

Fall 2022

Human Activity Recognition (HAR) Using Wearable Sensors and Machine Learning

Chrisogonas Otero Odhiambo

Follow this and additional works at: <https://scholarcommons.sc.edu/etd>



Part of the [Computer Sciences Commons](#)

Recommended Citation

Odhiambo, C. O.(2022). *Human Activity Recognition (HAR) Using Wearable Sensors and Machine Learning*. (Doctoral dissertation). Retrieved from <https://scholarcommons.sc.edu/etd/7057>

This Open Access Dissertation is brought to you by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact digres@mailbox.sc.edu.

HUMAN ACTIVITY RECOGNITION (HAR) USING WEARABLE SENSORS AND MACHINE LEARNING

by

Chrisogonas Otero Odhiambo

Bachelor of Science
Maseno University, 2004

Master of Science
University of South Carolina, 2008

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Computer Science

College of Engineering and Computing

University of South Carolina

2022

Accepted by:

Homayoun Valafar, Major Professor

Marco Valtorta, Committee Member

Michael Huhns, Committee Member

Forest Agostinelli, Committee Member

Cindy Corbett, Committee Member

Ray Bai, Committee Member

Cheryl L. Addy, Interim Vice Provost and Dean of the Graduate School

© Copyright by Chrisogonas Odera Odhiambo, 2022
All Rights Reserved.

DEDICATION

I wish to dedicate this work to all who have made it possible for me to get this far in my academic journey. I thank Almighty God for sound health and sane mind. I acknowledge my family – my wife, Everlyn Constance Nyaoro, and the children (Gracie Maya Achieng’ ‘Mama Gee’, Shanice Adhiambo Ochanji ‘Nyakadero’, and Griffin Elvis Odero ‘Jatelo’) – for their patience, love, and unwavering support throughout the journey. I recognize the support of all other family members who have been part of the story, one way or the other. I acknowledge the love and support from my angel mother (Grace Mary Atieno), though no longer with us. Equally, I recognize the support and encouragement from my father (Philip Odhiambo Okong’o), as well as all the siblings. I dedicate this work to all the teachers and professors who have been instrumental in my academic journey, including but not limited to, my greatest primary school Math and English teachers, Mr. Caleb Ochieng’ Okode (Bongu Pr. School) and the late Mr. Benson Anuro Ngicho (Kachar Pr. School), as well as my high school Director, and mentor, the late Dr. Geoffrey William Griffin, Founder of Starehe Boys’ Center.

Special mention to incredible brothers, families, and mentors, Dadaji (Paul Khelli) and Didi Doggette, and Lore Minudi. Their support has been authentic and unwavering. I cannot thank enough, Francine Platt, of Salt Lake City, UT, my Columbia, SC, hosts for the last five years, Mama Sara and Frank Santoro, and SC friends Toni Jones and Eileen Newman. Your kindness and friendship offered great support during this grueling journey.

ACKNOWLEDGEMENTS

This dissertation is a consummation of very many fronts. First and foremost, I wish to appreciate the support of all the members of ValafarLab, but particularly the colleagues we have worked with in the Human Activity Recognition (HAR) studies. I thank Andrew Smith, Musa Azeem, Luke Ablonczy, and Casey Cole. Thanks to every Machine Learning enthusiast in the lab, led by my good friend, Brendan Odigwe.

Very special thanks to Dr. Michael Huhns, my Master's Theses Committee Chair, and Dr. Homayoun Valafar, my Ph.D. Advisor and the incumbent Department Chair. Special mention to Dr. Valtorta, the incumbent Graduate Director, for being very flexible and accessible. I hugely thank Dr. Cindy Corbett, Director ACORN Center, College of Nursing, and her team (Dr. Pamela Wright and Sydney Reichardt, among others) for their support and collaboration. Thanks to Dr. Bai of Statistics Department and Dr. Agostinelli of the AI Institute for being a great resource.

Finally, I wish to register immense appreciation for the support of my entire committee members in a special way. Without their meticulous and rigorous tutelage, this work would not be possible. I feel privileged to have benefited from their combined diverse wealth of experience.

I thank all friends and amazing professors at The University of South Carolina who have been part of my academic and social life have been.

ABSTRACT

Humans engage in a wide range of simple and complex activities. Human Activity Recognition (HAR) is typically a classification problem in computer vision and pattern recognition, to recognize various human activities. Recent technological advancements, the miniaturization of electronic devices, and the deployment of cheaper and faster data networks have propelled environments augmented with contextual and real-time information, such as smart homes and smart cities. These context-aware environments, alongside smart wearable sensors, have opened the door to numerous opportunities for adding value and personalized services to citizens. Vision-based and sensory-based HAR find diverse applications in healthcare, surveillance, sports, event analysis, Human-Computer Interaction (HCI), rehabilitation engineering, occupational science, among others, resulting in significantly improved human safety and quality of life.

Despite being an active research area for decades, HAR still faces challenges in terms of gesture complexity, computational cost on small devices, and energy consumption, as well as data annotation limitations. In this research, we investigate methods to sufficiently characterize and recognize complex human activities, with the aim to improving recognition accuracy, reducing computational cost and energy consumption, and creating a research-grade sensor data repository to advance research and collaboration. This research examines the feasibility of detecting natural human gestures in common daily activities. Specifically, we utilize smartwatch accelerometer sensor data and structured

local context attributes and apply AI algorithms to determine the complex gesture activities of medication-taking, smoking, and eating.

This dissertation is centered around modeling human activity and the application of machine learning techniques to implement automated detection of specific activities using accelerometer data from smartwatches. Our work stands out as the first in modeling human activity based on wearable sensors with a linguistic representation of grammar and syntax to derive clear semantics of complex activities whose alphabet comprises atomic activities. We apply machine learning to learn and predict complex human activities. We demonstrate the use of one of our unified models to recognize two activities using smartwatch: medication-taking and smoking.

Another major part of this dissertation addresses the problem of HAR activity misalignment through edge-based computing at data origination points, leading to improved rapid data annotation, albeit with assumptions of subject fidelity in demarcating gesture start and end sections. Lastly, the dissertation describes a theoretical framework for the implementation of a library of shareable human activities. The results of this work can be applied in the implementation of a rich portal of usable human activity models, easily installable in handheld mobile devices such as phones or smart wearables to assist human agents in discerning daily living activities. This is akin to a social media of human gestures or capability models. The goal of such a framework is to domesticate the power of HAR into the hands of everyday users, as well as democratize the service to the public by enabling persons of special skills to share their skills or abilities through downloadable usable trained models.

TABLE OF CONTENTS

DEDICATION	iii
ACKNOWLEDGEMENTS	iv
ABSTRACT.....	v
LIST OF TABLES	x
LIST OF FIGURES	xii
LIST OF SYMBOLS	xv
LIST OF ABBREVIATIONS.....	xvi
CHAPTER 1: INTRODUCTION	1
1.1 Dissertation focus	3
1.2 Problem Purpose Statement	4
1.3 Research Challenges and specific objectives	5
1.4 Significance and Broader Impact	5
1.5 Innovation.....	8
1.6 Dissertation organization	8
1.7 Literature Review	10
1.8 Application Areas.....	12
1.9 Data Sources.....	24
1.10 Techniques	25
1.11 Challenges and Limitations	25
CHAPTER 2: STATE TRANSITION MODELING OF THE SMOKING BEHAVIOR USING LSTM RECURRENT NEURAL NETWORKS	30
2.1 Introduction	31
2.2 Background and Method	33
2.3 Neural Network Platform and Architecture	36
2.4 Results and Discussion.....	39
2.5 Conclusion.....	45

CHAPTER 3: MEDSENSOR: MEDICATION ADHERENCE MONITORING USING NEURAL NETWORKS ON SMARTWATCH ACCELEROMETER SENSOR

DATA	47
3.1 Introduction	48
3.2 Background	49
3.3 Data Collection and Method	50
3.4 Results and Discussion.....	57
3.5 Conclusions	62

CHAPTER 4: HUMAN ACTIVITY RECOGNITION ON TIME SERIES ACCELEROMETER SENSOR DATA USING LSTM RECURRENT NEURAL NETWORKS

4.1 Introduction	66
4.2 Background and Method	67
4.3 Neural Network Platform and Architecture	71
4.4 Results and Discussion.....	75
4.5 Conclusion.....	80

CHAPTER 5: DETECTING MEDICATION GESTURES USING MACHINE LEARNING AND ACCELEROMETER DATA COLLECTED VIA SMARTWATCH TECHNOLOGY: A FEASIBILITY STUDY

5.1 Abstract	82
5.2 Introduction	84
5.3 Methods	87
5.4 Development of the Artificial Neural Network.....	93
5.5 Results	97
5.6 Discussion	103
5.7 Conclusions	106

CHAPTER 6: TOWARD CONCURRENT IDENTIFICATION OF HUMAN ACTIVITIES WITH A SINGLE UNIFYING NEURAL NETWORK CLASSIFICATION: FIRST STEP

6.1 Abstract	109
6.2 Introduction	111
6.3 Methodology	117
6.4 Results	138

6.5 Discussion	150
6.6 Conclusions	153
CHAPTER 7: CONCLUSION	156
REFERENCES	160
APPENDIX A: Permission for Reprint	176

LIST OF TABLES

Table 2.1: Class assignment and one-hot encoding for each sub-gesture.....	35
Table 2.2: Mini-gesture detection using different ANN architectures	40
Table 2.3: The confusion matrix of the most optimal ANN architecture.	41
Table 2.4: Confusion matrix report for the most optimal ANN architecture.....	42
Table 2.5: A summary of LSTM’s performance in detection smoking mini-gesture as a function of different architectural parameters.....	43
Table 2.6: The confusion matrix of the most optimal LSTM architecture where Batch=100, Units=3, Activation=sigmoid, epochs=5000.....	45
Table 2.7: Confusion matrix report for the most optimal architecture of LSTM.	45
Table 3.1. Protocol-guided medication-taking activity.....	52
Table 3.2. Results of a two-layer ANN as a function of hidden neurons	58
Table 3.3: Training Accuracy Results of ANN training using a bootstrap approach after experimenting with 10 different hidden layer sizes for each excluded participant	60
Table 3.4: Testing Accuracy Results of ANN training using a bootstrap approach after experimenting with 10 different hidden layer sizes for each excluded participant.	60
Table 3.5: The average performance of ANNs for each architecture.	61
Table 3.6: Performance by participant.....	62
Table 4.1: Summary of all the datasets used in the study	76
Table 4.2: Metrics for the best performing configuration of window size 150	79

Table 4.3: Confusion matrix for the window size of 150 units	80
Table 5.1: Summary of all the datasets used in the study	92
Table 5.2: Time in seconds required for natural and scripted/protocol gestures.	99
Table 5.3: Outliers count for natural and scripted medication gestures, based on gestures longer than 100 seconds or shorter than 8 seconds as outliers.	99
Table 5.4: Analysis of upper-category outliers in seconds.	99
Table 5.5: Metrics for the best performing configuration of window size 150	101
Table 5.6: MTP/Non-MTP Confusion matrix for the window size of 150 units.....	102
Table 5.7: MTP/Non-MTP Confusion Matrix (in %).....	103
Table 6.1: Summary of all the datasets used in the study	128
Table 6.2: Visualization of pizza bite, eating salad with a fork, smoking and jogging activities	141
Table 6.3: Statistical description of natural and scripted/protocol gestures.	144
Table 6.4: Outliers count for natural and scripted medication gestures considered from gestures longer than 100 seconds or shorter than 8 seconds.....	145
Table 6.5: Analysis of upper-category outliers.....	145
Table 6.6: Summary metrics for the best performing configuration of window size 150 - Excludes outliers.....	147

LIST OF FIGURES

Figure 1.1: Taxonomy of Human Activities (based on [7] and [8])	4
Figure 1.2: Remote health monitoring architecture based on wearable sensors	7
Figure 1.3: Taxonomy of Human Activity Recognition (based on [1], [8], [13])	11
Figure 1.4: Frequency of application areas by HAR [7]	12
Figure 1.5: Frequency of data sources by existing literature on HAR [7]	24
Figure 1.6: Frequency of algorithms used in the existing literature on HAR [7]	25
Figure 2.1: An example of a smoking puff is shown where red indicates the X dimension of the accelerometer data, green the Y and blue the Z	34
Figure 2.2: An example of sub-gesture annotation of a puff where the first box denotes the hand-to-lip gesture and the second box the hand-off-lip gesture	35
Figure 2.3: A snippet of Python code used in Keras to define the used ANN	37
Figure 2.4: A snippet of Python code used in Keras to define the used LSTM	38
Figure 2.5: Most optimal ANN architecture loss function as a function of epochs	41
Figure 2.6: The measure of accuracy as a function of epochs for the training set (blue) and validation set (yellow)	41
Figure 2.7: Loss function of the most optimal LSTM architecture as a function of epochs, illustrated by blue and orange for training-loss and validation-loss, respectively	43
Figure 2.8: The measure of accuracy of the most optimal LSTM architecture	44
Figure 2.9: Comparing LSTM/ANN loss.	44

Figure 2.10: Comparing LSTM/ANN Accuracy.	44
Figure 3.1. MedSensor interface used for annotation at the edge.....	51
Figure 3.2: Visualization of a full medication gesture from one participant.	55
Figure 3.3: Visualization of the superimposition of three medication gestures from one participant.	55
Figure 3.4: Artificial Neural Network high-level architecture	57
Figure 4.1: Illustration of sensor axes on a typical smartphone and smartwatch.	69
Figure 4.2: The LSTM cell can process data sequentially and keep its hidden state through time (By Guillaume Chevalier - File: The_LSTM_Cell.svg, CC BY-SA 4.0) ...	72
Figure 4.3: An entire recording of consuming pizza consisting of 6 bites that in total took 147 seconds.	76
Figure 4.4: An eating gesture consisting of a single bite of pizza.	77
Figure 4.5: Single smoking gesture	77
Figure 4.6: A single medication-taking gesture that consists of multiple sub-events (opening bottle, dispensing pill, taking pill, drinking water, etc.)	78
Figure 4.7: Three consecutive steps during jogging activity.	78
Figure 4.8: Training plot for the window size of 150 units whose configuration produced the best performance among the different models.	79
Figure 5.1: An illustration of Smartphone and smartwatch accelerometer axes	89
Figure 5.2: The LSTM cell can process data sequentially and keep its hidden state through time	94
Figure 5.3: MTE complex activity atomic segmentations	97
Figure 5.4: Illustration of MTP intra-class differences.....	97

Figure 5.5: Training plot for the window size of 150 units.	101
Figure 6.1: General process of activity recognition model preparation.....	118
Figure 6.2: High-level framework of the data process from data origination to storage, to retrieval, and analysis.....	123
Figure 6.3: HAR as Personalized AI Models (HARP) high-level framework	124
Figure 6.4: An illustration of Smartphone and smartwatch accelerometer axes	125
Figure 6.5: Gesture context components	131
Figure 6.6: Complex activity signature.....	132
Figure 6.7: The repeating module in an LSTM contains four interacting layers [223] ..	135
Figure 6.8: The segments correspond to (A) open-bottle dispense-medicine, (B) Hand-to- mouth pill-to-mouth hand-off-mouth, (C) pick-up-water drink-water lower-cup-to-table close-bottle, and (D) post-medication.....	139
Figure 6.9: Illustration of MTP differences between a) scripted gesture from user1, b) natural gesture from user1, and c) natural gesture from user2.	140
Figure 6.10: Signal distribution along x-axis.....	142
Figure 6.11: Signal distribution along y-axis.....	142
Figure 6.12: Signal distribution along z-axis.....	143
Figure 6.13: Training plot for the window size of 150 units.	147
Figure 6.14: Training plot for the window size of 150 units	148
Figure 6.15: Confusion matrix for the window size of 150 units	149

LIST OF SYMBOLS

\odot	Denotes the Hadamard product (element-wise product).
σ	Standard deviation
\prod	Product notation for product summations.
T	Matrix transpose operation.
\in	Set membership i.e., some element on left is a member of set on the right.
\mathbb{R}	Set of real numbers
β	Weighting factor and a positive real number

LIST OF ABBREVIATIONS

HAR	Human Activity Recognition
HARP	HAR Personalized Models
CCTV	Closed-circuit Television
I.I.D	Independent Identical Distribution
MTE	Medication Taking Event
MTP	Medication Taking Protocol
nMTE	Non-Protocol MTE
sMTE.....	Scripted or Protocol MTE
ML.....	Machine Learning
AI	Artificial Intelligence
TFX.....	TensorFlow Extended
ANN	Artificial Neural Network
LSTM.....	Long Short-Term Memory
CNN	Convolutional Neural Network
CNN-DBNN	CNN- Discriminative Deep Belief Network
JITAI.....	Just-in-time adaptive interventions
YOLO	You Only Look Once real-time object-detection system

CHAPTER 1: INTRODUCTION

Human Activity Recognition (HAR) has been treated as a typical classification problem in computer vision and pattern recognition to recognize a variety of human activities in daily living, as well as anomalies. Anomaly refers to abnormal behavior or activities. Recent technological advancements, the miniaturization of electronic devices and the deployment of cheaper and faster data networks have propelled environments augmented with contextual and real-time information, such as smart homes and smart cities. These context-aware environments, alongside smart wearable sensors, have opened the door to numerous opportunities for adding value and personalized services to citizens. Vision-based HAR techniques rely on image and video data to recognize behavior while Sensor-based counterparts rely on sensor data to achieve the same goal. Sensors are converters that quantify the physical aspects of the world around us into electric values that can be perceived by a digital system [1]. This capability makes it possible to gain knowledge about human activities. This study considers the use of wearable sensors. Sensor-based HAR generally broadly comprises five steps: sensor selection, data collection, feature extraction, model training and testing[2].

A machine learning problem is focused on learning abstract relationships that allow consistent generalization when new samples are provided. HAR being a pattern recognition problem of specific actions uses classifiers and action detection methods. It aims to

understand daily behaviors of people through the analysis of observation sensor or vision data obtained from people and their context.

Over the past three decades, there has been a steady advancement, availability, and proliferation of miniature wearable sensor devices from necklaces to smart phones, smartwatches, etc. Human Activity recognition has so far found use in diverse domains such as healthcare, surveillance, sports, linguistics, event analysis, Human Computer Interactions (HCI)[3], among others. HAR has been used to characterize human behavior, understand human interactions, improve quality of life such as the case of continuous health monitoring, improving human safety and well-being, all over the world. Other areas of application include activity of daily living, physiological signals, quantified-self, postures detection, gestures detection, gait analysis, and indoor localization [4]–[6].

Wearable sensors come with significant benefits such as personalized health monitoring, removal of healthcare barrier, allowing for evenly distributed health monitoring, and affordability by majority populations. These benefits can be exploited for social and economic good such as remote healthcare monitoring and response systems. Correct interpretation of motions such as falls, medication, eating, smoking, etc. can be used to trigger appropriate response of caregivers, translating to reduced risks for the elderly and other vulnerable persons from potentially catastrophic tumbles, poor medication adherence, unhealthy eating habits, or inactive lifestyle activities. Accurate assessment of health behaviors in humans is necessary to evaluate health risk, and therefore, effectively intervene to facilitate behavior change, improve health, and reduce disease risk.

The next section of this work spells out the dissertation focus, problem purpose statement, challenges and specific objectives, significance and broader impact, contributions, innovation, as well as the analysis of relevant recent overarching literature. While the work

focuses on wearable sensors, it is imperative to comprehend and appreciate the place of Sensor-based HAR in the bigger scope of HAR.

1.1 Dissertation focus

In this research, we begin by focusing on understanding the challenges in developing an automated system to recognize complex human activities. Based on this, we then consider a complex activity recognition framework as well as various approaches to recognize the signature of complex human activity based on the sequencing and contextualization of atomic activities that form the basic building blocks of complex activities. We further describe our design and development of a smartwatch model-based activity recognizer. Our final study focuses on the designing and implementation of models that recognize complex functional activities using sensor triaxial values, and simple atomic human activity context. This study focuses on detection and identification of human activities coded dark maroon at the leaves of Figure 1.1.

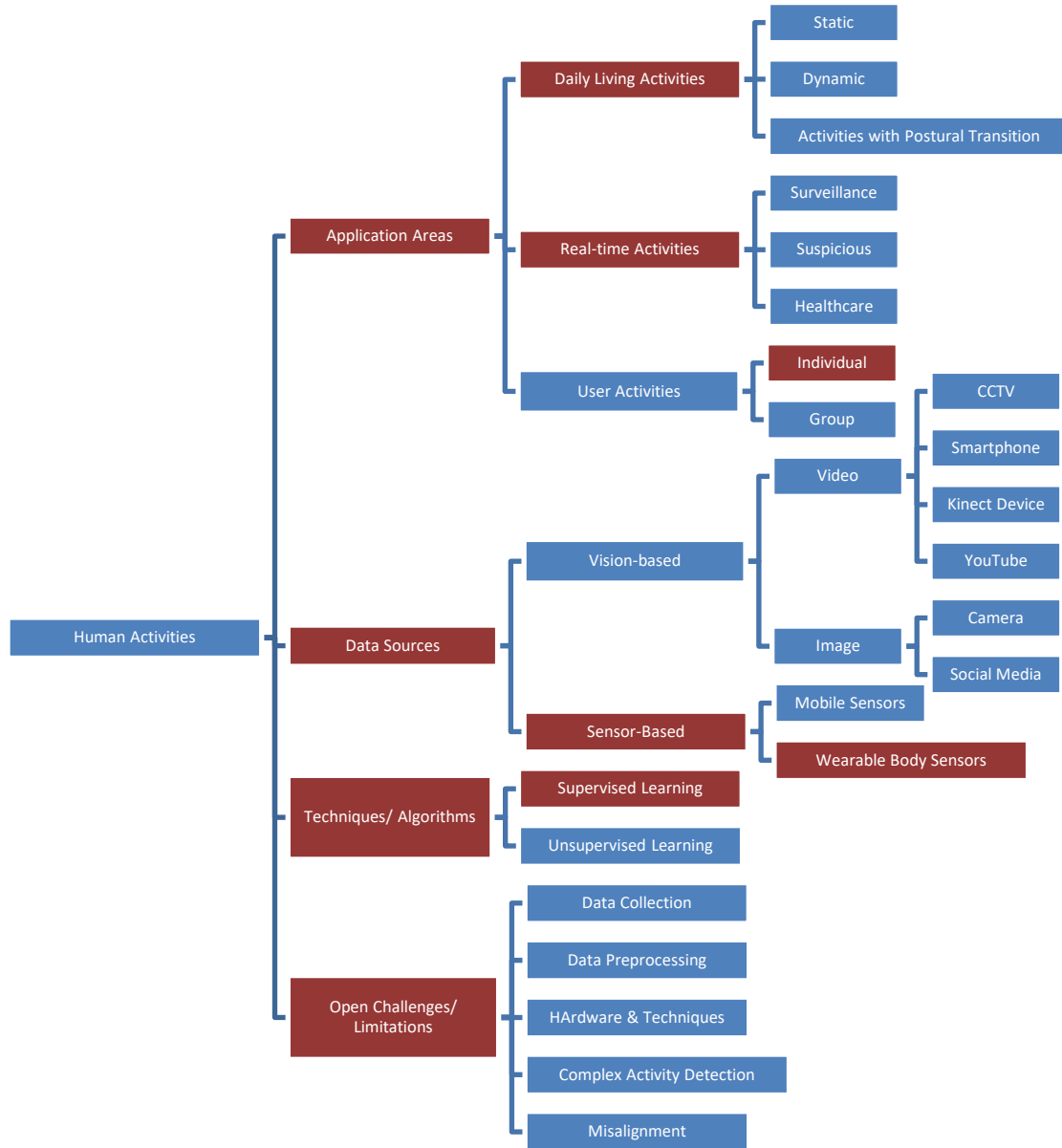


Figure 1.1: Taxonomy of Human Activities (based on [7] and [8])

1.2 Problem Purpose Statement

With the growth of IoT and proliferation of miniature affordable consumer sensor devices like smartwatches and smartphones, there is a ready platform to deploy tools that can take advantage of these sensors to provide useful data to improve healthcare services for vulnerable populations. The purpose of this study is to examine, evaluate, and implement machine learning models that can recognize signatures for various human activities,

making it possible for the use of wearable sensors in areas such as monitoring seniors living at home for falls, late-night activity, sleeping habits, medication adherence, healthy eating, or smoking habits. Although significant body of research has been done in the domain of HAR, but little has been done in complex human activity recognition based on the signature recognition of the sequenced sum recognition of the basic atomic activities.

1.3 Research Challenges and specific objectives

Objective #1: Develop a comprehensive, research-grade data acquisition and dissemination system, optimized for collection of sensor data from broad-range wearable devices.

Objective #2: Baseline gesture detection. Explore the effectiveness of CNN and LSTM neural networks to extract fine-grained features for detection and recognition of medication, smoking, eating, and exercise events. Design the framework of HAR as Personalized AI Models (HARP).

Objective #3: Evaluation of the proposed models on multiple different real datasets, using actual device to detect medication, smoking, eating, and selected exercise activities in a lab setting.

Objective #4: Analysis of device resource utilization, with a focus on memory, CPU, and battery life, based on number and type of active sensors, sampling rate and window size

1.4 Significance and Broader Impact

Technology: Understanding human behavior and its context is significant because it provides us with meaningful understanding of various social and health challenges. Certain human behaviors can be personally or/and socially harmful or beneficial. Smoking is harmful directly to the smoker, and indirectly to the secondary smokers. Smoking remains

a leading cause of preventable deaths in the United States. Overeating or poor eating can lead to obesity, diabetes, and sleep problems which can in turn result into cognitive disorders. Lack of regular exercise can lead to cardiovascular diseases. These behaviors are often induced by contexts. Our fundamental development of software technology and application of AI for recognizing human behaviors in context will enable the future investigations of these, as well as the social means for mitigating their harmful effects.

Importantly, taking advantage of commonplace wearable sensor technology such as smartwatches, our work stands great chance of improving existing healthcare solutions. Remote monitoring of an elderly, or any vulnerable dependent subject can be easily automated as is illustrated by the Figure 1.2 [9] in the 2015 study by Attal *et al.* The home supportive environment delivers trend data and detection of incidents using non-intrusive wearable sensors. This in turn facilitates a quick measurement and fast acceptance at the same time. Through real-time processing and data transmission, healthcare suppliers can monitor the subject's motions during daily activities as well as detect unpredictable events that may occur. The subject's records can be used in medical decision support, prediction, and prevention. Detecting and responding to life-threatening health situations such as falls among the elderly or children, detecting seizures and tremors in epilepsy patients [10]–[12] and intervening with timely communication to prevent injuries as well as respond with emergency services, are some of the ways in which our research can add significant impact in the healthcare space.

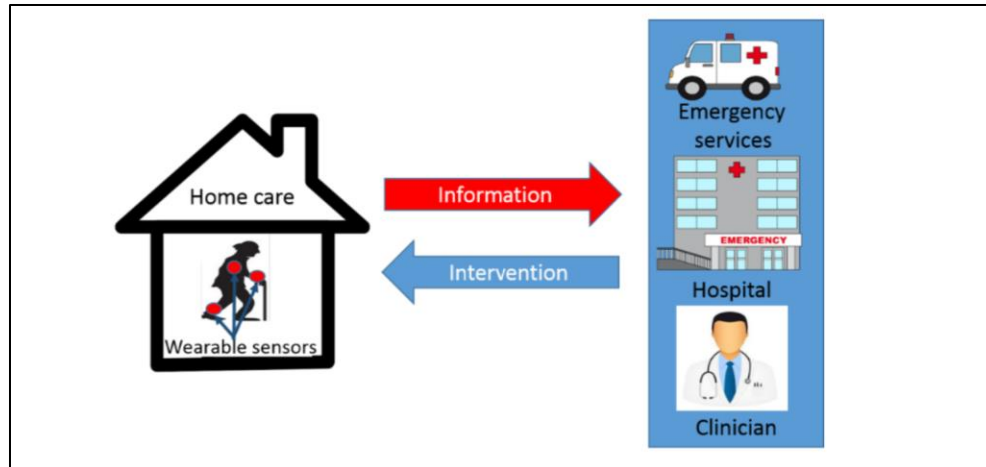


Figure 1.2: Remote health monitoring architecture based on wearable sensors

Dissemination: This research has produced a variety of rich HAR datasets, a comprehensive suite of tested software tools related to data acquisition, annotation, and management for smart watches. We have implemented a robust HAR database available to the public to support research and collaboration. We will distribute our software through Google App Store and a public version control sites such as Bitbucket or GitHub. The publicly available data and software will enable a wide range of investigations into the context-based detection of human behaviors, and the social interactions among humans.

Electronic documentation, tutorials, and representative data created will be made accessible to the international community through our research websites. All related publications are publicly available as a contribution to research. The HAR database also includes discussion features around datasets to make it possible for the online consumers to engage more directly in the interrogation of the data or related discussions among themselves or share feedback.

Energy Efficiency: Finally, our work in Objective #4 generates useful insights to enable end users and application developers to efficiently calibrate miniature sensor devices for better battery life. Good resource calibration translates into better performance.

1.5 Innovation

This research is novel in multiple ways:

- a) We tap into existing sensor data of smartwatches to analyze common complex human activities using modern predictive tools and technologies to understand relevant human behavior as well as create practical software tools to solve everyday problems. Based on the sensor data insights alongside other supportive characteristics such as time and context, we have created sophisticated Neural networks capable of extracting fine-grained features of human activity for detection and subsequent recognition. We investigate and implement machine learning models that recognize medication-taking gestures, smoking gestures, eating human activities. Through this, we have generated comprehensive knowledge of human behavior that can be exploited further for purposes of enhancing remote healthcare services, smoking cessation solutions, personal wellness assistance, and healthy eating support.
- b) Through assisted edge-annotation of self-activity data, we provide seamless aggregation and integration of user personalized data to centrally shareable repository and AI engine to train and deploy models, which can be refined generic baseline models or new personalized models. Based on this outcome and the proposed HARP Framework, the idea can be extended to serve *HAR as Personalized AI Models* (HARP) on affordable commodity sensor devices.
- c) This study turns affordable commodity devices, smartwatches in this case, to generate important human activity data, which is exploitable for useful analytics to further research, or provide a basis for custom HAR tools for smart devices users.

1.6 Dissertation organization

The dissertation is organized as follows:

In chapter 2, we examine state transition modeling of the smoking complex behavior using LSTM recurrent neural networks. We focus on modeling the smoking activity and the use of smartwatch sensors to recognize the activity. More specifically, we demonstrate how we reformulated one of our previous works in detection of smoking to include in-context recognition of smoking. We present a reformulation of the smoking gesture as a state-transition model comprising mini-gestures.

In chapter 3, we further examine the use of smartwatch sensors in medication adherence monitoring using neural networks on smartwatch accelerometer sensor data. We also consider the data process from watch to preprocessing, and network training. We demonstrate the use of smart applications to perform semi-annotation of data.

In chapter 4, we examine the problem of human activity recognition on time series accelerometer sensor data using LSTM recurrent neural networks. The use of sensors available through smart devices has pervaded everyday life in several applications including human activity monitoring, healthcare, and social networks. In this study, we focus on the use of smartwatch accelerometer sensors to recognize eating events from a dataset of ten participants.

In chapter 5, we look at a feasibility study on detecting medication gestures using machine learning and accelerometer data collected via smartwatch technology. We examine the use of wearable sensor devices in the recognition of human gestures and application of the same in addressing the costly problem of Medication adherence. Medication adherence is a complex human behavior associated with chronic condition self-management.

In chapter 6, we investigate and implement a first step in concurrent identification of activities with a single unifying neural network classification approach based on time series sensor data. We also examine activity modeling in a linguistic parallel. Finally, we explore a theoretical framework for domesticating and democratizing HAR models to smart devices users.

Chapter 7 is a conclusion that converges all the studies through a common framework as well as lays out the areas for future work.

1.7 Literature Review

The analysis of literature that puts this study into perspective is based on the conceptual framework illustrated by Figure 1.3. Just like in Figure 1.1, the dark maroon color in represents the HAR focus of this research. The Green color represents a partial focus of this research. We discuss the use of GPS data to determine context in smoking behavior, but we do not widely exploit the feature in determining context of human activity in the other human activity studies, such as medication, eating, and jogging exercise.

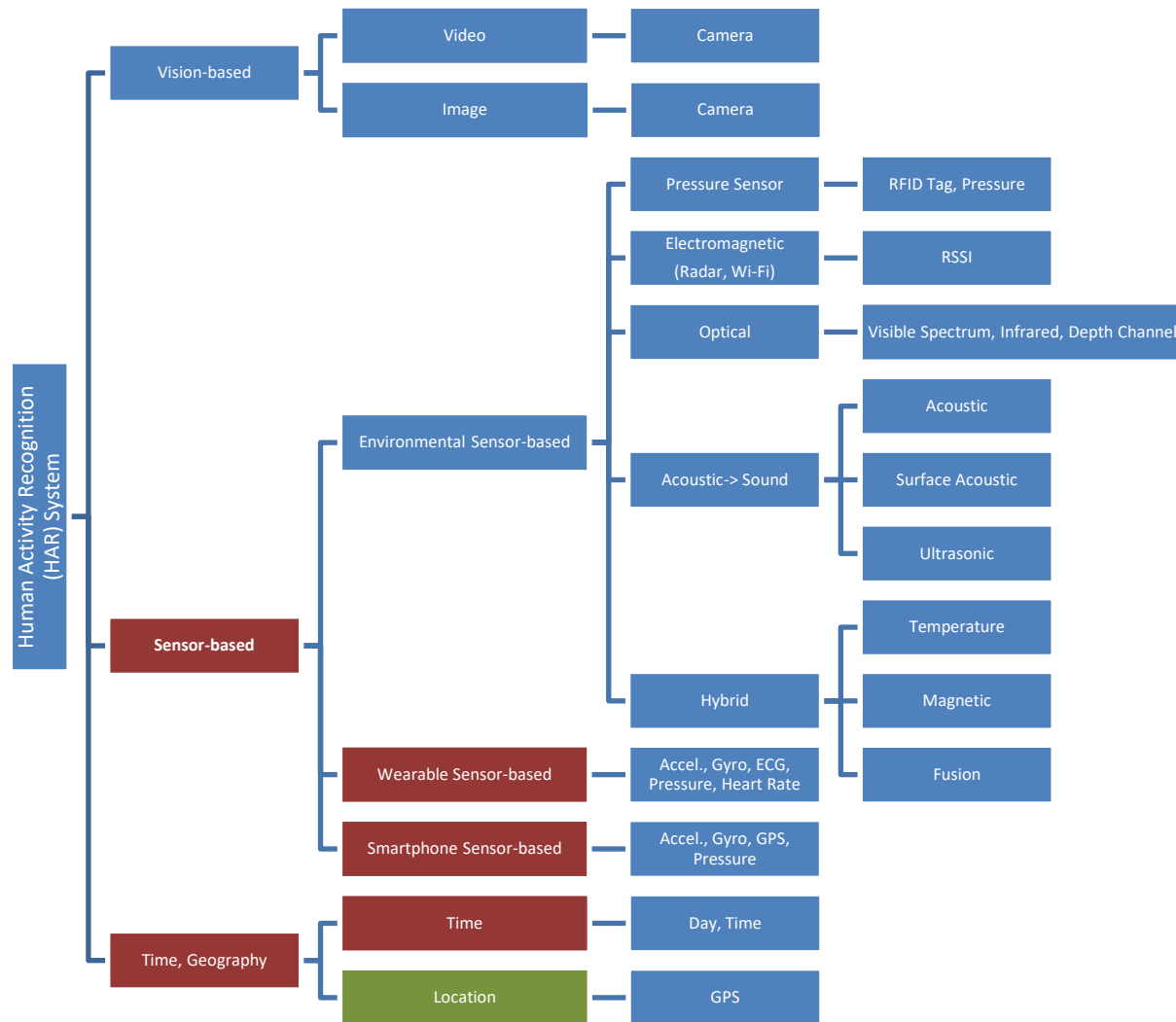


Figure 1.3: Taxonomy of Human Activity Recognition (based on [1], [8], [13])

1.8 Application Areas

The study [7] broadly classified HAR application areas under different categories of activities: daily living, real-time and user activities (individual and group-based). This classification is further illustrated by the work of A. Das Antar in [8]. According to Arshad, *et al* study, data showed that dynamic activities had the highest frequency of application, followed by static and group activities, as illustrated by Figure 1.4. This also shows that there is still immense room to come up with innovative HAR solutions in the healthcare domain.

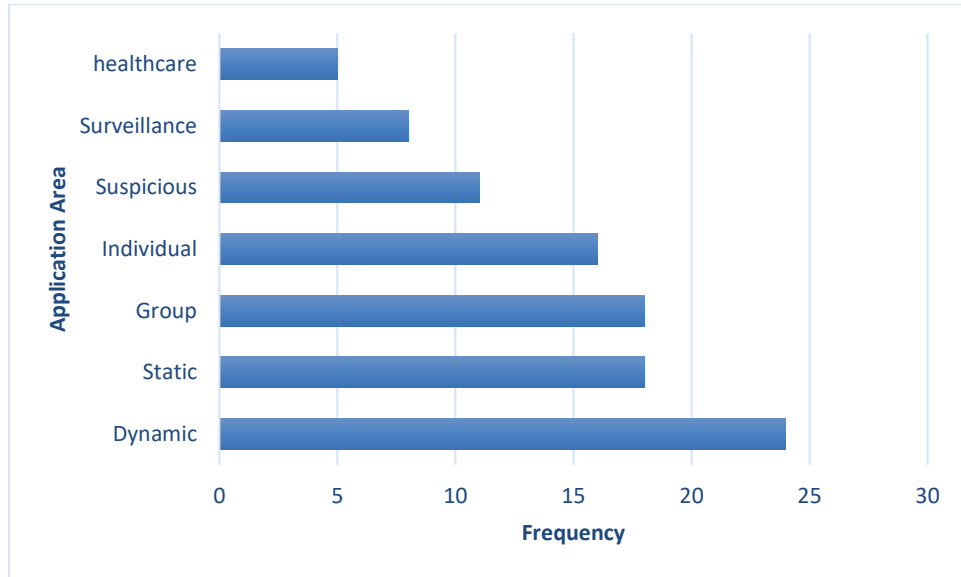


Figure 1.4: Frequency of application areas by HAR [7]

1.8.1 Daily Living Activities

Static Activities

The study[14] proposed SVM based framework to collect and persist data in a central repository. They used smartphones and eight different sensors. The data was then encoded into a feature vector. The proposed network showed, through experiments, its capacity to detect real-time static and dynamic activities with an accuracy of 87.1%. In another

study[15], Hasan *et al* introduced and trained a Deep Belief Network (DBN) and trained it on the same features. The DBN outperformed both the SVM and ANN for HAR. The work by Sukor *et al.* [16] achieved much superior performance of 96.11% accuracy using Principal Component Analysis (PCA) on a publicly available dataset. The study applied PCA to extract relevant features from mobile phone's tri-axial accelerometer and gyroscope sensor data.

Data annotation presents a tedious exercise in the HAR process. Bota *et al.* [17] helped mitigate this process through a Semi-Supervised Active Learning (SSAL) approach to relatively automate the annotation process for HAR on self-training (ST). In our previous studies on medication and smoking behaviors [18], [19], we automated the annotation process through user activity demarcation during data acquisition or activity performance. The marked data was persisted to cloud repository, and later downloaded and parsed through a further automated script to rapidly generate fully annotated data. This process made it possible to rapidly process large amounts of datasets.

Zhu *et al.* [20] proposed a semi-supervised deep learning approach to implement temporal ensemble of deep LSTM (DLSTM) on labeled and unlabeled data. In the study [21], Du *et al* proposed a three-stage framework that uses RFID tags for recognizing and forecasting HAR with LSTM: post-activity recognition, recognition in progress, and in advance prediction. Smart homes show enormous potential in healthcare, power-saving, etc. Further, it enables the operation of smart services according to human mind.

Chelli and Patzold [22] proposed a multi-activity recognition system. The system relied on mobile phone data comprising time and frequency domain features from acceleration and angular velocity. [23] introduced a pattern-based semi-supervised deep

recurrent convolutional attention network (RCAM) with wearable sensors to address the limitations and imbalanced distribution of labelled data as well as challenges of multimodal sensor data. Javed *et al* [24] proposed a 2-axis accelerometer Multilayer Perceptron (MLP) classifier to predict physical activities using smartphone accelerometer sensor data. The proposed model achieved 93% weighted accuracy using a publicly available WSDM dataset.

[25] argued that CNN with images will not burden the modern devices. [26] Using CNN and LSTM, a framework CNN-LSTM Model was proposed for multiclass wearable user identification while performing various activities. [27] and [28] applied some form of LSTM in their approaches. [27] Proposed a Contrastive Predictive Coding (CPC) framework based on CNN and LSTM for monitoring construction equipment activity. [28] Proposed a hybrid model that combines one-dimensional CNN with bidirectional LSTM (1D-CNNBiLSTM) to recognize individual actions using wearable sensors. To overcome limitations of some techniques such as ignoring data variability, having a large number of parameters, consuming a large number of resources, and being difficult to implement in real-time embedded devices, Pan *et al.* [29] employed the GRU network to address these issues. The approach collects valuable moments and temporal attention to minimize model attributes for HAR in the absence of I.I.D.

Dynamic Activities

Sani *et al* [30] proposed Coarse-to-Fine framework that uses Microsoft Kinect to capture activity sequences in 3D skeleton form, groups them into two forms, and then classifies them using the Bidirectional Long Short-Term Memory Neural Network (BLSTM-NN) classifier. Espinilla *et al* [31] proposed an online activity recognition with three temporal

sub-windows to recognize daily living activities such as showering, toilet, eating, etc. for predicting activity start time based on an activity's end label. The temporal sub-window registered an accuracy of 98.95% on the VanKasteren, Ordonez (<https://deeplearning.buzz/deep-learning-datasets/> (accessed on 16 July 2022)) dataset.

Alghyaline [32] proposed an approach based on YOLO object detection, Kalman Filter, and Homography to detect real-time static and dynamic activities from CCTV camera videos. Chen *et al.* [33] proposed an ensemble Extreme Learning Machine (ELM) algorithm to classify and recognize daily living human activities based on smartphone sensors, which used Gaussian random projection to initialize base input weights. Two experimental results achieved 97.35% and 98.88% accuracy on two public datasets. H. Ma *et al.* [34] proposed the AttnSense model to capture signal sensing dependencies with gyroscope and accelerometer sensors. A combination of CNN and a gated recurrent network (GRN) sense signals in spatial and temporal domains.

Almaadeed *et al.* [35] performed experiments using data representation from each person performing multiple activities in the same surveillance video, which is then used to detect the corresponding action. They used multiple human action recognition using 3Dimensional deep learning trained on KTH, Weizmann, and UCF-ARG datasets. Their experiment achieved 98% accuracy score. Gleason *et al.* [36] proposed a two-stage approach for HAR. The first stage generated dense spatio-temporal proposals on frame-wise object detection using hierarchical clustering and jittering techniques. Action classification and temporal refinement in untrimmed videos were performed in the second stage using the Temporal Refinement I3D (TRI-3D) network.

Wu *et al.* [37] presented AdaFrame, which included LSTM to select relevant frames for fast video recognition and reduced computational cost. Nadeem *et al.* [38] considered the significance of accurate detection of body parts in HAR. They proposed a framework that combined body part and discriminant analysis, with features extracted as displacement parameters that represent body part positions and processed using maximum entropy. Their experiment on the UCF dataset using the Markov model for markerless human pose estimation and physical activity recognition achieved 90.91% accuracy for body part detection. Ishikawa *et al.* [39] proposed the Action Segment Refinement Framework (ASRF) to improve performance on challenging datasets up to 13.7% in terms of segmental edit distance and 16.1% in terms of segmental F1 score. Temporal action segmentation is divided and refined into framewise action classifications with the Action Segmentation Branch (ASB) and action boundary regression with Boundary Regression Branch (BRB).

D'Arco *et al.* [40] used SVM to identify daily living activities by adjusting the size of the sliding window, reducing features, and implanting inertial and pressure sensors. Their system achieved an accuracy score of 94.66% while both sensors were used together. According to findings, inertial sensors are most suitable for dynamic actions, while pressure sensors are most suitable for static actions. Various researchers relied on pre-segmented sensor data to identify actions. However, Najeh *et al.* [41] relied on streaming sensors to perform HAR. They tried to establish whether the current action is a continuation of a previous action. They achieved this in three steps, namely: sensor correlation (SC), temporal correlation (TC), and determination of the activity activating the sensor.

1.8.2 Real-Time Activities

Surveillance

Jiang *et al.* [42] attempted to address real-time pedestrian detection by their proposal's approach that extracted static sparse features from each frame by feature pyramid and sparse dynamic features from successive frames to improve feature extraction speed, then combine them in Adaboost classification. [43] proposal spelled out a method to monitor traffic and unprecedented violence using CCTV cameras to identify object movements, and video synchronization to ensure proper detail alignment in the videos.

Basha *et al.* [44] approach employed CNN-DBNN to automatically track and detect criminal or brutal activities in videos. They used the CNN module to extract features from frames. The features were then subsequently passed to the Discriminative Deep Belief Network (DDBN). Experiment results showed increased accuracy at 90%. [45] proposed CNN model for multiple action detection, recognition, and summarization (such as two people fighting, walking, hand waving, etc.). This approach identified actions by comparing existing and generated Histogram of Oriented Gradients (HOG) of the frames of each shot's Temporal Difference Map (TDMap). Their experiments achieved an accuracy of 98.9% on a video dataset.

Qin *et al.* [46] also proposed a method for detecting and preventing criminal activities in shopping malls (DPCA-SM) using a video monitoring approach. The DPCA-SM could also trace people's routes and detect measures of store settings in real-time using surveillance cameras and generating alerts. The method evaluation yielded 92% accuracy in crowded conditions. Mahdi and Jelwy [47] proposed an approach for automatically detecting unusual situations in academic environments and alerting the appropriate authorities.

Suspicious Activities

Mohan *et al.* [48] used PCANet and CNN to overcome the challenges of manual detection of anomalies in videos and false alarms in public spaces. They employed PCA and SVM classifier to detect frame-wise anomalous occurrence. Tests on UCSD, UMN dataset, and Avenue Dataset showed significant performance score. [49] used Electronic Article Surveillance (EAS) systems, relaying CCTV real-time videos to CNN model to detect suspicious human activities in the store such as shoplifting, robbery, and break-in, and generate an alarm. This approach proved effective even where shoplifters removed labels from products.

Jyotsna and Amudha [50] used a deep learning approach based on a pretrained CNN model to obtain features from videos, followed by a feature classifier LSTM to detect behavior anomalies in an academic environment and alert the appropriate authorities. Pyataeva and Eliseeva [51] proposed a method for smoking event detection based on visual data, specifically leveraging video-based spatio-temporal features. Their method employed the ResNet three-dimensional CNN to recognize and detect smoking events. Riaz *et al.* [52] employed a pre-trained model to perform feature extraction from videos for pose estimation. The features were subsequently passed to a cascaded deep CNN to detect anomalies in examination halls.

Mudgal *et al.* [53] combined Gaussian Mixture Model (GMM) with the Universal Attribute Model UAM to distinguish between normal and abnormal activities such as hitting, slapping, and punching. W. Ullah *et al.* [54] approach employed a two-stream neural network using AIoT to recognize anomalies in Big Video Data. This method comprised a cloud component that received and analyzed frames for anomalies using a bi-directional long short-term memory (BD-LSTM) layer.

Healthcare

To address the limitation of Single sensing modality in a smart healthcare environment, Gumaei *et al.* [55] proposed a robust multi-sensor-based framework employing a hybrid deep learning model. The framework comprised simple recurrent units (SRUs) used to process the sequence of multimodal input data, and gated recurrent units (GRUs) used to store and learn from previous information. The framework recorded more than 90% accuracy performance on the MHEALTH dataset. Uddin and Hassan [56] employed a Deep CNN established on Gaussian kernel-based PCA to recognize activities based on features extracted from various body sensors. They tested their approach on the Mhealth dataset to determine its effectiveness and use for cognitive assistance.

While real-time monitoring can be performed using wearable sensors to recognize activity features like gait, falls, breathing, swallowing, etc., these devices can be a burden as well as source of discomfort to the wearer. Taylor *et al.* [57] attempted to offer a solution to this challenge; their method detected human motion with a quasi-real-time non-invasive method. They created a dataset of radio wave signals and developed an RF machine learning model to provide near-real-time classification between sitting and standing. [58] christened their approach as Ensem-HAR. They proposed a collection of models - “CNN-net”, “CNNLSTM-net”, “ConvLSTM-net”, and “StackedLSTM-net” models - based on one dimensional CNN and LSTM stacked predictions. They then trained a blender on them for final prediction.

Mobile health technology, including sensors worn on the body, can be used to passively and remotely collect and transmit objective data. These objective data can be much more valid and reliable compared to self-report, particularly for exercises such as walking [59]. The passive collection and transmission of data to researchers or clinicians

have other advantages, including a dramatic reduction in participant burden and the ability to process and provide feedback to participants automatically and in real-time or near real-time. This critical step provides a platform to develop and deliver ecological momentary interventions (EMI) [60] and just-in-time adaptive interventions (JITAI) [61]. EMI and JITAI deliver intervention strategies that are customized to address the specific needs of individual participants as soon as these needs are detected. Participant needs are identified by evaluation of the objective data from the remote sensors in real-time or near real-time. Indeed, EMI and JITAI can provide more automated and cost-effective approaches to intervene and improve health behavior remotely while maintaining efficacy [62]–[64].

1.8.3 User Activities

Individual Activities

Human behavior is a complex phenomenon in several aspects such as motion, magnitude, direction, and appearance. Hsu *et al.* [65] applied unsupervised learning, an SVM, and a Condition Random Field (CRF) to label video segments and detect anomalous events. Ko and Sim [66] proposed a deep convolutional framework to develop a unified framework for detecting behavior anomalies using LSTM in RGB images. They employed YOLO to determine the action of individuals in video frames, followed by VGG-16 classify them.

Territory or context transition is a challenge in the detection of HAR in live videos. [67] proposal provided a solution to this. The HOME FAST (i.e., Histogram of Orientation, Magnitude, and Entropy with Fast Accelerated Segment Test) spatiotemporal feature extraction approach based on optical flow information detects anomalies by obtaining low-level features with KLT feature extractor and then relaying the features to DCNN for

classification. To overcome the limitation of custom-designed algorithms of being fixed to detect only one specific type of behavior, Founta *et al.* [68] proposed a deep learning a unified architecture that utilized the available metadata and combined hidden patterns to detect multiple abusive norms that are highly interrelated. The architecture scored highly on performance at 98% on detecting behaviors like Cyberbullying, Hateful, Offensive, Sarcasm, and Abusive datasets. Dou *et al.* [69] used SVM to determine abnormal pedestrian behavior using extracted feature vectors and vector trajectories from the computed optical flow field of determined joint points determined by estimating the posture and optical flow field with a camera.

Moukafih *et al.* [70] proposed LSTM Fully Convolutional Network (LTSMFCN) to detect aggressive driving behavior sessions as time series classification with the aim of improving traffic safety. The LTSM-FCN scored 95.88% accuracy for a 5-min window length compared to other deep learning models on the UAH-DriveSet dataset collected using smartphone. Manual monitoring of CCTV video all of the time is an extremely difficult task. In a bid to address this challenge, Lee and Shin [71] proposed a deep learning model to automatically detect abnormal behavior such as assault, theft, kidnapping, drunkenness, etc. Their experiment results revealed I3D model as the best performer over the rest.

Xia and Li [72] applied a fully CNN and a pre-trained VGG-16 to extract static appearance features. The method used temporal attention mechanism to extract appearance features at the same position. They then used LSTM network to decode the features to predict feature anomalies in the moment to identify abnormal behavior in video frames. In a bid to address the challenge of the dynamism of human action, Bhagya Jyothi and

Vasudeva [73] proposed a Chronological Poor and Rich Tunicate Swarm Algorithm (CPRTSA)-based Deep Maxout Network to extract effective features to recognize human activities in different domains. Belhadi *et al.* [74] classified algorithms into two categories: The first performs data mining and knowledge discovery to investigate relationship between behaviors and identify anomalies; the second employs deep CNN that learns from historical data to detect the collective emergent abnormal behavior. Shu *et al.* [75] considered a graph LSTM-in-LSTM (GLIL) host-parasite architecture for group activity detection, which can be several person LSTM (P-LSTM) or graph LSTM (GLSTM). P-LSTM and GLSTM are based on interactions between persons. P-LSTM is integrated into G-LSTM, and residual LSTM learns person-level residual features comprising both the temporal and static features.

Group Activities

In an effort to improve performance, Ebrahimpour *et al.* [76] used both sensor and camera data sources to conduct crowd analysis based on three approaches: crowd video analysis, crowd spatio-temporal analysis, and social media analysis. [77] applied a deep convolutional relational machine (CRM) to recognizes group activities with an aggregation component that generated activity maps of individual and group activities with spatial information in the video. The map was generated through a multi-stage refinement reduces errors.

Q. Wang *et al.* [78] proposed the Multiview-based Parameter Free Framework (MPF) which comprised L1 and L2 norms. MPF helped in the characterization of structural properties of individuals in the crowd. The group overcame the limitation of crowd behavior analysis through their framework that automatically detected group numbers in

the crowd without using any parameters. H. Ullah *et al.* [79] method employed a two-stream CNN architecture that combined spatial and temporal networks to solve the problem of capturing information from still frames.

[80] implemented a Coherence Constrained Graph LSTM (CCGLSTM) based on spatio-temporal context coherence (STCC) and Global context coherence (GCC) constraint with a temporal and spatial gate to control the memory. The CCGLSTM component is used to recognize group activity at each time stamp by ignoring irrelevant features. Crowd density variation complicates crowd behavior analysis. In a bid to address this challenge, [81] implemented a two-stream convolutional network, incorporating LSTM, with density heat-maps and optical flow information to classify abnormal crowd behavior. T. Wang *et al.* [82] utilized hidden Markov model (HMM) to detect abnormal events in surveillance videos by relying on image descriptors derived from HOF orientations as feature extractor and a classification method.

The study "Violence in Crowds" [83] proposed a descriptor called Simplified Histogram of Oriented Tracklets (sHOT) based on spatio-temporal level and frame level orientation as well as magnitude extracted in spatio-temporal 3D patches at different levels. Behavior anomalies were localized in video sequences. To detect anomalies, Amraee *et al.* [84] created an approach that partitioned large non-overlapping cells, followed by the elimination of redundant information, and eventually the detection of abnormal events in crowded scenes using two distinct one-class SVM models. Liu *et al.* [85] used a predictive neural network to detect abnormal crowd behavior in public places by determining the difference between real frames and predictive frames in the moving object. Experiments on the UMN dataset scored highly at 97.7% accuracy. Khan [86] approach to congestion

detection used motion features extracted from optical flow and particle advection to show a pattern of increasing trajectory oscillation. The proposed method's evaluation showed that it could be used in real time.

[87] performed experiments using UMN dataset with the goal of determining crowd panic states based on entropy and enthalpy; enthalpy refers to the system's state and entropy is a measure of the degree of disorder. Crowd motion data can be obtained using optical flow technique. Gupta *et al.* [88] proposed a framework CrowdVAS-Net based on deep CNN for extracting features such as acceleration, velocity, etc., then trained with a RF classifier to differentiate between normal and abnormal behavior for effective crowd security and management in public places using videos.

1.9 Data Sources

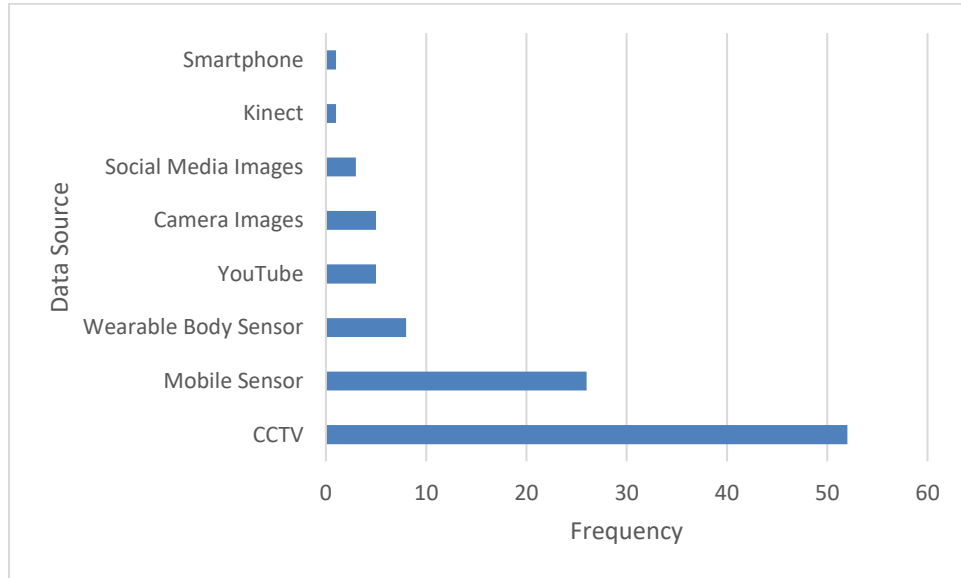


Figure 1.5: Frequency of data sources by existing literature on HAR [7]

1.10 Techniques

Machine Learning (supervised learning, unsupervised learning, and semi-supervised learning) form an integral part of HAR. The Figure 1.6 [7] provides a summary of frequency of techniques/algorithms used in the existing literature on HAR.

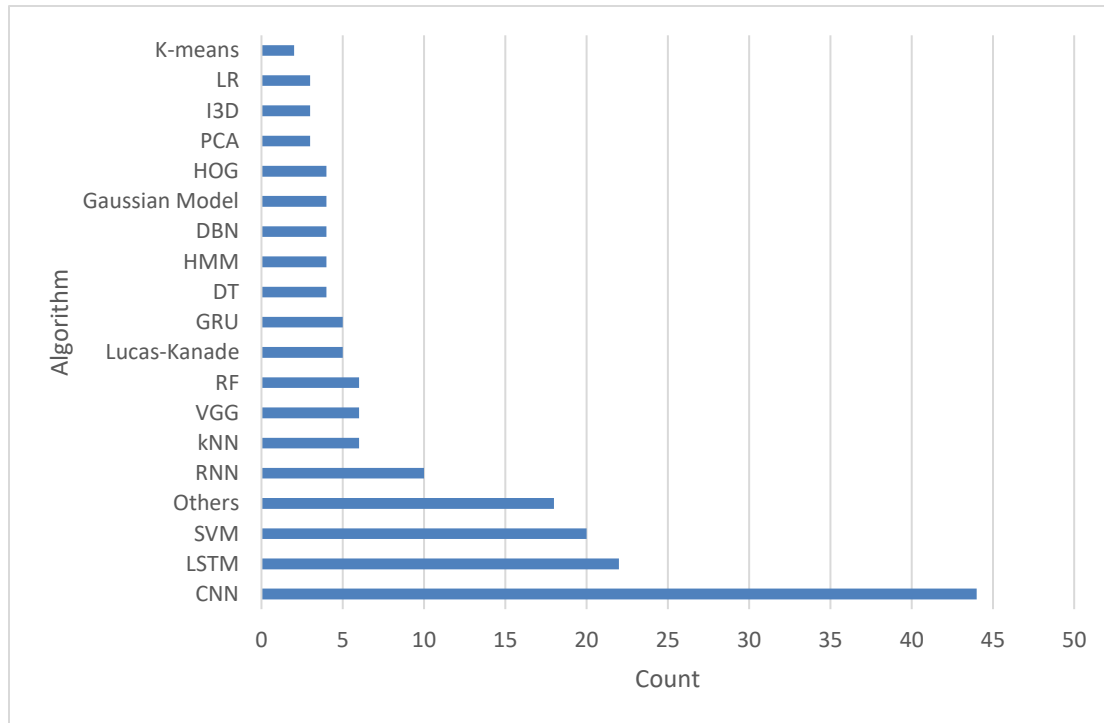


Figure 1.6: Frequency of algorithms used in the existing literature on HAR [7]

1.11 Challenges and Limitations

Despite the tremendous benefits of HAR in several domains, the existing solutions equally exhibit several limitations and challenges. These can be broadly categorized under data pre-processing, complex activities, activity misalignment, and hardware limitations.

1.11.1 Data Collection

The data collection process is pivotal in the research exercise. The quality of research output can be directly connected to the quality of data and the data process that supports the research. However, the process is not without challenges. Various studies

reveal data issues including unlabeled datasets, incorrectly labeled datasets, lack of temporal information, unknown class recognition, missing data, among others. These data limitations that must be addressed for good performance in activity recognition and prediction. Riaz *et al.* [52] investigated the impact of data labeling elimination on accuracy. On the contrary, Alafif *et al.* [89] studied the improvement of accuracy of the classifier by gathering more labeled data. Du *et al.* [21] study by spatial knowledge-based method established that managing temporal data activities is a challenge. Doshi and Yilmaz [90] demonstrated that algorithms may struggle to recognize activity when presented with unknown classes of data because they require a better understanding of objects and their context. Machine learning algorithms perform highest when trained on known data.

In their work, Zhu *et al.* [20] established that current designs are limited to only visible classes; that they cannot recognize unseen classes. The work of Hsu *et al.* [65] delved on intra-class variation challenges. They established that predicting an individual's actions is complicated and will vary depending on presence and presentation. [91] averred the need for a future project in artificial data generation to address the scarcity of large-scale image and video datasets. The works by Nadeem *et al.* [38], Köping *et al.* [14], Zhang *et al.* [92] and Lazaridis *et al.* [81] established that models trained on large datasets outperform counterparts trained on small datasets [49].

1.11.2 Data Preprocessing

After the collection of data, data preprocessing exercise helps to find the valuable input data for HAR Machine Learning training. Different studies identify “appearance and feature extraction” and “background reduction” as some of the key issues in data preprocessing. Ullah *et al.* [79] aimed to build hand-crafted characteristics and merge them

with deep learning architecture to increase implementation. Wang *et al.* [78] studied the need for more cooperative attributes in order to determine crowd actions. Amraee *et al.* [84] identified the significance of background images to be computed after a specific frequency and subsequently ingested into the proposed method. Jiang *et al.* [42] focused on how address cases of abrupt fluctuations of video backgrounds.

1.11.3 Hardware and Techniques

Hardware is the backbone and the ecosystem of all data and computations in the HAR process. Various disparate hardware may be involved concurrently, asynchronously, or synchronously. Where large size of data is involved, the hardware capacity becomes an important factor. The most common hardware are cameras for vision-based data, smartphones, smartwatches, and other different types of sensor implementations. This hardware may exhibit some limitations such as computational cost and hardware problems. Besides, some hardware is costly to acquire such as the case with good smartwatches. Xia and Li [72] investigated the intense computation power requirement for feature extraction by deep CNN. They attempted to come up with a simpler algorithm to extract appearance and motion features [93].

To shift away the burden from hardware as far as existing large-scale dataset is concerned, [74] investigated other possible representations of behavioral data to improve behavioral analysis time significantly. Yoon *et al.* [93] considered computational complexity as a constraint in their design of activity recognition as a multitask learning problem. Bhargava *et al.* [94] reported that DBN has a longer computation time when detecting anomalies. Ma *et al.* [95] established that the cost of computation load to determine the start and end of activity directly affects recognition performance, and this

should be optimized. Camera placement, both in location and orientation, determines the accuracy in HAR. The camera should be placed in such a way that it captures the required features. Where multiple views are necessary, multicamera are required such as analyzed in the study of crowd disasters control [88]. [96] looked at window size, network intensity, and breadth as factors that can be optimized for best recognition performance. J. Zhang *et al.* [97] investigated the perspective distortion compensation algorithm to obtain accurate weight calculation in a particular situation. Muhammad Shoaib *et al.* [98] investigated resource consumption of online activity recognition on smartphones and smartwatches on six different classifiers. They considered three different sensors, at different sampling rates and window sizes and analyzed the utilization of CPU, memory, and battery.

1.11.4 Complex Activities Detection

Most human activities are complex in the sense that they typically comprise multiple sub-activities, which can be decomposed further into more atomic or most basic activities. Cooking, eating, reading, medication, smoking, fighting, etc. are examples of complex activities. Complex Activities occur when multiple atomic activities occur sequentially or interleaved in time, whereas Atomic Activities are the simplified unit level activities that cannot be decomposed further, given the application semantics [99]. The complex activities comprise both the activities and real-time detection. Various studies have looked at the problem of concurrent activity datasets algorithms to improve performance of complex activity recognition [100]. Alafif *et al.* [89] work focused on complex attributes and application of deeper algorithms to process large data. With the proliferation of smartphones and smartwatches, there is great potential in the use of real-

time biometric data to recognize individual activities. Such an approach could generate real-time alerts from a live streaming camera [53].

1.11.5 Misalignment of Activities

Data annotation is one of the most tedious exercises of the data process in HAR, yet it remains fundamental to the process. Further, if not done well, annotation can result in inaccurate labeling and ambiguity in activity timing. Incorrectly labeled dataset translates to compromised performance. Furthermore, if the frame length of action is short, repeated instability in prediction can be observed due to frame limit uncertainty [101]. Misalignment is the outcome of an action's frame getting split into multiple frames, leading to the loss of some useful information during frame segmentation. Misalignment consequently leads to incorrect action detection, hence less reliability of the affected HAR solutions [30]. If synchronization of a system is based only on activity end labels, then there is a possibility that an activity may fall out of synchronization with the next activity in the frame.

CHAPTER 2: STATE TRANSITION MODELING OF THE SMOKING BEHAVIOR USING LSTM RECURRENT NEURAL NETWORKS¹

Abstract – The use of sensors has pervaded everyday life in several applications including human activity monitoring, healthcare, and social networks. In this study, we focus on the use of smartwatch sensors to recognize smoking activity. More specifically, we have reformulated the previous work in detection of smoking to include in-context recognition of smoking. Our presented reformulation of the smoking gesture as a state-transition model that consists of the mini-gestures hand-to-lip, hand-on-lip, and hand-off-lip, has demonstrated improvement in detection rates nearing 100% using conventional neural networks. In addition, we have begun the utilization of Long-Short-Term Memory (LSTM) neural networks to allow for in-context detection of gestures with accuracy nearing 97%.

Keywords: Smartwatch, IoT, Artificial Intelligence, Smoking detection, Mini-gesture, health, LSTM, ANN

¹ Related Publications and Authors:

Chrisogonas O. Odhiambo, Casey A. Cole, Alaleh Torkjazi, Homayoun Valafar. Proceedings of the 2019 International Conference on Computational Science and Computational Intelligence CSCI, "*State Transition Modeling of the Smoking Behavior Using LSTM Recurrent Neural Networks*", Reprinted here with permission from the publisher

2.1 Introduction

Cigarette smoking has remained the leading preventable cause of death in the world for the past several decades. In the US alone, 20% of the population report that they engage in smoking and diseases caused by smoking cost the population over \$170 billion in healthcare each year (www.cdc.gov, www.who.int)[102]. In addition, the majority of smokers report that they want to quit yet among those that make a quit attempt, the majority relapse at least once. Reducing the number of relapses is of great interest to the community of tobacco-related researchers. Many studies[103]–[106] have been conducted in an attempt to both properly describe smoking behavior as well as pinpointing the best-times to intervene such that a relapse does not occur. However, these studies are inherently limited due to the current methods of studying smoking behavior. Most studies conducted rely on participants to self-report their smoking behavior. In various studies[103], [107], [108], the accuracy of self-reporting has been shown to be no more than ~76%. To bypass the reliance on self-reporting, some studies have been conducted in laboratory-based settings. In these studies, participants are required to smoke while being recorded. In addition, some studies[103] also enforce that the participants insert their cigarette into a device that measures attributes, such as puff duration and the interval between puffs, as they smoke. These measures are extremely useful to researchers who study topics like craving and the effects of nicotine withdrawal. Whereas these studies bypass the limitation of self-reporting, participants often report that they felt uncomfortable in the lab environment or dissatisfaction of the smoking experience due to the incorporation of the measurement device.

A potential solution to these limitations is the use of smartwatch devices. The use of smartwatch technology allows the study to be conducted in a smoker's natural

environment, therefore eliminating the biases introduced in laboratory settings. In addition, commercial smartwatch devices come with a rich array of sensors that can be utilized, in conjunction with ML techniques, in order to detect a variety of human activities[109]–[111]. Automatic detection of behaviors allows for an unobtrusive and passive collection and characterization of human activities that does not rely on self-reporting. It also allows for “in-time” intervention techniques to be developed. Our previous work[107] in this field has indicated that smoking can be detected using only accelerometer data and single-layer artificial neural networks (ANNs) with an accuracy of ~95% in laboratory settings and ~90% in real-world settings[107], [112]. Independent reports[110], [113]–[116] also confirmed the usability of smartwatches in the study of human behavior.

Interpretation of human activity can substantially benefit from in-context analysis since there exist temporal relationship between activities. For instance, one smoking gesture clearly consists of a sequence of three consecutive mini-gestures initiated by hand-to-lip, followed by a duration of hand-on-lip, and concluded by a mini-gesture of hand-off-lip. To characterize smoking at this more fine-grain level of mini-gestures will require the reformulation of the classification problem while providing the benefit of improved gesture detection. The primary focus of the presented research is to investigate the impact of mini-gesture representation of smoking. To that end we will explore the performance of conventional and LSTM Neural Networks[117] and compare the results to the previously published work.

2.2 Background and Method

2.2.1 Previous and Related Work

Considering their rich array of sensors, the cost, accessibility, and ease of use, smartwatches have emerged as a compelling platform to study human activities unobtrusively. Smartwatches have been used as step-counters[116], sleep monitoring[118], diet monitoring[113] as well as general fitness tracking[119]. In the context of smoking, smartwatches have demonstrated to be usable for in-situ study of smoking[120], [121] with high accuracy[107], [120]–[122]. Smartwatches have been used to detect smoking gesture with 95% accuracy in laboratory environment[123] and 90% in-situ detection of smoking[107]. Study of smoking has also been demonstrated to be more accurate when compared to self-reporting (90% versus 78%)[107], [124].

In this experiment we have utilized the smoking data recorded from five smokers and compare our results to the previously published work. The previously reported detection of smoking was implemented using a hybrid approach that consisted of an ANN for low level detection of smoking puffs alongside a rule-based AI for overall classification of detected puffs into smoking sessions. Using this model, a session level detection accuracy of 81-90% was achieved. Whereas this demonstrated a high degree of success, improvements to the detection of individual puffs have the potential to increase this accuracy even further.

2.2.2 Data Annotation

In this experiment we used data from a previously published work that is available for download from <https://ifestos.cse.sc.edu>. This data consisted of 10 smoking events along with 15 non-smoking events collected across six individuals. Five of the individuals collected their data in a laboratory setting and one in real world settings. The non-smoking

activities collected ranging from eating to typing on the computer. In total, 172 individual smoking puffs were collected. An example of a smoking puff is shown in Figure 2.1.

In the previous study, each smoking puff was annotated by indicating the start and end of each puff by an expert. Using MATLAB, each puff gesture was further parsed into three sub-gestures: hand-to-lip, hand-on-lip, and hand-off-lip. The hand-to-lip gesture was defined as the motion of the cigarette from the resting position (hip, thigh, etc.) to the mouth. This region is shown in the left most box in Figure 2.2 and encompasses about 20 data points (or 0.8 seconds). The hand-off-lip gesture is then defined as the return of the participant's hand to a resting place. This region is shown in as the far-right box in Figure 2.2 and is also about 0.8 seconds in length. The hand-on-lip gesture was defined to be the region in which the person has the cigarette in or near their mouth (inhalation time) and is shown in between the two boxes in Figure 2.2. This region is typically longer than both the hand-to-lip and hand-off-lip regions and greatly varied across participants as well as within each participant. Each hand-on-lip gesture was further broken apart using a rolling window of size 20 to make them compatible with the other mini gestures. Non-smoking gestures were also extracted in this way.

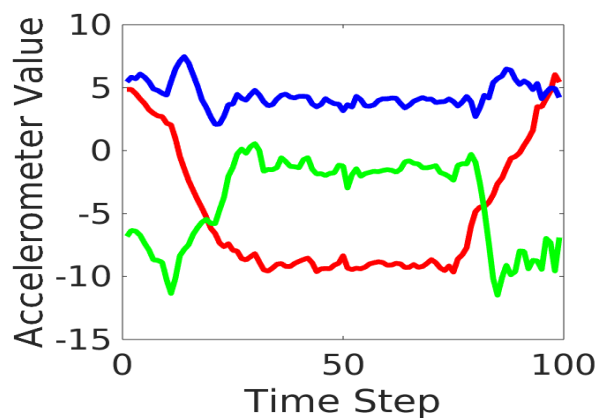


Figure 2.1: An example of a smoking puff is shown where red indicates the X dimension of the accelerometer data, green the Y and blue the Z.

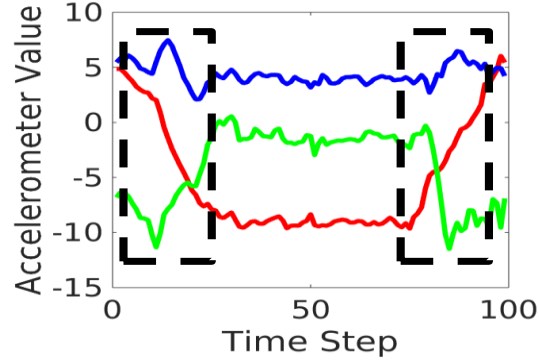


Figure 2.2: An example of sub-gesture annotation of a puff where the first box denotes the hand-to-lip gesture and the second box the hand-off-lip gesture.

2.2.3 Overview of Training/Validation/Testing Sets

The total number of gestures extracted was 172 hand-to-lip, 5054 hand-on-lip, 172 hand-off-lip and 5854 non-smoking. Due to the imbalance of data in each class, the hand-to-lip and hand-off-lip sub-gestures were duplicated 30 times to ensure nearly uniform number of observations in each training category. The final total number of gestures per sub-gesture were 5,160 hand-to-lip, 5,054 hand-on-lip, 5,160 hand-off-lip and 5,854 non-smoking. The target classes were coded as shown in Table 2.1 with non-smoking labeled as 1, hand-off-lip as 4, hand-on-lip as 3 and hand-to-lip as 2. One-hot encoding was used to generate a target matrix (one-hot encoding for each class shown in Table 2.1).

Table 2.1: Class assignment and one-hot encoding for each sub-gesture.

Sub-gesture	Class	One-hot Encoding
Non-Smoking	1	1000
Hand-to-lip	2	0100
Hand-on-lip	3	0010
Hand-off-lip	4	0001

The resulting dataset was split into the three traditional datasets of training, cross-validation, and testing sets in the ratios of 70%, 15%, 15%, respectively.

2.3 Neural Network Platform and Architecture

Using Keras and TensorFlow as the simulation platform of our ANN, we investigated the performance of the conventional feedforward perceptrons, and LSTM neural networks. Both networks were trained using the same dataset published by a previous study[123]. The following sections provide the details for each individual study. We trained both the conventional Artificial Neural Network and Recurrent Neural Network Long-Short-Term Memory (LSTM) types of networks. While ANN generally registered high accuracies, LSTM comes with unique advantages such as constant error backpropagation within memory cells which makes it possible to bridge very long-time lags. It also works well over a broad range of parameters. Importantly, the network's previous knowledge/output forms the input of the next unit. This means that LSTM's learning becomes better with every subsequent unit.

Conventional ANN – As the first step in our investigation, we explored the performance of a conventional (shallow) neural network in order to establish the impact of mini-gesture detection instead of detection of an entire smoking puff. Here we implemented a comparable architecture to the one from our previous study[123] and observed the implications of the reformulated study. In our previous study, we utilized a conventional ANN consisting of 300 input neurons, 10 hidden neurons and a single output neuron. The reformulation of the mini-gesture detection requires 60 input neurons and 4 output neurons representing 4 classes of mini-gestures. In this study we investigated the number of hidden layers and hidden neurons that will provide the optimal detection performance. During our

studies, we investigated activation function, number of batches, and the number of layers while using Adam optimization[125] method and Binary-cross entropy loss function for training the network. A summary of the Keras python code is shown in Figure 2.3. This code segment was modified to incorporate 2-layer, 3-layer, and 4 layers of hidden neurons.

```
model = Sequential()
model.add(Dense(12, input_dim = 60,
activation='relu'))
model.add(Dense(8, activation
= 'relu')) #... (ann-i)
model.add(Dense(4, activation='sigmoid'))
model.compile(optimizer='adam',
loss='binary_crossentropy',
metrics=['accuracy'])
```

Figure 2.3: A snippet of Python code used in Keras to define the used ANN.

LSTM-NN – While the conventional multi-layered feedforward ANNs remain excellent tools to be used in the prediction and classification tasks, they poorly incorporate temporal information. LSTM recurrent models[117], [126] have demonstrated success in incorporating temporal and historical information to their classification protocol and address a very critical aspect of the reformulated data i.e. the temporal aspect. For instance, a typical and permissible smoking puff should consist of the specific sequence of Hand-To-Lip, followed by Hand-On-Lip, and terminated by a Hand-Off-Lip mini-gestures. Furthermore, a typical puff should consist of approximately a reasonable duration of puff (identified by the Hand-On-Lip mini-gesture) that is no shorter than 0.5 second and no longer than 3 seconds. Any departure from this allowed range should disqualify the identification of a proper puff. All these relationships can define a smoking grammar based on the vocabulary of mini-gestures.

Our preliminary investigation of the LSTM-NN consisted of an exploration over the most optimal architecture (number of units) where the input and output of the network

consisted of 60 and 4 neurons respectively. Figure 2.4 illustrates our model of LSTM in the Keras environment. Using this model, we have investigated the performance of varying 2-unit, 3-unit, and 4-unit LSTM architectures.

```
model = Sequential()
model.add(LSTM(output_leng,
batch_input_shape=(None,1,input_leng),
return_sequences=True,
activation='sigmoid')) #... (lstm-i)
model.compile(loss='mse',
optimizer='adam',metrics=['accuracy'])
```

Figure 2.4: A snippet of Python code used in Keras to define the used LSTM.

2.3.1 Training and Testing Procedure

In our investigations, we used the hold-out strategy to set aside a section of the training dataset as a validation set that constitutes a fully independent data set. This strategy has a lower computational cost compared to k-fold strategy because it is only executed once. However, performance evaluation is subject to higher variance, given the smaller size of the data. The entire data set was partitioned using the ratios of 70:15:15 for training set, validation set, and test set respectively.

We evaluated our conventional and LSTM neural networks in terms of loss and accuracy, as a function of architecture. Accuracy measures the performance of the network, while the loss function helps in optimizing the parameters of the neural networks. The objective of the training is to minimize the loss by optimizing parameters i.e., weights. We calculate loss by matching the annotated target value and the predicted values by the network. We used Mean Squared Error (MSE) loss function to quantify the success of our ANN in predicting the desired outputs. We also relied on accuracy measure as an overall metric of classification success.

The primary objective of our exploration was to discover models of ANN that will outperform the previously reported performance of 95%, using the same data set. It is noteworthy, that although in this exercise we used the same data as before, the problem was reformulated such that the input size was reduced from 300 input neurons to 60, and the output neurons was increased from 1 to 4.

2.4 Results and Discussion

In this section we provide the results of our investigations for the optimal performing architectures for the conventional and LSTM neural networks.

2.4.1 Conventional Neural Network

In total, we examined the performance of more than 20 architectures of ANN in order to select the relative optimal architecture. Table 2.2 shows the performance outcomes at training and testing phases, under some representative architectural and training parameters (epochs, batches and units). Highlights show best performance configurations. The first entry in this table (with the yellow highlight) indicates the architecture with the most optimal performance based on the training set. The loss and accuracy functions for this network as a function of epochs are shown in Figure 2.5 and Figure 2.6 respectively. Careful examination of these figures indicates an overtraining of the network based on the increasing pattern of the loss of the validation set. Based on this observation, we imposed two additional constraints that relates to our application. The first criterion was to select the most optimal network based on the minimum combined loss of the training and validation set. The green entries in Table 2.2 denote all configurations that resulted in a combined loss value of 0.06. The second criterion selects against the larger ANN due to

the nature of our application that is cognizant of power consumption. Our final selected architecture is shown in blue that balances detection performance and power consumption.

Table 2.2: Mini-gesture detection using different ANN architectures

Epoch	Batch	Layers	Loss	Val. loss	Accuracy (%)	% Val. Accuracy	Test (%)
5000	100	4	0.01	0.07	99.54	98.84	98.79
5000	50	4	0.01	0.09	99.47	98.70	99.65
5000	50	3	0.02	0.06	99.34	98.72	99.28
3000	50	2	0.02	0.06	99.31	98.82	98.40
3000	50	3	0.03	0.07	99.07	98.69	99.17
2000	50	4	0.02	0.04	99.41	98.91	99.57
2000	100	4	0.03	0.06	99.13	98.65	99.38
1500	50	3	0.02	0.05	99.37	98.81	99.13
1500	50	4	0.02	0.04	99.35	98.85	99.42
1500	100	4	0.02	0.04	99.18	98.70	99.26
1000	50	4	0.02	0.05	99.09	98.55	98.69
1000	100	3	0.03	0.06	98.92	98.54	98.68

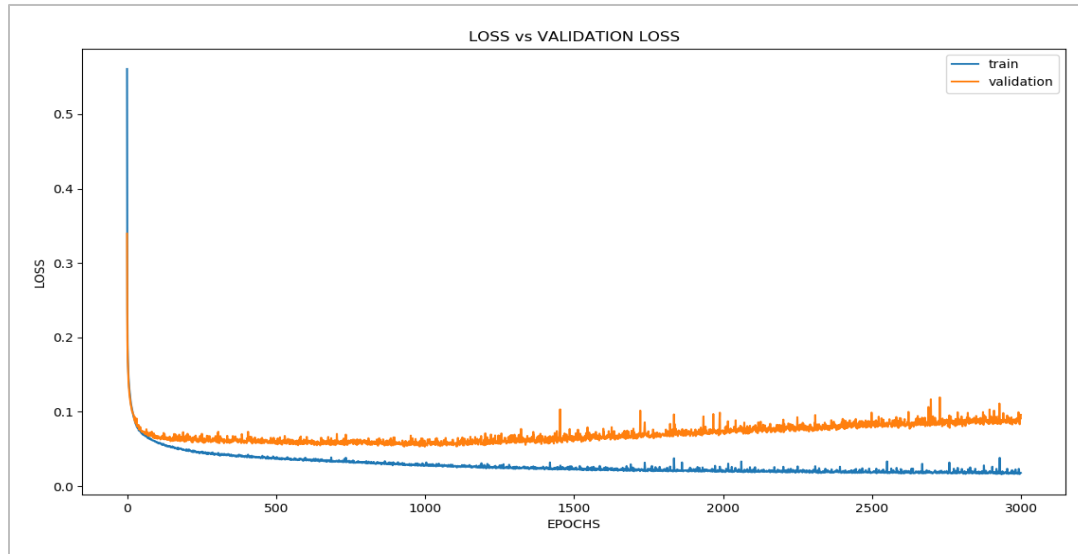


Figure 2.5: Most optimal ANN architecture loss function as a function of epochs

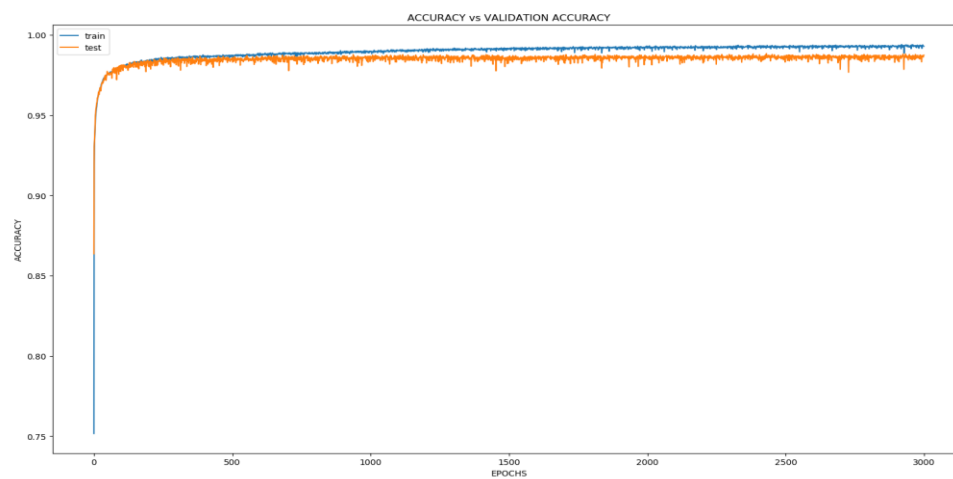


Figure 2.6: The measure of accuracy as a function of epochs for the training set (blue) and validation set (yellow).

Table 2.3: The confusion matrix of the most optimal ANN architecture.

		Prediction			
Actual		Rest	H-to-L	H-on-L	H-off-L
	Rest	1125	5	9	3
	H-to-L	14	1021	11	0
	H-on-L	59	7	930	9
	H-off-L	25	10	0	1018

Table 2.4: Confusion matrix report for the most optimal ANN architecture.

	Precision	Recall	f1-score	Support
0	0.92	0.99	0.95	1142
1	0.98	0.98	0.98	1046
2	0.98	0.93	0.95	1005
3	0.99	0.97	0.98	1053
micro avg	0.96	0.96	0.96	4246
macro avg	0.97	0.96	0.96	4246
weighted avg	0.97	0.96	0.96	4246

2.4.2 LSTM Neural Network

Table 2.5 provides an overview summary of our investigation of varying LSTM architectures in detection of smoking mini-gestures. Similar to the case of the conventional ANN, our first selections of the optimal architectures are highlighted in yellow. We imposed the minimalism of architecture without any compromise of the performance. As an indirect consequence of this selection, the performance of the network on the test set increased from approximately 93% to 95% indicating slight memorization by the network that can be remedied by smaller LSTM networks. The memorization phenomenon can also be confirmed after careful examination of the loss and accuracy functions of the training and validation sets shown in Figure 2.7 and Figure 2.8 respectively. Table 2.3 illustrates the confusion matrix of the most optimal ANN, while Table 2.4 provides a summary of the precision and recall of the confusion table.

Table 2.5: A summary of LSTM's performance in detection smoking mini-gesture as a function of different architectural parameters.

epoch	Batch	Units	Loss	Val. Loss	Accuracy (%)	Accuracy (%) Val.	Test Acc. (%)
5000	100	3	0.03	0.03	96.90	95.74	93.57
5000	50	3	0.03	0.03	96.00	95.29	94.85
3000	50	3	0.03	0.03	96.67	95.48	94.78
3000	50	2	0.03	0.03	96.21	95.20	95.02
1500	100	3	0.03	0.04	95.95	94.94	95.13
1500	50	3	0.04	0.04	94.67	94.23	94.92
1000	50	3	0.03	0.03	95.51	94.98	94.44
500	50	2	0.04	0.04	94.92	94.80	91.90

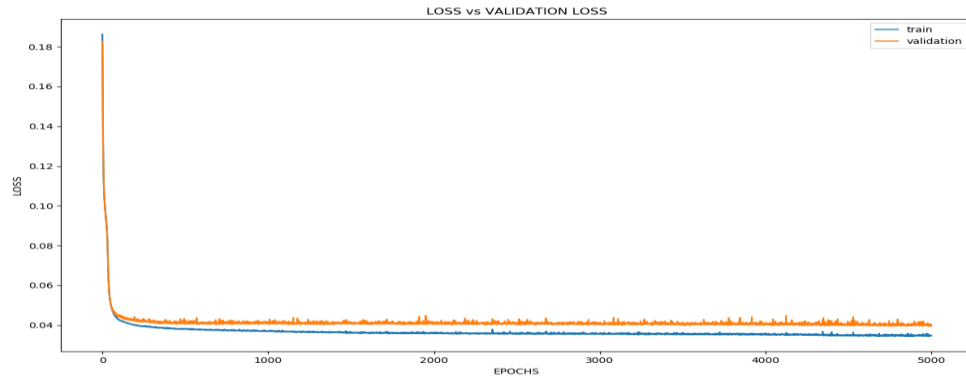


Figure 2.7: Loss function of the most optimal LSTM architecture as a function of epochs, illustrated by blue and orange for training-loss and validation-loss, respectively.

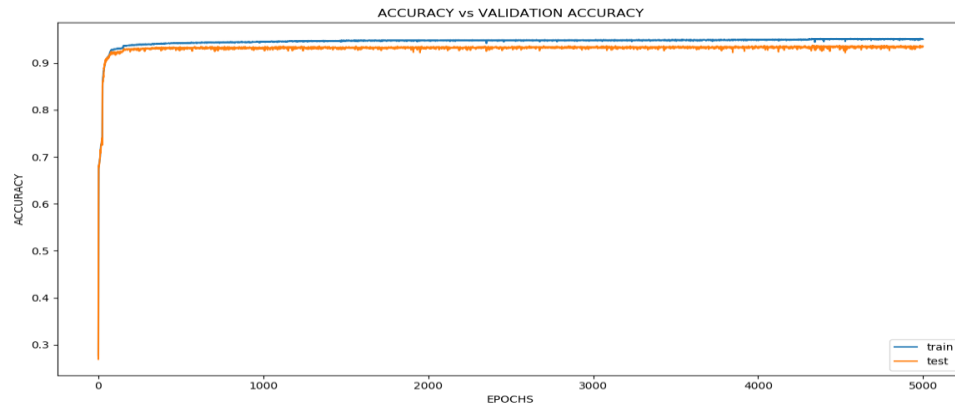


Figure 2.8: The measure of accuracy of the most optimal LSTM architecture

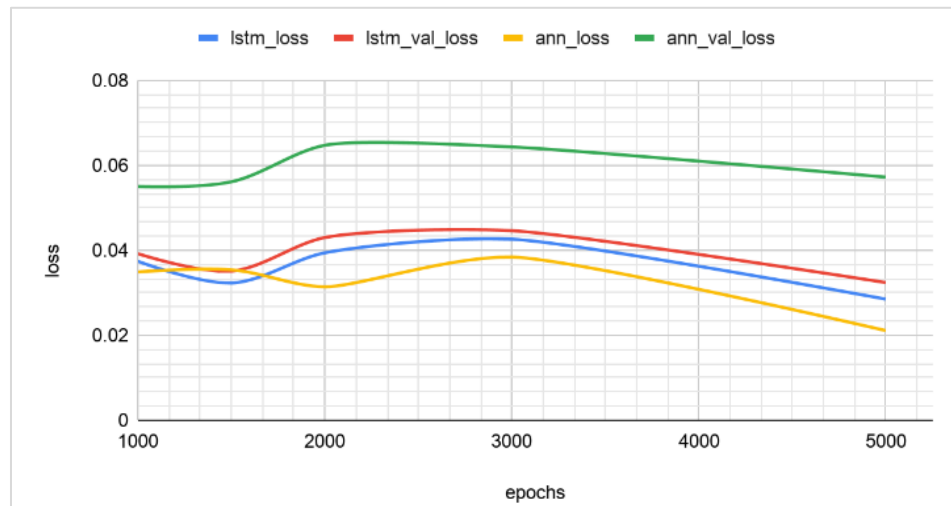


Figure 2.9: Comparing LSTM/ANN loss.

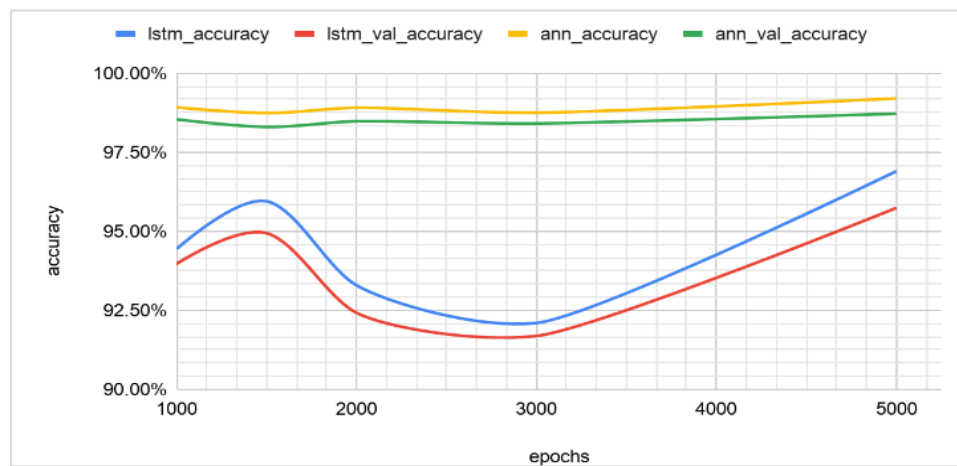


Figure 2.10: Comparing LSTM/ANN Accuracy.

Figure 2.9 shows ANN validation loss is generally higher than in LSTM for all trained models. Figure 2.10 shows both ANN validation accuracy and accuracy consistently higher than in LSTM for all trained models.

Table 2.6: The confusion matrix of the most optimal LSTM architecture where Batch=100, Units=3, Activation=sigmoid, epochs=5000

		Prediction			
Actual		Rest	H-to-L	H-on-L	H-off-L
	Rest	1090	7	33	12
	H-to-L	17	1017	12	0
	H-on-L	43	23	929	10
	H-off-L	10	25	24	994

Table 2.7: Confusion matrix report for the most optimal architecture of LSTM.

	Precision	Recall	f1-score	Support
0	0.94	0.95	0.95	1142
1	0.95	0.97	0.96	1046
2	0.93	0.92	0.93	1005
3	0.98	0.94	0.96	1053
micro avg	0.95	0.95	0.95	4246
macro avg	0.95	0.95	0.95	4246
weighted avg	0.95	0.95	0.95	4246

2.5 Conclusion

In this report, we have presented the reformulation of an entire smoking gesture (puff) as a combination of three time-dependent mini-gestures (hand-to-lip, hand-on-lip, and hand-off-lip). Using this reformulation, we demonstrated the success of conventional ANN

(99%) in improving upon the previously reported detection of smoking (95%) using the same set of data. Based on the results shown in Table 2.2, the reformulation of the smoking gesture as mini-gestures clearly reduces the complexity of detection as evidenced by the improved detection. Although we have achieved a near perfect detection of the smoking gesture, we anticipate unforeseen challenges during the live deployment of this technology for in-situ study of human smoking behavior. Furthermore, we remain cognizant of the battery requirement during the live deployment of this technology.

In order to incorporate the temporal dependency of human activities, including the mini-gestures, we have hypothesized that LSTM recurrent neural networks would exhibit a better performance. While our initial and in laboratory investigations have not supported this hypothesis, we anticipate that the true value of RNN will be exposed in live deployment of the system.

In summary, our state-transition approach to detection of smoking mini-gestures had demonstrated improvements over the previously reported approaches. We expect that the declaration of mini-gestures as the “vocabulary” of smoking is instrumental in the development of the smoking “grammar” that can be exploited by the incorporation of RNNs.

CHAPTER 3: MEDSENSOR: MEDICATION ADHERENCE MONITORING USING NEURAL NETWORKS ON SMARTWATCH ACCELEROMETER SENSOR DATA²

Abstract – Poor medication adherence presents serious economic and health problems including compromised treatment effectiveness, medical complications, and loss of billions of dollars in wasted medicine or procedures. Though various interventions have been proposed to address this problem, there is an urgent need to leverage light, smart, and minimally obtrusive technology such as smartwatches to develop user tools to improve medication use and adherence. In this study, we conducted several experiments on medication-taking activities, developed a smartwatch android application to collect the accelerometer hand gesture data from the smartwatch, and conveyed the data collected to a central cloud database. We developed neural networks, then trained the networks on the sensor data to recognize medication and non-medication gestures. With the proposed machine learning algorithm approach, this study was able to achieve average accuracy scores of 97% on the protocol-guided gesture data, and 95% on natural gesture data.

Keywords: Smartwatch, Wearable Sensors, Wearable Computing, Medication Protocol, Medication adherence, Neural Networks, Machine Learning

² Related Publications and Authors:

Chrisogonas O. Odhiambo, Pamela Wright, Cindy Corbett, Homayoun Valafar. Proceedings of the 2021 International Conference on Computational Science and Computational Intelligence CSCI, "*MedSensor: Medication Adherence Monitoring Using Neural Networks on Smartwatch Accelerometer Sensor Data*", Reprinted here with permission from the publisher

3.1 Introduction

Poor adherence to prescription medication is a major problem with a myriad health and economic implications. It can lead to compromised treatment effectiveness, medical complications, and even death especially when strict adherence to medication dosage and frequency is required. It can also lead to loss of billions of dollars in unnecessary health care expenses due to wasted medicine or further health complications arising from poor medication adherence [127]–[129]. Studies show that 33-69% of all medication-related hospital admissions in the United States (US) are caused by poor medication adherence, which translates to an annual cost of approximately \$100 billion [130], [131].

Annually in the US, non-adherence can account for up to 50% of treatment failures, approximately 125,000 deaths, and up to 25% of hospitalizations [132]. Typically, adherence rates of 80% or more are needed for optimal therapeutic efficacy. However, it is estimated that adherence to chronic medications is around 50% [132].

The two main causes of poor medication adherence are stress and the complexity of medication procedure or steps [133]. Both physical and emotional stress on a patient may result in depression, anger, denial of illness, or fear of medication. The complexity includes factors such as the dosage, frequency, duration, cost, and refill policy, which can demotivate the patient. While stress may be difficult to control by external factors or tools, the complexity burden may be reduced by technology.

With these economic and health implications, it is imperative to provide tools and means to enable medication adherence. The question then becomes: What are some of the readily available and affordable tools that can leverage modern technology to support adherence to medication? The purpose of this study was to explore the use of smartwatch

sensors in monitoring human hand motions to detect medication-taking, with the aim to help people adhere to their prescriptions, hence minimizing the negative effects of poor medication. This, in conjunction, with other messaging technologies such as Amazon Alexa, or simple SMS notifications, can provide useful medication reminders.

3.2 Background

3.2.1 Wearables in Human Activity Detection

Various studies demonstrate how smartwatches have been used to monitor and detect human motions, such as the case of smoking detection [122], [134]–[136], or fall-detection [110], [137]–[139]. Independent reports [113], [115], [140]–[143] also confirm the usability of smartwatches and other smart wearables in the study of complex human motion behaviors such as eating habits, physical activities, and foot motion [113], [115], [116], [118]–[120], [140]–[143]. Considering the rich array of sensors, cost, accessibility, and ease of use, smartwatches have emerged as a compelling platform to unobtrusively study human activities. Uses of Smartwatches include step-counters [116], sleep monitoring [118], diet monitoring [113] as well as general fitness tracking [119]. Smartwatches have demonstrated [120], [121] high accuracy for detecting smoking gestures [120]–[122]. Smoking gestures were detected with 95% accuracy in the laboratory environment[123] and 90% accuracy in-situ [122]. Studies to identify smoking via gestures has also been demonstrated to be more accurate when compared to self-reporting (90% versus 78%)[122], [124].

3.2.2 Wearables in Detection of Medication Adherence

Monitoring medication-taking can be broadly categorized as direct or indirect. The former involves observation of a person taking medicines or drug-testing in a laboratory [133]. The latter involves self-reporting, pill counting, medication refill tracking, and electronic tracking using smart wearables, cameras or pill caps with medication event monitoring systems [127], [130], [144]. Direct methods are most accurate, but generally more obtrusive, time-consuming, and expensive. Indirect methods are relatively inexpensive, efficient, and less obtrusive tools for monitoring and reporting medication-taking. This study focuses on indirect approaches of medication monitoring.

Smart wearable devices have been utilized in indirect observation of medication adherence in numerous ways including: (1) self-reporting facilitated by mobile devices[145], (2) sensors worn around neck such as the SenseCam[146] was originally envisaged for use within the domain of Human Digital Memory to create a personal lifelog or visual recording of the wearer's life, which can be helpful as an aid to human memory, (3) multi-axis inertial sensors worn on wrists[147], [148], and (4) the use of commodity smartwatches [149]. In summary, all the approaches have collectively demonstrated the potential of smart sensors to promote medication adherence, but leave potential for improvement in performance, cost, convenience, and usability.

3.3 Data Collection and Method

The overall approach to our investigation included data collection from human subjects (n=31) followed by developing and testing the performance of Artificial Neural Networks to identify medication-taking events. The following sections provide the details for each step of our studies.

3.3.1 Data Collection Protocol

The data collection was performed using Wear OS compatible smartwatches worn on each participant's right wrist. A custom software package named MedSensor was installed on the smartwatches to facilitate data collection, annotation, storage, and transmission. The acquired data consisted of time stamp and x, y, and z components of the accelerometer data sampled at 25Hz intervals. In addition, information regarding the start and end of each medication-taking session was provided by the user and recorded to assist with data annotation. Participants marked the beginning and end of each medication-taking activity by pressing the corresponding button (shown in Figure 3.1) on the MedSensor app. A total of 31 participants were included in the data collection activities. Each participant was directed to record 10 protocol-guided medication-taking activities per day for 5 days, followed by 5 days of recording 10 medication-taking activities per day using their natural medication-taking gestures. In total 1300 protocol-guided and 1300 natural medication-taking gestures were collected.







Figure 3.1. MedSensor interface used for annotation at the edge.


3.3.2 Medication Taking Activity

Our data collection consisted of two broad categories of protocol-guided and natural medication-taking gestures. The first phase of our study (presented here) focused on the

recognition of the protocol-guided medication-taking activity as a proof of concept. The protocol-guided medication-taking activity is defined as the five explicit consecutive steps shown in Table 3.1. The natural gesture medication-taking activity is purely defined by the participant and likely consists of many permutations of the sub-activities shown in Table 3.1 Table 3.1 and performed by any combination of left or right hands. It is important to note that medication-taking gestures cannot be performed by a single hand, and it must involve the use of both hands.

Table 3.1. Protocol-guided medication-taking activity.

Step	Description	Activity
1	Unscrew the medicine bottle cap with your right hand while holding the bottle by your left hand.	
2	Tip the medicine bottle with left hand to dispense pill(s) onto your right hand.	
3	Place/toss pill to mouth using the right hand.	
4	Pick up beverage/drink with the right hand, bring to mouth and drink to swallow "pill".	

5	Set glass down, hold the medicine bottle in left hand, and put its cap back using right hand.	
---	---	---

3.3.3 Data Consolidation and Transfer

Upon completion of medication-taking recording sessions, each participant submitted the medication-taking gesture data collected by the watch to their paired smartphone via Bluetooth. Further, the participant submitted the data from the phone to the cloud storage via an internet connection and the MedSensor phone application. The phone provided the bridge between the watch and cloud because the watches did not provide for direct file upload to the cloud. Besides this role, the phones were not necessary. Other than data transfer to cloud through the phone, it was also possible to access watch data directly via data cables or Wi-Fi. However, such a method would not be practical for some participants. To establish a homogenous protocol, the participants were directed to use the MedSensor interface to collect and submit data to a centralized cloud storage. This was a better locationally transparent process that also preserved the integrity of data from capture to dispatch. The data from the watch is a zip of two csv files: actual sensor data and annotation points that identify the medication gestures and the non-medication gestures. It is important to note that the data collected includes non-medication gestures. These are equally important in the network training since they ultimately help the models discern what is a valid and what is not a valid medication gesture.

3.3.4 Data Annotation Process

Careful and proper data annotation is one of the most critical, time-consuming, and challenging aspects of utilizing supervised learning. In our study, we used the self-reported start and end of each medication-taking event to easily expose a small section of a person's recorded medication-taking gestures. Figure 3.2 shows an example of one activity of interest embedded within a larger recording session. The self-reported start and end annotations are rough approximations and require further scrutiny by a trained supervisor. Aside from human error in marking the beginning and end of an activity by a user, the recorded data can include unrelated activities such as the gesture that is required to mark the START/END on the smartwatch. Therefore, we further refined the start and end of the medication-taking activity while confirming the existence of one. This process produces a more reliable and accurate set of training and testing data. A typical medication gesture is shown in Figure 3.2.

Although in our current investigation we use the entire medication taking gesture as one complete signal, in principle, it is possible to subdivide and analyze the signal as a temporal sequence of sub-gestures as denoted by segments A-D in Figure 3.2. These segments correspond to open-bottle dispense-medicine (A), Hand-to-mouth pill-to-mouth hand-off-mouth (B), pick-up-water drink-water lower-cup-to-table close-bottle (C), and post-medication (D), respectively. Figure 3.3 is a superimposition of three separate medication-taking events collected from the same participant.

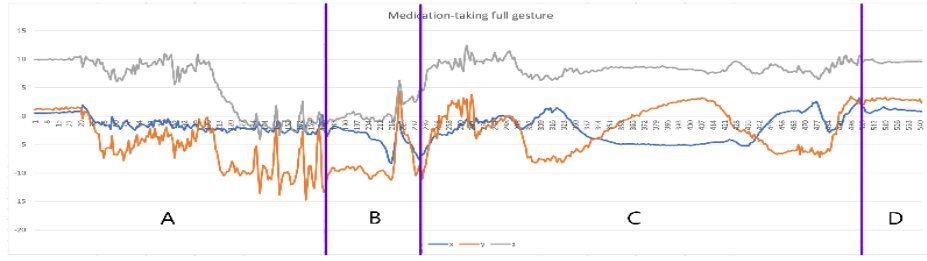


Figure 3.2: Visualization of a full medication gesture from one participant.

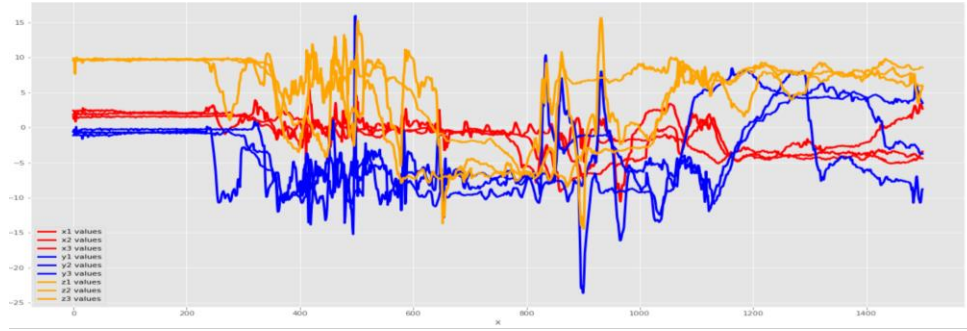


Figure 3.3: Visualization of the superimposition of three medication gestures from one participant.

3.3.5 ANN Training and Testing Methodology

We explored three related aims in our investigations. These aims are as follows and gradually span the gamut of the intended applications of our software:

Exp #1. Explore the capabilities of the network when trained and tested with protocol-guided data from all participants. The test data was a split in-sample dataset from the protocol-guided dataset.

Exp #2. Explore the capabilities of the network when trained with the protocol-guided data from $n-1$ participants to be tested with the n -th participant. Both the train and test datasets were protocol-guided gestures except the test dataset that was an out-of-sample.

Exp #3. A preliminary exploration of training a network using all gestures (both the protocol-guided and natural) from n-1 participants and tested on the n-th participant.

The training dataset also included the protocol guided dataset of the n-th participant.

In all these experiments, the explored ANNs were presented with an entire gesture. Therefore, the input size of ANNs consisted of the longest medication-taking gesture across all the training and testing sets. This consisted of 1500 consecutive accelerometer data points (approximately 20 seconds) that required an input size of 4500 neurons to accommodate the x, y, and z components of the accelerometer. The output layer consisted of a single neuron reporting the presence or absence of a medication-taking event. We employed a parsimonious strategy in determining the size and number of hidden layers. During each phase of our experimentation, an array of hidden neurons was explored to yield the optimal performance. In all the experiments, a single hidden layer sufficed (general architecture shown in Figure 3.4), and we therefore did not explore deep architectures. Generally, the number of hidden neurons started at 10 and was incremented in steps of 10 to as many as 100 hidden neurons. In each application, the optimal architecture was then selected to be carried for testing purposes.

In experiment #1 the training and testing data sets were randomly selected from all the data in the ratio of 80:20 respectively. Each bootstrapping exercise (Exp #2, and #3) was repeated n-times by excluding each participant in each round. We used Keras/TensorFlow platform for all our ANN simulations. A loss function of accuracy defined in Eq 1 was used to assess the performance of the trained ANN. In this equation the terms TP, TN, FP, and FN correspond to true positive, true negative, false positive, and false negative, respectively.

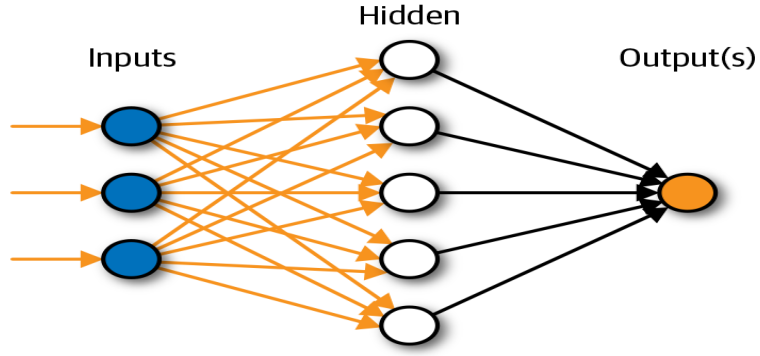


Figure 3.4: Artificial Neural Network high-level architecture

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad \text{Eq (1)}$$

3.4 Results and Discussion

3.4.1 Results of Experiment #1

The first experiment was the most fundamental test that could be conducted. The results of this experiment are shown in Table 3.2

Table 3.2. Each experiment was repeated three times to assess whether results varied due to randomization of the training/testing datasets. In general, the results did not vary noticeably and therefore, we report the results of one single instance of training/testing. Based on the results shown in this table, when developing a single ANN to detect a medication-taking event across the entire sample, a hidden layer size of 90 neurons is the optimal architecture. Although not a significant reduction, the performance of ANN slightly decreases with a larger or smaller hidden layer size.

Based on these results, it is possible that a single ANN (with optimal architecture of 4500, 90, 1) can detect a medication-taking event with as high as 97.8% accuracy if the 31 participants in our study provided a comprehensive representation of all medication-taking

events across the entire population. Skeptical of this conclusion, we embarked on evaluation of ANNs in detecting medication-taking events for new participants in the next experiment.

Table 3.2. Results of a two-layer ANN as a function of hidden neurons

Hidden Neurons	Accuracy Training	Accuracy Testing
100	98.90%	95.99%
90	98.95%	97.77%
80	98.58%	95.99%
70	98.96%	95.99%
60	98.25%	97.18%
50	98.93%	97.48%
40	98.41%	96.59%
30	99.08%	97.18%
20	98.24%	97.48%
10	99.06%	96.59%
Max	99.08%	97.77%
Min	98.24%	95.99%
Average	98.74%	96.82%

3.4.2 Results of Experiment #2

To better address the generalizability and practical applications of the presented detection mechanism, we tested the ability of ANN to identify the protocol-guided medication-taking event for a new participant. This approach allows the immediate use of the developed application on any new user without the need to retrain the network.

Therefore, ANNs were trained using bootstrapping to train a network on $n-1$ participants and testing with the n -th out-sample/participant dataset using the protocol-guided medication-taking events. This experiment was repeated 10 times for each of 31 participants by altering the hidden neurons from 10 to 100 in increments of 10; in total, 310 ANNs were examined.

Table 3.3 and Table 3.4 show summary of results for the first 10 of the 31 participants, for both training and testing accuracy scores, respectively, for the best and worst architectures. The **Hidden Neurons** column shows the number of hidden neurons in the configurations that produced the highest and lowest train or test accuracies. The averages in both tables refer to the average for the 31 participants. As an example, in Table 3.3 for Participant 1 (row 1), the best training performance was achieved by the configuration 4500-**20**-1 while best test performance (according to the corresponding Table 3.4) was achieved by the configuration 4500-**40**-1. The lowest corresponding train/test scores from the two tables were recorded by the configurations 4500-**30**-1 and 4500-**80**-1, respectively.

Several conclusions can be derived from the results shown in the two tables. First, detection of medication-taking events from new participants is possible with accuracies varying from 98% (participant 1) to 100% (participants 2 and 7) with an average of 99.7% across all participants. However, the optimal performance corresponds to a different number of hidden neurons for each participant. This anecdotal observation agrees with the general expectation of human behavior where some people may exhibit a more complex behavioral signature while others exhibit a simpler behavioral signature. The complex signatures require a more capable ANN, which translates to a greater number of hidden neurons. The results shown in Table 3.5 are the average performance of each ANN configuration across the entire sample.

Based on these results, an ANN with 60 or 100 hidden neurons exhibits an average performance of 96.8% across all participants and therefore, while not optimally configured for any one participant, they perform consistently well across our entire cohort. It further shows that no one model was the best fit for all participants. This was perhaps due to the fact that each participant's hand motions have some degree of uniqueness or signature as illustrated by Figure 3.2 and Figure 3.3. The latter shows that gestures from the same participant vary. However, there is clearly an emergent motion pattern in all the gestures.

Table 3.3: Training Accuracy Results of ANN training using a bootstrap approach after experimenting with 10 different hidden layer sizes for each excluded participant

Participant	Accuracy Scores		Hidden Neurons Count for	
	Highest	Lowest	Highest Accuracy	Lowest Accuracy
1	97.48%	96.30%	20	30
2	98.10%	94.44%	60	50
3	97.54%	92.78%	10	90
4	96.78%	95.55%	30	90
6	97.70%	97.09%	40	10
AVG	97.49%	93.60%		

Table 3.4: Testing Accuracy Results of ANN training using a bootstrap approach after experimenting with 10 different hidden layer sizes for each excluded participant.

Participant	Accuracy Scores		Hidden Neurons Count for	
	Highest	Lowest	Highest Accuracy	Lowest Accuracy
1	97.99%	31.66%	40	80

2	100.00%	84.82%	10	90
3	100.00%	79.44%	50	10
4	100.00%	100.00%	10	10
5	100.00%	100.00%	10	10
6	100.00%	100.00%	10	10
AVG	99.11%	86.79%		

Table 3.5: The average performance of ANNs for each architecture.

Hidden Neurons	Avg. Training Accuracy (%)	Avg. Testing Accuracy (%)
100	96.76%	93.25%
90	96.40%	96.49%
80	96.54%	95.42%
70	96.79%	97.25%
60	96.77%	97.12%
50	96.52%	97.79%
40	96.64%	96.48%
30	96.42%	96.59%
20	96.82%	96.90%
10	94.93%	94.50%

3.4.3 Results of Experiment #3

This experiment was conducted with an out-sample natural gesture dataset as follows: The training dataset comprised of all protocol-guided data of n-participants plus (n-1) natural datasets. The nth natural gesture dataset was used as the test set. Table 3.6 shows the highest, lowest and average accuracy scores as well as the number of observations used

in the training and testing procedure for the first cohort of participants who completed the data collection process successfully (n=10).

For experiment #3, using the data from participants' natural gestures, we also tested all the 10 configurations as was the case with the protocol-guided data.

Table 3.6: Performance by participant

Participant	Highest (%)		Lowest (%)		Averages (%)		Dataset Sizes	
	Train	Test	Train	Test	Train	Test	Train	Test
User1	97.7	98.3	96.5	94.7	97.3	97.7	5228	57
User2	97.8	98.2	96.9	69.6	97.3	89.1	5230	56
User3	97.3	98.2	77.5	46.3	94.7	91.3	5234	54
User4	97.1	100	95.0	69.6	96.4	94.8	5250	46
User5	97.5	100	95.6	84.8	96.7	96.5	5250	46
User6	97.4	100	96.3	77.8	97.0	85.8	5172	45
User7	97.2	100	96.1	100	96.9	100	5254	44
User8	97.5	100	96.0	84.0	96.8	98.4	5292	25
User9	97.5	100	95.7	95.8	96.8	96.3	5294	24
User10	97.6	100	96.0	82.4	96.7	93.5	5308	17
AVG	97.5	99.5	94.2	80.5	96.7	94.3		

Note that all tests were conducted on the trained models with out-sample natural gesture datasets; the sample was excluded from the model training.

3.5 Conclusions

Three experiments were conducted to identify participants' medication-taking gestures. We demonstrated that we could leverage smartwatches non-obtrusively to harness the

power of technology to consistently identify medication-taking gestures. Neural networks predicted medication-taking gestures that were greater than 97% accurate for protocol-guided data and 95% accurate for natural medication-taking gestures. Importantly, we were able to establish that every person has some uniqueness in medication-taking hand-motions. We trained different fitting models to suit each of these unique characteristics. The ability to accurately identify medication-taking gestures has the potential to improve medication adherence monitoring and translate to better population health outcomes and reduced health care costs. Combining medication reminders through SMS notifications or the use of conversational agents such as Amazon Echo may be particularly effective to improving medication adherence rates.

3.5.1 Future Work

In our future work, we intend to evaluate distinct parts of the medication-taking gestures as well as consider the full gesture. Correct recognition of parts of the whole may better distinguish medication-taking gestures from other similar gestures such as drinking in the absences of medication-taking. This may be better achieved using Long Short-Term Memory (LSTM) recurrent neural networks which are architecturally suited for order and sequence prediction problems. Finally, this study was done with a single smartwatch worn on the wrist of the right hand. This meant that majority of the recorded and analyzed hand motions were based on the right-hand motions. In a few cases, we observed gestures that showed less pronounced motions. It is possible that in such cases, the participant wore the watch on the wrist that was not executing the actual medication motions for the natural gesture. Therefore, for future studies and for a more comprehensive analysis and

characterization of medication gestures, it will be useful to consider concurrent data collection using two smart watches on both hands.

Acknowledgments

This work was supported by NIH grant number P20 RR-016461 to Dr. Homayoun Valafar, the SmartState Advancing Chronic Care Outcomes through Research and iNnovation (ACORN) Center, College of Nursing, University of South Carolina (Dr. Cynthia Corbett, Director), and Ms. Wright's NIH grant number 1F31NR019206-01A1.

CHAPTER 4: HUMAN ACTIVITY RECOGNITION ON TIME SERIES ACCELEROMETER SENSOR DATA USING LSTM RECURRENT NEURAL NETWORKS³

Abstract – The use of sensors available through smart devices has pervaded everyday life in several applications including human activity monitoring, healthcare, and social networks. In this study, we focus on the use of smartwatch accelerometer sensors to recognize eating activity. More specifically, we collected sensor data from 10 participants while consuming pizza. Using this information, and other comparable data available for similar events such as smoking and medication-taking, and dissimilar activities of jogging, we developed an LSTM-ANN architecture that has demonstrated 90% success in identifying individual bites compared to a puff, medication-taking or jogging activities.

Keywords: Smartwatch, Accelerometer, Sensors, Artificial Intelligence, Machine Learning, LSTM, Human Activity Recognition, Eating, Bite, Food Intake

³ Related Publications and Authors:

Chrisogonas O. Odhiambo, Sanjoy Saha, Corby K. Martin, Homayoun Valafar. Proceedings of the 2021 International Conference on Computational Science and Computational Intelligence CSCI), "*Human Activity Recognition on Time Series Accelerometer Sensor Data using LSTM Recurrent Neural Networks*", Reprinted here with permission from the publisher

4.1 Introduction

Accurately assessing health behaviors in humans is necessary to evaluate health risk and effectively intervene to facilitate behavior change, improve health, and reduce disease risk. Health behaviors, such as eating, smoking, exercise (e.g., jogging), and medication-taking are frequently assessed with subjective self-report methods, such as diaries, which participants complete throughout the day. The accuracy of self-report methods is poor, however, particularly for assessing food intake and physical activity [150]. Self-report methods are also burdensome for participants, particularly if health behaviors need to be assessed over the long term [151].

Mobile health technology, including sensors worn on the body, can be used to passively and remotely collect and transmit objective data. These objective data can be much more valid and reliable compared to self-report, particularly for exercises such as walking [59]. The passive collection and transmission of data to researchers or clinicians have other advantages, including a dramatic reduction in participant burden and the ability to process and provide feedback to participants automatically and in real-time or near real-time. This critical step provides a platform to develop and deliver ecological momentary interventions (EMI) [60] and just-in-time adaptive interventions (JITAI) [61]. EMI and JITAI deliver intervention strategies that are customized to address the specific needs of individual participants as soon as these needs are detected. Participant needs are identified by evaluation of the objective data from the remote sensors in real-time or near real-time. Indeed, EMI and JITAI can provide more automated and cost-effective approaches to intervene and improve health behavior remotely while maintaining efficacy [62]–[64].

Here we report a novel application of Artificial Neural Networks to, objectively and automatically, identify and discriminate eating activity from three other activities namely smoking, medication-taking, and jogging using accelerometer data acquired from a smartwatch. Validation of the algorithm would make it possible to develop and deploy novel EMI and JITAI to improve these four health behaviors. Machine Learning algorithms have been used to achieve great results in developing practical solutions in multiple domains: health diagnosis [152]–[154], sports [143], [155], human activity recognition [18], [156], [157], among many others.

4.2 Background and Method

4.2.1 Previous and Related Work

Considering their rich array of sensors, the cost, accessibility, and ease of use, smartwatches have emerged as a compelling platform to study human activities unobtrusively. Smartwatches have been used as step-counters[116], sleep monitoring[118], diet monitoring[113] as well as general fitness tracking[119]. In the context of smoking, smartwatches have been demonstrated to be usable for in-situ study of smoking[120], [121] with high accuracy[107], [120]–[122]. Smartwatches have been used to detect smoking gestures with 95% accuracy in a laboratory environment[123] and 90% in-situ detection of smoking[107]. The study of smoking has also been demonstrated to be more accurate when compared to self-report (90% versus 78%)[107], [124].

Wearable devices have been utilized in indirect observation of some activities with clear health implications, such as medication adherence in numerous ways, including (1) self-report the behavior via mobile devices[145], (2) sensors worn around the neck e.g. the SenseCam[146] was originally envisaged for use within the domain of Human Digital

Memory to create a personal lifelog or visual recording of the wearer's life, which can be helpful as an aid to human memory, (3) multi-axis inertial sensors worn on wrists[147], [148], and (4) the use of commodity smartwatches [149]. In summary, all the approaches have collectively demonstrated the potential of smart sensors to promote the study of human activities but leave potential for improvement in performance, cost, convenience, and usability.

In this experiment we have utilized our collected data for eating activity and previously available data for smoking, medication-taking, and jogging recorded from human subjects. We have used this data to train and test an ANN capable of detecting eating behavior at the bite level.

4.2.2 Data and the Acquisition Process

This study involved four sets of activities, namely eating, smoking, medication-taking, and jogging. The selection of these activities was influenced by several factors including the availability of data, and the degree of their similarity. In particular, smoking and medication-taking were included in this study to challenge the detection of eating by providing similar behaviors. Three of the four activities (eating, smoking, medication-taking) consist of hand-to-mouth, hand-off-mouth, and hand-on-mouth sequence of events and should, therefore, provide a reasonable assessment of the network's performance.

The following three behaviors were recorded using smartwatches (Polar M600, Asus Zenwatch, Motorola, TicWatch) running android operating system: eating pizza (without cutlery/utensils), medication-taking, and smoking. The participants wore the smartwatches on the wrist (right or left). Each watch is equipped with accelerometer sensor to measure

the 3D acceleration, and gyroscope sensor to measure the 3D angular velocity. In this study, we only utilized the accelerometer sensor data sampled at a frequency of 25 Hz.



Figure 4.1: Illustration of sensor axes on a typical smartphone and smartwatch.

The fourth behavior evaluated, jogging, relied on an open public data from Wireless Sensor Data Mining (WISDM) Lab (<http://www.cis.fordham.edu/wisdm/>) that was recorded using smartphone strapped to the waist location of the participant.

Data recording using the smartwatch was performed independently at the convenience of the participant. The paired phone was only necessary in data transmission. Both the phone and the watch were installed with Android App. On the watch side of the activity, our software performs the following: (1) initiate data collection, (2) allow the user to indicate start and end of an activity during data collection, (3) transmit data to the phone via Bluetooth. On the phone, the application performs the following: (1) receive data from watch, (2) uniquely name files and upload them to the research cloud repository. The cloud is a webservice that receives and logs data files from the phone.

4.2.3 Data Pre-processing

Data pre-processing is a crucial step in the data-mining process. It involves data-filtering, replacement of the missing and outlier's values, as well as feature

extraction/selection. The windowing technique is commonly used to extract features from raw data. The technique involves the segmentation of the sensor signals into small time blocks with overlapping [9], [158]. There are three types of windowing techniques namely: (i) sliding window, where the signals are divided into fixed-length blocks/windows, (ii) event-defined windows, where specific events are located/identified and restructured as successive data partitioning, and (iii) activity-defined windows, where partitioning is based on detection of specific activity changes [9]. This study applies the sliding window approach, which is well suited to real-time applications because it does not require any pre-processing.

The raw data logged at the cloud repository comprises zipped sensor files. Each zipped file comprises two files: actual raw data of tri-axial values and the corresponding timestamps. The second file contains annotation data provided by the user that identifies the approximate time of the activity of interest in the stream of raw data. These two files were processed using an in-house developed utility program that extracted gesture features from the raw data based on the approximate start/stop timestamps reported by the user in the second file. It is important to note that the timestamps reported by users are not directly useful for several reasons.

First, a participant may report false start and stop times for a variety of reasons including simple human error. When accurately reported, the start and end portions of the signal will include hand movements that are unrelated to the activity of interest. For instance, hand movements related to clicking the start/stop buttons will also be included as part of the activity of interest. Therefore, to reduce noise in the data, increase its integrity and improve network model performance, the gestures were visually confirmed and

trimmed at both ‘tail’ and ‘head’ ends of each activity by a “supervisor.” Using this pre-processing pipeline, the usable output files were generated for training and testing of Machine Learning.

4.3 Neural Network Platform and Architecture

Although in this study we evaluated numerous ML approaches, in the interest of brevity, here we report the most successful approach that consisted of LSTM-ANN. Long short-term memory (LSTM) is an artificial Recurrent Neural Network (RNN) architecture with feedback connections [159], [160]. An LSTM unit comprises a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. LSTM networks find suitable applications in classifying, processing, and making predictions based on time series data. LSTM architectures address the vanishing gradient problem that can be encountered when training traditional RNNs. Figure 4.2 is an illustration of a typical LSTM cell where x_t is the input vector to the LSTM unit, h_t is the hidden state vector (or LSTM unit output vector), c_t is the cell state vector, and c_{t-1} is the cell input activation vector.

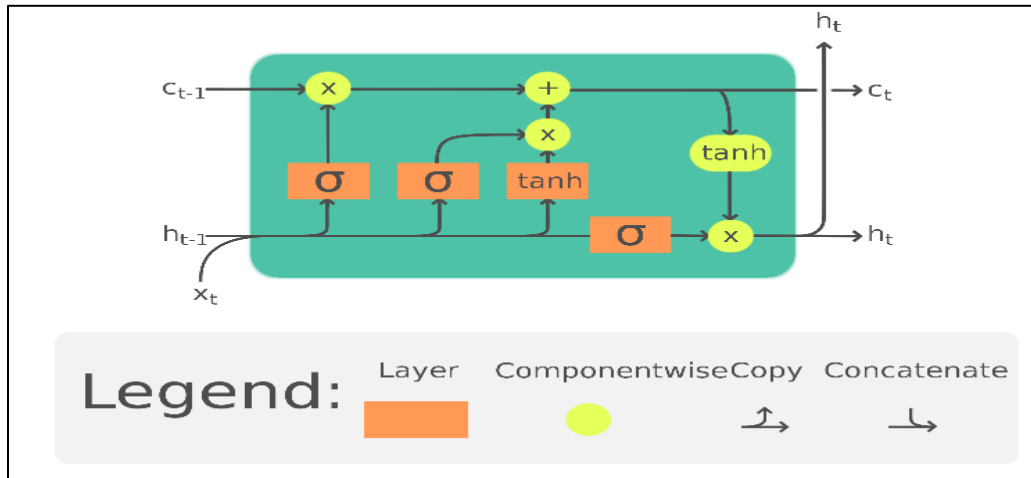


Figure 4.2: The LSTM cell can process data sequentially and keep its hidden state through time (By Guillaume Chevalier - File: The_LSTM_Cell.svg, CC BY-SA 4.0)

In the classification of sequential data, it is common to ignore the sequential aspect of data and treat the data as if it were independently and identically distributed (*iid*), and subsequently apply a standard machine learning classification algorithm that is designed for *iid* data [161]. It is important to note that the temporal dependence in the sequence of data will determine the effectiveness of the approximation. This, however, depends on how data is pre-processed i.e., is the raw data passed to the classifier as-is, or are features computed from the time-series and then in turn passed to the classifier. The human activities of interest in this study – eating-pizza, medication-taking, smoking, and jogging – are each sequence of mini-activities whose temporal aspect adds important component in the overall activity recognition. For example, the eating activity is made up of a series of mini-actions namely pick-pizza slice, raise slice to mouth, bite-pizza, lower slice from mouth. The sub-activities, and their sequence, is important. The past sub-features are useful and relevant as far as recognition of the full activity is concerned. For this reason, LSTMs find most relevant application in recognition of this human activity time-series data.

Our implementation of LSTM architecture comprised 2 fully-connected and LSTM layers (stacked on each other) with 64 units each. Stacking LSTM hidden layers makes the model deeper, more accurate, making it suitable for complex activities. Although the output of the system could have consisted of a single neuron to denote the presence or absence of eating event, we decided to design a two-neuron output. One of the two neurons is designated to the presence of an eating event and the other neuron indicates the presence of the other activities (smoking, medication-taking, jogging, or others). This design was used in light of our future expansion of the system to detect an array of activities. The input of the network (window size) was experimentally determined to consist of 150 consecutive points representing six seconds of recording time. The window step was maintained at 10 datapoints, and hence an overlap computed as $\text{Window-size} - \text{Step-size}$; for example, in the case of window-size of 50 units, the overlap is given by $(50 - 10 = 40)$. While we did not vary the step size, it is important to note that the smaller the step-size, the more real-time the data-series is. Smaller step-size improves performance, but increases window counts and slows down detection. The sliding window-with-overlaps process significantly transforms and reduces the training dataset. Further, the transformation assigns the most common activity (i.e., mode) as a label for the sequence; some windows comprise two or more activities, but the mode is considered the dominant or overriding activity. We transform the shape of our input into sequences of 150 rows, each containing x, y and z values (representing the accelerometer data). We also apply a one-hot encoding to the labels to transform them into numeric values that can be processed by the model [162]–[164].

4.3.1 Training and Testing Procedure

As the first step in training of a network, we balanced the data by randomly repeating a continuous segment of the recorded data so that all classes have an approximately equal representation. The train/test datasets were generated from the balanced dataset by partitioning in 80:20 ratio, respectively. During the training process, the learning rate was set at 0.0025 and the model was trained for 50 epochs while keeping track of accuracy and error. The batch size was maintained at 1024. We applied L2 regularization (Ridge Regression) to the model. The L2 penalty/force removes a small percentage of weights at each iteration, ensuring that weights never become to zero. The penalty consequently reduces the chance of model overfitting.

4.3.2 Evaluation

The study used accuracy to evaluate classifiers performances. The metric measures the proportion of correctly classified examples. Accuracy can be expressed as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (i)$$

Where TP (true positives) represent the correctly classified positive examples, TN (true negatives) represents the correctly classified negative examples, FP (false positives) represent negatives misclassified as positives, and FN (false negatives) represent positives misclassified as false. The accuracy measure does not take into account the bias arising from unbalanced datasets. Thus, the metric has a bias favor for the majority classes. For this reason, the study considered the following evaluation criteria: Precision, recall, F-measure, and specificity. Below are the formulae to compute the metrics:

$$precision = \frac{TP}{TP+FP} \quad (ii)$$

$$recall = \frac{TP}{TP+FN} \quad (iii)$$

F-measure is the combination of precision and recall. It is calculated as follows:

$$F\sim measure = \frac{(1 + \beta^2).recall.precision}{\beta^2 recall + precision} \quad (iv)$$

where β is a weighting factor and a positive real number i.e., the weighted harmonic mean of precision and recall, reaching its optimal value at 1 and its worst value at 0. The beta parameter determines the weight of recall in the combined score. It is used to control the importance of recall/precision. To give more weight to the Precision, we pick a Beta value in the interval $0 < \text{Beta} < 1$; To give more weight to the Recall, we pick a Beta Value in the interval $1 < \text{Beta}$. We applied a score of 1.

Specificity is computed as follows:

$$specificity = \frac{TN}{TN+FP} \quad (v)$$

4.4 Results and Discussion

4.4.1 Exploration and Visualization of the Data

Table 4.1 presents a summary of the total amount of data that was acquired or prepared for use in this study. In this table, the column denoted by Participants indicates the number of participants in each study, while Datapoints reports the total number of sampled data points collected across all participants. In essence, the number of Datapoints, when divided by 25 will correspond to the total duration of recording in seconds (e.g., a total of 3.03 hours of recording of the eating event). The column denoted as Patterns indicates the total number of presentations on the ANN that will contain the activity of interest (e.g., 5434 input patterns that contained an eating event). The large number of patterns is a result of a stream of data that presents different portions of the same activity (e.g., the same bite for

the same participant) to ANN. Therefore, one single gesture may consist of several input patterns to the ANN that consists of the activity of interest.

Table 4.1: Summary of all the datasets used in the study

Activity	Datapoints	Patterns	Participants
Eating pizza	272822	5434	10
Jogging	287461	5882	27
Medication	412798	3100	31
Smoking	62823	1279	15

As another critical step in annotation of data, it is important to visualize the data to become familiar with the intricate nature of each activity. This is a critical step for a “supervisor” to confirm and adjust the labeling of the outcome associated with each activity.

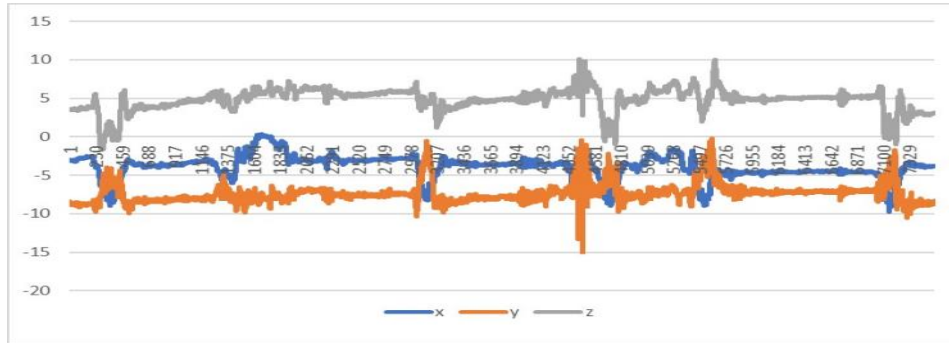


Figure 4.3: An entire recording of consuming pizza consisting of 6 bites that in total took 147 seconds.

Figure 4.4, Figure 4.5, Figure 4.6 and Figure 4.7 are visualizations of the tri-axial data for individual eating, smoking, medication-taking, and jogging activities, respectively. It is important to note that eating, smoking, and jogging consist of repetitious sub-activities

(a bite, a puff, a step) that are displayed in each of the corresponding graphs. Medication-taking activity, on the other hand, is composed of several sub-activities that appear in some temporal sequence and collectively appear only once. For instance, one medication-taking activity may consist of the sequence of opening a pill bottle, dispensing a pill, taking the pill, drinking water, putting everything back (pill bottle and water bottle, etc.). Previous work provides a detailed dissection of the medication-activity with an annotation of the sub-activities [18].

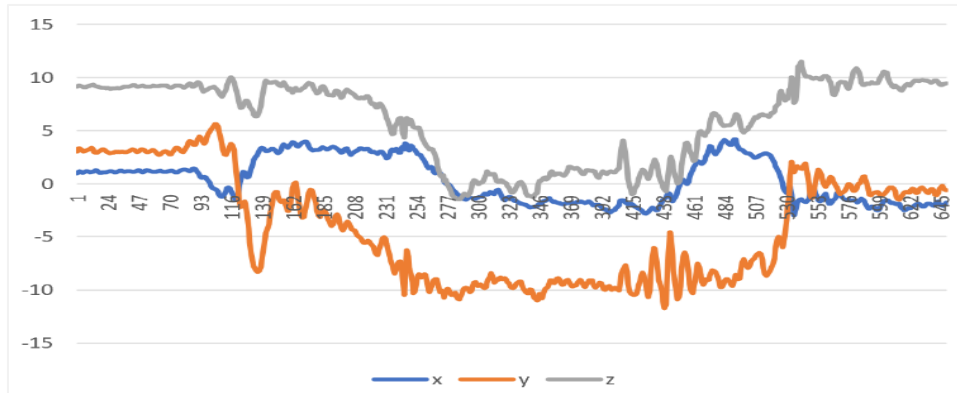


Figure 4.4: An eating gesture consisting of a single bite of pizza.

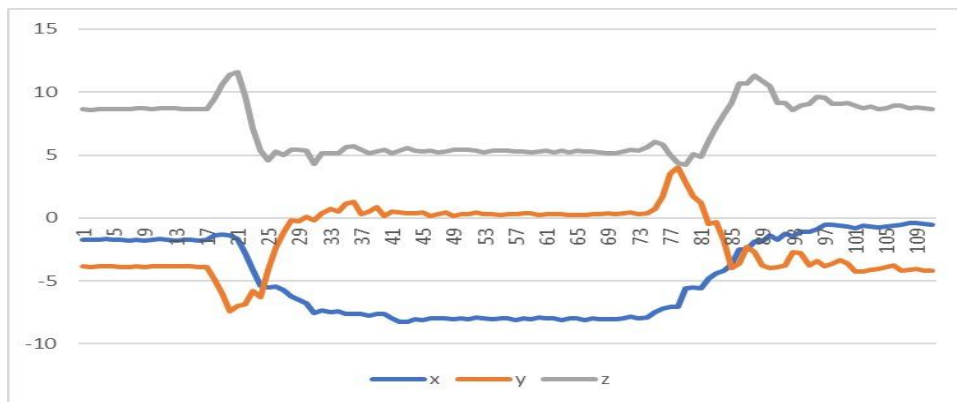


Figure 4.5: Single smoking gesture

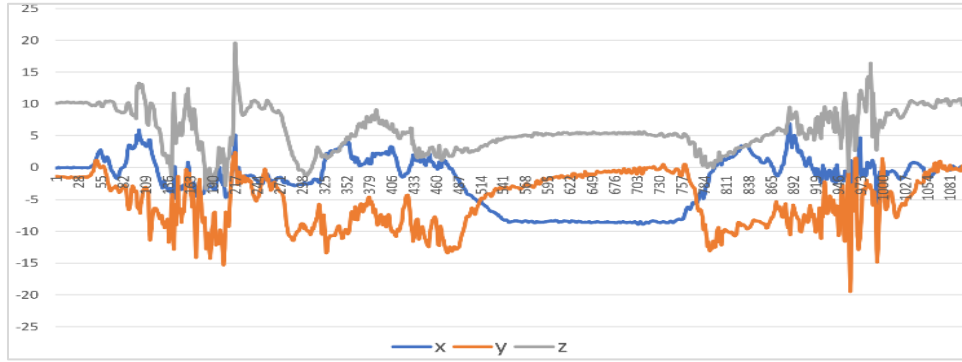


Figure 4.6: A single medication-taking gesture that consists of multiple sub-events (opening bottle, dispensing pill, taking pill, drinking water, etc.)

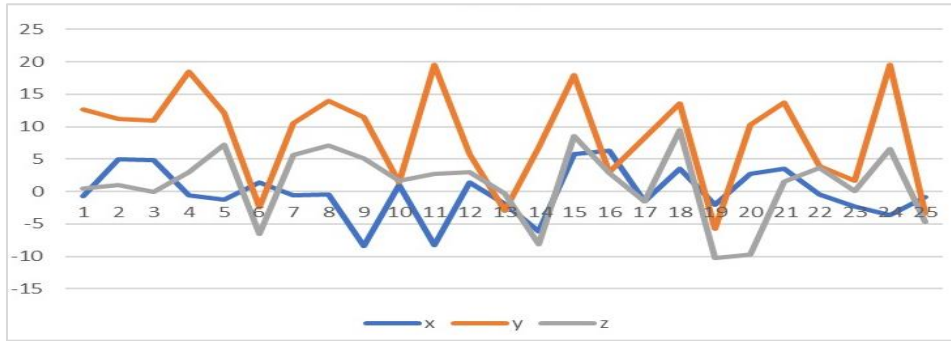


Figure 4.7: Three consecutive steps during jogging activity.

4.4.2 Training of ANN

The training of the LSTM-ANN proceeded upon the creation of the labeled input files (with balancing) using the hyper-parameters described in sections II.C and II.D. In this study we utilized TensorFlow/Keras as the primary ANN simulation platform. The training and testing process took place on Google Colab(atory). The Colab is a Google Research cloud service with a web IDE for python as well as computing services. The specifications of the computing resources we used were as follows: Intel(R) Xeon(R) CPU @ 2.20GHz, 13GB RAM, and 40GB of storage. The storage is upgradable to 108GB. The training of the system with 50 epochs required approximately 3 hours of dedicated compute time. The training and testing loss functions are illustrated in Figure 4.8. Based on the results shown

in this figure, it is clear that network has successfully learned the presented classification task.



Figure 4.8: Training plot for the window size of 150 units whose configuration produced the best performance among the different models.

4.4.3 Testing Results

Following the successful training of the LSTM-ANN, we embarked on testing of the system using 20% of the randomly selected original data set. Table 4.2 lists the results for the test set using different metrics. Based on the results shown in this table, the system achieved a performance of approximately 90% success. To explore the full nature of the misclassifications, the confusion table (shown in Table 4.3) was examined. The accuracy, precision, recall, F-measure, and specificity, as described by the formulae *i*, *ii*, *iii*, *iv* and *v*, respectively, are presented the Table 4.2.

Table 4.2: Metrics for the best performing configuration of window size 150

Activity	Precision	Recall	F-Measure	Specificity	Accuracy
Eating	0.89	0.97	0.93	0.96	0.96
Other	0.94	0.92	0.93	0.99	0.98

Table 4.3: Confusion matrix for the window size of 150 units

True label	Eating	5286	288
	Other	792	13587
		Eating	Other
	Predicted label		

4.5 Conclusion

Automated detection of eating activity is of critical importance in relation to obesity and healthy weight management. Automatic detection of eating sessions will help to remove the burden of self-reporting from the participants and therefore, provide a simpler way of tracking eating events. In this report, we demonstrated successful identification of individual bites with an accuracy of approximately 90% when tested against activities that significantly resemble to eating. In particular, smoking and medication-taking will share the common mini-gestures of hand-to-mouth, hand-on-mouth, and hand-off-mouth. Although not presented here, our initial investigation has confirmed very little confusion between eating and jogging. Therefore, we speculate the accuracy of the system to increase notably if other natural daily activities are included in our training and testing sets due to their dissimilarity to the eating gesture.

Although in this work we have achieved a reasonably high detection of the eating gesture, a number of additional investigations can be initiated to increase the performance and usability of the system. First, as an ultimate objective, we aim to develop one

application that can decipher numerous human activities to establish correlative or causative relationship between activities. For instance, eating at 1:00 PM may lead to a cigarette smoking soon after. The ability to monitor the temporal relationship between these two events would be very useful. To accomplish this, we need to engage in a formal investigation of the optimal viewable window size to and ANN that will be sufficient to successfully decipher between all activities of interest. Furthermore, there exists some inherent parallel between human activities and the principles of written language. To fully leverage this parallel analogy, human activities need to be examined in the more fundamental fashion by identifying the mini gestures that are the basis set of all complex activities. Here we will resort to some of the previous work [156] in order to understand the mini-gesture decomposition of the eating activity in relation to other similar activities such as smoking.

CHAPTER 5: DETECTING MEDICATION GESTURES USING MACHINE LEARNING AND ACCELEROMETER DATA COLLECTED VIA SMARTWATCH TECHNOLOGY: A FEASIBILITY STUDY ⁴

5.1 Abstract

Background: Medication adherence is a complex human behavior associated with chronic condition self-management. Medication adherence is a global public health challenge, as only about 50% of people adhere to their medication regimes. Smartphone apps and reminders have shown promising results in promoting medication adherence. However, practical mechanisms to determine whether a medication has been taken or not, once people are reminded, have been elusive. Emerging smartwatch technology may more objectively, unobtrusively, and automatically detect the medication-taking than currently available methods.

⁴ Related Publications and Authors:

Chrisogonas O. Odhiambo, Luke Ablonczy, Pamela J. Wright, Cynthia F. Corbett, Sydney Reichardt, Homayoun Valafar. Submitted to Journal of Medical Internet Research (JMIR) 2022, "*Detecting Medication Gestures using Machine Learning and Accelerometer Data Collected via Smartwatch Technology: A Feasibility Study*", Reprinted here with permission from the publisher

Objective: This study aimed to examine the feasibility of detecting natural medication-taking gestures using smartwatches.

Methods: Recruited participants (N=28) ranged in age (20 to 60 years) and comprised 57.0% males and 43.0% females. The majority were college students (71.4%), single (86.0%), and working at least part-time (61.0%). The sample represented racial diversity with 4% African American, 43% Asian, 43% White, and 10% reported two or more races [165]. Most participants were right-hand dominant (82%), while only one participant (4%) was ambidextrous. During data collection, each participant recorded at least five protocol-guided (scripted) medication-taking events (sMTE) and at least ten natural instances of medication-taking events (nMTE) per day for 5 days. Using a smartwatch, the accelerometer data was recorded for each session at 25Hz of sampling rate. The raw recordings were scrutinized by a team member to validate the accuracy of self-reports. The validated data were used to train an Artificial Neural Network (ANN) to detect a medication-taking activity. The training and testing data included previously recorded accelerometer data from smoking, eating, and jogging activities in addition to the medication-taking data recorded in this work. The accuracy of the model to identify medication-taking was evaluated by comparing the ANN's output to the actual output.

Results: In total, 2,800 medication-taking gestures (1400 natural plus 1400 scripted gestures) were used to train the network. During the testing session, 560 nMTE events that were not previously presented to the ANN were used to assess the network. Various metrics, such as accuracy, precision, and recall, were calculated to confirm the performance of the network. The trained ANN exhibited an average True-Positive performance of 96.5% and an average True-Negative performance of 94.5%. The network exhibited less

than 5% error in incorrect classification of the medication-taking gestures.

Conclusions: Smartwatch technology can provide an accurate, non-intrusive means of monitoring human behaviors such as natural medication-taking gestures. The use of machine learning algorithms combined with modern sensing devices may significantly improve medication adherence and monitoring.

Keywords: Machine Learning; Neural Networks; Automated Pattern Recognition; medication adherence; Ecological Momentary Assessment; Digital Signal Processing; Data Mining.

5.2 Introduction

Over three decades of international research has indicated that complete models of human health comprise complex interactions of biological, behavioral, and environmental factors. While there have been substantial technological advances in studying the biological and environmental bases of diseases, there have been relatively minor advances in technologies for characterizing human behaviors that influence health. Technological devices have pervaded and revolutionized much of our social and private lives, yet their implementation and utilization in healthcare remain sparse. In particular, the innovative use of existing, widely used, commercially available technology to influence health-promoting behaviors has been under-utilized. Adapting smart technologies, such as phones and watches, have the potential to initiate more effective health-promoting interventions for behaviors such as weight loss, physical activity, and chronic condition self-management. Adapting these devices to promote healthier behavior requires solving the crucial problem of characterizing and monitoring human behavior in a way that will be most useful, unobtrusive, and personally relevant. Once resolved, the subsequent step of

developing the optimal intervention mechanisms and personalized interventions can be explored.

Better understanding of daily activities such as eating, smoking, sleeping, exercising, and medication-taking can have a significant impact on population and individual health, with the potential to significantly reduce overall healthcare costs worldwide. In this study, we focus on the universal challenge of medication adherence. Medication adherence, defined as taking medicines according to decisions agreed upon between prescribing healthcare professionals and patients,[166], [167] is a complex human behavior associated with chronic condition self-management. Medication adherence is a global public health challenge, as only about 50% of people adhere to their medication regimes [168]. Studies have identified forgetfulness as the top reason for non-adherence to many long-term medicines.[169] To address forgetfulness, findings from a meta-analysis of medication adherence interventions among adults demonstrated that linking medication-taking with existing daily routines and using behavioral strategies (e.g., prompts to take medication) are the most effective approaches to promote adherence.[169] Smartphone apps and other technology-based reminders also have shown promising results for promoting medication adherence.[128], [170]–[172] However, practical mechanisms to determine whether a medication has been taken or not, once people are reminded, have been elusive.

Different methods, both direct and indirect, exist to measure medication adherence, however, none are considered a gold standard. Direct measurements, such as clinical biomarker specimens or metabolites from pharmaceutical metabolism and direct observations of medication-taking, can be expensive and impractical in large population

settings [173]. Indirect methods, such as pill counts, electronic monitoring, and self-report, offer simpler alternatives but, at best, approximate adherence through proxy data that can be initially overestimated with even less reliability over time[174]. An ideal method to measure medication adherence should be accurate, affordable, and practical (i.e., easy to implement).

Recent advances in sensor technology and artificial intelligence (AI) present an innovative opportunity to measure medication adherence objectively, unobtrusively, and conveniently. Wearable devices such as smartwatches may offer the platform to observe medication adherence, as well as people's other daily activities [115], [175], [176]. From a modest 37 million units in 2016, smartwatch shipments around the world are projected to grow to 253 million units by 2025 [177] As prices for smartwatches continue to decline, be equipped with additional sensors, and offer more mHealth applications, they are likely to become as pervasive as cell phones. Anticipating this trend, our team is investigating smartwatch use for human health promotion, including as a tool that can be used not only to remind people to take their medicine, but also to monitor medication adherence. In this report, we present an Artificial Neural Network (ANN) approach[18] that can detect the complex behavior of medication-taking, called the natural Medication Taking Event (nMTE), with as high as 95% accuracy using sensor data available from common smartwatches. To challenge the ability of the trained ANN in deciphering nMTE from other similar gestures, our experiments included sensor recordings of smoking, eating, and jogging. The eating and smoking gestures constitute a good basis for testing the system's ability to identify nMTE.

5.3 Methods

5.3.1 Overview

The use of sensors to automatically detect human activities was pioneered by the work of the Neural Network house and reported in the late 1990s. [178]–[181] in the more recent decade, inspired by the introduction of smart wearable devices, Human Activity Recognition has expanded to include activities such as detecting cigarette smoking [122], [135], falls,[182]–[184] and sleep [118], [185]. Sleep activity has been studied further using sensor data obtained from electroencephalogram (EEG) and electromyogram (EMG) devices to develop neural network models to examine and ascertain the sleep state of rodents in order to understand the sleep behavior in humans[186]. In relation to detecting medication-taking (MT), numerous approaches and technologies have been introduced including experimental devices worn on wrists, [18], [127], [131], [149] sensors worn around the neck to detect swallowing, [187]–[189] and vision modules embedded in smart environments such as the Microsoft’s EasyLiving project.[190], [191] The EasyLiving project showcased the early investigations into context-aware computing using an array of video-capture devices instead of more traditional physical sensors. By employing several vision modules in each room, the system could identify motion, people, gestures, and the surrounding environment. The project also focused on geometric relationships between people, places, and things to build context and form interaction information that would associate objects with their likely use, which could later be used in a more intelligent system for behavior prediction. Although vision-based medication adherence monitoring is a viable human activity recognition method, users’ privacy concerns and the identifiable nature of the data act as heavy deterrents in the practical adoption of this method. On the other hand, sensor-based smartwatches provide a scalable and practical platform for

conveniently, unobtrusively, and securely studying human behavior conveniently in natural settings (e.g., people's homes). Our previous work highlighted the potential for smartwatches to monitor medication taking events under protocol-guided (scripted) conditions (sMTE), where all participants followed the same method of taking their medication (e.g., use of right hand to perform most activities) [18]. However, participants' nMTE may significantly depart from the scripted method. For instance, a person may prefer to take the pill with their right hand while drinking water with their left hand. To establish a more practical application of this technology, we explored the feasibility of detecting unscripted, and nMTE that extends our previous work. The ability to detect nMTE that is applicable to the general population is a powerful tool in more accurately quantifying medication adherence rate. It has the potential to be an effective intervention tool that can increase adherence, reduce accidental over-medication instances, and be used for medication adherence monitoring by support persons or health professionals. To test the capabilities of our detection engine, we used sensor data from medication-taking events (sMTE and nMTE) and similar activities such as eating and smoking, and a dissimilar activity of jogging.

5.3.2 Participant Recruitment and Data Collection Process

This study was conducted during the height of the COVID-19 pandemic and required a substantial departure from a traditional means of engaging human participants in sensor recognition studies, which has primarily occurred in laboratory settings. Participants (n=28) were recruited using snowball sampling. An appointment was made with each potential participant to explain the purpose, benefits, and risks of the study, and address any questions or concerns. After obtaining informed consent, the participant

completed a demographic questionnaire and then received a packet with a Smartwatch and phone, two charger cables, user manual, pill bottle, and placebo medication. The user manual was presented and discussed in detail to the participant. Prior to data collection research team members had a video meeting with each participant to train them to collect and transfer the data, which culminated with participants properly demonstrating the activities.

Collected Data. Figure 5.1 shows the two devices - smartphone and smartwatch with triaxial sketches. The smartwatch was used to collect data and the phone was used to upload the data to cloud storage. The participants wore the watch on their right wrist in the case of sMTE, or on their wrist of preference in case of nMTE. The collected data comprised hand-motion accelerometer sensor logs of the triaxial values, recorded by the watch at a sampling rate of 25 Hz. The data included the timestamp and orientation and acceleration of the hand during the medication event taking activities. The xyz-sensor values were logged to a CSV file by the medication-taking app on the watch. The file was moved to the phone via Bluetooth, periodically and asynchronously.

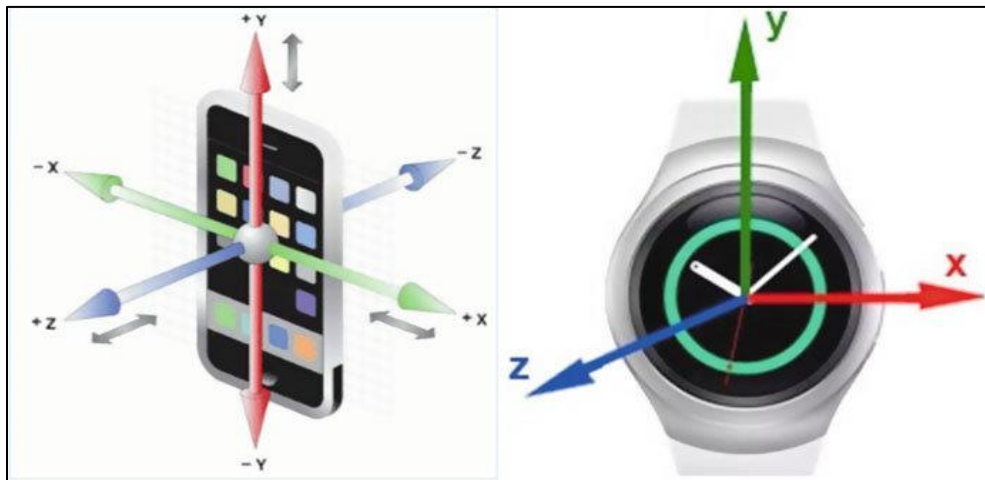


Figure 5.1: An illustration of Smartphone and smartwatch accelerometer axes

Data Collection Protocol. Following the medication-taking training sessions, participants

independently completed data collection in their homes. The exercise comprised one week (i.e., 5 days) of performing medication-taking behaviors using the participant's natural way of taking medications (nMTE) and a second week of performing medication-taking behaviors according to a scripted protocol (sMTE)[18]. Each participant (n=28) was directed to record 10 nMTE gestures per day for the first 5 days and then 10 sMTE gestures for the next 5 days. In total 1400 nMTE and 1400 sMTE were collected, tallying 2800 gestures.

To enable seamless transfer of data from the watches to cloud storage, each watch was paired with a smartphone. Both the watch and the phone were installed with respective versions of a custom android application called MedSensor, software developed by the research team. At the participant's convenience, watch data was relayed to the phone via Bluetooth connectivity.

Closure. After collecting and transferring 10 days of collected data, the participants returned the smartwatch and phone to the study project coordinator and received a \$25 gift card as an incentive. The smart devices were sanitized according to Centers for Disease Control (CDC) guidelines prior to use by other participants.

5.3.3 Data Preparation and Annotation

Proper use of data in supervised machine learning (ML) approaches requires reliable annotation of the data. The process requires a "supervisor" (an expert) to identify and define the gestures of interest to be used during the training of artificial intelligence (AI)/ML models. As a cumbersome process, the "supervisor" must have prior familiarity with the gesture of interest. To develop an understanding of what signal constitutes a medication taking activity, the scripted data were collected and used to understand the

individual components of the gesture. The exact details of the scripted gesture can be found in our previous report [18] and are briefly summarized in Figure 5.3. Using this information, team members then proceeded to identifying and annotating the individual gestures. The process of gesture identification and annotation was heavily accelerated by the self-reports that included time stamps indicating the beginning and ending of each MTE.

The raw data files logged at the cloud repository comprised a time stamp that included hour, minute, second, and millisecond; a date that included day, month, and year; and the x, y, and z components of the accelerometer data. A second file contained the time stamps corresponding to the start and end of each MTE reported by the participant. In theory, the self-reported MTE should be sufficient to identify the gesture of interest (i.e., medication-taking). However, in practice, participants may report the activity erroneously, or the time stamps may roughly indicate the start and end of a given activity. Therefore, visual confirmation of the gesture of interest is required to ensure high quality data. A separate utility program was developed to facilitate this process and to create the final usable data.[9] After this final step, the data files are presented in a usable format for the training and testing of the AI model.

5.3.4 Secondary Data Acquisition and Preparation

This study integrated accelerometer data from four different human activities, with the primary focus on recognizing MTE gestures as recorded by smartwatches (Polar M600, Asus Zenwatch, Motorola, TicWatch) as described above. The sMTE data was recorded by each participant wearing the smartwatch on the right wrist then sequentially performing the mini events of medication i.e., open bottle, dispense pills to right palm, toss pills to

mouth, drink water, and close bottle. For the nMTE, the participant performed the same mini events, in any sequence that they deemed natural i.e., how they would take medicine every other day. Smartwatch data for other behaviors (i.e., smoking, eating, and jogging) consisted of data reported in previous work [18], [192], while the MTE data were collected in this study using the protocol described above. The jogging dataset, on the other hand, is an open public data from Wireless Sensor Data Mining (WISDM) Lab[193] that was recorded using smartphone strapped to the waist location of the participant.

Table 5.1: Summary of all the datasets used in the study

Activity	Datapoints	Gestures	Participants
Medication	824,000	2,800	28
Eating	272,822	5,434	6
Smoking	62,823	1,279	12
Jogging	287,461	5,883	27

Data Preprocessing and Standardization. Prior to the integration of data from multiple studies, several normalization and standardization steps were performed. Specifically, attention was made to the consistent standardization of the accelerometer data and the sampling rate of the data. Because the devices used for data collection across all studies consisted of Android devices (versus Apple devices), the sensor data followed a common frame of x, y, and z axes. As the next step, all datasets were processed to adhere to a sampling rate of 25 Hz by excluding data points (in the case of oversampling) or resampling based on interpolation of the data (in the case of under sampling). To normalize for the different number of gestures per activity, the individual gestures were represented multiple times in our dataset to provide a balanced representation of activities.

5.4 Development of the Artificial Neural Network

5.4.1 Neural Network Platform and Architecture

The human activities of interest in this study – medication-taking, eating (pizza), smoking, and jogging – are each a sequence of mini-activities that have temporal properties, which is an important component in the overall activity recognition. For example, MTE comprise a series of mini actions namely (A) open-bottle and dispense-medicine, (B) Hand-to-mouth, pill-into-mouth and hand-off-mouth, (C) pick-up-water, drink-water, lower-cup-to-table and close-bottle (see Figure 5.3). The sub-activities, and their temporal sequence, are important. Therefore, the sub-features are useful and relevant for recognition of the full activity. Long-Short-Term Memory Neural Networks (LSTM-NN) are therefore, relevant in the application of human activity recognition because they contain internal memory of the past events in the analysis of time-series data. LSTM-NN is an artificial Recurrent Neural Network (RNN) architecture with feedback connections that facilitate awareness of past activities at the present time of the activity. [159], [160] Figure 5.2 is an illustration of a typical LSTM cell where x_t is the input vector to the LSTM unit, h_t is the hidden state vector (or LSTM unit output vector), c_t is the cell state vector, and c_{t-1} is the cell input activation vector. In this study, our model contained two fully connected and two LSTM layers (stacked on each other) with 64 units each. The learning rate was set at 0.0025.

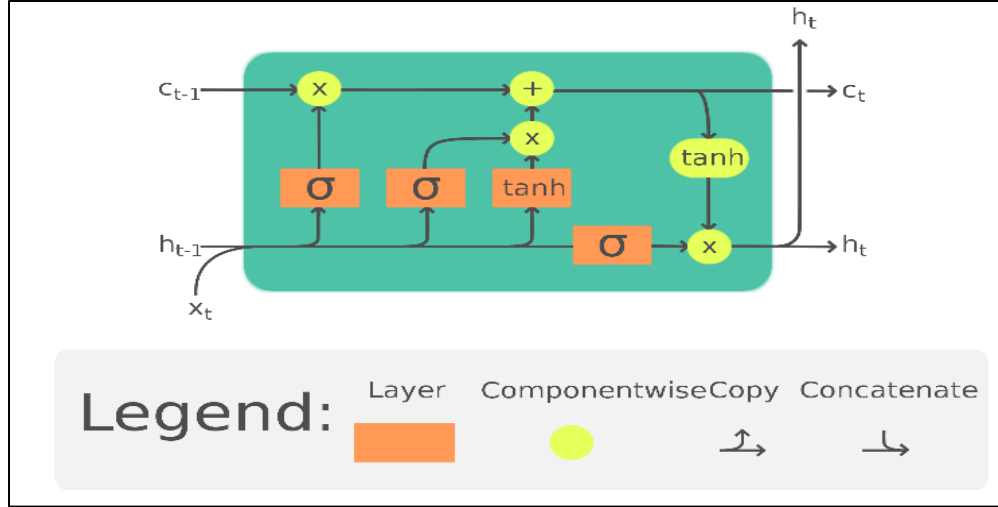


Figure 5.2: The LSTM cell can process data sequentially and keep its hidden state through time

5.4.2 Training and Testing Procedure

The LSTM network was trained for 150 epochs using the annotated data while keeping track of accuracy and error. The batch size was maintained at 1024. The train/test datasets were partitioned in the 80:20 ratio, respectively after the balancing procedure. We applied L2 regularization (Ridge Regression) to the model. The L2 penalty/force removes a small percentage of weights at each iteration, ensuring that weights never become zero. The penalty consequently reduces the chance of model overfitting.

The LSTM model expects the training data to be of fixed length. This fixed length of data that is directly presented to LSTM is referred to as the “window size.” In this study, a window size of 150 points was empirically determined to provide an acceptable performance. At a sampling frequency of 25Hz, a window size of 150 represents 6 seconds of recording. While the window size represents the portion of the raw data that should be in direct view of the ANN at the time of classification, any relevant past contextual information is saved in the internal cell of the LSTM architecture.

The temporal exposure of the LSTM-NN can be accomplished in a variety of ways including a sliding window of appropriate size. In this study a sliding window size of 10 points was selected as an optimal compromise between performance, simplicity, and responsiveness. The sliding window-with-overlap significantly transforms and reduces the training dataset. Further, the transformation assigns the most common activity (i.e., mode) in the exposed window of 150 points as a label for the sequence. This is necessary since some windows may comprise two or more activities, but the mode is considered the dominant or overriding activity. Consequential of the input definition, the data was reshaped into sequences of 150 rows, each containing x, y and z values with 10 points of overlap between two consecutive windows. The desired output of the system was based on one-hot encoding of the labels to transform them into numeric values that can be processed by the model [162], [164].

5.4.3 Evaluation of the Trained Network

During the training phase of an ANN, a single metric of performance needs to be defined to assess network performance. The network performance metric is used by the operator to direct the network to improve overall performance. In this study, we evaluated the performance of classifiers using the Accuracy metric as defined in Equation 1. In Equation 1, TP (true positives) represent the correctly classified positive examples, TN (true negatives) represents the correctly classified negative examples, FP (false positives) represent negatives misclassified as positives, and FN (false negatives) represent positives misclassified as false.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{eq(1)}$$

As a network performance measure, Accuracy does not account for the bias arising from unbalanced datasets. To remove the effect of unbalanced data (unequal representation of different activities), data within each activity were repeated to arrive at approximately equal number of representations for medication taking, eating, smoking, and jogging. Despite the adjustments to enforce dataset balance, some datasets remained larger than the rest, translating into a biased favor for the majority classes. For this reason, the study team considered the following additional evaluation criteria: Precision, recall, F-measure, and specificity. Precision indicates what fraction of positive predictions were truly positive. Recall (positive) indicates what fraction of all positive samples were correctly predicted as positive by the classifier (True Positive Rate). Recall (negative) indicates what fraction of all negative samples are correctly predicted as negative by the classifier (True Negative Rate).

Below are the formulae to compute the metrics:

$$precision = \frac{TP}{TP+FP} \quad \text{eq(2)}$$

$$recall = \frac{TP}{TP+FN} \quad \text{eq(3)}$$

F-measure is the combination of precision and recall. It is calculated as follows:

$$F\sim measure = \frac{(1+\beta^2).recall.precision}{\beta^2 recall+precision} \quad \text{eq(4)}$$

Where β is a weighting factor and a positive real number. It is used to control the importance of recall/precision.

$$specificity = \frac{TN}{TN+FP} \quad \text{eq(5)}$$

5.5 Results

5.5.1 Visualization of MTP Gesture

As the first step in performing activity recognition with wearable devices, a more detailed understanding of the gesture of interest needs to be developed. Figure 5.3 represents an example of an entire scripted MTE recorded from a right-hand dominant participant. After careful and repeated examination of the gesture, sequential segments of the gesture were identified (Figure 5.3). When considering the C portion of the gesture corresponding to the water-drinking (phase C), the gradual increase of the accelerometer's y-axis depicts the beginning of the drinking phase. It can be used both as a hallmark of an MTE and to quantifying drinking duration.

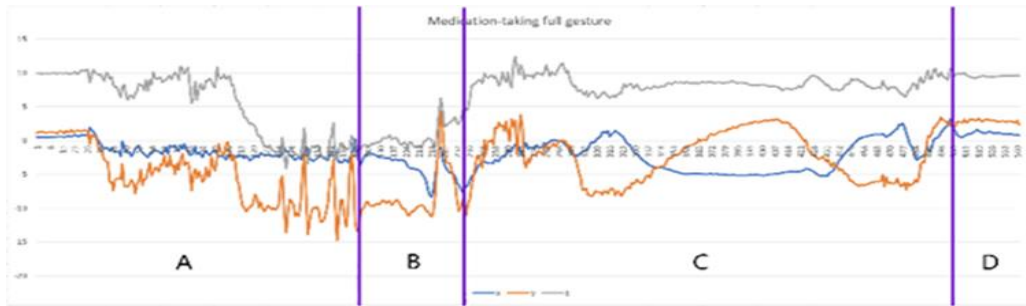


Figure 5.3: MTE complex activity atomic segmentations

Our medication-taking study consisted of scripted and natural methods of administering the medication. We based the scripted protocol on the natural behavior observed in most of our preliminary studies. Never-the-less, people's nMTE vary from the sMTEs as illustrated in Figure 5.4.

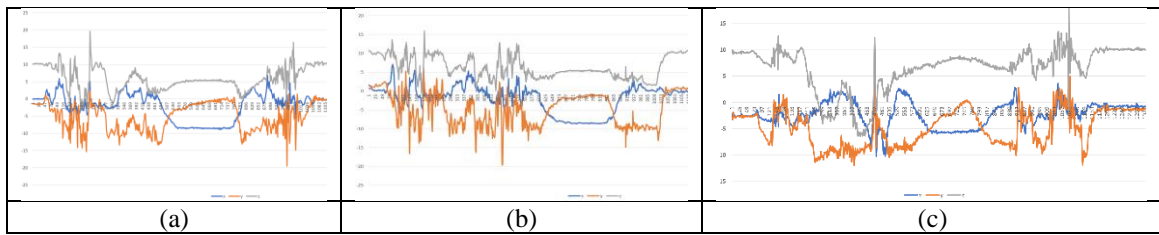


Figure 5.4: Illustration of MTP intra-class differences

It is important to highlight the challenging task of identifying nMTEs given gesture diversity in the mini activities among different participants. For instance, the simple method of drinking water between two participants can vary significantly as illustrated in Figure 5.4 panels b and c. The participant in panel b performs the task of drinking water with a sudden removal of glass from mouth, the participant in Figure 5.4 panel c removes the glass from their mouth more gradually. These differences in the individual mini activities are the root of the challenges associated with smartwatch gesture detection.

5.5.2 Validation and annotation of medication taking events

The study of human behavior with wearable devices has several advantages over the traditional self-report methods. Specific to medication-taking, self-reported adherence is known to be over-estimated[194]. In comparison to self-report, wearable devices can provide additional useful information such as the time of the day the medication was administered and the number of times the medication was taken in a day without incurring additional time, effort, or cost to the user. Table 5.2 displays additional data extracted from smartwatch MTE with MedSensor. Considering the average time needed to complete MTE, outliers can be examined for accuracy as shown in Table 5.3. In this study, we considered outliers (both natural and scripted) as gestures of duration 8 seconds and below for the lower category, or 100 seconds and above for the upper category. In determining the outliers, we considered the mean and standard deviations for natural gestures at 18.47 and 14.34, and scripted gestures at 20.11 and 20.65, respectively. The outliers were observably fewest in scripted than natural gestures. For the lower category, a random sample of 20 out of 103 gestures were examined, and all were invalid gestures, indicating that the users probably annotated start/stop of gestures in quick succession. On the other

hand, for outliers ≥ 100 seconds in duration, the majority contained one or more MTE gestures in most cases. To better understand the cause of these temporal discrepancies and therefore, validate or invalidate the reported gestures, each recording session was examined by a team member for validity. The results are depicted in Table 5.4.

Table 5.2: Time in seconds required for natural and scripted/protocol gestures.

	Mean (SD)	Median	Range (seconds)
Natural	18.47 (14.34)	17	5-331
Scripted	20.11 (20.65)	18	4-686

Table 5.3: Outliers count for natural and scripted medication gestures, based on gestures longer than 100 seconds or shorter than 8 seconds as outliers.

	Count	
Description/Category	Scripted Gestures	Natural Gestures
Duration ≥ 100 seconds	6	2
Duration ≤ 8 seconds	63	40

Table 5.4: Analysis of upper-category outliers in seconds.

Subject	Duration	Observations	Correction
U1	371	The participant reported 7 consecutively taken medications as one medication event.	Individual MTP events were separated by a ML supervisor.
U1, U2, U3, U4, U5, U6,	162	One MTP event was observed with some unrelated activities at the beginning or end of the recording.	The unrelated portions of the recording were trimmed.
U3	279	Comprises random activities that do not match medication gesture pattern	

5.5.3 ANN Training and Testing

As the first step in the training of a Neural Network, examination of the learning curve can be instrumental. Figure 5.5 provides an illustration of the learning curve of the designed LSTM-NN after the proper treatment of the outlier data. This figure illustrates the training accuracy (depicted in green) for the training and testing sets (dashed or solid lines respectively). Here, the consistently increasing values of accuracy is an indication that the network is successfully learning the classification task. The agreement of accuracies reported during the training and testing datasets indicate that the network is successful in generalizing the problem and not performing a memorization of the training patterns (avoid overfitting). The patterns shown in red describe the error function for the training and testing datasets (dashed versus solid lines). A decreasing value of error function is further indication of a successful learning with a gradually plateauing pattern that indicates a saturated training session. Table 5.5 summarizes the performance metrics for the final trained neural network that used a fixed window size of 150 epochs. The accuracy, precision, recall, F-measure, and specificity, as described by the equations 1, 2, 3, 4 and 5, respectively, are presented the Table 5.5. The test accuracies for eating, jogging, medication, and smoking were 94.3%, 100%, 93.6% and 98.6%, respectively. The average performance attained was 96.6%. To explore the full nature of misclassifications, the confusion tables (shown in Table 5.6 and Table 5.7) were examined.

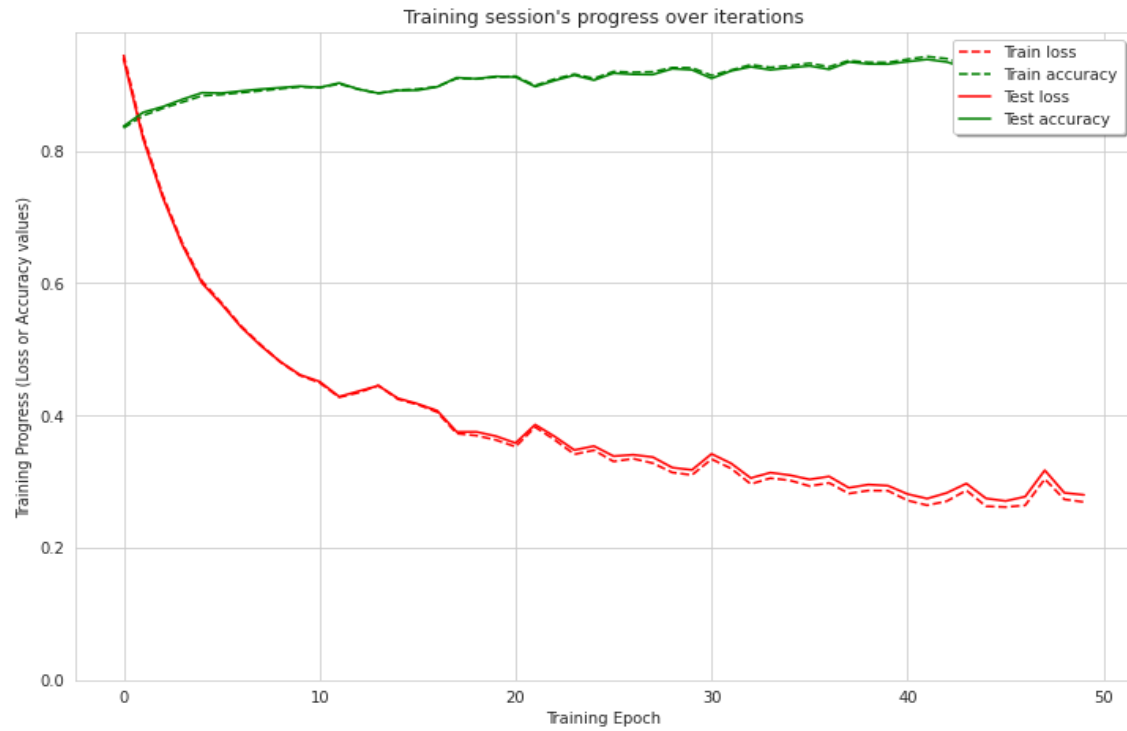


Figure 5.5: Training plot for the window size of 150 units.

Table 5.5: Metrics for the best performing configuration of window size 150

Activity	Precision	Recall	F-Measure	Specificity	Accuracy
Eating	0.802	0.920	0.857	0.948	0.943
Jogging	1.000	1.000	1.000	1.000	1.000
Medication	0.961	0.924	0.943	0.951	0.936
Smoking	0.888	0.769	0.824	0.996	0.986
Average	91.3%	90.3%	90.6%	97.4%	96.6%

Table 5.6: MTP/Non-MTP Confusion matrix for the window size of 150 units.

		Predicted label	
True label	True MTP	7412	708
	True Non-MTP	271	12224
		ANN_MTP	ANN_NonMTP

Table 5.7: MTP/Non-MTP Confusion Matrix (in %)

		Predicted label	
True label	True MTP	96.47273	5.474791
	True Non-MTP	3.527268	94.52521
		ANN_MTP	ANN_NonMTP

While the overall best performance of 96.6% accuracy is good, this study has not explored all possible nuanced configurations. Besides the sliding window size, it is possible to manipulate the hyperparameters such as the learning rate, adjust the number of LSTM units, windows step size, batch size, etc. to arrive at an even better performance. This is certainly an area for future exploration and further experiments.

5.6 Discussion

5.6.1 Principal Results

Studying human activities with smart and wearable devices has numerous advantages over the traditional approaches. Wearable devices provide the advantage of unobtrusively, continuously observing human behavior in their natural settings with little

burden to the user. The collected sensor data from these devices can be used to validate data reported by the user, therefore improving the accuracy and completeness of self-reports. In this work, participants used the smartwatch to report the beginning and end of their MTE. Errors in participants' self-reported MTE were identified. Some self-reported MTE had implausibly short or long durations. By validating the digitally recorded temporal gestures, we demonstrated the ability to correct erroneous self-reports, therefore improving the quality of the reports. Furthermore, the temporal signature that has been reported by the array of sensors available on wearable devices, can provide a plethora of additional information such as the temporal variation of an activity within a given user, or across a population of users. For example, in our study we demonstrated that nMTE were completed in an average of 18 seconds, but there were distinct differences across participants. Such comparison of behaviors provides several dimensions along which the study of human behavior can be expanded.

In addition to the expanded information that can be obtained from these devices, the ability to automatically detect and identify a nMTE with high accuracy will be beneficial. Automatic detection of an MTE event can be used as the foundation for both measuring and improving adherence. In the latter case, non-detection of an nMTE offers the opportunity to alert patients or support persons of the missed medications to improve adherence and the health outcomes associated with improved medication adherence. Improved medication adherence has the potential to significantly reduce morbidity [195]–[197], mortality [167], [197], and healthcare costs [195], [198]–[200]. Hence, detecting nMTE with smartwatches has exponential utility.

5.6.2 Limitations

The limitation of our approach given the current state of the technology is the availability of data from a single hand. Therefore, activities such as smoking that can be completed by a single hand, may not be detected by a watch that is worn on an opposite hand. Fortunately, because activities such as medication-taking are difficult to complete with a single hand, a residual portion of the activity will always be present from the perspective of a single watch. This problem can easily be overcome with the presence of a sensing device on each hand (or possibly each limb). Although not common presently, the arrival of smart wristbands, rings, and other forms of wearable is likely to provide a more complete picture of a person's daily activities. [141], [182], [184], [201]–[203]

A second limitation is the method by which people may wear their watch. A watch can be worn in four distinct ways of on left or right hand, and inside or outside of the wrist. In this study, we asked participants to wear their watch on their right hand, and on the outside of their wrists. However, sensor data recorded by the same watch in any of the other three configurations will produce related but undistinguishable signals by the ANN. Consistently wearing a watch on the outside right hand is critical at the current stage of our scientific development. However, using the existing human body symmetry, and the relationship between inside and outside of the wrists, mechanisms of unifying sensor signals collected from any mode of use can be developed as demonstrated previously [122]. This will produce high-capacity models with broader experience to recognize medication gestures regardless of watch-face orientation.

5.6.3 Proposed Improvements to Protocol

Generally, our current protocol worked very well. However, our future investigations will benefit from two additional steps to our existing protocol. The first step will collect calibration data during the initial orientation session. Currently, our orientation consists of familiarizing the participant with the watch, the app, and the use of the app. In the future, we will use this orientation session to collect data from a set of simple activities (e.g., touch toes, touch hip, and touch head) with the watch worn on each hand to obtain useful participant-specific data at baseline. By collecting data from left and right hands, we can establish a more precise relationship between the right and left hands for given participants. Although perfect human symmetry may indicate a 180-degree rotation between the two, natural human posture may create a departure from an ideal 180-degree symmetry. This information can be used to allow the user to wear the watch in any preferred method and provide the necessary information for the correction that is needed for the existing right-handed ANN.

5.7 Conclusions

Medication adherence is a complex human behavior associated with chronic condition self-management. It remains a global public health challenge, since nearly 50% of people fail to adhere to their medication regimes. Automated detection of medication-taking activity is of critical importance in relation to improved treatment effectiveness. The automatic detection of medication gestures will also help to eliminate the burden of self-reporting from the participants and therefore, provide a simpler way of tracking medication events. In this study, we demonstrated the use of LSTM-NN to detect and recognize mini activities, and the complex activities. We have demonstrated successful identification of

individual medication gestures with an accuracy of approximately 93.6% when tested against activities that significantly resemble medication, such as smoking and eating, which share the common mini-gestures of hand-to-mouth, hand-on-mouth, and hand-off-mouth. Our investigation also compared medication activity against jogging activity, and our models confirmed very little confusion between medication and jogging. This could be explained by the fact that the mini activities of both complex dynamic activities are largely different. By this observation, we speculate the accuracy of the system to increase notably if other natural daily activities are included in our training and testing sets due to their dissimilarity to the medication gesture.

Although in this work we have achieved a reasonably high detection of the medication-taking gesture, several additional investigations can be initiated to increase the performance and usability of the system. First, as an ultimate objective, we aim to develop one application that can decipher numerous human activities to establish correlative or causative relationship between activities. For instance, medication activity may occur at 8 PM before sleep activity, or at 7 AM before breakfast eating activity, or eating activity at 1 PM may be followed by cigarette smoking soon after. The ability to monitor the temporal relationship between these events would be very useful to provide the much-needed context to further understand human behavior and therefore model useful health-related solutions or provide real-time intervention reminders. To accomplish this, we need to engage in a formal investigation of the optimal viewable window size to an ANN that will be sufficient to successfully decipher between all activities of interest. Additionally, there exists some inherent parallel between human activities and the principles of written language. To fully leverage this parallel analogy, human activities need to be examined in the more

fundamental fashion by decomposing complex activities further to their most basic building blocks, referred in this study as mini gestures. Our previous work[18], [192] illustrates the mini-gesture decomposition of the eating activity in relation to other similar activities such as smoking.

Acknowledgements

Funding for this work was provided by NIGMS grant number P20 RR-01646100 (Dr. Homayoun Valafar), College of Computer Science and Engineering, and Center for Advancing Chronic Care Outcomes through Research and Innovation, College of Nursing, University of South Carolina.

CHAPTER 6: TOWARD CONCURRENT IDENTIFICATION OF HUMAN ACTIVITIES WITH A SINGLE UNIFYING NEURAL NETWORK CLASSIFICATION: FIRST STEP⁵

6.1 Abstract

Background: Medication adherence, smoking, and unhealthy eating are complex human behaviors associated with chronic conditions self-management and mental health. About 50% of people in the United States fail to adhere to their medication regimes, creating a major public health challenge. Smoking remains a leading preventable killer. Obesity epidemic is a public health crisis associated with poorer mental health outcomes and reduced quality of life. Smartphone apps and reminders have shown promising results in promoting medication adherence. However, practical mechanisms to determine whether a medication has been taken or not, or determine excess consumption of nicotine, or determine unhealthy eating, have been elusive. Emerging smartwatch technology may more objectively, unobtrusively, and automatically address this gap.

⁵ Related Publications and Authors:

Chrisogonas O. Odhiambo, Cynthia F. Corbett, Homayoun Valafar. To be submitted to Sensors (MDPI) 2022, "*Toward Concurrent Identification of Human Activities with a Single Unifying Neural Network Classification: First Step*"

Objective: This study aimed to examine the feasibility of detecting natural medication-taking gestures using smartwatches.

Methods: Recruited participants (N=28) ranged in age (20 to 60 years) and comprised 57.0% males and 43.0% females. The majority were college students (71.4%), single (86.0%), and working at least part-time (61.0%). The sample represented racial diversity with 4% African American, 43% Asian, 43% White, and 10% reported two or more races¹. Most participants were right-hand dominant (82%), while only one participant (4%) was ambidextrous. During data collection, each participant recorded at least five protocol-guided (scripted) medication-taking events (sMTE) and at least ten natural instances of medication-taking events (nMTE) per day for 5 days. Using a smartwatch, the accelerometer data was recorded for each session at 25Hz of sampling rate. The raw recordings were scrutinized by a team member to validate the accuracy of self-reports. The validated data were used to train a unified Artificial Neural Network (ANN) to detect a medication-taking, smoking, eating, and jogging activities. The training and testing data included previously recorded accelerometer data from smoking (12 participants), eating (6 participants), and jogging (27 participants) activities in addition to the medication-taking data recorded in this work. The accuracy of the model to identify medication-taking was evaluated by comparing the ANN's output to the actual output.

Results: A total of 2,800 medication-taking gestures (1400 natural plus 1400 scripted gestures), 5,434 eating gestures, 1,279 smoking gestures, and 5,883 jogging gestures were used to train the network. The training and testing datasets were split in the ration of 80:20, respectively. Various metrics, such as accuracy, precision, and recall, were used to measure the network performance. The trained ANN exhibited an average True-Positive

performance of 96.5% and an average True-Negative performance of 94.5%. The network exhibited less than 5% error in incorrect classification for the overall average accuracy.

Conclusions: Smartwatch technology can provide an elegant, accurate, non-intrusive means of monitoring human behaviors such as natural medication-taking gestures, smoking activities, as well as unhealthy eating events. The use of machine learning algorithms combined with modern sensing devices may significantly improve healthcare support and healthy lifestyles.

Keywords: Machine Learning; Neural Networks; Automated Pattern Recognition; Smart Healthcare; Ecological Momentary Assessment; context-aware environments, Human Activity Recognition.

6.2 Introduction

Over three decades of international research has indicated that complete models of human health comprise complex interactions of biological, behavioral, and environmental factors. Technological devices have pervaded and revolutionized much of our lives, yet their implementation and utilization in healthcare remain sparse. The innovative exploitation of existing, widely used, commercially available technology to influence health-promoting behaviors has been under-utilized. There is a huge potential in adapting smart technologies, such as phones and watches, to develop more effective health-promoting interventions for behaviors such as weight loss, physical activity, and chronic condition self-management. The successful adoption of these devices to promote healthier behavior requires solving the crucial problem of characterizing and monitoring human behavior in a way that will be most useful, less obtrusive, and personally relevant. When this is sufficiently realized, the

subsequent step of developing the optimal intervention mechanisms and personalized interventions can be explored.

Human Activity Recognition (HAR) has been treated as a typical classification problem in computer vision and pattern recognition to recognize an array of human activities in daily living as well as anomalies i.e., abnormal behavior or activities. Vision-based HAR techniques rely on image and video data to recognize behavior while Sensor-based counterparts rely on sensor data to achieve the same goal. The patterns are related to specific actions[1]. It aims to understand daily behaviors of people through the analysis of observation sensor or vision data obtained from people and their neighboring environments of living. Recent developments have seen a steady advancement, availability, and proliferation of miniature wearable sensor devices from necklaces to smart phones, smartwatches, etc. The accuracy of these devices can be affected by various factors such as lighting, background, crowded scenes, camera location and behavior complexity[204].

Human Activity recognition enables us to solve many human-centered problems, such as health care, individual assistance, the need to infer various simple to complex human activities. So far, HAR has found use in diverse domains such as healthcare, surveillance, sports, linguistics, event analysis, Human Computer Interactions (HCI)[3], among others. HAR has been applied to characterize human behavior, understand human interactions, improve daily life such as the case of continuous health monitoring, improving human safety and well-being, all over the world. Other areas of application include activity of daily living, physiological signals, quantified-self, postures detection, gestures detection, gait analysis, and indoor localization [4]–[6].

Wearable sensors come with significant benefits such as (i) Personalized health monitoring, (ii) Removal of healthcare barrier, allowing for evenly distributed health monitoring, and (iii) Affordability by majority populations. Further, it they make it possible for automatic crowd analysis in detecting threats and anomalous activities by analyzing crowd modeling, crowd tracking, density estimation and crowd behavior understanding. These sensors can help to recognize complex human actions, translating into significant social and economic benefits such as remote healthcare monitoring and response systems. Correct interpretation of motions such as falls, medication, eating, smoking, etc. can be used to trigger appropriate response of caregivers. This can be used to reduce risks for the elderly persons from potentially catastrophic tumbles, monitor medication adherence, eating habits or exercise activities.

Abnormal behavior depends on context; for example, throwing fists can be normal during an exercise, but abnormal during a domestic exchange or a physical bullying context. Machine learning models refer to mathematical systems that share many common features. Machine learning algorithms work with data to create associations, find relationships, discover patterns, generate new samples, as well as work with well-defined datasets. A machine learning problem is focused on learning abstract relationships that allow consistent generalization when new samples are provided [205], [206]. HAR being a pattern recognition problem of specific actions uses classifiers and action detection methods divided into three main categories: (1) Generative models which is a probability-based method to learn the statistical distribution of the underlying data distribution, (2) Deterministic models, which are static classifiers trying to learn the hidden feature representations from labeled data, (3) Non-parametric methods[1].

Accurate assessment of health behaviors in humans is necessary to evaluate health risk, and therefore, effectively intervene to facilitate behavior change, improve health, and reduce disease risk. Health behaviors, such as eating, smoking, exercise (e.g., jogging), and medication-taking are frequently assessed with subjective self-report methods, such as diaries, which participants complete throughout the day. The accuracy of self-report methods is poor, however, particularly for assessing food intake and physical activity [150]. Self-report methods are also burdensome for participants, particularly if health behaviors need to be assessed over the long term [151].

In this study, we aim to develop a unifying Neural Network model that can recognize multiple signatures of various human activities, making it possible for the use of wearable sensors to aid daily living activities such as monitoring seniors living alone at home for falls, late-night activity, sleeping habits, medication adherence, healthy eating, or smoking habits. Falls kill thousands of elderly adults each year and injure millions more. Highlighting risk factors or timely response could save lives, reduce insurance premiums, and help caregivers serve this population more efficiently. According to CDC, falls among adults aged sixty-five and older are very costly. It reports that each year the United States spends about \$50 billion on medical costs related to non-fatal fall injuries and \$754 million on injuries related to fatal falls [207]. According to [208], only one-third of those who fall seek medical care. In Netherlands, according to a study [209] on Health care costs of injury in the older population (trauma patients aged 65 years and above), the main cost drivers were the post-hospital costs due to home care and stay at an institution. Falls (72%) and traffic injury (15%) contributed most to the total health care costs, although costs of cause of trauma varied with age and sex. The study showed that in-hospital costs were especially

high in patients with high injury severity, frailty, and comorbidities. Several patient and injury characteristics including age, high injury severity, frailty and comorbidity were associated with post-hospital health care costs.

Poor medication adherence presents serious economic and health challenges, including compromised treatment effectiveness, medical complications, and loss of billions of dollars in wasted medicine or treatment procedures. Studies show that 33-69% of all medication-related hospital admissions in the United States (US) are caused by poor medication adherence, which translates to an annual cost of approximately \$100 billion [130], [131]. Annually in the US, non-adherence can account for up to 50% of treatment failures, approximately 125,000 deaths, and up to 25% of hospitalizations [132].

Smoking remains the leading cause of preventable death. Annually, tobacco use causes more than seven million deaths worldwide. According to CDC, over sixteen million Americans are living with a disease caused by smoking. In the US alone, 20% of the population report that they engage in smoking; diseases caused by smoking cost the population over \$170 billion in healthcare each year according to CDC.

In the United States, more than two-thirds of adults are overweight or obese [210], [211]. According to CDC, obesity epidemic is a public health crisis because it is associated with poorer mental health outcomes and reduced quality of life. Obesity is also associated with the leading causes of death in the US and worldwide, including diabetes, heart disease, stroke, and some types of cancer. Apart from proper health education and counterforce health care providers, healthy eating (foods) and regular exercise make for crucial preventive measures against obesity.

There are social and economic benefits of HAR such as assisting elderly people during their daily activities, increasing their safety, well-being, and autonomy. Activity recognition based on new wearable technologies remains an important research challenge. The elderly people living alone, physically, or mentally disabled persons, and children need continuous monitoring of their activities to detect abnormal situations or prevent unpredictable events such as falls [9]. Recognizing and monitoring human activities are fundamental functions to provide healthcare and assistance services to these populations. The new technologies of health monitoring devices range from on-body wearable sensors to *in vivo* sensors.

Sensors are converters that quantify the physical aspects of the world around us into electric values that can be perceived by a digital system [1]. This capability makes it possible to gain knowledge about human activities. This study considers the use of wearable sensors in HAR. Human activity recognition plays a critical role in helping to solve human-centered problems. Better understanding daily activities such can have a significant impact on population and individual health, translating into significantly reduced overall cost of healthcare worldwide. Sensor-based HAR generally broadly comprises five steps: sensor selection, data collection, feature extraction, model training and testing[2].

Vision-Based Systems embedded in smart environments or "Smart Homes [204], [212] . Although vision-based monitoring of medication adherence may be the primary choice of pursuit, the identifiable nature of the data acts as a heavy deterrent in the practical adaptation of this technology. On the other hand, smartwatches provide a scalable and practical platform for conveniently, unobtrusively, and securely studying human behavior

conveniently in natural settings such as homes. Recent advances in sensor technology and artificial intelligence (AI) present an innovative opportunity to measure medication adherence objectively, and conveniently. Wearable devices such as Smartwatches may offer the platform to observe medication adherence, as well as people's other daily activities [7]–[9].

From a modest 37 million units in 2016, smartwatch shipments around the world are projected to grow to 253 million units by 2025, making them as pervasive as cell phones. In complementary to this vast computing landscape is the entrenchment of the Internet of Things (IoT), Internet of Everything (IoE), as well as the convenience of miniaturization of wearable computing devices. These developments have essentially devolved significant computing power to things or objects. Wearables, fashion technology, smart wear, tech togs, skin electronics or fashion technology electronics are smart electronic devices that are worn close to and/or on the surface of the skin, where they detect, analyze, and transmit information concerning e.g., body signals such as vital signs, and/or ambient data and which allow in some cases immediate biofeedback to the wearer. In other cases, the feedback is returned to remote monitoring system such as the case with the remote health monitoring of the vulnerable persons like the elderly [213]–[215].

6.3 Methodology

6.3.1 Generic Procedure of Recognition

The recognition process involved the following steps, as further illustrated by Figure 6.1:

- (i) Data acquisition from inertial sensors, (ii) Data preprocessing for feature extraction, (iii) Model training and optimization, as well as model tests, (iv) Performance evaluation, and (v) Perform prediction.

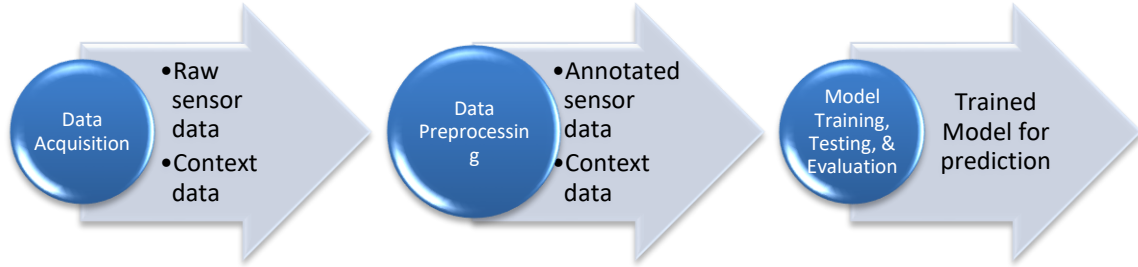


Figure 6.1: General process of activity recognition model preparation

6.3.2 The Data Process: Origination, Pre-processing, Storage, Analysis, and Retrieval.

Figure 6.2 is a high-level representation of the data process as proposed, developed, and implemented in this study. We christened this process as METASENS to loosely label it as the versatile data ocean for all wearable sensor data and related metadata artefacts. The implementation details are described in the next sections. In this study, we consider the data process as a critical component of research because the quality of research outcome is directly dependent on the quality of the data and the data process that supports the research. The data process comprises the following components:

A: Data Origination is the source of raw sensor data recorded by smartwatches. Besides the tri-axial values, we also capture timestamp and activity performed by the user. These features are written to a flat datafile saved locally in the smartwatch. Besides the data file, there is a second metadata flat file used to record annotation values indicating the start and end of user activity. Each of these points has date and time in the format

DD/MMM/YY HH:MM:SS. For example, the line below shows the user activity ran between 2nd February 2022 06:15:38 and 2nd February 2022 06:15:53 i.e., the activity took a total of 15 seconds.

REPORT_SELF: Recording, 02-Feb-10 06:15:38, 02-Feb-10 06:15:53

B: Data Route or Extended Data Origination Platform. This is represented by the smartphone. The smartwatch is limited in resources such as CPU power. In cases where more CPU power is necessary, such as map and GPS resources are part of the data required, some of the watch operations may be offloaded to the phone. More importantly, the phone provides a larger storage that may hold the data routed from watch to the phone. This may be useful in cases where participants may need to collect data over a period and delay to submit it to the cloud perhaps due to weak data speeds or purely as local backup. The phone may also be used to conduct surveys that do not need to be real time with the watch data activity recording events. The phone provides larger user interface that is much easier to interact with especially when it comes to typing and navigation. Most importantly, for our purposes, the phone provided a means to route data over from the smartwatch to the cloud storage. The watches we used for the exercise did not have cellular capabilities, meaning we had to pair them with smartphones, and use Bluetooth to move data from watches to phones, then from phones to cloud, we relied on internet connectivity via Wi-Fi or cellular data services. In some cases where we conducted experiments in the lab, we used the alternative means of direct data extraction from phone to PC as shown by the dashed line in Figure 6.2. The watch is a minicomputer with a file system just like any other computer. This is accessible via ADB commands or Android Studio. Through these means, the files can be accessed directly and copied to the PC.

C: Cloud Server and Web Services. This component provides the web hosting, file storage and database services. The phone routes files to the cloud server via the webservices that process and log the files into the remote storage. All files are assigned unique names before storage to avoid conflicts. The web server manages and controls web requests. The database server stores both the metadata as well as various user security data. The file system also allows users to upload datasets and their related videos or/and images. These extra files can be useful in describing a dataset, e.g., where a user records a video of how he/she performed an activity. The same applies to images, where user may share photos of the recorded activities.

D: Data Pre-processing and Analysis represents any local network access to the data for further preprocessing, particularly where a user needs to prepare the dataset(s) for a neural network training process or custom analysis.

E: Public Access Interface (PAI) provides any member of public the user interface to interact with the data system to perform activities such as creating their own projects, uploading, and managing their own datasets, uploading, and managing dataset related artefacts or metadata such as videos, images, related statistics, etc. PAI provides the user end of the METASENS that allows not only the onboarding, but also the consumption of data as a research resource by the research community. The datasets also allow for open public discussions around them. This is one of the ways by which meaningful active and open discussion can be encouraged among the users of these datasets. The METASENS forms part of a larger framework as described below and illustrated by Figure 6.3.

6.3.3 Proposed HARP Architecture

The HARP framework aims to define a workflow framework that allows a user to record and upload own HAR dataset using wearable sensor devices, upload it to METASENS cloud database, request for models trained on the personal dataset, and then download and deploy the models to own device for activity recognition. Below is a description of the framework:

A: METASENS Portal: This component is described above and is implemented in Python Django web framework and MySQL database. It stores and manages user datasets, data security, integrity as well as user accounts. The portal is illustrated by Figure 6.2.

B: ML Pipeline: represents the Machine Learning pipeline engine. It can be implemented as TensorFlow Extended (TFX) or MLflow. These two solutions are available as open-source tools to the public. The ML Pipeline manages the Machine Learning workflow from data input, to training, testing and deployment.

C: HARP Models Web server: When the ML Engine is done by training and testing, it deploys the ready models to the HARP Models Web Server that hosts the models for the users. The web server runs web server applications that serves content to the public through web browser or webservice access.

D: User: This component represents any public user of the workflow who wants to record and share datasets about HAR activities. The user can upload own datasets to METASENS. The user can also send a request to get models trained on own datasets. For example, if a user wants to monitor medication, the user can record data using smartwatch, upload the data to METASENS, use the same interface to request for a model trained based on the uploaded medication dataset, and deploy that model to local device, smartwatch in

this case. Once deployed, user can use the model or simply run HAR tests. This means that the workflow will require model installer into the smartwatch. It is possible to turn the web server into a public repository resource or library of human activity models called HARPs, which can be shared by portal users, if HARP owner avails own models to the public. This can be seen in a sense as a social media of Human Activities.

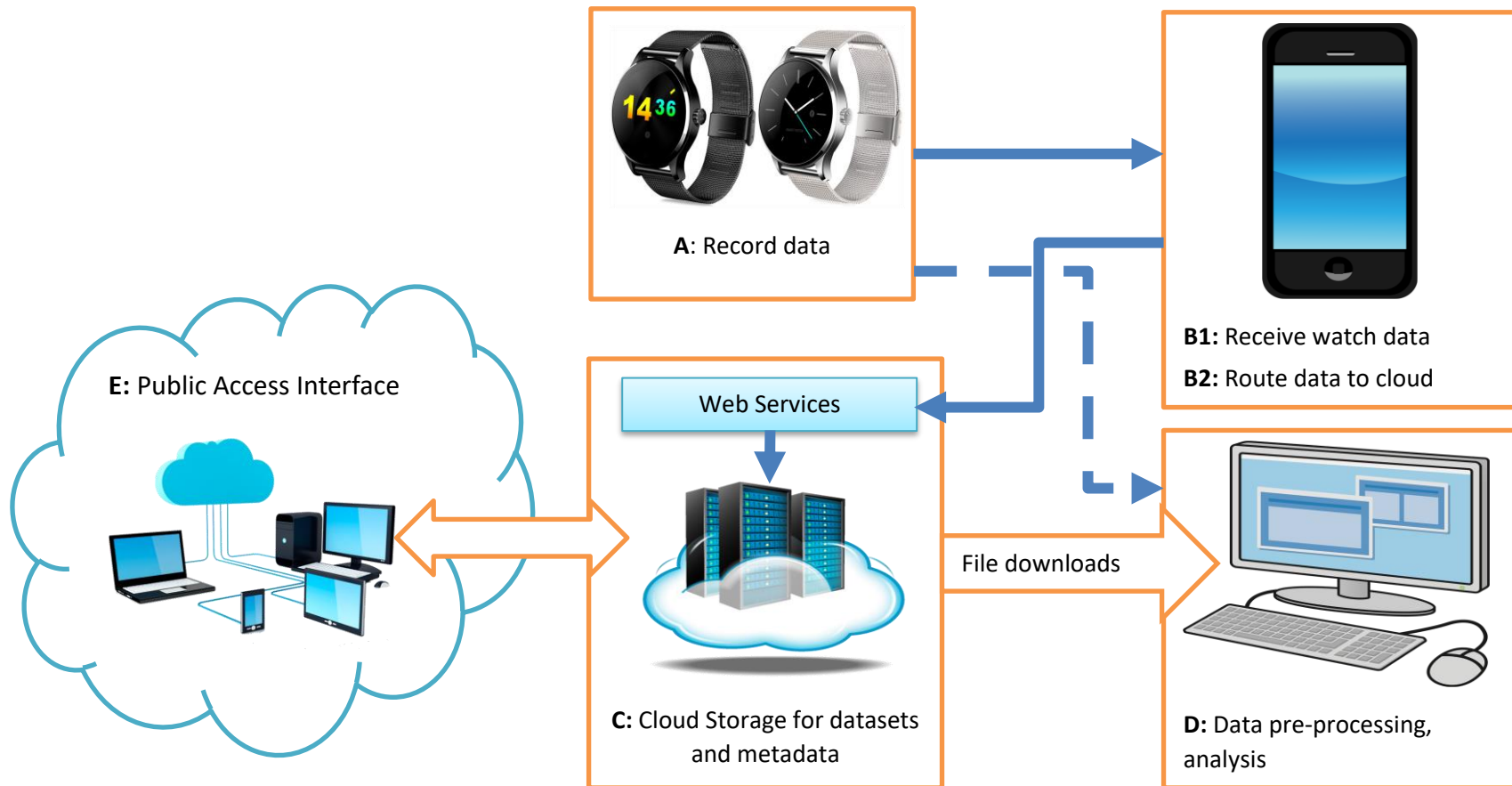


Figure 6.2: High-level framework of the data process from data origination to storage, to retrieval, and analysis.

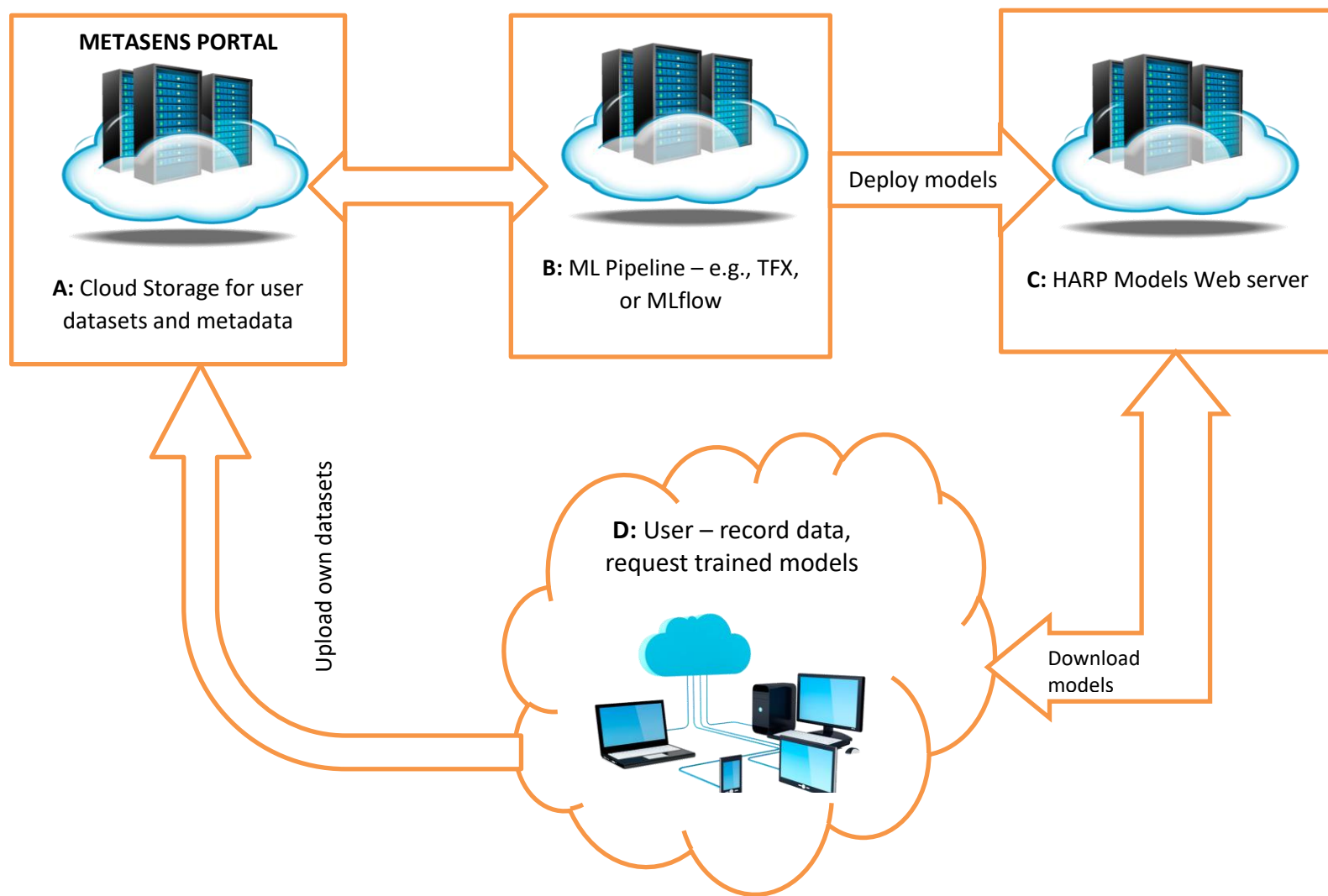


Figure 6.3: HAR as Personalized AI Models (HARP) high-level framework

6.3.4 Data Acquisition

Figure 6.4 shows a smartphone and a smartwatch with triaxial sketches. The smartwatch was used for actual data collection while the phone was used to upload the data to the cloud storage. The participants wore the watch on their right wrist in the case of scripted gestures approach, or on their wrist of preference in case of natural gestures approach. The collected data comprised hand-motion accelerometer sensor logs of the triaxial values, recorded by the watch at a sampling rate of 25 Hz. The data included the timestamp, orientation, and acceleration of the hand during an activity – medication-taking, smoking, or eating. The xyz-orientation values were logged to a CSV file by the MedSensor App on the watch. The file was moved to the phone via Bluetooth, periodically and asynchronously.

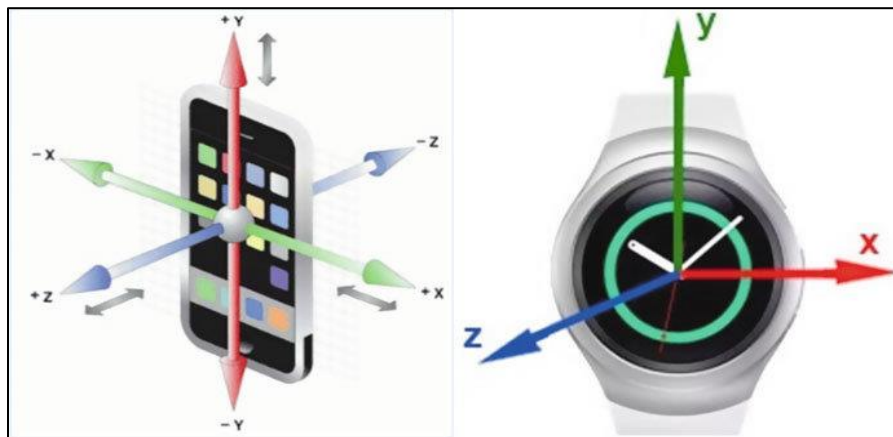


Figure 6.4: An illustration of Smartphone and smartwatch accelerometer axes

6.3.5 Participant Recruitment and Data Collection Process

a. Medication Data Acquisition

Participants (n=28) were recruited using snowball sampling. An appointment was made with each potential participant to explain the purpose, benefits, and risks of the study and address any questions or concerns. After obtaining informed consent, the participant completed a demographic questionnaire and then received a packet with a Smartwatch and

phone, two charger cables, user manual, pill bottle, and placebo medication. The user manual was presented and discussed in detail to the participant either in person or remote meetings. Additional training included an online question/answer session and demonstration to ensure the participant was properly collecting and transferring data.

Data Collection Protocol. Following the medication-taking training sessions, participants independently completed data collection in their homes. The exercise comprised one week (i.e., 5 days) of performing medication-taking behaviors using the participant's natural way of taking medications (nMTE) and a second week of performing medication-taking behaviors according to a scripted protocol (sMTE) [18]. Each participant (n=28) was directed to record 10 nMTE gestures per day for the first 5 days and then 10 sMTE gestures for the next 5 days. In total 1400 nMTE and 1400 sMTE were collected, tallying 2800 gestures. To enable seamless transfer of data from the watches to cloud storage, each watch was paired with a smartphone. Both the watch and the phone were installed with respective versions of a custom android application called MedSensor, software developed by the research team. At the participant's convenience, watch data was relayed to the phone via Bluetooth connectivity.

Closure. After collecting and transferring 10 days of collected data, the participants returned the smartwatch and phone to the study project coordinator and received a \$25 gift card as an incentive. The smart devices were sanitized according to Centers for Disease Control (CDC) guidelines prior to use by other participants.

b. Eating Data Acquisition

The eating activity was performed in a laboratory-controlled environment. For the purposes of the study, the activity was a pizza eating exercise for all. Where a participant did not want to eat pizza, they simply simulated using a hard paper without executing any bites.

Participants were scheduled to visit the lab in at different slots convenient to their schedules. There were before and after noon slots. Each participant emailed in to confirm their preferred slot. Each participant turned up at the lab. They went through a consent form, and after they confirmed their agreement to the exercise, they were each ran through a mock exercise to demonstrate of how to go about the data process. That was followed by the actual exercise of eating activity. The participant was handed an android watch, which he/she wore on the right wrist with the watch face facing outwards. The eating event comprised picking a slice of pizza, raising it to mouth level, taking a bite, and lowering it off mouth. This event was repeated multiple times until participant was done eating, or participant decided he/she had had enough. Upon completion of the exercise, the participant handed back the watch to the lab data team. The data was directly extracted from the watches via android studio and backed up in a local structured file repository. The data comprised watch ID, timestamp for each triaxial data logs, normal date/time log, tri-axial values, and the activity label.

c. Smoking Data Acquisition

The smoking data was based on data as collected in one of our previous studies [156]

d. Jogging Data Acquisition

The jogging dataset came from open public dataset, Wireless Sensor Data Mining (WISDM) Lab [193] that was recorded using smartphone strapped to the waist location of the participant.

6.3.6 Data Preparation and Annotation

This study integrated accelerometer sensor data from four different human activities recorded by smartwatches (Polar M600, Asus Zenwatch, Motorola, TicWatch E and E3

series). Table 6.1 provides a summary of the four different datasets, datapoints, and the number of points for each category.

Table 6.1: Summary of all the datasets used in the study

Activity	Datapoints	Gestures	Participants
Medication	824,000	2,800	28
Eating	272,822	5,434	6
Smoking	62,823	1,279	12
Jogging	287,461	5,883	27

Data Preprocessing and Standardization. Prior to the integration of data from multiple studies, several normalization and standardization steps needed to be performed. Attention needed to be made to the consistent standardization of the accelerometer data and the sampling rate of the data. Since all the devices used in data collection across all studies consisted of Android devices (versus Apple devices), the sensor data followed a common frame of x, y, and z axes. As the next step, all datasets were processed to adhere to a sampling rate of 25 Hz by excluding data points (in the case of oversampling) or resampling based on interpolation of the data (in the case of under sampling). To normalize for the different number of gestures per activity, the individual gestures were represented multiple times in our dataset to provide a balanced representation of activities.

Data Annotation: Proper use of data in supervised Machine Learning approaches requires annotation of the data. This is the process in which a supervisor (an expert) will

identify and define the gestures of interest. This is generally a cumbersome process and requires prior familiarity of the “supervisor” with the gesture of interest. To develop an understanding of what signal constitutes a MT activity, the scripted data were collected and used to understand the individual components of the gesture. The exact details of the scripted gesture can be found in our previous report [18]. Using this information, team members then proceeded to identifying and annotating the individual gestures. The process of gesture identification and annotation was heavily accelerated by the self-reports that included time stamps indicating the beginning and ending of each MT activity.

The raw data files logged at the cloud repository comprised a time stamp that included Hour, Minute, Second, and Millisecond; a Date the included day, month, and year; the x, y, and z components of the accelerometer data. A second file contained the time stamps corresponding to the start and end of each activity reported by the participant. In theory, the self-report of activities should be sufficient to identify the gesture of interest (e.g., MT). However, in practice, participants may report the activity erroneously, and the time stamps may roughly indicate the start and end of a given activity. Therefore, visual confirmation of the gesture of interest is executed to ensure high quality data. A separate utility program was developed to facilitate this process and to create the final usable data using a sliding window, which is well suited to real-time applications because it does not require any pre-processing. After this final step, the data files are presented in a usable format for the training and testing of our AI model.

6.3.7 Definition of Context

Context is an important aspect of gesture recognition in this study. The richness of activity characterization lies in the attention that it gives to multiple dimensions of analyzing human

engagement with the world. Saguna *et al.* [216]–[219] proposed a model for recognizing multiple concurrent and interleaved complex activities using a Context Driven Activity Theory (CDAT), which uses probabilistic data analysis for recognizing sequential, concurrent, and interleaved activities. As described earlier in the study, this research considers HAR as a language based on the alphabet of atomic activities. These atomic units form the building blocks of complex activities. For example, smoking activity can be decomposed into hand-to-lip, hand-at-lip for about three seconds, hand-off-lip [156]. These simple activities could be decomposed even further. Some of the atomic activities such as hand-to-mouth occur in many complex activities such as scratching face, taking medicine, eating, among others. To address this ambiguity and successfully recognize each complex activity correctly, there is need to determine the context of each activity. The context ultimately determines the signature or uniqueness of each complex activity. This is what is achieved through supervised training with a carefully calibrated sliding window and other hyperparameters.

Context can be either physical or cognitive. The former can be defined as the environmental information or the sensor data e.g., GPS location, time, temperature, etc. The latter refers to the mental states, preferences, tasks, and social affinities of the users [220]. Additionally, situation can be considered as an important factor in HAR. Human activity is situation-driven. This is also averred by Saguna *et al.*, “Situations are set of circumstances in which a human or an object may find itself” Physical context can be useful where geographic location and/or time add critical value to interpreting human activity or delivering a service to the subject. Alternatively, context can also be determined by in situ aspects of an activity such as sequence, frequency, time, delays, etc. In this study, we

implement and investigate this alternative in situ context. We consider context to be a sentence of atomic activities alphabet in specific sequence, or simply a sequence of atomic alphabets. This sequence vector also includes the frequency of an activity, temporal data, sensor orientation as well as the length of time the hand stays at a location i.e., how long sensor orientation remains unchanged. Figure 6.5 is the illustration of context as applied in this study:



Figure 6.5: Gesture context components

Figure 6.6 is a representation of a complex behavior signature as the fusion of a set of atomic activities, activity sequence data, and context attributes i.e.:

Activity Signature = Atomic Activities Set + Sequence data + Context Attributes

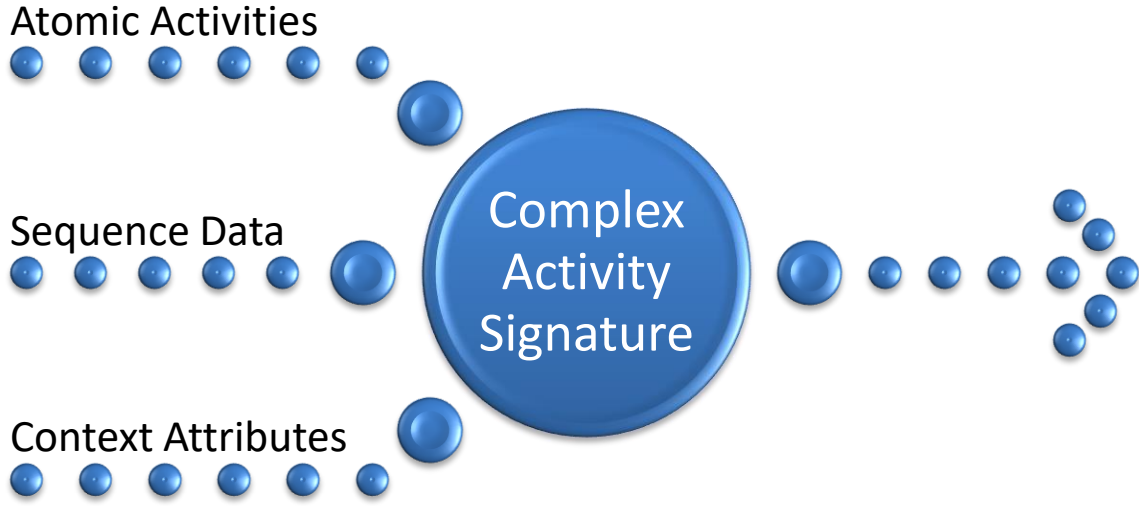


Figure 6.6: Complex activity signature

6.3.8 Tri-axial Sensor Data Definition

To understand the nature and structure of sensor data processed for neural network input, consider an accelerometer sensor that is attached to a human body and takes samples (at time index t) of the form:

$$r_t = r_t^* + w_t, \quad t = 1, 2, \dots \quad (1)$$

Where: $r_t = [r_t^x \ r_t^y \ r_t^z]^T$ is a 3D accelerometer data point generated at time t and comprises r_t^x , r_t^y and r_t^z which represents the x, y, and z acceleration components, respectively. The correct acceleration value in each axis channel is a floating-point value bounded to some known constant $B > 0$ such that $r_t^x \leq B$, $r_t^y \leq B$, and $r_t^z \leq B$. For example, for an accelerometer with $B = 2g$ units, it means that it can record proper acceleration up to twice the gravitational acceleration (recall that $1g = 9.8 \text{ meter/second}^2$) [221]. An accelerometer that is placed on a flat surface records a vertical acceleration value of $\pm 1g$ upward.

$r_t^* \in \mathbb{R}^3$ refers to a tri-axial noiseless vector of acceleration readings. $w_t \in \mathbb{R}^3$ is a noise vector of independent, zero-mean Gaussian random variables with variance σ_w^2 such that $w_t \sim N(0, \sigma_w^2 \mathbb{I}_3)$. Examples of noise during signal acquisition include the effect of temperature drifts and electromagnetic fields on electrical accelerometers [222].

Three channel frames s_t^x , s_t^y and $s_t^z \in \mathbb{R}^N$ are then formed to contain the x, y, and z acceleration components, respectively. These channel frames are structured as a sliding window as follows:

$$s_t^x = [r_t^x \dots r_{t+N-1}^x]^T \quad (2)$$

$$s_t^y = [r_t^y \dots r_{t+N-1}^y]^T \quad (3)$$

$$s_t^z = [r_t^z \dots r_{t+N-1}^z]^T \quad (4)$$

The sequence size N i.e., window size, should be carefully calibrated to ensure the most accurate activity recognition. Where we assume that the system supports M different activities: let $A = \{a_1, a_2, \dots, a_M\}$ be a finite activity space. Based on the windowed excerpts s_t^x , s_t^y and s_t^z , the activity recognition method infers the occurrence of an activity $y_t \in A$ [221]. Where the window is smaller than the whole activity, it is considered to be recognizing atomic sequences of the larger activity in specific order based on the time they were recorded, i.e., First In First Out (FIFO) order, such that by the end of the window slide across the atomic sequences of the complex activity, either the model will have positively identified one of the M activities or none.

6.3.9 Artificial Neural Network Architecture and Design - LSTMs

The human activities of interest in this study – eating-pizza, medication-taking, smoking, and jogging – are each composed of sequences of atomic-activities whose temporal aspect adds important component in the overall activity recognition. For example, the medication

activity is made up of a series of atomic actions namely (a) Open bottle, (b) Dispense medicine, (c) Toss pill to mouth, and (d) Drink water. The sub-activities, and their temporal sequence, are important. Therefore, the past sub-features are useful and relevant as far as recognition of the full activity.

Human Activity Recognition requires context that remembers the past array of atomic activities. The complex activity is the array sequence. By their own architecture, LSTMs lend themselves suitable for the classification hence recognition of human activities. This chain-like nature of RNNs intimately relate them to sequences and lists. They are the natural architecture of neural network to use for such data. Unlike other forms of RNNs, LSTMs are explicitly designed to avoid the long-term dependency problem. Remembering information for long periods of time is practically their default behavior. They can connect previous information to the present task, such as using previous video frames might inform the understanding of the present frame.

LSTMs (Long Short-Term Memory) find most relevant application in recognition of human activity because they contain internal memory of the past events in the analysis of time-series data. LSTM-NN is an artificial Recurrent Neural Network (RNN) architecture with feedback connections that facilitates awareness of past activities at present time of the activity[223]. An LSTM unit comprises a cell, an input gate, an output gate and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. The cell can process data sequentially and keep its hidden state through time. Figure 6.7 is an illustration of a typical LSTM cell where x_t is the input vector to the LSTM unit, h_t is the hidden state vector (or LSTM unit output vector), c_t is the cell state vector, and c_{t-1} is the cell input activation

vector. Our model contained two fully connected and two LSTM layers (stacked on each other) with 64 units each. The learning rate was set at 0.0025.

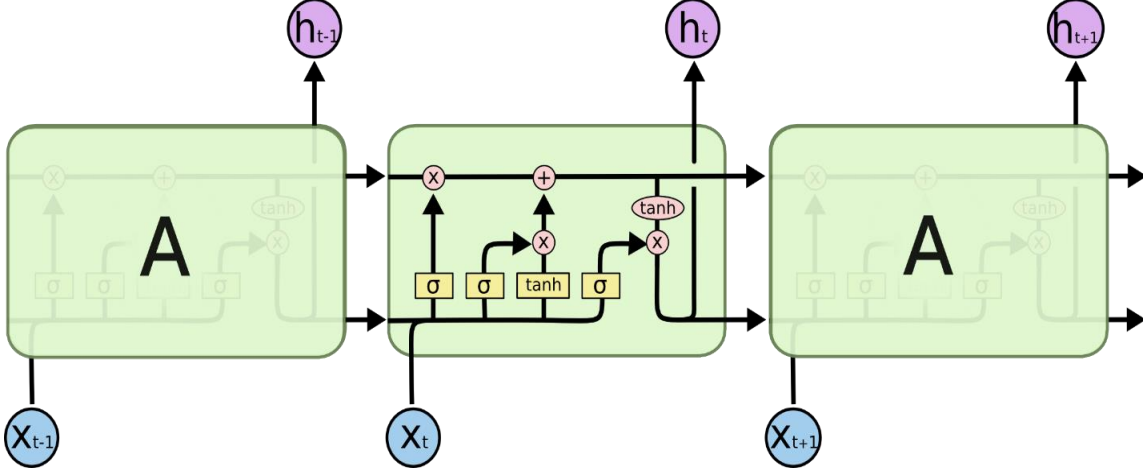
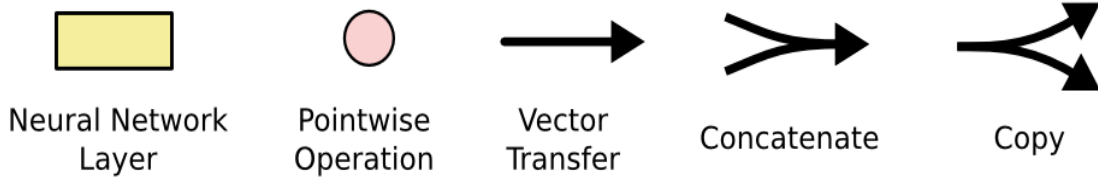


Figure 6.7: The repeating module in an LSTM contains four interacting layers [223]



Below are compact forms of the equations for the forward pass of an LSTM cell with a forget gate [160], [224].

- $f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$
- $i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$
- $o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$
- $\bar{c}_t = \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$
- $c_t = f_t \odot c_{t-1} + i_t \odot \bar{c}_t$
- $h_t = o_t \odot \sigma_h(c_t)$

where the initial values are $c_0 = 0$ and $h_0 = 0$ and the operator \odot denotes the Hadamard product (element-wise product). The subscript t indexes the time step.

Variables

- $x_t \in \mathbb{R}^d$: input vector to the LSTM unit
- $f_t \in (0,1)^h$: forget gate's activation vector
- $i_t \in (0,1)^h$: input/update gate's activation vector
- $o_t \in (0,1)^h$: output gate's activation vector
- $h_t \in (-1,1)^h$: hidden state vector also known as output vector of the LSTM unit
- $\bar{c}_t \in (-1,1)^h$: cell input activation vector
- $c_t \in \mathbb{R}^h$: cell state vector
- $W \in \mathbb{R}^{h \times d}$, $U \in \mathbb{R}^{h \times h}$ and $b \in \mathbb{R}^h$: weight matrices and bias vector parameters to be learnt during training. The superscripts d and h refer to the number of input features and number of hidden units, respectively.

Activation Functions

- σ_g : sigmoid function
- σ_c : hyperbolic tangent function
- σ_h : hyperbolic tangent function

6.3.10 Network Training and Testing Procedure

The LSTM network was trained for 50 epochs while keeping track of accuracy and error. The batch size was maintained at 1024. The train/test datasets were partitioned in the 80:20 ratio, respectively. We applied L2 regularization (Ridge Regression) to the model. The L2 penalty/force removes a small percentage of weights at each iteration, ensuring that weights never become zero. The penalty consequently reduces the chance of model overfitting.

The LSTM model expects the training data to be of fixed length. This fixed length of data that is directly presented to LSTM is referred to as the “window size.” In this study, a window size of 150 points was empirically determined to provide an acceptable performance. At a sampling frequency of 25Hz, a window size of 150 represents 6 seconds

of recording. While the window size represents the portion of the raw data that should be in direct view of the ANN at the time of classification, any relevant past contextual information is saved in the internal cell of the LSTM architecture. The temporal exposure of the LSTM to the data occurred in sliding window size of 10 points as an optimal compromise between performance, responsiveness, and without increasing classification error. The sliding window-with-overlaps process significantly transforms and reduces the training dataset. Further, the transformation assigns the most common activity (i.e., mode) as a label for the sequence; some windows may comprise two or more activities, but the mode is considered the dominant or overriding activity. We transform the shape of our tensor into sequences of 150 rows, each containing x, y and z values. We also apply a one-hot encoding to the labels to transform them into numeric values that can be processed by the model.

6.3.11 Evaluation of the Trained Network

During the training phase of a Neural Network, a single metric of performance needs to be defined to assess network performance. The network performance metric is used by the operator to direct the network to improve overall performance. In this study, we evaluated the performance of classifiers using the Accuracy metric as defined in Equation 1. In Equation 1, TP (true positives) represent the correctly classified positive examples, TN (true negatives) represents the correctly classified negative examples, FP (false positives) represent negatives misclassified as positives, and FN (false negatives) represent positives misclassified as false.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{eq(6)}$$

Accuracy measure does not consider the bias arising from unbalanced datasets. To remove the effect of unbalanced data (unequal representation of different activities), data within each activity were repeated to arrive at approximately equal number of representations. Thus, the metric has a bias favor for the majority classes. For this reason, the study considered the following evaluation criteria: Precision, recall, F-measure, and specificity. Precision indicates what fraction of positive predictions were positive. Recall (True Positive Rate) indicates what fraction of all positive samples were correctly predicted as positive by the classifier. Recall (True Negative Rate) indicates what fraction of all negative samples are correctly predicted as negative by the classifier.

Below are the formulae to compute the metrics:

$$precision = \frac{TP}{TP+FP} \quad \text{eq(7)}$$

$$recall = \frac{TP}{TP+FN} \quad \text{eq(8)}$$

F-measure is the combination of precision and recall. It is calculated as follows:

$$Fmeasure = \frac{(1+\beta^2).recall.precision}{\beta^2 recall + precision} \quad \text{eq(9)}$$

Where β is a weighting factor and a positive real number. It is used to control the importance of recall/precision.

$$specificity = \frac{TN}{TN+FP} \quad \text{eq(10)}$$

6.4 Results

6.4.1 Visualization of MTP Gesture

As the first step in performing activity recognition with wearable devices, a more detailed understanding of the gesture of interest needs to be developed. Figure 6.8 represents an

example of the entire MTP activity recorded from a right-hand dominant participant. After careful and repeated examination of the gesture, sequential segments of the gesture can be identified as shown in the figure. It is of notable interest to highlight the portion of the gesture corresponding to the water-drinking phase of the activity illustrated as the latter portion of the phase C in Figure 6.8. The gradual increase of the y-axis of the accelerometer denotes the beginning of the drinking phase and can therefore be used to quantify the duration of the drinking event.

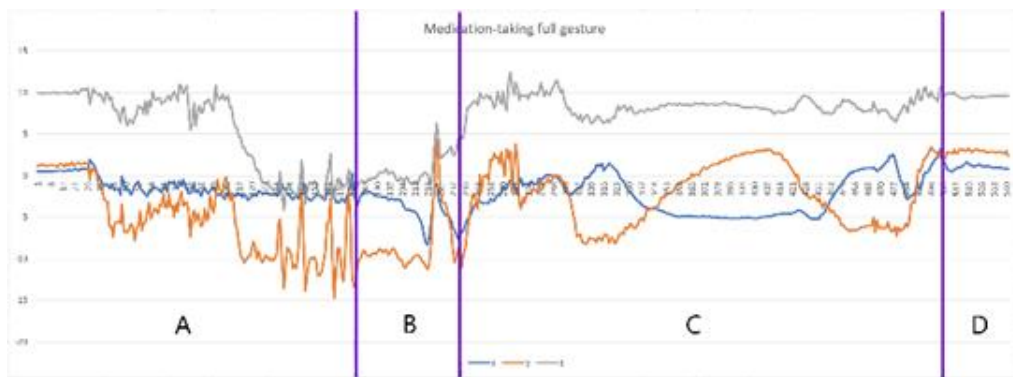


Figure 6.8: The segments correspond to (A) open-bottle dispense-medicine, (B) Hand-to-mouth pill-to-mouth hand-off-mouth, (C) pick-up-water drink-water lower-cup-to-table close-bottle, and (D) post-medication

Our study of MTP consisted of scripted and natural methods of administering the medication. Although our scripted version was based on the behavior observed in most of our preliminary studies, it is possible that participant's natural behavior to deviate from the scripted version. Figure 6.9 provides visual illustration of the differences between scripted and natural MTP activity from the same participant (panels a and b), and a natural MTP activity from a second participant.

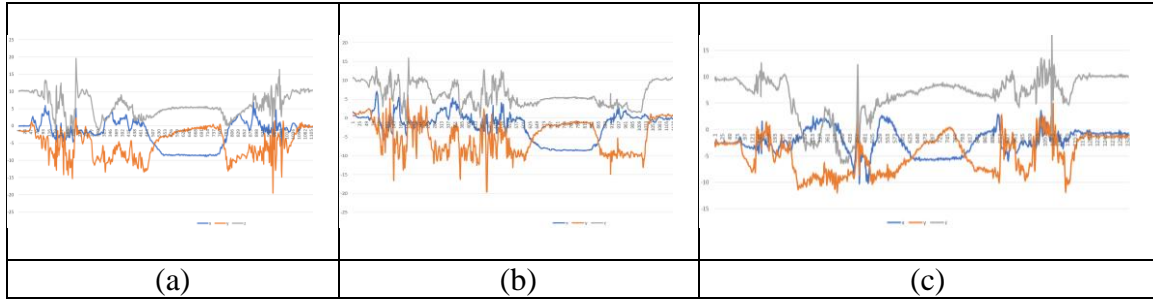


Figure 6.9: Illustration of MTP differences between a) scripted gesture from user1, b) natural gesture from user1, and c) natural gesture from user2.

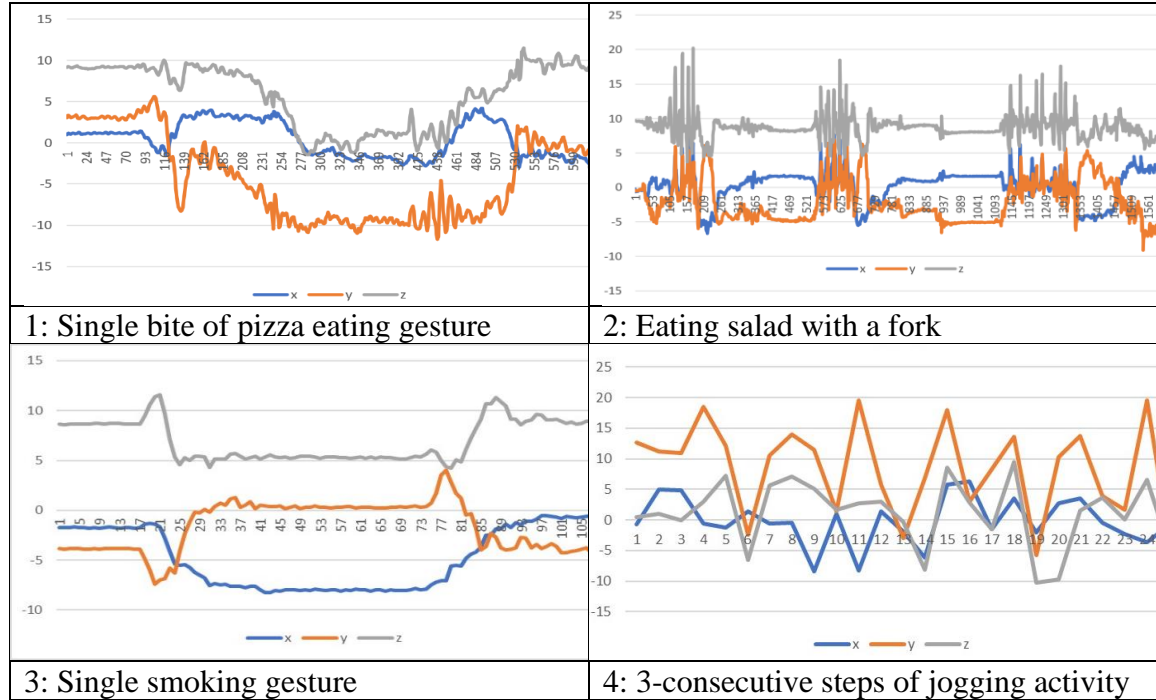
It is important to highlight the challenging task of identifying an MTP activity given the diversity in the activity among different participants. Numerous variations can be expected in the way different participants complete any given activity. For instance, the simple method of drinking water between two participants can vary significantly as illustrated in Figure 6.9 panels b and c. While the first participant (Figure 6.9 panel b) performs the task of drinking water with a sudden removal of glass from mouth, the second participant (Figure 6.9 panel c) performs a more gradual removal of the glass from mouth. These differences in the individual sub-gestures are the root of the challenges associated with detection of any gesture.

6.4.2 Visualization of other Gestures

Table 6.2 represents instances of four different activities as described by the caption under each figure. The gestures 1 and 3 comprise single instance of the respective complex activity. Gestures 2 and 4 comprise three instances of the respective complex activity. Gesture number two, the salad eating activity, shows two distinctive alternating regions. The sections of increased rapid activity and higher magnitude along y-axis indicate moments when the subject was repeatedly forking the salad. The zones immediately after

the forking shows hand-to-mouth to deliver fork ‘payload’ to mouth, followed by hand-off-mouth, followed by a rest period before the cycle repeats.

Table 6.2: Visualization of pizza bite, eating salad with a fork, smoking and jogging activities



6.4.3 Activity-wise distribution of the signal data

As seen in Figure 6.10, Figure 6.11, and Figure 6.12 below, it is observed that there is very high overlap in the data among activities such as Medication, Eating and Smoking on all the axes. Jogging appears to have distinctive values along y-axis and z-axis. The large overlaps indicate that there are similar/common atomic activities in the complex activities that show interclass similarities. This similarity can also be seen in the Confusion matrix where there are significantly higher number of medication and eating gestures similar than there are between medication and jogging.

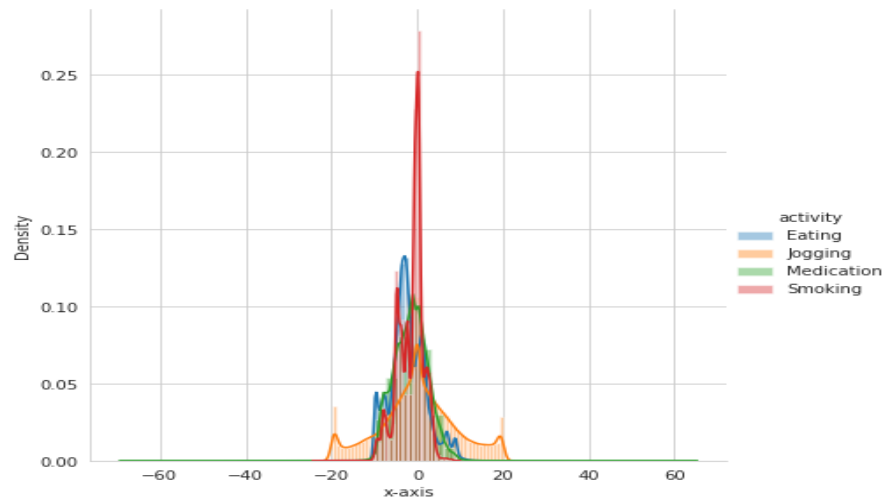


Figure 6.10: Signal distribution along x-axis

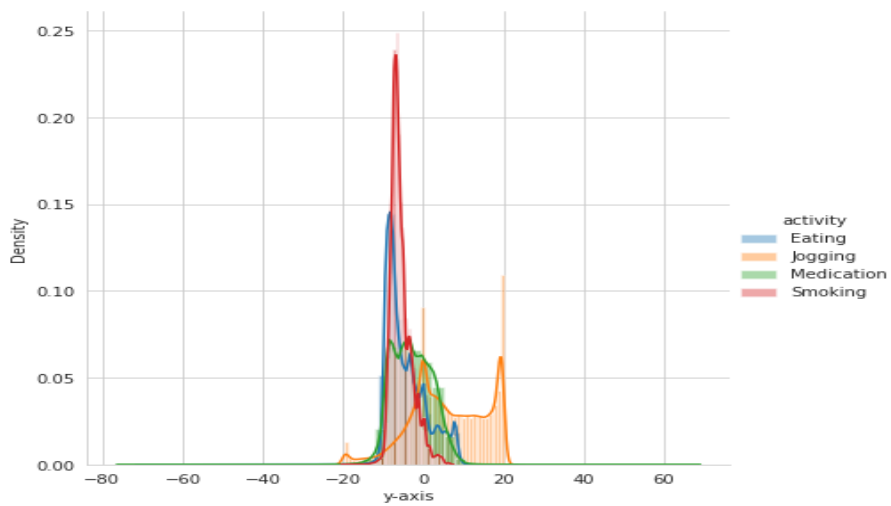


Figure 6.11: Signal distribution along y-axis

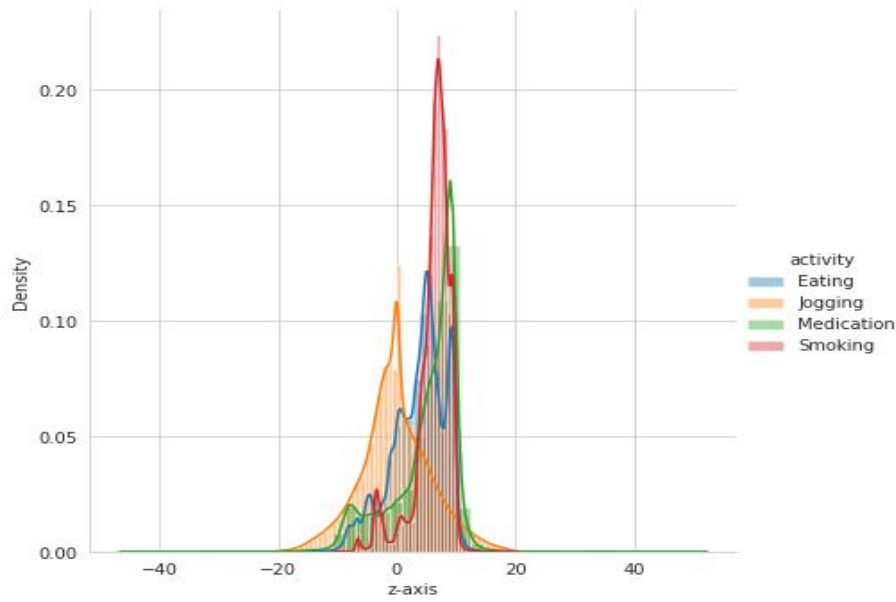


Figure 6.12: Signal distribution along z-axis

6.4.4 Validation and Annotation of Medication Taking Events

The study of human behavior with wearable devices has several advantages over the traditional self-report methods. Specific to medication-taking, self-reported adherence is known to be over-estimated[194]. In comparison to self-report, wearable devices can provide additional useful information such as the time of the day the medication was administered and the number of times the medication was taken in a day without incurring additional time, effort, or cost to the user. Table 6.2 displays additional data extracted from smartwatch MTE with MedSensor. Considering the average time needed to complete MTE, outliers can be examined for accuracy as shown in Table 6.3.

In this study, we considered outliers (both natural and scripted) as gestures of duration 8 seconds and below for the lower category, or 100 seconds and above for the upper category. In determining the outliers, we considered the mean and standard deviations for natural gestures at 18.47 and 14.34, and scripted gestures at 20.11 and 20.65, respectively. The outliers were observably fewest in scripted than natural gestures. For the

lower category, a random sample of 20 out of 103 gestures were examined, and all were invalid gestures, indicating that the users probably annotated start/stop of gestures in quick succession. On the other hand, for outliers ≥ 100 seconds in duration, the majority contained one or more MTE gestures in most cases. To better understand the cause of these temporal discrepancies and therefore, validate or invalidate the reported gestures, each recording session was examined by a team member for validity. The results are depicted in Table 6.4.

Table 6.3: Statistical description of natural and scripted/protocol gestures.

	Min	Mean	Median	Std. Dev	Max
Natural	5	18.47	17	14.34	331
Scripted	4	20.11	18	20.65	686

Table 6.3 summarizes the number of extreme outliers that were identified in our pool of acquired data. In this study, we considered outliers (both natural and scripted) as gestures of duration 10 seconds and below for the lower category, or 100 seconds and above for the upper category. For the lower category, a random sample of 20 out of 103 gestures, all turned out to be invalid gestures, indicating that the user probably annotated start/stop of gestures in quick succession, probably to cancel unintended activity recording, or the user simply played start/stop without recording activity. On the other hand, for the category, i.e., outliers ≥ 100 seconds in duration, it turned out that majority of them contained one or more medication gestures in most cases, and none in a few cases. Table 5.3Table 6.3 provides a summary of the outliers' category and counts. The Table 6.4 shows the analysis and observations of the upper-category outliers in terms of the gesture duration in seconds as well as the number of gestures identified therein. All the upper-category

outliers with at least one valid gesture at the beginning, the first valid gesture activity, or at least 24 seconds of activity, was extracted and included in the pre-processed input data to the neural network training; the rest of the data was dropped.

Table 6.4: Outliers count for natural and scripted medication gestures considered from gestures longer than 100 seconds or shorter than 8 seconds

Description/Category	Count	
	Scripted Gestures	Natural Gestures
Duration equal to or greater than 100 seconds	6	2
Duration equal to, or less than 10 seconds	63	40

To better understand the cause of these temporal discrepancies and therefore, validate or invalidate the reported gestures, each recording session was examined by a team member for validity. Table 6.4 provides some examples of additional insights that the recordings from MedSensor provided and the corrections that were made by a ML supervisor. An additional benefit of this manual examination of the data was a more accurate representation of the start and end of the MTP activity while ensuring high quality of the training data.

Table 6.5: Analysis of upper-category outliers.

Subject	Duration	Observations	Correction
U1	371	The participant reported 7 consecutively taken medications as one medication event.	Individual MTP events were separated by a ML supervisor.

U1, U2, U3, U4, U5, U6,	162	One MTP event was observed with some unrelated activities at the beginning or end of the recording.	The unrelated portions of the recording were trimmed.
U3	279	Comprises random activities that do not match medication gesture pattern	

6.4.5 Neural Network Training and Testing

Using a fixed window size of 150 produced the performance shown in averages, as shown by the tables below. The test accuracies for eating, jogging, medication, and smoking were 94.3%, 100%, 93.6% and 98.6%, respectively. The average performance attained was 96.63%.

Table 6.6 shows the summary of performance metrics for the neural network trained on data that excluded outliers. A similar experiment for data that included outliers produced an accuracy performance of 97.4%. As observed in the averages, there was a slight performance difference. This could possibly be attributed to the fact that the outliers made up a minute percentage of the training data, such that their exclusion did not amount to significant performance difference. It can also be seen from **Table 6.4** that the outliers were a further minor data of a few subjects, meaning that the network would still generalize to majority observations.

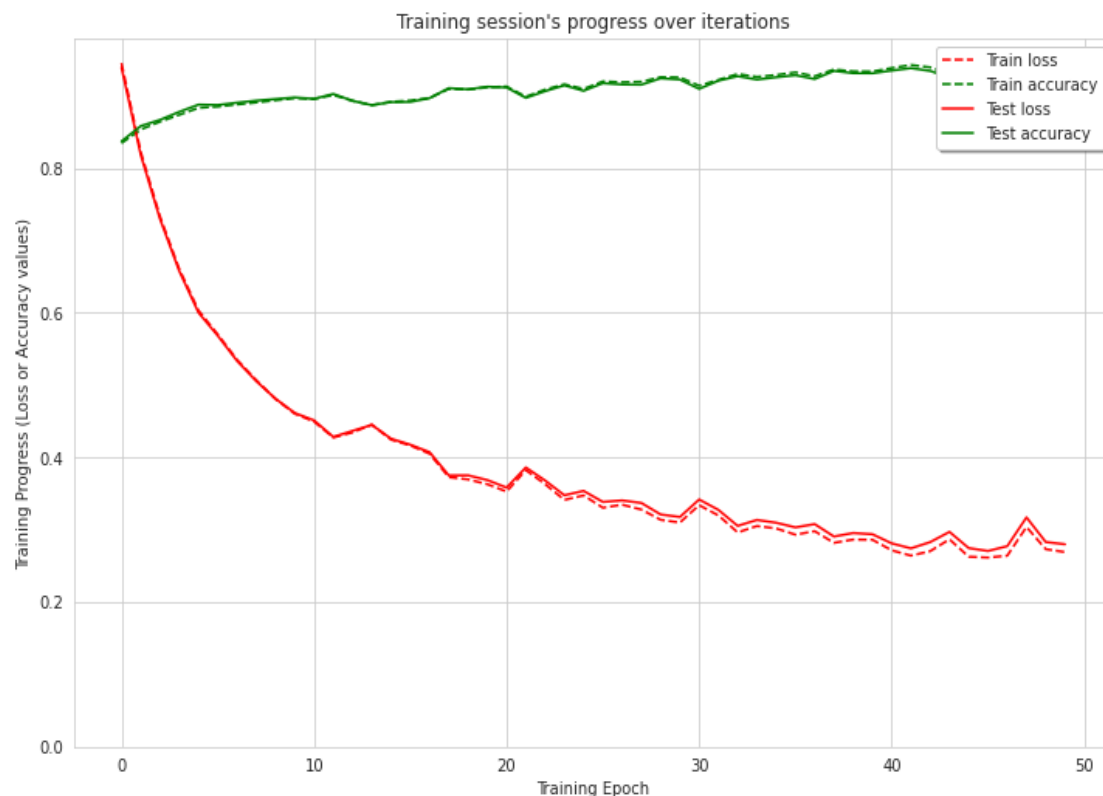


Figure 6.13: Training plot for the window size of 150 units.

Table 6.6: Summary metrics for the best performing configuration of window size 150 - Excludes outliers

Activity	Precision	Recall	F-Measure	Specificity	Accuracy
Eating	0.802	0.920	0.857	0.948	0.943
Jogging	1.000	1.000	1.000	1.000	1.000
Medication	0.961	0.924	0.943	0.951	0.936
Smoking	0.888	0.769	0.824	0.996	0.986
Average	91.29%	90.33%	90.59%	97.37%	96.63%

From the experiments with different sliding window sizes, we established that smaller windows, (other hyper-parameters remaining fixed), produced lower performances (95.5%). The reverse was true for wider windows, with the best performance of 97.4%

occurring at 150 units window size. To explore the full nature of misclassifications, the confusion matrix shown in Figure 6.15 was examined. The accuracy, precision, recall, F-measure, and specificity, as described by the equations 1, 2, 3, 4 and 5, respectively, are presented the Table 6.5.

While the overall best performance of ~97% is good, this study has not explored all possible nuanced configurations. Besides the sliding window size, it is possible to manipulate the hyperparameters such as the learning rate, adjust the number of LSTM units, windows step size, batch size, etc. to arrive at an even better performance. This is certainly an area for future exploration and further experiments.

Training Results without Outliers

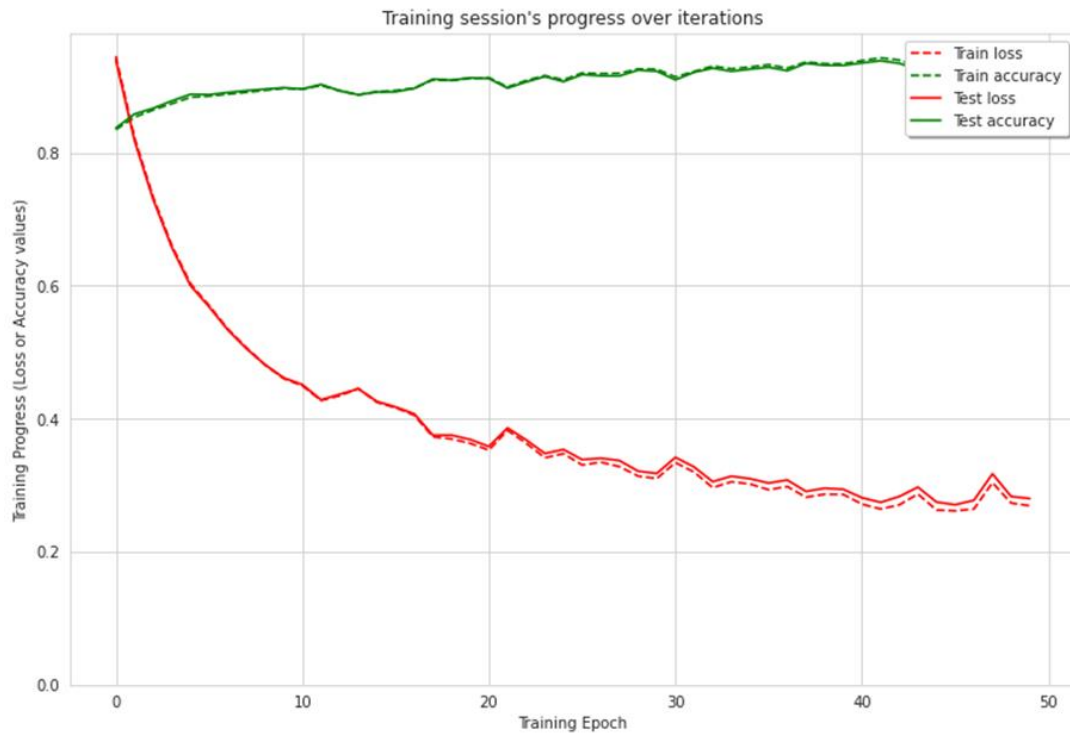


Figure 6.14: Training plot for the window size of 150 units

		Predicted label			
True label	Eating	4958	0	411	18
	Jogging	0	5867	0	0
	Medication	1147	0	15240	98
	Smoking	78	0	200	923
		Eating	Jogging	Medication	Smoking

Figure 6.15: Confusion matrix for the window size of 150 units

6.4.6 Summary of Data

Participant Demographics

The participants comprised 57.0% males and 43.0% females, all of whom were mostly college students (71.4%), single (86.0%), and working at least part-time (61.0%). The sample represented racial diversity with 4% African American, 43% Asian, 43% White, and 10% Mixed. The right hand was dominant for 82%, while only 1 participant was ambidextrous.

6.5 Discussion

6.5.1 Principal Results

Studying human activities with smart and wearable devices has numerous advantages over the traditional approaches. Wearable devices provide the unobtrusive, continuous observation of human behavior in their natural settings with little burden to the user. The collected sensor data from these devices can be used to validate data reported by the user, therefore improving the accuracy and completeness of self-reports. In this work, participants used the smartwatch to report the beginning and end of their MTE. Errors in participants' self-reported MTE were identified. Some self-reported MTE had implausibly short or long durations. Participants also used smartwatch to report the beginning and end of pizza eating activity. In the same way, the participants recorded their smoking activities. We also examined a public dataset for jogging exercise. While these activities in many ways involve the use of two hands, we relied on one smartwatch per participant, worn on one wrist. This approach certainly ignores data that could prove useful in improving gesture richness for activity recognition. We, however, assume for now that the minimal approach which depict how most people use the wearable devices like smart watch is still sufficient in interpreting human activity, especially when the subject wears the device upon the most dominant or active hand.

By validating the digitally recorded temporal gestures, we demonstrated the ability to correct erroneous self-reports, therefore improving the quality of the reports. Furthermore, the temporal signature that has been reported by the array of sensors available on wearable devices, can provide a plethora of additional information such as the temporal variation of an activity within a given user, or across a population of users. For example, in our study we demonstrated that nMTE were completed in an average of 18 seconds, but

there were distinct interclass differences across participants; equally, there were also intra-class differences where gesture uniqueness also occurs even among the gestures of the same individual. Such comparison of behaviors provides several dimensions along which the study of human behavior can be expanded.

In addition to the expanded information that can be obtained from these devices, the ability to automatically detect and identify an nMTE with high accuracy will be beneficial. Automatic detection of an MTE event can be used as the foundation for both measuring and improving adherence. In the latter case, non-detection of an nMTE offers the opportunity to alert patients or support persons of the missed medications to improve adherence and the health outcomes associated with improved medication adherence. Improved medication adherence has the potential to significantly reduce morbidity [195]–[197], mortality [167], [197], and healthcare costs [195], [198]–[200]. Hence, detecting nMTE with smartwatches has exponential utility. Equally significant, automatic detection of eating, and smoking gestures opens an array of opportunity to create innovative solutions such as smoking cessation tools, healthy eating assistants, among others. These tools can empower everyday persons with simple tools available freely as mobile apps without the need of costly expert consultancies.

6.5.2 Limitations

The limitation of our approach given the current state of the technology is the availability of data from a single hand. Therefore, activities such as smoking that can be completed by a single hand, may not be detected by a watch that is worn on an opposite hand. Fortunately, because activities such as medication-taking are difficult to complete with a single hand, a residual portion of the activity will always be present from the perspective of a single

watch. This problem can easily be overcome with the presence of a sensing device on each hand (or possibly each limb). Although not common presently, the arrival of smart wristbands, rings, and other forms of wearable is likely to provide a more complete picture of a person's daily activities [141], [182], [184], [201]–[203]. This limitation can also be mitigated in future studies through the augmentation of wearable sensor data with vision-based data, particularly in environments where vision devices are mounted. For example, it would be easier to verify eating, smoking or medication activities if there is a video or picture taken of the same event. This can corroborate and confirm the detected activity. With increased environmental sensors embedded in many public places, as well as CCTV cameras in the public, and where the laws allow, this approach can be significantly complimentary to wearable sensor activity recognition.

A second limitation is the method by which people may wear their watch. A watch can be worn in four distinct ways of on left or right hand, and inside or outside of the wrist. In this study, we asked participants to wear their watch on their right hand, and on the outside of their wrists. However, sensor data recorded by the same watch in any of the other three configurations will produce related but undistinguishable signals by the Neural Network. Consistently wearing a watch on the outside right hand is critical at the current stage of our scientific development. However, using the existing human body symmetry, and the relationship between inside and outside of the wrists, mechanisms of unifying sensor signals collected from any mode of use can be developed as demonstrated previously [122]. This will produce high-capacity models with broader experience to recognize medication gestures regardless of watch-face orientation.

6.5.3 Proposed Improvements to Protocol

Generally, our current protocol worked very well. However, our future investigations will benefit from two additional steps to our existing protocol. The first step will collect calibration data during the initial orientation session. Currently, our orientation consists of familiarizing the participant with the watch, the app, and the use of the app. In the future, we will use this orientation session to collect data from a set of simple activities (e.g., touch toes, touch hip, and touch head) with the watch worn on each hand to obtain useful participant-specific data at baseline. By collecting data from left and right hands, we can establish a more precise relationship between the right and left hands for given participants. Although perfect human symmetry may indicate a 180-degree rotation between the two, natural human posture may create a departure from an ideal 180-degree symmetry. This information can be used to allow the user to wear the watch in any preferred method and provide the necessary information for the correction that is needed for the existing right-handed Neural Network.

6.6 Conclusions

Medication activity, Eating event, and Smoking activity are all complex human activities that require finer decomposition into atomic sub-activities in order to recognize the underlying building blocks of each behavior. Medication adherence challenge is associated with chronic condition self-management. Smoking remains a leading cause of preventable death, and obesity remains a major health challenge. Automated detection of medication-taking activity is of critical importance in relation to improved treatment effectiveness. Equally, automated detection of smoking or Eating events hold a huge potential for exploitation to improve human health as well as prevent death. In this study, we demonstrated the use of LSTM-NN based on in situ context data to detect, and

recognize atomic activities, and the complex activities of four complex activities. We have demonstrated successful identification of individual medication gestures with an accuracy of approximately 93.6% when tested against activities that significantly resemble medication, such as smoking and eating, which share the common atomic-gestures of hand-to-mouth, hand-on-mouth, and hand-off-mouth. By the same unified model, we achieved accuracies of 94.3% for Eating, 100% for Jogging, and 98.6% for Smoking. Our investigation showed almost no confusion between jogging and the other three activities. This could be explained by the fact that the atomic activities and context attribute composition of the two categories are largely different. By this observation, we speculate the accuracy of the system to increase notably if other natural daily activities are included in the training and testing sets due to their dissimilarity.

Although in this work we have achieved a reasonably high detection of the four activities, several additional investigations can be initiated to increase the performance and usability of the system. First, as an ultimate objective, we aim to develop one application that can decipher numerous human activities to establish correlative or causative relationship between activities. For instance, medication activity may occur at 8 PM before sleep activity, or at 7 AM before breakfast eating activity, or eating activity at 1 PM may be followed by cigarette smoking soon after. The ability to monitor the temporal relationship between these events would be very useful to provide the much-needed context to further understand human behavior and therefore model useful health-related solutions or provide real-time intervention reminders. To accomplish this, we need to engage in a formal investigation of the optimal viewable window size to an ANN that will be sufficient to successfully decipher between all activities of interest. Additionally, there exists some

inherent parallel between human activities and the principles of written language. To fully leverage this parallel analogy, human activities need to be examined in the more fundamental fashion by decomposing complex activities further to their most basic building blocks, referred in this study as atomic gestures. Our previous work[18], [192] illustrates the mini-gesture decomposition of the eating activity in relation to other similar activities such as smoking.

6.6.1 Acknowledgements

Funding for this work was provided by NIGMS grant number P20 RR-01646100 (Dr. Homayoun Valafar), College of Computer Science and Engineering, and Center for Advancing Chronic Care Outcomes through Research and Innovation, College of Nursing, University of South Carolina.

6.6.2 Conflicts of Interest

The authors declare no conflict of interest

6.6.3 Abbreviations

ML: Machine Learning

AI: Artificial Intelligence

ANN: Artificial Neural Network

TPR: True Positive Rate

FPR: False Positive Rate

CHAPTER 7: CONCLUSION

Human activity recognition is a complex pattern recognition problem. In this research, we have examined human activity theory as well as the various techniques and approaches used to decipher the semantics of human activities grammar. We have studied a variety of activities, specifically smoking behavior, medication taking behavior, eating behavior, and jogging exercise activities. Medication adherence is a complex human behavior associated with chronic condition self-management. It remains a global public health challenge; nearly 50% of people fail to adhere to their medication regimes. Automated detection of medication-taking activity is of critical importance in relation to improved treatment effectiveness. The automatic detection of medication gestures will also help to eliminate the burden of self-reporting from the participants and therefore, provide a simpler way of tracking medication events. In this study, we demonstrated the use of LSTM-NN to detect and recognize atomic activities, and the complex activities.

We have demonstrated successful identification of individual medication gestures with an accuracy of approximately 93.6% when evaluated against activities that significantly resemble medication, such as smoking and eating, which share the common atomic gestures of hand-to-mouth, hand-on-mouth, and hand-off-mouth. Our investigation also compared medication activity against jogging activity, and our model confirmed little confusion between medication and jogging. This could be explained by the fact that the mini activities of both complex dynamic activities are different. By this observation, we

speculate the accuracy of the system to increase notably if other natural daily activities are included in our training and testing sets due to their dissimilarity to the medication gesture. We further evaluated our unified classifier model against eating, jogging, medication, and smoking, with performance outcomes of 94.3%, 100%, 93.6% and 98.6%, respectively, and an overall average performance of 96.63%. To reduce the error margins, which can be significant when we consider the activities such as medication, the context data can be enriched with external attributes like time and location. Where necessary, visual aspects can be captured to augment sensor data and make it possible to corroborate triaxial orientations.

Although in this work we have achieved a high detection of the medication-taking gesture, several additional investigations can be initiated to increase the performance and usability of the system. Future work on the unified model should be aimed to both decipher numerous human activities as well as establish correlative or causative relationship between activities. For instance, medication activity may occur at 8 PM before sleep activity, or at 7 AM before breakfast eating activity, or eating activity at 1 PM may be followed by cigarette smoking soon after. The ability to monitor the temporal relationship between these events would be particularly useful to provide the much-needed context to further understand human behavior and therefore model useful health-related solutions or provide real-time intervention reminders. To accomplish this, we need to engage in a formal investigation of the optimal viewable sliding window size to the Neural Network that will be sufficient to successfully decipher between all activities of interest. Additionally, there exists great potential in exploiting the inherent linguistic parallel between human activities and the principles of written language. To fully leverage this

parallel analogy, human activities need to be examined in the more fundamental fashion by decomposing complex activities further to their most basic building blocks akin to alphabets, otherwise referred in this study as mini or atomic gestures. In two of our studies, we have applied this parallel to model HAR as a language.

By the language concept, we have presented the reformulation of an entire smoking gesture (puff) as a combination of three time-dependent mini-gestures (hand-to-lip, hand-on-lip, and hand-off-lip). Based on the results shown in the reformulation of the smoking gesture as mini gestures, we addressed one major HAR challenge of by reducing the complexity of detection as evidenced by the improved detection. Although we have achieved a near perfect detection of the four activities, we anticipate unforeseen challenges during the live deployment of this technology due to other noise data such as background or intra-class gesture variations.

To incorporate the temporal dependency of human activities, including the mini-gestures, we implemented LSTM. We expect that with more data, and better annotation, these performance metrics can be improved significantly. We observe that our state-transition approach to detection of atomic gestures had demonstrated improvements over the previously reported approaches. We expect that the declaration of atomic gestures as the “vocabulary” of human activities is instrumental in the development of the activity “grammar” that can be exploited by the incorporation of RNNs and transfer learning.

Further, our research addresses two major challenges associated with HAR sensor data. Through edge-enabled annotation at the device, we significantly improve data annotation and preprocessing, making it possible to automate more processing downstream. Although this approach does not guarantee 100% gesture integrity, it makes data process more

efficient than conventional tedious manual data annotation. Most important, and tied to the data process, we implement a comprehensive online Sensor Database that streamlines the data process, enables research and collaboration, and meaningfully organizes datasets in a way that allows controlled, secure access. This is what we call the METASENS project. We have already deployed it for lab use, but ultimately, it will be a resource to the public to open for a wider HAR research community.

This research also spent significant time investigating efficient energy utilization by smartwatches. Energy use, and hence battery life, remains one key problem especially for small devices. Thankfully, recent industry development has improved small battery capacity making most watches last for more than 24 hours on full use. We emerged with crucial analysis that would still help users and application developers to get more life out of small devices when the devices are correctly calibrated. In most cases, there is no need to run all sensors since accelerometer data alone is sufficient in the characterization of human motions. Where there is need to fuse sensor data, other sensors can be initialized, otherwise they can be deactivated. We share some of these findings in our presentations.

Finally, HAR remains active research seeking more efficient solutions particularly in data acquisition and labeling, computation overheads, data augmentation, intra-class and inter class variability challenges. It is our expectation that our work adds value to the research community, and we hope to come up with more innovative solutions to the existing domain challenges.

REFERENCES

- [1] B. Fu, N. Damer, F. Kirchbuchner, and A. Kuijper, "Sensing technology for human activity recognition: A comprehensive survey," *IEEE Access*, vol. 8, pp. 83791–83820, 2020.
- [2] Y. Wang, S. Cang, and H. Yu, "A survey on wearable sensor modality centred human activity recognition in health care," *Expert Syst. Appl.*, vol. 137, pp. 167–190, Dec. 2019.
- [3] L. Minh Dang, K. Min, H. Wang, M. Jalil Piran, C. Hee Lee, and H. Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognit.*, vol. 108, p. 107561, Dec. 2020.
- [4] H. S. Alzubi, S. Gerrard-Longworth, W. Al-Nuaimy, Y. Goulermas, and S. Preece, "Human activity classification using a single accelerometer," in *2014 14th UK Workshop on Computational Intelligence, UKCI 2014 - Proceedings*, 2014.
- [5] R. Rajagopalan, I. Litvan, and T. P. Jung, "Fall prediction and prevention systems: Recent trends, challenges, and future research directions," *Sensors (Switzerland)*, vol. 17, no. 11. 2017.
- [6] L. Jarvis, S. Moninger, J. Pavon, C. Throckmorton, and K. Caves, "Accelerometer-Based Machine Learning Categorization of Body Position in Adult Populations," *Comput. Help. people with Spec. needs ... Int. Conf. ICCHP ... proceedings. Int. Conf. Comput. Help. People with Spec. Needs*, vol. 12377, p. 242, 2020.
- [7] M. H. Arshad, M. Bilal, and A. Gani, "Human Activity Recognition: Review, Taxonomy and Open Challenges," *Sensors (Basel)*, vol. 22, no. 17, Sep. 2022.
- [8] A. Das Antar, M. Ahmed, and M. A. R. Ahad, "Challenges in sensor-based human activity recognition and a comparative analysis of benchmark datasets: A review," *2019 Jt. 8th Int. Conf. Informatics, Electron. Vision, ICIEV 2019 3rd Int. Conf. Imaging, Vis. Pattern Recognition, icIVPR 2019 with Int. Conf. Act. Behav. Comput. ABC 2019*, pp. 134–139, May 2019.
- [9] F. Attal, S. Mohammed, M. Dedabrishvili, F. Chamroukhi, L. Oukhellou, and Y. Amirat, "Physical Human Activity Recognition Using Wearable Sensors," *Sensors 2015, Vol. 15, Pages 31314-31338*, vol. 15, no. 12, pp. 31314–31338, Dec. 2015.
- [10] J. Lockman, R. S. Fisher, and D. M. Olson, "Detection of seizure-like movements using a wrist accelerometer," *Epilepsy Behav.*, vol. 20, no. 4, pp. 638–641, Apr. 2011.
- [11] D. J. Wile, R. Ranaway, and Z. H. T. Kiss, "Smart watch accelerometry for analysis and diagnosis of tremor," *J. Neurosci. Methods*, vol. 230, pp. 1–4, Jun. 2014.
- [12] A. L. Patterson *et al.*, "SmartWatch by SmartMonitor: Assessment of Seizure Detection Efficacy for Various Seizure Types in Children, a Large Prospective Single-Center Study," *Pediatr. Neurol.*, vol. 53, no. 4, pp. 309–311, Oct. 2015.

- [13] O. Gani, "A Novel Approach to Complex Human Activity Recognition," 2009.
- [14] L. Köping, K. Shirahama, and M. Grzegorzec, "A general framework for sensor-based human activity recognition," *Comput. Biol. Med.*, vol. 95, pp. 248–260, Apr. 2018.
- [15] M. M. Hassan, M. Z. Uddin, A. Mohamed, and A. Almogren, "A robust human activity recognition system using smartphone sensors and deep learning," *Futur. Gener. Comput. Syst.*, vol. 81, pp. 307–313, Apr. 2018.
- [16] A. S. Abdull Sukor, A. Zakaria, and N. Abdul Rahim, "Activity recognition using accelerometer sensor and machine learning classifiers," *Proc. - 2018 IEEE 14th Int. Colloq. Signal Process. its Appl. CSPA 2018*, pp. 233–238, May 2018.
- [17] P. Bota, J. Silva, D. Folgado, and H. Gamboa, "A Semi-Automatic Annotation Approach for Human Activity Recognition," *Sensors 2019, Vol. 19, Page 501*, vol. 19, no. 3, p. 501, Jan. 2019.
- [18] C. Odhiambo, P. Wright, C. Corbett, and H. Valafar, "MedSensor: Medication Adherence Monitoring Using Neural Networks on Smartwatch Accelerometer Sensor Data."
- [19] C. O. Odhiambo, A. Torkjazi, C. A. Cole, and H. Valafar, "State transition modeling of the smoking behavior using LSTM recurrent neural networks," *arXiv*, 2020.
- [20] Q. Zhu, Z. Chen, and Y. C. Soh, "A Novel Semisupervised Deep Learning Method for Human Activity Recognition," *IEEE Trans. Ind. Informatics*, vol. 15, no. 7, pp. 3821–3830, Jul. 2019.
- [21] Y. Du, Y. Lim, and Y. Tan, "A Novel Human Activity Recognition and Prediction in Smart Home Based on Interaction," *Sensors 2019, Vol. 19, Page 4474*, vol. 19, no. 20, p. 4474, Oct. 2019.
- [22] A. Chelli and M. Patzold, "A Machine Learning Approach for Fall Detection and Daily Living Activity Recognition," *IEEE Access*, vol. 7, pp. 38670–38687, 2019.
- [23] K. Chen, L. Yao, D. Zhang, X. Wang, X. Chang, and F. Nie, "A Semisupervised Recurrent Convolutional Attention Model for Human Activity Recognition," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 31, no. 5, pp. 1747–1756, May 2020.
- [24] A. R. Javed, M. U. Sarwar, S. Khan, C. Iwendi, M. Mittal, and N. Kumar, "Analyzing the Effectiveness and Contribution of Each Axis of Tri-Axial Accelerometer Sensor for Accurate Activity Recognition," *Sensors 2020, Vol. 20, Page 2216*, vol. 20, no. 8, p. 2216, Apr. 2020.
- [25] C. Tong, S. A. Tailor, and N. D. Lane, "Are accelerometers for activity recognition a dead-end?," *HotMobile 2020 - Proc. 21st Int. Work. Mob. Comput. Syst. Appl.*, vol. 20, pp. 39–44, Mar. 2020.
- [26] S. Mekruksavanich and A. Jitpattanakul, "Biometric User Identification Based on Human Activity Recognition Using Wearable Sensors: An Experiment Using Deep Learning Models," *Electron. 2021, Vol. 10, Page 308*, vol. 10, no. 3, p. 308, Jan. 2021.
- [27] H. Haresamudram, I. Essa, and T. Plötz, "Contrastive Predictive Coding for Human Activity Recognition," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 2, p. 26, Jun. 2021.
- [28] S. Mekruksavanich, A. Jitpattanakul, P. Youplao, and P. Yupapin, "Enhanced Hand-Oriented Activity Recognition Based on Smartwatch Sensor Data Using

- LSTMs,” *Symmetry* 2020, Vol. 12, Page 1570, vol. 12, no. 9, p. 1570, Sep. 2020.
- [29] J. Pan, Z. Hu, S. Yin, and M. Li, “GRU with Dual Attentions for Sensor-Based Human Activity Recognition,” *Electron. 2022, Vol. 11, Page 1797*, vol. 11, no. 11, p. 1797, Jun. 2022.
 - [30] R. Saini, P. Kumar, P. P. Roy, and D. P. Dogra, “A novel framework of continuous human-activity recognition using Kinect,” *Neurocomputing*, vol. 311, pp. 99–111, Oct. 2018.
 - [31] M. Espinilla, J. Medina, J. Hallberg, and C. Nugent, “A new approach based on temporal sub-windows for online sensor-based activity recognition,” *J. Ambient Intell. Humaniz. Comput.* 2018, pp. 1–13, Mar. 2018.
 - [32] S. Alghyaline, “A Real-Time Street Actions Detection,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 2, pp. 322–329, Dec. 2019.
 - [33] Z. Chen, C. Jiang, and L. Xie, “A Novel Ensemble ELM for Human Activity Recognition Using Smartphone Sensors,” *IEEE Trans. Ind. Informatics*, vol. 15, no. 5, pp. 2691–2699, May 2019.
 - [34] H. Ma, W. Li, X. Zhang, S. Gao, and S. Lu, “Attnsense: Multi-level attention mechanism for multimodal human activity recognition,” *IJCAI Int. Jt. Conf. Artif. Intell.*, vol. 2019-August, pp. 3109–3115, 2019.
 - [35] N. Almaadeed, O. Elharrouss, S. Al-Maadeed, A. Bouridane, and A. Beghdadi, “A Novel Approach for Robust Multi Human Action Recognition and Summarization based on 3D Convolutional Neural Networks,” Jul. 2019.
 - [36] J. Gleason, R. Ranjan, S. Schwarcz, C. D. Castillo, J. C. Chen, and R. Chellappa, “A proposal-based solution to spatio-temporal action detection in untrimmed videos,” *Proc. - 2019 IEEE Winter Conf. Appl. Comput. Vision, WACV 2019*, pp. 141–150, Mar. 2019.
 - [37] Z. Wu, C. Xiong, C.-Y. Ma, R. Socher, and L. S. Davis, “AdaFrame: Adaptive Frame Selection for Fast Video Recognition.”
 - [38] A. Nadeem, A. Jalal, and K. Kim, “Accurate Physical Activity Recognition using Multidimensional Features and Markov Model for Smart Health Fitness,” *Symmetry* 2020, Vol. 12, Page 1766, vol. 12, no. 11, p. 1766, Oct. 2020.
 - [39] Y. Ishikawa, S. Kasai, Y. Aoki, and H. Kataoka, “Alleviating over-segmentation errors by detecting action boundaries,” *Proc. - 2021 IEEE Winter Conf. Appl. Comput. Vision, WACV 2021*, pp. 2321–2330, Jan. 2021.
 - [40] L. D’arco, H. Wang, and H. Zheng, “Assessing Impact of Sensors and Feature Selection in Smart-Insole-Based Human Activity Recognition,” *Methods Protoc.* 2022, Vol. 5, Page 45, vol. 5, no. 3, p. 45, May 2022.
 - [41] F. Lamonaca, D. L. Carnì, H. Najeh, C. Lohr, and B. Leduc, “Dynamic Segmentation of Sensor Events for Real-Time Human Activity Recognition in a Smart Home Context,” *Sensors* 2022, Vol. 22, Page 5458, vol. 22, no. 14, p. 5458, Jul. 2022.
 - [42] Y. Jiang, J. Wang, Y. Liang, and J. Xia, “Combining static and dynamic features for real-time moving pedestrian detection,” *Multimed. Tools Appl.*, vol. 78, no. 3, pp. 3781–3795, Feb. 2019.
 - [43] R. Dhaya, “CCTV Surveillance for Unprecedented Violence and Traffic Monitoring,” *J. Innov. Image Process.*, 2020.
 - [44] A. Basha, P. Parthasarathy, and S. Vivekanandan, “Detection of Suspicious

Human Activity based on CNN-DBNN Algorithm for Video Surveillance Applications,” *2019 Innov. Power Adv. Comput. Technol. i-PACT 2019*, Mar. 2019.

- [45] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, A. Bouridane, and A. Beghdadi, “A combined multiple action recognition and summarization for surveillance video sequences,” *Appl. Intell.*, vol. 51, no. 2, pp. 690–712, Feb. 2021.
- [46] Z. Qin, H. Liu, B. Song, M. Alazab, and P. M. Kumar, “Detecting and preventing criminal activities in shopping malls using massive video surveillance based on deep learning models,” *Ann. Oper. Res.*, pp. 1–18, Sep. 2021.
- [47] M. S. Mahdi, A. Jelwy Mohammed, A. waedallah Abdulghafour, and D. Al-Waqf Al-Sunni, “Detection of Unusual Activity in Surveillance Video Scenes Based on Deep Learning Strategies,” *J. Al-Qadisiyah Comput. Sci. Math.*, vol. 13, no. 4, p. Page 1 – 9-Page 1 – 9, Dec. 2021.
- [48] A. Mohan, M. Choksi, and M. A. Zaveri, “Anomaly and Activity Recognition Using Machine Learning Approach for Video Based Surveillance,” *2019 10th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2019*, Jul. 2019.
- [49] O. M. Rajpurkar, S. S. Kamble, J. P. Nandagiri, and A. V. Nimkar, “Alert Generation on Detection of Suspicious Activity Using Transfer Learning,” *2020 11th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2020*, Jul. 2020.
- [50] C. V. Amrutha, C. Jyotsna, and J. Amudha, “Deep Learning Approach for Suspicious Activity Detection from Surveillance Video,” *2nd Int. Conf. Innov. Mech. Ind. Appl. ICIMIA 2020 - Conf. Proc.*, pp. 335–339, Mar. 2020.
- [51] A. V Pyataeva and M. S. Eliseeva, “Video based human smoking event detection method,” 2021.
- [52] H. Riaz, M. Uzair, H. Ullah, and M. Ullah, “Anomalous Human Action Detection Using a Cascade of Deep Learning Models,” *Proc. - Eur. Work. Vis. Inf. Process. EUVIP*, vol. 2021-June, Jun. 2021.
- [53] M. Mudgal, D. Punj, and A. Pillai, “Suspicious Action Detection in Intelligent Surveillance System Using Action Attribute Modelling,” *J. Web Eng.*, vol. 20, no. 1, pp. 129–146–129–146, Feb. 2021.
- [54] W. Ullah *et al.*, “Artificial Intelligence of Things-assisted two-stream neural network for anomaly detection in surveillance Big Video Data,” *Futur. Gener. Comput. Syst.*, vol. 129, pp. 286–297, Apr. 2022.
- [55] A. Gumaei, M. M. Hassan, A. Alelaiwi, and H. Als Salman, “A Hybrid Deep Learning Model for Human Activity Recognition Using Multimodal Body Sensing Data,” *IEEE Access*, vol. 7, pp. 99152–99160, 2019.
- [56] M. Z. Uddin and M. M. Hassan, “Activity Recognition for Cognitive Assistance Using Body Sensors Data and Deep Convolutional Neural Network,” *IEEE Sens. J.*, vol. 19, no. 19, pp. 8413–8419, Oct. 2019.
- [57] W. Taylor, S. A. Shah, K. Dashtipour, A. Zahid, Q. H. Abbasi, and M. A. Imran, “An Intelligent Non-Invasive Real-Time Human Activity Recognition System for Next-Generation Healthcare,” *Sensors 2020, Vol. 20, Page 2653*, vol. 20, no. 9, p. 2653, May 2020.
- [58] D. ; Bhattacharya *et al.*, “Ensem-HAR: An Ensemble Deep Learning Model for Smartphone Sensor-Based Human Activity Recognition for Measurement of Elderly Health Monitoring,” *Biosens. 2022, Vol. 12, Page 393*, vol. 12, no. 6, p.

393, Jun. 2022.

- [59] K. Kastelic, M. Dobnik, S. Löfler, C. Hofer, and N. Šarabon, “Validity, Reliability and Sensitivity to Change of Three Consumer-Grade Activity Trackers in Controlled and Free-Living Conditions among Older Adults,” *Sensors* 2021, Vol. 21, Page 6245, vol. 21, no. 18, p. 6245, Sep. 2021.
- [60] K. E. Heron and J. M. Smyth, “Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments,” *Br. J. Health Psychol.*, vol. 15, no. Pt 1, pp. 1–39, Feb. 2010.
- [61] I. Nahum-Shani *et al.*, “Just-in-Time Adaptive Interventions (JITAI)s in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support,” *Ann. Behav. Med.*, vol. 52, no. 6, pp. 446–462, May 2018.
- [62] C. K. Martin, A. C. Miller, D. M. Thomas, C. M. Champagne, H. Han, and T. Church, “Efficacy of SmartLoss, a smartphone-based weight loss intervention: results from a randomized controlled trial,” *Obesity (Silver Spring)*, vol. 23, no. 5, pp. 935–942, May 2015.
- [63] C. K. Martin, L. A. Gilmore, J. W. Apolzan, C. A. Myers, D. M. Thomas, and L. M. Redman, “Smartloss: A Personalized Mobile Health Intervention for Weight Management and Health Promotion,” *JMIR Mhealth Uhealth* 2016;4(1)e18 <https://mhealth.jmir.org/2016/1/e18>, vol. 4, no. 1, p. e5027, Mar. 2016.
- [64] L. M. Redman *et al.*, “Effectiveness of SmartMoms, a Novel eHealth Intervention for Management of Gestational Weight Gain: Randomized Controlled Pilot Trial,” *JMIR mHealth uHealth*, vol. 5, no. 9, Sep. 2017.
- [65] S. C. Hsu, C. H. Chuang, C. L. Huang, P. R. Teng, and M. J. Lin, “A video-based abnormal human behavior detection for psychiatric patient monitoring,” *2018 Int. Work. Adv. Image Technol. IWAIT 2018*, pp. 1–4, May 2018.
- [66] K. E. Ko and K. B. Sim, “Deep convolutional framework for abnormal behavior detection in a smart surveillance system,” *Eng. Appl. Artif. Intell.*, vol. 67, pp. 226–234, Jan. 2018.
- [67] K. G. Gunale and P. Mukherji, “Deep Learning with a Spatiotemporal Descriptor of Appearance and Motion Estimation for Video Anomaly Detection,” *J. Imaging* 2018, Vol. 4, Page 79, vol. 4, no. 6, p. 79, Jun. 2018.
- [68] A.-M. Founta, D. Chatzakou, N. Kourtellis, J. Blackburn, and A. Vakali, “A Unified Deep Learning Architecture for Abuse Detection,” *Proc. 10th ACM Conf. Web Sci.*, p. 10, 2019.
- [69] Y. Dou, C. Fudong, J. Li, and C. Wei, “Abnormal Behavior Detection Based on Optical Flow Trajectory of Human Joint Points,” *Proc. 31st Chinese Control Decis. Conf. CCDC 2019*, pp. 653–658, Jun. 2019.
- [70] Y. Moukafih, H. Hafidi, and M. Ghogho, “Aggressive Driving Detection Using Deep Learning-based Time Series Classification,” *IEEE Int. Symp. Innov. Intell. Syst. Appl. INISTA 2019 - Proc.*, Jul. 2019.
- [71] J. Lee and S.-J. Shin, “A Study of Video-Based Abnormal Behavior Recognition Model Using Deep Learning,” *Int. J. Adv. smart Converg.*, vol. 9, no. 4, pp. 115–119, 2020.
- [72] L. Xia and Z. Li, “A new method of abnormal behavior detection using LSTM network with temporal attention mechanism,” *J. Supercomput.*, vol. 77, no. 4, pp. 3223–3241, Apr. 2021.

- [73] K. L. Bhagya Jyothi and Vasudeva, "Chronological Poor and Rich Tunicate Swarm Algorithm integrated Deep Maxout Network for human action and abnormality detection," *2021 4th Int. Conf. Electr. Comput. Commun. Technol. ICECCT 2021*, 2021.
- [74] A. Belhadi, Y. Djenouri, G. Srivastava, D. Djenouri, J. C. W. Lin, and G. Fortino, "Deep learning for pedestrian collective behavior analysis in smart cities: A model of group trajectory outlier detection," *Inf. Fusion*, vol. 65, pp. 13–20, Jan. 2021.
- [75] X. Shu, L. Zhang, Y. Sun, and J. Tang, "Host-Parasite: Graph LSTM-in-LSTM for Group Activity Recognition," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 32, no. 2, pp. 663–674, Feb. 2021.
- [76] Z. Ebrahimpour, W. Wan, O. Cervantes, T. Luo, and H. Ullah, "Comparison of Main Approaches for Extracting Behavior Features from Crowd Flow Analysis," *ISPRS Int. J. Geo-Information 2019, Vol. 8, Page 440*, vol. 8, no. 10, p. 440, Oct. 2019.
- [77] S. M. Azar, M. G. Atigh, A. Nickabadi, and A. Alahi, "Convolutional relational machine for group activity recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 7884–7893, Jun. 2019.
- [78] Q. Wang, M. Chen, F. Nie, and X. Li, "Detecting Coherent Groups in Crowd Scenes by Multiview Clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 46–58, Jan. 2020.
- [79] H. Ullah, I. U. Islam, M. Ullah, M. Afaq, S. D. Khan, and J. Iqbal, "Multi-feature-based crowd video modeling for visual event detection," *Multimed. Syst.*, vol. 27, no. 4, pp. 589–597, Aug. 2021.
- [80] J. Tang, X. Shu, R. Yan, and L. Zhang, "Coherence Constrained Graph LSTM for Group Activity Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 2, pp. 636–647, Feb. 2022.
- [81] L. Lazaridis, A. Dimou, and P. Daras, "Abnormal behavior detection in crowded scenes using density heatmaps and optical flow," *Eur. Signal Process. Conf.*, vol. 2018-September, pp. 2060–2064, Nov. 2018.
- [82] T. Wang *et al.*, "Abnormal event detection based on analysis of movement information of video sequence," *Optik (Stuttg.)*, vol. 152, pp. 50–60, Jan. 2018.
- [83] H. Rabiee, H. Mousavi, M. Nabi, and M. Ravanbakhsh, "Detection and localization of crowd behavior using a novel tracklet-based model," *Int. J. Mach. Learn. Cybern. 2017 912*, vol. 9, no. 12, pp. 1999–2010, Apr. 2017.
- [84] S. Amraee, A. Vafaei, K. Jamshidi, and P. Adibi, "Abnormal event detection in crowded scenes using one-class SVM," *Signal, Image Video Process.*, vol. 12, no. 6, pp. 1115–1123, Sep. 2018.
- [85] Y. Liu, K. Hao, X. Tang, and T. Wang, "Abnormal Crowd Behavior Detection Based on Predictive Neural Network," *Proc. 2019 IEEE Int. Conf. Artif. Intell. Comput. Appl. ICAICA 2019*, pp. 221–225, Mar. 2019.
- [86] S. D. Khan, "Congestion detection in pedestrian crowds using oscillation in motion trajectories," *Eng. Appl. Artif. Intell.*, vol. 85, pp. 429–443, Oct. 2019.
- [87] X. Zhang, X. Shu, and Z. He, "Crowd panic state detection using entropy of the distribution of enthalpy," *Phys. A Stat. Mech. its Appl.*, vol. 525, pp. 935–945, Jul. 2019.
- [88] T. Gupta, V. Nunavath, and S. Roy, "CrowdVAS-Net: A deep-CNN based

- framework to detect abnormal crowd-motion behavior in videos for predicting crowd disaster,” *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern.*, vol. 2019-October, pp. 2877–2882, Oct. 2019.
- [89] T. Alafif, B. Alzahrani, Y. Cao, R. Alotaibi, A. Barnawi, and M. Chen, “Generative adversarial network based abnormal behavior detection in massive crowd videos: a Hajj case study,” *J. Ambient Intell. Humaniz. Comput.* 2021 138, vol. 13, no. 8, pp. 4077–4088, Jun. 2021.
 - [90] K. Doshi and Y. Yilmaz, “Rethinking Video Anomaly Detection - A Continual Learning Approach,” *Proc. - 2022 IEEE/CVF Winter Conf. Appl. Comput. Vision, WACV 2022*, pp. 3036–3045, 2022.
 - [91] A. Sunil, M. H. Sheth, E. Shreyas, and Mohana, “Usual and Unusual Human Activity Recognition in Video using Deep Learning and Artificial Intelligence for Security Applications,” *2021 4th Int. Conf. Electr. Comput. Commun. Technol. ICECCT 2021*, 2021.
 - [92] X. Zhang, Q. Zhang, S. Hu, C. Guo, and H. Yu, “Energy Level-Based Abnormal Crowd Behavior Detection,” *Sensors 2018, Vol. 18, Page 423*, vol. 18, no. 2, p. 423, Feb. 2018.
 - [93] D. H. Yoon, N. G. Cho, and S. W. Lee, “A novel online action detection framework from untrimmed video streams,” *Pattern Recognit.*, vol. 106, p. 107396, Oct. 2020.
 - [94] A. Bhargava, G. Salunkhe, and K. Bhosale, “A comprehensive study and detection of anomalies for autonomous video surveillance using neuromorphic computing and self learning algorithm,” *2020 Int. Conf. Conver. to Digit. World - Quo Vadis, ICCDW 2020*, Feb. 2020.
 - [95] C. Ma, W. Li, J. Cao, J. Du, Q. Li, and R. Gravina, “Adaptive sliding window based activity recognition for assisted livings,” *Inf. Fusion*, vol. 53, pp. 55–65, Jan. 2020.
 - [96] Z. N. Khan and J. Ahmad, “Attention induced multi-head convolutional neural network for human activity recognition,” *Appl. Soft Comput.*, vol. 110, p. 107671, Oct. 2021.
 - [97] J. Zhang, C. Wu, Y. Wang, and P. Wang, “Detection of abnormal behavior in narrow scene with perspective distortion,” *Mach. Vis. Appl.*, vol. 30, no. 5, pp. 987–998, Jul. 2019.
 - [98] M. Shoaib, O. D. Incel, H. Scolten, and P. Havinga, “Resource consumption analysis of online activity recognition on mobile phones and smartwatches,” in *2017 IEEE 36th International Performance Computing and Communications Conference, IPCCC 2017*, 2018, vol. 2018-January, pp. 1–6.
 - [99] S. Chakraborty, C. Y. Han, X. Zhou, and W. G. Wee, “A Context Driven Human Activity Recognition Framework.”
 - [100] A. M. Helmi, M. A. A. Al-Qaness, A. Dahou, R. Damaševičius, T. Krilavičius, and M. A. Elaziz, “A Novel Hybrid Gradient-Based Optimizer and Grey Wolf Optimizer Feature Selection Method for Human Activity Recognition Using Smartphone Sensors,” *Entropy 2021, Vol. 23, Page 1065*, vol. 23, no. 8, p. 1065, Aug. 2021.
 - [101] H. Kwon, G. D. Abowd, and T. Plötz, “Handling Annotation Uncertainty in Human Activity Recognition,” *Proc. 23rd Int. Symp. Wearable Comput.*

- [102] K. E. Warner, T. A. Hodgson, and C. E. Carroll, "Medical costs of smoking in the United States: estimates, their validity, and their implications," *Tob. Control*, vol. 8, pp. 290–300, 1999.
- [103] S. De Jesus, A. Hsin, G. Faulkner, and H. Prapavessis, "A systematic review and analysis of data reduction techniques for the CReSS smoking topography device," *J. Smok. Cessat.*, vol. 10, no. 1, pp. 12–28, Oct. 2013.
- [104] E. T. Moolchan, C. S. Parzynski, M. Jaszyna-Gasior, C. C. Collins, M. K. Leff, and D. L. Zimmerman, "A link between adolescent nicotine metabolism and smoking topography," *Cancer Epidemiol. Biomarkers Prev.*, vol. 18, no. 5, pp. 1578–1583, May 2009.
- [105] F. H. Franken, W. B. Pickworth, D. H. Epstein, and E. T. Moolchan, "Smoking rates and topography predict adolescent smoking cessation following treatment with nicotine replacement therapy," *Cancer Epidemiol. Biomarkers Prev.*, vol. 15, no. 1, pp. 154–157, Jan. 2006.
- [106] L. W. Frederiksen, P. M. Miller, and G. L. Peterson, "Topographical components of smoking behavior," *Addict. Behav.*, vol. 2, no. 1, pp. 55–61, 1977.
- [107] C. A. Cole, D. Anshari, V. Lambert, J. F. Thrasher, and H. Valafar, "Detecting Smoking Events Using Accelerometer Data Collected Via Smartwatch Technology: Validation Study.," *JMIR mHealth uHealth*, vol. 5, no. 12, p. e189, Dec. 2017.
- [108] D. L. Patrick, A. Cheadle, D. C. Thompson, P. Diehr, T. Koepsell, and S. Kinne, "The validity of self-reported smoking: A review and meta-analysis," *Am. J. Public Health*, vol. 84, no. 7, pp. 1086–1093, 1994.
- [109] M. Ashfak Habib, M. S. Mohktar, S. Bahyah Kamaruzzaman, K. Seang Lim, T. Maw Pin, and F. Ibrahim, "Smartphone-based solutions for fall detection and prevention: Challenges and open issues," *Sensors (Switzerland)*, vol. 14, no. 4. MDPI AG, pp. 7181–7208, 22-Apr-2014.
- [110] T. R. Mauldin, M. E. Canby, V. Metsis, A. H. H. Ngu, and C. C. Rivera, "SmartFall: A Smartwatch-Based Fall Detection System Using Deep Learning," *Sensors (Basel)*, vol. 18, no. 10, Oct. 2018.
- [111] C. Ma, D. Wong, W. Lam, A. Wan, and W. Lee, "Balance Improvement Effects of Biofeedback Systems with State-of-the-Art Wearable Sensors: A Systematic Review," *Sensors*, vol. 16, no. 4, p. 434, Mar. 2016.
- [112] C. A. Cole, J. F. Thrasher, S. M. Strayer, and H. Valafar, "Resolving ambiguities in accelerometer data due to location of sensor on wrist in application to detection of smoking gesture," in *2017 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2017*, 2017, pp. 489–492.
- [113] S. Sen, V. Subbaraju, A. Misra, R. K. Balan, and Y. Lee, "The case for smartwatch-based diet monitoring," in *2015 IEEE International Conference on Pervasive Computing and Communication Workshops, PerCom Workshops 2015*, 2015, pp. 585–590.
- [114] G. M. Weiss, J. L. Timko, C. M. Gallagher, K. Yoneda, and A. J. Schreiber, "Smartwatch-based activity recognition: A machine learning approach," in *3rd IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2016*, 2016, pp. 426–429.
- [115] M. Shoaib, S. Bosch, O. Incel, H. Scholten, and P. Havinga, "Complex Human

- Activity Recognition Using Smartphone and Wrist-Worn Motion Sensors,” *Sensors*, vol. 16, no. 4, p. 426, Mar. 2016.
- [116] V. Genovese, A. Mannini, and A. M. Sabatini, “A Smartwatch Step Counter for Slow and Intermittent Ambulation,” *IEEE Access*, vol. 5, pp. 13028–13037, 2017.
 - [117] D. Arifoglu and A. Bouchachia, “Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks,” in *Procedia Computer Science*, 2017, vol. 110, pp. 86–93.
 - [118] X. Sun, L. Qiu, Y. Wu, Y. Tang, and G. Cao, “SleepMonitor,” *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–22, Sep. 2017.
 - [119] A. Pfannenstiel and B. S. Chaparro, “An investigation of the usability and desirability of health and fitness-tracking devices,” in *Communications in Computer and Information Science*, 2015, vol. 529, pp. 473–477.
 - [120] N. Saleheen *et al.*, “PuffMarker: A multi-sensor approach for pinpointing the timing of first lapse in smoking cessation,” in *UbiComp 2015 - Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 999–1010.
 - [121] A. L. Skinner, C. J. Stone, H. Doughty, and M. R. Munafò, “StopWatch: The preliminary evaluation of a smartwatch-based system for passive detection of cigarette smoking,” *Nicotine and Tobacco Research*, vol. 21, no. 2. Oxford University Press, pp. 257–261, 01-Feb-2019.
 - [122] C. A. Cole, D. Anshari, V. Lambert, J. F. Thrasher, and H. Valafar, “Detecting Smoking Events Using Accelerometer Data Collected Via Smartwatch Technology: Validation Study,” *JMIR mHealth uHealth*, vol. 5, no. 12, p. e189, Dec. 2017.
 - [123] “(16) (PDF) Recognition of Smoking Gesture Using Smart Watch Technology.” [Online]. Available: https://www.researchgate.net/publication/315720056_Recognition_of_Smoking_Gesture_Using_Smart_Watch_Technology. [Accessed: 08-Nov-2019].
 - [124] L. E. Wagenknecht, G. L. Burke, L. L. Perkins, N. J. Haley, and G. D. Friedman, “Misclassification of smoking status in the CARDIA study: A comparison of self-report with serum cotinine levels,” *Am. J. Public Health*, vol. 82, no. 1, pp. 33–36, 1992.
 - [125] D. P. Kingma and J. Lei Ba, “ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION.”
 - [126] J. C. Núñez, R. Cabido, J. F. Vélez, A. S. Montemayor, and J. J. Pantrigo, “Multiview 3D human pose estimation using improved least-squares and LSTM networks,” *Neurocomputing*, vol. 323, pp. 335–343, Jan. 2019.
 - [127] J. Ma, A. Ovalle, and D. M. K. Woodbridge, “Medhere: A Smartwatch-based Medication Adherence Monitoring System using Machine Learning and Distributed Computing,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2018, vol. 2018-July, pp. 4945–4948.
 - [128] M. Aldeer, M. Javanmard, and R. Martin, “A Review of Medication Adherence Monitoring Technologies,” *Appl. Syst. Innov.*, vol. 1, no. 2, p. 14, May 2018.
 - [129] M. T. Brown, J. Bussell, S. Dutta, K. Davis, S. Strong, and S. Mathew,

- “Medication Adherence: Truth and Consequences,” *Am. J. Med. Sci.*, vol. 351, no. 4, pp. 387–399, Apr. 2016.
- [130] L. Osterberg and T. Blaschke, “Adherence to Medication,” *N. Engl. J. Med.*, vol. 353, no. 5, pp. 487–497, Aug. 2005.
- [131] H. Kalantarian, N. Alshurafa, E. Nemati, T. Le, and M. Sarrafzadeh, “A smartwatch-based medication adherence system,” in *2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks, BSN 2015*, 2015.
- [132] “Medication Adherence: The Elephant in the Room.” [Online]. Available: <https://www.uspharmacist.com/article/medication-adherence-the-elephant-in-the-room>. [Accessed: 30-Mar-2021].
- [133] D. Fozoonmayeh *et al.*, “A Scalable Smartwatch-Based Medication Intake Detection System Using Distributed Machine Learning,” *J. Med. Syst.*, vol. 44, no. 4, pp. 1–14, Apr. 2020.
- [134] A. Ortis, P. Caponnetto, R. Polosa, S. Urso, and S. Battiato, “A report on smoking detection and quitting technologies,” *International Journal of Environmental Research and Public Health*, vol. 17, no. 7. MDPI AG, 01-Apr-2020.
- [135] C. A. Cole, S. Powers, R. L. Tomko, B. Froeliger, and H. Valafar, “Quantification of Smoking Characteristics Using Smartwatch Technology: Pilot Feasibility Study of New Technology,” *JMIR Form. Res.*, vol. 5, no. 2, p. e20464, Feb. 2021.
- [136] M. H. Imtiaz, R. I. Ramos-Garcia, S. Wattal, S. Tiffany, and E. Sazonov, “Wearable sensors for monitoring of cigarette smoking in free-living: A systematic review,” *Sensors (Switzerland)*, vol. 19, no. 21. MDPI AG, 01-Nov-2019.
- [137] A. K. Bourke and G. M. Lyons, “A threshold-based fall-detection algorithm using a bi-axial gyroscope sensor,” *Med. Eng. Phys.*, vol. 30, no. 1, pp. 84–90, Jan. 2008.
- [138] J. J. Guiry, P. van de Ven, and J. Nelson, “Multi-sensor fusion for enhanced contextual awareness of everyday activities with ubiquitous devices,” *Sensors (Switzerland)*, vol. 14, no. 3, pp. 5687–5701, Mar. 2014.
- [139] S. B. Khojasteh, J. R. Villar, C. Chira, V. M. González, and E. de la Cal, “Improving fall detection using an on-wrist wearable accelerometer,” *Sensors (Switzerland)*, vol. 18, no. 5, May 2018.
- [140] M. Shoaib, S. Bosch, O. Durmaz Incel, H. Scholten, and P. J. M. Havinga, “Fusion of smartphone motion sensors for physical activity recognition,” *Sensors (Switzerland)*, vol. 14, no. 6, pp. 10146–10176, Jun. 2014.
- [141] A. Moschetti, L. Fiorini, D. Esposito, P. Dario, and F. Cavallo, “Recognition of daily gestures with wearable inertial rings and bracelets,” *Sensors (Switzerland)*, vol. 16, no. 8, Aug. 2016.
- [142] R. I. Ramos-Garcia, E. R. Muth, J. N. Gowdy, and A. W. Hoover, “Improving the recognition of eating gestures using intergesture sequential dependencies,” *IEEE J. Biomed. Heal. Informatics*, vol. 19, no. 3, pp. 825–831, May 2015.
- [143] A. A. Aguilera, R. F. Brena, O. Mayora, E. Molino-Minero-re, and L. A. Trejo, “Multi-sensor fusion for activity recognition—a survey,” *Sensors (Switzerland)*, vol. 19, no. 17. MDPI AG, 01-Sep-2019.
- [144] M. El Alili, B. Vrijens, J. Demonceau, S. M. Evers, and M. Hilgsmann, “A scoping review of studies comparing the medication event monitoring system (MEMS) with alternative methods for measuring medication adherence,” *Br. J.*

- Clin. Pharmacol.*, vol. 82, no. 1, pp. 268–279, 2016.
- [145] S. L. West, D. A. Savitz, G. Koch, B. L. Strom, H. A. Guess, and A. Hartzema, “Recall Accuracy for Prescription Medications: Self-report Compared with Database Information,” *Am. J. Epidemiol.*, vol. 142, no. 10, pp. 1103–1112, Nov. 1995.
 - [146] D. Byrne, A. R. Doherty, G. J. F. Jones, A. F. Smeaton, S. Kumpulainen, and K. Järvelin, “The SenseCam as a Tool for Task Observation,” 2008.
 - [147] M. Straczekiewicz, N. W. Glynn, and J. Harezlak, “On Placement, Location and Orientation of Wrist-Worn Tri-Axial Accelerometers during Free-Living Measurements,” *Sensors (Basel)*, vol. 19, no. 9, May 2019.
 - [148] A. Doherty *et al.*, “Large scale population assessment of physical activity using wrist worn accelerometers: The UK biobank study,” *PLoS One*, vol. 12, no. 2, Feb. 2017.
 - [149] I. Maglogiannis, G. Spyroglou, C. Panagopoulos, M. Mazonaki, and P. Tsanakis, “Mobile reminder system for furthering patient adherence utilizing commodity smartwatch and Android devices,” in *Proceedings of the 2014 4th International Conference on Wireless Mobile Communication and Healthcare - “Transforming Healthcare Through Innovations in Mobile and Wireless Technologies”*, *MOBIHEALTH 2014*, 2015, pp. 124–127.
 - [150] N. V. Dhurandhar *et al.*, “Energy balance measurement: when something is not better than nothing,” *Int. J. Obes. (Lond)*, vol. 39, no. 7, pp. 1109–1113, Jul. 2015.
 - [151] C. Höchsmann and C. K. Martin, “Review of the validity and feasibility of image-assisted methods for dietary assessment,” *Int. J. Obes. (Lond)*, vol. 44, no. 12, pp. 2358–2371, Dec. 2020.
 - [152] B. E. Odigwe, J. S. Eyitayo, C. I. Odigwe, and H. Valafar, “Modelling of Sick Cell Anemia Patients Response to Hydroxyurea using Artificial Neural Networks,” Nov. 2019.
 - [153] L. Zhao, B. Odigwe, S. Lessner, D. Clair, F. Mussa, and H. Valafar, “Automated analysis of femoral artery calcification using machine learning techniques,” *Proc. - 6th Annu. Conf. Comput. Sci. Comput. Intell. CSCI 2019*, pp. 584–589, Dec. 2019.
 - [154] B. E. Odigwe, F. G. Spinale, and H. Valafar, “Application of Machine Learning in Early Recommendation of Cardiac Resynchronization Therapy,” Sep. 2021.
 - [155] V.-Y. H, H. P, S. J, V. T, and S. H, “Reliable recognition of lying, sitting, and standing with a hip-worn accelerometer,” *Scand. J. Med. Sci. Sports*, vol. 28, no. 3, pp. 1092–1102, Mar. 2018.
 - [156] C. O. Odhiambo, C. A. Cole, A. Torkjazi, and H. Valafar, “State transition modeling of the smoking behavior using lstm recurrent neural networks,” *Proc. - 6th Annu. Conf. Comput. Sci. Comput. Intell. CSCI 2019*, pp. 898–904, Dec. 2019.
 - [157] W. Jiang and Z. Yin, “Human activity recognition using wearable sensors by deep convolutional neural networks,” in *MM 2015 - Proceedings of the 2015 ACM Multimedia Conference*, 2015, pp. 1307–1310.
 - [158] P. Casale, O. Pujol, and P. Radeva, “Human Activity Recognition from Accelerometer Data Using a Wearable Device,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6669 LNCS, pp. 289–296, 2011.
 - [159] Y. Yu, X. Si, C. Hu, and J. Zhang, “A Review of Recurrent Neural Networks:

- LSTM Cells and Network Architectures,” *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019.
- [160] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
- [161] N. Twomey *et al.*, “A Comprehensive Study of Activity Recognition Using Accelerometers,” *Informatics 2018, Vol. 5, Page 27*, vol. 5, no. 2, p. 27, May 2018.
- [162] “TensorFlow-on-Android-for-Human-Activity-Recognition-with-LSTMs.” [Online]. Available: https://github.com/curiously/TensorFlow-on-Android-for-Human-Activity-Recognition-with-LSTMs/blob/master/human_activity_recognition.ipynb. [Accessed: 09-Nov-2021].
- [163] “Human Activity Recognition using LSTMs on Android | TensorFlow for Hackers (Part VI) | Curiously - Hacker’s Guide to Machine Learning.” [Online]. Available: <https://curiously.com/posts/human-activity-recognition-using-lstms-on-android/>. [Accessed: 09-Nov-2021].
- [164] Y. Zhao, R. Yang, G. Chevalier, X. Xu, and Z. Zhang, “Deep Residual Bidir-LSTM for Human Activity Recognition Using Wearable Sensors,” *Math. Probl. Eng.*, vol. 2018, Aug. 2017.
- [165] United States Census Bureau, “Race.” [Online]. Available: <https://www.census.gov/topics/population/race.html>. [Accessed: 13-Nov-2022].
- [166] W. H. Organization, “Adherence to Long-Term Therapies - Evidence for action,” 2003.
- [167] M. T. Brown and J. K. Bussell, “Medication adherence: WHO cares?,” *Mayo Clin. Proc.*, vol. 86, no. 4, pp. 304–314, 2011.
- [168] K. Kvarnström, A. Westerholm, M. Airaksinen, H. Liira, P. Kardas, and J. Lee, “pharmaceutics Factors Contributing to Medication Adherence in Patients with a Chronic Condition: A Scoping Review of Qualitative Research,” 2021.
- [169] V. S. Conn and T. M. Ruppar, “Medication adherence outcomes of 771 intervention trials: Systematic review and meta-analysis,” *Prev. Med. (Baltim.)*, vol. 99, pp. 269–276, Jun. 2017.
- [170] M. J. Stirratt, J. R. Curtis, M. I. Danila, R. Hansen, M. J. Miller, and C. A. Gakumo, “Advancing the Science and Practice of Medication Adherence,” *J. Gen. Intern. Med.*, vol. 33, no. 2, pp. 216–222, Feb. 2018.
- [171] M. Changizi and M. H. Kaveh, “Effectiveness of the mHealth technology in improvement of healthy behaviors in an elderly population—a systematic review,” *mHealth*, vol. 3, pp. 51–51, Nov. 2017.
- [172] N. Singh, U. Varshney, F. Rowe, and R. Klein, “European Journal of Information Systems IT-based reminders for medication adherence: systematic review, taxonomy, framework and research directions IT-based reminders for medication adherence: systematic review, taxonomy, framework and research directions,” 2019.
- [173] L. A. Anghel, A. M. Farcas, and R. N. Oprean, “An overview of the common methods used to measure treatment adherence,” *Med. Pharm. REPORTS*, vol. 92, no. 2, pp. 117–122, 2019.
- [174] A. L. McRae-Clark, N. L. Baker, S. C. Sonne, C. L. DeVane, A. Wagner, and J. Norton, “Concordance of Direct and Indirect Measures of Medication Adherence

- in A Treatment Trial for Cannabis Dependence,” *J. Subst. Abuse Treat.*, vol. 57, pp. 70–74, Oct. 2015.
- [175] S. C. Mukhopadhyay, “Wearable sensors for human activity monitoring: A review,” *IEEE Sens. J.*, vol. 15, no. 3, pp. 1321–1330, 2015.
 - [176] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, “A Survey on Activity Detection and Classification Using Wearable Sensors,” *IEEE Sens. J.*, vol. 17, no. 2, pp. 386–403, Jan. 2017.
 - [177] “Smartwatches - Statistics & Facts | Statista.” [Online]. Available: <https://www.statista.com/topics/4762/smartwatches/#dossierKeyfigures>. [Accessed: 27-Jan-2022].
 - [178] M. C. Mozer, “The Neural Network House: An Environment that Adapts to its Inhabitants,” 1998.
 - [179] U. Leonhardt and J. Magee, “Multi-sensor location tracking,” pp. 203–214, 1998.
 - [180] A. R. Golding and N. Lesh, “Indoor navigation using a diverse set of cheap, wearable sensors,” *Int. Symp. Wearable Comput. Dig. Pap.*, pp. 29–36, 1999.
 - [181] A. Ward, A. Jones, and A. Hopper, “A new location technique for the active office,” *IEEE Pers. Commun.*, vol. 4, no. 5, pp. 42–47, Oct. 1997.
 - [182] S. Jung *et al.*, “Wearable Fall Detector using Integrated Sensors and Energy Devices,” *Sci. Rep.*, vol. 5, 2015.
 - [183] K. Chaccour, R. Darazi, A. H. El Hassani, and E. Andres, “From Fall Detection to Fall Prevention: A Generic Classification of Fall-Related Systems,” *IEEE Sens. J.*, vol. 17, no. 3, 2017.
 - [184] J. Qu, C. Wu, Q. Li, T. Wang, and A. H. Soliman, “Human fall detection algorithm design based on sensor fusion and multi-threshold comprehensive judgment,” *Sensors Mater.*, vol. 32, no. 4, pp. 1209–1221, Apr. 2020.
 - [185] J. M. Peake, G. Kerr, and J. P. Sullivan, “A critical review of consumer wearables, mobile applications, and equipment for providing biofeedback, monitoring stress, and sleep in physically active populations,” *Frontiers in Physiology*, vol. 9, no. JUN. 2018.
 - [186] A. Smith, H. Anand, S. Milosavljevic, K. M. Rentschler, A. Pocivavsek, and H. Valafar, “Application of Machine Learning to Sleep Stage Classification,” in *2021 International Conference on Computational Science and Computational Intelligence (CSCI)*, 2021, pp. 349–354.
 - [187] “Sensor Necklace Records When Pill In Swallowed, And Prompts Patient When It Is Time To Take Another -- ScienceDaily.” [Online]. Available: <https://www.sciencedaily.com/releases/2008/03/080305111857.htm>. [Accessed: 31-Jan-2022].
 - [188] H. Kalantarian, B. Motamed, N. Alshurafa, and M. Sarrafzadeh, “A wearable sensor system for medication adherence prediction,” *Artif. Intell. Med.*, vol. 69, pp. 43–52, May 2016.
 - [189] H. Kalantarian, N. Alshurafa, and M. Sarrafzadeh, “A Survey of Diet Monitoring Technology,” *IEEE Pervasive Comput.*, vol. 16, no. 1, pp. 57–65, Jan. 2017.
 - [190] J. Bennett, O. Rokas, and L. Chen, “Healthcare in the Smart Home: A Study of Past, Present and Future,” *Sustain. 2017, Vol. 9, Page 840*, vol. 9, no. 5, p. 840, May 2017.
 - [191] I. Iancu and B. Iancu, “Elderly in the Digital Era. Theoretical Perspectives on

- Assistive Technologies,” *Technol.* 2017, Vol. 5, Page 60, vol. 5, no. 3, p. 60, Sep. 2017.
- [192] C. O. Odhiambo, S. Saha, C. K. Martin, and H. Valafar, “Human Activity Recognition on Time Series Accelerometer Sensor Data using LSTM Recurrent Neural Networks.”
- [193] “WISDM Lab: Wireless Sensor Data Mining.” [Online]. Available: <https://www.cis.fordham.edu/wisdm/>. [Accessed: 14-Sep-2022].
- [194] M. J. Stirratt *et al.*, “Self-report measures of medication adherence behavior: recommendations on optimal use,” *Transl. Behav. Med.*, vol. 5, no. 4, p. 470, Dec. 2015.
- [195] S. Bansilal *et al.*, “Assessing the Impact of Medication Adherence on Long-Term Cardiovascular Outcomes,” *J. Am. Coll. Cardiol.*, vol. 68, no. 8, pp. 789–801, Aug. 2016.
- [196] L. Du *et al.*, “Determinants of Medication Adherence for Pulmonary Tuberculosis Patients During Continuation Phase in Dalian, Northeast China,” *Patient Prefer. Adherence*, vol. 14, pp. 1119–1128, 2020.
- [197] W. H. Polonsky and R. R. Henry, “Poor medication adherence in type 2 diabetes: recognizing the scope of the problem and its key contributors,” *Patient Prefer. Adherence*, vol. 10, pp. 1299–1306, Jul. 2016.
- [198] E. K. Buysman, F. Liu, M. Hammer, and J. Langer, “Impact of medication adherence and persistence on clinical and economic outcomes in patients with type 2 diabetes treated with liraglutide: a retrospective cohort study,” *Adv. Ther.*, vol. 32, no. 4, pp. 341–355, Apr. 2015.
- [199] T. Kennedy-Martin, K. S. Boye, and X. Peng, “Cost of medication adherence and persistence in type 2 diabetes mellitus: a literature review,” *Patient Prefer. Adherence*, vol. 11, pp. 1103–1117, Jun. 2017.
- [200] M. C. Roebuck, J. N. Liberman, M. Gemmill-Toyama, and T. A. Brennan, “Medication adherence leads to lower health care use and costs despite increased drug spending,” *Health Aff. (Millwood)*, vol. 30, no. 1, pp. 91–99, Jan. 2011.
- [201] S. B. Khojasteh, J. R. Villar, C. Chira, V. M. González, and E. de la Cal, “Improving fall detection using an on-wrist wearable accelerometer,” *Sensors (Switzerland)*, vol. 18, no. 5, May 2018.
- [202] A. Watanabe, H. Noguchi, M. Oe, H. Sanada, and T. Mori, “Development of a Plantar Load Estimation Algorithm for Evaluation of Forefoot Load of Diabetic Patients during Daily Walks Using a Foot Motion Sensor,” *J. Diabetes Res.*, vol. 2017, 2017.
- [203] Y. H. Lee *et al.*, “Wearable textile battery rechargeable by solar energy,” *Nano Lett.*, vol. 13, no. 11, pp. 5753–5761, Nov. 2013.
- [204] D. R. Beddiar, B. Nini, M. Sabokrou, and A. Hadid, “Vision-based human activity recognition: a survey,” *Multimed. Tools Appl.*, vol. 79, no. 41–42, pp. 30509–30555, Nov. 2020.
- [205] M. K. Seal, S. M. A. Noori Rahim Abadi, M. Mehrabi, and J. P. Meyer, “Machine learning classification of in-tube condensation flow patterns using visualization,” *Int. J. Multiph. Flow*, vol. 143, p. 103755, Oct. 2021.
- [206] S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, “Machine learning: a review of classification and combining techniques,” *Artif. Intell. Rev.* 2007 263, vol. 26, no.

- 3, pp. 159–190, Nov. 2007.
- [207] “Cost of Older Adult Falls | Fall Prevention | Injury Center | CDC.” [Online]. Available: <https://www.cdc.gov/falls/data/fall-cost.html>. [Accessed: 24-Oct-2022].
 - [208] C. S. Florence, G. Bergen, A. Atherly, E. Burns, J. Stevens, and C. Drake, “The Medical Costs of Fatal Falls and Fall Injuries among Older Adults,” *J. Am. Geriatr. Soc.*, vol. 66, no. 4, p. 693, Apr. 2018.
 - [209] M. van der Vlegel, J. A. Haagsma, A. J. L. M. Geraerds, L. de Munter, M. A. C. de Jongh, and S. Polinder, “Health care costs of injury in the older population: a prospective multicentre cohort study in the Netherlands,” *BMC Geriatr.* 2020 201, vol. 20, no. 1, pp. 1–11, Oct. 2020.
 - [210] S. Feller, H. Boeing, and T. Pischon, “Body-mass-index, taillenumfang und risiko für diabetes mellitus typ 2: Konsequenzen für den medizinischen alltag,” *Dtsch. Arztebl.*, vol. 107, no. 26, pp. 470–476, Jul. 2010.
 - [211] “Obesity Facts in America.” [Online]. Available: <https://www.healthline.com/health/obesity-facts#5.-Your-waist-size-increases-your-risk-for-diabetes>. [Accessed: 24-Oct-2022].
 - [212] A. M. Tabar, A. Keshavarz, and H. Aghajan, “Smart Home Care Network using Sensor Fusion and Distributed Vision-based Reasoning General Terms Algorithms,” 2006.
 - [213] P. Düking, S. Achtzehn, H.-C. Holmberg, and B. Sperlich, “Integrated Framework of Load Monitoring by a Combination of Smartphone Applications, Wearables and Point-of-Care Testing Provides Feedback that Allows Individual Responsive Adjustments to Activities of Daily Living,” *Sensors* 2018, Vol. 18, Page 1632, vol. 18, no. 5, p. 1632, May 2018.
 - [214] P. Düking, A. Hotho, H.-C. Holmberg, F. K. Fuss, and B. Sperlich, “Comparison of Non-Invasive Individual Monitoring of the Training and Health of Athletes with Commercially Available Wearable Technologies,” *Front. Physiol.*, vol. 0, no. MAR, p. 71, Mar. 2016.
 - [215] O’DonoghueJohn and HerbertJohn, “Data Management within mHealth Environments,” *J. Data Inf. Qual.*, vol. 4, no. 1, Oct. 2012.
 - [216] Saguna, A. Zaslavsky, and D. Chakraborty, “Building activity definitions to recognize complex activities using an online activity toolkit,” *Proc. - 2012 IEEE 13th Int. Conf. Mob. Data Manag. MDM 2012*, pp. 344–347, 2012.
 - [217] Saguna, A. Zaslavsky, and D. Chakraborty, “Recognizing concurrent and interleaved activities in social interactions,” *Proc. - IEEE 9th Int. Conf. Dependable, Auton. Secur. Comput. DASC 2011*, pp. 230–237, 2011.
 - [218] Saguna, “Inferring multiple activities of mobile users with activity algebra,” *Proc. - IEEE Int. Conf. Mob. Data Manag.*, vol. 2, pp. 23–26, 2011.
 - [219] Saguna, A. Zaslavsky, and D. Chakraborty, “Complex activity recognition using context-driven activity theory and activity signatures,” *ACM Trans. Comput. Interact.*, vol. 20, no. 6, Dec. 2013.
 - [220] A. K. Dey, G. D. Abowd, and D. Salber, “A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications,” https://doi.org/10.1207/S15327051HCII6234_02, vol. 16, no. 2–4, pp. 97–166, 2009.
 - [221] M. A. Alsheikh, A. Selim, D. Niyato, L. Doyle, S. Lin, and H.-P. Tan, “Deep

Activity Recognition Models with Triaxial Accelerometers.”

- [222] A. Fender *et al.*, “Two-axis temperature-insensitive accelerometer based on multicore fiber Bragg gratings,” *IEEE Sens. J.*, vol. 8, no. 7, pp. 1292–1298, Jul. 2008.
- [223] C. Olah, “Understanding LSTM Networks -- colah’s blog.” [Online]. Available: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>. [Accessed: 17-Oct-2022].
- [224] F. A. Gers, J. Schmidhuber, and F. Cummins, “Learning to forget: continual prediction with LSTM,” *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000.

APPENDIX A: Permission for Reprint

Publication reported at Footnote 1: Chrisogonas O. Odhiambo, Casey A. Cole, Alaleh Torkjazi, Homayoun Valafar. Proceedings of the 2019 International Conference on Computational Science and Computational Intelligence (CSCI), "*State Transition Modeling of the Smoking Behavior Using LSTM Recurrent Neural Networks*", Reprinted here with permission from the publisher as shown on the following website maintained by IEEE. It states that the work can be reposted anywhere without permission as long as the reader is directed to the IEEE copyright policy page: <https://www.ieee.org/publications/rights/copyright-policy.html>.



State Transition Modeling of the Smoking Behavior Using LSTM Recurrent Neural Networks

Conference Proceedings:

2019 International Conference on Computational Science and Computational Intelligence (CSCI)

Author: Chrisogonas O. Odhiambo

Publisher: IEEE

Date: December 2019

Copyright © 2019, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

- 1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
- 2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
- 3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

- 1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
- 2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
- 3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

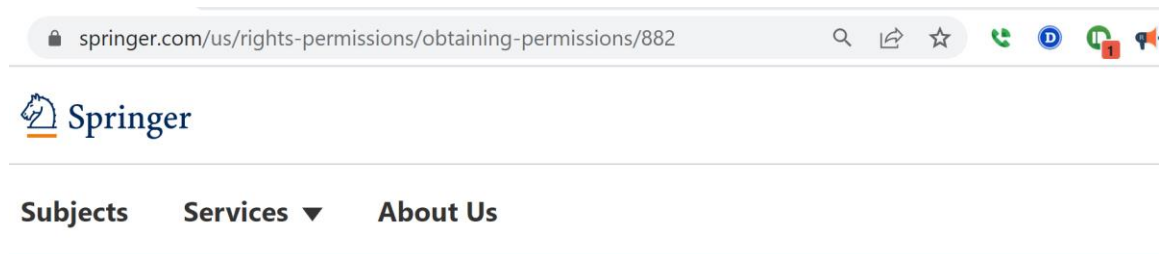
If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

Publication reported at Footnote 2: Chrisogonas O. Odhiambo, Pamela Wright, Cindy Corbett, Homayoun Valafar. Proceedings of the 2021 International Conference on Computational Science and Computational Intelligence CSCI), "*MedSensor: Medication Adherence Monitoring Using Neural Networks on Smartwatch Accelerometer Sensor Data*",

And

Publication reported at Footnote 3: Chrisogonas O. Odhiambo, Sanjoy Saha, Corby K. Martin, Homayoun Valafar. Proceedings of the 2021 International Conference on Computational Science and Computational Intelligence CSCI), "*Human Activity Recognition on Time Series Accelerometer Sensor Data using LSTM Recurrent Neural Networks*",

Reprinted here with permission from the publisher



Permissions

Get permission to reuse Springer Nature content

Springer Nature is partnered with the Copyright Clearance Center to meet our customers' licensing and permissions needs.

Author reuse

Please check the Copyright Transfer Statement (CTS) or Licence to Publish (LTP) that you have signed with Springer Nature to find further information about the reuse of your content.

Authors have the right to reuse their article's Version of Record, in whole or in part, in their own thesis. Additionally, they may reproduce and make available their thesis, including Springer Nature content, as required by their awarding academic institution. Authors must properly cite the published article in their thesis according to current citation standards. Material from: 'AUTHOR, TITLE, JOURNAL TITLE, published [YEAR], [publisher - as it appears on our copyright page]'

If you are any doubt about whether your intended re-use is covered, please contact journalpermissions@springernature.com for confirmation.

Publication reported at Footnote 4: Chrisogonas O. Odhiambo, Luke Ablonczy, Pamela J.

Wright, Cynthia F. Corbett, Sydney Reichardt, Homayoun Valafar. Submitted to Journal of

Medical Internet Research (JMIR) 2022, "*Detecting Medication Gestures using Machine*

Learning and Accelerometer Data Collected via Smartwatch Technology: A Feasibility Study",

Reprinted here with permission from the publisher

Articles in this section

How to add an author after publication

Can I order reprints of my article?

Is JMIR Cancer in PubMed / PubMed Central yet?

Is JMIR Rehabilitation & Assistive Technologies in PubMed / PubMed Central yet?

Why does my article not have a PMCID yet?

Can you give me permission to publish my article as part of my thesis or book?

Can you give me permission to publish my article as part of my thesis or book?



Editorial Director
16 days ago · Updated

Follow

Authors of articles published by JMIR Publications do not need our permission to use part or all of an article in their thesis or book. JMIR Publications' articles are published under a Creative Commons Attribution license (see copyright statement on each article), meaning that authors retain copyright ([What is a "Creative Commons License"?](#)). Authors can publish their article in a thesis, or as a book chapter, or anywhere else, as long as they adhere to the requirements of the Creative Commons Attribution license, which is basically to cite the original source and to state that it was published (and can be reproduced) under the terms of Creative Commons Attribution license 2.0 (or 4.0 as of June 2017).

Authors should however be aware that -- while they retain copyright and may be able to use their work as book chapter, etc, as long as they disclose that it has been published as journal article in a JMIR journal --, [duplicate publication in other journals may be considered scientific misconduct](#).

Related:

Publication reported at Footnote 5: Chrisogonas O. Odhiambo, Cynthia F. Corbett, Homayoun Valafar. To be submitted to Sensors (MDPI) 2022, "*Toward Concurrent Identification of Human Activities with a Single Unifying Neural Network Classification: First Step*"

MDPI Terms of Use URL: <https://www.mdpi.com/about/termsfuse>

mdpi.com/about/termsfuse

MDPI Journals Topics Information Author Services Initiatives About Sign In / Sign Up

Search for Articles: Title / Keyword Author / Affiliation All Journals All Article Types Search

Legal Information

Privacy Policy

Terms and Conditions

Terms of Use

Terms of Use

§ 1 These Terms of Use govern the use of the MDPI websites or any other MDPI online services you access. This includes any updates or releases thereof. By using our online services, you are legally bound by and hereby consent to our Terms of Use and Privacy Policy. These Terms of Use form a contract between MDPI AG, registered at St. Alban-Anlage 66, 4052 Basel, Switzerland ("MDPI") and you as the user ("User"). These Terms of Use shall be governed by and construed in accordance with Swiss law, applicable at the place of jurisdiction of MDPI in Basel, Switzerland.


§ 2 Unless otherwise stated, the website and affiliated online services are the property of MDPI and the copyright of the website belongs to MDPI or its licensors. You may not copy, hack or modify the website or online services, or falsely claim that some other site is associated with MDPI. MDPI is a registered brand protected by the Swiss Federal Institute of Intellectual Property.

§ 3 Unless otherwise stated, articles published on the MDPI websites are labeled as "Open Access" and licensed by the respective authors in accordance with the Creative Commons Attribution (CC-BY) license. Within the limitations mentioned in §4 of these Terms of Use, the "Open Access" license allows for unlimited distribution and reuse as long as appropriate credit is given to the original source and any changes made compared to the original are indicated.

Creative Commons URL: <https://creativecommons.org/licenses/by/4.0/>

creativecommons.org/licenses/by/4.0/ English

This page is available in the following languages: English



Attribution 4.0 International (CC BY 4.0)


This is a human-readable summary of (and not a substitute for) the [license](#). [Disclaimer](#).

You are free to:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.



Under the following terms:



Attribution — You must give [appropriate credit](#), provide a link to the license, and [indicate if changes were made](#). You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions — You may not apply legal terms or [technological measures](#) that legally restrict others from doing anything the license permits.

Notices:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable [exception or limitation](#).

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as [publicity, privacy, or moral rights](#) may limit how you use the material.