in the system's success. Considering the previous studies, many studies agreed on the form of the

system in terms of the type of common components, while others differed in the number of components.

Among the studies that limited the number of system components to four parts or stages are (Dang et al., 2020; Gupta et al., 2022; Vidya & Sasikumar, 2022; Wang et al., 2019). Figure 2.3 shows these stages as follows:

1. Signal Acquisition: Where the raw signal data are captured using several types of sensors.
2. Data preparation: Cleansing the captured data and feature extraction for better utilization in the next step.
3. Model training: Cleansing the captured data and feature extraction for better utilization in the next step.
4. Activity recognition: It is done based on the trained model to predict human activities.

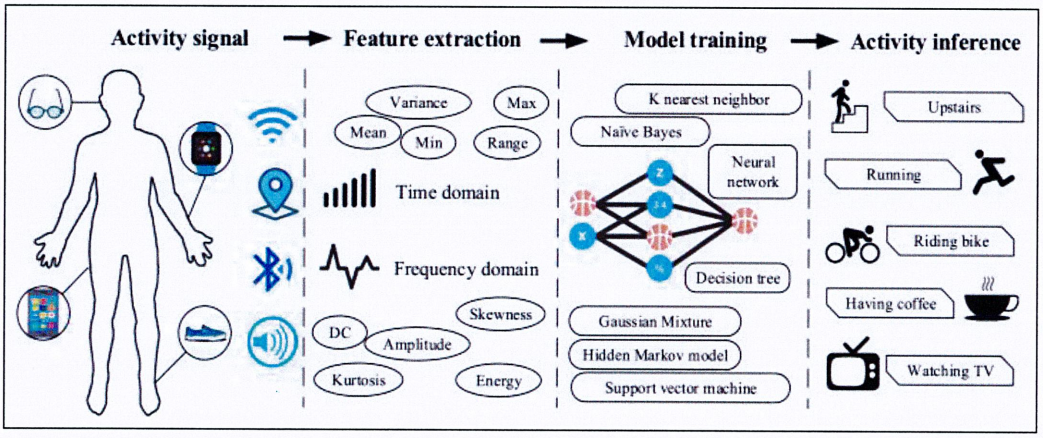


Figure 2.3: Stages of HAR System (Wang et al., 2019)

However, it can be shortened to three stages, as the engineering and feature extraction stage can be combined with the model training stage. It is applied if DL methods are used instead of traditional ML methods. The

potential of these methods lies in the ability to extract features without the usual manual extraction at this stage in the system. (Wang et al., 2019)

2.2. Artificial Intelligence in HAR

In Coinciding with the development of data science and analytics, Artificial Intelligence (AI) has significantly aided the processing of massive volumes of data in many fields. The strength of this technique is its ability to classify the data into several classes employed in a particular dataset. Moreover, this is the core of the human activity recognition system because the resulting classification model using ML or DL methods is critical to the system's success. (Alanazi et al., 2022)

2.2.1. The Impact of Artificial Intelligence

Artificial Intelligence has been more important in recent years in many domains. For example, it enables computers to build prediction models based on the experience of input data processing. Therefore, it clearly can be found in HAR systems which have gained some benefits such as:

- **Automated, accurate and real-time predictions:**
  Many human activity tracking systems benefit from real-time follow-up and guidance from trained models produced by Artificial Intelligence techniques. These models can automatically analyze data in real time without post-processing. Meanwhile, the notable instances of implementing this technology are the devices and applications for Fitness; they are extensively used worldwide. Another instance is monitoring older people's activities and lifestyles (Hayat et al., 2022; Shiraly, 2021).

- **Recognition of Hidden patterns in data:**

  The learning techniques, such as data mining, can obtain meaningful information from vast amounts of data by looking at it from varying distances. It can recognize and identify hidden or untraceable data patterns. In addition, it can make classification and predictions of unseen data. Therefore, it helps make useful suggestions in applying computer vision and Biometric devices (Great Learning, 2023).

Many other benefits have become visible due to the development of artificial intelligence tools with other dimensions, such as the Internet of Things and Big Data. These tools started through traditional ML methods and later developed into DL techniques, where each type has different features.

2.3. HAR using Machine Learning Algorithms

Traditional ML algorithms were the Base of the early human activity recognition systems as the performance metrics have shown that these approaches are effective in prediction accuracy and processing speed (Ambati & El-Gayar, 2021). Therefore, the utilization of these algorithms has been found in many applications of human activity recognition.

Using Significant Skeletal Joints data, K-Nearest Neighbour (KNN) algorithm with extracted spatial and temporal features has shown a higher accuracy performance with low time consumption for processing (Abdul Lateef Haroon P. et al., 2022). While in the field of video sequences, Kushwaha et al. (2021) have used Support Vector Machine (SVM) to process magnitude and orientation information extracted from the optical

flow of video sequences in different datasets. The proposed method performs well along with the existing state-of-the-art methods.

2.3.1. Application On Public Wearable Devices Datasets

Meanwhile, the data of embedded sensors in wearable devices, such as accelerometers and gyroscopes, also have been processed with different ML algorithms.

Bozkurt (2021) has applied several algorithms on the UCI dataset to classify eight activities based on accelerometer and gyroscope measurements. More than 500 features were included in the dataset Due to feature extraction operations. Therefore, The results stated that a hybrid form of SVM and KNN scored 96.71% accuracy among other hybrid and normal algorithms. On the other hand, the SVM algorithm has also been the key to gaining 98.03% in terms of accuracy in the study by Navita & Mittal (2022) using the same dataset. Meanwhile, in a study by Radhika et al. (2022), the Random Forest (RF) algorithm achieved an accuracy of 98.3%, which actually outperforms the outcome of the first study for the same algorithm with the same dataset.

The classification of 11 activities from the IM-WSHA dataset was done by Palimkar et al. (2021) using the KNN algorithm to have the highest accuracy of 98.56%. Also, the other algorithms have scored above 90%, which might effectively recognize human actions in terms of accuracy and efficiency.

2.3.2. Traditional ML On WISDM Dataset

The ML algorithms, which have been applied to the WISDM dataset in recent studies, produced various results because of the difference in the feature extraction methods and techniques.

KNN classifier with Particle Swarm Optimization (PSO) method has the best performance among other classifiers in the study by Al-Taei et al. (2021). The proposed model achieved 91.07% accuracy, which is considered to be near the highest scores that have been recorded, as per the authors. In addition, Abidine and Fergani (2022) have used the KNN algorithm and Support Vectors (SV) as a feature extraction method in a compatible find. The outcome of the proposed approach shows efficiency in accuracy and F-score results with 96.1% for both.

Meanwhile, Amer (2021) found that the SVM classifier performs best in processing the combined data of embedded sensors. For example, the prediction accuracy of the phone and watch models were 96.4% and 94.1%, respectively. Saleh et al. (2020) have tried three algorithms by combining all sensor data and using a feature selection method. The reported result stated that the RF model has the best accuracy, with a score of 79.85%. As per the researchers, the low result is due to experimenting setting such as the number of classes was 18.

The summary of the recent studies in Table (2.1) shows that manual feature extraction was necessary to achieve high accuracy despite the utilized algorithm. Traditional ML algorithms usually suffer in multiclass problems with limited data. Moreover, it has confusing regarding similar activities or different simulations for the same activity. Hence, DL methods

have been used more to overcome these issues. In addition, it can keep the relationships between the instances in data sequences (Ramanujam et al., 2021).

Table (2.1): Summary of recent studies applied to the WISDM dataset using ML algorithms

Study	**Sensor**	**Classes**	**Best Method**	**Notes**
Al-Taei et al. (2021)	Phone Accelerometer (Acc.)	6	KNN	- 561 features extracted from raw data in the source. - 91.07% accuracy. - Optimize with PSO
Hou (2020)	Phone Acc.	6	RF	- No information about features extraction. - 82.7% accuracy.
Abidine & Fergani (2022)	Phone Acc.	6	KNN with SVM	- six features (raw data) - SVM used to extract additional features (SV). - 96.1% accuracy.
Amer (2021)	Phone/Watch Acc./ Gyro.	5	SVM	- 43 features extracted from raw data in the source. - 97.7% accuracy.
Saleh et al. (2020)	Phone/Watch Acc./ Gyro.	18	RF	- 43 features extracted from raw data in the source. - 79.85% accuracy.

2.4. HAR using Deep Learning Methods

Since the beginning, the fast development in DL methods allows a complete activity recognition task. This results in the gradual creation and widespread usage of HAR applications. For example, the hierarchy of the neural networks helps to complete the feature extraction and model learning processes from the original raw data without manual interference. Compared to traditional procedures, this can be done easily and rapidly (Li et al., 2023).

2.4.1. Most Common DL Methods

Following the variety of types of HAR systems, several DL methods have been used to suit the objectives of those systems. Sarveshwaran et al. (2022) have categorized the methods into two main types:

1. Supervised methods: such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN).
2. Unsupervised methods: such as Self Organizing Model (SOM) and Deep Belief Network (DBN).

Regarding the second type, Ige & Noor (2022) focused in their survey on activity recognition in wearable sensor-based datasets, where the scope was the clustering methods. In the first type, the most common methods were Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which can provide excellent results. Furthermore, they are recommended approaches because of their capacity for analysis and awareness of power use (Nguyen et al., 2021).

Although the design of RNN gives the ability to extract temporal information from time series data, it is limited to a short period due to the issue of gradient vanishing. Therefore, Gated Recurrent Unit (GRU) method has been developed based on RNN to overcome this issue. The GRU cell has two gates, one for receiving and updating while the other for resetting or ignoring. In the same context, the LSTM method has been designed with three gates, where it shares the updating and forgetting functions (Gupta S. , 2021).

Based on the above, a recent study has used a combination of Bidirectional GRU and interception layers to record above 98.25% accuracy with the WISDM dataset and above 95% with other datasets (Tong et al, 2022). In contrast, in the study of Lu et al. (2022), a hybrid model of CNN and GRU methods has been used in three public datasets for human activity recognition to achieve above 96% accuracy for all of them. As for the LSTM side, it will be reviewed in the next section.

Through its design, the strength of the CNN method lies in the ability to extract patterns from the data, even complex ones. The convolutional layer takes the mission of feature mapping or extraction; meanwhile, the pooling layer will reduce all the extracted features. Then a fully connected layer is responsible for making the predictions. As a result, adding more convolutional layers means a higher ability to extract complex features (Jaberi & Ravanmehr, 2022).

Dhammi and Tewari (2021) designed a model that has four convolutional layers with different kernel sizes to define the depth of the extraction process. The model gave a 95% accuracy with the WISDM dataset (six classes). With a similar result, a model with five convolutional

layers by Varshney et al. (2022) has a little better score with 96.52 % accuracy with the same dataset also. However, in seeking more efficiency by decreasing the complexity of the model, some researchers invented one dimensional CNN model to suit wearable and smart devices (Yi & Hwang, 2022). For example, Muralidharan et al. (2021) applied this idea to the previous dataset to earn an enhancement over the base model with + 3.0% to reach above 96%.

In conclusion, DL methods have shown excellent performance in classifying human activity using raw data. Moreover, it decreases reliance on domain experts in the feature extraction process. In addition, it has accurate feature detection in temporal data. Hence, the next section will review the subject methods in recent studies.

2.4.2. HAR using LSTM Method

Because of its unique memory cells, LSTM became one of the best RNN variants that have solved the gradient vanishing problem. It outperforms CNN regarding feature extraction from sequence data (Xia et al., 2020). It has been used in many different HAR fields like Video Recognition (Razmah et al., 2022) and Recognition of Suspicious actions by skeleton features (Nale et al., 2021).

In the field of HAR using the embedded sensors of wearable devices, many recent studies have utilized the WISDM (six classes) dataset to train LSTM models. A simple LSTM model by the study of Kolkar et al. (2022) was in second place with 93.34% accuracy below the GRU model and above the CNN model. However, in a confirming study of the previous result, a model developed by Shiranthika et al. (2020) could not exceed an

accuracy of 85% and 76% F1-score. The model was tested with the same dataset and contained one LSTM layer followed by a dropout layer, then a Dense layer, with a size of 100 neurons for all layers.

While in the opposite direction, the result of a study by Sa-nguannarm et al. (2020) stated that the LSTM-based model performs better than the CNN-based model by 12%. The model started with a fully connected layer, then two stacked LSTM layers, followed by the output layer with Softmax activation function for classification.

Furthermore, Pienaar and Malekian (2019) designed a model for real-time classification consisting of three stacked layers LSTM layers to give a 94% prediction accuracy and below 30% loss value. In addition, they applied regularization techniques like adding a dropout layer to avoid overfitting effectively.

In the same way as using stacked layers, a study found that a descending layer size for three LSTM layers scored 99 % accuracy (Benhaili et al., 2022). Also, it found that the performance gets enhanced with a smaller batch size. As a different result for the stacking layers Model, Hou (2020) found that a model with two stacked LSTM layers had an 81% accuracy score. It was better than models with one or three stacked layers by 1% and 3%, respectively.

In addition, some resrarchers have used the smartwatch data only to predict 18 classes. The model has one LSTM layer with two fully connected layers (one as an output layer). The reported result was 89.6% accuracy (Mekruksavanich et al., 2020).

To improve the LSTM model performance, some studies have used additional optimization methods such as Particle Swarm Optimization (PSO) to increase the accuracy by up to 97% in this study (Gajjala & Chakraborty, 2021), while another study used "SigOpt" model with baseline LSTM Model to have 98.26 % accuracy and 99.55 % F1-score (Rojanavasu et al., 2021).

In conclusion, there were different results for the recent studies that applied the LSTM model using the WISDM dataset due to the various implementation techniques. In addition, as per the author's knowledge, the standard LSTM model was not implemented on the same dataset using 18 classes or all sensors' data. Hence there is a need to explore these aspects of the data when using the LSTM method. Table (2.2) shows more details about some configurations in recent studies.

Table (2.2): Summary of recent studies on the WISDM dataset using LSTM method

Study	Sensor Type	Classes Count	Model Layer's Structure	Notes & Results
Kolkar et al. (2022)	Phone Acc.	6	Not provided	- No Normalization reported. - 93.34 % accuracy
Shiranthi-ka et al. (2020)	Phone Acc.	6	LSTM, Dropout, Dense.	- Maximum absolute value data scaling. - 84.71% accuracy, 76% F1-score.
Sa-nguannarm et al. (2020)	Phone Acc.	6	Dense, (2) LSTM, Dense.	- No Normalization reported. - Data segmented by Sliding window. - 95.82% accuracy.
Pienaar & Malekian (2019)	Phone Acc.	6	(3) LSTM, Dense.	- No Normalization reported. - Data segmented by Sliding window. - 94% accuracy, 30% loss
Benhaili et al. (2022)	Phone Acc.	6	(3) LSTM, Dropout, Dense.	- No Normalization reported. - Data segmented by Sliding window. - 99 % accuracy.
Hou (2020)	Phone Acc.	6	(2) LSTM, Dense.	- No Normalization reported. - 81 % accuracy.
Mekruksav-anich et al. (2020)	Watch Acc. & Gyro.	18	LSTM, Dropout, Dense.	- No Normalization reported. - Data segmented by Sliding window. - 89.6 % accuracy.

2.4.3. HAR using GCN Method

Graph neural network (GNN) recently achieved progress in the fields of spatiotemporal relations and time series forecasting. Furthermore, compared to other networks that can only study data standalone, this method may gather deeper information by Finding structural linkages between the data elements. Therefore, it has successful applications in traffic flow, weather forecasting, fraud detection, and medical fields. While the CNN model cannot reflect the complex topological structure, a generalization was made with GNN to form the Graph Convolutional Network (GCN), which can handle any arbitrary graph-structured input. It has the ability to work with data that have spatial and temporal dependencies. (Lira et al., 2022; Liu et al., 2022; Zhao et al., 2020)

The usage of the GCN method focuses on the applications of the skeleton modality despite the diversity of systems for identifying human activities. It is because the link between the spatial and temporal domains is clearly described. In a survey by Ahmad et al. (2021), the authors reviewed many types and techniques of the GCN method that have recently been used in skeleton-based activity recognition. In contrast, few recent studies have explored this method in the applications of human activity recognition using embedded sensors in wearable and smart devices.

To build a deep transfer learning model, Liao et al. (2022) introduced a multi-layer residual structure, including a graph convolutional neural network termed as "ResGCNN". The model is designed to deal with various datasets with different activities or sensor settings. It mainly comprises blocks with Chebyshev graph filtering, graph normalization, and activation layers. The data have been sliced into smaller parts, each one

considered as graphs with an adjacency matrix describing the connection relationship between data channels. The results of the experiments on mHealth and PAMAP2 datasets recorded a high average accuracy of 98.18% and 99.07%, respectively.

Meanwhile, another study tried a novel training method that requires the model to predict unknown activity labels using the known labels (Mohamed et al., 2022). Labeled activity measurements were combined with a few unlabeled ones as vertices to form a graph using their relationship as edges. The data have been passed to a Graph CNN layer connected to a set of CNN layers with a classification layer at the end, see Figure 2.4. The model was compared against LSTM and CNN baseline models. As a result of using the Extra-Sensory dataset, it has about 25% and 68% improvement in performance, respectively. Meanwhile, with the PAMAP dataset, the model gave nearly 2% F1-score enhancement with almost 100% accuracy.

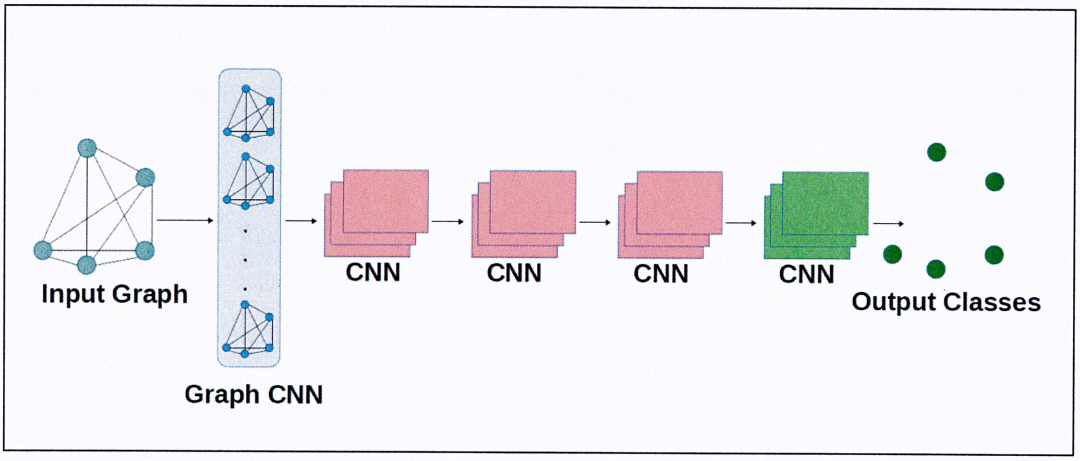


Figure 2.4: HAR – GCN Model Description (Mohamed et al., 2022)

The study of Mondal et al. (2021) introduced a new framework for HAR based on smartphone sensors using a graph neural network to consider the

structural relations among data samples. First, the data of six public datasets were sliced individually based on users and activities. Then the resulting slices (graphs) were divided to none overlapping windows, where every window was considered a node. The proposed model was composed of blocks of "GraphicConv" followed by a "Top-KPooling" layer, and then their output is passed to a linear layer and a final classification layer. As a general result of model performance, the model has achieved almost 100% prediction accuracy among all six datasets, including WISDM (6 classes).

To maintain data privacy, Sarkar et al. (2021) proposed a novel method using GNN and federated learning, where the model weights can be shared to help prediction for unlabeled activity data. The WISDM dataset with nine classes has been preprocessed the same as the earlier study, where the nodes were connected based on "Mahalanobis" distance with a threshold. While the method has two models, centralized, and federated, they share the same structure with a convolution layer, a fully connected layer, and a SoftMax activation layer for classification. The result of the centralized model scored 91.0% accuracy, slightly lower than the baseline model (forward neural network). In contrast, the federated model has higher accuracy than the baseline model, with 81.7%. The authors attribute the performance variation to the problem of imbalanced classes in the dataset.

Based on all the above, the GCN method has been implemented using some datasets with notable performance against other methods. However, the WISDM dataset has shown different results due to the variation in the number of classes. In addition, there were no comparisons to measure the performance of the GCN method against state-of-art methods like LSTM. Hence, this research aims to fill this area of knowledge.

2.5. Conclusion

The recognition of daily human activities has gained importance from researchers since it deals with many applications that help people make their lives easier. Therefore, HAR systems have been developed with embedded sensors, such as accelerometers and gyroscopes. Accordingly, Artificial Intelligence can get helpful information from sequence data of sensors for activity recognition by using machine and DL methods.

Despite its efficiency, traditional ML methods require a feature extraction process before the implementation step, which is usually done manually. On the other side, DL methods can do that inside the training stage, so they can deal with the raw data directly by considering the sequence aspect. The LSTM method has a large share of use among these methods due to its proven efficiency in such data types. At the same time, the GCN got considerable attention in recent years due to its notable performance with these data.

Although, there was a shortage of studies in terms of exploring and comparing the performance of both methods with a dataset with many classes and features like WISDM. For that, this research aims to develop an accurate model for both methods using the subject dataset and then compare their performance results.

Chapter Three

Methods and Procedures

3.0. Overview

3.1. Research Methodology

3.1. Research Dataset

3.2. Implemented Models

3.4. Evaluation Metrics

Chapter Three
Methods and Procedures

3.0. Overview

The first section of this chapter opens with a review of the research methodology; it deals with the research design. While the second section describes the research's dataset, the third section covers the algorithms used to create the models of this research. Finally, the last section discusses the metrics for evaluating and comparing the performance of the implemented models.

3.1. Research Methodology

This research adopts the Cross Industry Standard Process for Data Mining (CRISP-DM) as a reference methodology. That is because it suits the research idea and has flexibility between its steps, as it is an essential feature for such type of research. The illustration of the adopted methodology can be found in Figure 3.1, showing four stages to be followed during this research.

The first stage is defining the research idea, including the required questions and objectives, then conducting the literature review for previous studies and efforts. Many aspects related to HAR were covered in Chapter 2, including the recently applied methods and algorithms with their results. Therefore, The output of the literature review leads to selecting the necessary data and tools for research implementation.

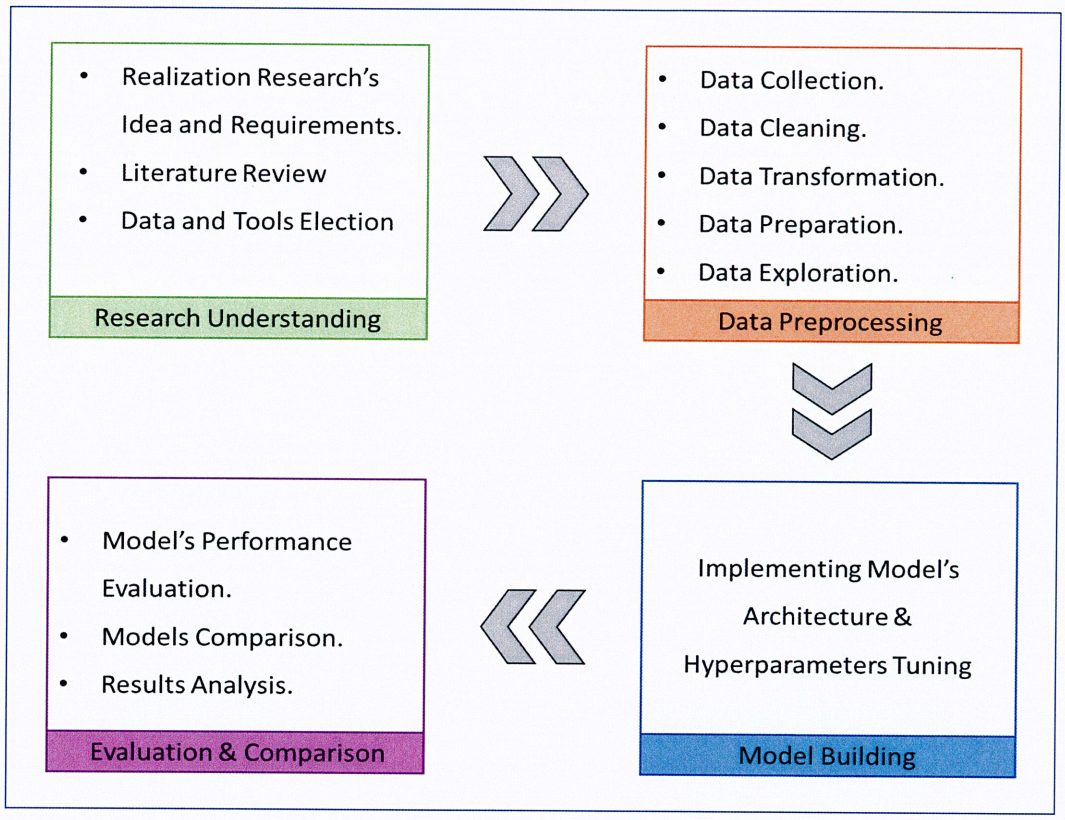


Figure 3.1: The Illustration of Research Methodology

While the second stage is related to the selected data (WISDM dataset), it describes the collecting process, then many pre-processing steps starting with data cleaning operations such as dealing with missing and null values and checking for outliers. After that, the transformation of some features, such as the one-hot encoding for the labels. The next step is data preparation for feeding the models, as each model's input data format may differ. Finally, the exploration and analysis of the data are done during this stage to get a better understanding of it.

In the third stage, the selected methods, LSTM and GCN, are trained with the result of previous stage data through set upping a suitable architecture and tuning the hyperparameters. This stage will be applied for four models as each method will have two respecting the provided data.

This stage contains many experiments and iterations to get an accurate model.

The last stage is depending on the former stage, as it is the evaluation process for the resulted trained models. The evaluation is done using many metrics such as Accuracy and loss scores. The comparison part in this stage is aiming to find the best models that have the best performance by using also suitable metrics such as F1-score.

3.2. Research Dataset

3.2.1. Dataset overview

The selected dataset for this research is the WISDM dataset. It was produced by Wireless Sensor Data Mining Lab at Fordham University (Weiss, 2019) and downloaded from the "UCI" repository. It was selected since it is a time series dataset with high number of instances and classes, so it can be used to explore the classification performance with such big numbers. In addition, it contains measurements of two embedded sensors in smart devices (Accelerometer and Gyroscope).

A smartphone (Google Nexus 5/5x and Samsung Galaxy S5) and a smartwatch (LG G Watch) were worn by 51 persons while conducting 18 activities (classes) for 3 minutes for each activity. The smartphone was in the participant's pocket, while the smartwatch was in the dominant hand. The data were logged at a rate of 20 Hz (one instance every 50ms). The list of activities can be found in Table (3.1), where every activity was coded and grouped into one of 3 groups based on hand orientation during the performing process.

Table (3.1): List of Performed Activities (Classes)

Activity	Code	Group
Walking	A	Non-hand-oriented activities
Jogging	B	
Stairs	C	
Sitting	D	
Standing	E	
Typing	F	Hand-oriented activities (General)
Brushing Teeth	G	
Eating Soup	H	Hand-oriented activities (Eating)
Eating Chips	I	
Eating Pasta	J	
Drinking from Cup	K	
Eating Sandwich	L	
Kicking (Soccer Ball)	M	Non-hand-oriented activities
Playing Catch w/ Tennis Ball	O	Hand-oriented activities (General)
Dribbling (Basketball)	P	
Writing	Q	
Clapping	R	
Folding Clothes	S	

The dataset is composed of 204 raw data files, where each person has four files as there are two sensors (Accelerometer and Gyroscope) in each device. The total number of measurements is 15,630,426 lines, and each line is formatted to define six features, as shown in Table (3.2). A sample of two lines from the raw data is below:

$$1600, A, 252207666810782, -0.36476135, 8.793503, 1.0550842;$$
$$1600, A, 252207717164786, -0.8797302, 9.768784, 1.0169983;$$

Table (3.2): Description of Features for Raw Measurements

Feature Name	Type	Definition
Subject-ID	Numeric	Subject Identifier, Range:1600-1650.
Activity code	Text "Letter"	Range: A-S (no "N" value) (Class)
Timestamp	Numeric	Linux time
x	Numeric	Sensor value for x axis
y	Numeric	Sensor value for y axis
z	Numeric	Sensor value for z axis

3.2.2. Dataset Preprocessing

The preprocessing phase consists of three stages, Data Combination, Data Transformation and Cleaning, and Data preparation. It was done using Python libraries in Google Colab environment.

Data Combination: The first step in this stage was to group the raw data files based on device type, which means there are two datasets, one for the smartphone and the other for the smartwatch. In the second step, for each resulting dataset, the files were merged based on sensor type. Then, the resulting sub-datasets were merged again based on the Timestamp value to synchronize the measurement values of both sensors. The final data size for the smartphone is 2909149, while for the smartwatch is 3370860, in addition to nine features for both. Furthermore, a sorting operation was done using the "Subject-ID" and Timestamp to reserve the data sequence.

Data Transformation and Cleaning: The starting step was to convert the data type of two columns from text to numeric values, which were the z-axis value of both sensors. Then the data were checked for

missing, null, and inappropriate values, but no records were found. Regarding the normalization step, the data have not been normalized as it is affecting the performance of DL methods as per the author's experiment and based on the result of a recent study (Pires et al., 2020).

Data Preparation: In order to feed the model with the data, the overlapped sliding windows method was implemented due to its performance in DL methods. Furthermore, it helps extract the patterns in sequence data (Dehghani et al., 2019). Therefore, A sliding window of size equal to 100 instances (5 seconds) was chosen with ten instances as a time step to create an overlapping percentage of 90% as calculated in Equation (1):

$$O = \frac{(w-s)}{w} \tag{1}$$

Where $O, w,$ and s are Overlapping percentage, Window size, and Step size, respectively. Then each window assigned its class by taking the mode of class value for included instances. After that, the data were converted to a graph dataset as an extra step for the GCN method only. The conversion process is covered in the next section.

The last step in the data preparation was splitting the data. The data split was done using 70% for training, 10% for validation, and 20% for testing. The results of the preprocessing stage will be in the next chapter.

3.3. Implemented Models

This section presents the subject methods in the research through a theoretical overview first, then the declaration of the models' setting, which have been implemented during the experiments. Finally, the development process of the models has been done using the TensorFlow framework in the Google Colab environment.

3.3.1. Long Short-Term Memory (LSTM)

3.3.1.1. Overview:

Long Short-Term Memory (LSTM) is type of Recurrent Neural Networks (RNNs), which have been created specially to process sequential data, because they can learn long-term relationships in the data. LSTM networks are admirably adapted for tasks like translating between languages, recognition of speech, and time series predictions.

The problem of Vanishing Gradient problem in RNN is solved by LSTM with its memory cell. It is a storage unit that has three gates to hold information for lengthy period (Ingolfsson, 2021). The structure of this cell as presented in Figure 3.2, shows the three gates as following:

1. **Forget Gate:** the controlling gate which gets rid of the data that is no longer helpful in the cell state. It has two inputs, the first is the input of current time (X_t), while the second is the output of previous cell (h_{t-1}). The filtration process is done by Sigmoid activation.

2. **Input Gate:** It takes the mission of addition of information to the memory cell state. The input of this gate is the useful output of Forget gate. The activation functions in this gate which they are Sigmoid and Tanh, are responsible for processing the inputs.

3. **Output Gate:** It determines what information is released from the memory cell either to the next cell or as the final output. The task is done by Sigmoid activation function for hidden state, while the Tanh activation function is for the cell state.

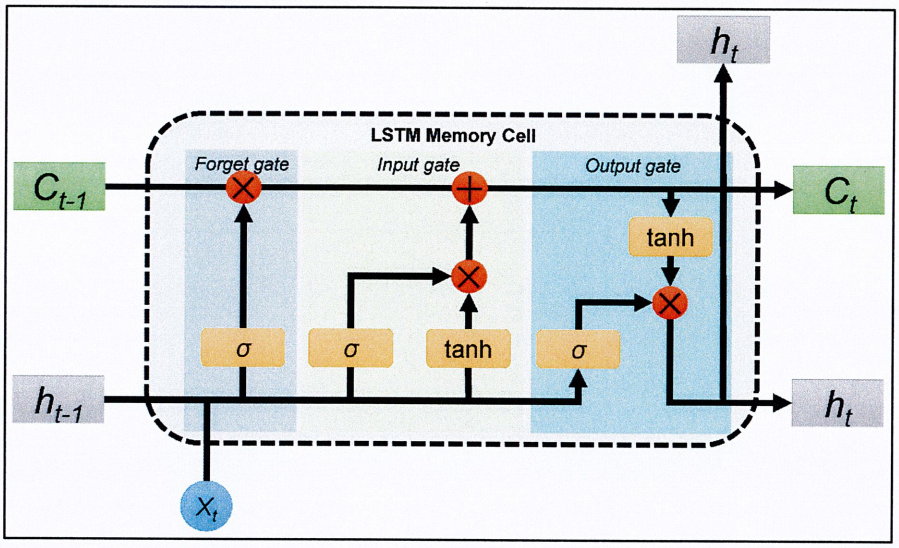


Figure 3.2: LSTM Cell Architecture (Fan et al., 2020)

3.3.1.2. Implemented Model Settings

Since there were two sub-datasets resulted from the parent dataset based on the device type, a two LSTM models have been implemented to suit each one. The model of smartphone sub-dataset is formed as showing in the Table (3.3), while the form of the model for smartwatch sub-dataset is presented in Table (3.4). Both models were fitted using Adam optimizer and Cross-Entropy loss function.

Table (3Error! No text of specified style in document..3): LSTM Model Settings for Smartphone Sub-dataset

Layer Type	Size	Notes
LSTM	128	Activation Function (Tanh)
Dropout	0.5	-
Fully connected	64	Activation Function (Relu)
Fully connected	18	Output with Activation Function (Softmax)

Table (3.4): LSTM Model Settings for Smartphone Sub-dataset

Layer Type	Size	Notes
LSTM	128	Activation Function (Tanh)
Dropout	0.5	-
Fully connected	64	Activation Function (Relu)
Fully connected	18	Output with Activation Function (Softmax)

3.3.2. Graph Convolutional Network (GCN)

3.3.2.1. Overview:

The Graph Convolutional Network (GCN) is an emerging DL method which the researchers have demonstrated its usefulness in many fields, such as social networking analysis and computer vision. It has been developed by generalizing convolution neural models to deal with structured graph data (Baroud et al., 2021). The graph data is passed to feeding a graph convolutional layer to deal with such type of data as shown in Figure 3.3.

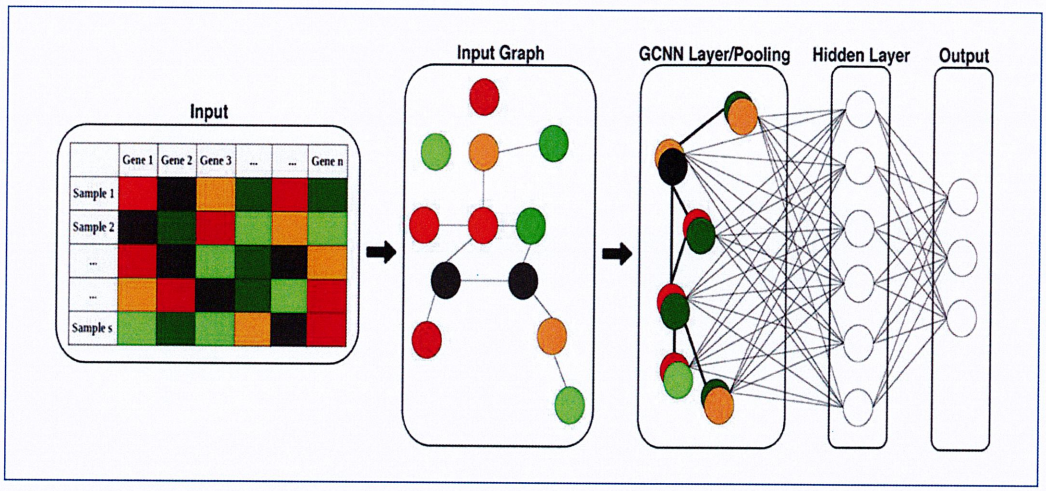


Figure 3.3: Simple graph convolutional network design (Ramirez et al., 2020)

The main component in such type of data is the Graph which can be defined a graph $G = (V, E)$ where V means set of vertices (nodes) and E refer to the set of edges. Any two nodes $v_i, v_j \in$ V are connected by an edge $e_{ij} = (v_i, v_j)$ as shown in Figure 3.4. The common way to describe the relations of nodes in a graph is the adjacency matrix $A \in R^{N \times N}$ where $N = |V|$. An edge has value $A_{ij} = 1$ if there is connection between the node v_i and the node v_j , otherwise it has value $A_{ij} = 0$. While the Degree matrix D is representing number of neighbors for each node. Edges can be described as directed when the source and the target nodes are defined, while it is undirected when there is no definition.

The nodes (also edges) can have a set of assigned features or properties in a form of (vector) $a = (a_1, a_2, a_3, \ldots\ldots, a_n)$ where an individual feature $a_i \in \Omega_A$ of a feature domain. The features matrix represented as $X \in R^{z \times n}$ where z is number of nodes and n is number of assigned features for each node (Bloemheuvel et al., 2022).

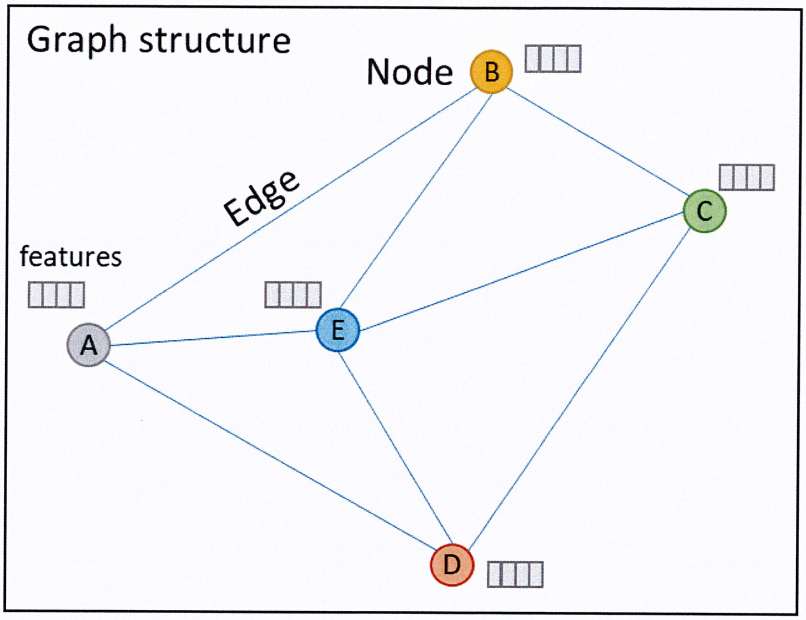


Figure 3.4: Simple Graph Data structure

The classification problems using graph structure data could be on graph level, node level and link (edge) level, where the required labels for prediction are assigned to it. GCN model used the filter parameters everywhere in the graph structure to learn a function of assigned features. In graph level, a nonlinear function can be written for GCN layer as:

$$\mathrm{H}^{(\mathrm{L}+1)} = f\left(\mathrm{H}^{(\mathrm{L})}, \mathrm{A}\right) \tag{2}$$

Where $H^{(0)} = X$ and $H^{(L)} = Z$ which is node-level output of past layer, L is number of Layers. Other models will have difference in $f(.,.)$ in selection of inputs and parameters. More theoretical details about GCN are presented in (Kipf & Welling, 2016).

3.3.2.2. Implemented Model Settings

The adopted classification level for the GCN models is the graph level. Therefore, each resulting window from the preprocessing stage is considered a graph, and the sequenced instances are selected as nodes. Meanwhile, the sensor measurements of each instance represent the node features. The combined features in each graph are the feature matrix (X). The class of each graph is the same as the class of the matched window, while the class matrix (Y) is constructed using the one-hot encoding technique. Regarding the formation of the adjacency matrix (A), each graph is considered an undirected graph, where the degree of the first and last nodes is one while the rest is two.

The graph dataset is formed by combining all graphs where each graph is represented as $G = (X, A, Y)$. The GCN model and the dataset are built based on a Python library for graph DL called Spektral. It has many tools, such as convolutional, pooling layers, and data management functions (Grattarola & Alippi, 2020).

Regarding the structure of the two built models for the GCN method, Table (3.5) displayed the model of the Smartphone sub-dataset, followed by Table (3.6) for the model of the Smartwatch sub-dataset.

Table (3.5): GCN Model Settings for Smartphone Sub-dataset

Layer Type	Size	Notes
"GCNConv"	256	Activation Function (Tanh),
"GlobalAvgPool"	-	Pooling layer
Fully connected	256	Activation Function (Relu)
Fully connected	18	Output with Activation Function (Softmax)

Table (3.6): GCN Model Settings for Smartwatch Sub-dataset

Layer Type	Size	Notes
"GCNConv"	256	Activation Function (Tanh),
"GlobalAvgPool"	-	- Pooling layer
Fully connected	256	Activation Function (Relu)
Dropout	0.1	-
Fully connected	18	Output with Activation Function (Softmax)

As for model fitting settings, both models were fitted using Adam optimizer and Cross-Entropy loss function.

3.4. Evaluation Metrics

The need to evaluate the performance of the models under study lies in two points; the first is to know the quality and reliability of these models when used in prediction. Meanwhile, the second is to make a comparison between them as stated in the subject of the research. The definition of used metrics for evaluating the classification models is as follows.

- **Accuracy:** It is the ratio of the correct predictions to the total predictions.
- **Precision:** It is defined by the ratio of the correct prediction to the total predictions relevant to a particular class.
- **Recall:** Refers to the ratio of the correct predictions to the actual labeled instances of a specific class.
- **F1-Score:** The harmonic mean of the recall and precision values.

- **Receiver Operating Characteristic (ROC):** the plotted graph of the relationship between the actual positive rate (Y-axis) to

the false positive rate (X-axis). Then the Area Under the Curve (AUC) is computed from the resulting graph, which measures the separability of one class against others.

In addition to the help of the confusion matrix. Which is a table layout that compares predicted instances against actual ones for all classes. Interested readers can refer to (Hossin & Sulaiman, 2015) for more details about the mentioned metrics.

Chapter Four

Results and Discussion

4.0. Overview

4.1. Exploratory Data Analysis Results

4.2. Evaluation of Smartphone Sub-dataset Models

4.3. Evaluation of Smartwatch Sub-dataset Models

4.4. Conclusion

Chapter Four
Results and Discussion

4.0. Overview

This chapter is dedicated to reviewing the results of this research, followed by a discussion and then the conclusion at the end. The results section begins with presenting the results of the Exploratory Data Analysis (EDA) and data pre-processing process. Then it goes through the results of implemented models for each sub-dataset with the evaluation and discussion at the end.

4.1. Exploratory Data Analysis Results

Since there are two sub-datasets based on the data collection device and their embedded sensors, it is important to explore the data to view and compare the distribution and relationships for each one.

4.1.1. Smartphone Sub-dataset EDA Results

As passed in the previous chapter, this sub-dataset contains 2909149 instances, each with six features of accelerometer and gyroscope measurements on three axes. For example, Figure 4.1 shows samples of Accelerometer measurements over 20 seconds for Walking and Jogging activities for 20 seconds (400 instances). At the same time, Gyroscope measurements are shown in Figure 4.2 for the same activities and period.

The contrast of oscillations in each axis is notable when comparing the two activities. While each graph seems to have almost a consistent shape, the accelerometer graph for Jogging activity is an exception.

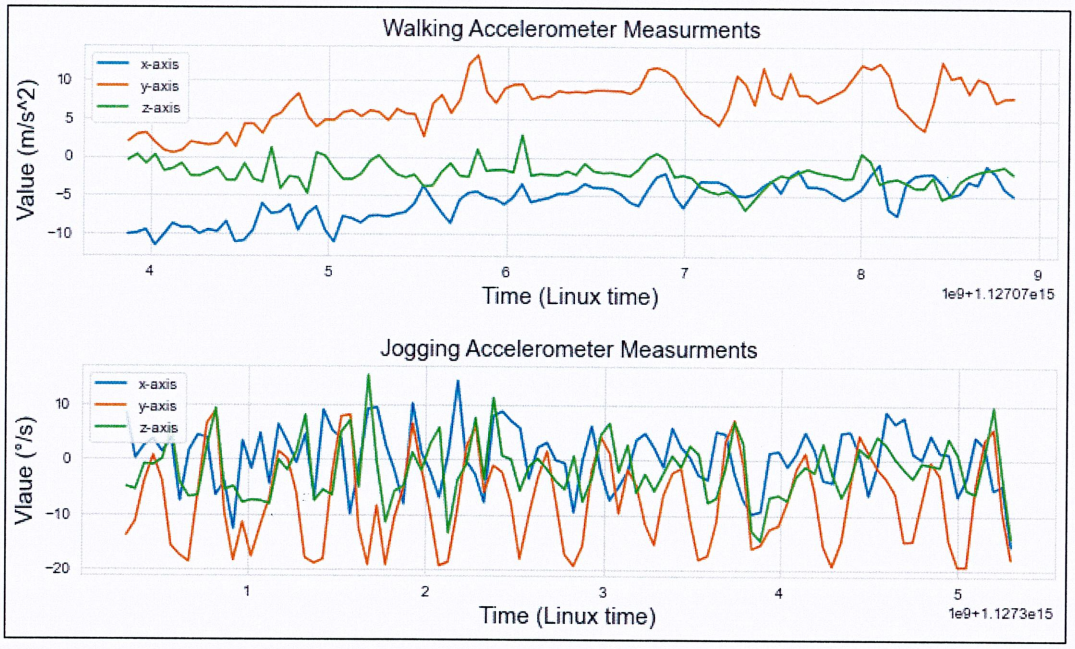


Figure 4.1: Smartphone's Accelerometer Measurements for Some Activities

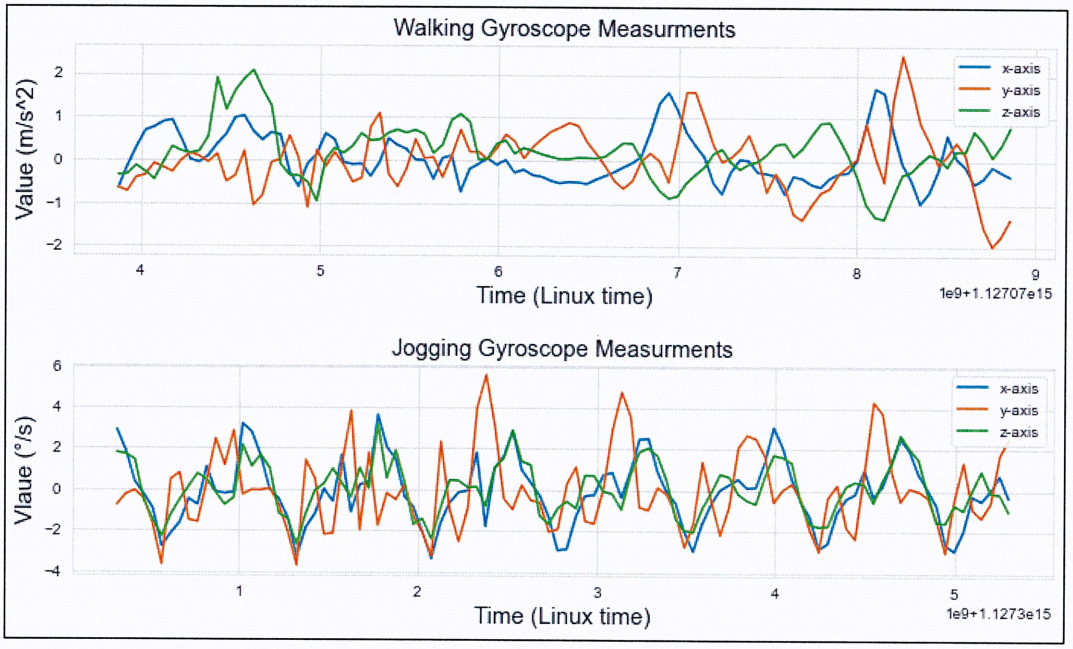


Figure 4.2: Smartphone's Gyroscope Measurements for Some Activities

For example, Figure 4.3 depicts the correlation matrix for Kicking activity to show the interactions of each sensor feature. In addition, to examine their internal and external relationships. For example, the accelerometer sensor's X and Z axes have the highest correlation values.

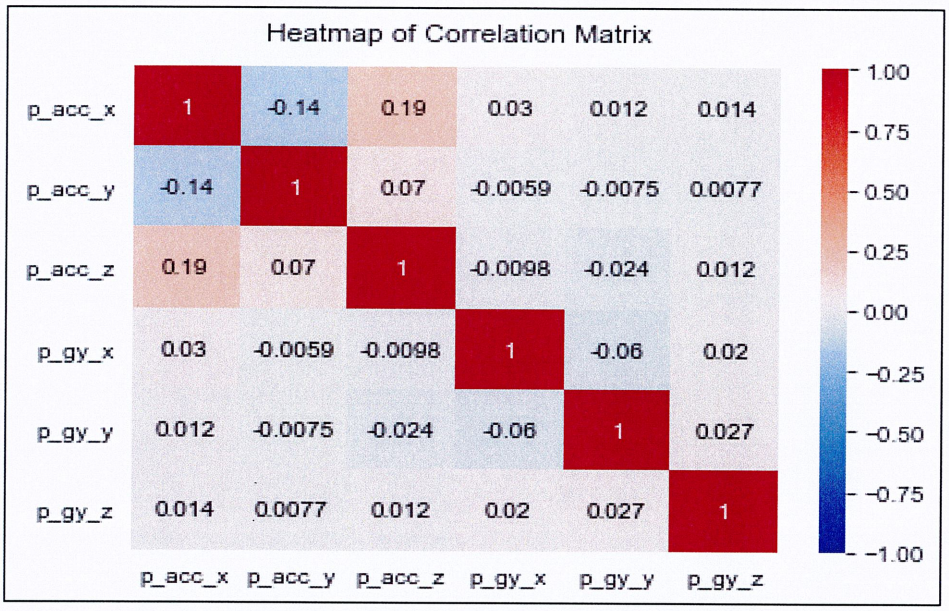


Figure 4.3: Correlation of Sensor's Measurements for Kicking Activity

It is essential to scan the values range of each feature to detect any anomalies. The box plot in Figure 4.4 shows the spread of the values for each axis in the subject sensors. Undoubtedly, both sensors have outlier values located out of the fourth quartiles as circles. Nevertheless, since the dataset is an outcome of a controlled lab experiment, these values are unlikely to be measurement errors. Furthermore, there is a slight skewness in the z-axis of the Accelerometer to the positive direction, while the others look normal.

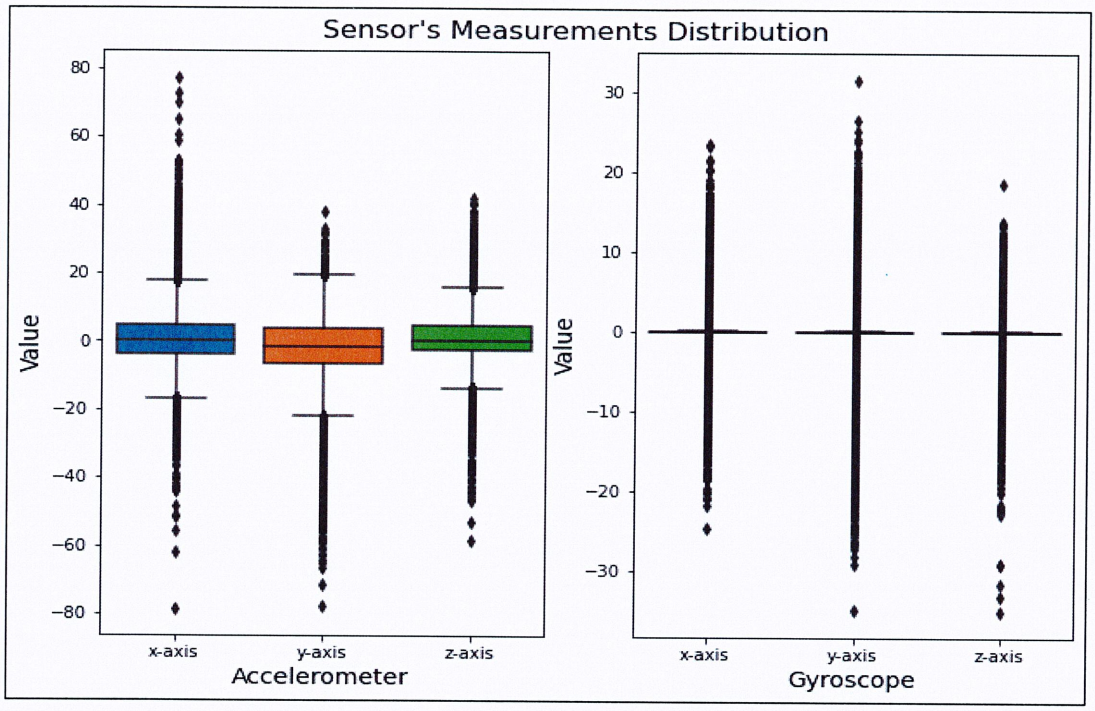


Figure 4.4: The Spread of Sensor's Values in Smartphone Sub-dataset

Regarding the count of classes in this sub-dataset, which are the activities, the distribution of the classes over the instances is plotted in Figure 4.5. there is a slight but notable variation in the count of each class, although the classes seem balanced.

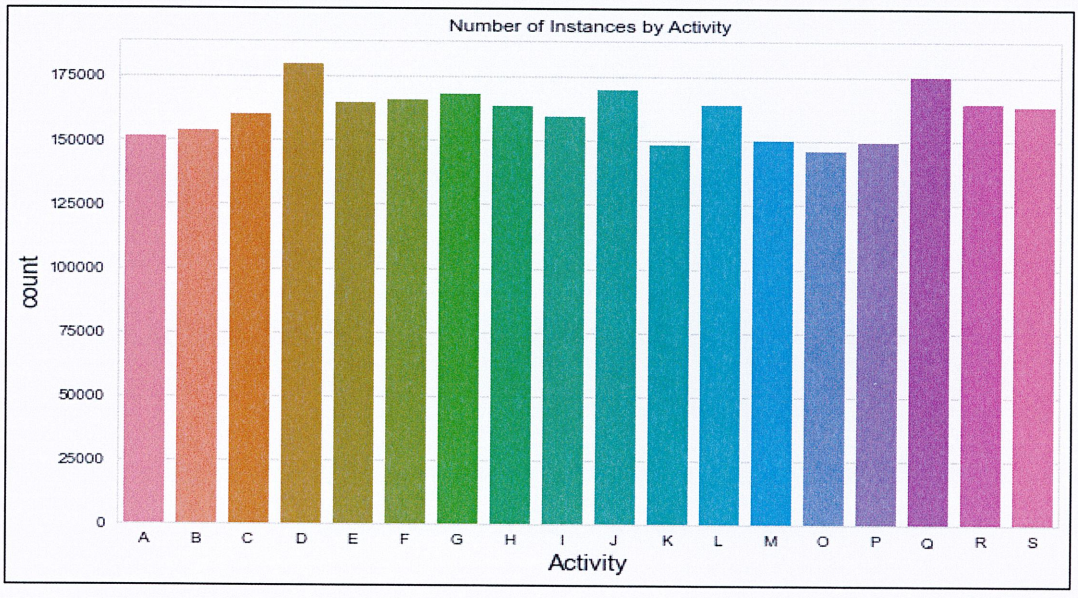


Figure 4.5: Distribution of Activities in Smartphone Sub-dataset

4.1.2. Smartwatch Sub-dataset EDA Results

The resulting size of this sub-dataset is 3370860, which is higher than the smartphone sub-dataset. By looking at Figures 4.6 and 4.7, which represent samples of Accelerometer and Gyroscope measurements for Clapping and Brushing Teeth activities, it can be noted that a dominant axis has the highest or lowest fluctuations than others.

While looking for the relationships in terms of numbers for Clapping Activity, the correlation matrix of the measurements for both sensors is shown in Figure 4.8 as a heat map graph. The y-axis in Gyroscope has a notable relationship with the z and x-axis in the same sensor and the x-axis in Accelerometer but in a different direction.

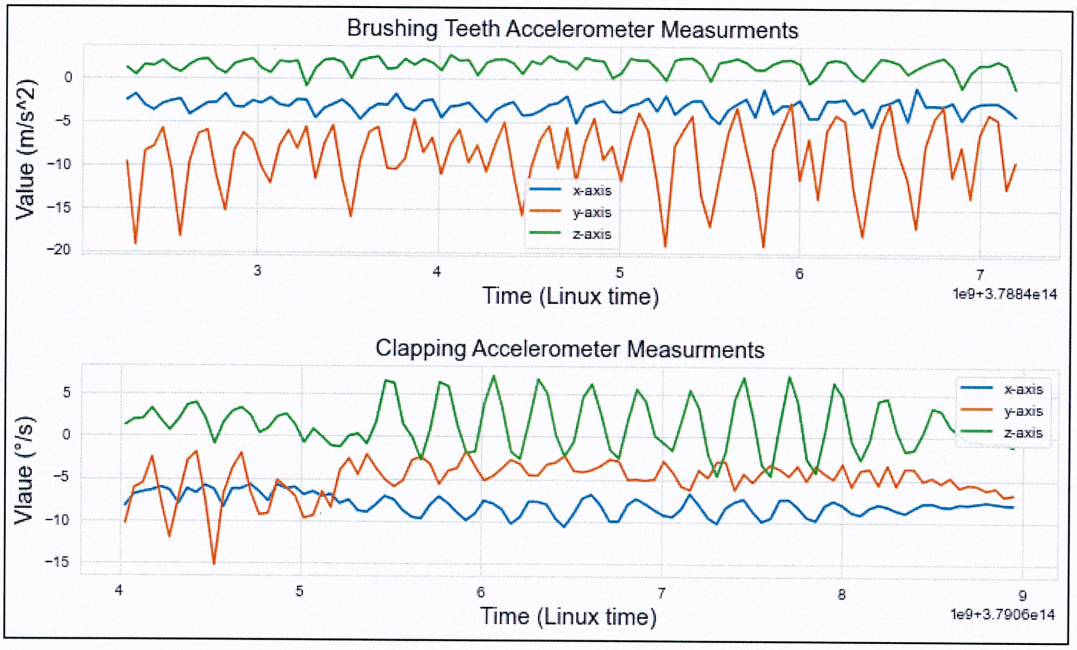


Figure 4.6: Smartwatch's Accelerometer Measurements for Some Activities

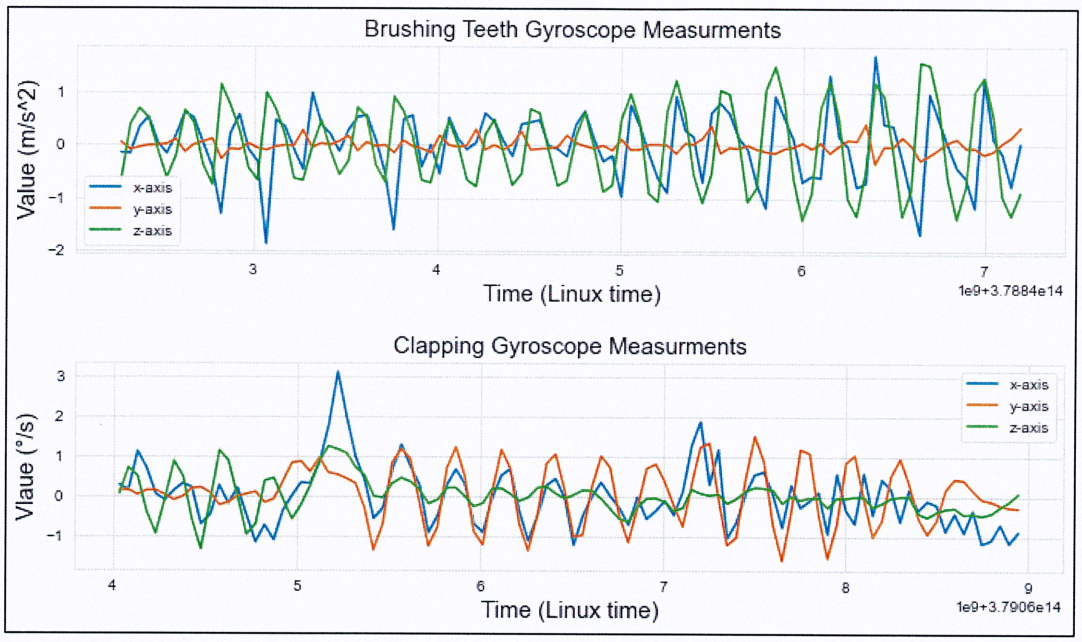


Figure 4.7: Smartwatch's Gyroscope Measurements for Some Activities

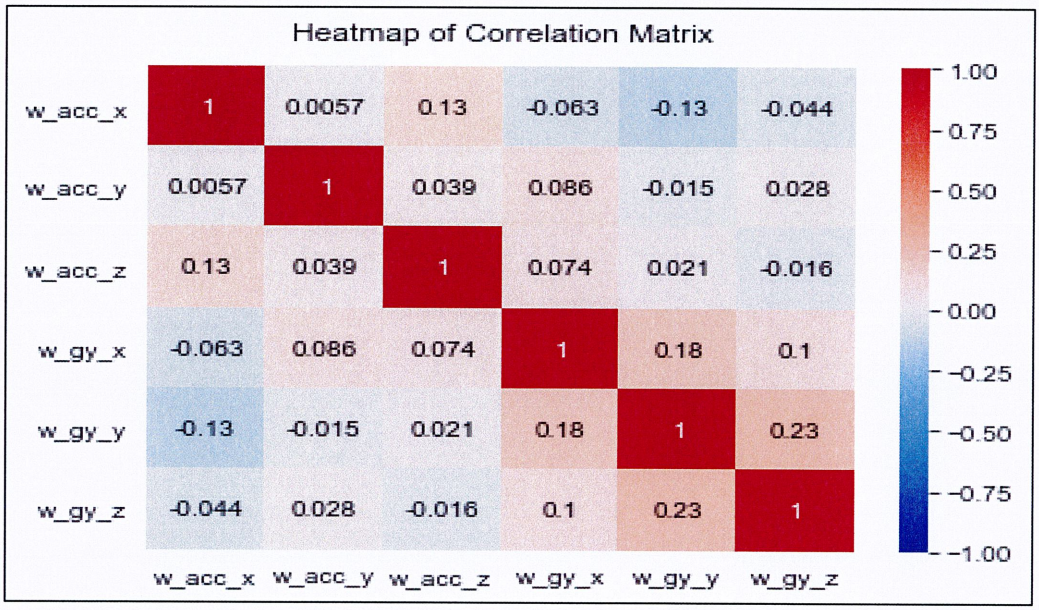


Figure 4.8: Correlation of Sensor's Measurements for Clapping Activity

For the spread of data in this sub-dataset, Figure 4.9 comes with the same previous note regarding the outlier values. While for data skewness, the x-axis of the Accelerometer has an apparent skewness to the positive direction, the median value is around 0, opposite to the other axes where the medians are above and below the 0 value.

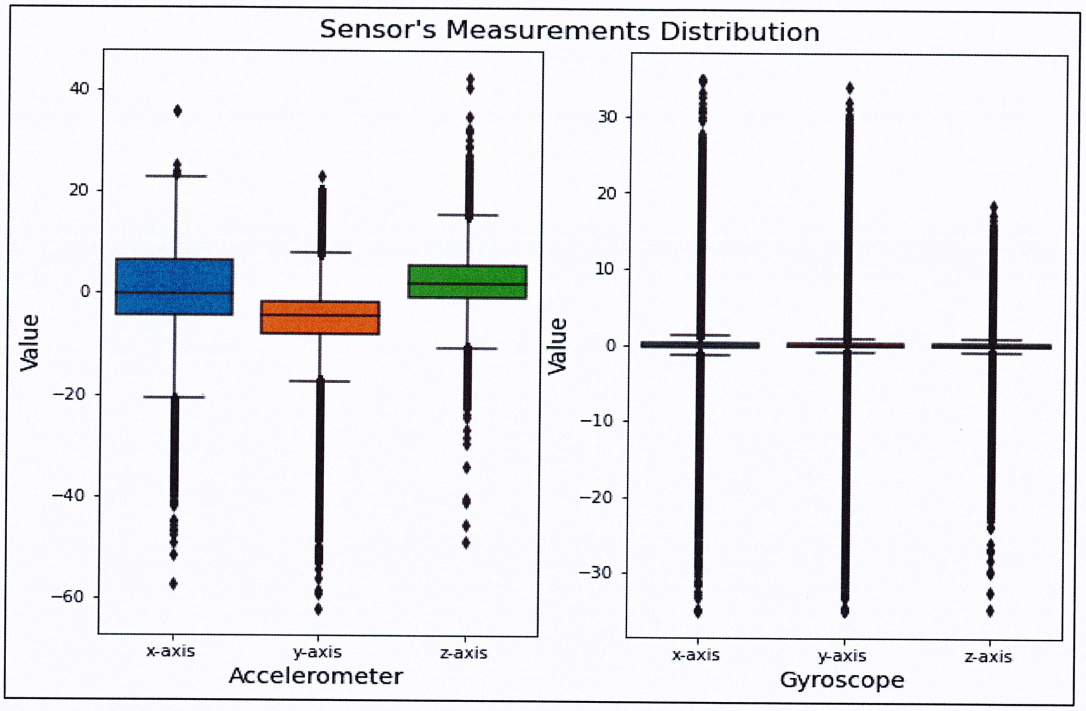


Figure 4.9: The Spread of Sensor's Values in Smartwatch Sub-dataset

In comparison to the smartphone sub-dataset, the subject sub-dataset seems to have more balanced classes, although there is a slight variation in the count of each class also. Figure 4.10 shows the distribution of classes in the subject sub-dataset.

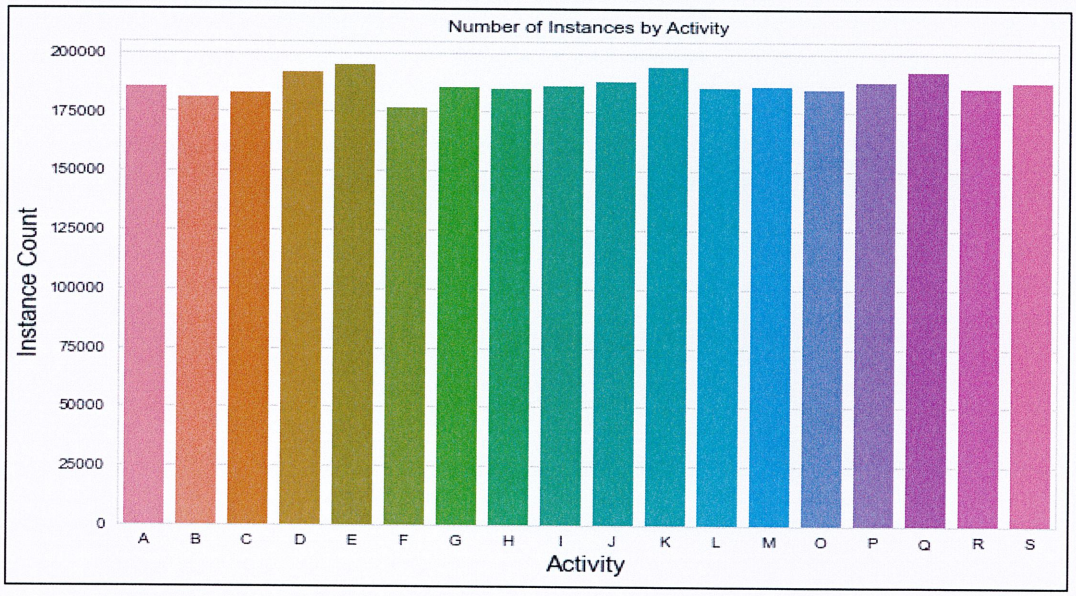


Figure 4.10: Distribution of Activities in Smartwatch Sub-dataset

4.2. Evaluation of Smartphone Sub-dataset Models

This section will present the evaluation result of implemented models on the smartphone sub-dataset. For each model, the results of some hyperparameters tuning, model training, and model testing will be presented, followed by the comparison between the models and the discussion.

4.2.1. Performance of LSTM Model

4.2.1.1. Hyperparameters Tuning

Based on initial experiments with related literature review configurations, the hyperparameters under focus will be the layer size, the number of epochs, and batch size. In contrast, the number of layers is limited to one method layer and two fully connected layers, where one of them is the output layer.

The effect of layer size on this model is shown in Figure 4.11, where a size of 128 is recommended. There was no significant improvement in accuracy and loss values in case of increasing the layer size; contrarily, it will increase model complexity. Regarding the batch size, Figure 4.12 shows that the best batch size is 64, as the highest and lowest values of accuracy and loss, respectively, at this size. While for epochs size, the largest size was 50 epochs, as the early stopping method was applied to prevent overfitting while training.

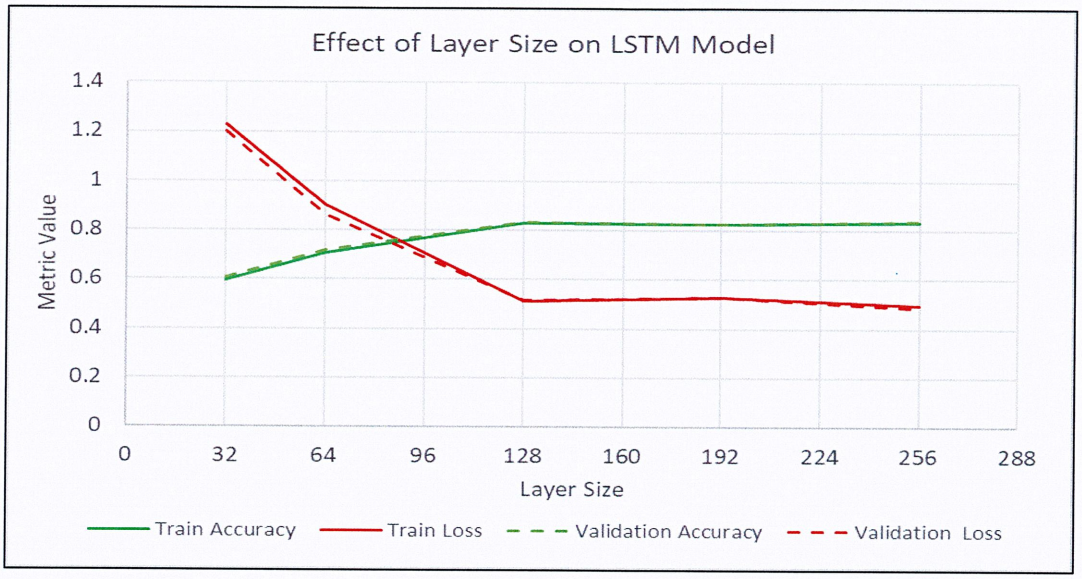


Figure 4.11: Effect of Layer Size on LSTM Model of Smartphone Sub-dataset

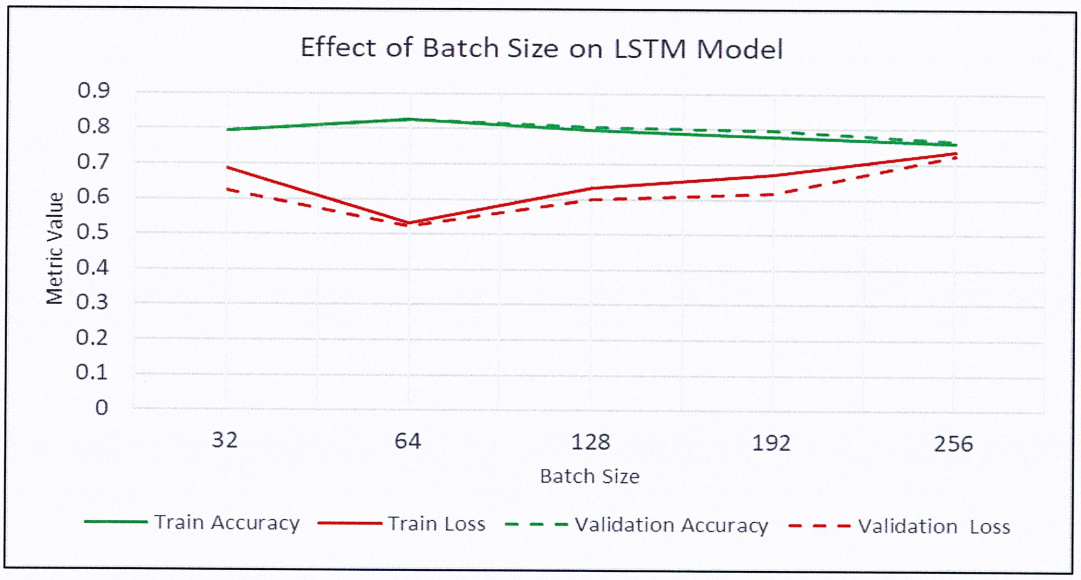


Figure 4.12: Effect of Batch Size on LSTM Model of Smartphone Sub-dataset

4.2.1.2. Model Performance Results

The LSTM model was trained based on the best-gained hyperparameters mentioned above. The result of model training, which is presented in Figure 4.13, shows that the achieved 91.40% training accuracy and 0.2648 as loss value. Meanwhile, the validation accuracy was 90.12%, and the loss value was 0.3183.

Although the difference between training and validation accuracies is minor, the difference in loss value might not be, especially since there is some fluctuation in the validation loss. Therefore, the model is to be evaluated on unseen data. The result of model testing has been compared with training ones to evaluate the training performance, as shown in Figure 4.14. While the testing accuracy was 90.73%, it is lower than training accuracy but also slightly higher than validation accuracy. The same situation happened to the testing loss value, which was 0.3019.

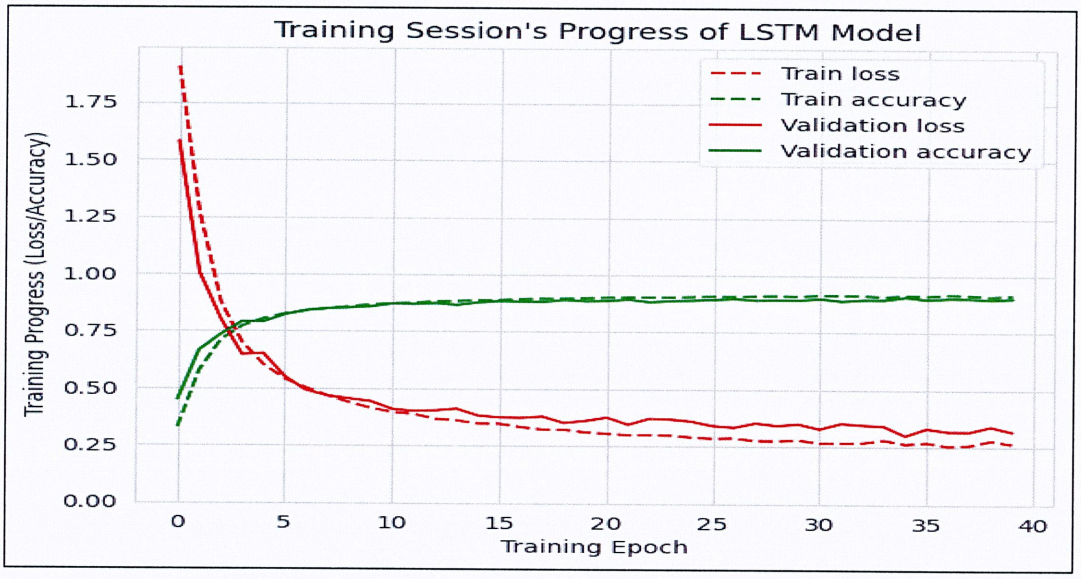


Figure 4.13: Result of Training LSTM Model Using Smartphone Sub-dataset

The result of model testing has been compared with training ones to evaluate the training performance, as shown in Figure 4.14. While the testing accuracy was 90.73%, it is lower than training accuracy but also slightly higher than validation accuracy. The same situation happened to the testing loss value, which was 0.3019.

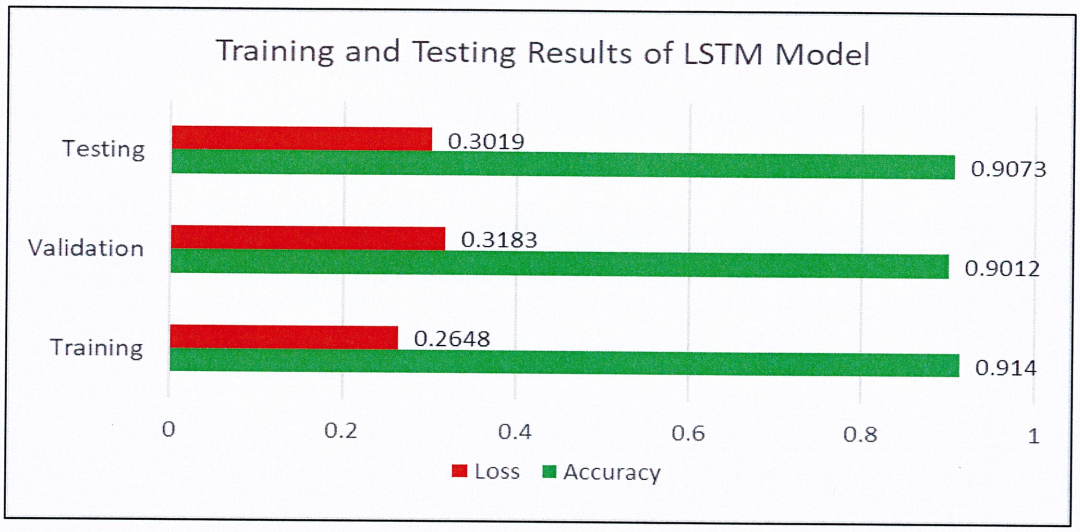


Figure 4.14: LSTM Model Performance using Smartphone Sub-dataset

However, the model's performance can be further evaluated by examining its predictions for each class through the confusion matrix to know how correct they are. The confusion matrix of the applied LSTM model can be found in Appendix A. It can be noted that the Eating Soup class was misclassified with Walking, Sitting, and Kicking classes more than the others, in addition to a few other high misclassifications.

This can be explained by the similarity in the measurements of the sensor axes while performing this activity with other activities during specific time windows. For example, the phone position in the Eating Soup class is similar to the Sitting class in some periods.

For more evaluation details about the model's performance on testing data, Table (4.1) gives a detailed report of classification performance. For each class, the scores of precision, recall, and F1-score evaluation metrics are provided. The support column is the total samples for each class in the test dataset.

Table (4.1): Performance of LSTM Model Across Each Class in Smartphone Sub-dataset

Class	Precision	Recall	F1-Score	Support
Walking	0.8558	0.858	0.8569	3001
Jogging	0.8946	0.878	0.8862	3172
Stairs	0.914	0.9195	0.9168	3330
Sitting	0.9076	0.8586	0.8825	3056
Standing	0.8869	0.8854	0.8861	3010
Typing	0.8946	0.8874	0.891	3260
Brushing Teeth	0.987	0.9731	0.98	3119
Eating Soup	0.7967	0.8744	0.8337	3002
Eating Chips	0.871	0.9155	0.8927	3362
Eating Pasta	0.8969	0.8671	0.8817	3302
Drinking from Cup	0.9458	0.9315	0.9386	3652
Eating Sandwich	0.8861	0.877	0.8815	3228
Kicking	0.9125	0.9093	0.9109	3198
Playing Catch	0.9123	0.9193	0.9158	3223
Dribbling (Basketball)	0.9123	0.9271	0.9196	3332
Writing	0.9385	0.9357	0.9371	3361
Clapping	0.9619	0.9632	0.9626	3070
Folding Clothes	0.9541	0.9383	0.9462	3503

The best-predicted class was Brushing Teeth, as it had the highest values across all metrics with 0.987 precision, 0.9731 recall, and 0.98 F1-score. This class is distinguished from others in the accelerometer measurements as seen in EDA results, making it easy to identify.

In contrast, the lowest values belonged to the Eating Soup class, with 0.7967, 0.8744, and 0.8337 for precision, recall, and F1-score, respectively. Due to the mentioned similarity above, the model has predicted this class well, but it is accompanied by false predictions also at the same time.

Also, it was noticed that half of the classes (nine classes) had scored above 0.9 (three of them above 0.95) in all metrics. Meanwhile, the rest of the classes scored below 0.9, all above 0.85, except the Eating Soup class, as mentioned above.

In conclusion, it seems that the LSTM has a good performance in general, despite the low values in some classes.

4.2.2. Performance of GCN Model

4.2.2.1. Hyperparameters Tuning

The process of hyperparameters tuning was done based on some settings recommended in the literature and following the same procedures as passed in the LSTM model. The tunned hyperparameters are layer size, batch size, and epoch size. Concurrently, the same limitation of the number of layers has been applied here also.

Regarding the impact of layer size, Figure 4.15 shows that a size of 256 has the highest accuracy rate and the lowest loss value even though 192 looks the same, but it has a higher validation loss value. On the other hand, the best size for batch size was 32, as presented in Figure 4.16, where the best accuracy and loss values are obtained, regardless of the semi-closeness result for a size of 64.

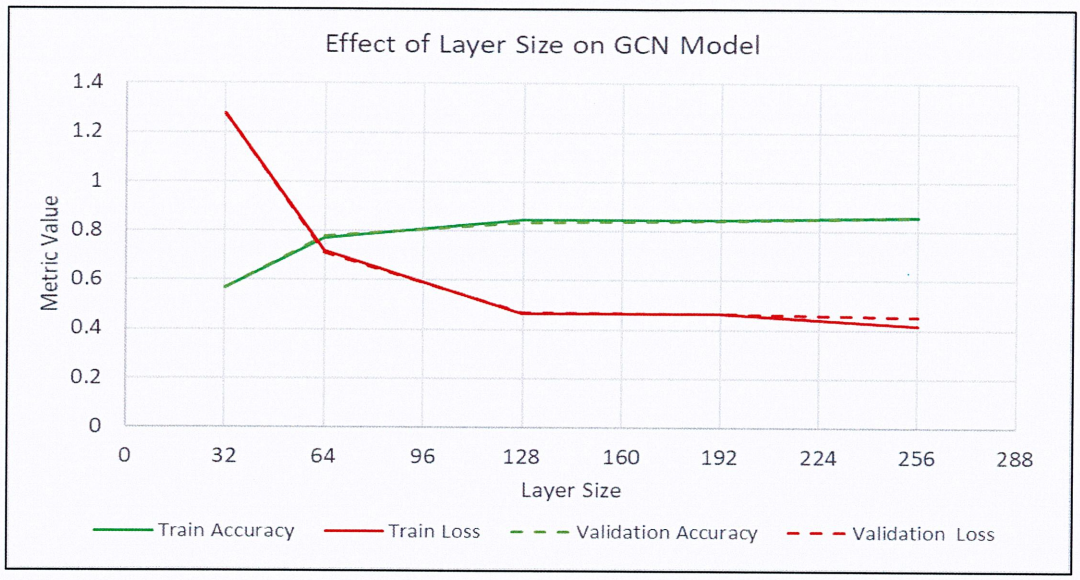


Figure 4.15: Effect of Layer Size on GCN Model of Smartphone Sub-dataset

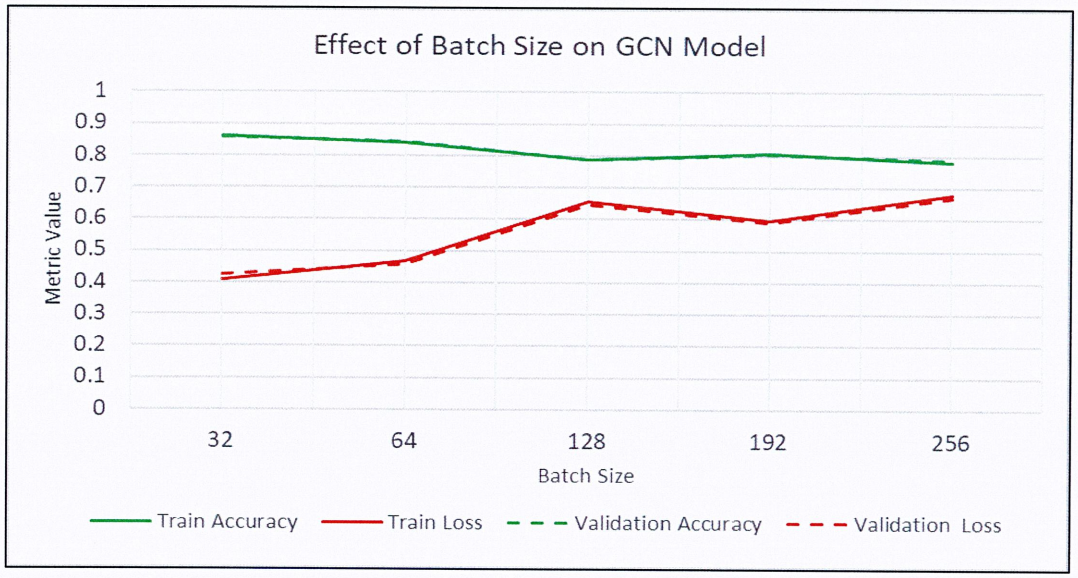


Figure 4.16: Effect of Batch Size on GCN Model of Smartphone Sub-dataset

4.2.2.2. Model Performance Results

The best-gained hyperparameters listed above were used to train the GCN model. The trained model has achieved high training scores where the accuracy and loss values were 95.16%, 0.1377 in sequence. However, the validation scores were slightly lower, as the accuracy was 94.28% and the loss was 0.175. The progress of training is shown in Figure 4.16.

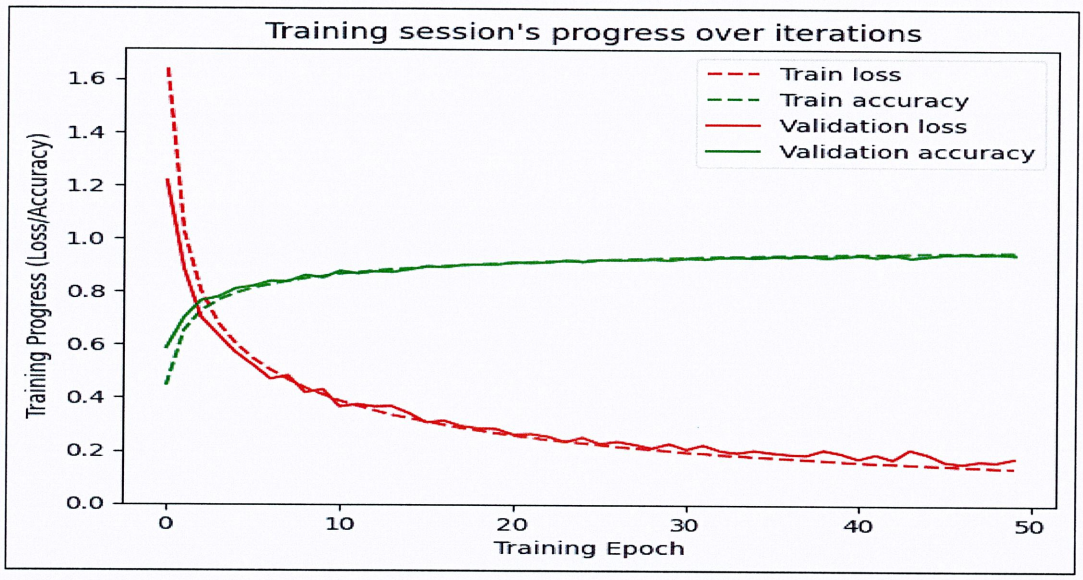


Figure 4.17: Result of Training GCN Model Using Smartphone Sub-dataset

Hence, as an extra verification step, the unseen testing data were passed to the GCN model to measure its performance. The outcome of the testing process was almost similar to validation result in favour of the testing as presented in Figure 4.18. The testing accuracy was 94.42% while the loss value was 0.1716.

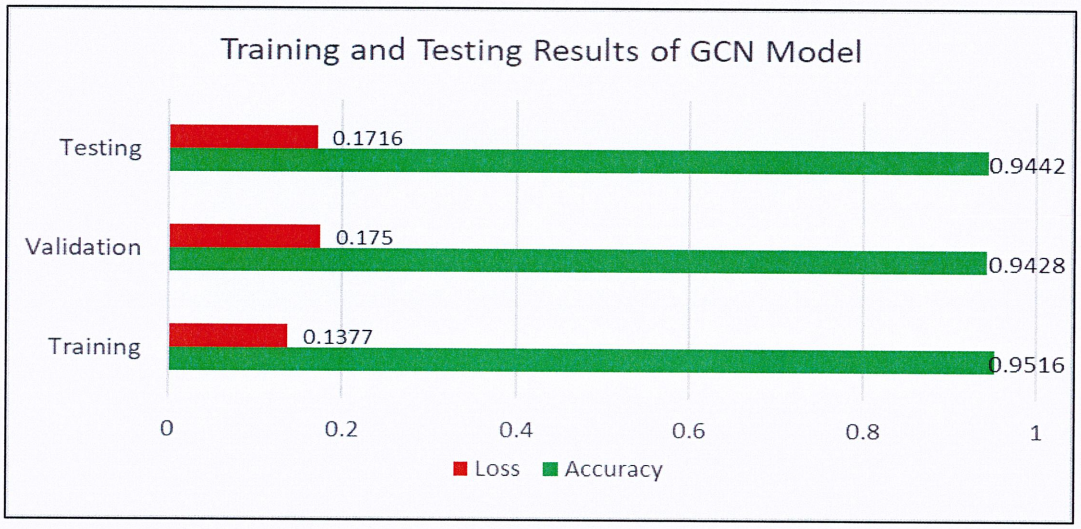


Figure 4.18: GCN Model Performance using Smartphone Sub-dataset

Furthermore, looking at the confusion matrix can help in getting more detailed prediction performance for each class. The resulted confusion matrix of the applied GCN model on Smartwatch Sub-dataset is presented in Appendix B.

Generally, the performance of prediction is good, but some classes have been assigned wrongly more than the others, the top three classes are Brushing Teeth, Eating Pasta and Folding cloth. Meanwhile, the most misclassified classes are Catching and Eating Chips; they have been predicted with the same nature classes. For example, the model predicted the Catching class mainly as kicking and dribbling.

However, to measure the performance in terms of evaluation metrics, Table (4.2) describes the prediction model's performance for each class. The provided metrics are precision, recall, and F1-score, while the support value is each class's sample count.

The Jogging class has the highest values with 0.9967 for precision, 0.9981 for recall, and 0.9891 for F1-score, while the overall lowest values belong to the Playing Catch class, as it has 0.8465 recall value and 0.8941 F1-score value despite its 0.9475 precision value. Considering F1-score only, all classes have above 0.9 values (six classes are above 0.95) except the Playing Catch as passed.

Table (4.2): Performance of GCN Model Across Each Class in Smartphone Sub-dataset

Class	Precision	Recall	F1-Score	Support
Walking	0.9876	0.9905	0.9891	3065
Jogging	0.9967	0.9881	0.9924	3032
Stairs	0.9596	0.9885	0.9738	3122
Sitting	0.9594	0.9519	0.9556	3574
Standing	0.9511	0.9363	0.9436	3362
Typing	0.9565	0.9766	0.9664	3284
Brushing Teeth	0.9048	0.9678	0.9353	3389
Eating Soup	0.9213	0.9179	0.9196	3240
Eating Chips	0.9469	0.8932	0.9193	3296
Eating Pasta	0.916	0.951	0.9332	3451
Drinking from Cup	0.9643	0.9259	0.9447	3065
Eating Sandwich	0.9363	0.924	0.9301	3277
Kicking	0.907	0.9492	0.9276	2990
Playing Catch	0.9475	0.8465	0.8941	2834
Dribbling (Basketball)	0.9168	0.9328	0.9248	2978
Writing	0.9633	0.9728	0.9681	3535
Clapping	0.9514	0.9287	0.9399	3351
Folding Clothes	0.9199	0.9436	0.9316	3336

Overall, it appears that the GCN model performs very good also. Hence, the next section is for the comparison with the LSTM model.

4.2.3. Analysis and Comparison

In order to accomplish the objectives of this research, this section is presenting the analysis and comparison of the performance for both models using the Smartphone sub-dataset.

The overall comparison between both models of the smartphone sub-dataset is demonstrated in Figure 4.19 in terms of evaluation metrics. It

clarifies that the GCN model outperforms the LSTM model across all metrics by a rate of approximately 4%, except the ROC Score, where the difference is about 0.0027. Indeed, the GCN model has benefited from the convolution performed by the GCN layer to capture the temporal and spatial dependencies in the data (Islam et al., 2022). It supported by large size of the layer (256 neurons) to get more features then to gain a better score than the LSTM model (128 neurons).

Conversely, the increased layer size led to a generalization gap for the LSTM model, even with the trials of minimization. Nevertheless, this came at the cost of model complexity, as the GCN model is more computationally expensive due to convolution phase where the size of input data (matrices) is affecting. Therefore, the training time of the LSTM model is nearly half of the needed time for the GCN model, although it can be fixed by using parallel computing to reduce computation time for convolution phase (Li, et al., 2023).

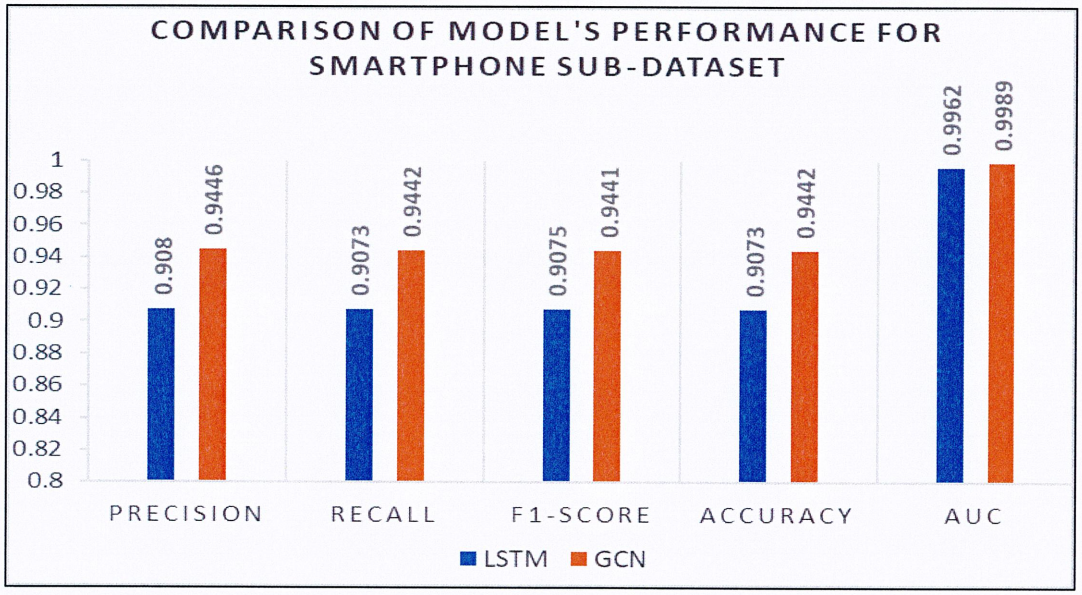


Figure 4.19: Overall Comparison Between LSTM and GCN Models

Since the classes have been grouped in the parent dataset into three groups, Table (4.3) is comparing the performance of both models for each group using the average score of evaluation metrics. The GCN model has made around 7% improvement in performance over the LSTM model in the Non-hand-oriented activities and Hand-oriented activities (Eating) groups. While with the Hand-oriented activities (General) group, there was a slight improvement in Recall & F1-score values.

Table (4.3): Group-wise Performance Comparison for LSTM and GCN Models

Group	Model	Precision	Recall	F1-Score	Accuracy
Non-hand-oriented activities	LSTM	0.8952	0.8848	0.8899	0.8952
	GCN	0.9602	0.9674	0.9637	0.9602
Hand-oriented activities (General)	LSTM	0.9372	0.9349	0.9360	0.9372
	GCN	0.9372	0.9384	0.9372	0.9372
Hand-oriented activities (Eating)	LSTM	0.8793	0.8931	0.8856	0.8793
	GCN	0.9370	0.9224	0.9294	0.9370

Moreover, the Non-hand-oriented activities group has been considered a subset dataset which was used in many of studies (Hou, 2020; Kolkar et al., 2022; Pienaar & Malekian, 2019; Sa-nguannarm et al., 2020). Therefore, the comparison of the results between the mentioned studies and the related subset proves the superiority of the GCN model over their LSTM models.

Similarly, the LSTM model of this research has been compared with the GCN model considering the Non-hand-oriented activities only. The results in Table (4.4) illustrate the superior performance of the GCN model in each class.

While the comparison with the GCN model in the study of Sarkar et al. (2021) indicates the exceed performance of the current GCN model. On the other hand, the added features of the gyroscope have improved the performance of the LSTM model with approximately 5% comparing to the LSTM model of Shiranthika et al. (2020).

Table (4.4): Class-wise Performance Comparison for LSTM and GCN Models

Group	Model	Precision	Recall	F1-Score
Walking	LSTM	0.8558	0.858	0.8569
	GCN	0.9876	0.9905	0.9891
Jogging	LSTM	0.8946	0.878	0.8862
	GCN	0.9967	0.9881	0.9924
Stairs	LSTM	0.914	0.9195	0.9168
	GCN	0.9596	0.9885	0.9738
Sitting	LSTM	0.9076	0.8586	0.8825
	GCN	0.9594	0.9519	0.9556
Standing	LSTM	0.8869	0.8854	0.8861
	GCN	0.9511	0.9363	0.9436
Kicking (Soccer Ball)	LSTM	0.9125	0.9093	0.9109
	GCN	0.907	0.9492	0.9276

In summary, considering the Smartphone sub-dataset, the GCN model proves its efficiency against the LSTM model of current research, in addition to other LSTM and GCN models in some previous studies.

4.3. Evaluation of Smartwatch Sub-dataset Models

This section will present the evaluation results of implemented models on the smartwatch sub-dataset. It will start with the results of hyperparameter tuning then model training, and model testing for each model, then models' comparison and discussion.

4.3.1. Performance of LSTM Model

4.3.1.1. Hyperparameters Tuning

Following the hyperparameters tuning procedure of the past LSTM model passed in section 4.2.1.1, the tuning process will focus on layers size, number of epochs, and batch size. As for the number of layers, it will be one method layer followed by two fully connected layers, where the second is the output layer with Softmax activation function.

As shown in Figure 4.20, the best value for layer size was 192, where at this point, the highest and lowest values for accuracy and loss were obtained at this size, respectively.

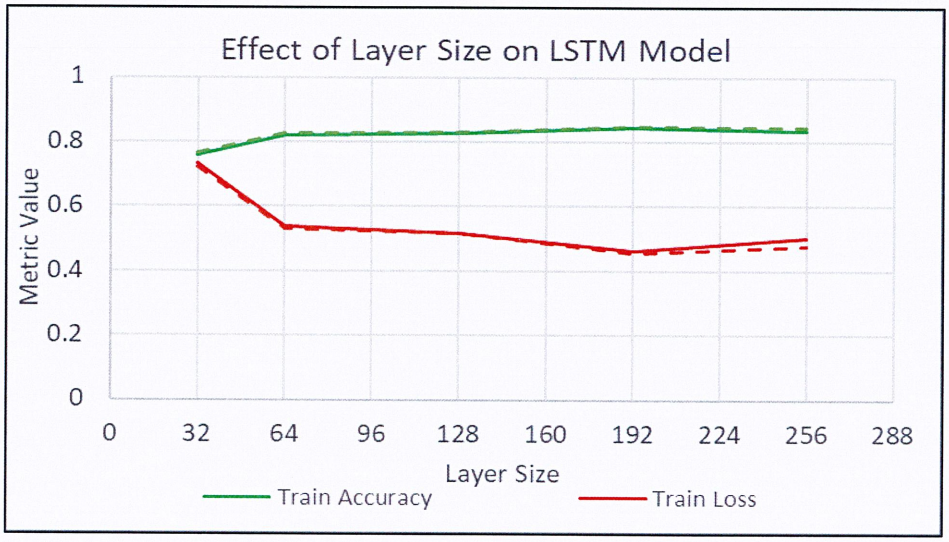


Figure 4.20: Effect of Layer Size on LSTM Model of Smartwatch Sub-dataset

While for batch size, the best batch size was also 192, as shown in Figure 4.21. there was no significant difference across the accuracy value, but the loss value was lowest at this size.

Regarding the epoch size, due to the early stopping method, the maximum size was limited to 50 to avoid overfitting in model training.

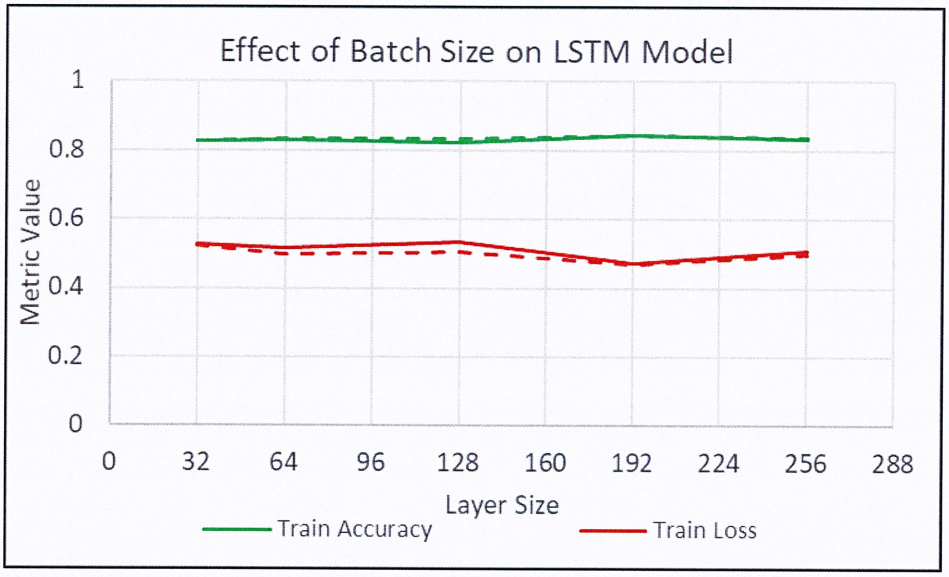


Figure 4.21: Effect of Batch Size on LSTM Model of Smartwatch Sub-dataset

4.3.1.2. Model Performance Results

The training process of the subject model has been done by following the tunned hyperparameters. As a result, the training accuracy was 93.01%, and the training loss was 0.2087. Meanwhile, the validation values were notably lower; for the accuracy, it was 89.94%, and for the loss, it was 0.3226. Figure 4.22 shows the training progress over many epochs.

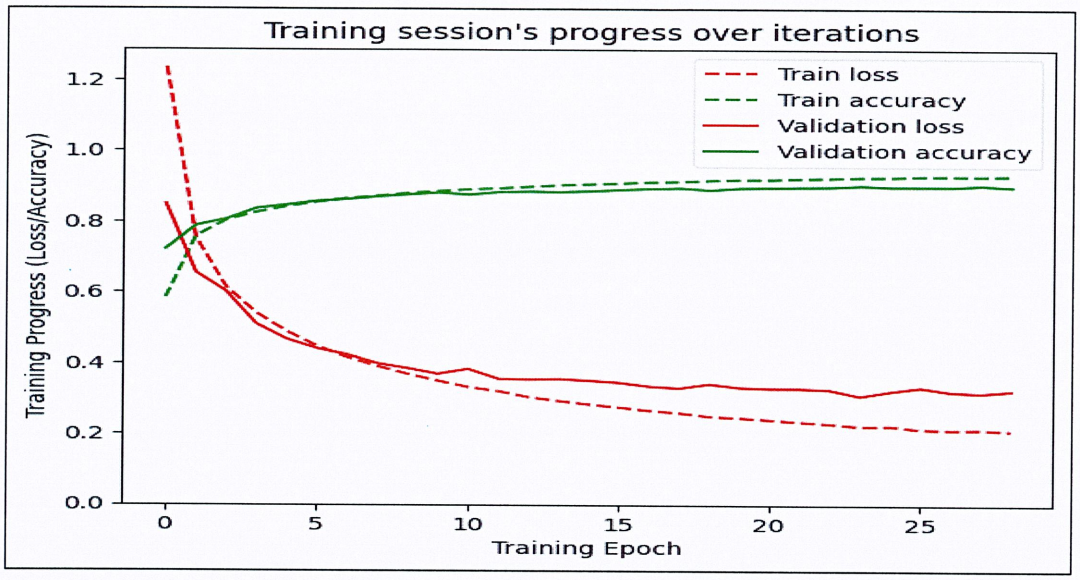


Figure 4.22: Result of Training LSTM Model Using Smartwatch Sub-dataset

In light of this, the model has been tested on new test data to verify the training results and to compare the prediction with actual classes. Hence, the testing process yielded 90.15% accuracy and 0.3192 loss scores, almost the same as the validation scores. Figure 4.23 shows the comparison between the three results.

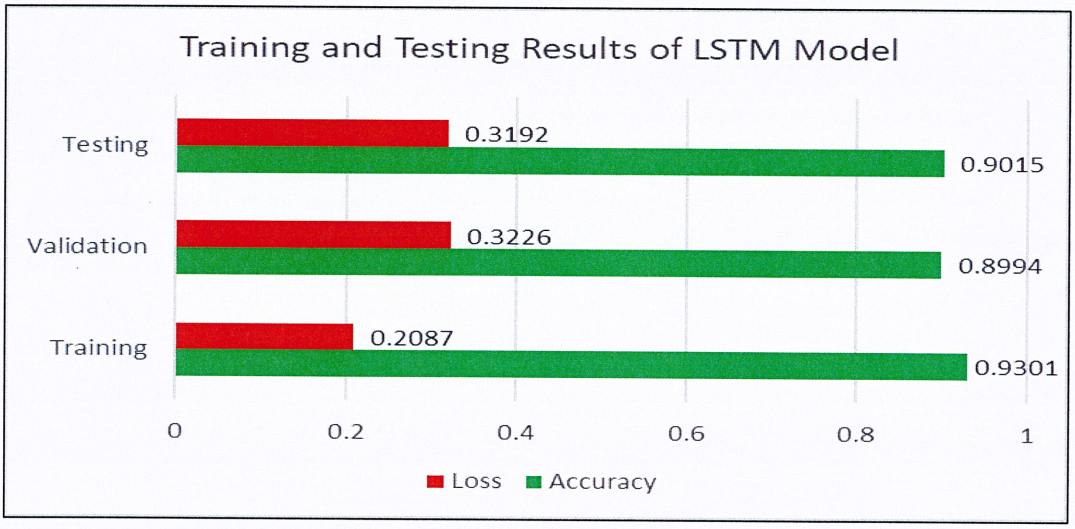


Figure 4.23: LSTM Model Performance using Smartwatch Sub-dataset

Besides the above results, the assessment of the model could be more transparent with the confusion matrix, as it measures the performance of model prediction for each class. Referring to the confusion matrix of the LSTM model available in Appendix C, the model has predicted some classes very well but not for other classes. For example, it poorly classified the class of Eating Pasta instead of the actual one with a high number of times in many cases, such as 474 times in place of Jogging class. It can be referred to some similarity in hand movement at certain window time. Therefore, the top three classes that were assigned wrongly were Eating Pasta, Eating Chips, and Jogging with 1251, 882, and 863 times.

Table (4.5) shows a class-wise assessment of model prediction based on the confusion matrix results in terms of precision, recall, and F1-score metrics. Brushing Teeth was the best-predicted class, achieving the highest scores across all metrics, 0.9933 for precision, 0.9791 for recall, and 0.9861 for F1-score. Similarly, the Stairs, Sitting, and Dribbling classes have also scored above 0.97 in all metrics. The total number of classes that have above 0.90 scores is nine.

On the other hand, the worst predicted class was Eating Pasta, with the lowest values of 0.6715, 0.7015, and 0.6862 for precision, recall, and F1-score, respectively. It has likely mixed hand movements that are similar to other classes. At the same time, the following classes were Jogging and Eating Chips, where they joined a group of 9 classes with below 0.90 metric scores. By considering F1-score only, more than half of the classes have above 0.90; meanwhile, the reset is above 0.8 expect two classes that were below.

Table (4.5): Performance of LSTM Model Across Each Class in Smartwatch Sub-dataset

Class	Precision	Recall	F1-Score	Support
Walking	0.9229	0.9486	0.9356	3696
Jogging	0.7471	0.6906	0.7177	3691
Stairs	0.9884	0.983	0.9857	3712
Sitting	0.9732	0.9605	0.9668	3745
Standing	0.8588	0.8261	0.8422	3917
Typing	0.8764	0.9389	0.9065	3745
Brushing Teeth	0.9933	0.9791	0.9861	3634
Eating Soup	0.8945	0.9132	0.9037	3789
Eating Chips	0.7906	0.8782	0.8321	3793
Eating Pasta	0.6715	0.7015	0.6862	3645
Drinking from Cup	0.8714	0.829	0.8497	3849
Eating Sandwich	0.9009	0.8484	0.8739	3838
Kicking	0.9212	0.8324	0.8745	3651
Playing Catch	0.9073	0.9304	0.9187	3965
Dribbling (Basketball)	0.9756	0.9654	0.9705	3761
Writing	0.929	0.9561	0.9424	3488
Clapping	0.9123	0.9291	0.9206	3750
Folding Clothes	0.941	0.9528	0.9468	3747

In light of above results, the LSTM model has predicting issue with certain group of classes (Eating classes) more than the other classes. However, the comparison should give more details about that.

4.3.2. Performance of GCN Model

4.3.2.1. Hyperparameters Tuning

Following the previous hyperparameters tunning sections, the layer size, the number of epochs, and the batch size will be the hyperparameters that are the subject of tunning process. Meanwhile the number of layers limited to two layers as mentioned earlier.

The impact of layer size is shown in Figure 4.24 where the best size was 256. It is determined by taking the lowest loss value due to the unclarity of the best accuracy value with respect to the layer size. On the other hand, in Figure 4.25, the best batch size was 32 where the loss value was the significance key value. The closeness of accuracy value did not help in selecting the best batch size. In order to prevent overfitting in model training, the maximum epoch size was set at 50 due to the early stopping approach.

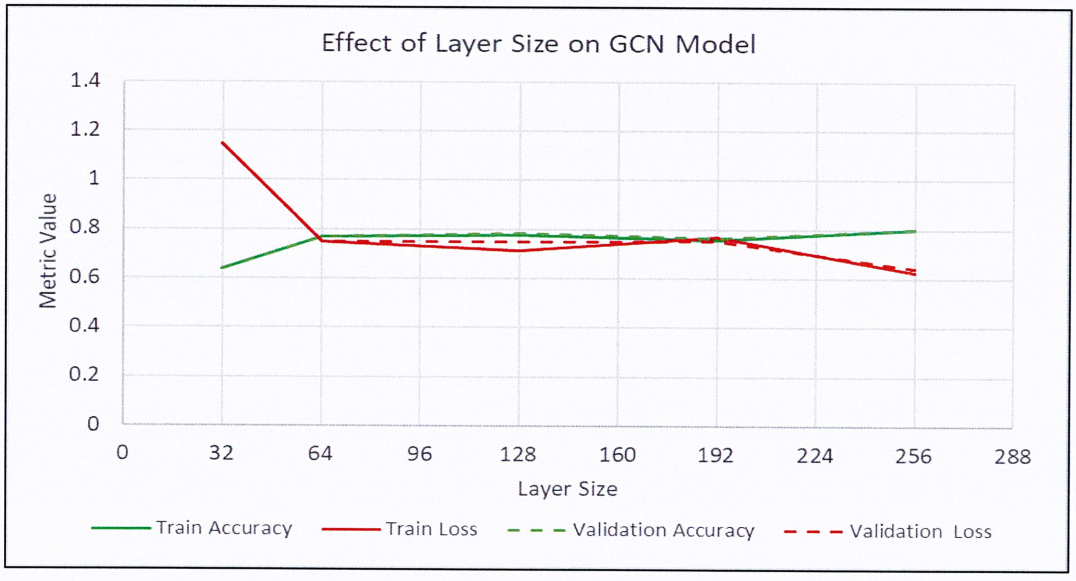


Figure 4.24: Effect of Layer Size on GCN Model of Smartwatch Sub-dataset

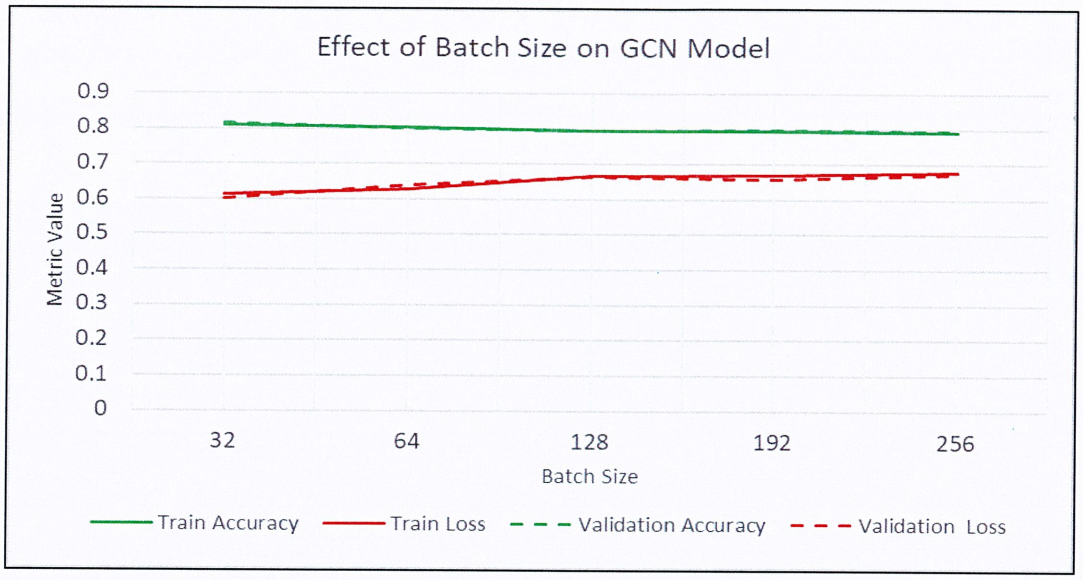


Figure 4.25: Effect of Batch Size on GCN Model of Smartwatch Sub-dataset

4.3.2.2. Model Performance Results

The progress of training process of the GCN model is presented in Figure 4.26. the result of training in terms of accuracy and loss values were 90.54% and 0.2848, respectively. In parallel, the validation scores were 90.50% and 0.3032 for accuracy and loss values, in sequence.

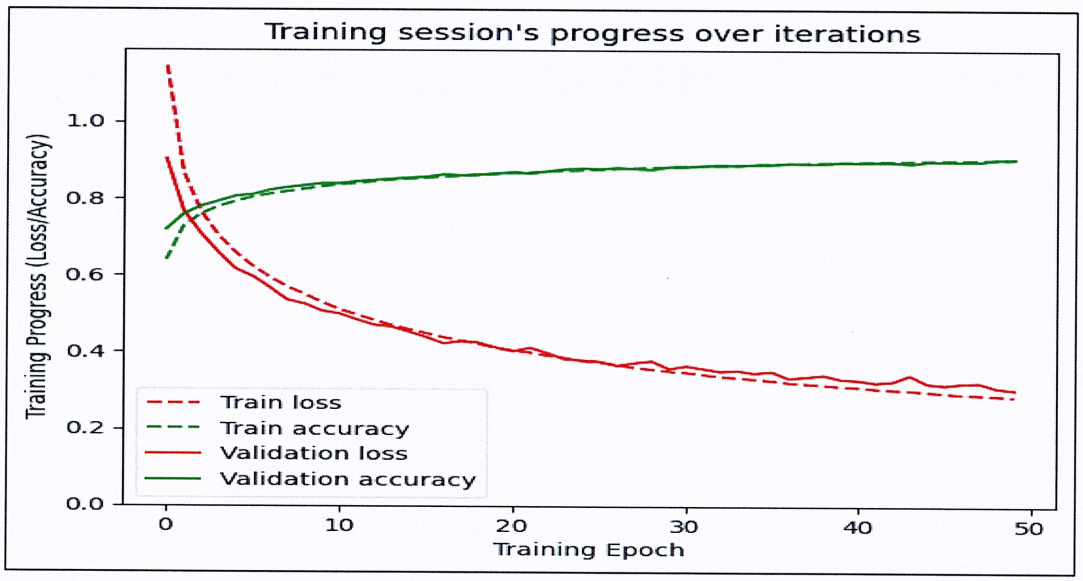


Figure 4.26: Result of Training GCN Model Using Smartwatch Sub-dataset

The closeness of training and validation scores need to be verified by testing the model on other unseen data. Accordingly, the testing step yielded almost similar results to the training and validation results. the testing scores were slightly lower with 90.40% accuracy and 0.313 loss value, as shown in Figure 4.27.

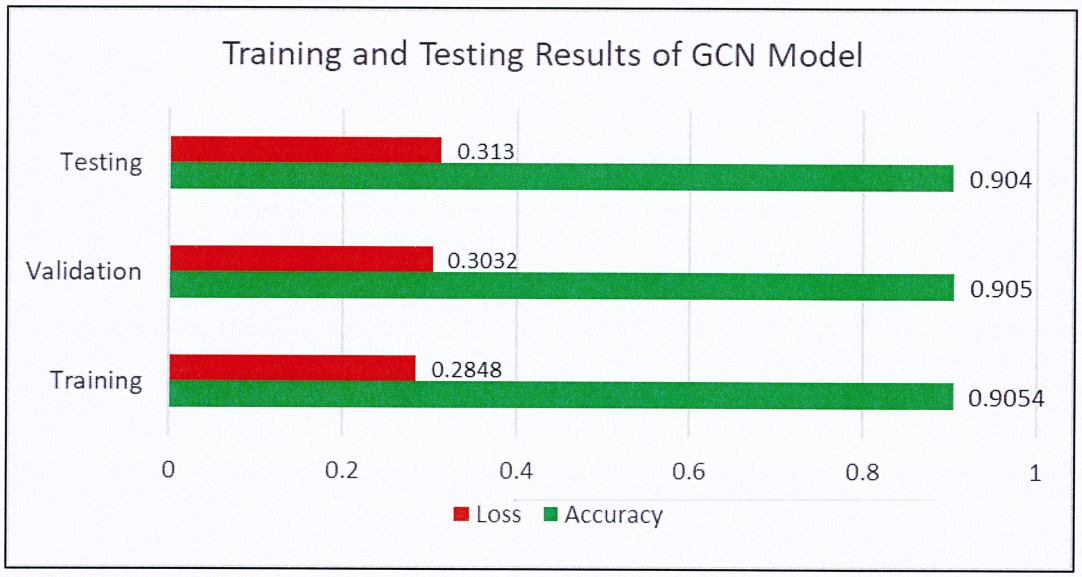


Figure 4.27: GCN Model Performance using Smartwatch Sub-dataset

However, the exploration of the model's predictions gives a better assessment of the classification performance of the model. Therefore, the confusion matrix can show the efficiency of each class classification. The confusion matrix of the subject model can be found in Appendix D. The produced matrix indicates a classification issue with some classes; for example, the class Eating Sandwich is the most misclassified class with 975 times, then the class Eating Chips with 853 times.

In the same way, the above classes are the most wrongly assigned classes in addition to the Eating Pasta class. Therefore, It can be seen that the model does not perform well with the group of Eating classes, as it is

likely that the movement of the hand associated with the measuring device is shared in a certain way in each class of this group.

Based on the above, the evaluation metrics can represent the prediction performance measurably. Hence, Table (4.6) shows the prediction performance of each class in terms of Precision, Recall, and F1-score values. As a result, the highest values refer back to the Jogging Class, which is the best-predicted class; it has values of 0.9780, 0.9890, and 0.9835 for precision, recall, and F1-score in sequence.

Table (4.6): Performance of GCN Model Across Each Class in Smartwatch Sub-dataset

Class	Precision	Recall	F1-Score	Support
Walking	0.9070	0.9548	0.9303	3716
Jogging	**0.9780**	**0.9890**	**0.9835**	**3643**
Stairs	0.9376	0.8652	0.9000	3666
Sitting	0.8444	0.8457	0.8451	3928
Standing	0.8912	0.9406	0.9153	3807
Typing	0.8863	0.9653	0.9241	3601
Brushing Teeth	0.9313	0.9453	0.9383	3714
Eating Soup	0.9166	0.8785	0.8972	3729
Eating Chips	0.8146	0.7708	0.7921	3722
Eating Pasta	0.8270	0.8668	0.8465	3740
Drinking from Cup	0.8807	0.8318	0.8556	3913
Eating Sandwich	**0.7697**	**0.7345**	**0.7517**	**3672**
Kicking	0.8966	0.9167	0.9066	3662
Playing Catch	0.9577	0.9538	0.9558	3726
Dribbling (Basketball)	0.9773	0.9773	0.9773	3870
Writing	0.9396	0.9254	0.9324	3832
Clapping	0.9794	0.9816	0.9805	3689
Folding Clothes	0.9347	0.9334	0.9341	3786

On the other hand, the Eating Sandwich class was the worst; it has 0.7697 for precision, 0.7345 for recall, and 0.7517 for f1-score. Regarding F1-score values only, there are 12 classes above 0.95 and six classes below 0.90 (two classes below 0.80).

4.3.3. Analysis and Comparison

This section presents the performance analysis and comparison for both models utilizing the Smartphone sub-dataset to meet the objectives of this research.

The comparison of overall performance for the LSTM and the GCN models indicates a slight advantage for the GCN, as presented in Figure 4.28. However, The GCN model did not benefit much from its large layer size (256 neurons) compared to the LSTM's layer size (128 neurons). It is opposite to the Smartphone sub-dataset result. Indeed, this came at the cost of model complexity and thus was more time-consuming in model training.

Although the additive performance seems minor, which does not exceed 0.1, it is still noticeable as the scores have jumped to the nineties. Therefore, the GCN model has minimized the generalization gap better than the LSTM model by comparing the passed Figures 4.22 and 4.26.

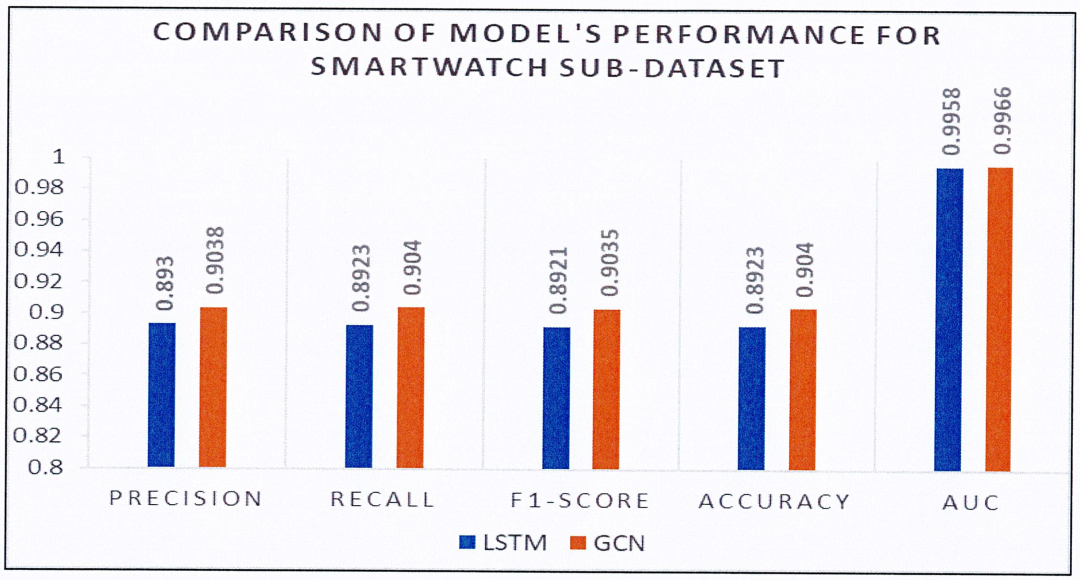


Figure 4.28: Overall Comparison Between LSTM and GCN Models

However, the inspection of performance using the axis of grouped classes can give more information about the gained improvement. Hence, a group-wise comparison between both models is demonstrated in Table (4.7). The GCN model has boosted the correct predictions in the group of the Non-hand-oriented activities, as the recall score raised 0.9187 from 0.8735. In addition, the enhancement in the Hand-oriented General activities group across all metrics was small but considerable.

Table (4.7): Group-wise Performance Comparison for LSTM and GCN Models

Group	Model	Precision	Recall	F1-Score	Accuracy
Non-hand-oriented activities	LSTM	0.9019	0.8735	0.8871	0.9019
	GCN	0.9091	0.9187	0.9135	0.9091
Hand-oriented activities (General)	LSTM	0.9336	0.9503	0.9417	0.9336
	GCN	0.9438	0.9546	0.9489	0.9438
Hand-oriented activities (Eating)	LSTM	0.8258	0.8341	0.8291	0.8258
	GCN	0.8417	0.8165	0.8286	0.8417

In contrast, the group of Hand-oriented Eating activities had conflicting results, where the scores of precision and accuracy increased, while the scores of recall and F1-score decreased. This conflict means that the model has been less sensitive to certain classes, leading to lower wrong predictions (false positives) in the cost of missing the correct ones (true positives).

To address this issue, the class-wise comparison in Table (4.8) shows that the model has a classification problem with the Eating Sandwich class, as it has a significant drop in all metrics. In addition, the drop in recall scores for the classes Eating Soup and Eating Chips might support that also. Nevertheless, the GCN model performs well but has some issues with the group of Hand-oriented Eating activities.

Table (4.8): Class-wise Performance Comparison for LSTM and GCN Models

Group	Model	Precision	Recall	F1-Score
Eating Soup	LSTM	0.8945	0.9132	0.9037
	GCN	0.9166	0.8785	0.8972
Eating Chips	LSTM	0.7906	0.8782	0.8321
	GCN	0.8146	0.7708	0.7921
Eating Pasta	LSTM	0.6715	0.7015	0.6862
	GCN	0.8270	0.8668	0.8465
Drinking from Cup	LSTM	0.8714	0.8290	0.8497
	GCN	0.8807	0.8318	0.8556
Eating Sandwich	LSTM	0.9009	0.8484	0.8739
	GCN	0.7697	0.7345	0.7517

However, both models have struggled with this group, which might belong to noisy data in these classes as per this study (Mekruksavanich et al., 2022). Therefore, the study has applied an LSTM model on this group's data with noise removal to achieve an above 93% accuracy score.

Finally, the performance of the LSTM model compared to the model by Mekruksavanich et al. (2020) is slightly lower by around 0.4 in each metric. Indeed, the mentioned study has cut off some data due to the small sample size. However, despite that, it proves the improved performance of the GCN model over both LSTM models.

Chapter Five

Conclusion and Recommendation

5.0. Overview

5.1. Research Summary

5.1. Research Limitations

5.3. Recommendations and Future work

Chapter Five
Conclusion and Recommendation

5.0. Overview

This chapter deals with the summarization of the research. It starts with a brief research summary, including the objectives, methods, models, and results. Then it goes on to highlight the limitations of the research. Finally, it provides some recommendations and future work directions.

5.1. Research Summary

This research aims to compare the performance of the LSTM and GCN methods in classifying and recognizing human activities. The comparison was made on the WISDM dataset, a time series dataset containing 18 daily human activities.

The interest of humans in the daily performed activities was through the observation and analysis by their power. Due to the recent development of technology in computer hardware, the task of analyzing sensor data to classify those activities is called Human Activity Recognition. Therefore, The importance of this field came from its power in problem-solving and information extraction. In addition, it has a wide range of the applications, such as healthcare, security, fitness, and human-computer interaction.

However, the human activity recognition systems have been categorized based on the data collection tools into Vision-based, sensor-

based, and time geography-based systems. Despite the variation, all the systems share the same stages, including raw signal acquisition, data preparation, Model training, and activity recognition. Hence, AI has a role in model training through ML and DL methods.

Using traditional ML algorithms has achieved high-performance scores in human activity recognition. However, despite that, it came with some issues, including the manual feature extraction from the raw data and the suffering in the time-series data processing. Therefore, DL methods were introduced to overcome these issues.

Among the DL methods, the most common were Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) methods. While the RNN method cannot handle long sequenced data, the LSTM method has proved its efficiency in dealing with it, supported by its cell structure. On the other hand, the CNN method can extract the most complex features but not with the topological structure that contain spatial and temporal dependencies. Hence, the GCN could extract the spatiotemporal relations of time series data using the graph data structure.

Although many studies have explored the performance of the LSTM and GCN methods in human activity recognition using the WISDM dataset, they were limited to some activities of the dataset. In addition, there was no comparison between both methods. Therefore, this research was conducted to fill this gap of knowledge.

The dataset was divided into two sub-datasets based on the collection device (smartphone and smartwatch). For each sub-dataset, the data of the accelerometer and gyroscope sensors were combined using the timestamp

values. Then, the sub-datasets were pre-processed and prepared for model feeding using the overlapped sliding windows. Therefore, for each sub-dataset, two trained models (LSTM and GCN) with best-tunned hyperparameters existed.

As a result, the performance comparison between the models of the smartphone sub-dataset stated that the GCN model outperforms the LSTM model by around 4%. Furthermore, the GCN model scored 94.4% accuracy and F1-score, while the LSTM model had 90.7% accuracy and 90.8% F1-score. In addition, the GCN model performed better than some previous LSTM models in the literature review.

On the other hand, the results of the smartwatch sub-dataset gave a slight advantage to the GCN model over the LSTM model by nearly 1%. The accuracy and F1-scores for the GCN model were 90.4%; meanwhile, it was 89.2% for the LSTM model. Thus, both models have struggled with eating activities where their performance was conflicted.

In conclusion, the GCN method showed a better and competitive and convergent performance in both sub-datasets comparing to the LSTM method, where It had a notable outperformance in the smartphone sub-dataset. Therefore, it can be considered as a rival method for the LSTM method in such data type with taking the training time in consideration.

5.2. Research Limitations

The conducted research was seeking human activity recognition using the WISDM dataset only, so the obtained results are limited to this dataset and the applied methodology. Moreover, the utilized classification level for the GCN method is the graph level, and other levels might have different results. In addition, the cross validation was not considered in training and testing the models due to time limitation. Finally, the model structure for both methods was limited to have three layers as described formerly; hence the other types of structures were not tested.

Meanwhile, the primary obstruction in implementing this research was the need for a fully supported Python library that can deal with graph data in the TensorFlow framework.

5.3. Recommendations and Future Work

Based on research results, this research recommends expanding the utilization of the GCN methods in more similar datasets, including the non-controlled or lab productions, to gain more generalization. In addition, it is also recommended to include the GRU method in the comparison as well as to add the timing aspect of each method with the cross-validation technique in such a comparison.

The result of implemented models in this research can be enhanced for future work using other optimization techniques. In addition, the comparison should be more focused, where every group of activities can be a standalone sub-dataset.

References

Abdul Lateef Haroon P., S., Rashmi, P., & Supriya M., C. (2022). Human Activity Recognition Using Significant Skeletal Joints. *International Journal of e-Collaboration, 18(1)*, 1–16. https://doi.org/10.4018/ijec.304377

Abidine, M. B., & Fergani, B. (2022). An Efficient Kernel KNN classifier for Activity Recognition on Smartphone. *IEEE 21st international Ccnference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, Sousse, Tunisia, 73-78. IEEE https://doi.org/10.1109/sta56120.2022.10019099

Ahmad, T., Jin, L., Zhang, X., Lai, S., Tang, G., & Lin, L. (2021). Graph Convolutional Neural Network for Human Action Recognition: A Comprehensive Survey. *IEEE Transactions on Artificial Intelligence, 2(2)*, 128–145. https://doi.org/10.1109/tai.2021.3076974

Alanazi, M., Aldahr, R. S., & Ilyas, M. (2022). Effectiveness of Machine Learning on Human Activity Recognition Using Accelerometer and Gyroscope Sensors: A Survey. In N. Callaos, E. Gaile-Sarkane, S. Hashimoto, B. Sánchez (Eds.), *Proceedings of the 26th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI 2022*, 2(2022),1-5. International Institute of Informatics and Cybernetics https://doi.org/10.54808/WMSCI2022.02.1

Alexan, A., Alexan, A., & Oniga, S. (2022). Smart Watch Activity Recognition Using Plot Image Analysis. *IEEE 2nd Conference on Information Technology and Data Science (CITDS)*, Debrecen, Hungary, 1-6. IEEE. https://doi.org/10.1109/citds54976.2022.9914230

Ali, S. E., Khan, A. N., Zia, S., & Mukhtar, M. (2020). Human Activity Recognition System using Smart Phone based Accelerometer and Machine Learning. *IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*, Bali, Indonesia, 69-74. IEEE. https://doi.org/10.1109/iaict50021.2020.9172037

Al-Taei, A., Ibrahim, M. F., & Habeeb, N. J. (2021). Optimizing the Performance of KNN Classifier for Human Activity Recognition. In *Communications in Computer and Information Science (ICACDS 2021), 1440*(2021), 373–385. Springer International Publishing, https://doi.org/10.1007/978-3-030-81462-5_34

Ambati, L. S., & El-Gayar, O. (2021). Human Activity Recognition: A Comparison of Machine Learning Approaches. *Journal of the Midwest Association for Information Systems (JMWAIS), 2021*(1), 48-60.

Amer, M. M. (2021). Smartphone and Smartwatch for Human Activity Recognition. *Annals of Advanced Biomedical Sciences, 4*(1), 1-11. https://doi.org/10.23880/aabsc-16000159

Athota, R. K., & Sumathi, D. (2022). Human Activity Recognition Based On Hybrid Learning Algorithm For Wearable Sensor Data. *Measurement: Sensors, 24*(2022), 1-13. https://doi.org/10.1016/j.measen.2022.100512

Baroud, S., Chokri, S., Belhaous, S., & Mestari, M. (2021). A Brief Review of Graph Convolutional Neural Network Based Learning for classifying remote sensing images. *Procedia Computer Science,* 191(2021), 349–354. https://doi.org/10.1016/j.procs.2021.07.047

Benhaili, Z., Kabbaj, I., Balouki, Y., & Moumoun, L. (2022). Human Activity Recognition Using Stacked LSTM. In *Advances in Information, Communication and Cybersecurity,* 357(2022) , 33–42.

Springer International Publishing. https://doi.org/10.1007/978-3-030-91738-8_4

Bhambri, P., Kaur, H., Gupta, A., & Singh, J. (2021). Human Activity Recognition System. *Oriental journal of computer science and technology, 13*(2-3), 91–96. https://doi.org/10.13005/ojcst13.0203.05

Bian, S., Liu, M., Zhou, B., & Lukowicz, P. (2022). The State-of-the-Art Sensing Techniques in Human Activity Recognition: A Survey. *Sensors,* 22(2), 4596. https://doi.org/10.3390/s22124596

Bloemheuvel, S., van den Hoogen, J., Jozinović, D., Michelini, A., & Atzmueller, M. (2022). Graph Neural Networks for Multivariate Time Series Regression with Application to Seismic Data. *International Journal of Data Science and Analytics,*1-16. https://doi.org/10.1007/s41060-022-00349-6

Bozkurt, F. (2021). A Comparative Study on Classifying Human Activities Using Classical Machine and Deep Learning Methods. *Arabian Journal for Science and Engineering, 47*(2), 1507–1521. https://doi.org/10.1007/s13369-021-06008-5

Challa, S. K., Kumar, A., & Semwal, V. B. (2021, August). A Multibranch CNN-BiLSTM Model for Human Activity Recognition Using Wearable Sensor Data. *The Visual Computer, 38(12)*, 4095–4109. https://doi.org/10.1007/s00371-021-02283-3

Dang, L. M., Min, K., Wang, H., Piran, M. J., Lee, C. H., & Moon, H. (2020). Sensor-Based and Vision-based Human Activity Recognition: A Comprehensive Survey. *Pattern Recognition, 108*(2020), 1-24. https://doi.org/10.1016/j.patcog.2020.107561

Dehghani, A., Sarbishei, O., Glatard, T., & Shihab, E. (2019). A Quantitative Comparison of Overlapping and Non-Overlapping

Sliding Windows for Human Activity Recognition Using Inertial Sensors. *Sensors, 19*(22), 1-19. https://doi.org/10.3390/s19225026

Dhammi, L., & Tewari, P. (2021). Classification Of Human Activities Using Data Captured Through A Smartphone Using Deep Learning Techniques. *3rd International Conference on Signal Processing and Communication (ICPSC)*, Coimbatore, India, 689-694. IEEE. https://doi.org/10.1109/icspc51351.2021.9451772

Edemekong, P. F., Bomgaars, D. L., Sukumaran, S., & Schoo, C. (2022). *Activities of Daily Living.* StatPearls Publishing, Treasure Island (FL). http://europepmc.org/books/NBK470404

Fan, H., Jiang, M., Xu, L., Zhu, H., Cheng, J., & Jiang, J. (2020). Comparison of Long Short Term Memory Networks and the Hydrological Model in Runoff Simulation. *Water, 12*(1), 175-190. https://doi.org/10.3390/w12010175

Fra, V., Forno, E., Pignari, R., Stewart, T. C., Macii, E., & Urgese, G. (2022). Human Activity Recognition: Suitability of A Neuromorphic Approach For On-Edge Aiot Applications. *Neuromorphic Computing and Engineering, 2*(2022), 1-15. https://doi.org/10.1088/2634-4386/ac4c38

Gajjala, K. S., & Chakraborty, B. (2021). Human Activity Recognition based on LSTM Neural Network Optimized by PSO Algorithm. *IEEE 4th International Conference on Knowledge Innovation and Invention (ICKII)*, Taichung, Taiwan, 128-133. IEEE https://doi.org/10.1109/ickii51822.2021.9574788

Grattarola, D., & Alippi, C. (2020). Graph Neural Networks in TensorFlow and Keras with Spektral. *Graph Neural Networks in TensorFlow and Keras with Spektral.* arXiv. https://doi.org/10.48550/ARXIV.2006.12138

Gupta, N., Gupta, S. K., Pathak, R. K., Jain, V., Rashidi, P., & Suri, J. S. (2022). Human Activity Recognition In Artificial Intelligence Framework: A Narrative Review. *Artificial Intelligence Review, 55*(6) , 4755–4808. https://doi.org/10.1007/s10462-021-10116-x

Gupta, S. (2021). Deep Learning Based Human Activity Recognition (HAR) Using Wearable Sensor Data. *International Journal of Information Management Data Insights, 1*(2), 1-18. https://doi.org/10.1016/j.jjimei.2021.100046

Hayat, A., Morgado-Dias, F., Bhuyan, B., & Tomar, R. (2022). Human Activity Recognition for Elderly People Using Machine and Deep Learning Approaches. *Information, 13(6)*, 275-288. https://doi.org/10.3390/info13060275

Hossin, M., & Sulaiman, M. N. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process, 5*(2), 1–11. https://doi.org/10.5121/ijdkp.2015.5201

Host, K., & Ivašić-Kos, M. (2022). An overview of Human Action Recognition in Sports Based on Computer Vision. *Heliyon, 8*(6), 1-25. https://doi.org/10.1016/j.heliyon.2022.e09633

Hou, C. (2020). A study on IMU-Based Human Activity Recognition Using Deep Learning and Traditional Machine Learning. *5th International Conference on Computer and Communication Systems (ICCCS),* Shanghai, China, 225-234. IEEE. https://doi.org/10.1109/icccs49078.2020.9118506

Ige, A. O., & Noor, M. H. (2022). A Survey on Unsupervised Learning for Wearable Sensor-Based Activity Recognition. *Applied Soft Computing, 127*(2022), 1-22. https://doi.org/10.1016/j.asoc.2022.109363

Ingolfsson, T. M. (2021). Insights Into LSTM Architecture | Thorir Mar Ingolfsson. *Insights into LSTM architecture | Thorir Mar Ingolfsson.*

Great Learning (2023). Introduction to Pattern Recognition in Machine Learning. *Introduction to Pattern Recognition in Machine Learning.* https://www.mygreatlearning.com/blog/pattern-recognition-machine-learning/

Isakava, T. (2022). Human Activity Recognition: Everything You Should Know. *Human Activity Recognition: Everything You Should Know.* Retrieved January 4, 2023, from https://indatalabs.com/blog/human-activity-recognition

Islam, M. M., Nooruddin, S., Karray, F., & Muhammad, G. (2022). Human Activity Recognition Using Tools of Convolutional Neural Networks: A State of The Art Review, Data Sets, Challenges, And Future Prospects. *Computers in Biology and Medicine*, *149*(2022), 1-20. https://doi.org/10.1016/j.compbiomed.2022.106060

Jaberi, M., & Ravanmehr, R. (2022). Human Activity Recognition Via Wearable Devices Using Enhanced Ternary Weight Convolutional Neural Network. *Pervasive and Mobile Computing*, *83*(2022),1-18 . https://doi.org/10.1016/j.pmcj.2022.101620

Kipf, T. N., & Welling, M. (2016). Semi-Supervised Classification with Graph Convolutional Networks. *arXiv preprint arXiv:1609.02907.*

Kolkar, R., Tomar, R. P., & Vasantha, G. (2022). IoT-based Human Activity Recognition Models based on CNN, LSTM and GRU. *IEEE Silchar Subsection Conference (SILCON),* Silchar, India, 1-7. IEEE. https://doi.org/10.1109/silcon55242.2022.10028803

Kushwaha, A., Khare, A., & Khare, M. (2021). Human Activity Recognition Algorithm in Video Sequences Based on Integration of Magnitude and Orientation Information of Optical Flow.

International Journal of Image and Graphics, *22*(1), 1-19. https://doi.org/10.1142/s0219467822500097

Li, Y., Yang, G., Su, Z., Li, S., & Wang, Y. (2023). Human Activity Recognition Based on Multienvironment Sensor Data. *Information Fusion, 91(2023)*, 47–63. https://doi.org/10.1016/j.inffus.2022.10.015

Liao, T., Zhao, J., Liu, Y., Ivanov, K., Xiong, J., & Yan, Y. (2022). Deep Transfer Learning with Graph Neural Network for Sensor-Based Human Activity Recognition. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM),* Las Vegas, USA, 2445-2452. IEEE. https://doi.org/10.1109/bibm55620.2022.9995660

Lira, H., Martí, L., & Sanchez-Pi, N. (2022). A Graph Neural Network with Spatio-Temporal Attention for Multi-Sources Time Series Data: An Application to Frost Forecast. *Sensors, 22(4)*, 1486-1505. https://doi.org/10.3390/s22041486

Liu, P., Li, Z., Cong, Y., & Xu, Y. (2022). Sensor Network Prediction Based on Spatial and Temporal GNN. (L. Nguyen, Ed.) *ITM Web of Conferences, 47*(2022), 1-10. https://doi.org/10.1051/itmconf/20224701003

Lu, L., Zhang, C., Cao, K., Deng, T., & Yang, Q. (2022). A Multichannel CNN-GRU Model for Human Activity Recognition. *IEEE Access, 10*(2022), 66797–66810. https://doi.org/10.1109/access.2022.3185112

Mekruksavanich, S., Jantawong, P., Hnoohom, N., & Jitpattanakul, A. (2022). Deep Learning Networks for Eating and Drinking Recognition based on Smartwatch Sensors. *Research, Invention, and Innovation Congress: Innovative Electricals and Electronics (RI2C),* Bangkok, Thailand, 106-111. IEEE. https://doi.org/10.1109/ri2c56397.2022.9910318

Mekruksavanich, S., Jitpattanakul, A., Youplao, P., & Yupapin, P. (2020). Enhanced Hand-Oriented Activity Recognition Based on Smartwatch Sensor Data Using LSTMs. *Symmetry, 12*(9), 1570-1589. https://doi.org/10.3390/sym12091570

Mohamed, A., Lejarza, F., Cahail, S., Claudel, C., & Thomaz, E. (2022). HAR-GCNN: Deep Graph CNNs for Human Activity Recognition From Highly Unlabeled Mobile Sensor Data. *IEEE International Conference on Pervasive Computing and Communications Workshops and other Affiliated Events (PerCom Workshops),* Pisa, Italy, 335-340 IEEE. https://doi.org/10.1109/percomworkshops53856.2022.9767342

Mondal, R., Mukherjee, D., Singh, P. K., Bhateja, V., & Sarkar, R. (2021). A New Framework for Smartphone Sensor-Based Human Activity Recognition Using Graph Neural Network. *IEEE Sensors Journal, 21*(10), 11461–11468. https://doi.org/10.1109/jsen.2020.3015726

Moradi, B., Aghapour, M., & Shirbandi, A. (2022). Compare of Machine Learning and Deep Learning Approaches for Human Activity Recognition. *30th International Conference on Electrical Engineering (ICEE),* Tehran, Iran, 592-596. IEEE. https://doi.org/10.1109/icee55646.2022.9827335

Muralidharan, K., Ramesh, A., Rithvik, G., Prem, S., Reghunaath A, A., & Gopinath, D. M. (2021). 1D Convolution Approach to Human Activity Recognition Using Sensor Data and Comparison with Machine Learning Algorithms. *International Journal of Cognitive Computing in Engineering, 2*(2021), 130–143. https://doi.org/10.1016/j.ijcce.2021.09.001

Najeh, H., Lohr, C., & Leduc, B. (2022). Real-Time Human Activity Recognition in Smart Home on Embedded Equipment: New Challenges. In *Participative Urban Health and Healthy Aging in the*

Age of AI: 19th International Conference (ICOST), Paris, France, 125-138. Springer International Publishing. https://doi.org/10.1007/978-3-031-09593-1_10

Nale, R., Sawarbandhe, M., Chegogoju, N., & Satpute, V. (2021). Suspicious Human Activity Detection Using Pose Estimation and LSTM. *2021 International Symposium of Asian Control Association on Intelligent Robotics and Industrial Automation (IRIA).* Goa, India, 197-202. IEEE. https://doi.org/10.1109/iria53009.2021.9588719

Navita, & Mittal, P. (2022). Machine Learning (ML) based Human Activity Recognition Model using Smart Sensors in IoT Environment. *12th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 330-334.* IEEE. https://doi.org/10.1109/confluence52989.2022.9734152

Nguyen, B., Coelho, Y., Bastos, T., & Krishnan, S. (2021). Trends in human activity recognition with focus on machine learning and power requirements. *Machine Learning with Applications, 5*(2021). https://doi.org/10.1016/j.mlwa.2021.100072

Palimkar, P., Bajaj, V., Mal, A. K., Shaw, R. N., & Ghosh, A. (2021). Unique Action Identifier by Using Magnetometer, Accelerometer and Gyroscope: KNN Approach. In *Advanced Computing and Intelligent Technologies*, *218*(2021) , 607–631. Springer Singapore. https://doi.org/10.1007/978-981-16-2164-2_48

Pienaar, S. W., & Malekian, R. (2019). Human Activity Recognition using LSTM-RNN Deep Neural Network Architecture. *IEEE 2nd Wireless Africa Conference (WAC),* Pretoria, South Africa, 1-5. IEEE. https://doi.org/10.1109/africa.2019.8843403

Pires, I. M., Hussain, F., Garcia, N. M., Lameski, P., & Zdravevski, E. (2020). Homogeneous Data Normalization and Deep Learning: A

Case Study in Human Activity Classification. *Future Internet, 12*(11), 194. https://doi.org/10.3390/fi12110194

Radhika, V., Prasad, C. R., & Chakradhar, A. (2022). Smartphone-Based Human Activities Recognition System using Random Forest Algorithm. *International Conference for Advancement in Technology (ICONAT), Goa, India, 1-4.* IEEE. https://doi.org/10.1109/iconat53423.2022.9726006

Ramanujam, E., Perumal, T., & Padmavathi, S. (2021). Human Activity Recognition with Smartphone and Wearable Sensors Using Deep Learning Techniques: A Review. *IEEE Sensors Journal, 21*(12), 13029–13040. https://doi.org/10.1109/jsen.2021.3069927

Ramirez, R., Chiu, Y.-C., Hererra, A., Mostavi, M., Ramirez, J., Chen, Y., Huang, Y., & Jin, Y.-F. (2020). Classification of Cancer Types Using Graph Convolutional Neural Networks. *Frontiers in Physics, 8*(2020), 203. https://doi.org/10.3389/fphy.2020.00203

Razmah, M., Senthil G., A., Prabha, R., Divya, B., Sridevi, S., & Naveen, A. (2022). LSTM Method for Human Activity Recognition of Video Using PSO Algorithm. *International Conference on Power, Energy, Control and Transmission Systems (ICPECTS),* Chennai, India, 1-6 IEEE. https://doi.org/10.1109/icpects56089.2022.10046783

Rojanavasu, P., Jitpattanakul, A., & Mekruksavanich, S. (2021). Comparative Analysis of LSTM-based Deep Learning Models for HAR using Smartphone Sensor. *Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering,* Cha-am, Thailand, 269-272. IEEE. https://doi.org/10.1109/ectidamtncon51128.2021.9425733

Saleh, S. J., Ali, S. Q., & Zeki, A. M. (2020). Random Forest vs. SVM vs. KNN in classifying Smartphone and Smartwatch sensor data using

CRISP-DM. *International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI),* Sakheer, Bahrain, 1-4. IEEE. https://doi.org/10.1109/icdabi51230.2020.9325607

Sa-nguannarm, P., Elbasani, E., Kim, B., Kim, E.-H., & Kim, J.-D. (2020). Experimentation of Human Activity Recognition by Using Accelerometer Data Based on LSTM. In *Lecture Notes in Electrical Engineering, 716*(2020), 83–89. Springer Singapore. https://doi.org/10.1007/978-981-15-9309-3_12

Sarkar, A., Sen, T., & Roy, A. K. (2021). GraFeHTy: Graph Neural Network using Federated Learning for Human Activity Recognition. *20th IEEE International Conference on Machine Learning and Applications (ICMLA),* Pasadena, USA, 1124-1129. IEEE. https://doi.org/10.1109/icmla52953.2021.00184

Sarveshwaran, V., Joseph, I. T., Maravarman, M., & Karthikeyan, P. (2022). Investigation on Human Activity Recognition using Deep Learning. *Procedia Computer Science, 204*(2022), 73–80. https://doi.org/10.1016/j.procs.2022.08.009

Shao, Z., Guo, J., Zhang, Y., Zhu, R., & Wang, L. (2021). LightBGM for Human Activity Recognition Using Wearable Sensors. *International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS),* Xi'an, China, 668-671. IEEE. https://doi.org/10.1109/icitbs53129.2021.00169

Shiraly, K. (2021). *Latest Advances in Video-Based Human Activity Recognition | Modern HAR in Video & Images*. Retrieved January 10, 2023, from WIDTH.AI: https://www.width.ai/post/human-activity-recognition

Shiranthika, C., Premakumara, N., Chiu, H.-L., Samani, H., Shyalika, C., & Yang, C.-Y. (2020). Human Activity Recognition Using CNN &

LSTM. *5th International Conference on Information Technology Research (ICITR),* Moratuwa, Sri Lanka, 1-6. IEEE. https://doi.org/10.1109/icitr51448.2020.9310792

Suwannarat, K., & Kurdthongmee, W. (2021). Optimization of deep neural network-based human activity recognition for a wearable device. *Heliyon, 7*(8), 1-10. https://doi.org/10.1016/j.heliyon.2021.e07797

Tina, Sharma, A. K., Tomar, S., & Gupta, K. (2021). Various Approaches of Human Activity Recognition: A Review. *5th International Conference on Computing Methodologies and Communication (ICCMC),* Erode, India, 1668-1676. IEEE. https://doi.org/10.1109/iccmc51019.2021.9418226

Tong, L., Ma, H., Lin, Q., He, J., & Peng, L. (2022). A Novel Deep Learning Bi-GRU-I Model for Real-Time Human Activity Recognition Using Inertial Sensors. *IEEE Sensors Journal, 22*(6), 6164–6174. https://doi.org/10.1109/jsen.2022.3148431

Varshney, N., Bakariya, B., Kushwaha, A. K., & Khare, M. (2022). Human Activity Recognition by Combining External Features with Accelerometer Sensor Data Using Deep Learning Network Model. *Multimedia Tools and Applications, 81(2022)*, 34633–34652. https://doi.org/10.1007/s11042-021-11313-0

Vidya, B., & Sasikumar, P. (2022). Wearable Multi-Sensor Data Fusion Approach for Human Activity Recognition Using Machine Learning Algorithms. *Sensors and Actuators A: Physical, 341*(2022), 1-16. https://doi.org/10.1016/j.sna.2022.113557

Vrigkas, M., Nikou, C., & Kakadiaris, I. A. (2015). A Review of Human Activity Recognition Methods. *Frontiers in Robotics and AI, 2*(2015), 1-28. https://doi.org/10.3389/frobt.2015.00028

Wang, J., Chen, Y., Hao, S., Peng, X., & Hu, L. (2019). Deep Learning for Sensor-Based Activity Recognition: A Survey. *Pattern Recognition*

Letters, *119*(2019), 3–11. https://doi.org/10.1016/j.patrec.2018.02.010

Wang, M. (2022). A Comprehensive Survey on Human Activity Recognition Using Sensing Technology. *Highlights in Science, Engineering and Technology,* *9*(2022), 376–389. https://doi.org/10.54097/hset.v9i.1868

Weiss, G. M. (2019). WISDM Smartphone and Smartwatch Activity and Biometrics Dataset. *UCI Machine Learning Repository: WISDM Smartphone and Smartwatch Activity and Biometrics Dataset Data Set,* *7*(2019), 133190–133202.

World Health Organization. (2022). Physical activity. *Physical activity.* World Health Organization. Retrieved January 4, 2023, from https://www.who.int/news-room/fact-sheets/detail/physical-activity

Xia, K., Huang, J., & Wang, H. (2020). LSTM-CNN Architecture for Human Activity Recognition. *IEEE Access,* *8*(2020), 56855–56866. https://doi.org/10.1109/access.2020.2982225

Yi, M.-K., & Hwang, S. O. (2022). Smartphone Based Human Activity Recognition Using 1D Lightweight Convolutional Neural Network. *2022 International Conference on Electronics, Information, and Communication (ICEIC),* Jeju, Korea, 1-3. IEEE. https://doi.org/10.1109/iceic54506.2022.9748312

Zhang, S., Li, Y., Zhang, S., Shahabi, F., Xia, S., Deng, Y., & Alshurafa, N. (2022). Deep Learning in Human Activity Recognition with Wearable Sensors: A Review on Advances. *Sensors,* *22*(4), 1476. https://doi.org/10.3390/s22041476

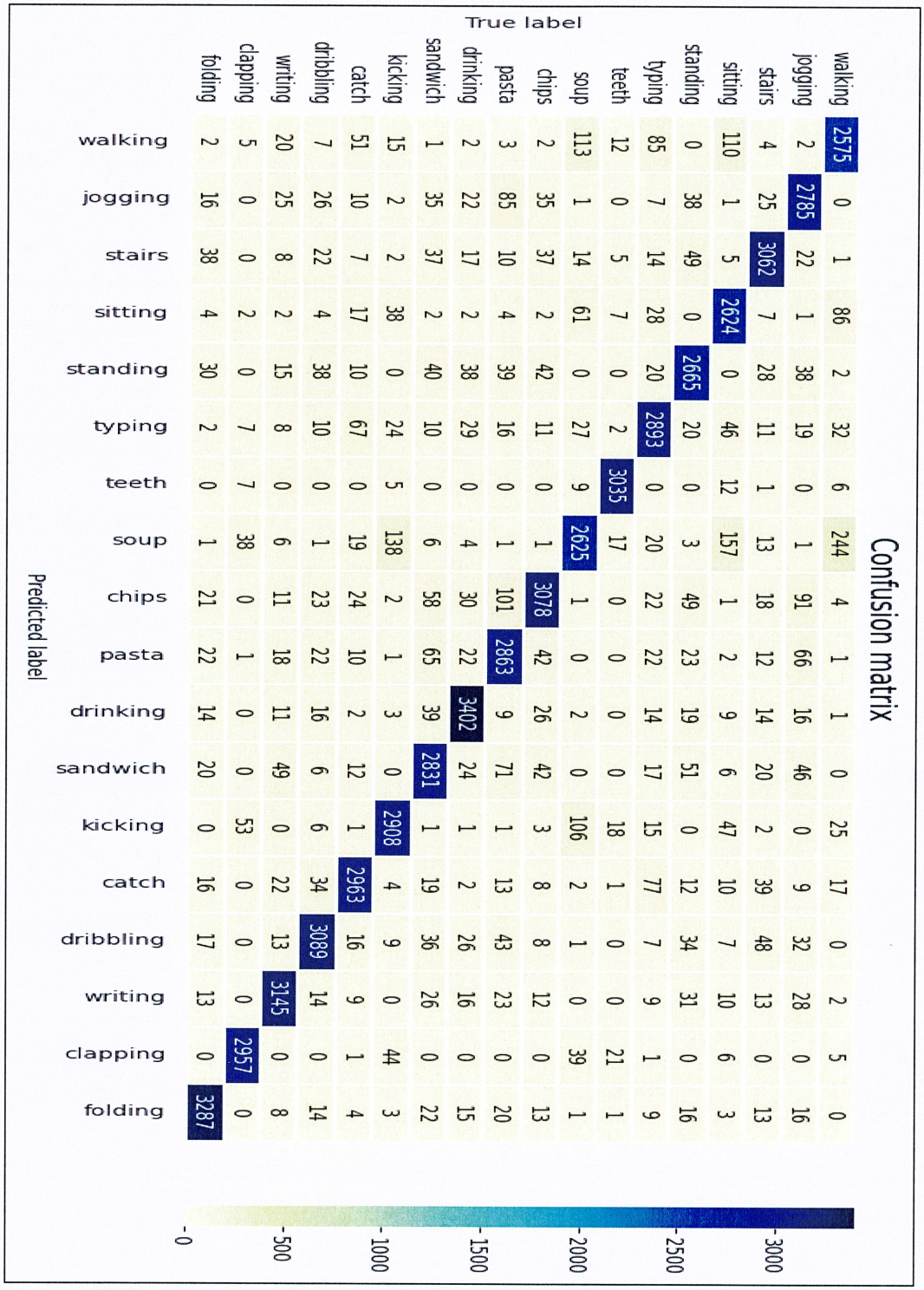
Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., & Li, H. (2020). T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Transactions on Intelligent Transportation*

Systems, *21*(9), 3848–3858. https://doi.org/10.1109/tits.2019.2935152

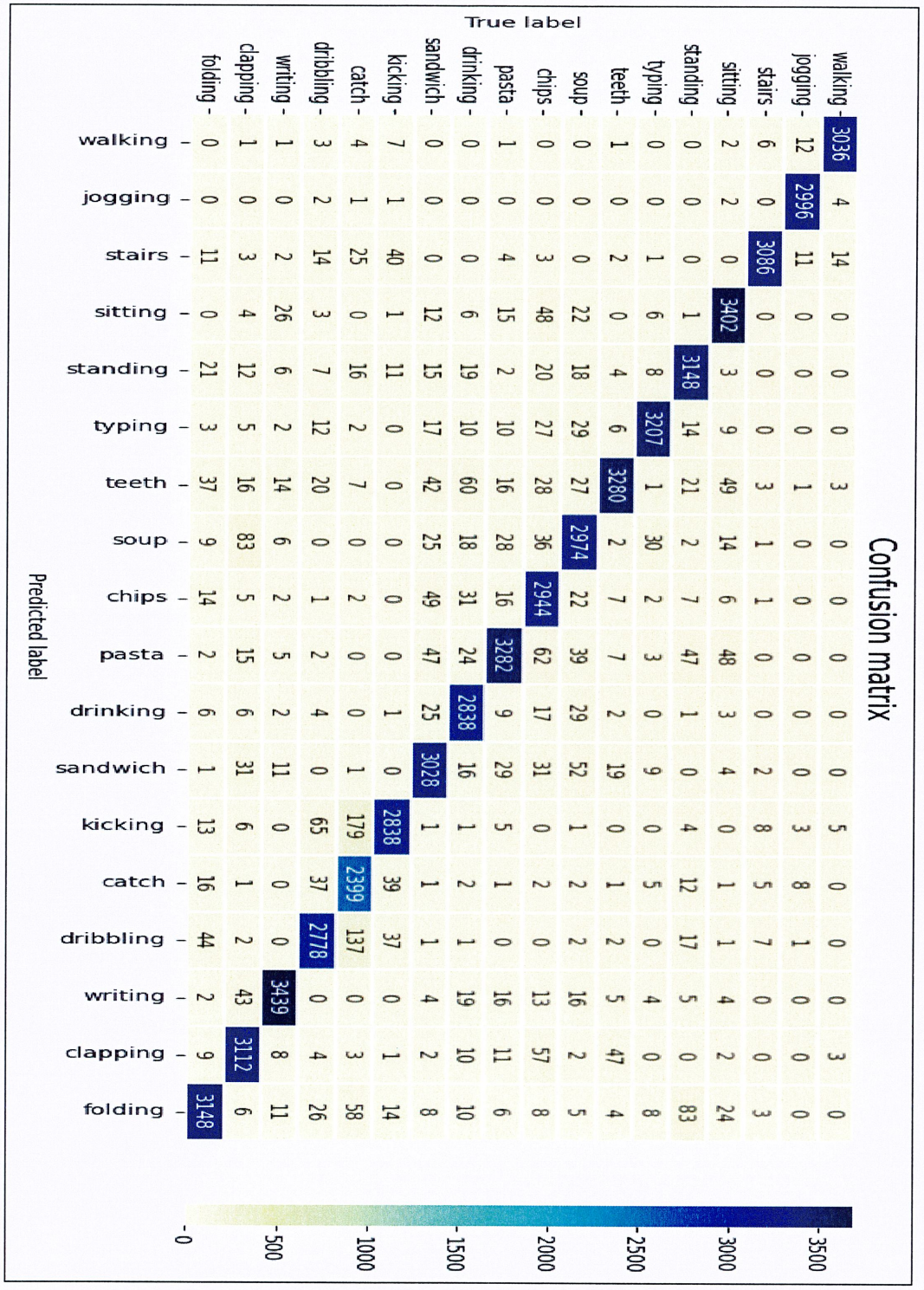
Zheng, L., Wu, D., Ruan, X., Weng, S., Peng, A., Tang, B., Lu, H., Shi, H., & Zheng, H. (2017). A Novel Energy-Efficient Approach for Human Activity Recognition. *Sensors,* *17*(9), 2064. https://doi.org/10.3390/s17092064

Appendices

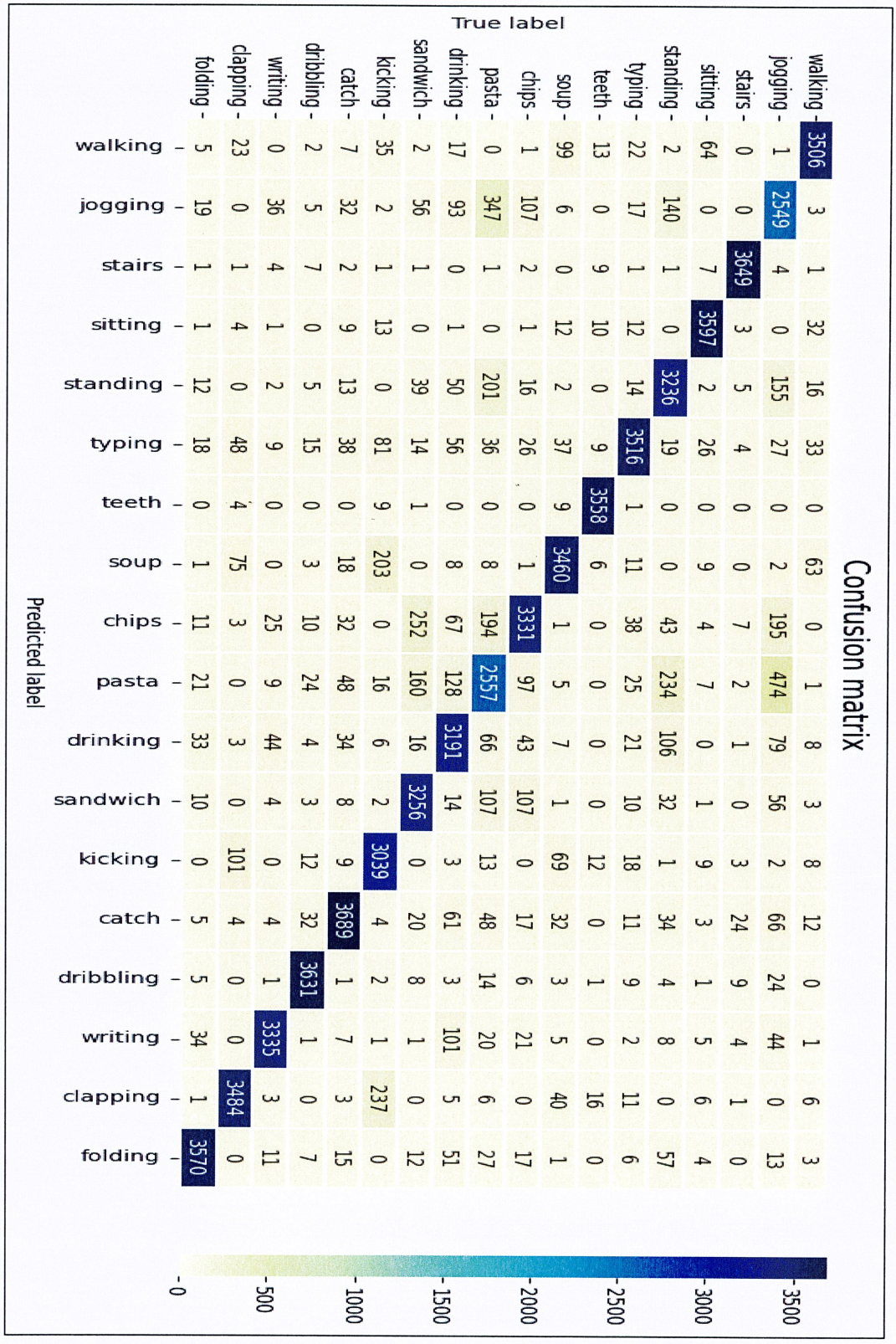
Appendix (A) The Confusion Matrix of LSTM Model using Smartphone Sub-dataset



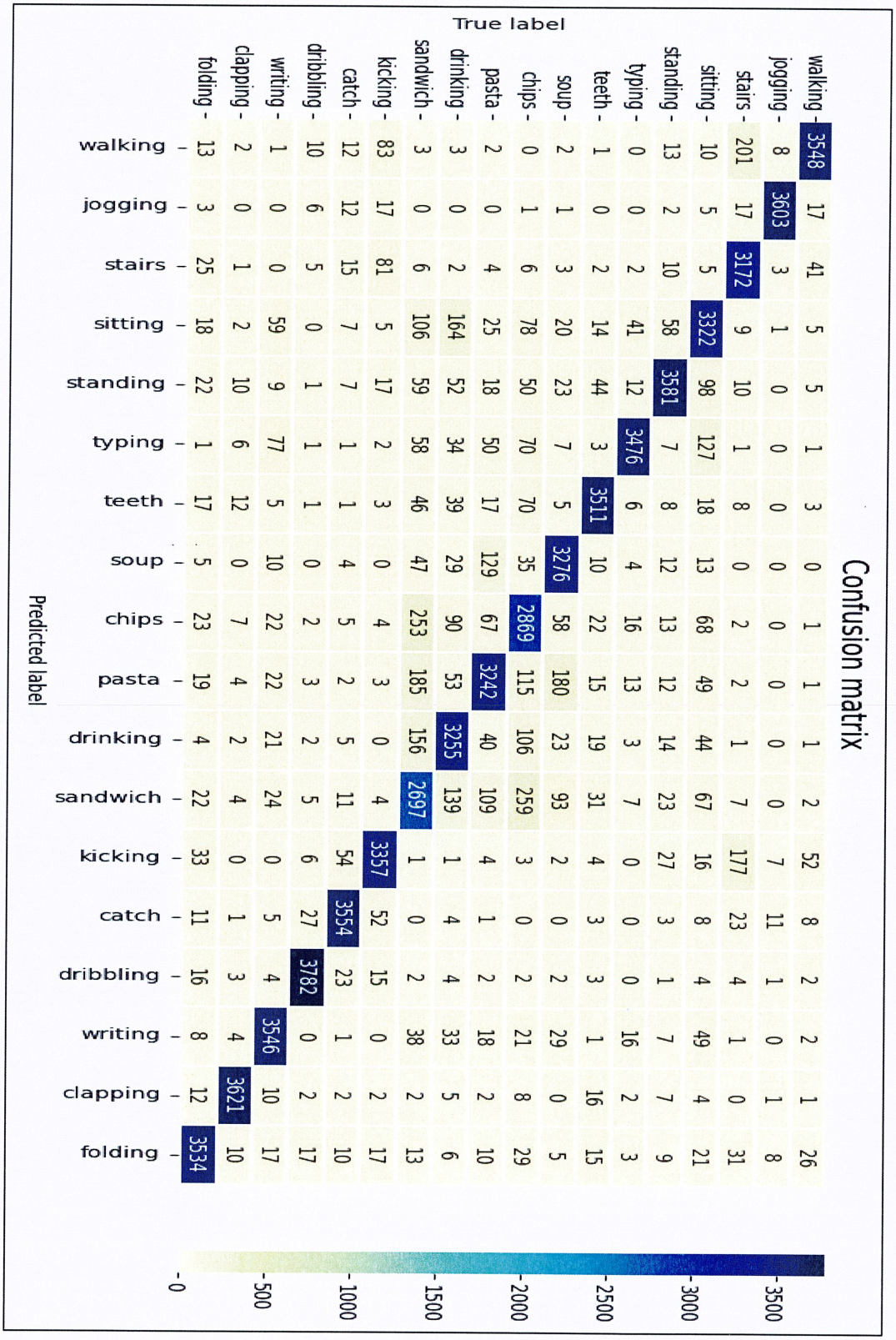
Appendix (B) The Confusion Matrix of LSTM Model using Smartwatch Sub-dataset.



Appendix (C) The Confusion Matrix of LSTM Model using Smartwatch Sub-dataset



Appendix (D) The Confusion Matrix of GCN Model using Smartwatch Sub-dataset



الملخص

أدى تطوير التكنولوجيا في أجهزة الحاسوب إلى جعل مراقبة أنشطة الإنسان اليومية وتحليلها مَهمة سهلة. لذلك أفادت أنظمة التعرف على النشاط البشري من المستشعرات المدمجة في الأجهزة الذكية في عدة مجالات، مثل الرعاية الصحية والأمن واللياقة البدنية. علاوة على ذلك، يمكن لهذه المستشعرات إنشاء بيانات متسلسلة ذات علاقات زمنية ومكانية. لذلك، تم استخدام تقنيات الذكاء الاصطناعي، مثل التعلم الآلي والتعلم العميق، لمعالجة البيانات وتصنيف النشاط. في حين أن طرق التعلم الآلي التقليدية قد حققت أداءً جيدًا، إلا أنها تتطلب استخراج الميزات يدويًا من البيانات الأولية.

في الوقت نفسه، يمكن لأساليب التعلم العميق استخراج الميزات المطلوبة التي تراعي التبعيات المكانية والزمانية. من بين تلك الطرق، أثبتت طريقة الذاكرة طويلة المدى (LSTM) فعاليتها في التعامل مع بيانات السلاسل الزمنية. بالإضافة إلى ذلك، فإن الشبكة التلافيفية للرسم البياني (GCN) لديها القدرة على استخراج العلاقات الزمانية المكانية باستخدام بنية بيانات الرسم البياني. ومع ذلك، على حد معرفة الباحث، لم تقارن الدراسات السابقة أداء كلتا الطريقتين، لذلك يهدف هذا البحث إلى مقارنة كلتا الطريقتين باستخدام مجموعة بيانات WISDM. تحتوي مجموعة البيانات على بيانات من مقاييس التسارع وأجهزة استشعار الجيروسكوب المضمنة في الهواتف الذكية والساعات الذكية لـ 18 نشاطًا منفذًا.

لكل جهاز ذكي، تم دمج بيانات المستشعرين لتشكيل مجموعتي بيانات فرعيين تم تجهيزهما مسبقًا وإعدادهما لتغذية النماذج باستخدام النوافذ المنزلقة المتداخلة. ومن ثم تم تطبيق كلتا الطريقتين على مجموعات البيانات الفرعية الناتجة. نتيجة لذلك، بالنسبة لمجموعة البيانات الفرعية للهواتف الذكية، تفوق نموذج GCN على نموذج LSTM بحوالي 4٪، مسجلاً 94.4٪ من حيث الدقة ودرجة F1. بالإضافة إلى ذلك، كان أداء نموذج GCN أفضل من بعض نماذج LSTM السابقة مع الأخذ في الاعتبار مجموعة فرعية من الأنشطة. من ناحية أخرى، أعطت مجموعة البيانات الفرعية للساعة الذكية ميزة طفيفة لنموذج GCN على نموذج LSTM بنسبة 1٪ تقريبًا، حيث سجلت 90.4٪ من حيث الدقة ودرجة F1. ومع ذلك، يمكن اعتبار طريقة GCN منافسة لطريقة LSTM في هذا النوع من البيانات، والتي يجب تطبيقها على مجموعات البيانات الأخرى في الأعمال المستقبلية للتعميم.

جامعة البحرين

كلية العلوم **قسم برامج الدراسات العليا**



مقارنة نموذجي الذاكرة طويلة المدى والشبكة التلافيفية للرسم البياني للتعرف على النشاط البشري

أطروحة مقدمة كجزء من متطلبات الحصول على درجة الماجستير
في علوم وتحليل البيانات الضخمة

إعداد

مجيد محمد حسن إبراهيم

202101787

إشراف

د .أحمد محمد زكي
أستاذ مساعد

جامعة البحرين

مملكة البحرين
أبريل 2023