ENHANCED CONDITIONAL GENERATIVE ADVERSARIAL NETWORK FOR HANDLING SUBJECT VARIABILITY IN HUMAN ACTIVITY RECOGNITION

ALI OLOW JIMALE

UNIVERSITI SAINS MALAYSIA

2023

ENHANCED CONDITIONAL GENERATIVE ADVERSARIAL NETWORK FOR HANDLING SUBJECT VARIABILITY IN HUMAN ACTIVITY RECOGNITION

by

ALI OLOW JIMALE

Thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

July 2023

DEDICATION

To my father "Olow Jimale Sabriye"

To my mother "Madina Barrow Abdi"

To my wife "Muna Abdulle Bariise"

To my kids

DECLARATION

Name **:Ali Olow Jimale**

Matric Number**:P-COD0002/20(R)**

School **:School of Computer Sciences**

Thesis Title**: Enhanced Conditional Generative Adversarial Network For Handling Subject Variability In Human Activity Recognition**

I hereby declare that this thesis which I have submitted to *School of computer sciences* on 28/01/2023 is my own work. I have stated all references used for the completion of my thesis.

I agree to prepare electronic copies of the said thesis to the examiner (s) for the determination of amount of words used or to check on plagiarism should a request be made.

I make this declaration with believe that what is stated in this declaration is true and the thesis as forwarded is free from plagiarism as provided under Rule 6 of the Universities and University Colleges (Amendment) Act 2008, Universiti Sains Malaysia Rules (Student Discipline) 1999.

I conscientiously believe and agree that the University can take disciplinary actions against me under Rule 48 of the Act should my thesis be found to be the work or ideas of other persons.

Students Signature: ______________________________ Date: __28__/__01__/2023__

ACKNOWLEDGEMENT

بسم الله الرحمن الرحيم

Praise be to Almighty ALLAH who created man from semen (fluid). We bear witness that there is no deity worthy of worship except God and prophet Muhammad (S.A.W) is His slave and messenger. May peace and blessing of ALLAH be upon him, his family, and his companions.

I would like to express heartfelt gratitude to my supervisor Dr. Mohd Halim Bin Mohd Noor for giving me the opportunity to learn from him vast knowledge despite his hectic schedule and commitment. His encouragement, convivial manner, advice, correction, suggestion, and tireless effort that brought this thesis into existing will remain in my memory forever. My heartfelt gratitude also goes to my internal panel members, which provided me with helpful suggestions during the proposal and research review.

Also, I acknowledged the support of my parents (beloved mother and father), brothers, sisters, my beloved wife, and children. They always pray for successful completion of this research work and my heart, thoughts, and prayers are always with them.

The contribution of my colleagues especially Fawaz Hameed Hazzaa Mahyoub (Yemen), Haruna Abdu (Nigeria), and others will never be forgotten. I really enjoyed your company. I would also like to show my gratefulness to Universiti Sains Malaysia (USM) and SIMAD university who gave me financial assistance during this course study.

Lastly, I offer my regards to all of those who supported me in any respect during the completion of this study.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS

μ	Mu
σ	Alpha
λ	Lambda
$\sum$	Sigma

LIST OF ABBREVIATIONS

CGAN	Conditional Generative Adversarial Network
FCGAN	Fully Connected Generative Adversarial Network
FN	False Negative
FP	False Positive
GAN	Generative Adversarial Network
HAR	Human Activity Recognition
TN	True Negative
TP	True Positive
USM	Universiti Sains Malaysia
VAE	Variational Autoencoder
X	x axis of accelerometer
Y	y axis of accelerometer
Z	z axis of accelerometer

PENINGKATAN RANGKAIAN ADVERSARIAL GENERATIF BERSYARAT UNTUK MENGENDALIKAN KEBOLEHUBAHAN SUBJEK DALAM PENGIKTIRAFAN AKTIVITI MANUSIA

ABSTRAK

Semasa memisahkan set data, penyelidik menganggap bahawa set latihan boleh ditukar dengan set ujian dan mengharapkan prestasi pengelasan yang baik. Andaian ini tidak sah kerana kebolehubahan subjek disebabkan perbezaan umur. Model klasifikasi yang dilatih menggunakan data aktiviti daripada satu kumpulan umur tertentu seperti orang dewasa tidak boleh digeneralisasikan kepada data aktiviti yang dikumpul daripada kumpulan umur yang berbeza seperti warga tua. Kebolehubahan subjek dalam konteks umur adalah masalah sah yang mengurangkan prestasi pengecaman aktiviti, tetapi masih belum diterokai. Kajian sedia ada yang menyiasat kebolehubahan subjek dalam pengecaman aktiviti mengabaikan masalah ini dan hanya menumpukan pada kebolehubahan subjek mengikut konteks dan kebolehubahan intra-subjek. Kajian ini menyiasat kesan kebolehubahan subjek terhadap prestasi pengecaman aktiviti berasaskan sensor. Set data warga emas dan dewasa digunakan untuk menilai teknik penilaian. Eksperimen pada set data dewasa sahaja, eksperimen pada set data warga tua sahaja dan eksperimen pada set data dewasa (sebagai latihan) dan warga tua (sebagai ujian) telah dijalankan menggunakan pembelajaran mesin dan pembelajaran mendalam. Keputusan menunjukkan penurunan prestasi yang ketara dalam pengiktirafan aktiviti pada subjek yang berbeza dengan kumpulan umur yang berbeza. Secara purata, penurunan ketepatan pengecaman masing-masing adalah 9.75% dan 12% untuk pembelajaran mesin dan model

pembelajaran mendalam. Rangkaian Adversarial Generatif Bersyarat (CGAN) adalah penyelesaian yang ideal untuk menangani kebolehubahan subjek. Walau bagaimanapun, CGAN tercanggih mengehadkan penggunaan rangkaian bersambung sepenuhnya dalam lapisan input penjananya dan lapisan keluaran diskriminatornya yang menjejaskan kualiti sampel. Kajian ini mencadangkan Rangkaian Adversarial Generatif Tersambung Penuh (FCGAN), seni bina CGAN yang berkesan yang mensintesis sampel berkualiti tinggi dengan menggabungkan lapisan konvolusi dengan berbilang rangkaian bersambung sepenuhnya dalam input penjana dan keluaran diskriminator. Penilaian visual, ukuran persamaan dan penilaian kebolehgunaan sedang digunakan untuk menilai dan mengesahkan kualiti sampel yang dihasilkan. Penilaian visual dan ukuran persamaan menunjukkan bahawa data sintetik model FCGAN lebih tepat mewakili data sebenar dan mencipta lebih banyak variasi dalam setiap data sintetik daripada CGAN tercanggih masing-masing. Penilaian kebolehgunaan menunjukkan peningkatan prestasi ketepatan sebanyak 2.5% (peringkat eksperimen I), 2.5%, (peringkat eksperimen II), 3.1% (peringkat percubaan III) dan 4.4% (peringkat eksperimen IV) berbanding CGAN tercanggih. Model FCGAN tidak dapat meningkatkan kebolehubahan data sintetik kerana ia hanya menambah hingar Gaussian pada data; maka data sintetik mewarisi pengedaran data sebenar yang sama daripada set data latihan. Kajian ini mengubah suai fungsi objektif FCGAN untuk mensintesis sampel dengan peningkatan subjek yang berbeza umur. Ini dicapai dengan menambahkan algoritma kekangan yang menggunakan nilai ambang dipacu data untuk memastikan kebolehubahan sampel sintetik kepada fungsi objektif FCGAN. Prestasi FCGAN yang dipertingkatkan dinilai menggunakan penilaian visual, ukuran kebolehubahan dan penilaian kebolehgunaan. Penilaian visual dan ukuran persamaan menunjukkan bahawa FCGAN yang dipertingkatkan menjana sebarang data yang diberikan latihan. Penilaian kebolehgunaan menunjukkan bahawa FCGAN yang dipertingkatkan mengatasi ketepatan FCGAN sebanyak 0.9% (peringkat percubaan I), 0.5% (peringkat eksperimen II), 0.7% (peringkat eksperimen III), dan 1.1% (peringkat eksperimen IV) menggunakan isyarat dewasa, dan 2% (peringkat eksperimen I), 6.5% (peringkat eksperimen II), 10.8% (peringkat eksperimen III) dan 12.6% (peringkat eksperimen IV) menggunakan isyarat warga tua dalam keseluruhan peringkat percubaan.

ENHANCED CONDITIONAL GENERATIVE ADVERSARIAL NETWORK FOR HANDLING SUBJECT VARIABILITY IN HUMAN ACTIVITY RECOGNITION

ABSTRACT

While splitting datasets, researchers assume that training set is exchangeable with test set and expect good classification performance. This assumption is invalid due to subject variability due to age differences. Classification models trained on activity data from one particular age group such as adults cannot generalize to activity data collected from a different age group such as elderly. Subject variability in the context of age is a valid problem that degrades the performance of activity recognition, but remains unexplored. Existing studies that investigated subject variability in activity recognition overlooked this problem and only focused on contextual subject variability, and intra-subject variability. This study investigates the effects of subject variability on the performance of sensor-based activity recognition. Elderly and adult datasets were used to evaluate the assessment techniques. Experiments on adult dataset only, experiments on elderly dataset only, and experiments on adult (as training) and elderly (as test) datasets were conducted using machine learning and deep learning. The results show a significant performance drop in activity recognition on different subjects with different age groups. On average, the drop in recognition accuracy is 9.75% and 12% for machine learning and deep learning models respectively. Conditional Generative Adversarial Network (CGAN) is an ideal solution to address subject variability. However, state-of-the-art CGANs restrict the use of fully connected networks in the input layer of its generator and the output layer of its discriminator which affects sample quality. This study proposed Fully-connected Generative Adversarial Network (FCGAN), an effective CGAN architecture that synthesizes higher-quality samples by combining convolutional layers with multiple fully connected networks in the generator's input and discriminator's output. Visual evaluation, similarity measure, and usability evaluation are being used to assess and validate the quality of generated samples. The visual evaluation and similarity measure demonstrates that the FCGAN models' synthetic data more accurately represents actual data and creates more variations in each synthetic data than the state-of-the-art CGAN respectively. The usability evaluation shows an accuracy performance gain of 2.5% (experimental stage I), 2.5%, (experimental stage II), 3.1% (experimental stage III), and 4.4% (experimental stage IV) over the state-of-the-art CGAN. The FCGAN model is not able to increase the synthetic data variability as it only adds Gaussian noise to the data; hence the synthetic data inherits the same real data distribution from the training dataset. This study modifies FCGAN's objective function to synthesize samples with increased subjects of different ages. This is achieved by adding a constraint algorithm that uses data-driven threshold value to ensure the variability of synthetic samples to FCGAN's objective function. The enhanced FCGAN's performance is assessed using visual evaluation, variability measure, and usability evaluation. The visual evaluation and similarity measure demonstrates that the enhanced FCGAN generates any data given the training. The usability evaluation shows that the enhanced FCGAN outperforms the FCGAN accuracy by 0.9% (experimental stage I), 0.5% (experimental stage II), 0.7% (experimental stage III), and 1.1% (experimental stage IV) using adult signals, and 2% (experimental stage I), 6.5% (experimental stage II), 10.8% (experimental stage III) , and 12.6% (experimental stage IV) using elderly signals in the overall experimental stages.

CHAPTER 1

INTRODUCTION

1.1 Background

Over recent decades, microelectronics and computer systems have been making outstanding development empowering sensor devices and mobiles with unprecedented characteristics (Lara & Labrador, 2013). Because of their low cost, small size, and high computational resources, microelectronics, and computer systems provide human daily living interactions with devices. It provides excellent pervasive systems that can be used to collect personal and environmental data (Davoudi et al., 2019). That was the formation of Pervasive Computing, which is a hot research area aimed at extracting knowledge from data collected by pervasive sensors (Perez et al., 2010).

Pervasive Computing, also known as Ubiquitous Computing, is a system in which computing resources can be accessed from everywhere, at any time of user demand (Ramli et al., 2006). It contains computers and cellular phones together with an internet connection, GPS chips, location information, and sensors (Perez et al., 2010). Pervasive computing and sensing technologies have advanced dramatically, enabling automatic analysis and recognition of human behavior and activities (A. Wang et al., 2021). It has a wide range of applications to enhance industries' quality and human aspects of care: from the manufacturing domain and smart cities to supporting elderly care, fitness tracking, and lifelogging (Becker et al., 2019; Maekawa et al., 2012b). Human Activity Recognition (HAR) is one of the important tasks in the applications of pervasive computing systems (Maekawa et al., 2012a).

Human activity recognition (HAR) is a classification problem that addresses how computers can understand what human is doing (Abdu-Aguye & Gomaa, 2019; Bulling et al., 2014; Ho et al., 2016; Kerdjidj et al., 2020). It can be defined as a process of identifying predefined activities of interest performed by a human through monitoring human activities and/or surrounding environments using sensors or digital cameras (Chiang et al., 2019). Existing HAR approaches can be divided into two categories: vision-based HAR and sensor-based HAR (L. Chen et al., 2012; Cook et al., 2008).

Traditionally, the vision-based HAR approach which analyzes digital images and or videos with human motions from cameras has been at the forefront of the human activity recognition field (Bulling et al., 2014; K. Chen et al., 2021). This is due to its potential applications in sports video analysis, surveillance systems, smart rooms, video retrieval, and human-computer interfaces (Ali & Shah, 2010).

However, the vision-based HAR modalities pose privacy, space, and light dependency issues (Ramasamy Ramamurthy & Roy, 2018). The location, angle, obstruction, illumination, and privacy invasion restricted the usage of vision-based HAR scenarios. Because of the defined drawbacks of vision-based HAR approaches, technological advancements, and low prices of sensor devices, HAR researchers shifted to work on a sensor-based HAR approach (Hussain et al., 2019).

In the sensor-based HAR approach, human activities are recognized using either environmental or wearable sensors (Jindong Wang et al., 2019). Environmental sensors are sensors that are either tagged to a certain location and human activity inference is based on the user's interaction with the tagged object, or they are deployed in an environment where no tag/device is required. Passive Infrared sensors, pressure sensors, and contact switches are examples of these sensors. Wearable sensors are sensors that are worn by users or attached to portable devices such as mobile phones, smartwatches, and many other textile products (Labrador & Yejas, 2011; Tufek et al., 2020). Accelerometers and gyroscopes are examples of wearable sensors used for activity recognition. Wang et al (Jindong Wang et al., 2019) has added a hybrid sensor to the list of human activity recognition sensor modalities. Table 1.1 summarizes the descriptions of sensor-based HAR modalities.

Table 1.1 HAR Sensor Modality Description

Sensor Type		**Description**
Wearable		The most common modality sensor that users wear to define human body movements.
Environmental	Object	A sensor that is attached to objects to sense movements of objects.
	Ambient	A sensor that is applied in the environment to reflect the interaction of the user
Hybrid		Combination of two or three sensor types.

Nevertheless, wearable sensors have recently dominated the Human Activity Recognition field due to their outstanding advantages over the rest of the sensor modalities (Cornacchia et al., 2017). First of all, wearable sensors are cheap, easy to use, ubiquitous, and unobtrusive. These features made smartphones and wearable sensors become part of human daily life and a popular method for human activity recognition (Nweke et al., 2018). Examples of wearable sensors that can be embedded with smartphones and watches include accelerometers and gyroscopes. Several studies confirm that combining these two sensors improves the performance of sensor-based HAR (M. Zhang & Sawchuk, 2012).

The second advantage of mobile and wearable sensors is their capability to support the real-time implementation of HAR systems. In addition to that, mobile and wearable sensors are easy to deploy and do not pose any health hazard to their users (Cornacchia et al., 2017). Because of these advantages, several machine learning and deep learning methods have been explored to classify and recognize human activities using wearable sensors such as accelerometers, and gyroscopes.

Machine learning methods for sensor-based HAR use hand-crafted features that are manually extracted with the help of domain experts (Farias et al., 2016; Jindong Wang et al., 2019). However, expert-driven feature extraction methods have several issues (Nweke et al., 2018). First of all, domain experts can only learn very limited features (J. B. Yang et al., 2015) related to some statistical information, including mean, frequency, variance, and amplitude which cannot fully support the dynamic nature of today's ubiquitous and seamless collection of wearable and mobile senor streams (Hasan & Roy-Chowdhury, 2015). These shallow features also fail to support modeling complex activities (Q. Yang, 2009) and involve time-consuming feature selections (Ronao & Cho, 2016). Second, manually engineered features are error-prone which may result in the loss of important information for activity recognition (D. Shi et al., 2015). This affects the performance and accuracy of the human activity recognition system (Nweke et al., 2018). Third, the current manual feature extraction is application-dependent or problem-specific that cannot be transferred to another activity with similar patterns. Finally, there is no universal rule for selecting appropriate human activity features. That is the reason why several studies resort to large-scale heuristic domain knowledge to create and select appropriate features for a given human activity recognition system (Safaei et al., 2019).

To deal with manual feature extraction issues, researchers have recently applied deep learning to automate feature extraction and extract higher-level representation to recognize human activities (Shaheen et al., 2016; Jindong Wang et al., 2019). Although deep learning methods have achieved outstanding performance in feature classification and are being adopted to automatically learn features in sensor-based human activity recognition, it is still premature for pervasive and wearable sensor data.

1.2 Motivation

The increased life expectancy, together with declining birth rates led to an aged population structure. The population of the world is rapidly aging (Lee et al., 2020). Approximately all countries in the world are experiencing growth in the percentage of elderly in their population. For instance, the current number of elderly people (60 years and older) in the world is higher than the number of children younger than 5 years old. By 2050, it is expected that 1 in 6 persons in the globe will be over 65 years old (Agudelo-Cifuentes et al., 2019).

This increased longevity is a threat to the stability of every society due to its negative effects on elderly health and social care (Howdon & Rice, 2018) including loss of physical, mental, and cognitive abilities causing impaired actions and greater vulnerability to morbidity and mortality (Chang et al., 2019). Aged people are always vulnerable to many age-related problems including diabetes, stroke, Parkinson's, Alzheimer's, dementia, cardiovascular, osteoarthritis, and other chronic diseases (Subasi et al., 2020; Vepakomma et al., 2015).

These diseases together with the weak cognitive and physical ability of the elderly prevent them from independent living and barrier them from performing daily activities (i.e. toileting, bathing, cooking, etc.) (Van Kasteren et al., 2010). To assist elderly people, some family members and governments provide high nursing spending on elderly care (Vepakomma et al., 2015; Yao et al., 2018).

However, with the increase in the elderly population, caregivers' assistance is becoming scarce and the caregivers become overburdened with the continuous monitoring responsibility (Piyathilaka & Kodagoda, 2015; Richter et al., 2017). Therefore, there is a primary need for a system that can early detect the elderly gradual cognitive changes and automatically recognize elderly activities to monitor their health conditions and provide evidence-based nursing assistants (Nambu et al., 2000; Vijayaprabakaran et al., 2020).

This has recently attracted many scientists who proposed human activity recognition systems aimed at promoting and assisting the living independence of elderly people through developing techniques and systems that recognize the mobility, daily life activities, and physiological signs of elderly people (Khusainov et al., 2013). This is one reason why activity recognition is becoming a hot research area in sensor-rich and ubiquitous mobile devices (Zahin et al., 2019b) that is specially applied in the elderly healthcare domain (Dinarević et al., 2019).

In this time of crisis, it has become a fundamental need to recognize and monitor human activities and behaviors to maximize human protection from COVID-19 and prevent its possible spread to them. In this regard, researchers conducted sensor-based activity recognition studies that can help in the fight against the virus including studies for avoiding face touching (Michelin et al., 2021), contact tracing (Angelo & Palmieri, 2021), supporting consistent home workout schedules (Matthies et al., 2021), monitoring social distancing (X. Wang et al., 2022), and assessing a confined person's body temperature (Hoang et al., 2021).

Understanding the different kinds of human activities also has an extensive contribution to solving other real-world problems such as security and military (Labrador & Yejas, 2011; Lara & Labrador, 2013), entertainment, surveillance, gaming, remote monitoring, intelligent environments (Hussain et al., 2019), health tracking and monitoring, rehabilitation and assisted living (Rezaie & Ghassemian, 2018), home behavior analysis (Satapathy & Das, 2016), gait analysis (Hammerla et al., 2016), gesture recognition (Y. Kim & Toomajian, 2016), assistive technologies & manufacturing. It may easily change the way people sense, monitor, recognize and predict human physical activities and surrounding environments (Campbell et al., 2008; Chiang et al., 2019).

Subject Variability generated by age differences among subjects is one key challenge that affects the recognition accuracy of deep learning models for sensor-based human activity recognition. The motivation of this research is to provide techniques to investigate and bridge the recognition accuracy gap across different subjects of different ages in sensor-based HAR.

1.3 Problem Statement

The way human activities are performed and their durations vary from one person to another (Akbari & Jafari, 2020). Subject variability can be caused by several factors such as age, sex, fitness level, and environmental state. This variation (referred to as subject variability) changes the pattern of the sensory data from one subject to another and limits the generalization of the classification models to new subjects, hence reducing the recognition accuracy of the models.

Subject variability generated by age differences among subjects occurs whenever the classification models are trained with sensor data from one particular age group such as adults and tested the trained model with sensor data from another different age group e.g., elderly.

The signals of elderly activities vary from the signals of adult activities even when the same activity is being performed. Typically, the acceleration (magnitude) is lower, and the activity signals have a longer duration. This is because, elderly people have a lower intensity of dynamic (i.e., walking, running, jogging) and transitional (i.e. stand-to-sit, sit-to-stand) activities and less stable static activity (i.e. standing, sitting) than adults.

This variation originates from the fact that adults are stronger, more confident, and more active than elderly people in performing the activities. Consequently, a classification model that is trained on activity data that is collected from adults is not able to generalize to the elderly’s dataset. Although subject variability (age) is a real problem in activity recognition, it remains largely unexplored. The existing public HAR datasets such as UCI HAR (Anguita et al., 2013),mHealth (Banos et al., 2014), OPPORTUNITY (Roggen et al., 2010), PAMAP2 (Reiss & Stricker, 2012), and WISDM (Kwapisz et al., 2011) do not contain subjects with different age groups. As dataset creation with hundreds of subjects with different age groups are expensive and time-consuming with great effort requirements (Ono & Suzuki, 2020), existing datasets for activity recognition are collected from subjects in the same age group. In particular, publicly available datasets in use today are collected in fully controlled or semi-natural settings, and elderly people's activities are not considered (Abdu et al., 2021).

One ideal solution for overcoming subject variability generated by age differences among subjects in sensor-based HAR is collecting more training data with various subjects of different ages. However, collecting large, annotated training data is expensive and time-consuming. In particular, it is an obstacle in sensor-based HAR. This is due to certain limitations of wearable sensors, including battery lifetime and storage capacity constraints.

Data augmentation can be another ideal solution to deal with subject variability generated by age differences among subjects in activity recognition. Generative Adversarial Network (GAN) is the most popular method for overcoming limited datasets problems in activity recognition used in the literature recently due to its ability to generate verisimilar synthetic samples (Ian J. Goodfellow∗, Jean Pouget-Abadie†, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair‡, Aaron Courville, 2014).

However, a key limitation of the current GANs architecture is generating better synthetic data with rich subjects of different ages. It is not able to increase the synthetic data variability as it only adds Gaussian noise to the data; hence the synthetic data inherits the same real data distribution from the training dataset. Another key issue concerning existing GANs is the design of network architecture that did not implement fully connected layers (Hong et al., 2020). Not including fully connected layers in GAN architectures results in poor quality of the synthetic samples (Barua et al., 2019). Therefore, there is a need to develop an enhanced GAN model that generates high-quality synthetic data with increased artificial subjects of different ages.

1.4 Research Questions

These research questions aim to find specific issues to be addressed to improve human activity recognition performance on different subjects with different age groups. These questions include:

1. How subject variability generated by age differences among subjects affects the performance of sensor-based activity recognition?

2. How to enhance the network architecture of GANs to generate better synthetic data with rich subjects of different ages for sensor-based activity recognition?

3. How to enhance the training algorithm of GANs to generate verisimilar subject variability-rich sensor data of different ages for sensor-based activity recognition?

1.5 Research Objectives

This research aims to address subject variability generated by age differences among subjects in sensor-based HAR. The main objectives of this research are:-

1. To investigate the effects of subject variability generated by age differences on sensor-based HAR. The focus of this objective is to demonstrate that subject variability is a real issue that contributes to the performance decline of sensor-based activity recognition.

2. To design an improved GANs network architecture that synthesizes verisimilar human activity signals and achieves faster model learning convergence and training stability. The focus of this objective is to present an improvement to the state-of-the-art GANs for data augmentation by changing the network structure.

3. To design an algorithm for training GANs to generate verisimilar subject variability-rich sensor data of different ages for sensor-based HAR. The focus of this objective is to enhance the proposed GANs by modifying the loss function.

1.6 Research Contributions

The main contribution of this research is proposing a GAN approach, which is designed to produce more quality synthetic samples with increased artificial subjects of different ages. The contributions of this thesis are summarized as follows; -

1. The first study to investigate the performance degradation caused by subject variability contributed by age differences among subjects in activity recognition using various machine learning and deep learning techniques.

2. FCGAN, an effective Conditional GAN architecture that combines convolutional layers with multiple fully connected networks in the input and output layers

of the generator and discriminator respectively to generate more quality synthetic samples for sensor-based activity recognition.

3. An enhanced FCGAN architecture that modifies the objective function of the FCGAN to produce synthetic data with increased artificial subjects of different ages.

Synthetic subject-rich data to improve the classification accuracy of sensor-based HAR.

1.7 Scope of the study

The scope of this research will be limited to investigating and handling subject variability contributed by age differences among subjects in sensor-based HAR using GANs architecture. Two different datasets – one locally collected from 15 elders and another public from 30 adults with 8 types of human activities namely walking, standing, sitting, lying down, sit-to-lie, sit-to-stand, lie-to-sit, and stand-to-sit – will be used to evaluate the assessment techniques and to generate artificial subjects.

1.8 Research Process

This section presents a systematic plan implemented to conduct this research. As shown in Figure 1.1, the steps of this study are divided into four main phases: preliminary investigation & analysis, experimental dataset collection & pre-processing, design and modeling, and performance evaluation.

In phase 1, existing research studies related to the study will be reviewed and analyzed. This phase starts with an investigation of how classification models are trained, followed by the types of subject variability in sensor-based HAR. After reviewing the literature related to this study, problems with existing methods will be analyzed. Then, the effects of subject variability on activity recognition using inertial sensors will be investigated. As a result, the problem statement of this work is identified to outline the proposed approach.

In phase 2, two activity recognition datasets are used which are an internally collected (local) dataset and a public dataset. Both datasets will be preprocessed using fixed sliding window segmentation.

In phase 3, modified and enhanced GANs models for generating more quality synthetic data with various subjects for sensor-based HAR using limited sensory data will be designed. The proposed models will be implemented using a python programming language with scientific computation, visualization, and deep learning libraries such as NumPy, pandas, pyplot, Keras, Tensorflow, etc.

In phase 4, an evaluation will be carried out to validate the research findings and the proposed approaches. To do the evaluation, a comparison between the proposed approaches' findings and the existing studies is carried out. The usability of the synthetic data is also assessed. Standard evaluation metrics to be measured include recognition accuracy. Finally, conclusions and further work improvements are listed.

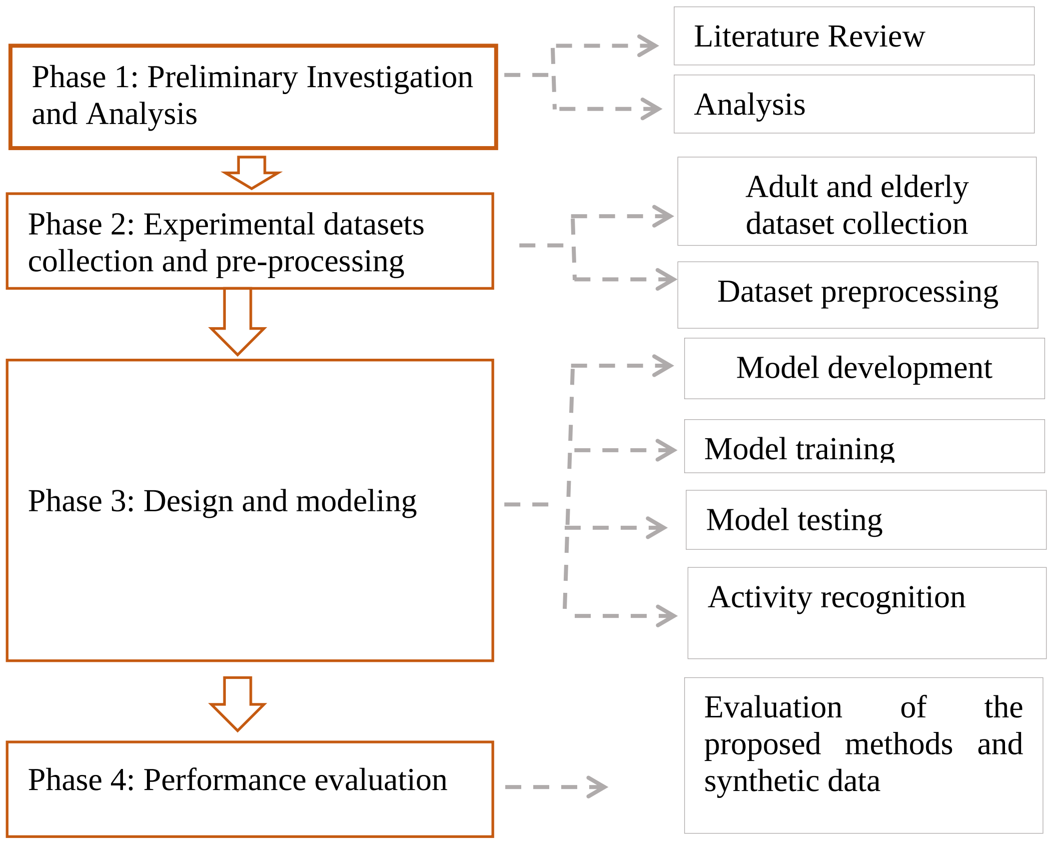

Figure 1.1 Overview of the research process

1.9 Thesis Organization

The remaining chapters of this thesis are organized as follows. CHAPTER 2 presents the background and literature review of the study. CHAPTER 3 presents the research methodology. It contains the proposed approaches in this research. CHAPTER 4 investigates the effects of subject variability generated by age differences on activity recognition. CHAPTER 5 reports the experiments, results, evaluation, and discussion of the FCGAN. CHAPTER 6 contains the experiments, results, evaluation, and discussion of the enhanced FCGN. CHAPTER 7 provides a conclusion, recommendations, and future work for this research. It explains the achievements of the research objectives and questions. Also, the limitations of this research and the future work related to this study are discussed.

CHAPTER 2

LITERATURE REVIEW

The previous chapter has given an overview of the research. This chapter provides an overview of human activity recognition, human activity sensing, and modeling. It also reviews the existing approaches and methods related to sensor-based HAR, in particular subject variability generated by age differences among subjects. The chapter is organized as follows.

Section 2.1 gives an overview of human activity recognition. Section 2.2 discusses human activity sensing. Section 2.3 contains activity modeling and recognition. Section 2.4 and section 2.5 explain human activity recognition using machine learning and deep learning techniques, respectively. Section 2.6 introduces the methods used to address subject variability problems in activity recognition. It also discusses the state-of-the-art. Section 2.7 summarizes the chapter.

2.1 Overview of Human Activity Recognition

Human Activity can be defined as the physical movements of the human body (Beddiar et al., 2020). Human Activity has 5 basic attributes: subject, action, time, object, and location (The et al., 2010). The subject is the person performing the activity, action is the type of human activity itself, time is the duration of a specific human activity, an object is the kind of object has the activity affected, and location is where the activity is performed.

The time attribute is one of the basic approaches currently being adopted in research to classify human activities. Considering the time duration, human activities can be classified into two main categories: transitional and basic human activities (J. H. Li et al., 2019).

Transitional human activities are simple events with a small duration in the order of seconds. These activities are further divided into two subcategories: gesture and transition. Gesture refers to the visual movements of a part of the human body such as the arm, hand, head, and finger to communicate nonverbally (J. Liu et al., 2019), while transitions are the activities that connect the transitions between two different human activities such as lie-to-sit, lie-to-stand, stand-to- sit, stand-to-lie, sit-to-lie, and sit-to-stand (J. H. Li et al., 2019).

On the other hand, basic activities are human activities with a long duration in the order of minutes. These activities can be characterized as either dynamic activity or static activity. Dynamic activities are continuous activities with periodicity (i.e., walking, running), while static activities are activities with static posters (i.e., sitting, standing). Figure 2.1 shows the taxonomy of human activity categories.

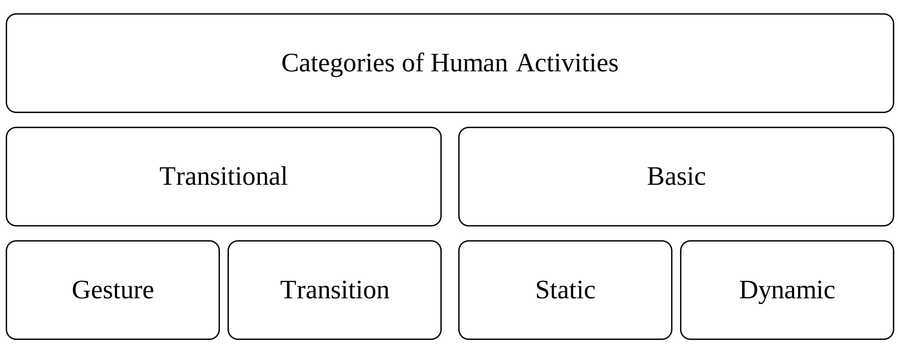

Figure 2.1 Taxonomy of human activity categories

Understanding the different kinds of human activities can have an extensive contribution to solving various real-world problems such as supporting elderly care, lifelogging, demotics, security, entertainment, surveillance, gaming, remote monitoring, intelligent environments, and human-computer interaction (Hussain et al., 2019; Labrador & Yejas, 2011; Maekawa et al., 2012a). The research field which aims at understanding human activities and behavior is called Human Activity Recognition (HAR) (J. H. Li et al., 2019). It has recently been applied in the fight against the COVID-19 virus. There are studies for enhancing COVID-19 contact tracking (Angelo & Palmieri, 2021), supporting consistent home workout schedules (Angelo and Palmieri 2021), measuring a confined person's body temperature (Hoang et al., 2021), recognizing face touching (C. Bai et al., 2021), and monitoring social distancing (X. Wang et al., 2022).

Human Activity Recognition (HAR) refers to the process of identifying human physical movements, actions, and surrounding environments based on motion data collected through digital cameras and sensor devices (Abdu-Aguye & Gomaa, 2019; Bulling et al., 2014; Campbell et al., 2008; Chiang et al., 2019).

Existing HAR approaches can be divided into two categories: vision-based HAR and sensor-based HAR (L. Chen et al., 2012; Cook et al., 2008). To recognize human activities, the vision-based HAR approach analyzes digital images and or video sequences with human motions from cameras (K. Chen et al., 2021). Initially, this approach has been a hot scientific topic. Many researchers investigated human activity recognition from images and videos (Bulling et al., 2014). This is due to its wide applications in sports, surveillance systems, health care, smart rooms, video retrieval, and human-computer interfaces (Ali & Shah, 2010).

Later, the Sensor-based HAR approach became a popular and fast-growing topic. (Hussain et al., 2019). Many researchers are working on it. This is due to technological advancements and low prices of sensor devices as well as the issues of the vision-based approach including privacy, space, and light dependency issues (Ramasamy Ramamurthy & Roy, 2018).

2.2 Human Activity Sensing

Human Activity Recognition Systems with sensing capabilities can improve human quality of life, especially for people having any type of limitation and lack of well-being. For this purpose, various sensor-based HAR techniques have been developed to assist the limitations and well-being of humans.

The sensor-based HAR approach makes use of either an environmental sensor or a wearable sensor to recognize human activities (Jindong Wang et al., 2019). In the former, sensor devices are either fixed to a defined point (object tagged) or deployed in an environment (dense sensing/ambient). In object tagged, sensors are attached to objects and human activity inference depends on the user's interaction with these objects. In ambient sensing, sensors are applied in the environment and human activities are recognized based on the environmental changes surrounding the subjects.

On the other hand, the sensor devices are worn by humans or embedded with portable devices such as smartphones and watches, clothes, etc. These sensors are cheap, easy to deploy, easy to use, ubiquitous, unobtrusive, and capable to support real-time HAR. Unlike wireless signals-based, wearable sensors do not pose any health hazard to their users (Cornacchia et al., 2017). Because of these features, several studies have explored different machine learning and deep learning algorithms to classify and recognize human activities using wearable sensors such as accelerometers, magnetometers, and gyroscopes. In particular, accelerometer and gyroscope motion sensors are the most frequently used wearable sensor in human activity recognition due to their effectiveness (Straczkiewicz & Onnela, 2019).

An accelerometer is a wearable sensor that can be integrated into smartphones, clothes, watches, and even bands. This sensor is used to measure object acceleration in the form of meters per second square ($^{m}/_{s^2}$) or gravitational force (g) units. Its sampling rate is in the range of tens to hundreds of Hz (K. Chen et al., 2021). To recognize human activities, an accelerometer can be mounted on the waist, arm, ankle, and wrist. Three axes accelerometer is often used to collect tri-variate time-series data with the x-axis, y-axis, and z-axis (K. Chen et al., 2021). The accelerometer device's location and acceleration are represented by these axis coordinates.

Unlike an accelerometer, a gyroscope sensor measures orientation and angular velocity in the form of degrees per second ($^{\circ}/_{s}$) units. However, it has a similar sampling rate, axes, and data dimension to the accelerometer. Gyroscope sensors are usually combined with accelerometer sensors and mounted on the same body parts such as the chest, waist, arm, ankle, wrist, etc. (K. Chen et al., 2021).

Wang et al (Jindong Wang et al., 2019) added a hybrid sensor to the list of human activity recognition sensor modalities. It combines different sensor modalities to perform activity modeling and recognition.

2.3 Activity Modeling and Recognition

HAR is a typical pattern recognition problem that has achieved good progress by utilizing various machine learning and deep learning classifiers such as Support Vector Machine, Decision Tree, k-Nearest Neighbors, Naive Bayes, Hidden Markov Models, Adaboost, Random Forest, XGBoost, Convolutional Neural Network, Restricted Boltzmann Machine Restricted, Deep Autoencoder, Sparse coding, and Recurrent neural network.

Most activity recognition systems follow four regular phases (data collection, data pre-processing, feature extraction, and training and activity classification (Nweke et al., 2018; Straczkiewicz & Onnela, 2019)) with slight variations based on the model (machine learning vs. deep learning), application domain, and dataset.

2.3.1 Data collection

HAR datasets can be accomplished by collecting and storing humans' motion data and or surrounding environments' data via smart sensors and video cameras (Anguita et al., 2013; L. Chen et al., 2012; Cook et al., 2008; Nweke et al., 2018). The HAR data is collected in either naturalistic or controlled experimental environments (Anguita et al., 2013). In the naturalistic data collection protocol, the participants receive no specific instructions on how to perform the physical activities, while the controlled data collection protocol is subject to laboratory conditions that participants are required to follow while performing the physical activities (Labrador & Yejas, 2011).

The collected data is either manually annotated by the data collectors or automatically annotated by designated applications. Some HAR systems perform self-data collection while others utilize existing datasets (Jindong Wang et al., 2019) mainly created by a group of research assistants for either HAR public dataset or private dataset for specific HAR problems.

Usually, data comes in an unclean format that may not be feasible for feature extraction and modeling (Zhaohui Wang et al., 2018). Data is unclean if it is missing features, or feature values, contains noise, outlier, duplicate, or wrong data. The presence of any of these produces misleading results. Data preprocessing which will be discussed in the next section is an important step in any machine learning project to prepare data for further analysis (Attal et al., 2015).

2.3.2 Preprocessing and signal segmentation

Data preprocessing of HAR refers to data cleaning (Zhaohui Wang et al., 2018). Data cleaning aims at denoising the data, detecting and filtering outlier values in the raw data, replacing or eliminating raw data missing, and performing class balancing (Attal et al., 2015; Fong et al., 2016; Zhu et al., 2019). Possible errors in human activity data include noise from sensors, data collection problems, incorrect labeling of human activities, and environmental noise (Aminikhanghahi & Cook, 2019).

On the other hand, signal segmentation refers to the process of segmenting the data into independent human activity fragments for feature extraction and modeling purposes (J. H. Li et al., 2019). Through data segmentation, feature vector input for feature extraction could be provided (Ma et al., 2020; Suto et al., 2018). A fixed sliding window is the most common signal segmentation technique in sensor-based activity recognition (Noor et al., 2017).

2.3.3 Feature extraction and selection

Feature extraction is the most important part of HAR which transforms high-dimensional data into a reduced and meaningful data dimensionality (Attal et al., 2015). This process expresses preprocessed data as some features that maximize preprocessed data into meaningful features for activity recognition (Straczkiewicz & Onnela, 2019; Zhaohui Wang et al., 2018). It identifies lower sets of human activity features from input data to reduce human activity classification errors and computational complexity (Nweke et al., 2018).

There are two types of feature extraction: manual feature extraction and automatic feature extraction. Manual feature extraction is an expert-driven approach that requires domain knowledge to extract relevant features such as time-domain features, frequency domain features, and Hulbert- Huang features from physical activity data (Zhelong Wang et al., 2016; Zdravevski et al., 2017). Examples of features include mean median, variance, and skewness. Unlike manual feature extraction, automatic feature extraction employs deep learning techniques to automatically generate human features and select the best ones (Zdravevski et al., 2017). Through feature extraction techniques feature set for HAR activity is created.

Some of these features become irrelevant, useless, or redundant based on the type of activity and the classifier to be applied in the HAR system (Malaise et al., 2019). This is where feature selection techniques play an important role in selecting the most significant features from the feature set. Feature selection is a technique of automatically selecting relevant features from a large feature set to improve the recognition generalization performance and reduce the dimensionality of the features (Malaise et al., 2019). These features will be utilized to train the HAR system that can detect and recognize human activities (K. Kim et al., 2019).

There are three general categories of feature selection methods: filter, wrapper, and embedded method. Filter methods use statistical measures to rank features. Examples of filter feature selection methods are the Chi-squared test, correlation coefficient scores, and information gain. The wrapper feature selection method selects a set of features like a search problem and compares it with other features to assign a score. Recursive feature elimination algorithm is an example of the wrapper feature selection method. The embedded feature selection method selects the best features during the model creation. Random forest feature importance and regularization methods are the most common type of embedded feature selection techniques.

2.3.4 Training and Classification

The final stage of HAR is model training, and activity classification. In this stage, similar to any other pattern recognition, the activity dataset is divided into two subsets: a training set and a testing set. The training dataset is designed for training the classifier, while the test dataset is for testing the performance of the classifier (Soleimani & Nazerfard, 2021). After that, the human activity classifier will be trained and tested (Zhaohui Wang et al., 2018).

There are several machine learning and deep learning algorithms available for this purpose. This study experiments on the most common machine learning and deep learning classifiers in activity recognition for subject variability investigation and usability evaluation of the proposed methods. It applies single machine learning classifiers, including logistic regression, support vector machine (SVM), and decision trees, as well as ensemble classifiers, including random forest (RF), adaboost, and extremely gradient boosting trees (XGBoost). This is because these classifiers are appropriate for resource-limited devices due to their low memory and computational requirements. Among these machine classifiers, support vector machine (SVM) ranked the most used machine learning model for sensor-based activity recognition followed by k-nearest neighbor (kNN), decision tree (DT), and random forest (RF). It also experimented with convolutional neural networks (CNN) which ranks the most used deep learning model for sensor-based activity recognition. Figures 2.2 and 2.3 show the distribution of machine learning and deep learning models for activity recognition respectively (Abdu et al., 2021).

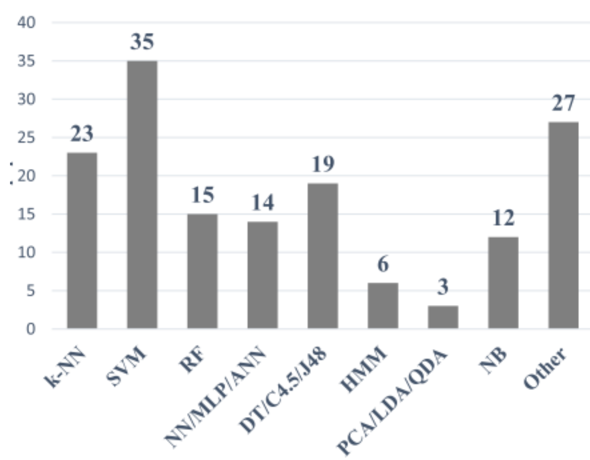

Figure 2.2 Distribution of machine learning classifiers for activity recognition