

MOBILE HEALTH ANALYTICS FOR SENIOR CARE:
A DATA MINING AND DEEP LEARNING APPROACH

by

Shuo Yu

Copyright © Shuo Yu 2019

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF MANAGEMENT INFORMATION SYSTEMS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2019


THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by *Shuo Yu*, titled *Mobile Health Analytics for Senior Care: A Data Mining and Deep Learning Approach* and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.



Hsinchun Chen

Date: 05/02/2019



Jay F. Nunamaker, Jr.

Date: 05/02/2019



Susan A. Brown

Date: 05/02/2019

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.



Dissertation Director: Hsinchun Chen

Date: 05/02/2019

**Regents' Professor, Thomas R. Brown Chair in Management and Technology
Department of Management Information Systems**

ACKNOWLEDGEMENTS

I hold my utmost appreciation to my dissertation committee members, Drs. Hsinchun Chen, Jay F. Nunamaker, Jr., and Susan A. Brown, for their guidance, encouragement, and inspiration. I am especially grateful to my advisor, Dr. Hsinchun Chen, for his support and feedback during my senior year at Tsinghua University in China and throughout my five-year doctoral study at the University of Arizona. The lessons I learned from him about respect, attitude, vision, and balance will have a persistent influence on my life.

I sincerely thank my colleagues at the Artificial Intelligence Lab for their generous assistance in various projects. Special thanks to my Ph.D. peers and good friends, Hongyi Zhu and Sagar Samtani, for their academic communication and support. I thank MIS department and AI Lab staff members, especially Cathy Larson and Riley McIssac, for their logistic assistance. I also owe my deepest gratitude to my dear parents, Hongbin Yu and Qiuxiang Song, and my loving wife, Luozhe Lu, for supporting me through this special phase of life.

I gratefully acknowledge that the studies in this dissertation have been partly supported by grants from the National Science Foundation (SES-1314631, DUE-1303362, and IIP-1622788).

DEDICATION

This dissertation is dedicated to my family.

TABLE OF CONTENTS

LIST OF FIGURES	8
LIST OF TABLES	9
ABSTRACT	10
1. INTRODUCTION	11
2. ESSAY I: HIDDEN MARKOV MODEL-BASED FALL DETECTION WITH MOTION SENSOR ORIENTATION CALIBRATION	17
2.1. Introduction	17
2.2. Literature Review	20
<i>2.2.1. Fall Detection: Features and Systems</i>	20
<i>2.2.2. Hidden Markov Models for Health Applications</i>	24
<i>2.2.3. Sensor Orientation Calibration</i>	28
2.3. Research Gaps and Questions	31
2.4. Research Design	32
<i>2.4.1. Data Collection</i>	33
<i>2.4.2. Sensor Orientation Calibration</i>	36
<i>2.4.3. Hidden Markov Model Design</i>	37
<i>2.4.4. Evaluation</i>	40
2.5. Results and Discussion	42
<i>2.5.1. Dataset: POOL</i>	42
<i>2.5.2. Dataset: IND</i>	45
<i>2.5.3. Dataset: FAR</i>	46
2.6. Conclusion and Future Directions	47
3. ESSAY II: RISK PREDICTION FOR MOBILE HEALTH: A DEEP LEARNING APPROACH	49
3.1. Introduction and Essay Structure	49
3.2. Literature Review	52
<i>3.2.1. IS Works on Health and Mobile Topics</i>	52
<i>3.2.2. Motion Sensor-Based Risk Prediction: Scenarios and Approaches</i>	55
<i>3.2.3. Theoretical Support in Deep Learning</i>	57
<i>3.2.4. Convolutional Neural Networks in Health Contexts</i>	59
3.3. Research Gaps and Questions	62
3.4. Research Design	63
<i>3.4.1. Data Collection and Preprocessing</i>	63

3.4.2. 2D-hetero CNN Architecture	66
3.4.3. Evaluation	69
3.5. Experiment Results	71
3.5.1. Classic Machine Learning Algorithms	71
3.5.2. Alternative Neural Networks.....	71
3.5.3. Ablation Analyses.....	73
3.5.4. Robustness Checks	75
3.5.5. Case Study.....	77
3.6. Discussion and Conclusions	78
3.6.1. Contributions to IS Knowledge Base	79
3.6.2. Practical Implications.....	82
3.6.3. Limitations and Future Research.....	83
4. ESSAY III: MOTION SENSOR-BASED HEALTH CONDITION RISK AND SEVERITY ASSESSMENT: DEEP MULTISOURCE MULTITASK LEARNING FOR MOBILE HEALTH ANALYTICS.....	85
4.1. Introduction and Essay Structure	85
4.2. Literature Review	89
4.2.1. IS Research on Health Information Technology.....	89
4.2.2. Computational Design Science Guidelines.....	91
4.2.3. Motion Sensor-Based Health Condition Risk and Severity Assessment	92
4.2.4. Computational Analytical Model: A Deep Learning Approach	94
4.2.5. Deep Multitask Learning and Deep Multisource Learning.....	97
4.3. Research Gaps and Questions	100
4.4. Research Design	101
4.4.1. Data Collection and Preprocessing.....	102
4.4.2. The Deep Multisource Multitask Learning for Mobile Health Analytics (DMML-MHA) Framework.....	106
4.4.3. Evaluation	111
4.5. Experiment Results.....	114
4.5.1. Single Feature-Based Models	114
4.5.2. Classic Single Task Machine Learning Models.....	115
4.5.3. Classic Multitask Machine Learning Models	116
4.5.4. Alternative Deep Learning Models	117
4.6. Discussion and Conclusions	118
4.6.1. Contributions to IS Knowledge Base	119

4.6.2. <i>Practical Implications</i>	120
4.6.3. <i>Limitations and Future Research</i>	121
5. ESSAY IV: MOTION SENSOR-BASED HEALTH PROFILING FOR SENIOR CARE: ADAPTIVE TIME-AWARE CONVOLUTIONAL LONG SHORT TERM MEMORY (ATCLSTM)	123
5.1. Introduction and Essay Structure	123
5.2. Literature Review	125
5.2.1. <i>Motion Sensor-Based Health Profiling</i>	126
5.2.2. <i>Deep Learning on IoT Streaming Data Analytics</i>	129
5.2.3. <i>Deep Learning for Temporal Health Profiling</i>	132
5.3. Research Gaps and Questions	136
5.4. Research Design	137
5.4.1. <i>Data Collection</i>	138
5.4.2. <i>Data Preprocessing</i>	139
5.4.3. <i>Adaptive Time-aware Convolutional Long Short Term Memory (ATCLSTM)</i> ...	140
5.4.4. <i>Evaluation</i>	145
5.5. Experiment Results	148
5.5.1. <i>Single Feature-Based Models</i>	148
5.5.2. <i>Classic Machine Learning Models</i>	149
5.5.3. <i>Alternative Deep Learning Models</i>	150
5.6. Discussion and Conclusions	151
6. CONCLUSIONS AND FUTURE DIRECTIONS.....	153
6.1. Research Contributions	153
6.2. Future Research Directions	155
7. REFERENCES	158

LIST OF FIGURES

Figure 2.1: Illustration of Motor States Derived from a Simulated Fall Observation	25
Figure 2.2: Signal Series for Simulated Falls Collected from Two Sensor Orientations	29
Figure 2.3: Differences between the Two Signal Series Shown in Figure 2.2	29
Figure 2.4: Research design for our HMM-based fall detection system	33
Figure 2.5: (Left) Sensor Shape and Size; (Right) Sensor Locations (Frontal View)	34
Figure 2.6: Comparison of SPEC and GEN HMM Designs.....	38
Figure 2.7: F-measure Comparison on POOL Dataset: (Left) SPEC; (Right) GEN	43
Figure 2.8: F-measure across Locations on IND Dataset: (Left) SPEC; (Right) GEN	46
Figure 3.1: (Left) 1D CNN; (Middle) 2D Homogeneous CNN; (Right) 2D Heterogeneous CNN	62
Figure 3.2: Research Design for Our Deep Learning Architecture: 2D-hetero CNN	63
Figure 3.3: (Left) Sensor Shape and Size; (Right) Sensor Locations (Frontal View)	64
Figure 3.4: 2D-hetero CNN architecture	66
Figure 3.5: Two Examples Misclassified by Benchmark Models (Chest Location)	77
Figure 4.1: (Left) Single Task Learning; (Right) Multitask Learning.....	98
Figure 4.2: (Left) Single Source Learning; (Right) Multisource Learning	99
Figure 4.3: Research Design of Our Proposed the Deep Multisource Multitask Learning for Mobile Health Analytics (DMML-MHA) Framework.....	102
Figure 4.4: (Left) Sensor Shape and Size; (Right) Sensor Locations (Frontal View)	103
Figure 4.5: An Abstract DMML-MHA Framework	106
Figure 4.6: The DMML-MHA Instantiation for the Internal SilverLink Dataset.....	107
Figure 4.7: The DMML-MHA Instantiation for the External mPower Dataset	108
Figure 5.1: (Left) LSTM Cell; (Right) TLSTM Cell.....	134
Figure 5.2: Research Design for Motion Sensor-Based Health Profiling.....	138
Figure 5.3: Adaptive Time-aware Convolutional Long Short Term Memory (ATCLSTM)	140
Figure 5.4: Convolutional Module.....	141
Figure 5.5: Adaptive TLSTM Module.....	143
Figure 5.6: (Left) Uniform Time Decay Factors in TLSTM; (Right) Adaptive Time Decay Factors in Adaptive TLSTM.....	144

LIST OF TABLES

Table 1.1: Overview of the Four Essays in the Dissertation.....	14
Table 2.1: Precision, Recall, and F-measures for POOL Dataset Evaluation.....	44
Table 2.2: Accuracies of the Activity Categories in the POOL Dataset.....	45
Table 2.3: Precision, Recall, and F-measures for the FAR Dataset.....	47
Table 3.1: Experiment Results for Classic Machine Learning Algorithms	71
Table 3.2: Experiment Results for Alternative Neural Networks.....	72
Table 3.3: Experiment Results for Sensitivity Analyses	74
Table 3.4: Experiment Results for Robustness Checks	76
Table 4.1: Precision, Recall, F-measure, and RMSE for Single Feature-Based Models.....	114
Table 4.2: Precision, Recall, F-measure, and RMSE for Classic Single Task Machine Learning Models	115
Table 4.3: Precision, Recall, F-measure, and RMSE for Classic Multitask Machine Learning Models.....	116
Table 4.4: Precision, Recall, F-measure, and RMSE for Alternative Deep Learning Models	117
Table 5.1: Selected Recent Research in Motion Sensor-Based Health Profiling	127
Table 5.2: Selected Major Deep Learning Research in a Health Context	130
Table 5.3: Selected Deep Learning Literature for Temporal Health Profiling.....	133

ABSTRACT

Senior citizens confront numerous challenges to their independent living, including chronic physical health conditions and a decline in mobility. With the advancement of mobile sensing technologies, medical professionals and information systems (IS) researchers have sought to apply data mining techniques to provide precise, prompt, and personalized assessment for falls and health conditions including Parkinson's disease. Given the societal importance of senior care, my dissertation aims to address the following four research questions: (1) how can we promptly detect senior citizens' adverse events, e.g., falls, to alleviate consequences, (2) how can we precisely assess senior citizens' health risks, e.g., fall risks, to provide proper interventions, (3) how can we leverage multiple data sources and assess senior citizens' health risks in a more holistic manner, and (4) how can we profile senior citizens' long-term health progression for more personalized care.

This dissertation presents four essays to answer these questions. The essays develop state-of-the-art data mining and deep learning techniques to address selected senior care inquiries. The first essay focuses on a novel hidden Markov model with sensor orientation calibration to detect falls. The second essay presents a two-dimensional heterogeneous convolutional neural network to precisely assess fall risks. The third essay leverages deep multisource multitask learning to achieve sensor fusion and assess multiple health risks and disease severities. The final essay develops an adaptive time-aware convolutional long short term memory model that enables long-term health profiling with time irregularities. Presented frameworks, systems, and design principles not only advance mobile health analytics and deep learning methodologies, but also guide future computational design science research in IS.

1. INTRODUCTION

People are enjoying longer life due to healthy lifestyles and advances in medicine and health services. Meanwhile, population aging has also been a growing concern, with 47.8 million US citizens (14.9 percent of the total population) aged 65 or more in 2015 (U.S. Census Bureau 2017). Senior citizens face many challenges to their independent living, including a decline in mobility or cognition and having chronic physical health conditions. Such conditions may include frailty, diabetes, Parkinson's disease, dementia, stroke, falls, etc. Among these, falls are one of the most severe threats that affect senior citizens' lives, as approximately 30 percent of adults aged 65 and over fall each year (Bergen et al. 2016). Falls threaten senior citizens' living both physically and psychologically. The estimated costs of fatal and nonfatal falls combined totaled approximately \$50.0 billion (Florence et al. 2018). Among the chronic conditions, Parkinson's disease (PD) is the second most common neurodegenerative disorder in the United States, costing \$15.5 billion per year (Gooch et al. 2017).

To better manage chronic conditions and corresponding adverse events, two general approaches have been explored in literature (Howcroft et al. 2013). The first approach detects adverse events after they happen as soon as possible to call for prompt assistances from caregivers or hospitals, such that more severe consequences could be prevented. For instance, many fallers are unable to get up again without assistance, leading to a "long lie" situation. This prolonged immobilization can lead to serious medical complications such as hypothermia, dehydration, and skin injury (Rubenstein and Josephson 2002; Tinetti et al. 1993). Timely detection of falls with quick assistances reduces the risk of hospitalization by 26 percent and the risk of death by 80 percent (Scheffer et al. 2008). The second approach assesses the risk or severity of health conditions, aiming to provide interventions to prevent future adverse events

or alleviate the progression of health conditions. Such interventions include exercise, assistive devices, and increased surveillance and care by caregivers (Weinstein and Booth 2006). This approach is perhaps of greater value than the first approach as it provides a view of senior citizens' health status and can potentially eliminate the damage caused by the deterioration of existing health conditions or occurrences of future adverse events.

Traditionally, clinical assessments for falls and other mobility-related conditions are conducted via survey-based numerical evaluations of risk factors and performance of timed mobility tests. Survey-based evaluations include Fried's Frailty Criteria (Fried et al. 2001), STRATIFY Score (Oliver et al. 1997), Physiological Profile Assessment (PPA) (Lord et al. 2003), and Tinetti Performance Oriented Mobility Assessment (POMA) (Tinetti 1986), among others. Widely accepted timed mobility tests include 10-meter ground walking, timed up and go (TUG), sit-to-stand transitions (STS), alternating step test (AST), etc. However, such assessments confront the following limitations. Survey-based evaluations rely on interviews and physicians' subjective assessment. Timed mobility tests are associated with simple thresholds on the completion time, which oversimplifies the analysis of body motion. To address the above concerns, gait laboratories with specialized equipment (e.g., motion sensors, ground force plates, and electromyographic signals (Nelson-Wong et al. 2012)) provide objective and quantitative measures for risk prediction. However, due to high costs and time-consuming set-ups, it is less practical to integrate such gait lab experiments into typical clinic routines.

With the advancement of mobile health and sensing technologies, miniature wearable wireless motion sensors have emerged as a proxy for gait labs to efficiently capture and analyze quantitative mobility data for advanced adverse event detection and fine-grained condition risk

assessment (Wang et al. 2017; Greene et al. 2017). Motion sensors are attached to a senior citizen's body for a short period of time (5 to 10 minutes) to collect data from mobility tests. A data analytics engine processes motion data and provides risk prediction for a single or multiple health conditions. This can help the physician provide more timely interventions to alleviate or even eliminate the condition threats (Weinstein and Booth 2006), which can thus improve the senior citizen's quality of life.

However, developing effective approaches for adverse event detection and risk assessment is still an emerging domain under exploration. The immense amount of data generated by sensors (millions of data points per day) require advanced processing and analytics methods to provide precise, prompt, and personalized senior care. Despite these challenges, mobile health analytics powered by sensor-based technologies has been highlighted as a key area for business intelligence and analytics in Information Systems (IS) literature (Chen et al. 2012). As such, I develop four essays utilizing wearable sensor data in conjunction with state-of-the-art data mining and deep learning techniques for advanced senior care. Table 1.1 outlines the key aspects of each essay.

Table 1.1: Overview of the Four Essays in the Dissertation

Essay	Research Question	Methodology	Algorithms/Models
I	How to promptly detect senior citizens' adverse events to alleviate consequences?	Data Mining; Stochastic Process	Hidden Markov Model with Sensor Orientation Calibration (HMM-CAL)
II	How to precisely assess senior citizens' health risks to provide proper interventions?	Data Mining; Deep Learning	Two-Dimensional Heterogeneous Convolutional Neural Network (2D-hetero CNN)
III	How to leverage multiple data sources and assess senior citizens' health risks in a more holistic manner?	Data Mining; Deep Learning	Deep Multitask Multisource Learning with Mobile Health Analytics (DMML-MHA)
IV	How to profile senior citizens' long-term health progression for more personalized care?	Data Mining; Deep Learning	Adaptive Time-aware Convolutional Long Short Term Memory (ATCLSTM)

The first essay focuses on the prompt detection of adverse events for senior care. This essay presents a principled hidden Markov model (HMM)-based fall detection system (HMM-CAL) to detect falls automatically using a single motion sensor for real-life home monitoring scenarios (Yu et al. 2018). A new representation for acceleration signals is proposed in HMMs to avoid feature engineering and developed a sensor orientation calibration algorithm to resolve sensor misplacement issues (misplaced sensor location and misaligned sensor orientation) in real-world scenarios. To evaluate the system, a dataset is collected from experiments of simulated falls and normal activities, and another dataset is acquired from a real-world fall repository (FARSEEING). The HMM-CAL system is able to precisely detect falls in real-life home scenarios with a reasonably low false alarm rate.

Essay II takes a slightly different angle on senior care, aiming to predict fall risks in a long term. With grounding in motion sensor technologies and deep learning methodologies, a Two-Dimensional Heterogeneous Convolutional Neural Network (2D-hetero CNN) architecture with cross-axial and cross-locational convolutions is proposed to optimize the

model effectiveness in a mobile health analytics context. 10-meter ground walking data is collected at a neurology clinic to evaluate our model for fall risk prediction. Five sensors are attached to 52 patients with Parkinson’s disease to collect motion data throughout the walking test. Extensive sensitivity analyses and robustness checks are conducted on the proposed 2D-hetero CNN model. A case study is discussed to illustrate the advantage of applying the model for precise risk prediction. The design of our 2D-hetero CNN can provide insights into future deep learning research using motion sensor data for health predictive analytics and many other related topics.

Essay III extends Essay II’s goal by combining multiple data sources to assess multiple condition risks and severities in a more holistic manner. A Multisource Multitask Learning for Mobile Health Analytics (DMML-MHA) framework is proposed for transferable health condition risk and severity assessment. Source-specific, general, and task-specific deep learning layers are designed to extract salient features from motion sensor signals and alleviate model overfitting to improve assessment results. An internal clinical experiment dataset and an external publicly available dataset are collected to rigorously evaluate our proposed framework against state-of-the-art benchmark models. Practical implications to senior citizens, their families, and health professionals are discussed. The proposed DMML-MHA framework can contribute to improved life quality for senior citizens and provide design guidelines for future IS research in mobile health analytics.

Essay IV is motivated by the need for long-term health profiling to provide more personalized senior care, especially for modeling multiple test cases taken at variable time intervals. A novel deep learning model, Adaptive Time-aware Convolutional Long Short Term Memory (ATCLSTM), is proposed to model motion sensor data with time irregularities. A

convolutional module is developed to automatically learn features from sensor signals, and an adaptive time-aware long short term memory module to model time irregularities for temporal sequences. A publicly available dataset is collected to evaluate the proposed model. The experimental evaluations demonstrate that the ATCLSTM attains an improved predictive performance when compared with alternative benchmark models for health profiling. More importantly, the ATCLSTM not only provides more accurate health profiling for early disease detection and intervention with motion sensor data, but also can be generalized to model any time series data with time irregularities, such as Electronic Health Records (EHR), social media timeline, etc. The generalizability of the proposed ATCLSTM is a fundamental methodological contribution to the deep learning, predictive health analytics, and the IS community.

2. ESSAY I: HIDDEN MARKOV MODEL-BASED FALL DETECTION WITH MOTION SENSOR ORIENTATION CALIBRATION

2.1. Introduction

Advances in medicine and public health services have significantly improved human life expectancy. With an average life length of 79.3 years in the United States (World Health Organization 2016a), however, senior citizens face many challenges to their independent living, including a decline in mobility or cognition and chronic physical health conditions that compromise their ability to maintain their independence. Among those conditions, falls are a major cause of fatal injury and create a serious obstruction to independent living. According to a report from the World Health Organization (2008), approximately 28 to 35 percent of people aged 65 and over fall each year, and this proportion increases to 32 to 42 percent for those over 70 years of age. Falls threaten senior citizens' autonomous living both physically and psychologically. On one hand, in addition to the direct injuries from the fall, many fallers are unable to get up again without assistance, leading to a "long lie" situation. This subsequent situation can lead to hypothermia, dehydration, bronchopneumonia, and pressure sores, which increases the severity of falls (Rubenstein and Josephson 2002; Tinetti 1993). More than 20 percent of patients admitted to hospitals because of a fall were on the ground for an hour or more, and their morbidity rates within 6 months were very high. On the other hand, senior citizens with a history of falling live in fear of future falls and may be at increased risk of falling again. Such fear has been associated with negative consequences such as avoidance of activities, less physical activity, falling, depression, decreased social contact, and lower quality of life (Scheffer et al. 2008). According to a report from the Centers for Disease Control and Prevention (CDC) (Burns et al. 2016), there were 24,190 fatal falls and 3.2 million non-fatal

fall injuries in 2012; the average cost was \$26,340 for a fatal fall and \$9,780 for a non-fatal fall. In the US, fatal falls were projected to cost \$637.2 million and non-fatal falls \$31.3 billion in 2015.

The consequences of falls can be severe. Nevertheless, obtaining quick assistance after a fall reduces the risk of hospitalization by 26 percent and the risk of death by 80 percent (Scheffer et al. 2008). Timely help from caregivers not only provides first aid for direct injuries, but also alleviates long lie situations. Senior citizens living alone require supportive devices to request prompt care after falls. Two current solutions for home care systems are SOS buttons and sensor-based automatic fall detection. SOS buttons are active emergency systems. They are small electronic devices carried by senior citizens that are connected to care giving facilities. When an emergency occurs, senior citizens can press the SOS button to notify the care giving facility, and caregivers will be dispatched to the house. On the other hand, sensor-based fall detection systems provide active and unobtrusive emergency support. They leverage data collected by sensors set in the senior citizen's house or attached to the senior citizen's body. When a fall event is identified by the underlying fall detection algorithm, the system notifies the care giving facility on behalf of the senior citizen, even if the wearer becomes unconscious or hurts his/her hands after the fall. This feature makes sensor-based automatic fall detection systems more proactive than SOS buttons. According to market research (Persistence Market Research 2016), the global market for fall detection systems accounted for \$321.3 million in 2014 and is expected to increase to \$448.1 million by 2021.

A variety of active sensors for fall detection systems have been studied in past research, including cameras (video/Kinect) (Stone and Skubic 2015), radars (Su et al. 2015), pressure sensors (Daher et al. 2016), and wearable motion sensors (Bourke et al. 2016; Lee et al. 2015).

Among those options, cameras, radars, and pressure sensors require special room settings, and falls occurring in rooms without the devices cannot be detected. Cameras bring further privacy concerns. In contrast, miniature motion sensors do not face the constraints of room settings. They are unobtrusive to carry as a necklace, keychain, or simply put in a pocket. In terms of a real-life home environment, motion sensor-based systems are preferred for senior citizens living independently. Although there are studies on motion sensor-based fall detection (Bourke et al. 2016; Lee et al. 2015; Kau and Chen 2015; Shen et al. 2015; Özdemir and Barshan 2014), there is little research about how the system interacts with the uncertainty of the real-life environment. In realistic scenarios, multi-sensor systems are cumbersome for users and reduce their applicability. Furthermore, past studies mostly focused on fixed sensor placement (location and orientation). For instance, the sensor was always attached to the chest, and the y-axis was always aligned with gravity. However, controlling the actual usage of fall detection systems is challenging. Failing to place and orient the sensor properly can lead to dramatically worse results.

To resolve such issues, we propose a hidden Markov model (HMM)-based fall detection system with motion sensor orientation calibration (HMM-CAL). We use tri-axial acceleration signals to describe the direction and magnitude of human motion patterns. We develop an orientation calibration algorithm to compensate for errors introduced by inconsistent sensor placement, which are common in real-world scenarios. The main novelty of this work is as follows. First, we propose a new representation for acceleration signals and HMMs to avoid feature engineering. Gaussian distributions are selected as observation distributions to model multi-dimensional acceleration signals. Second, we develop an

orientation calibration algorithm to resolve the misplacement issues in real-life contexts, which significantly improved the performance of our fall detection system.

2.2. Literature Review

2.2.1. Fall Detection: Features and Systems

Sensors and mobile health technologies have been widely applied in various health contexts. For instance, Abdullah et al. (2016) proposed an approach to automatically assess the Social Rhythm Metric (SRM) in bipolar disorder based on passively collecting sensor data from smartphones, including accelerometer, microphone, location, and communication information, which was used to infer behavioral and contextual patterns. Heintzman and Kleinberg (2016) collected glucose levels, insulin dosing, physical activity, and sleep data to infer the causal relationships between other variables versus glucose levels for type I diabetes patients. Neinstein et al. (2015) discussed creating a cloud-based platform to integrate data from diabetes devices for type I diabetes patients, including insulin pumps, continuous glucose monitors, fitness trackers, etc. Jonas et al. (2016) designed a smartphone application to provide instructions and analysis of the Congo Red Dot (CRD) test, showing promise as a diagnostic and prognostic tool for preeclampsia. In addition, motion sensor-based fall detection has been an active research area, as detecting falls and providing prompt assistance is important for senior citizens living independently.

Based on the definition from the World Health Organization, a fall is an event which results in a person coming to rest inadvertently on the ground or floor or other lower level (WHO 2016b). Fall events are closely related to various motor and cognitive disorders, including Parkinson's disease, stroke, dementia, frailty, and diabetes. The goal of fall detection systems is to accurately detect falls when they occur. There are special characteristics

distinguishing fall events from normal activities of daily living (ADL). Three phases occur sequentially during a fall event: collapse, impact, and inactivity. In the collapse phase, the faller accelerates towards the floor. In the impact phase a shock occurs between the faller and the floor. In the inactivity phase, the faller lies on the floor with no signs of activity. Environmental and wearable sensors have been utilized to detect indoor falls, including cameras (Kinect) (Stone and Skubic 2015), radars (Su et al. 2015), pressure sensors (Daher et al. 2016), accelerometers (Lee et al. 2015), gyroscopes, and magnetometers. Several studies combined multiple types of motion sensors to provide inference (Bourke et al. 2016; Shen et al. 2015; Kau and Chen 2015; Özdemir and Barshan 2014). Gyroscopes and magnetometers were typically used with accelerometers to provide additional information. As discussed above, environmental sensor-based fall detection has a heavy overhead on device deployment and limited area of usage, which is less preferred by home-dwelling senior citizens. Wearable sensors such as gyroscopes and magnetometers are difficult to design and use a lot of power, leading to frequent battery recharging (recharge every day or every other day). An ideal fall detection device would use accelerometers that allow the sensor to be continuously used for weeks to months, alleviating the users' burden of battery recharging.

Features: General and Fall-Specific

Two categories of features were extracted from the collected signals, including general signal processing features and fall-specific features. General signal processing features are those features widely used in many signal processing tasks. Özdemir and Barshan (2014) extracted time domain features (mean, standard deviation, zero crossing rate, etc.) from six accelerometers, gyroscopes, and magnetometers. Su et al. (2015) extracted frequency domain

features (coefficients of discrete Fourier transform, wavelet transform) from Doppler radar signals. Fall-specific features are those features aimed at recognizing the three specific phases of a fall event, namely collapse, impact, and inactivity. The collapse phase was identified by a low acceleration magnitude (loss of weight) and a high vertical velocity (Bourke et al. 2016; Lee et al. 2015). The impact phase was identified by a high acceleration magnitude indicating a shock (Bourke et al. 2016; Lee et al. 2015; Kau and Chen 2015). The inactivity phase was identified by a large rotation angle, indicating a posture change (lying on the floor) (Bourke et al. 2016; Stone and Skubic 2015).

Systems: Threshold-Based and Classifier-Based

Threshold-based (rule-based) and classifier-based fall detection systems were constructed based on those features. Threshold-based systems create a set of thresholds for the features for fall inference. For instance, one prevalent approach for fall detection is to set a threshold for the maximum acceleration magnitude (Bourke et al. 2007). This approach has a relatively good distinguishability for less intensive activities (standing, sitting) versus fall events, but triggers more false alarms for more intensive ones (sit down, lie down). Daher et al. (2016) set thresholds for floor pressure and acceleration sensors. This approach has difficulty distinguishing human falls from object drops. Other studies aimed at recognizing multiple phases in fall events (Lee et al. 2015; Kau and Chen 2015). However, Kangas et al. (2008) found that involving more phases does not consistently improve fall detection results. The accuracies of those threshold-based systems vary on sensor locations and the number of rules. Furthermore, most of those systems suffer from the following two design deficiencies. First, they typically did not involve a training set/test set separation when delivering their

parameter thresholds. Namely, they delivered the thresholds and tested the effectiveness of the thresholds on the same dataset. This may lead to the well-known over-fitting issue, which reduces the actual prediction power. Second, the values of thresholds are location-specific, and the effectiveness of each type of threshold (collapse, impact, posture change, etc.) is relevant to the location. For instance, the impact threshold for the waist is different from that for the chest; impact and posture change are the most effective indicators for the waist, whereas impact and collapse are best for the chest. Those limitations greatly affect the performance of fall detection systems in the real world, in that sensor locations may vary from time to time. In addition, threshold-based systems are prone to false alarms. In practice, Bagalà et al. (2012) applied several threshold-based systems on a real-life fall dataset, reporting that the detection accuracies were considerably lower (57 to 83 percent), and the number of false alarms generated by the algorithms during one-day monitoring of three representative fallers ranged from 3 to 85. This fact supports our idea that *ad hoc* threshold-based systems are oversimplified for valid and effective fall detection.

Classifier-based systems leverage machine learning algorithms for more accurate fall detection. Bourke et al. (2016) constructed a decision tree for collapse, impact, and posture change features collected by accelerometers and gyroscopes attached at the waist. Stone and Skubic (2015) used an ensemble of decision trees for collapse and posture change features from Kinect sensors. Su et al. (2015) applied k-nearest neighbors (kNN) for time-frequency features from wavelet transformations on Doppler radars. Özdemir and Barshan (2014) tested various machine learning algorithms, including kNN, support vector machines (SVM), artificial neural networks (ANN), and dynamic time warping (DTW), and concluded that kNN provided the best performance. Those classifier-based systems showed better fall detection

results. However, most of them were conducted in lab experiments with some assumptions that may not hold in real-life contexts. First, past studies generally assumed that the sensor placement is fixed, meaning unvarying sensor location and orientation. In real-world scenarios, real users may not consistently wear the sensor in the same place, and it is even more difficult to maintain sensor orientation (e.g., outwards vs. inwards, upwards vs. downwards, etc.). Second, most studies conducted controlled experiments to collect simulated (acted) fall events by stunt actors or volunteers. As Bourke et al. (2016) argued, “less than 7% of studies used real falls recorded from elderly people.” There may be large discrepancies between real-life falls and simulated falls, and the generalizability of lab experiment results still needs confirmation from the field. Third, most prior studies on fall detection did not take the temporal factor of fall events into account. The extracted features provide a one-shot description of the events. However, fall events consist of three distinct phases in a sequential order. Missing or changing the order of any of those phases may lead to different scenarios. For instance, a missing inactivity phase may indicate that the fall is of no or minimal damage. Therefore, a temporal pattern recognition model may be needed to reduce the false alarms in real-life scenarios.

2.2.2. Hidden Markov Models for Health Applications

A hidden Markov model (HMM) is a Markov process with unobserved (hidden) states, plus observed (visible) outputs. A first-order HMM involves a double random process in which each state has a probability distribution over the following states and a distribution over the possible outputs. The states are those underlying attributes in the system that cannot be directly observed. The outputs are realizations of those states that can be observed. HMMs provide an

approach to infer the dynamics of the hidden states (e.g., human motor states) based on the information of observable outputs (e.g., acceleration signals). They have been applied in a wide range of signal pattern recognition tasks, such as speech and handwriting recognition, as well as health-related tasks. Using gait data (acceleration signals), Mannini et al. (2015) applied HMMs for gait segmentation for hemiparetic and Huntington’s disease patients. Khorasani and Daliri (2014) applied HMMs to separate Parkinson’s patients from healthy subjects. Li et al. (2015) and Trabelsi et al. (2013) applied HMMs to recognize activities of daily living (standing, sitting, lying, etc.). HMMs were also used for non-gait signal patterns. Akhbari et al. (2016) performed analysis on ECG segmentation and fiducial point extraction. For fall detection systems, human motor states (e.g., falling state, impact state, lying down state, etc.) can be modeled based on observed acceleration signal series. Each state is characterized as a specific pattern of acceleration signals. Figure 2.1 gives an intuitive illustration of how the states may be derived from human acceleration signals. Falls and normal activities can be modeled as transitions of those states.

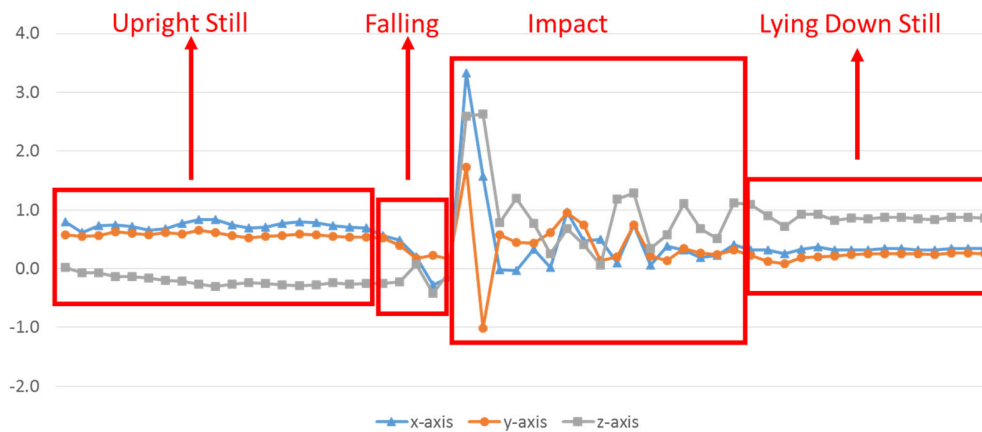


Figure 2.1: Illustration of Motor States Derived from a Simulated Fall Observation

Hidden Markov Model Designs

HMMs have been applied on fall detection systems using cameras (Hagui et al. 2015; Dai et al. 2013; Jiang et al. 2013), radars (Wu et al. 2013), and wearable sensors (Khan et al. 2014; Lim et al. 2014; Tong et al. 2013). After training the parameters of HMMs, we can calculate the likelihood for the models to generate a certain series of signals and make inference accordingly. Various HMM designs were proposed regarding the activities to recognize. Tong et al. (2013) and Jiang et al. (2013) designed a single model for detecting fall events, based on similarities between incoming data and known fall signals. Khan et al. (2014) designed a single model for detecting normal activities. It alerts falls based on dissimilarities between incoming data and known fall signals, which is essentially anomaly detection. These HMM designs aim to recognize generalized characteristics across multiple types of fall events and/or normal activities. On the other hand, Hagui et al. (2015), Lim et al. (2014), Wu et al. (2013), and Dai et al. (2013) designed multiple models for sub-categories of fall events and normal activities. These generate fall alerts when the maximum likelihood is generated by one of the fall models. This design of HMMs focuses on the specialized characteristics that belong to each specific category of activities, e.g., fall forward, fall backward, walk, lie down, etc., allowing for more fine-grained analysis of fall events and normal activities.

Representations for Observation Distributions

Regarding the representations for observation distributions, most prior studies extracted feature vectors from signals and discretized them as observation outputs, via manual segmentation or k-means clustering (codebook approach). This approach is subjective, as researchers select the features, segmentation thresholds and number of clusters.

Tong et al. (2013) and Khan et al. (2014) developed an alternative approach to resolve the subjectivity issue. Instead of extracting features from acceleration signals, Tong et al. (2013) used resultant (magnitude) acceleration signal series from just before fall events to provide fall prediction (200 to 400 milliseconds before the collision). They divided the acceleration signals into 8 segments based on their magnitude and used discrete observation probabilities for hidden states. After the model parameters were trained, given a new acceleration signal series, the likelihood that the model generates the series was used as the criteria distinguishing falls vs. normal activities. They computed two thresholds for optimal fall prediction and fall detection, respectively. Essentially, they constructed a model to recognize the collapse phase of a fall event, not considering the impact or inactivity phases. This approach may lead to an incomplete view of fall events and introduces false alarms in real-world scenarios. Furthermore, they mainly tested their system on forward falls. Although forward falls are a main category of falls, ignoring backward or lateral falls may adversely affect the effectiveness of the system as they typically exhibit different patterns (Kangas et al. 2008). Although this work used manual segmentation to discretize acceleration signals, it suggests an alternative way to avoid hand-crafted features.

Khan et al. (2014) applied HMMs for fall detection in the absence of fall-specific training data. They used Gaussian distributions as their observation distributions for hidden states and trained HMMs on normal activities and “inflated” the covariance matrices of Gaussian distribution parameters to capture the anomalies. This method is essentially an anomaly detection method, and falls are only one type of potential anomaly. It may not be preferred for a specialized fall detection system. However, their application of Gaussian distributions as observation distributions provides an effective approach for handling

continuous acceleration data, as discrete observation probabilities of HMMs may lose the ordinal information of accelerations and are sensitive to the manual threshold selection for segmentation. Inspired by this work, we chose Gaussian distributions as the observation distributions for our HMMs to provide a smooth representation of acceleration signals.

2.2.3. *Sensor Orientation Calibration*

Acceleration signals collected by tri-axial accelerometers are represented by tuples of three orthogonal acceleration components (x, y, z) in the sensor reference frame. However, the sensor reference frame is not always well-aligned with the body reference frame denoted by the vertical (VT), medio-lateral (ML), and antero-posterior (AP) directions. The correspondence between the two reference frames may change from time to time and from subject to subject. For instance, if the x -axis of the accelerometer is aligned vertically, a $+1\text{ G}$ (9.81 m/s^2) of acceleration (due to gravity) will be reflected on the value of the x -acceleration. However, if we tilt the sensor and let the y -axis be aligned vertically, the gravity component will be reflected on the value of the y -acceleration. This inconsistency provides much complexity in data processing, in that two series of signals may not be directly comparable without the prior knowledge of sensor orientation, which is a common case in less controlled real-life senior living scenarios.

Past studies in gait analysis (Zhong and Deng 2014; Palermo et al. 2014) and activity recognition (Prathivadi et al. 2014; Morales et al. 2014) observed large discrepancies in signal patterns affected by inconsistent or unknown sensor reference frames. Figure 2.2 provides a comparison of sensor signal snippets collected from two different orientations for a simulated fall. Figure 2.3 further shows the differences (the first chart subtracted from the second)

between the two series. Most of the data points deviate from the zero line, which indicates large discrepancies. This can lead to substantially lower classification or recognition accuracies. In a study using dynamic time warping (DTW) for activity recognition, Kale et al. (2012) concluded that the sequence matching algorithm achieves an accuracy of less than 50 percent if the sensor is misoriented by more than 15 degrees on any of the three axes.

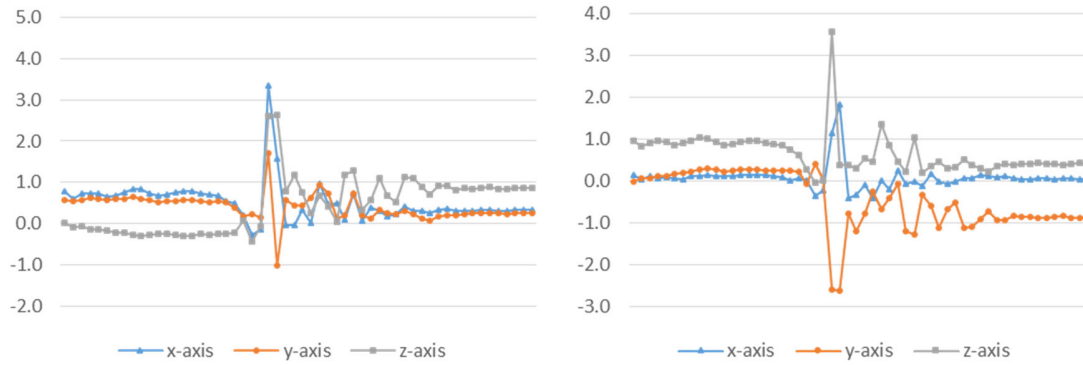


Figure 2.2: Signal Series for Simulated Falls Collected from Two Sensor Orientations

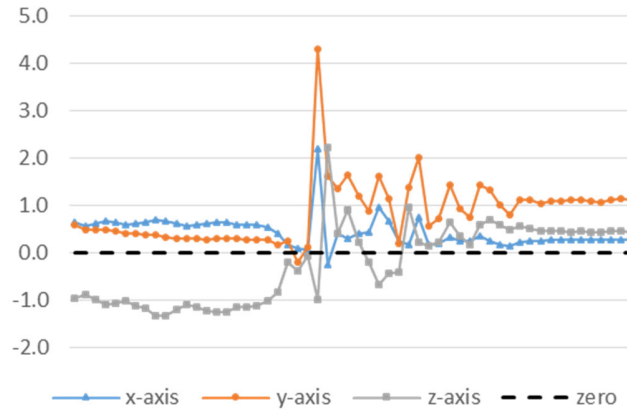


Figure 2.3: Differences between the Two Signal Series Shown in Figure 2.2

According to Henpraserttae et al. (2011), there are three potential approaches to address this issue. The first approach is to use orientation-invariant, robust features. Using acceleration magnitude is one common approach (Kau and Chen 2015; Shen et al. 2015; Tong et al. 2013).

It is also feasible to use the inner products of vectors to provide inference (Zhong and Deng, 2014). This approach avoids the orientation issue of loss of direction information and may provide inferior results, as posture change is an important indicator for detecting falls. The second approach is to create a set of models specific to each detected sensor orientation. This approach is infeasible in our case as there are potentially infinite sensor orientations set by users. The third approach is to calibrate the signals to a unified reference frame as a preprocessing procedure. A transformation matrix can be applied on the raw acceleration signals for calibration. Equivalently, two orthogonal bases for the unified body reference frame should be identified as vectors in the original sensor reference frame (the third base can be delivered by a cross product of the first two). In cases where two orthogonal bases cannot be obtained, identifying only one base can also act as partial orientation calibration. This approach is most applicable in our case, in that it can provide acceptable consistency across the signal sequences without losing direction information.

We further investigated prior studies in sensor orientation calibration. The calibration can be conducted explicitly, e.g., using a camera (Wu and Jafari 2014), or by asking the user to perform specific calibration actions (Prathivadi et al. 2014). However, both methods are intrusive to users and thus less applicable in real scenarios. We elect to consider implicit calibration methods without extra human effort, based on the characteristics of sensor signals. Morales et al. (2014) constructed a new reference frame by performing principal component analysis (PCA) on acceleration signals. The drawback of this approach is that the PCA-decomposed reference frame does not guarantee consistent correspondence with the body reference frame. A common method for estimating the vertical axis of the body reference frame is to leverage the fact that the accelerometer is constantly measuring gravity. Mizell (2003)

proposed an approach to produce a good estimate of gravity (i.e., the vertical axis) by taking the average of acceleration signals over a reasonable time period. Henpraserttae et al. (2011) further estimated the antero-posterior axis from the horizontal components by performing eigen-decomposition, assuming that most of the activities were conducted along a certain direction. This method can be applied only when walking is the major activity that provides a good estimation of the antero-posterior axis. If not, an alternative is to use the magnitude of the horizontal component after estimating the vertical component (Yang 2009; Gao et al. 2014), i.e., decompose the tri-axial acceleration signal into a vertical component and a horizontal magnitude. This approach is simple to calculate (only estimating the gravity) yet reserves direction information along the vertical axis, which is sufficient for our analysis of posture changes.

2.3. Research Gaps and Questions

As discussed above, most prior studies in fall detection extracted hand-crafted features for their threshold-based, classifier-based, or HMM-based systems, which are often subjective and labor intensive. In this research, we propose a new representation that applies HMMs for fall detection. We use original acceleration signals as inputs, which is feature-less and able to avoid *ad hoc* feature engineering, and adopt Gaussian distributions as continuous observation distributions for HMMs, which avoids manual segmentation for discrete observation outputs to provide a smoother representation.

Most prior studies in fall detection applied a setting of known sensor placement (known location, e.g., waist or chest; known orientation, e.g., y -axis aligned with gravity) in their experiments. However, in reality it is difficult to obtain sensor placement details. Alternatively,

orientation-invariant features (e.g., acceleration magnitude) were used as input to avoid placement issues. However, losing the direction information weakens the system's ability to detect posture changes. Both approaches lead to inferior fall detection accuracy due to the inability to model fall events. Also, to the best of our knowledge, no prior research has applied sensor orientation calibration on HMMs for motion sensor-based fall detection. In this research, we propose to develop sensor orientation calibration algorithms to resolve the misplacement issues (misplaced locations and misaligned orientations) in real-life home monitoring scenarios. Furthermore, most prior studies did not include real-world fall events for evaluation. The generalizability of such fall detection systems needs further confirmation. We ask the following research questions:

1. Can the hidden Markov model-based fall detection system achieve high precision and recall with our new representation?
2. Can the orientation calibration algorithm improve the precision and recall for HMM-based fall detection systems, given that the sensor orientation is unknown?
3. Can the system achieve high performance on both location-irrelevant and location-specific data?
4. Is the system generalizable to real-world fall events?

2.4. Research Design

Figure 2.4 illustrates our research design. The major components, data collection, sensor orientation calibration, hidden Markov model design, and evaluation, are discussed in detail in the following subsections.

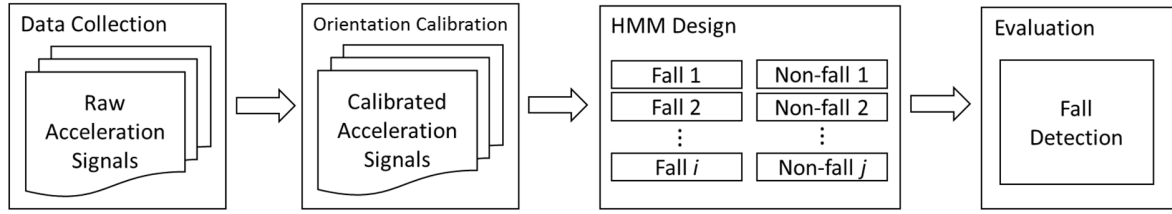


Figure 2.4: Research design for our HMM-based fall detection system

2.4.1. Data Collection

Two datasets were used to train and evaluate our proposed hidden Markov model-based fall detection system. We collected the first dataset in controlled lab experiments. College student volunteers were asked to perform simulated fall events and normal activities. The second dataset was acquired from the FARSEEING project, which is a real fall repository project funded by the European Union. The FARSEEING dataset had been used in past studies (Bagalà et al. 2012; Bourke et al. 2016) and is suitable for real-life scenario assessment by our system.

For the first dataset, we used tri-axial accelerometers on MetaWear C sensors (inertial unit model: Bosch BMI160¹) in our experiments. The size of the sensor is illustrated in Figure 2.5 (Left). The resolution was set to 12 bits, measuring acceleration signals between +4 G and -4 G. The sampling frequency was set to 12.5 Hz. Ten student subjects (aged 23 to 29) from a public university participated in our controlled experiments. We attached five sensors on subjects to investigate the influence of sensor locations on our fall detection system. The five locations were neck (necklace pendant), waist (keychain), chest (left shirt pocket), right side of body (right trouser pocket), and left side of body (left trouser pocket), as illustrated in Figure 2.5 (Right). Those locations are commonly used locations in past studies (Bourke et al. 2016;

¹ https://www.bosch-sensortec.com/bst/products/all_products/bmi160, retrieved on May 1, 2019.

Shen et al. 2015; Liu and Lockhart 2014). Note that although we attach five sensors simultaneously, we only use single sensor data to train our models, as we treat the data collected from different sensors as separate samples.



Figure 2.5: (Left) Sensor Shape and Size; (Right) Sensor Locations (Frontal View)

Four categories of simulated fall events were performed by each subject (Bourke et al. 2007). The subjects intentionally fell onto a large mattress in four directions from a still, upright posture: forward, backward, left laterally, and right laterally. Eight categories of common non-fall activities of daily living from prior studies (Shen et al. 2015; Özdemir and Barshan 2014) were performed as well: quiet sitting, quiet standing, quiet lying, walking, standing up, sitting down, lying down, and getting up from bed.

Different from prior studies, we collected signals from five locations on the human body, but used them independently. We aimed to investigate whether sensor locations can affect fall detection. Thus, we created two datasets based on the setting. The first is the pooled, location-irrelevant dataset (denoted as POOL), where we pooled the data collected from different locations together and ignored the actual location. This gave us 200 fall samples and 385 non-fall samples (15 samples were dropped due to incomplete data). We used this dataset to investigate the performance of our system with a location-irrelevant setting. The second is

the individual, location-specific datasets (denoted as IND), where we treated data collected from different locations as individual datasets. This gave us 40 fall samples and 80 (or 75) non-fall samples for 5 locations. We used these datasets to investigate the performance of our system on separate locations and reveal the best location to attach the motion sensor.

Our second dataset (denoted as FAR) is a real-life fall dataset of 22 falls by 22 subjects from the FARSEEING project, which involved universities, medical and research centers, and enterprises from Italy, France, Norway, UK, and Germany. Over 300 real-world fall events were recorded from 2012 to 2015, among which 22 are accessible for research purposes. They used sensors attached to the lower back or right thigh. Accelerometers were used for all subjects, with gyroscopes and magnetometers used by a few of them. For consistency, we used only accelerometer data from the dataset for our system evaluation. The sampling frequency was 20 or 100 Hz. Data for each recorded fall contains signals and a fall report describing the context of the fall, subject profile, sensor information, caused injuries, etc. An example for fall report is “walking down the corridor, hooked into a handrail then fell down.” Each fall record contains a two-minute snippet of tri-axial accelerations. The data point where a fall occurred has a fall mark. It is crucial for a fall detection system to identify not only *whether* a fall occurred, but also *when* it occurred. Therefore, a better time granularity for the fall samples is necessary.

To simulate real-life usage of the fall detection system, we segmented the signal snippets into non-overlapping, one-second samples, treating them as independent events. This resulted in 2,640 samples (22 x 120 seconds). Those samples with a fall mark inside were used as a dataset of fall events (22 fall events). The others constituted a dataset of normal activities

(2,618 normal activities). This highly skewed dataset is consistent with real-life situations, where falls are very rare events in daily life.

2.4.2. Sensor Orientation Calibration

We developed a transformation on acceleration signals to perform sensor orientation calibration. Each data point, \mathbf{a}^T , is a tri-axial acceleration signal represented by a 1×3 row vector,

$$\mathbf{a}^T = [a_x \quad a_y \quad a_z].$$

Each sample, S , is an $n \times 3$ matrix, consisting of n acceleration vectors,

$$S = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_n^T \end{bmatrix} = \begin{bmatrix} a_{x,1} & a_{y,1} & a_{z,1} \\ a_{x,2} & a_{y,2} & a_{z,2} \\ \vdots & \vdots & \vdots \\ a_{x,n} & a_{y,n} & a_{z,n} \end{bmatrix}.$$

We applied the calibration process introduced in Yang (2009) to decompose the tri-axial data into a vertical component and a horizontal magnitude. This idea was also applied by Lee et al. (2015) and Henpraserttae et al. (2011). The gravity (or the vertical axis, \mathbf{a}_V^T) is estimated by an average of acceleration signals, \mathbf{a}_i , over a reasonable time period, t_0 to $t_0 + \Delta t$,

$$\mathbf{a}_V^T = \frac{1}{\Delta t} \sum_{i=t_0}^{t_0+\Delta t} \mathbf{a}_i^T.$$

We used the first second of each data sample to provide the estimation, as those data were recorded in an upright and still posture.

For each acceleration signal, \mathbf{a}^T , the vertical component, v , is computed as the projection of \mathbf{a}^T on the direction of \mathbf{a}_V^T ,

$$v = \frac{\mathbf{a}^T \mathbf{a}_V}{\|\mathbf{a}_V^T\|},$$

where $\|\mathbf{a}_V^T\|$ is the L₂-norm of \mathbf{a}_V . Note that v is a signed scalar. The remaining (horizontal) component of \mathbf{a}^T , \mathbf{a}_H^T , is given by

$$\mathbf{a}_H^T = \mathbf{a}^T - v \frac{\mathbf{a}_V^T}{\|\mathbf{a}_V^T\|}.$$

We use $\|\mathbf{a}_H^T\|$, which is the L₂-norm of \mathbf{a}_H^T , as an indicator of the horizontal component of the acceleration vector. We further denote $\|\mathbf{a}_H^T\|$ as h .

This approach transforms the acceleration vector, \mathbf{a}^T , into a combination of vertical component and horizontal magnitude, $[v \ h]$. The vertical component is the signed magnitude of the vector on the estimated gravity direction. The horizontal magnitude is the unsigned magnitude of the vector on the plane perpendicular to the estimated gravity direction,

$$\mathbf{a}^T = [a_x \ a_y \ a_z] \rightarrow \mathbf{a}'^T = [v \ h].$$

For sensor data collected in high sampling frequencies (e.g., >100 Hz), it is preferred to first apply sliding windows on the signal series to extract local features (e.g., maximum) before applying HMMs (Tong et al. 2013; Khan et al. 2014). The goal is to reduce the length of inputs and avoid unnecessary complexity. However, we skipped this procedure because the effect of our low frequency setting of 12.5 Hz is similar to extracting local features using sliding windows. Thus, the combination of vertical component and horizontal magnitude of acceleration signals are treated as the observation outputs for HMMs.

2.4.3. Hidden Markov Model Design

As discussed, two major types of HMM designs exist in past literature. One design trains a separate HMM for each category of fall events and non-fall normal activities (as in

Lim et al., 2014) with the idea of capturing the details of special characteristics for each category of activities. We denote the first design as *specialized models* (SPEC). The other design trains a general HMM for all fall events and another for all non-fall normal activities (as in Tong et al. 2013; Khan et al. 2014) with the idea of capturing the generalized characteristics of falls and non-falls. We denote the second design as *generalized models* (GEN). A comparison of the two design types is illustrated in Figure 2.6. We applied both designs to further investigate which is better.

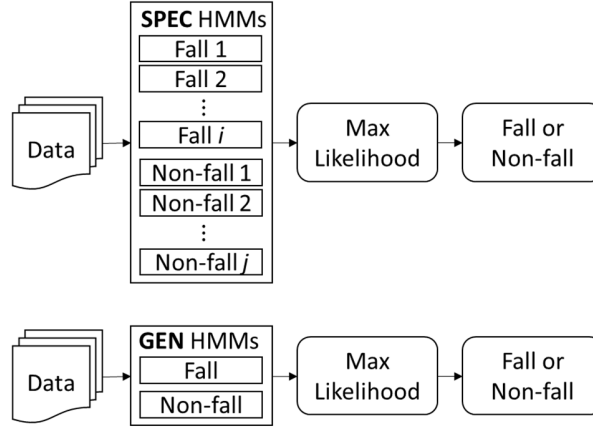


Figure 2.6: Comparison of SPEC and GEN HMM Designs

We first introduce the formulation of HMMs. A HMM, λ , with n hidden states, can be represented as

$$\lambda = \{A, B, \pi\},$$

where A is the $n \times n$ state transition matrix, B is the emission (observation) probability distributions for each state, and π is the start state probability distribution. We adopt Gaussian distributions as the emission probability distributions, as it can handle continuous acceleration data (Khan et al., 2014),

$$B_i \sim N(\mu_i, \Sigma_i) \text{ for } i = 1, 2, \dots, n.$$

Note that $\boldsymbol{\mu}_i$ is a $p \times 1$ column vector, and Σ_i is a $p \times p$ covariance matrix, where p is the dimension of observations ($p = 3$ for tri-axial acceleration signals). All components of $\boldsymbol{\mu}_i$ and Σ_i will be estimated during the training stage for each hidden state. In order to reduce the parameters to estimate, we let Σ_i be diagonal ($\Sigma_i = \text{diag}(\sigma_{11}^2, \sigma_{22}^2, \dots, \sigma_{pp}^2)$), assuming the dimensions of observations are uncorrelated.

The unknown parameters, A, B, π , are randomly initialized and estimated through a state-of-the-art Expectation Maximization (EM) algorithm (Baum-Welch algorithm) (Rabiner 1989) until convergence. Note that the number of hidden states, n , is a parameter to be set. Past literature used $n = 3$ or 4 (Tong et al. 2013; Khan et al. 2014). To explore more settings on n , we conducted tests on $n = 3, 4, 5$, and 6 . Detailed results are disclosed in the discussion section.

For the specialized model (SPEC) design, we construct a set of HMMs for fall events,

$$\Lambda_f = \{\lambda_{f1}, \lambda_{f2}, \dots, \lambda_{fj}\},$$

corresponding to the j categories of fall events; and a set of HMMs for non-fall activities,

$$\Lambda_a = \{\lambda_{a1}, \lambda_{a2}, \dots, \lambda_{ak}\},$$

corresponding to the k categories of non-fall activities. In our case, $j = 4$ and $k = 8$. Denote

$$\Lambda = \Lambda_f \cup \Lambda_a.$$

Given a new acceleration signal sample, S , we calculate

$$\lambda^* = \text{argmax}_{\lambda_i \in \Lambda} P(S|\lambda_i),$$

where $P(S|\lambda_i)$ is the log-likelihood for model λ_i to generate S . We classify S as a fall event if $\lambda^* \in \Lambda_f$; otherwise it is a non-fall normal activity.

For the generalized model (GEN) design, we construct λ_f for all fall events and λ_a for all non-fall activities. Given a new acceleration signal sample, S , we classify it as a fall event if

$$P(S|\lambda_f) > P(S|\lambda_a),$$

otherwise it is a non-fall normal activity.

2.4.4. Evaluation

To show the advantage of applying sensor orientation calibration in fall detection, we compare our orientation calibrated system (OUR) with two benchmark systems without calibration.

Benchmark 1 (BEN1): a HMM-based fall detection system using acceleration magnitude as observation outputs. Tri-axial acceleration signals are reduced to an orientation-invariant, one-dimensional series ($[\|\mathbf{a}^T\|], \mathbf{a}^T = \sqrt{a_x^2 + a_y^2 + a_z^2}$). This approach was applied in Tong et al. (2013) and other threshold- and classifier-based fall detection systems (Kau and Chen 2015; Shen et al. 2015; Tong et al. 2013).

Benchmark 2 (BEN2): a HMM-based fall detection system using raw tri-axial acceleration signals ($[a_x, a_y, a_z]$) as observation outputs. No modification or transformation was conducted on the data, assuming there are no orientation issues. This idea was prevalent in threshold- and classifier-based fall detection systems (Lee et al. 2015; Özdemir and Barshan 2014).

In addition, we implemented a basic threshold-based method (BASE) as a reference baseline to show the advantage of using a temporal pattern recognition model compared to a threshold-based model. A simple heuristic is applied in this method, in that if the maximum

magnitude of acceleration signals exceeds some threshold, a fall alert is issued. Essentially, this method, widely applied in past literature (Bourke et al. 2016; Lee et al. 2015; Kau and Chen 2015), identifies the impact phase during a fall event. To find the optimal threshold, we tested baselines varying from 1.0 G to 4.0 G, using the F-measure as the metric. We found 2.5 G to be the optimal separating threshold, achieving an average F-measure of 0.851.

Both the SPEC and GEN HMM designs were evaluated and compared. Sensitivity (recall rate) is the focus of fall detection systems, since failing to capture a fall event can lead to severe outcomes, such as long-lie situations. However, frequent false alarms (low precision rate) also reduce users' willingness to use such systems. For simplicity, we report our performance evaluation in the form of F₁-measure (F-measure), which is the harmonic mean of precision and recall,

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}, \quad F_1 - measure = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall},$$

where TP is the true positive (detected as a fall and is a real fall), FP is the false positive (detected as a fall but is not a real fall), and FN is the false negative (not recognized as a fall but is a real fall). The F-measure provides overall evaluation of the system's ability to detect potential falls while reducing false alarms.

We evaluated our systems by using the POOL, IND, and FAR datasets. The POOL evaluation aims to examine whether that our new representation for HMMs can handle varying sensor locations and orientations, and whether models with sensor orientation calibration can achieve higher performance. Evaluation on the IND dataset aims to examine the performance of the system handling location-specific sensor data. Evaluation on the FAR dataset aims to examine the generalizability of our system on real-life fall data.

All systems were evaluated using ten-fold cross-validation to avoid over-fitting. As the Expectation Maximization algorithm finds local maxima when estimating HMM parameters, we repeated the training procedures five times and report the average performance.

2.5. Results and Discussion

This section presents evaluation results on three separate datasets: POOL, IND, and FAR. Precision, recall, F-measure, and accuracy are reported for respective datasets.

2.5.1. Dataset: POOL

To illustrate that our system is robust on sensor misplacement, we conducted evaluations for POOL data (location irrelevant, orientation unknown) on both the specialized model (SPEC) design and the generalized model (GEN) design. N is a tuning parameter controlling the complexity of HMMs (the number of hidden states). Figure 2.7 (Left) shows the results for the SPEC design. It is clear that our system with sensor orientation calibration (F-measures 0.984 to 0.986) outperforms the two benchmarks (F-measures 0.809 to 0.887). Benchmarks without orientation calibration (BEN1, BEN2) achieved similar performances compared to the baseline (BASE) (0.851), which indicates failure to fully utilize the direction information in acceleration signals. This result supports our argument that magnitude of resultant acceleration data (BEN1) and raw tri-axial acceleration data (BEN2) provide inferior fall detection results compared to our system. Using raw tri-axial acceleration data suffers from the inconsistent reference frames for the signal samples, which leads to the system's inability to identify distinguishable patterns along each body axis. Using the magnitude of resultant

acceleration data avoids the orientation concerns at the cost of losing direction information, which is improper for the fall detection task where posture changes are an important indicator.

Figure 2.7 (Right) shows the results for the GEN design. Our system provides F-measures ranging from 0.814 to 0.899, which did not significantly outperform the benchmarks (0.732 to 0.893), and are considerably lower than the SPEC design measures (0.984 to 0.986). A possible reason is that a generalized model averages the characteristics of different categories of fall events (e.g., falling forward and backward) and normal activities (e.g., standing and walking), making it less sensitive to each of the categories.

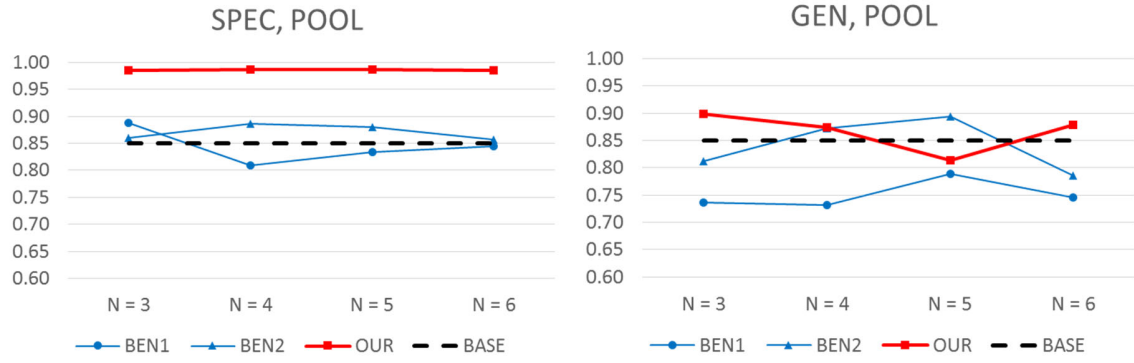


Figure 2.7: F-measure Comparison on POOL Dataset: (Left) SPEC; (Right) GEN

We show our detailed results in precision, recall, and F-measures in Table 2.1. The model parameter N was set to 4, which was the best-performing setting. Our proposed SPEC design model achieved promising results of precision at 0.981 and recall at 0.992, indicating less than 2 percent false alarms and less than 1 percent missing fall events in the POOL dataset. This result significantly outperforms the benchmarks, which have 10 to 33 percent false alarms and 4 to 13 percent missing fall events.

Table 2.1: Precision, Recall, and F-measures for POOL Dataset Evaluation

		Precision	Recall	F-measure
SPEC	OUR	0.981	0.992	0.986
	BEN1	0.826	0.958	0.887
	BEN2	0.906	0.867	0.886
GEN	OUR	0.822	0.992	0.899
	BEN1	0.667	0.967	0.789
	BEN2	0.896	0.892	0.894
BASE		0.804	0.903	0.851

We then analyze our results based on category of fall events and normal activities. For illustration, we chose one of the five repetitions of model training and show the classification accuracies (whether it is a fall event or a normal activity) in Table 2.2. The first eight rows show results from normal activities, while the last four rows are results from simulated fall events. In the SPEC design, our system perfectly recognized all the fall events, with only a few accidental misclassifications of a quiet standing instance as a fall event. Benchmark 1 also recognized all the fall events, at the cost of a relatively low accuracy on normal activities (as low as 0.86). This indicates a high false alarm rate, which conforms to our analysis that using only acceleration magnitude is sub-optimal for a real-life fall detection system. Benchmark 2 produced relatively good results (0.95 to 1.00) on normal activities, but the performance on detecting falls (as low as 0.88) was poor. This also conforms to our observation that raw tri-axial acceleration signals may suffer from inconsistent reference frames. A similar situation exists in the GEN design, except that our model generated false alarms on normal walking activities. This may be because the walking signal pattern exhibits similar characteristics to the generalized model of four of the fall event models, or a generalized model for normal activities provides too little support for the walking activity. The baseline method (BASE) has an

acceptable performance detecting falls (over 0.92) and classifying normal activities, except for “sit down” and “lie down” (over 0.95). The poor performance for those two activities is due to the relatively large acceleration magnitudes when performing those activities, which also shows the deficiency of threshold-based fall detection systems to correctly identify those events.

Table 2.2: Accuracies of the Activity Categories in the POOL Dataset

		SPEC			GEN			
Type	Activity Category	OUR	BEN1	BEN2	OUR	BEN1	BEN2	BASE
Normal	Quiet standing	0.96	0.88	1.00	1.00	0.84	1.00	1.00
	Quiet sitting	1.00	0.91	1.00	0.95	0.86	1.00	1.00
	Walking	1.00	0.86	0.95	0.36	0.73	0.82	0.95
	Get up from bed	1.00	0.91	0.95	0.81	0.91	0.91	0.95
	Quiet lying	1.00	0.86	1.00	0.95	0.82	1.00	1.00
	Stand up	1.00	0.95	1.00	1.00	0.95	1.00	1.00
	Sit down	1.00	1.00	1.00	1.00	0.76	0.96	0.40
	Lie down	1.00	0.88	1.00	1.00	0.40	1.00	0.60
Fall	Fall forward	1.00	1.00	0.88	1.00	1.00	0.92	0.92
	Fall left laterally	1.00	1.00	0.92	1.00	1.00	0.84	0.92
	Fall backward	1.00	1.00	0.88	0.96	0.96	1.00	1.00
	Fall right laterally	1.00	1.00	0.92	1.00	0.96	0.88	0.96

2.5.2. Dataset: IND

To assess the system performance on location-specific sensor data, we compared the performances based on acceleration data collected from individual locations (IND). These locations included neck, waist, chest, right side, and left side. Separate models were constructed for each of the locations. Figure 2.8 (Left) shows that all five locations achieved F-measures of over 0.897 for the SPEC design. The chest and the waist locations provide the

best results (0.980 to 1.000), showing the advantage of attaching motion sensors to these locations for fall detection. Chest and waist sensors are close to the center of body mass, providing a more accurate description of body movement as a whole. Using the GEN design, the performance ranged from 0.755 to 1.000 (Figure 2.8 (Right)), which varied based on the location and the selection of model parameter N . This may be due to a lack of data samples for the model to generalize, as we only have 40 fall events and 80 (or 75) normal activities for each sensor location. Nevertheless, chest and waist locations still show better performance.

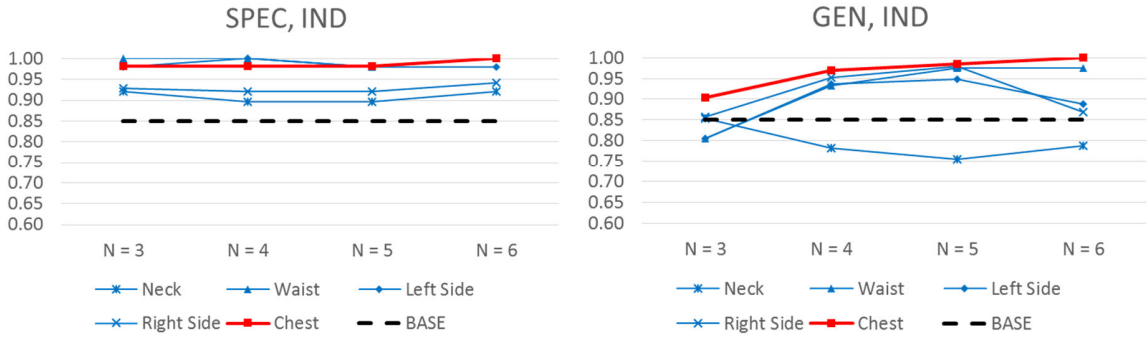


Figure 2.8: F-measure across Locations on IND Dataset: (Left) SPEC; (Right) GEN

2.5.3. Dataset: FAR

To assess the generalizability of our system, we also present the evaluation results using the fall event samples from the real-life fall repository, FARSEEING. This dataset is highly skewed, with 22 fall events and 2,618 normal activities. Our system achieved an F-measure of 0.880, clearly outperforming the benchmarks (0.123 to 0.694) (Table 2.3). In further analysis, a recall of 1.000 indicates no missing falls and a precision of 0.786 given 22 true positives (fall events) indicates that 6 false alarms are issued. An inspection of the false alarms reveals that those signal patterns are very similar to the marked fall events. There is an interesting case that two false alarms are alerted right before a recognized fall event (1 second and 3 seconds ahead,

respectively). A possible reason is that the patient fell during the three seconds and experienced multiple shocks, which is correctly captured by our system but not properly annotated in the dataset. The benchmark systems (BEN1, BEN2) generate too many false alarms (precision from 0.066 to 0.380) with missing fall events (recall from 0.864 to 0.909).

Table 2.3: Precision, Recall, and F-measures for the FAR Dataset

		Precision	Recall	F-measure
SPEC	OUR	0.786	1.000	0.880
	BEN1	0.211	0.909	0.342
	BEN2	0.380	0.864	0.528
GEN	OUR	0.500	1.000	0.667
	BEN1	0.066	0.909	0.123
	BEN2	0.170	0.864	0.284
BASE		0.629	0.773	0.694

2.6. Conclusion and Future Directions

Falls are one of the most severe threats faced by senior citizens living independently. We developed a hidden Markov model-based fall detection system that can automatically detect falls using a single motion sensor for real-life home monitoring scenarios. The main novelty in this work is that we proposed a new representation for acceleration signals in HMMs to avoid hand-crafted features. Gaussian distributions were chosen as HMM observation distributions to model multi-dimensional acceleration signals. Furthermore, we applied sensor orientation calibration on HMMs to resolve misplacement issues (misplaced sensor location and misaligned sensor orientation) encountered in real-life scenarios. We conducted student experiments to collect signal samples for simulated fall events and non-fall normal activities on five body locations, the neck, chest, waist, right side, and left side. We also used fall event

samples from a real fall repository, FARSEEING, as part of our evaluation dataset. Evaluation results showed that our system achieved a maximum F-measure of 0.986 for our pooled experiment dataset (POOL) and a maximum F-measure of 0.880 for the real fall dataset (FAR), clearly outperforming benchmark HMM-based systems without orientation calibration. The chest and the waist are the most appropriate locations to attach the sensor based on results of the location-specific datasets (IND). In conclusion, our system successfully leveraged acceleration data collected from multiple locations and fused them into an integrated model, which is crucial for fall detection applications in real-life home monitoring scenarios.

There are several limitations in our work. Although we selected multiple sensor locations on the human body, those locations were around the body trunk, following prior studies on fall detection. Extremities (hands, feet) were not investigated; since they are far away from the center of body mass their motion patterns are much more complicated and less related to the body trunk. We leave this challenge for future research. Another promising future research direction is about fall prediction or fall risk assessment in the context of senior care, which may involve representations and analytical models different from those used by fall detection systems.

3. ESSAY II: RISK PREDICTION FOR MOBILE HEALTH: A DEEP LEARNING APPROACH

3.1. Introduction and Essay Structure

Senior citizens face many challenges to their independent living, including a decline in mobility or cognition and chronic physical health conditions. Among those, falls are one of the most severe threats. Falls threaten senior citizens' living both physically and psychologically. In addition to direct injuries from a fall, many fallers are unable to get up again without assistance, leading to a "long lie" situation that can lead to serious medical complications. Also, senior citizens with a history of falling live in fear of future falls, which has been associated with negative consequences such as avoidance of activities, depression, and an increased risk of future falls.

Current clinical fall risk assessments typically involve survey-based numerical evaluations of fall risk factors and timed mobility tests. Marier et al. (2016) used electronic medical records (EMR) to improve fall risk prediction among nursing home residents. Widely accepted timed mobility tests include 10-meter ground walking and timed up and go (TUG). However, such assessments face the following limitations. Survey-based evaluations rely on interviews and the physician's subjective assessment. Timed mobility tests are associated with simple thresholds on completion time, which lacks detailed analysis of body motion and oversimplifies geriatric fall risks. Specialized equipment, including motion sensors, ground force plates, and electromyographic signals (Nelson-Wang et al. 2012), provides objective and quantitative measures for fall risks, but it is impractical to integrate into typical clinic schedules due to its costs and time-consuming setups. As a proxy, wearable sensor-based systems have

emerged to efficiently capture and analyze quantitative mobility data for fine-grained fall risk assessment. Miniature sensors are attached to the human body to collect data for analytics.

The significance of health predictive analytics has been recognized in prior information systems (IS) literature. Healthcare is a domain in which prediction is perhaps more important than explanation, considering the daunting cost of delay in diagnosis and treatment (Agarwal and Dhar 2014). Chen et al. (2012) discussed mobile analytics for healthcare with sensor-based technologies in the context of business intelligence and analytics. One focus of IS research is to develop and utilize information technology (IT) artifacts, including models, techniques, and systems, to address practical needs (Hevner et al. 2004). This stream of design science research is getting more attention and becoming even more relevant with the increasing interest in health predictive analytics and the emergence of mobile sensor technologies (Chen et al. 2012). With the immense amount of data generated by sensors (for example, a single motion sensor with a sampling frequency of 12.5 Hz can generate one million acceleration records per day, equivalent to 20 megabytes of data volume), the research motivations lie in providing more precise and prompt insights through the theory and development of advanced analytics techniques, especially the techniques that are effective in dealing with motion sensor data. Deep learning is one such promising technique for health predictive analytics given high-dimensional motion sensor data (LeCun et al. 2015; Goodfellow et al. 2016).

Consistent with the design science paradigm and prior IS research on health predictive analytics (Lin et al. 2017; Bardhan et al. 2015), we have developed and evaluated a novel deep learning approach in the context of health risk prediction. To the best of our knowledge, this is the first IS study focusing on motion sensor data with deep learning techniques. Our goal is to improve physicians' and patients' decision making and provide precision healthcare with

mobile health analytics. Specifically, we have utilized high-granularity sensor data and propose a two-dimensional heterogeneous convolutional neural network (2D-hetero CNN) to predict senior citizens' risks of fall events. Our 2D-hetero CNN could assist physicians in identifying patients with high risks to provide timely interventions and improve senior citizens' self-assessment to reduce or ultimately eliminate fall events. We chose predicting fall risks for Parkinson's disease (PD) patients as our research case due to the prevalence and significance of fall events for PD patients. We rigorously evaluated the 2D-hetero CNN against baseline approaches with a series of benchmark experiments. We also demonstrated the utility of the 2D-hetero CNN through two case studies of sensor signals.

The major contributions of our proposed 2D-hetero CNN are as follows. First, we propose a novel variation of convolutional neural network (CNN) that specializes in health risk prediction. By designing cross-axial and cross-locational convolution stages, the 2D-hetero CNN is able to extract both cross-axial (e.g., x -, y -, and z -axes in a sensor) and cross-locational (e.g., left and right parts of the human body) features that are semantically meaningful and effective for human gait and balance analysis. Second, the proposed 2D-hetero CNN serves as a good template for future motion sensor-related research. The extensive experiments and discussion in this work can guide future research in choosing model designs and hyperparameters on similar applications. Third, due to the feature engineering-less solution being grounded in deep learning, our proposed 2D-hetero CNN is generalizable and transferrable to the risk prediction of many mobility-related chronic conditions (e.g., Parkinson's disease, dementia, etc.), as well as other applications that involve multiple motion sensors to assess human motion (e.g., athlete performance assessment, rehabilitation therapy assessment, etc.).

The remainder of the essay is organized as follows. In the next section, we review the literature on existing IS works on mobile and health analytics, motion sensor-based fall risk prediction, and applications of convolutional neural networks in health contexts. We then present our research design of developing our 2D-hetero CNN architecture for fall risk prediction. Following that, we set up benchmark systems, experiments, and robustness checks to compare the results and show the utility of our model. Finally, we discuss the contribution of this study to the IS knowledge base, consider the practical implications, and suggest future directions.

3.2. Literature Review

Our research background is informed by the following four streams of literature: (1) existing IS works on health and mobile topics to understand the prevailing IS research directions, (2) scenarios and approaches applied in prior motion sensor-based risk prediction studies, (3) theoretical support for the analytical approach of deep learning, and (4) convolutional neural networks (CNN) as the state-of-the-art approach for signal pattern recognition, including risk prediction based on motion sensor signals. We then summarize research gaps and questions.

3.2.1. IS Works on Health and Mobile Topics

IS Works on Health Topics

Health and healthcare have long been areas of interest for IS researchers. Early studies on healthcare focused on health information systems (Venkatesh et al. 2011) and telemedicine (Miscione 2007). They answered questions about how health information systems were being

used by professionals (e.g., doctors, nurses, etc.) and their impacts. Recently, health information sharing has emerged as a new research area. Ayabakan et al. (2017) discussed how health information can be shared among health providers to reduce duplicate medical tests. Yaraghi et al. (2014) discussed how a health information exchange platform can be adopted and used by practitioners. Angst et al. (2017) focused on how IT security investments influence healthcare data breaches. There is also an increasing interest in understanding the privacy concerns of the disclosure of personal health information from the patients' perspective. Anderson and Agarwal (2011) examined the circumstances under which individuals are willing to disclose identified personal health information for digitalization. Angst and Agarwal (2009) studied how individuals can be persuaded to adopt electronic health records (EHR) given consumer privacy issues.

The above research streams followed the positivist paradigm and focused on how professionals and patients can adopt and use health information systems and data. With those as foundations, healthcare predictive analytics aims to predict certain outcomes or events given current data on clinical trials or records. Such predictions are often of great practical importance, as they are related to adverse events (Lin et al. 2017), hospital readmissions (Bardhan et al. 2015), or patient mortality (Tabak et al. 2014), among others. There is a significant distinction between predictive analytics and traditional explanatory research. As quoted by Lin et al. (2017),

“While predictive analytics may be used to inform causal inference, the primary goal of prediction models is not to unbiasedly explain whether a factor contributes to an outcome, but to predict the outcome in new observations as accurately as possible. This important difference between prediction research and explanatory research drives distinctive principles

for model development and evaluation given that explanatory power does not imply predictive power.”

Motivated by this same goal, we focus on designing more effective systems (i.e., IT artifacts) that can accurately predict the occurrence of adverse events in the future.

IS Works on Mobile Topics

With the number of mobile devices surpassing PCs in 2011 (Chen et al. 2012), mobile and sensor-based studies are catching more attention in the IS community. Early studies treated mobile devices as extra Web entry points beyond computers, focusing on mobile Internet (Ghose et al. 2012) and mobile Web browsing activities (Adipat et al. 2011). Mobile apps are another research focus for IS researchers. Hoehle and Venkatesh (2015) adopted Apple’s user experience guidelines to conceptualize the usability of mobile apps. Kwon et al. (2016) examined addictive behaviors on mobile social apps at an individual level. Sun et al. (2017) discussed how to maximize the expected revenue generated from mobile in-app advertising with an optimal sequencing strategy. Recently, with the popularity of the concept of Internet of Things (IoT), sensor technology is starting to receive interest. Venkatesh et al. (2017) designed and evaluated six shopping assistance artifacts in which they manipulated the hardware design – barcode scanner versus radio frequency identification (RFID) reader. The above studies have established a landscape of interest in mobile devices and sensors.

Nevertheless, how the area of health predictive analytics can benefit from the support of mobile and sensor technologies to provide more precise and personalized health services is still an underexplored topic. Although there have been smart wearable devices on the market for some time (e.g, Apple Watch, Fitbit, etc.), the research on the design, adoption, and use of

related IT artifacts is still a new realm. We focus on the design science perspective of IS, whose main objective is to address practical problems with IT artifacts (Gregor and Hevner 2013; Hevner et al. 2004). We position ourselves at the fusion of mobile sensor technologies and health predictive analytics, and we aim to explore how the design science discipline can provide a unique contribution to this field.

3.2.2. Motion Sensor-Based Risk Prediction: Scenarios and Approaches

In this section, we review the literature on risk prediction based on motion sensors. We focus on the scenarios of setting up sensors and the analytical approaches in prior studies.

Application Scenarios: Clinical and Home Settings

There are two major scenarios of risk prediction: clinical setting and home setting. In a clinical setting, subjects wear motion sensors to perform controlled mobility tests as instructed by physicians. Typical mobility tests include timed up and go (TUG) test (Silva and Sousa 2016; Greene et al. 2017; Najafi et al. 2014; Schwenk et al. 2014), sit-to-stand (STS) (Ejupi et al. 2017), and 10-meter ground walking (Wang et al. 2017). The test duration is usually short-term (less than 15 minutes) and provides a snapshot of the subject's gait and balance. The structured nature of clinical tests enables us to accurately annotate sensor signals based on ground-truth human motions. Multiple sensors are often used to provide a more comprehensive assessment across locations (chest, leg, ankle, etc.) and sides (left and right) of human body. Sensor locations and orientations are usually fixed and consistent across tests to ensure the comparability of collected data.

In a home setting, subjects wear motion sensors at home while performing their regular daily activities (e.g., sitting, walking, lying, etc.) (Silva and Sousa, 2016; Greene et al. 2017;

Schwenk et al. 2014). Typically, no specific instructions are provided for the subjects; they perform their daily routines as naturally as possible. The assessment duration is relatively long (more than 24 hours) compared to clinical tests. It is unstructured in the sense that the true labels of subjects' activities are difficult to obtain. Thus, a preprocessing algorithm is necessary for grouping the signals into snippets of activity categories; the algorithm then provides analysis on each activity category. Single-sensor systems are preferred, as multiple sensors are cumbersome for people's daily use.

Approaches: General, Gait, and Physical Activity Features

Most prior studies applied an approach of (1) deriving features from sensor signals, and then (2) applying statistical tests (e.g., t-test, ANOVA) to test their effectiveness of distinguishability. Various categories of features have been extracted to reflect different aspects of health risks and used for further statistical analyses or machine learning algorithms. Past studies used three major categories of features: general signal processing features, gait features, and physical activity features. General signal processing features are those common features that are also applied in other applications of signal processing and are not constrained by the domain of risk prediction or gait analysis. Such features include signal mean, standard deviation, maximum, inter-quartile range, energy, mean-crossing rate, fast Fourier transform (FFT) coefficients, etc. (Silva and Sousa 2016). Gait features reflect subjects' walking patterns, which are used in prior gait analysis in laboratory settings. Examples are walking speed, stride time, stride length, gait variability, gait symmetry, etc. (Van Schooten et al. 2015). Physical activity features are mostly derived from home monitoring scenarios, which reflect subjects' activities of daily living. For instance, Van Schooten et al. (2015) and Schwenk et al. (2014) used durations of walking, standing, sitting, and lying to measure subjects' activity level. This

approach focused on testing whether a single feature from their feature set could significantly predict risks. Only a few involved statistical models, such as logistic regression (Wang et al. 2017) or regularized discriminant classifier (Greene et al. 2017), to investigate the synergy of multiple features.

The major drawback of feature-based approaches is that the selections of features are ad hoc across researchers. Each study develops and evaluates on its own feature set, reducing the transferability between studies. There are also situations where contradictory results were reported by different groups of researchers who used the same set of features (Hubble et al. 2015). It is costly to cross-reference various feature sets and verify the most effective features. We thus turn to the feature engineering-less solution, i.e., deep learning, to provide risk prediction from sensor signals.

3.2.3. Theoretical Support in Deep Learning

Deep learning is a significant and emerging branch of research in machine learning. As noted by Abbasi et al. (2010), machine learning has its grounding in statistical learning theory. In a formal representation, (supervised) machine learning aims at finding a mapping, f ,

$$f: X \rightarrow \mathbf{y},$$

where $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i]$ are the input vectors (in our case, acceleration signals collected from sensors), and $\mathbf{y} = [y_1, y_2, \dots, y_i]$ are possible labels (in our case, high or low predicted risks). One intuitive and common approach is linear models, i.e., the mapping f is in the family of linear functions,

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x},$$

where \mathbf{w} is the vector of parameters (the same as $\boldsymbol{\beta}$ in common notations for linear regressions). Linear models are succinct and appealing in practice, as they can be fit efficiently and reliably. However, linear models also have the obvious defect that they cannot model complex relations between \mathbf{x} and y (e.g., higher order of \mathbf{x} or the interaction between components of \mathbf{x}) (Goodfellow et al. 2016, p. 164).

As further discussed in Goodfellow et al. (2016, pp. 164-165), to enable the models to represent nonlinear functions of \mathbf{x} , we can apply the linear model not to \mathbf{x} itself, but to a transformed input $\phi(\mathbf{x})$, where ϕ is a nonlinear transformation. There are three general options of choosing ϕ . The first option is to apply a very generic ϕ , such as the infinite-dimensional ϕ that is implicitly used by kernel machines based on the radial basis function (RBF) kernel. This way we can model arbitrary complexity, but the generalization of the model often remains poor. That is the reason why no prior studies directly apply models on the raw sensor signals \mathbf{x} for risk prediction. The second option is to manually engineer ϕ , just as the traditional approaches of extracting features in the signal processing literature we reviewed. This is the dominant approach that requires immense human efforts of feature engineering with little transferability across domains (e.g., risk prediction for falls vs. risk prediction for Parkinson’s disease). The third option is to *learn* ϕ as the strategy applied by deep learning. In this approach, we have a model

$$y = f(\mathbf{x}; \boldsymbol{\theta}, \mathbf{w}) = \mathbf{w}^T \phi(\mathbf{x}; \boldsymbol{\theta}),$$

where we have two sets of parameters: $\boldsymbol{\theta}$, the parameters we use to learn ϕ from a broad class of functions; and \mathbf{w} , the parameters that map from $\phi(\mathbf{x})$ to the desired output, y . This approach can capture the benefit of the first two options, especially the second option. Human experts can encode their knowledge to help generalization by designing families $\phi(\mathbf{x}; \boldsymbol{\theta})$ that are

expected to perform well. The advantage is that we need to find only the right general function family, $\phi(\mathbf{x}; \boldsymbol{\theta})$, and then let the algorithm learn its parameters $\boldsymbol{\theta}$, instead of directly designing the function $\phi(\mathbf{x})$ itself. This approach has been widely applied in complicated sensory tasks such as image and speech recognition (LeCun et al. 2015). We believe that deep learning is also applicable in our context of sensor-based health predictive analytics due to the representation learning ability of deep learning that can automatically identify effective underlying features from sensor signals.

3.2.4. Convolutional Neural Networks in Health Contexts

Convolutional neural networks (CNNs) are a type of deep learning model specializing in recognizing sensor signals. There are multiple layers stacked in CNNs for automatic feature extraction, i.e., learning $\phi(\mathbf{x}; \boldsymbol{\theta})$ (LeCun et al. 2015). CNNs are inspired by the mammalian visual cortex. There are two basic cell types in the cortex, simple and complex. Simple cells have relatively small receptive fields and respond maximally to specific edge-like patterns. Complex cells are larger receptive fields and are locally invariant to the exact position of the pattern. CNNs have a similar architecture. Given a single or multi-dimensional input (e.g., arrays or matrices), low-level layers in their architecture extract local features, and high-level layers learn global features. This characteristic makes CNNs a robust solution for various pattern recognition tasks without human effort of feature engineering.

There are three major types of layers in CNNs: convolution layers, non-linear layers, and pooling layers. Convolution layers use filters (i.e., matrices of weight parameters) to apply on a small portion of input data to extract local features. They are followed by non-linear layers, which increase the non-linearity of the entire CNN architecture. Then pooling layers are used

to reduce the number of parameters, keep salient features, and provide local scale invariance. The three layers can be stacked onto other layers, making the model “deep.” A typical CNN architecture has two or more sets of the three types of layers to successfully extract local and global features.

In health contexts, CNNs have been applied in various domains and applications. One-dimensional (1D) CNNs have been used on sequential data such as ECG (beat type classification, Kiranyaz et al. 2016) and accelerometer signals (activity recognition, Yang et al. 2015; Zeng et al. 2014). Two-dimensional (2D) CNNs have been used on image data such as radiology, medical imaging, handwriting, and camera image recognition. Wimmer et al. (2017) applied CNNs on endoscopy images to provide automated diagnosis of celiac diseases. Pereira et al. (2016) recognized patients’ handwriting to assess the severity of Parkinson’s disease. Anthimopoulos et al. (2016) applied CNNs on CT images to diagnose interstitial lung diseases. Three-dimensional (3D) CNNs have been used on cubic radiology (Alzheimer’s disease diagnostics, Hosseini-Asl et al. 2016) or video data for activity recognition (Ji et al. 2013).

For sensor signal data, 1D CNNs have been applied in past literature (Yang et al. 2015; Zeng et al. 2014; Chen and Xue 2015). The convolutions are based on temporal adjacencies in signals, as the sensors measure accelerations through a continuous period of time (Figure 3.1 (Left)). The limitation of 1D CNNs is that they can only capture features along a single axis. When multiple sensors are involved, 1D CNNs cannot capture features across sensor axes. To capture features along different dimensions, 2D CNNs have been widely used in image recognition. In this case, the two dimensions have the same semantics (homogeneity), i.e., spatial adjacencies (Figure 3.1 (Middle)). Convolving along one dimension is as meaningful

as convolving along the other dimension. However, dimension heterogeneity exists in our scenario, with one dimension on temporal adjacencies and the other on cross-axial adjacencies (or cross-locational adjacencies, if sensors are attached to different locations on the human body) (Figure 3.1 (Right)). Assume we have collected 100 acceleration signals from 3 sensors, each sensor with 3 axes (x , y , z). The three axes in the same sensor are clearly related to each other, and convolving its x -, y -, and z -axes has the semantics of capturing cross-axial features within a sensor set at a specific body location. However, convolving the y - and z -axes of one sensor with the x -axis of another sensor is less semantically meaningful, as they belong to part of two different sensors set at two locations. Similarly, convolving signals from sensors set on the left and right thighs (or feet) has a specific semantics of assessing balance, while other combinations do not. Although general CNN solutions have been proposed on multiple sensor axes for sensor fusion applications (Yao et al. 2017; Münzner et al. 2017), the architecture proposed by Yao et al. (2017) does not support temporal convolutions, which are necessary for capturing features in repetitive gait cycles from walking tests. The work proposed by Münzner et al. (2017) does not support selective sensor fusion (i.e., sensor fusion only between locations of interest such as left and right thighs), which is critical for human gait and balance analysis (how the left part of human body is coordinating with the right part). This observation has motivated us to propose a new CNN architecture that can capture both cross-axial and cross-locational features, specifically between the left and right parts of the human body, for advanced gait and balance analysis and health risk prediction.

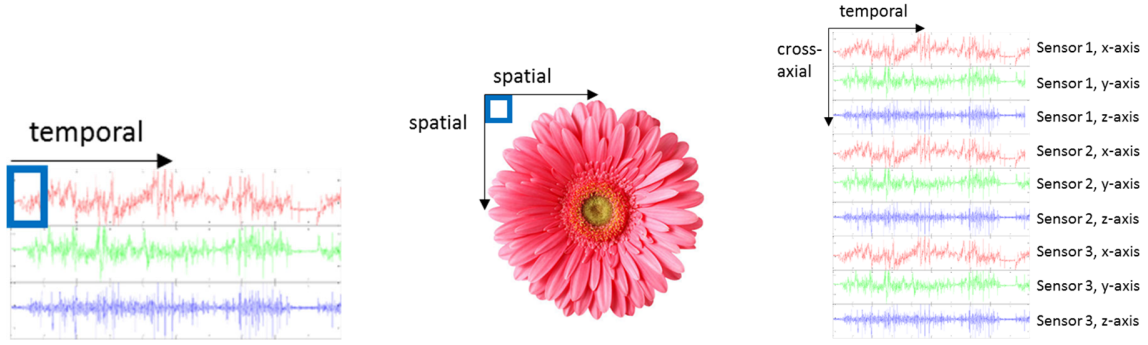


Figure 3.1: (Left) 1D CNN; (Middle) 2D Homogeneous CNN; (Right) 2D Heterogeneous CNN

3.3. Research Gaps and Questions

As discussed above, most prior studies on motion sensor-based risk prediction extracted ad hoc features from signals and applied univariate statistical tests to identify effective indicators, where the selection of features was laborious, subjective, and did not come to an agreement. To the best of our knowledge, no prior studies have developed deep learning systems on motion sensor-based risk prediction. Although CNN models have been proposed in radiology and image recognition as well as for sensor signals, existing architectures are less applicable in our context of gait and balance analysis for risk prediction. We designed our study to ask the following research questions:

1. How can we design a deep learning model that can fully leverage cross-axial and cross-locational information to improve fall risk prediction?
2. What are the generalizable guidelines that we can summarize from our design science practice?
3. How can the guidelines be used to benefit general health risk prediction and other related applications?

3.4. Research Design

In this essay, we propose our architecture of the 2D-hetero CNN to specialize and optimize in our context of multi-sensor risk prediction. Figure 3.2 illustrates our research design. The major components, data collection, data preprocessing, CNN architecture design, and evaluation, are discussed in detail in the following subsections.

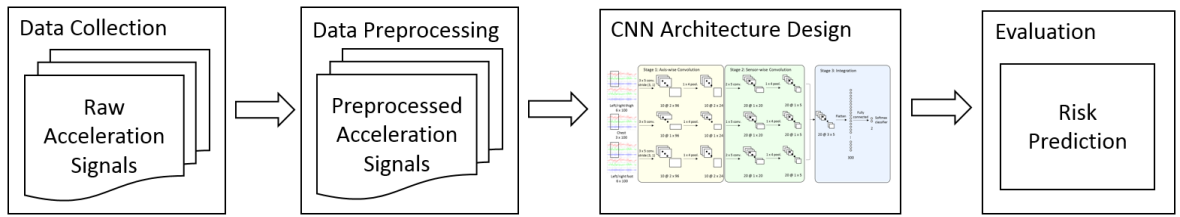


Figure 3.2: Research Design for Our Deep Learning Architecture: 2D-hetero CNN

3.4.1. Data Collection and Preprocessing

This study is part of the SilverLink project, a project aiming at remote senior care with motion sensing technologies. We have developed sensors and gateways for data collection from the human body and the environment. Accelerometers, gyroscopes, and thermometers are embedded in the sensors (inertial unit model: Bosch BMI160). The shape and size of the sensor can be found in Figure 3.3 (Left). In this study, we utilized only the accelerometers for our analysis.



Figure 3.3: (Left) Sensor Shape and Size; (Right) Sensor Locations (Frontal View)

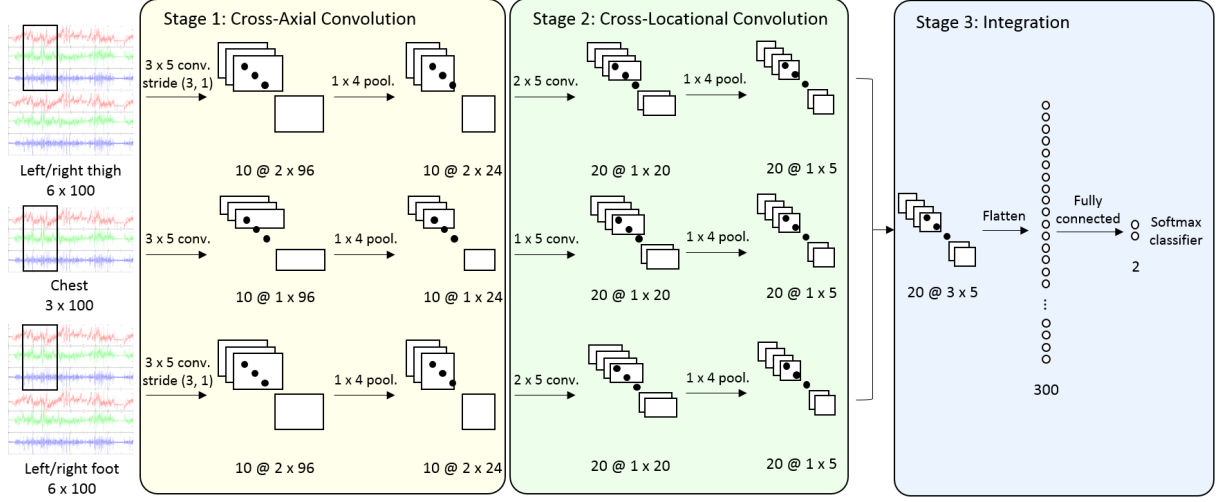
We recruited 52 subjects at a neurology clinic for our experiments. All of them were Parkinson’s disease patients at or over age 65. We used prospective fall occurrence as the criterion of fall risks. If the patient fell in the 12 months following the test, he/she was marked as a faller (predicted high fall risk), otherwise a non-faller (predicted low fall risk). Among the 52 subjects, 22 were fallers. This criterion has been widely used by prior studies, and the number of subjects is on a par with prior studies as well (Wang et al. 2017; Van Schooten et al. 2015; Schwenk et al. 2014). As a common practice in deep learning (Goodfellow et al. 2016), we augmented our dataset by 90-degree-wise rotations on each of the sensor axes (x , y , z), yielding a total of 3,328 samples. This is equivalent to asking the patients to wear the sensors rotated in given degrees and perform the mobility test again.

We attached five SilverLink sensors to each subject. The five sensors were set at the chest, left and right thighs, and left and right feet to capture body and lower extremity motions (Wu et al. 2013), as shown in Figure 3.3 (Right). The sampling resolution was set to 12 bits measuring acceleration signals between +4 G and -4 G (1 G equals 9.81 m/s^2) with a sampling frequency of 25 Hz. We chose 25 Hz because that was the maximum possible sampling frequency for our gateway to communicate with five sensors simultaneously. We conducted

10-meter ground walking tests in a hallway without obstacles to collect motion data for gait and balance assessment. The subjects wore sensors and walked at their own pace for 10 meters as instructed by the researchers. Walking aids such as sticks and walkers were allowed as long as the subjects could walk without others' help. This test setting is common in past gait analysis studies (Wang et al. 2017; Wu et al. 2013).

The subjects completed the 10-meter walking tests based on their specific conditions. The completion time varied from 10 to 25 seconds. However, CNNs require fixed-length signals as its input. In past literature on activity recognition, Zeng et al. (2014) used a window of 64 points, while Yang et al. (2015) used 30 or 32. We believe that a wider window was needed in our scenario of fall risk assessment to identify patterns spanning over multiple gait cycles. To get more stable gait patterns, we subsampled the middle four seconds of walking trials for all samples. This is equivalent to a window of 100 data points, as we sampled at a frequency of 25 Hz. As we had five sensors with three axes for each sensor, we ended up with matrices of 15 x 100 values as input for the analytical model.

3.4.2. 2D-hetero CNN Architecture



Note: conv. = convolutional layer; pool. = pooling layer. The notation “ $x @ y \times z$ ” denotes x feature maps with height y and width z .

Figure 3.4: 2D-hetero CNN architecture

Figure 3.4 illustrates our proposed 2D heterogeneous CNN architecture. The notation “ $x @ y \times z$ ” denotes x feature maps with height y and width z . We partitioned our data into three parts: left/right thigh, chest, and left/right foot, and fed each part of data into a separate CNN. This is to avoid the features extracted at one location influencing other locations on a low level of the model, as they are expected to have large discrepancies due to the nature of human motions (e.g., chest motions are more stable compared to extremity motions). There are three major stages in the architecture. Stage 1 provides cross-axial convolutions and aims to extract cross-axial features within a sensor set at a specific body location. Stage 2 provides cross-locational convolutions and aims to extract cross-locational features between left and right thighs and between left and right feet. Stage 3 integrates the extracted pattern to provide final inference on risk prediction. The main novelty of this architecture compared to traditional

2D homogeneous CNNs is that it provides semantically meaningful convolutions along the non-temporal dimension to specialize in gait and balance assessment for risk prediction.

For the first two stages, each stage comprised of the following three layers: convolution, non-linear, and pooling layers. Non-linear layers are omitted in Figure 4 due to space constraints.

Convolution Layers

A convolution layer convolves the data input or the previous layer's output with a set of filters. Each filter is essentially a matrix of weights applicable to a small proportion of the input data. The operation in convolution layers is represented by the following formula:

$$a_{x,y}^{i,j} = b^{i,j} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} w_{p,q}^{i,j,m} z_{x+p,y+q}^{i-1,m},$$

where $a_{x,y}^{i,j}$ is the output value of this layer at row x and column y for feature map j in layer i , $b^{i,j}$ is the bias for feature map j in layer i , m is the number of feature maps in the previous layer, P_i and Q_i are the height and width of the convolution filter in layer i , $z_{x+p,y+q}^{i-1,m}$ is the output from the previous layer (layer $i - 1$) for feature map m at row $x + p$ and column $y + q$, and $w_{p,q}^{i,j,m}$ is the weight for position (p, q) pointing from feature map m in layer $i - 1$ to feature map j in layer i , which is invariant to the values of x or y and is to be learned through the training process. The values of i, j, P_i, Q_i , and other parameters can be found in Figure 3. Note that a stride of (3, 1) is included in Stage 1 to ensure that only the three axes in the same sensor can be convolved.

Non-linear Layers

A rectified linear unit (ReLU) layer maps the output of the previous layer to generate the non-linearity of the model by the following function:

$$f(a) = \max \{a, 0\}.$$

The ReLU function is the most widely used non-linear function for CNNs due to its simplicity and effectiveness (LeCun et al. 2015).

Pooling Layers

In the pooling layers, the resolution of feature maps is reduced to keep salient features and increase local scale invariance. A maximum pooling layer is added after each non-linear layer axis by the following function:

$$a_{x,y}^{i,j} = \max_{\substack{0 \leq r \leq R_i-1 \\ 0 \leq s \leq S_i-1}} z_{x+r,y+s}^{i-1,j}$$

where $a_{x,y}^{i,j}$ is the output value of this layer at row x and column y for feature map j in layer i , R_i and S_i are the height and width of the pooling filter in layer i , and $z_{x+r,y+s}^{i-1,j}$ is the output from the previous layer (layer $i - 1$) for feature map j at row $x + r$ and column $y + s$. The exact values of parameters can be found in Figure 4.

We used dropouts on convolution layers to control overfitting as a common regularization technique (LeCun et al. 2015). For each pass of data fed into the CNN, each layer disables a random part of its neurons that participate in the CNN. The proportion of disabled neurons is known as the dropout rate. We tested our model for dropout rates from 0 to 0.5, and chose 0.25 as a balanced value between model convergence and overfitting avoidance. We also used early-stopping as a regularization strategy.

Stage 3 has a flattening layer and a fully connected layer, leading to a softmax classifier (high or low fall risk). The flattening layer flattens the feature maps into a one-dimensional array. The fully connected layer is linked to the softmax classifier to provide final inferences on the fall risks. All weights across the three stages are randomly initialized.

3.4.3. Evaluation

We compared the performance of our 2D-hetero CNN model with state-of-the-art benchmarks for risk prediction. We set up four sets of benchmarks.

1. Classic Machine Learning Algorithms. We compared our architecture with classic machine learning algorithms based on combined features extracted from sensor data. The algorithms included support vector machines (SVM), random forest (RF), and AdaBoost (AB). We extracted the most widely used features in the fall risk prediction literature, including acceleration mean, acceleration standard deviation, acceleration root mean square (RMS), jerk mean, jerk RMS, stride time variability, and walking speed (Hubble et al. 2015).

2. Alternative Neural Networks. We compared our architecture with existing state-of-the-art neural network benchmarks, including 2D homogeneous CNN (2D-homo CNN) (Wimmer et al. 2016; Pereira et al. 2016), 1D CNN (Kiranyaz et al. 2016; Yang et al. 2015), vanilla recurrent neural networks (RNN), long short-term memory (LSTM), and multilayer perceptron (MLP) with one hidden layer. To demonstrate neural networks' ability of automatic feature extraction, raw sensor data were used as model inputs without feature engineering.

3. Ablation Analyses. We disabled one of the two stages (cross-axial and cross-locational convolutions) of the 2D-hetero model to investigate the consequences. We also examined on leaving out part of the input data, either leaving out one location (chest, right

thigh, left thigh, right foot, or left root) or one sensor axis (x-, y-, or z-axis) to identify their impacts towards model performance.

4. Robustness Checks. We provided robustness checks on varying a series of hyperparameters and design choices of our 2D-hetero CNN, including the dropout rate, activation function, number of fully connected layers, convolution length, pooling length, and window size, and discussed their influences on the model.

Ten-fold cross-validation was applied as this is a common approach to avoid overfitting. We conducted two-sample t-tests to identify whether our 2D-hetero CNN significantly outperformed the benchmark models. The significant levels were labeled in the cells of benchmark models. For simplicity, we reported our performance evaluation in the form of F₁-measure (F-measure), which is the harmonic mean of positive predictive value (PPV, aka, precision) and sensitivity (aka, recall). The F-measure provides an overall evaluation of the system's ability to identify patients who are at risk of falling while reducing false alarms. We also reported the specificity as a common referential metric. The formulas for the above metrics are as follows:

$$Sensitivity = \frac{TP}{TP + FN}, \quad Specificity = \frac{FP}{FP + TN}, \quad PPV = \frac{TP}{TP + FP},$$

$$F_1 - measure = \frac{2 \cdot Sensitivity \cdot PPV}{Sensitivity + PPV},$$

where TP is the true positive (recognized as at risk and is at risk), FP is the false positive (recognized as at risk but is not), TN is the true negative (recognized as not at risk and is not at risk), and FN is the false negative (recognized as not at risk but is actually at risk).

3.5. Experiment Results

3.5.1. Classic Machine Learning Algorithms

We compared our proposed 2D-hetero CNN with three classic machine learning algorithms, SVM, RF, and AB, with a combination of features manually extracted from sensor data (Table 3.1). The two ensemble models (RF and AB) performed better than SVM, with AB performing the best across the three benchmark models. However, our model significantly outperformed all benchmark models in PPV, sensitivity, and F-measure. This indicated that manual feature engineering may overlook salient features in sensor data, while our deep learning-based model is able to extract such hidden features. With a proper architecture design, the neural network is able to learn the effective representation of the data and provide improved prediction results.

Table 3.1: Experiment Results for Classic Machine Learning Algorithms

	Specificity	PPV	Sensitivity	F-measure
2D-hetero CNN	0.920	0.927	0.963	0.943
SVM	0.636 ^{***}	0.600 ^{***}	0.667 ^{***}	0.632 ^{***}
RF	0.855 [*]	0.837 [*]	0.733 ^{***}	0.767 ^{***}
AB	0.909	0.889 [*]	0.889 ^{***}	0.889 ^{***}

Note: ^{*}: p-value < 0.05; ^{**}: p-value < 0.01; ^{***}: p-value < 0.001. The asterisks indicate the significance level of 2D-hetero CNN outperforming the corresponding benchmark.

3.5.2. Alternative Neural Networks

We compared our proposed 2D-hetero CNN with alternative neural networks, including two alternative CNN designs (2D-homo CNN and 1D CNN), two RNN designs (vanilla RNN and LSTM), and a classic artificial neural network (MLP) (Table 3.2). The inputs for the neural networks were raw data without feature engineering. The results showed that

MLP faces difficulties in extracting features from raw data for effective risk predictions. Although LSTM and RNN are able to capture temporal dependencies to some extent, they are typically inferior to CNN models in representation learning for sensor signal data. This observation conforms to a prior study (Hammerla et al. 2016). Among CNN models, our 2D-hetero CNN significantly outperformed 2D-homo CNN and 1D CNN. The results showed the advantage of extracting features on the non-temporal axis with explicit semantics (cross-axial and cross-locational convolutions). 1D CNN does not extract features on the non-temporal axis. Although 2D-homo CNN extracts features across such axis, they convolve one axis of a sensor with two axes of another, which is less semantically meaningful and might introduce irrelevant features. 2D-hetero CNN leverages the strength of temporal convolutions of 1D CNN and includes semantically meaningful non-temporal convolutions, and is thus preferable in our context of predicting fall risks.

Table 3.2: Experiment Results for Alternative Neural Networks

	Specificity	PPV	Sensitivity	F-measure
2D-hetero CNN	0.920	0.927	0.963	0.943
2D-homo CNN	0.684**	0.765***	0.950	0.843***
1D CNN	0.575***	0.668***	0.894	0.770***
RNN	0.441***	0.493***	0.547***	0.517***
LSTM	0.389***	0.559***	0.775***	0.649***
MLP	0.791***	0.604***	0.319***	0.417***

Note: *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001. The asterisks indicate the significance level of 2D-hetero CNN outperforming the corresponding benchmark.

3.5.3. Ablation Analyses

We provided sensitivity analyses on the two stages of our proposed 2D-hetero CNN. We disabled one of the two convolution stages and inspected their respective effects on the model. The full model outperformed the two benchmark models (cross-axial only and cross-loc. only) significantly in specificity, PPV, and F-measure. This showed the value of involving cross-axial and cross-locational convolutions simultaneously. Cross-axial convolutions extract features among axes within a sensor, while cross-locational convolutions extract features between the left and right sides of the human body. As a result, the full architecture detected subjects with high fall risks more accurately than the architectures with only one stage in effect.

In addition, we also inspected how leaving out part of the input data may impact the model performance. Leaving out data from one location lead to lower model performance in F-measure in general, although the decreases were not significant for right thigh or right foot locations. Leaving out data from one axis (F-measure 0.799 to 0.823) brought larger decreases than leaving out a location (F-measure 0.830 to 0.936). A possible reason is that missing one axis in sensor signals damages the integrity of sensor data and results in incomplete data in describing the motion of a particular location.

Table 3.3: Experiment Results for Sensitivity Analyses

	Specificity	PPV	Sensitivity	F-measure
2D-hetero CNN	0.920	0.927	0.963	0.943
Leaving out one of the first two stages in the model				
Cross-axial only	0.748 ^{**}	0.807 ^{**}	0.885	0.879 ^{**}
Cross-loc. only	0.844 ^{**}	0.869 ^{**}	0.986	0.922 [*]
Leaving out data collected from one location				
No Chest	0.866 [*]	0.878 [*]	0.947	0.910 [*]
No Right Thigh	0.907	0.906	0.941	0.919
No Left Thigh	0.858	0.868 [*]	0.916 [*]	0.890 [*]
No Right Foot	0.897	0.905	0.972	0.936
No Left Foot	0.650 ^{***}	0.735 ^{***}	0.956	0.830 ^{***}
Leaving out data collected from one sensor axis				
No x-axis	0.594 ^{***}	0.707 ^{***}	0.931 [*]	0.799 ^{***}
No y-axis	0.691 ^{***}	0.752 ^{***}	0.913 ^{**}	0.823 ^{***}
No z-axis	0.672 ^{***}	0.742 ^{***}	0.919 [*]	0.819 ^{***}

Note: *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001. The asterisks indicate the significance level of 2D-hetero CNN outperforming the corresponding benchmark.

3.5.4. Robustness Checks

We conducted a series of robustness checks to investigate how the changes of hyperparameters and design choices of the deep learning architecture may influence model performance (Table 3.4). The table rows in boldface are the design choices we have made in our 2D-hetero CNN. We compared our model with alternative settings in dropout rates, activation functions, numbers of fully connected layers, convolution lengths, pooling lengths, and window sizes. The results showed that a low dropout rate may lead to model overfitting, while a high dropout rate has an adverse effect on model convergence. Choosing ReLU as the activation function outperformed alternatives including sigmoid and tanh, which is consistent with common practices in deep learning (LeCun et al. 2015). In terms of the design of fully connected layers, a 2D-hetero CNN with one or two such layers generated similar results, whereas not including such layers at all may lead to significantly less effective models. The selection of convolution lengths did not play a crucial role in model performance, while a pooling length of four significantly outperformed that of one (i.e., no pooling) or two. Finally, a wider window size such as 100 was preferred compared to 32 or 64 as in the literature, which echoes our prior speculation.

Table 3.4: Experiment Results for Robustness Checks

	Specificity	PPV	Sensitivity	F-measure
2D-hetero CNN	0.920	0.927	0.963	0.943
Dropout Rate				
0.00	0.684 ^{***}	0.760 ^{***}	0.986	0.858 ^{***}
0.25	0.920	0.927	0.963	0.943
0.50	0.497 ^{***}	0.683 ^{***}	0.870 ^{**}	0.745 ^{***}
Activation Function				
ReLU	0.920	0.927	0.963	0.943
Sigmoid	0.764 ^{**}	0.806 ^{**}	0.956	0.873 ^{***}
Tanh	0.655 ^{***}	0.739 ^{***}	0.928 [*]	0.819 ^{***}
Number of Fully Connected Layers				
0	0.533 ^{***}	0.663 ^{***}	0.866 ^{**}	0.745 ^{***}
1	0.920	0.927	0.963	0.943
2	0.942	0.943	0.938 [*]	0.940
Convolution Length				
3	0.886	0.893	0.930	0.909
5	0.920	0.927	0.963	0.943
7	0.934	0.935	0.934	0.934
Pooling Length				
1	0.580 ^{***}	0.704 ^{***}	0.975	0.816 ^{***}
2	0.737 ^{***}	0.796 ^{***}	0.994	0.882 ^{**}
4	0.920	0.927	0.963	0.943
Window Size				
32	0.363 ^{***}	0.578 ^{***}	0.862 ^{***}	0.690 ^{***}
64	0.770 ^{**}	0.792 ^{**}	0.820 ^{***}	0.803 ^{***}
100	0.920	0.927	0.963	0.943

Note: *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001. The asterisks indicate the significance level of 2D-hetero CNN outperforming the corresponding benchmark.

3.5.5. Case Study

To better illustrate the nature of signals and the power of our proposed model, we selected two signal examples from two female Parkinson’s patients who were at risk of falling, but were misclassified by the best performing benchmark model (Figure 4.5). The peaks marked by red crosses roughly represent the steps walked by the patients. The vertical axes are measured in G. The horizontal axes are in a temporal order (numbers omitted). The first patient was aged 70 with a Unified Parkinson Disease Rating Scale (UPDRS) score of 28. The second patient was aged 72 with a UPDRS score of 17. They shared the pattern of fast paces (≥ 1.25 m/s), not frequently seen in people who are at risk of falling. Feature-based models tend to mark fast paces as indicators of low fall risks, which is the major reason for misclassifying these two cases. Our 2D-hetero CNN correctly recognized the two high-fall-risk cases by its strengths of extracting both cross-axial and cross-locational features, which may be difficult to recognize for humans or non-deep learning models.

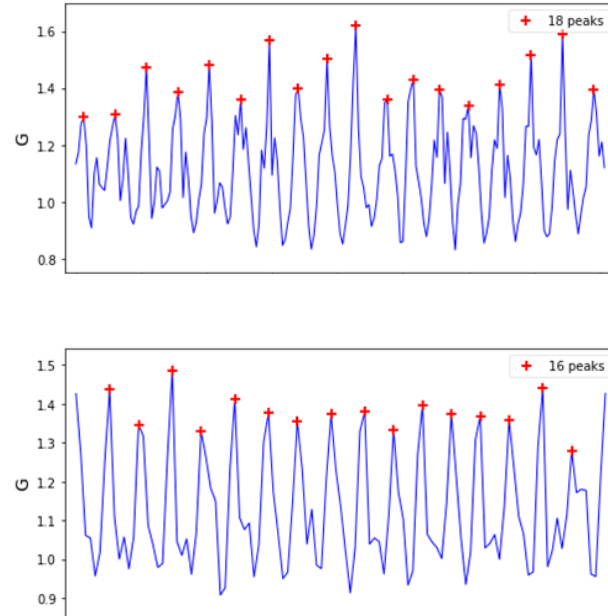


Figure 3.5: Two Examples Misclassified by Benchmark Models (Chest Location)

3.6. Discussion and Conclusions

Mobile technologies have been leading a new round of innovation and reformation in the IS community (Chen et al. 2012). Mobile sensing technologies provide us an always-on option to obtain health data from individuals via smartphones and wearable sensors. With the vast amount of data that we are potentially able to collect, however, prior studies on motion sensor-based health risk prediction (e.g., Wang et al. 2017; Silva and Sousa 2016; Greene et al. 2017) focused on manual feature engineering, which is laborious, ad hoc, and inconclusive. With the development and success of deep learning on image and speech recognition tasks, we were intrigued by the possibility of applying deep learning techniques on motion sensor data as a feature engineering-less and generalizable approach. In this work, we developed a convolutional neural network (CNN) to automatically extract features from raw sensor signals. We further designed a novel two-dimensional heterogeneous (2D-hetero) architecture to extract cross-axial and cross-locational features, which specializes and optimizes in gait and balance assessment for health risk prediction. To demonstrate the effectiveness of our proposed 2D-hetero CNN, we collected motion sensor data from Parkinson's patients as our research base, and chose fall risks as the prediction task. Our experimental results showed that the 2D-hetero CNN significantly and consistently outperformed the competitive benchmark models in terms of F-measure. The case studies illustrated the potential advantage of applying deep learning approaches to identify patients with high risks that are indistinguishable by human looking into the signals. Our proposed model further enables precise automatic risk alert without human involvement, which can potentially lead to early detection of disease signals and corresponding treatment.

3.6.1. Contributions to IS Knowledge Base

Our study makes multiple unique contributions to the IS discipline. We summarize the contributions of our work as an IT artifact as well as diverse design guidelines for future research.

IT Artifact for Health Predictive Analytics

Health predictive analytics is a topic receiving increasing interest in the IS community (e.g., Lin et al. 2017; Bardhan et al. 2014). Our study advances health predictive analytics research by proposing the 2D-hetero CNN, a deep learning model that enables health risk prediction based on motion sensor data. To the best of our knowledge, the intersection of health predictive analytics and motion sensor technology is an exciting but underexplored research area, both inside and outside the IS community. Methodologically, we contribute to deep learning by proposing a two-dimensional heterogeneous convolutional neural network (2D-hetero CNN) for fall risk prediction, which differs from and outperforms existing 1D or 2D CNN architectures. By designing cross-axial and cross-locational convolution stages, the 2D-hetero CNN is able to extract both cross-axial (e.g., x -, y -, and z -axes in a sensor) and cross-locational features (e.g., left and right parts of the human body) that are meaningful and effective for human gait and balance assessment. Our model serves as an IT artifact for advanced clinical decision support, which can assist healthcare providers and senior citizens in better assessing their risks for more prompt and personalized care.

Generalizable Design Principles

Design science contributes to the IS knowledge base not only in the form of IT artifact, but also as a search process (Hevner et al. 2014). Based on our extensive experiments, comparisons, and robustness checks, we discuss our generalizable design principles for future research from the following three perspectives: data and application characteristics, deep learning method selection, and model building suggestions.

Data and application characteristics. Sensor data is generated in an inherently high-velocity manner. For example, a sensor with a regular sampling rate of 25 Hz can lead to 2 million records per day. This characteristic prevents classic statistical machine learning methods from being directly applied on the raw sensor data due to the excessive variable dimensionality (a one-minute snippet can contain 1,500 ($= 25 \text{ per second} \times 60 \text{ seconds}$) independent variables). Deep learning serves as an effective tool of representation learning (Goodfellow et al. 2016), i.e., dimension reduction with salient features, without human intervention. When multiple motion sensors are attached to a human body to investigate human gait and balance patterns, it is of greater interest to focus on how the left part of a human body is coordinating with the right part (Hubble et al. 2015), i.e., how the data collected one sensor is interacting with another particular sensor. Our proposed 2D-hetero CNN is specifically designed to tackle this genre of problems. Meanwhile, benefiting from its feature engineering-less design, no disease-specific domain knowledge has been explicitly encoded in the 2D-hetero CNN. This enables the model to generalize and transfer to health risk prediction tasks other than falling, such as PD, dementia, frailty, etc., by altering the output labels from the risk of falling to the risk of PD/dementia/frailty, and altering the input to the data collected from desirable mobility tests (e.g., timed up and go test). The application domains can be further

extended to any prediction or assessment tasks that involve multiple motion sensors to capture human motion characteristics, such as athlete performance assessment or rehabilitation therapy assessment.

Deep learning method selection. Deep learning comprises of a family of multilayer backpropagation neural networks, including multilayer perceptron (MLP), recurrent neural networks (RNN), and convolutional neural networks (CNN), among others. As noted by Goodfellow et al. (2016),

“CNNs are a specialized kind of neural network for processing data that has a known grid-like topology. Examples include time-series data, which can be thought of as a 1-D grid taking samples at regular time intervals, and image data, which can be thought of as a 2-D grid of pixels.”

CNNs are suitable for modeling motion sensor data because the multiple axes of sensor data form a 2-D grid with one dimension on temporal adjacencies and the other on cross-axial and cross-locational adjacencies. Although RNNs have an inherent ability of modeling time series data, they are limited in representation learning and automatic feature extraction that are critical for processing the high-velocity, streaming motion sensor data. Our experimental results confirmed the advantage of CNNs over RNNs on motion sensor data, which echoes prior studies (Münzner et al. 2017; Hammerla et al. 2016).

Model building suggestions. Deep learning models including CNNs introduce flexibility in choosing the network architecture and hyperparameters. We followed the guidelines suggested by Goodfellow et al. (2016) in stacking multiple sets of a convolution layer, a non-linear (detector) layer, and a pooling layer to extract both local and global features embedded in motion sensor data. In addition, we conducted extensive robustness checks to

identify the optimal model designs and hyperparameters in our context. Results showed that adding fully connected layers and pooling layers do have positive impacts on model performance, while convolution lengths do not play a significant role. Choosing ReLU as the activation function outperforms alternatives including sigmoid and tanh, which is consistent with common practices in deep learning (LeCun et al. 2015). In addition, a moderate dropout rate is beneficial in avoiding model overfitting while ensures model convergence. We hope our above exploration can guide more rapid and effective model building in future research.

3.6.2. Practical Implications

Healthcare is inherently becoming mobile and ubiquitous. With the advancement of sensing technologies, people are able to collect and record their motion and physiological data on a daily basis. This creates a unique opportunity for people, especially senior citizens, to track themselves more precisely and promptly. Physicians, senior citizens, and their families can all benefit from a more comprehensive and timely health predictive model such as 2D-hetero CNN. We discuss key practical implications for those stakeholders in the following subsections.

Physicians. The 2D-hetero CNN model is a tool of advanced clinical decision support at the point of care for physicians. As illustrated in the case studies, although physicians are highly trained professionals, there are circumstances where the sensor signals are difficult for human eyes to distinguish. With the assistance of advanced analytics models such as 2D-hetero CNN, physicians can identify patients with high risks more accurately and provide more timely interventions, leading to a reduction of errors and delays.

Senior citizens and families. The concentration of healthcare has been moving from hospitals to homes. The 2D-hetero CNN model enables the possibility of home monitoring systems that allow senior citizens and their families to assess their health risks by themselves. The implications of such systems are two-fold. First, for those individuals who are patients with certain chronic conditions, e.g., Parkinson’s disease, the systems can provide real-time risk assessments, which can indicate the progression of the condition or the effect of drugs or therapies. Second, and perhaps more importantly, for those individuals who are not patients, the systems can predict the risks of certain conditions or events (e.g., falls). This can potentially guide senior citizens and their families to take steps for preventive solutions, as “prevention is better than cure.”

3.6.3. Limitations and Future Research

As with any other scientific research in an emerging field, there are two important limitations in our work. First, only mobility test data have been used in training the predictive model. It would be interesting to further explore how other types of data can be integrated into the framework (e.g., demographical data, physiological data, medical imaging data, etc.), as the cross-axial and cross-locational convolutions may not be directly applicable in such situations. Second, the approach has been verified only on Parkinson’s disease patients in predicting fall risks. Although cross-validation has been conducted to demonstrate the effectiveness and utility of our model, it is still possible that the performance of 2D-hetero CNN is limited to the dataset. Researchers are encouraged to experiment with the 2D-hetero CNN approach on other conditions or adverse events. Despite the above limitations, this study pioneers the application of deep learning approaches on motion sensor data for health risk

prediction. We believe that this principled approach can further increase the accuracy of risk prediction and create a better living environment and quality of life for senior citizens.

4. ESSAY III: MOTION SENSOR-BASED HEALTH CONDITION RISK AND SEVERITY ASSESSMENT: DEEP MULTISOURCE MULTITASK LEARNING FOR MOBILE HEALTH ANALYTICS

4.1. Introduction and Essay Structure

Recent years have witnessed a significant increase in life expectancy in the United States (79.3 years in 2016, WHO 2016) due to healthy lifestyles and advances in medicine and health services. Meanwhile, the aging population has also become a growing societal concern, with 47.8 million U.S. citizens (14.9 percent of the total population) aged 65 years old or older in 2015 (U.S. Census Bureau 2017). Falls, a common type of adverse event experienced by senior citizens, are receiving increasingly more practical and academic attention due to the fact that approximately 30 percent of this population falls each year (Bergen et al. 2016). Falls threaten senior citizens' living both physically and psychologically. In 2015, there were more than 28,000 fatal falls and the estimated costs of fatal and nonfatal falls combined totaled approximately \$50 billion (Florence et al. 2018). Among the various chronic conditions, Parkinson's disease (PD) is the second most common neurodegenerative disorder in the U.S. with symptoms of tremor, rigidity, bradykinesia, and postural instability, costing \$15.5 billion per year (Gooch et al. 2017). Postural instability is generally not improved with medical treatment and thus patients at all stages of PD remain at risk for falls even though other motor symptoms can be largely alleviated. Additional factors, such as cognition and higher level motor planning, may also contribute to fall risk. This complex interplay is not easily assessed by healthcare providers, and even studies that use combinations of disease-specific and functional tests are not very successful (Kerr et al. 2010). Assessing patients' PD severity and which patients are at risk for a fall will require an innovative approach. Such an approach will

allow proactive interventions such as physical and occupational therapy to prevent falls and alleviate the progression of PD, as well as provide insight into general mobility-related health condition management.

In current clinical practice, health professionals conduct surveys, examinations, or timed mobility tests with patients to identify risk factors for PD and other related health conditions. However, such approaches face the following limitations. Surveys rely on patients' recall of recent events (e.g., "over the past week, do you usually have trouble turning over in bed?"), which may include omissions and mistakes. Examinations rely on health professionals' subjective assessment of patients' health and activity status and thus may not be comparable across different health professionals. The assessments with timed mobility tests are typically associated with completion time thresholds. For instance, Shumway-Cook et al. (2000) suggested that for a timed up and go (TUG) test, times that exceed 14 seconds indicate increased fall risk for community dwelling elderly without neurological disorders. Such assessments may oversimplify the complexity of body motion and may not be sufficient to provide a comprehensive analysis.

With the development of mobile sensing technologies, wearable sensors equipped with accelerometers have been widely accepted by researchers and practitioners to address the above concerns (Wang et al. 2017; Greene et al. 2017). Such sensors can capture quantitative mobility data with high sensitivity (precision up to $1/4096$ G, where 1 G equals 9.81 m/s^2) and high granularity (sampling frequency up to 100 Hz, equal to 100 data points per second). They can be attached to a senior citizen's body in a clinical or home setting to collect an immense amount of detailed data (millions of data points per day). With a data analytics engine, the motion sensor data can be analyzed to assess health condition risks and severities, which

empowers senior citizens and their families by revealing early signals of adverse events or health condition progression without clinical visits.

In the past 15 years, health information technology (health IT) research has become a major focus of the Information Systems (IS) community (Baird et al. 2018). The topics of health IT research have been moving from traditional studies on adoption and use of IT artifacts in a healthcare context to analytical models for precise health predictions. With the availability of large-scale motion sensor signal data, a novel IT artifact is urgently needed to obtain insights from such data and to provide more precise, prompt, and personalized senior care. Following the computational design science paradigm (Rai 2017) and prior IS research on health analytics (Lin et al. 2017; Bardhan et al. 2015), we propose and evaluate a novel deep learning approach, Deep Multisource Multitask Learning for Mobile Health Analytics (DMML-MHA), in the context of mobile health analytics to assess senior citizens' risks and severities of health conditions and adverse events. By identifying the magnitude of risks and severities, medical interventions including exercise, assistive devices, review and modification of medication, and increased surveillance and care by caregivers (Weinstein and Booth 2006) can be provided to reduce or eliminate adverse events and alleviate the progression of adverse health conditions. There are two core novelties in this study. First, compared to prior studies relying on feature engineering, our proposed DMML-MHA approach avoids subjective and laborious feature engineering and creates a featureless solution grounded in deep learning, which brings value in more generalizable and transferable health condition risk and severity assessment. Second, we are among the first who incorporate both deep multisource learning and deep multitask learning in an integrated framework for mobile health analytics. We design source-specific, general, and task-specific deep learning layers to improve the effectiveness of features

automatically extracted from multiple data sources of motion sensor signals, which leads to improved assessment results.

We choose PD and falls as our research cases due to their pervasiveness and societal significance. We collected two datasets for internal and external evaluations on our proposed framework. We used five SilverLink motion sensors (a National Science Foundation (NSF)-funded system previously developed by the University of Arizona) attached to PD patients' chest, left and right thighs, and left and right feet to collect motion data from controlled clinical experiments and form our internal dataset. We also obtained an external, publicly available dataset, the mPower dataset (Bot et al. 2016), which contains iPhone motion sensor data collected from PD patients. Consistent with the computational design science guidelines, we rigorously evaluate the DMML-MHA framework against a series of state-of-the-art benchmark models. In addition to the designed IT artifacts, this work provides design principles that can in part guide future design and research in the areas of health analytics, big data, and multitask learning.

The rest of the essay is organized as follows. First, we review the existing literature on IS works in health IT, computational design science guidelines, motion sensor-based health condition risk and severity assessment, deep learning, and deep multisource learning and deep multitask learning in the health contexts. Second, we present our DMML-MHA framework for health condition risk and severity assessment and its two instantiations on the two datasets. Third, we set up benchmark models and discuss the experiment results. Finally, we discuss the contributions of this study to the IS knowledge base, its practical implications, and future research directions.

4.2. Literature Review

Our research is guided by the following five streams of literature: (1) current IS research on health information technology to inform the need for mobile health analytics, (2) computational design science guidelines that guide our IT artifact design, (3) scenarios of and tests conducted in prior studies on motion sensor-based health condition risk and severity assessment, (4) deep learning as a computational analytical model, and (5) multitask learning and multisource learning in a deep learning context and their applications. We then discuss the research gaps and questions.

4.2.1. *IS Research on Health Information Technology*

As quoted in a research curation in *MIS Quarterly* (Baird et al. 2018), health information technology (health IT, defined as “a broad concept that encompasses an array of technologies to store, share, and analyze health information”) research is conducted at the critical intersection between societies, organizations, and consumers. Among the 41 articles selected by this research curation over the 15 years from 2003 to 2018, three time periods have been identified with a progression of research themes. In the period prior to 2007, much of the health IT research evaluated traditional IT artifacts in a healthcare context, e.g., group support systems (Dennis and Garfield 2003), relationship between IT and customer service (Ray et al. 2005), and application integration used to address fragmentation of specialized knowledge (Mitchell 2006). Essentially, many of the primary terms and concepts during this time period were consistent with IS research conducted in traditional organizational contexts. In the 2007 to 2012 period, due to the government- and policy-driven growth in health IT markets, the research focus moved to explicitly considering health IT artifacts and their impacts, e.g.,

consumer decision-making processes associated with health IT use (Angst and Agarwal 2009) and health IT investment decision making and governance (Xue et al. 2008). More diverse topics of interest have been observed in the mainstream IS community since 2013. In addition to traditional research streams examining the adoption and use of IT artifacts (e.g., how the use of health information impacts outcomes such as duplicate testing (Ayabakan et al. 2017) and how consumers are playing a role in impacting perceptions of medical provider quality (Gao et al. 2015)), analytical models are being applied in health IT contexts to identify more salient influencing factors or provide more precise health predictions. Examples of such analytical models include growth-mixture models (Angst et al. 2017a), sequence analysis (Angst et al. 2017b), and Bayesian multitask learning models (Lin et al. 2017). In the current literature, however, analytical approaches have only been adopted in health data analytics to predict adverse events (Lin et al. 2017), hospital readmissions (Bardhan et al. 2015), or patient mortality (Tabak et al. 2014), among others. Mobile health analytics, which applies health analytics in a mobile or sensor context, remains an underexplored but increasingly high-impact area.

Mobile and sensor analytics are attracting more attention in the mainstream IS community since the number of mobile devices surpassed PCs in 2011 (Chen et al. 2012). However, recent interest in mobile and sensor analytics mostly focused on mobile Internet and Web browsing activities (Adipat et al. 2011; Ghose et al. 2012), mobile app design and behavior (Hoehle and Venkatesh 2015; Kwon et al. 2016; Sun et al. 2017), and customer service innovation (Venkatesh et al. 2017; Ye and Kankanhalli 2018). Mobile health analytics in the IS community is still a nascent domain, which critically requires innovative IT artifacts

designed to address practical problems for relevant stakeholders including patients, their families, and health professionals.

4.2.2. Computational Design Science Guidelines

A systematic approach is required to develop effective IT artifacts (e.g., constructs, models, methods, and instantiations) in mobile health analytics, which can be guided by the design science paradigm (Hevner et al. 2004). Rai (2017) summarized the rapidly evolving realm of design science in four genres: computational, optimization, representation, and economics. Among the four, the computational aspect of design science addresses the design and development of novel algorithms, computational models, and systems to resolve practical and emerging business problems. Due to the nature of mobile health analytics where millions of sensor records can be generated per person per day, a computational analytical aspect is needed to discover knowledge from the immense amount of data.

Three critical guidelines can be distilled from the computational aspect of the design science paradigm. First, the design of the IT artifact can be informed by domain knowledge and data characteristics, where a relevant mature theory is often absent due to the nascence of the problem domain. Second, the novelty and utility of the designed artifact has to be demonstrated by comprehensive evaluations vis-à-vis selected state-of-the-art benchmark approaches. Third, the artifact's design should contribute to the IS knowledge base (Rai 2017) through situated implementations of the artifact (e.g., software or system), nascent design theory (e.g., design principles), and other forms (Gregor and Hevner 2013). For example, Lin et al. (2017) incorporated these three guidelines by developing Bayesian multitask learning on electronic health record (EHR) data to predict critical hospital adverse events.

Consistent with the emerging data analytics trend, our research aims to follow the computational design science paradigm to design novel and impactful IT artifacts for mobile health analytics. However, it is necessary for us to first obtain a firm understanding of existing motion sensor-based analytics approaches for health condition risk and severity assessment.

4.2.3. Motion Sensor-Based Health Condition Risk and Severity Assessment

Motion sensors have been used for health condition risk and severity assessment in various contexts, including PD (Hubble et al. 2015; Weiss et al. 2014), diabetes (Watanabe et al. 2017; Najafi et al. 2013), frailty (Millor et al. 2017; Schwenk et al. 2015), and falls (Wang et al. 2017; Silva and Sousa 2016; Van Schooten et al. 2015). The objective of assessing health condition risk and severity is to obtain a risk or severity score for the health conditions of interest, such that proper interventions can be provided to alleviate the progression of adverse health conditions or eliminate the future occurrence of adverse events. Two major scenarios have been identified for health condition risk and severity assessment: clinical setting and home setting. In a clinical setting, motion sensors are attached to patients to collect data from controlled mobility tests instructed by health professionals. Limited by the time constraints of clinical visits, clinical tests can only provide infrequent snapshots of patients' mobility patterns. However, health professionals and researchers are able to observe the full process of clinical tests, ensuring that test criteria are met and necessary annotations are recorded. Using multiple sensors is often preferred to collect more comprehensive data across multiple locations on the human body (e.g., chest, waist, leg, ankle, etc.). Sensor locations and orientations are usually fixed and consistent across tests to ensure data comparability. Consequently, the motion sensor data collected from clinical tests are small in volume and relatively short-term, but more

structured and comprehensive. In contrast, motion sensor data collection in a home setting requires participants to wear motion sensors while performing regular daily activities (e.g., sitting, walking, lying, etc.) (Silva and Sousa 2016; Greene et al. 2017; Schwenk et al. 2014) or unsupervised mobility tests. Although the assessment time period can be several days or months, the data acquired is often less structured with lower quality as the performed tests may not exactly conform to the standards of clinical tests.

Typical mobility tests include timed up and go (TUG) test (Silva and Sousa, 2016; Greene et al. 2017; Najafi et al. 2013; Schwenk et al. 2014) and ground walking test (Watanabe et al. 2017; Millor et al. 2017; Wang et al. 2017). In a TUG test, the participant is asked to sit on a chair and perform the following instructions in a continuous manner: (1) stand up, (2) walk straight for 3 meters, (3) turn around, (4) walk straight for 3 meters, (5) turn around, and (6) sit down on the chair. In a ground walking test, the participant is asked to walk straight at his/her comfortable pace for 10 meters or 20 steps. The TUG test is considered more comprehensive in reflecting the participant's mobility as it includes activities other than walking, while the ground walking test requires less effort to conduct and is more transferable from clinical tests to home living settings. However, current clinical practice for mobility tests typically uses the completion time as the only criterion to assess the participant's health condition risk or severity. For instance, Shumway-Cook et al. (2000) suggested that for a TUG test, times that exceed 14 seconds indicate increased fall risk for community dwelling elderly without neurological disorders. Due to the large amount of data points collected from motion sensors, a computational analytical model is critically needed for more fine-grained and comprehensive health condition risk and severity assessment.

4.2.4. Computational Analytical Model: A Deep Learning Approach

For general health condition risk and severity assessment, a computational analytical model accepts motion sensor data collected from mobility tests as the input and generates a numerical or categorical value as the estimated level of health condition risk (e.g., fall risk: high or low) or severity (e.g., PD severity represented by a metric score). In a formal representation, the computational analytical model aims at finding a mapping, f ,

$$f: X \rightarrow \mathbf{y},$$

where $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i]$ is the input, i.e., motion sensor data collected from mobility tests, and $\mathbf{y} = [y_1, y_2, \dots, y_i]$ is the output, i.e., estimated levels of health condition risk or severity. This formulation is consistent with prior predictive analytics studies in a health context (Lin et al. 2017).

Linear models are the most intuitive and succinct models for choosing f , i.e.,

$$\hat{y} = f(\mathbf{x}) = \mathbf{w}^T \mathbf{x},$$

where \mathbf{w} is the vector of related parameters and \hat{y} is the estimated output. Although linear models are intuitive, they are unable to represent complex, nonlinear relations between \mathbf{x} and y , which may be present in many applications (Goodfellow et al. 2016). To represent nonlinear functions of \mathbf{x} in the model, we can apply the linear function not to \mathbf{x} itself, but to a transformed input $\phi(\mathbf{x})$, where ϕ is a nonlinear transformation. Three directions for choosing ϕ have been explored in literature. The first direction applies a generic ϕ on \mathbf{x} (e.g., infinite-dimensional ϕ used by the radial basis function (RBF) kernel in SVMs (Chang et al. 2010; Vert et al. 2004)). Although arbitrary complexity can be modeled, its generalizability often remains unsatisfactory due to its tendency to overestimate the complexity of the given problem. The second direction leverages domain knowledge on the design of ϕ . Also known

as feature engineering, this process has been adopted in many classical machine learning approaches. Three common categories of features have been applied in prior health studies related to motion sensor applications: (1) general signal processing features (e.g., signal mean, standard deviation, maximum, mean-crossing rate, fast Fourier transform (FFT) coefficients, etc.) (Silva and Sousa 2016); (2) gait features that reflect walking patterns (e.g., walking speed, stride time, stride length, gait variability, gait symmetry, etc.) (Van Schooten et al. 2015); (3) physical activity features that reflect activities of daily living (e.g., durations of walking, standing, sitting, and lying, etc.) (Van Schooten et al. 2015; Schwenk et al. 2014). Despite the maturity of these categories, scholars have noted that feature engineering requires extensive human labor and is often inconsistent across researchers (Hubble et al. 2015). The third direction aims to address the drawbacks of the first two directions by learning ϕ with domain experts' knowledge as the guidance to constrain the function family of ϕ . In this direction, the model is rewritten as:

$$y = f(\mathbf{x}; \boldsymbol{\theta}, \mathbf{w}) = \mathbf{w}^T \phi(\mathbf{x}; \boldsymbol{\theta}),$$

where $\boldsymbol{\theta}$ is a set of parameters learned by the algorithm to represent $\phi(\mathbf{x}; \boldsymbol{\theta})$. The advantage of this approach is that domain knowledge is considered in the design of features $\phi(\mathbf{x}; \boldsymbol{\theta})$ without specifically engineering the exact features $\phi(\mathbf{x})$. The flexibility and generalizability of the third direction has been demonstrated by modern deep learning algorithms in complex image and speech recognition tasks, which also provide insights into pattern recognition tasks for motion sensors (LeCun et al. 2015).

Deep learning is a significant and emerging branch of machine learning that originates from the Artificial Neural Network (ANN). Contrast to the “shallow” three-layer ANN, deep learning incorporates multiple layers and specialized architectures to achieve effective learning

and generalizability. Advancements in deep learning have yielded complex ANN architectures, such as convolutional neural networks (CNN) and recurrent neural networks (RNN). Such architectures have been developed to address complex emerging problems in image recognition (CNN) and natural language processing (RNN) (LeCun et al. 2015). Essentially, deep learning algorithms rely on the back-propagation error-correction procedure to learn the function mapping from the input to the desired output. In each pass, a data point is fed forward in the deep learning structure. The training error is then back-propagated throughout the structure to adjust the weights connecting the layers to the direction that reduces the error with optimization algorithms such as stochastic gradient descent (SGD) (Goodfellow et al. 2016). After achieving convergence, the neurons in the deep learning structure act as hidden features that predict the desired output. With a properly designed deep learning structure, $\phi(\mathbf{x}; \boldsymbol{\theta})$ is trained to represent a particular function family relevant to the problem domain. With the training process, $\boldsymbol{\theta}$ is learned automatically without human intervention, significantly reducing labor cost without losing prediction accuracy. However, classical deep learning only generates one predicted value as the output and accepts one data source as the input, which is recognized as single task and single source learning. Consistent with our research goal of incorporating multiple existing data sources and multiple mobility assessment tasks to provide more comprehensive and accurate assessment of health conditions (such as PD), we review deep multitask learning and deep multisource learning as a promising future direction in the following subsection.

4.2.5. Deep Multitask Learning and Deep Multisource Learning

Multitask learning is a machine learning strategy in which multiple related tasks are trained jointly to improve learning performance and model generalizability (Caruana 1998). There is a shared model structure to tie individual tasks together in a unified training process (e.g., simultaneously assess PD severity and fall risk). There are three general approaches for multitask learning (Lin et al. 2017): regularization-based, Bayesian framework-based, and ANN-based multitask learning. The first approach applies a regularization function over the tasks. This approach can be adopted across multiple underlying machine learning algorithms, e.g., regressions (Huang et al. 2012), support vector machines (SVM) (Cai and Cherkassky 2012), tree-based models (Simm et al. 2014), etc. The second approach imposes common prior distributions over tasks in a Bayesian framework (Lin et al. 2017; Archambeau et al. 2011). The parameters for task models are correlated based on prior distributions to transfer information. This approach is flexible for baseline machine learning models that have Bayesian representations to share information between tasks (e.g., ANNs, SVMs, decision trees, regressions, etc. (Lin et al. 2017)). The third approach uses ANNs with each output node configured as an individual task (Bakker and Heskes 2003; Caruana 1998). Hidden layers are shared in the ANN architectures to achieve multitask learning.

Among the above three approaches, however, prior mobile health analytics studies show that manual feature engineering is necessary before applying classical machine learning algorithms such as regressions, SVM, tree-based models, etc. (Wang et al. 2017; Silva and Sousa 2016; Van Schooten et al. 2015). This is because the number of independent variables in raw sensor signal data may be very large. If we consider a single tri-axial motion sensor with its sampling frequency set to 25 Hz, recording 20 seconds of motion data is equivalent to

generating a sample signal snippet with 1,500 ($= 3 \times 25 \times 20$) independent variables, while the number of data samples may be a magnitude less. This fact prevents classical machine learning algorithms in regularization-based and Bayesian framework-based multitask learning from working properly and introduces the necessity of feature engineering to reduce the number of independent variables. With the recent advancement in deep learning, ANN-based multitask learning is a promising approach for mobile health analytics because feature representations are automatically extracted from sensor signal data in deep learning architectures such as CNNs, as discussed in Section 4.2.4.

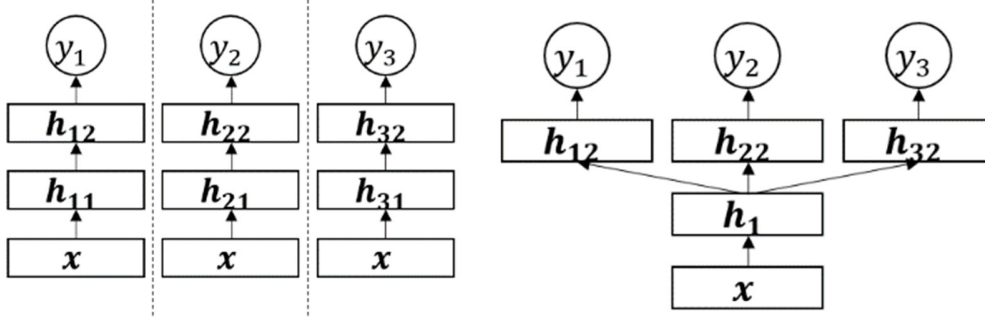


Figure 4.1: (Left) Single Task Learning; (Right) Multitask Learning

Compared to deep single task learning with task-specific layers only, deep multitask learning contains both general and task-specific layers in its architecture. Figure 4.1 (Left) shows a simple but typical single task learning architecture. The hidden layers, h_{ij} ($i = 1, 2, 3$, $j = 1, 2$) are trained separately for three different tasks, e.g., y_1 , y_2 , and y_3 . Figure 4.1 (Right) shows a multitask learning architecture with the same three tasks. In addition to the task-specific layers, h_{i2} ($i = 1, 2, 3$), the general layer h_1 acts as the shared model structure to tie the individual tasks together. h_{i2} represents the features specific to each task, while h_1 represents the general features across all tasks. This shared structure forces the model to learn

general features (\mathbf{h}_1), thus alleviating overfitting and improving the performance for each task (Goodfellow et al. 2016).

The motivation for multisource learning is based on data and sensor fusion. People often make decisions based on multiple data sources. For instance, when making a diagnosis, health professionals rely on multiple radiological images (MRI, CT, ultrasound, etc.). When judging human mobility, health professionals typically assess multiple types of activities (e.g., walking, standing up, turning around, etc.). Each type of activity has its unique features, while there can be high-level general features across multiple types of activities. For instance, tremors are common across multiple types of activities for some PD patients. A deep multisource learning architecture can be designed to capture specific features as well as general features to improve learning performance.

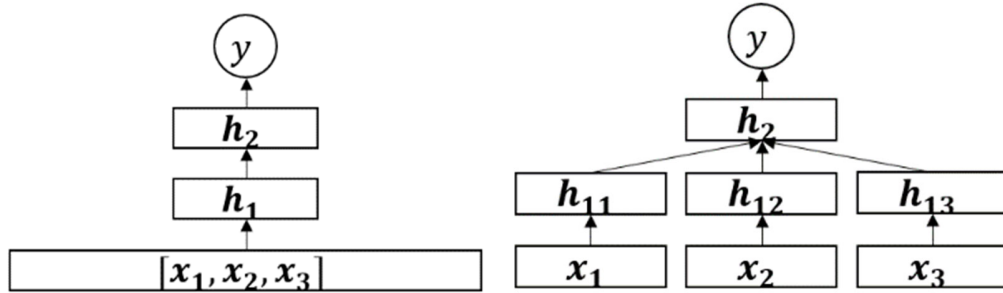


Figure 4.2: (Left) Single Source Learning; (Right) Multisource Learning

Assume we have one task, y , to learn. As illustrated in Figure 2 (Left), a single source learning architecture simply concatenates the three sources of data, \mathbf{x}_1 , \mathbf{x}_2 , \mathbf{x}_3 . Only general features across the three sources (\mathbf{h}_1 , \mathbf{h}_2) can be extracted. In contrast, in a multisource learning architecture illustrated in Figure 2 (Right), hidden layers \mathbf{h}_{1j} ($j = 1, 2, 3$) specific to source \mathbf{x}_j are designed to extract source-specific features. Similar to multitask learning, the shared

structure forces the model to learn general features (h_2) while preserving specific features (h_{1j}), thus alleviating overfitting and improving overall model performance.

Current applications of deep multitask learning and deep multisource learning include radiology on biomedical cell tracking (He et al. 2017), breast cancer diagnosis (Samala et al. 2017), Alzheimer's disease diagnosis (Suk et al. 2016; Xiang et al. 2013; Yuan et al. 2012), human pose estimation (Fan et al. 2015; Ouyang et al. 2014), and image or object recognition (Yu et al. 2017; Zhang et al. 2017; Huang et al. 2017). Most of these studies incorporated CNN for feature extraction from sensor data. However, it is unclear how deep multitask learning and deep multisource learning can be applied on motion sensor data to assess health condition risks and severities. In addition, more research is needed on how deep multitask learning and deep multisource learning can work together in constructing an integrated and enhanced model.

4.3. Research Gaps and Questions

Based on the above literature review on IS research in health IT, motion sensor-based health condition risk and severity assessment, and deep multitask learning and deep multisource learning, we have identified the following research gaps. First, past motion sensor-based health condition risk and severity assessment studies employ a feature engineering approach for their context. While valuable, this process is laborious, ad hoc, and the features extracted do not apply to different health conditions. Second, although deep single task learning and deep single source learning are capable in automatically extracting salient features, they only generate one predicted value as the output and accept one data source as the input. These restrictions limit their applicability on multiple data sources and multiple assessment tasks. Third, past studies on deep multitask learning and deep multisource learning focus

primarily on radiology and image recognition. To the best of our knowledge, we are unaware of any framework that integrates both deep multitask learning and deep multisource learning on motion sensor data to improve health condition risk and severity assessment. Based on the above research gaps, we ask the following research questions:

1. How can we utilize deep learning techniques to improve motion sensor-based health condition risk and severity assessment?
2. How can we leverage multiple data sources to provide more comprehensive health condition risk and severity assessment?
3. How can we design a deep learning framework that integrates multisource and multitask learning for mobile health analytics?

4.4. Research Design

Guided by prior studies and research gaps in motion sensor-based health condition risk and severity assessment, deep multitask learning, and deep multisource learning, we propose a novel Deep Multisource Multitask Learning for Mobile Health Analytics (DMML-MHA) framework (Figure 4.3). The main novelty of our proposed DMML-MHA framework is that it integrates multisource and multitask learning in a deep learning architecture specialized for featureless and transferable mobile health analytics. We discuss data collection and preprocessing on two datasets, our proposed DMML-HMA framework and its instantiations, and evaluation by benchmark experiments in detail in the following subsections.

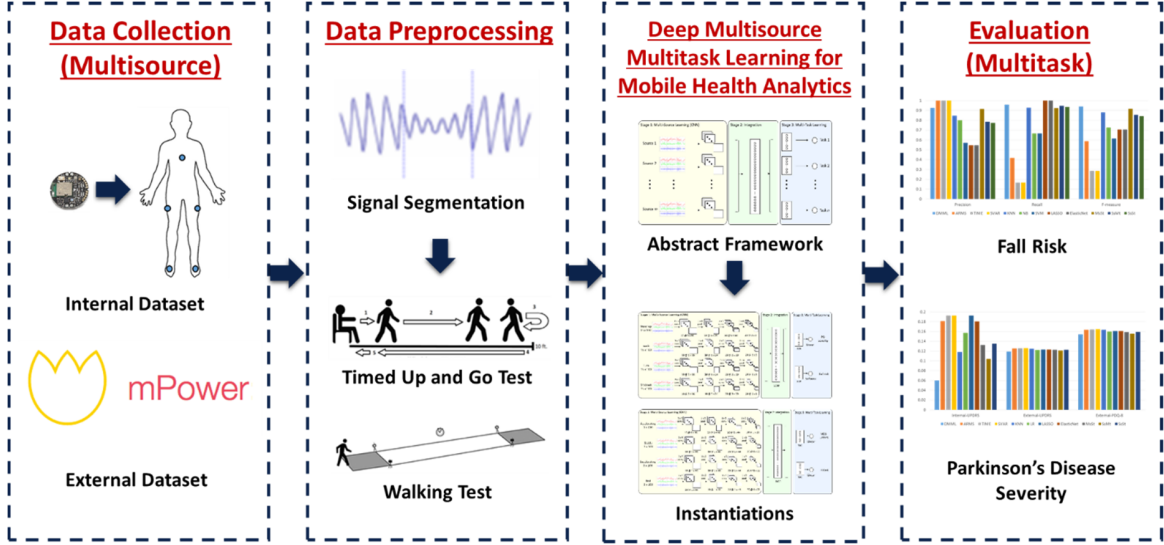


Figure 4.3: Research Design of Our Proposed the Deep Multisource Multitask Learning for Mobile Health Analytics (DMML-MHA) Framework

4.4.1. Data Collection and Preprocessing

We collected two datasets for internal and external evaluations on our proposed DMML-MHA framework. The first dataset is an internal clinical experiment dataset collected by motion sensors attached to PD patients' chest, left and right thighs, and left and right feet for timed up and go (TUG) tests. The second dataset is an external publicly available dataset, the mPower dataset (Bot et al. 2016), which contains iPhone motion sensor data collected from walking tests of participants with and without Parkinson's disease.

Internal Dataset: SilverLink

Funded by the National Science Foundation (NSF), SilverLink was a project developed by the Artificial Intelligence Lab at the University of Arizona (Maimoon et al. 2016; Chuang et al. 2015). The SilverLink project aimed to develop sensing technologies and health analytics to remotely manage senior care. We developed sensors and gateways for collecting data from the human body and the environment. Accelerometers are embedded in the sensors (inertial

unit model: Bosch BMI160). The shape and size of the sensor are illustrated in Figure 4.4 (Left).



Figure 4.4: (Left) Sensor Shape and Size; (Right) Sensor Locations (Frontal View)

We recruited 22 participants at Dr. Scott Sherman’s neurology clinic for our experiments. Consistent with guidelines provided in past studies, only those patients who could walk independently (autonomous walking aids allowed) were included as our participants (Najafi et al. 2013; Shumway-Cook et al. 2000). All participants are PD patients aged 65 or over. We use retrospective fall history as the criterion of fall risk (Silva and Sousa 2016; Ejupi et al. 2016; Greene et al. 2017). If a patient had any falls in the past 12 months, he/she is marked as a faller. 12 out of 22 participants met this criterion; thus they were labeled as fallers. The remaining 10 were denoted as non-fallers. Irrespective of category, we augmented the dataset by simulating 0, 90, 180, and 270-degree rotations on one or more of the sensor axes (x , y , and z), yielding a total of 1,408 signal snippets. This configuration ensures that the collected data is not constrained to one specific sensor orientation setting.

We attached 5 SilverLink sensors to each participant. The 5 sensors were set at chest, left and right thighs, and left and right feet to capture body and lower extremity motions (Wu et al. 2013), as shown in Figure 4 (Right). The sampling resolution was set to 12 bits, measuring acceleration signals between +4 G and -4 G (1 G equals 9.81 m/s^2) with a sampling frequency

of 25 Hz. We chose 25 Hz because it is the maximum possible sampling frequency for our gateway to communicate with 5 sensors simultaneously. We conducted TUG tests in a hallway without obstacles to collect motion data for gait and balance assessment. The participants got sensors attached, sat on a chair, and performed the following test components in a continuous manner: (1) stand up, (2) walk straight for 3 meters, (3) turn around, (4) walk straight for 3 meters, (5) turn around, and (6) sit down on the chair. We used components (1), (2), (3), and (6) to constitute the dataset of standing up, walking, turning around, and sitting down activities. The four activity components are the four data sources used in our proposed DMML-MHA framework.

The participants completed the TUG tests in different times, thus generating varying-length data inputs. However, our proposed model requires fixed-length data inputs. In past literature on activity recognition, Zeng et al. (2014) suggested segmenting sensor signals using a sliding window of 64 points, while Yang et al. (2015) suggested the window size to be 30 or 32. We believe that a wider window, e.g., 100 data points, is needed in our scenario of health condition risk and severity assessment to identify patterns spanning over multiple gait cycles. Therefore, we sampled the middle 100 points for each activity component to ensure capturing the performed activity.

External Dataset: mPower

mPower is an observational smartphone-based study developed using Apple's ResearchKit library to evaluate the feasibility of remotely collecting sensor data reflecting PD patients' health condition severities. The study was open to individuals diagnosed with and without PD. In terms of motion sensor data, participants were instructed to walk 20 steps in a

straight line (walking) and stand still for 30 seconds (standing). All participants were requested (although not required) to respond to standard surveys used for PD assessment, including Parkinson Disease Questionnaire 8 (PDQ-8) and a subset of questions from the Movement Disorder Society Universal Parkinson Disease Rating Scale (MDS-UPDRS) (Bot et al. 2016). All data are available upon request on Sage Bionetwork’s Synapse platform².

To attain a comprehensive and fine-grained analysis for the walking test, we divided the walking test into three states: accelerating state (first 100 data points), stable state (middle 100 data points), and decelerating state (last 100 data points). We also sampled the middle 100 data points from the standing test and denote them as the rest state. The four states (accelerating, stable, decelerating, and rest) are the four data sources used in our proposed DMML-MHA framework.

The mPower dataset contains 3,101 participants who completed the walking and standing test, 2,024 participants who responded to the MDS-UPDRS questionnaire, and 1,334 participants who responded to the PDQ-8 questionnaire. There were 506 participants who completed the test and responded to the two questionnaires, with 2,191 walking bouts that we used for our evaluations. The normalized total scores of the MDS-UPDRS and the PDQ-8 questionnaires linked to the corresponding participants are the two tasks assessed in our proposed DMML-MHA framework, which reflect PD severity from two different aspects.

² <https://www.synapse.org/>

4.4.2. The Deep Multisource Multitask Learning for Mobile Health Analytics (DMML-MHA)

Framework

Figure 4.5 shows an abstract illustration of our proposed DMML-MHA framework. The DMML-MHA framework is flexible and adaptive to the actual number of data sources (m) to learn from and the actual number of health tasks (n) to assess. Conceptually, there are three stages in the DMML-MHA framework. Stage 1 is designed for multisource learning. A separate CNN is applied on each data source to extract source-specific features from the m sources, as CNN excels in automatic signal feature extraction. Stage 2 is designed to learn general features across all the sources. Stage 3 is designed for multitask learning. In Stage 3, each task learns its specific features from the general features in Stage 2. By training the model jointly, we are able to extract source-specific, general, and task-specific features from data, thus improving model performance.

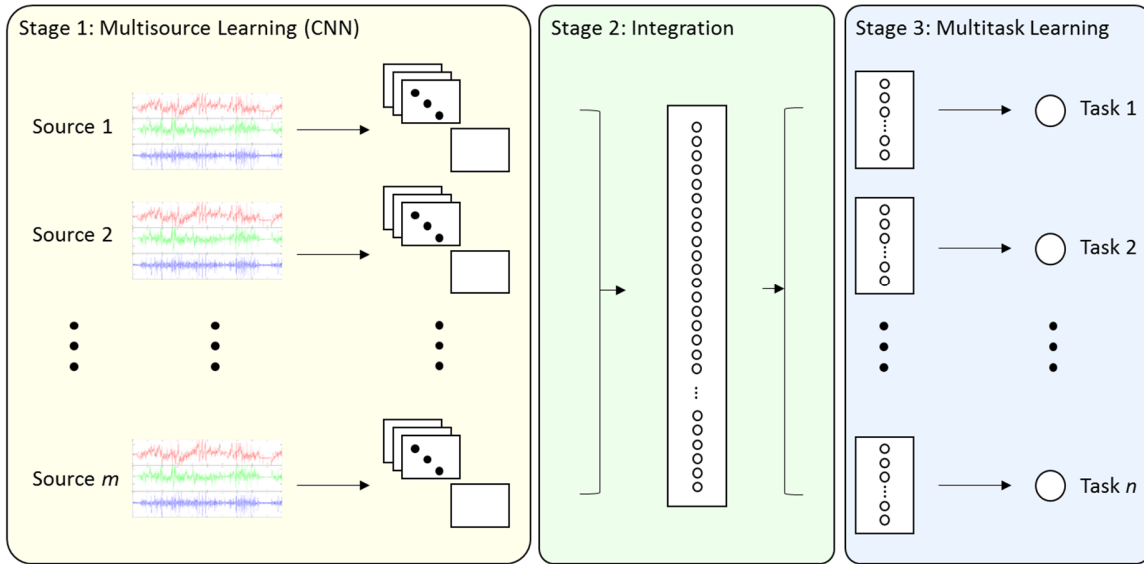
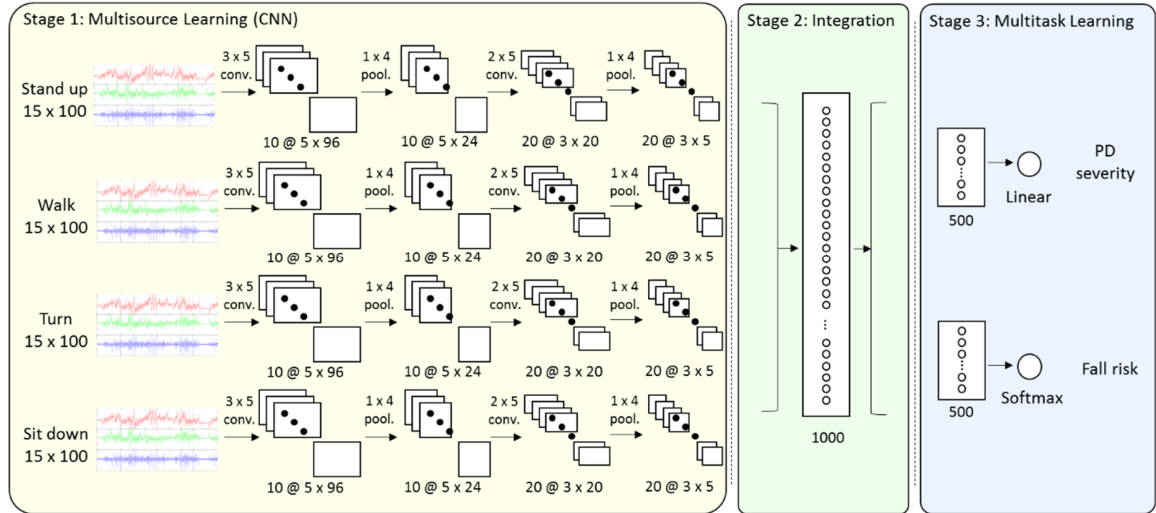


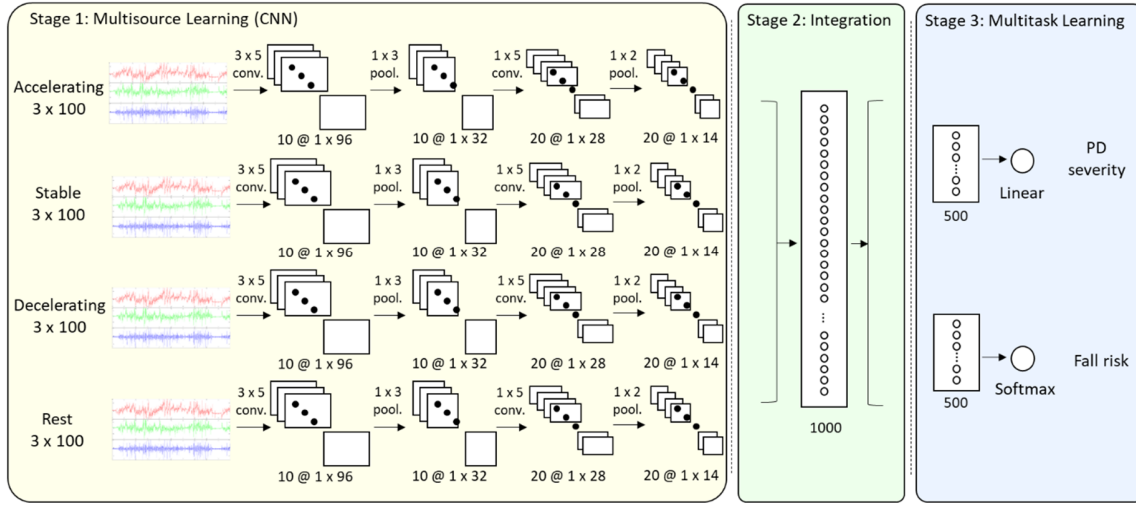
Figure 4.5: An Abstract DMML-MHA Framework

The actual instantiation of the DMML-MHA framework (m , n , and design of the deep learning layers) depends on the application and problem setting. Figures 4.6 and 4.7 are the detailed designs applicable to the internal SilverLink dataset and the external mPower dataset, respectively. In both designs, Stage 1 is constituted by multiple sets of the following three types of layers: convolution, nonlinear, and pooling layers. Nonlinear layers are omitted in Figures 6 and 7 due to space constraints. Stage 2 is constituted by fully connected layers. Stage 3 is constituted by a fully connected layer and an output layer. Stage 3 also contains loss functions that guide the neural network back-propagation and the model learning process. The model hyperparameters (model structure, numbers of neurons in each layer, etc.) are optimized through our experiments as suggested in state-of-the-art deep learning design guidelines (Goodfellow et al. 2016). Detailed algorithm designs for each layer follow below.



Note: conv. = convolutional layer; pool. = pooling layer. The notation “ $x @ y \times z$ ” denotes x feature maps with height y and width z .

Figure 4.6: The DMML-MHA Instantiation for the Internal SilverLink Dataset



Note: conv. = convolutional layer; pool. = pooling layer. The notation “ $x @ y \times z$ ” denotes x feature maps with height y and width z .

Figure 4.7. The DMML-MHA Instantiation for the External mPower Dataset

Convolutional Layers

A convolutional layer convolves the data input or the previous layer’s output with a set of convolution kernels. A convolution kernel is a matrix of weights applicable to small proportions of the input data. Compared to a general convolution kernel, a customized convolution kernel is preferred because an accelerometer records three orthogonal axes simultaneously, unlike other applications. It is semantically more meaningful to apply the convolution operation on the three axes from an identical sensor, rather than on one axis from one sensor and two axes from another. Therefore, we design a stride of 3 on the dimension of axes, with the following formula:

$$a_{x,y}^{i,j} = b^{i,j} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} w_{p,q}^{i,j,m} z_{3x+p,y+q}^{i-1,m},$$

where $a_{x,y}^{i,j}$ is the output value of this layer at row x and column y for feature map j in layer i , $b^{i,j}$ is the bias for feature map j in layer i , m is the number of feature maps in the previous layer, P_i and Q_i are the height and width of the convolution filter in layer i , $z_{3x+p,y+q}^{i-1,m}$ is the output from the previous layer (layer $i - 1$) for feature map m at row $3x + p$ and column $y + q$, and $w_{p,q}^{i,j,m}$ is the weight for position (p, q) pointing from feature map m in layer $i - 1$ to feature map j in layer i , which is invariant to the values of x or y and is to be learned through the training process.

Non-linear Layers

A rectified linear unit (ReLU) layer maps the output of the previous layer to generate the nonlinearity of the model by the following function:

$$f(a) = \max \{a, 0\}.$$

The ReLU function is the most widely used nonlinear function for CNNs due to its simplicity and effectiveness (LeCun et al. 2015).

Pooling Layers

In pooling layers, the resolution of feature maps is reduced to keep salient patterns and increase local scale invariance. A maximum pooling layer is added after each nonlinear layer axis by the following function:

$$a_{x,y}^{i,j} = \max_{\substack{0 \leq r \leq R_i-1 \\ 0 \leq s \leq S_i-1}} z_{x+r,y+s}^{i-1,j},$$

where $a_{x,y}^{i,j}$ is the output value of this layer at row x and column y for feature map j in layer i , R_i and S_i are the height and width of the pooling filter in layer i , and $z_{x+r,y+s}^{i-1,j}$ is the output

from the previous layer (layer $i - 1$) for feature map j at row $x + r$ and column $y + s$. The exact values of parameters can be found in Figures 6 and 7.

Fully Connected Layers

A fully connected layer first calculates the weighted sum of all outputs in its previous layer, then applies a sigmoid function to normalize the output to the range of -1 and 1:

$$a_j = \text{sigmoid}(\sum w_{ij}x_i),$$

where a_j is the j -th output of this layer, x_i is the i -th input of this layer, and w_{ij} is the weight pointing the i -th input to the j -th output.

Output Layers

An output layer links prior learning layers to a task to assess. A softmax layer is added if the task is a classification task (e.g., fall risk assessment: high risk or low risk):

$$a = \text{softmax}(\sum w_i x_i),$$

and a linear layer is added if the task asks for a numeric value (e.g., health condition severity assessment: MDS-UPDRS or PDQ-8 scores):

$$a = \sum w_i x_i,$$

where a is the output of this layer, x_i is the i -th input of this layer, and w_i is the weight pointing from the i -th input to the output.

Loss Functions

The selection of loss functions, $L(\cdot)$, depends on the nature of the tasks. A cross-entropy loss function is applied if the task is a classification task:

$$L(y, \hat{y}) = \text{cross_entropy}(y, \hat{y}),$$

and a mean squared error (MSE) loss function is applied if the task asks for a numeric value:

$$L(y, \hat{y}) = \text{MSE}(y, \hat{y}),$$

where y is the ground-truth output and \hat{y} is the output predicted by the model.

Following state-of-the-art deep learning practices (LeCun et al. 2015), we use dropouts on convolutional layers and fully connected layers to control overfitting. For each pass of data fed into the DMML-MHA, the dropout disables a random part of the neurons that participate in the current layer. The proportion of disabled neurons is known as the dropout rate. We tested our model for dropout rates from 0 to 0.5 and chose 0.25 as an optimal value between model convergence and avoiding overfitting. We also used early stopping as a regularization strategy.

4.4.3. Evaluation

Adhering to the computational design science paradigm guidelines, we systematically evaluated our proposed DMML-MHA framework for health condition risk and severity assessment with four sets of benchmark experiments. The first experiment evaluates whether DMML-MHA can outperform widely-applied single feature-based assessment models. The second experiment benchmarks the performance of DMML-MHA against classical single task machine learning approaches. The third experiment compares the performance of DMML-MHA with classical multi-task machine learning models. Finally, the fourth experiment aims to demonstrate how deep multitask learning and deep multisource learning work jointly to improve the performance of DMML-MHA. The intuition, procedures, and metrics used in each experiment are detailed below.

Single Feature-Based Models. Features are extracted from signals as indicators of fall risk, which is the most widely applied approach in past health literature. We created three benchmark models based on the three most commonly investigated features: SVAR, ARMS, and TIME. (Howcroft et al. 2013; Hubble et al. 2015).

Stride-time variability (SVAR). SVAR is the ratio of the standard deviation of gait cycle durations over their mean. If there are N gait cycles in the test instance and we denote the duration of gait cycle i as t_i , SVAR is defined as follows:

$$SVAR = \frac{1}{\bar{t}} \sqrt{\frac{1}{N} \sum_{i=1}^N (t_i - \bar{t})^2},$$

where $\bar{t} = \frac{1}{N} \sum_{i=1}^N t_i$. Higher SVAR indicates more variability and less stability in gait patterns.

Acceleration root mean square (ARMS). If there are M acceleration signals in the test instance and we denote the magnitude of signal j as a_j , ARMS is defined as follows:

$$ARMS = \sqrt{\frac{1}{M} \sum_{j=1}^M a_j^2}.$$

A lower ARMS shows smaller changes in acceleration magnitudes, which indicates milder and smaller steps.

Test Completion Time (TIME). For the clinical TUG and walking tests, the completion time is widely used as a criterion to assess the participant's fall risk. For the internal SilverLink dataset, we record the actual time that participants spent on the task. For the external mPower dataset, we record the amount of time to finish walking 20 steps as instructed by the mPower study protocol.

Classical Single Task Machine Learning Models. Several machine learning models have been applied in prior risk and severity assessment studies, including k-nearest neighbor (**KNN**), support vector machine (**SVM**), and naïve Bayes (**NB**) (Cabe et al. 2011; Greene et al.

2012). We create three benchmark models (KNN, SVM, NB) for the classification task and two benchmark models (KNN regression, linear regression (**LR**)) for the regression task.

Classical Multitask Machine Learning Models. Several classical multitask machine learning models have been applied in general machine learning studies, including multitask LASSO (**LASSO**) and multitask elastic net (**ElasticNet**) (Cui et al. 2018; Li et al. 2015; Zhang et al. 2012). We create the above two benchmark systems for both classification and regression tasks.

Alternative Deep Learning Architectures. In our DMML-MHA framework, both Stage 1 (multisource learning) and Stage 3 (multitask learning) contribute to the overall model performance. This evaluation provides sensitivity analysis for settings where we use single source learning instead of multisource learning, single task learning instead of multitask learning, or both. The three benchmarks are denoted as multisource single task learning (**MsSt**), single source multitask learning (**SsMt**), and single source single task learning (**SsSt**), respectively.

Ten-fold cross-validation is applied as a common machine learning approach to avoid overfitting. We conduct two-sample *t*-tests to identify whether our DMML-MHA framework significantly outperforms the benchmark models. The significant levels are labeled in the form of asterisks.

For the fall risk assessment task, we report our performance evaluation in the form of F₁-measure (F-measure), which is the harmonic mean of precision and recall. The F-measure provides overall evaluation of the system’s ability to identify patients who are fallers while reducing false alarms. The formulas for the above metrics are as follows:

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN}, F_1 - measure = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall},$$

where TP is the true positive (recognized as a faller and is a faller), FP is the false positive (recognized as a faller but is not), and FN is the false negative (recognized as a non-faller but is actually a faller). A higher F-measure indicates that the model assesses fall risk more accurately.

For the PD severity assessment task, we report root mean squared error (RMSE) to reflect how accurately the model is able to assess the severity. The RMSE is calculated as:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2},$$

where y is the ground-truth questionnaire score, \hat{y} is the score predicted by the model, and n is the number of samples. A lower RMSE indicates that the model assesses PD severity more accurately.

4.5. Experiment Results

4.5.1. Single Feature-Based Models

Table 4.1: Precision, Recall, F-measure, and RMSE for Single Feature-Based Models

	Internal Dataset				External Dataset	
	Fall Risk			PD Severity	MDS-UPDRS	PDQ-8
	Precision	Recall	F-measure	RMSE	RMSE	RMSE
DMML-MHA	0.925	0.958	0.940	0.0601	0.1191	0.1543
ARMS	1.000	0.417***	0.588***	0.1812***	0.1253***	0.1636***
TIME	1.000	0.167***	0.286***	0.1927***	0.1257***	0.1643***
SVAR	1.000	0.167***	0.286***	0.1927***	0.1260***	0.1650***

Note: *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001. The asterisks indicate the significance level of DMML-MHA outperforming the corresponding benchmark model.

We compare our proposed DMML-MHA framework with three widely used single feature-based benchmark models, ARMS, TIME, and SVAR (Table 1). Our model achieves

F-measure of 0.940 and RMSE of 0.0601 on the internal dataset, and RMSE of 0.1191 and 0.1543 on the external dataset, significantly outperforming all three benchmark models. Although the feature-based approaches achieve perfect precision, their low recall indicates that they failed to identify many fallers. This result shows the advantage of applying deep learning techniques over single feature-based models in sensor-based health condition risk and severity assessments, which tend to focus on overall features (e.g., test completion time) and reflect less fine-grained local features (e.g., patterns from step to step). As a result, our approach detects fallers and assesses PD severities more accurately than the single feature-based benchmark models.

4.5.2. Classic Single Task Machine Learning Models

Table 4.2: Precision, Recall, F-measure, and RMSE for Classic Single Task Machine Learning Models

	Internal Dataset				External Dataset	
	Fall Risk			PD Severity	MDS-UPDRS	PDQ-8
	Precision	Recall	F-measure	RMSE	RMSE	RMSE
DMML-MHA	0.925	0.958	0.940	0.0601	0.1191	0.1543
KNN	0.846***	0.917*	0.880***	0.1181***	0.1247*	0.1636*
NB	0.800***	0.667***	0.727***	-	-	-
SVM	0.571***	0.667***	0.615***	-	-	-
LR	-	-	-	0.1569***	0.1222*	0.1599

Note: *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001. The asterisks indicate the significance level of DMML-MHA outperforming the corresponding benchmark model.

We compare our proposed DMML-MHA framework with four widely recognized classic single task machine learning models, KNN, NB, SVM, and LR (Table 2). Our model outperforms all the benchmark models, most with significant margins. The classic single task machine learning approaches show significant improvement over single feature-based models.

However, the advantage of applying DMML-MHA is still considerable. The best-performing classic machine learning model is KNN for the internal dataset (F-measure 0.880, RMSE 0.1181), and LR for the external dataset (RMSE 0.1222 and 0.1599). This result suggests that the information extracted by classic machine learning models is insufficient to handle high-dimensional motion sensor data. In contrast, our DMML-MHA framework comprehensively utilizes sensor data and automatically learns salient patterns, resulting in improved assessment results compared to the classic single task machine learning benchmark models.

4.5.3. Classic Multitask Machine Learning Models

Table 4.3: Precision, Recall, F-measure, and RMSE for Classic Multitask Machine Learning Models

	Internal Dataset				External Dataset	
	Fall Risk			PD Severity	MDS-UPDRS	PDQ-8
	Precision	Recall	F-measure	RMSE	RMSE	RMSE
DMML-MHA	0.925	0.958	0.940	0.0601	0.1191	0.1543
LASSO	0.546***	1.000	0.706***	0.1926***	0.1230**	0.1610*
ElasticNet	0.546***	1.000	0.706***	0.1804***	0.1229**	0.1608*

Note: *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001. The asterisks indicate the significance level of DMML-MHA outperforming the corresponding benchmark model.

We compare our proposed DMML-MHA framework with two classic multitask machine learning models, multitask LASSO and multitask ElasticNet (Table 3). Our model significantly outperforms both the benchmark models in terms of F-measure and RMSE. We observed that ElasticNet slightly outperformed LASSO. Such a result seems intuitive because ElasticNet is a generalization of the LASSO and Ridge methods. Although both LASSO and ElasticNet achieve perfect recall in identifying fallers, their low precision leads to high false alarm rates. A high false alarm rate negatively affects health professionals' and patients'

confidence and willingness to apply the model in daily life. The multitask learning mechanism in the two benchmark models has difficulty in identifying salient features hidden in motion sensor signals, whereas the multisource and multitask learning layers in DMML-MHA are able to extract such features.

4.5.4. Alternative Deep Learning Models

Table 4.4: Precision, Recall, F-measure, and RMSE for Alternative Deep Learning Models

	Internal Dataset				External Dataset	
	Fall Risk			PD Severity	MDS-UPDRS	PDQ-8
	Precision	Recall	F-measure	RMSE	RMSE	RMSE
DMML-MHA	0.925	0.958	0.940	0.0601	0.1191	0.1543
MsSt	0.915	0.923	0.917*	0.1325***	0.1226***	0.1590***
SsMt	0.785***	0.947	0.855***	0.1040***	0.1211***	0.1556**
SsSt	0.772***	0.934	0.842**	0.1352***	0.1227***	0.1593***

Note: *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001. The asterisks indicate the significance level of DMML-MHA outperforming the corresponding benchmark model.

Lastly, we compare our proposed DMML-MHA framework with alternative deep learning models, MsSt, SsMt, and SsSt, which can be treated as partially functioning DMML-MHA frameworks (Table 4). Our model significantly outperforms all three benchmark models in terms of F-measure and RMSE. An interesting observation is that, among the alternative models, we observe that MsSt generates the best F-measure (0.917) for the fall risk assessment task, while SsMt generates the best RMSE (0.1040 for the internal dataset, 0.1211 and 0.1556 for the external dataset) for the PD severity assessment task. Our proposed DMML-MHA framework combines and leverages the strengths of both MsSt and SsMt with improved results on both fall risk assessment and PD severity assessment tasks.

4.6. Discussion and Conclusions

Benefiting from the recent advancement of sensing technologies and success of deep learning on complex image recognition and natural language processing tasks, this research leverages multiple data sources for multiple related health assessment tasks to enhance and advance mobile health analytics. In particular, we contribute to the state-of-the-art research by designing a novel Deep Multisource Multitask Learning for Mobile Health Analytics (DMML-MHA) framework for more precise, prompt, and personalized senior care. By extracting source-specific, general, and task-specific feature layers, the DMML-MHA framework is able to alleviate significant laborious and ad hoc feature engineering and automatically learn salient features from multiple data sources with sensor fusion. This innovation leads to improved health condition risk and severity assessment results and further helps stakeholders' healthcare management.

To demonstrate the effectiveness of the proposed DMML-MHA framework, we collect an internal sensor signal dataset from our SilverLink project and an external dataset from the mPower project, both related to PD patients, and assess fall risks and PD severities. Our experiment results show that our DMML-MHA framework consistently outperformed the four sets of competitive benchmark models in terms of F-measure and RMSE. This novel framework can result in significantly reduced total expenditure by accurately identifying senior citizens who are fallers and those at early stages of PD, as well as providing proper preventive interventions to alleviate health condition deterioration.

4.6.1. Contributions to IS Knowledge Base

Our contribution to IS knowledge base is two-fold: the two instantiations of the DMML-MHA framework based on the two datasets, and design insights in multiple domains related to IS analytics. The design insights are further discussed below.

Health analytics. Our proposed DMML-MHA framework is designed for featureless and transferable mobile health analytics. Only limited modifications (e.g., numbers of data sources and health assessment tasks) are required for the framework to apply on new diseases (e.g., from Parkinson's disease to dementia or diabetes) without the need of disease-specific features.

Big data. The DMML-MHA framework is able to obtain insights from multisource sensor data that are generated in high velocity (e.g., 25 data points per second) and high variety (e.g., activities such as walking, standing up, sitting down, and turning around). This framework can be extended to obtain value from other big data applications sharing the above two characteristics, e.g., multisensor fusion for Internet of Things (IoT), multisite fusion for online social media analysis, etc.

Multitask learning. Prior studies in multitask learning mostly rely on classical machine learning algorithms (e.g., regressions, SVM, tree-based models, etc.), which have difficulty in handling high-dimensional sensor signal data. The DMML-MHA framework contributes to the multitask learning methodology by extending its applicability to sensor signal data and other types of data sharing similar characteristics.

4.6.2. Practical Implications

Since Former President Obama launched the Precision Medicine Initiative in 2015, healthcare research and practice are increasingly becoming more precise, prompt, and personalized. With the development of sensing technologies, researchers and practitioners are able to collect, store, and analyze various types of sensor data from senior citizens to assess their health condition risks and severities and provide medical intervention. Senior citizens, their families, and health professionals can all benefit from DMML-MHA, a more generalizable and precise framework for mobile health analytics. We discuss major practical implications for those stakeholders below.

Senior citizens. Senior citizens face difficulties in their independent living, partly due to the lack of a convenient and reliable approach to better understand their adverse event risks and health condition severities. Assistive tools such as instantiations of the DMML-MHA framework can empower senior citizens by enabling them to conduct mobility tests at home and receive assessment reports from the analytical engine, which relieves them from the burden of visiting clinics. Senior citizens can obtain precise assessments of the progression of existing health conditions or the risks of potential health conditions, improving their confidence for independent living.

Senior citizens' families. Traditionally, home-dwelling senior citizens require one or more caregivers to ensure that their health conditions are properly monitored, and unforeseen accidents are resolved in a timely manner (e.g., falling down, acute onset of symptoms, etc.). With a remote motion sensor system equipped with analytical engines such as the DMML-MHA framework, senior citizens' families can attain real-time reports of their loved ones'

daily routines with health condition risks and severities without disrupting them, leading to peace of mind.

Health professionals. The DMML-MHA framework is a flexible tool for advanced clinical decision support at the point-of-care for health professionals. It can be integrated into health professionals' clinical routines by conducting mobility tests (including timed up and go and walking tests) to support their diagnosis. Although health professionals are highly trained, the DMML-MHA framework may reveal significant additional patterns in sensor signals that are overlooked by humans, leading to an improvement of overall healthcare quality.

4.6.3. Limitations and Future Research

As with other academic studies in emerging applications, there are limitations in this research. First, although we demonstrated the generalizability of our proposed DMML-MHA framework by internal and external evaluations, we only tested our framework in the context of falls and PD. Future researchers can instantiate the DMML-MHA framework based on their applications and contexts (e.g., dementia, diabetes, etc.) and explore the boundaries of such a framework. Second, the instantiation of the DMML-MHA framework may require significant tuning and experimentation on framework structure and hyperparameters, which is an issue commonly faced by the deep learning community. Although we followed the general guidelines for constructing deep learning frameworks, such design issue is still an open research topic in the field. Third, although we imposed source-specific, general, and task-specific deep learning layers in our DMML-MHA framework, the actual features learned by the framework may not be directly interpretable by humans. Medical practitioners may find difficulty in adopting the framework in clinical routines. Future research in identifying

clinically meaningful and interpretable features is needed. We believe that the above directions of research are of great interest to researchers and practitioners to further enhance senior citizens' life quality.

5. ESSAY IV: MOTION SENSOR-BASED HEALTH PROFILING FOR SENIOR CARE: ADAPTIVE TIME-AWARE CONVOLUTIONAL LONG SHORT TERM MEMORY (ATCLSTM)

5.1. Introduction and Essay Structure

Modern medicine has enabled humans to combat numerous deadly diseases and live more robust lives well into their 60's, 70's, and 80's. However, chronic conditions such as Parkinson's disease, frailty, and dementia severely obstruct senior citizens' independent daily living. Parkinson's disease, costing \$15.5 billion per year for the U.S., is the second most common neurodegenerative disorder with symptoms of tremor, rigidity, bradykinesia, and postural instability.

The management of Parkinson's disease is of great significance. Specific patterns can be demonstrated in the progression of Parkinson's disease (Visanji et al. 2013). A clear understanding of the PD progression might lead to a better comprehension of the pathophysiology of the disease, opening new ways for the development of neuroprotective therapies (Maetzler et al. 2009). In addition, a study shows that a 20% reduction in PD progression would result in net monetary benefits of \$60,657 per patient (Johnson et al. 2013).

One major management tool for Parkinson's disease is long-term disease progression assessment, or health profiling. By tracking the a citizen's health status in a long term (e.g., 6 to 12 months), health profiling aims to calculate a personalized summative score that reflects disease progression or severity for the senior citizen, which provides patient self-empowerment and assists physicians and caregivers for timely interventions. In health profiling, the senior citizen's health status is sampled and collected multiple times on different dates, and may

contain time irregularities. This creates a trajectory for predicting the senior citizen's disease progression, which is an indicator for future therapies and interventions.

Two conventional health profiling approaches exist. The first surveys a patient's disease history, medications, medical attention, walking steadiness, gait, and other risk factors (Phelan et al. 2015). However, surveys are subjectively conducted by physicians. Thus, they may be inaccurate and inconsistent. The second approach evaluates patients using timed mobility tests. These include the 4-meter walking test (Maggio et al. 2016), balance test (Franchignoni et al. 2010), and Timed Up and Go (TUG) test (Shumway-Cook et al. 2000), among others. Despite their strengths, such tests require controlled laboratory environments, are expensive, and the results may not generalize.

These limitations have motivated the rapid emergence of wearable motion sensor-based information systems (IS) for health profiling. These systems provide physicians and caregivers more timely, objective, accurate, convenient, and inexpensive remote home monitoring capabilities than traditional approaches. However, current analytics engines for motion sensor-based health profiling mostly rely on manual feature engineering, which is labor-intensive, ad hoc, and may reach inconclusive results (Hubble et al. 2015). Also, they focus on one-shot test cases, which is unable to effectively integrate data collected at different times or dates. Although deep learning models have been proposed on motion sensor signals to resolve manual feature engineering issues, such models are primarily proposed for human activity recognition, which may not be directly applicable for health profiling. Additionally, to the best of our knowledge, current works on deep learning-based temporal health profiling only focus on Electronic Health Records (EHR) data, which leaves health profiling on motion sensor data unexplored.

To address the above gaps, we propose a novel deep learning model, Adaptive Time-aware Convolutional Long Short Term Memory (ATCLSTM), to provide long-term health profiling with motion sensor data. The ATCLSTM consists of two main modules: convolutional module, which aims to automatically learn feature representations from raw motion sensor data, and adaptive time-aware long short term memory module, which adaptively models temporal sequences with time irregularities. The main novelty of ATCLSTM is as follows. First, to the best of our knowledge, this is the first integrated framework combining Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) for long-term health profiling. Second, the proposed adaptive time-aware module allows learning fully data-driven time decay functions, which enhances the model's ability in modeling time irregularities and is generalizable to time series data other than motion sensor signals.

The remainder of this essay is organized as follows. First, we review motion sensor-based health profiling methods, deep learning studies on motion sensor data, and deep learning models for temporal health profiling. From our review, we identify key research gaps and pose research questions for study. Subsequently, we detail our research design and datasets. We then summarize evaluation results. Finally, we highlight our research contributions and conclude this research.

5.2. Literature Review

Three literature streams guide this research. First, we review motion sensor-based health profiling in medical literature to identify state-of-the-art practices in clinics. Second, we examine deep learning studies on Internet of Things (IoT) streaming data analytics to identify

the unique data characteristics of IoT streaming data and how prevailing deep learning methods operate. Subsequently, we review deep learning models for temporal health profiling to understand how to model time irregularities in deep learning.

5.2.1. Motion Sensor-Based Health Profiling

Motion sensors have emerged as a major trend for health profiling, including assessing severity of Parkinson's disease (Warlop et al. 2018; Hubble et al. 2015; Weiss et al. 2014), stage of frailty (Millor et al. 2017; Schwenk et al. 2015), risk of falling (Wang et al. 2017; Silva and Sousa 2016; Van Schooten et al. 2015), among others. Based on data collected by motion sensors, motion sensor-based health profiling aims to generate a score that reflects the progression of certain diseases, such that proper interventions can be provided to alleviate disease progression or eliminate the future occurrence of adverse events. Table 5.1 summarizes selected recent research based on device, mobility test, application context, and analytics.

Table 5.1: Selected Recent Research in Motion Sensor-Based Health Profiling

Year	Author	Device	Mobility Test	Context	Analytics
2018	Warlop et al.	Acc	Treadmill walking	Parkinson's disease	Statistical analysis
2017	Wang et al.	Acc	Walking, up/downstairs	Fall	Logistic regression
2017	Watanabe et al.	Acc, Gyro	Walking	Diabetes	Statistical analysis
2017	Millor et al.	Acc, Gyro, Mag	30s chair stand test, walking	Frailty	Statistical analysis
2017	Silva & Sousa	Acc, Gyro	TUG	Fall	Statistical analysis
2015	Hubble et al.	Acc	Walking, TUG	Parkinson's disease	Statistical analysis
2015	Schwenk et al.	Acc	Walking, standing	Frailty	Statistical analysis
2015	Van Schooten et al.	Acc	8-day home monitoring	Fall	Statistical analysis
2014	Weiss et al.	Acc	Walking	Parkinson's disease	Statistical analysis
2013	Najafi et al.	Acc	TUG	Diabetes	Statistical analysis

Note: Acc=Accelerometer; Gyro=Gyroscope; Mag=Magnetometer; TUG=Timed Up and Go test

Two major scenarios have been identified for health profiling: clinical setting and home setting. In a clinical setting, motion sensors are attached to patients to collect data from controlled mobility tests instructed by health professionals. Limited by the time constraints of clinical visits, clinical tests can only provide infrequent snapshots of patients' mobility patterns. However, health professionals and researchers are able to observe the full process of more complicated clinical mobility tests, ensuring that test criteria are met and necessary annotations are recorded. Timed Up and Go (TUG) tests are one of such mobility tests, which consist of six continuous components: (1) stand up, (2) walk straight for 3 meters, (3) turn around, (4) walk straight for 3 meters, (5) turn around, and (6) sit down on the chair. In addition, multiple sensors are often preferred to collect more comprehensive data across multiple locations on the human body (e.g., chest, waist, leg, ankle, etc.). Consequently, the motion sensor data collected

from clinical tests are small in volume and relatively short-term, but more structured and comprehensive. In contrast, motion sensor data collection in an uncontrolled home setting requires participants to wear motion sensors or hold smartphones while performing regular daily activities (e.g., sitting, walking, lying, etc.) (Silva and Sousa 2016; Greene et al. 2017; Schwenk et al. 2014) or unsupervised mobility tests (e.g., walking for 20 steps) (Bot et al. 2016). The health profiling period of this approach can span over days to months, which allows collecting more frequent mobility test samples (e.g., walking for 20 steps per day) and modeling disease progression in a long term (e.g., 6 to 12 months) for more precise and personalized health profiling.

Regardless of health profiling scenarios, most prior studies derived features from sensor signals (primarily acceleration values from accelerometers) followed by statistical tests (e.g., t-test, ANOVA) to test the statistical significance of the features. Numerous features have been proposed to represent different aspects of disease progression, including two major categories: general signal processing features and gait features. General signal processing features are common features such as mean, standard deviation, maximum, inter-quartile range, energy, entropy, mean-crossing rate, Fast Fourier Transform (FFT) coefficients, etc. (Silva and Sousa 2016). Those feature are applicable to general signal processing domains not limited to sensor-based mobility analysis. Gait features are specialized features such as gait speed, stride time, stride length, gait variability, gait symmetry, etc. (Van Schooten et al. 2015; Hubble et al. 2015), which directly reflect a subject's gait patterns. However, such approaches require manual feature engineering, which is labor-intensive, ad hoc, and inconclusive (Hubble et al. 2015). Additionally, extant works primarily focus on one-shot tests without modeling temporal progression of diseases. As deep learning models have shown their capabilities in automatic

feature learning and temporal modeling (LeCun et al. 2015; Goodfellow et al. 2016), we next review deep learning-based methods on IoT streaming data analytics, which motion sensor data analytics belongs to.

5.2.2. Deep Learning on IoT Streaming Data Analytics

Recently, deep learning has brought astounding breakthroughs in numerous areas that were considered too complex for machines, including image recognition, speech recognition, machine translation, among others (LeCun et al. 2015), which shows a potential to be applied on streaming data. Streaming data refers to the data generated or captured within tiny intervals of time and need to be promptly analyzed to extract immediate insights and/or make fast decisions (Mohammadi et al. 2018). Applications include smart homes (Manic et al. 2016), smart city (Song et al. 2016; Liang et al. 2016), intelligent transportation systems (Ma et al. 2015; Tian and Pan 2015), healthcare and wellbeing (Wang et al. 2017; Lipton et al. 2016), education (Piech et al. 2015; Yang et al. 2017), sport (Kautz et al. 2017; Ibrahim et al. 2016), among others. Motion sensors have been used in multiple areas such as smart homes, healthcare and wellbeing, and sport. We summarize selected major deep learning literature on healthcare based on device, application context, and deep learning model in Table 5.2.

Table 5.2: Selected Major Deep Learning Research in a Health Context

Year	Author	Device	Context	Model
2018	Kim et al.	Acc, Gyro	Gesture recognition	CNN + GRU
2018	Zhu et al.	Acc, Env	Activity recognition	Seq2Seq GRU
2018	Hassan et al.	Acc	Activity recognition	Deep Belief Network
2017	Murad & Pyun	Acc	Activity recognition	LSTM
2016	Eskofier et al.	Acc	Parkinson's disease	CNN
2016	Ordóñez & Roggen	Acc, Gyro, Mag	Activity recognition	Convolutional LSTM
2015	Yang et al.	Acc, Env	Activity recognition	CNN
2014	Zeng et al.	Acc	Activity recognition	CNN

Note: Acc=Accelerometer; Gyro=Gyroscope; Mag=Magnetometer; Env=Environmental Sensor;
 CNN=Convolutional Neural Network; GRU=Gated Recurrent Unit; Seq2Seq=Sequence to Sequence;
 LSTM=Long Short Term Memory

We notice that the majority of deep learning studies using motion sensor data have been proposed on activity recognition, which aims to identify human activities of daily living (ADL, e.g., walking, standing up, sitting down, etc.) based on collected motion sensor data. Accelerometers are the most frequent data source due to their ability in capturing translational movements and their prevalence in embedded devices. Regarding deep learning analytics, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are the two most successful deep learning models, with CNNs excelling in automatic feature learning for grid-like data (e.g., images) and RNNs designed for temporal modeling (e.g., time series). As multi-axial sensor signals are considered both grid-like data (1-D or 2-D) and time series, both CNNs and RNN variants (Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU)) have been proposed to model motion sensor data. The successful implementation of CNNs for activity recognition is due to their capability for learning powerful and discriminative features, as well as utilizing convolutions across temporal sequences to capture

local dependencies between nearby input samples (Murad and Pyun 2017). However, the temporal dynamics of activity sequences are not extensively captured through the convolution operations because CNNs do not explicitly model temporal dependencies (a later signal is generated *after* an earlier signal). To address this issue, RNNs have been proposed to capture such temporal dynamics exploiting their internal memories. Due to the training difficulties (vanishing and exploding gradients) in vanilla RNNs, gated architectures such as LSTM and GRU are proposed to maintain a stable internal memory representation and allow the gradient to flow smoothly in the training process. However, applying RNNs directly on motion sensor data is considered suboptimal due to the lack of necessary abstraction on the fine-grained raw data.

To fully leverage the advantages of CNNs and RNNs, Ordóñez and Roggen (2016) proposed DeepConvLSTM (referred to as CLSTM in this essay), which combines convolutional and recurrent layers in a deep learning framework. The convolutional layers act as feature extractors and provide abstract representations of the input sensor data in feature maps, whereas the recurrent layers model the temporal dynamics of the activation of the feature maps. The main difference between CLSTM and a plain CNN is that CLSTM uses LSTM cells in place of fully connected layers. Improved performance in activity recognition is confirmed in Ordóñez and Roggen (2016). In addition, Kim et al. (2018) proposed a CNN with GRU architecture for gesture recognition following the CLSTM design.

However, current deep learning studies on motion sensor signals focus on one-shot, short-term experiments (data collected within a day), which may not be applicable to long-term health profiling (data collected through days or months with irregularities). This is because RNNs (including LSTM and GRU) assume the time intervals between data samples

are constant, which is not designed to tackle time irregularities. This observation also applies to most deep learning studies on streaming data analytics as they either ignore the time intervals between two samples and treat the sequences as ordinal (e.g., Lipton et al. 2016; Piech et al. 2015), or segment samples into constant-interval buckets (e.g., Liang et al. 2016). To the best of our knowledge, none of the above studies explicitly considered time irregularities in their proposed models. We next turn to deep learning literature on temporal health profiling to seek potential solutions.

5.2.3. Deep Learning for Temporal Health Profiling

Extant deep learning literature for temporal health profiling has been mostly proposed on electronic health record (EHR) data because there is a need to model the time irregularities in multiple hospital admission cases. Table 3 summarizes their application contexts, deep learning models, and temporal modeling strategies.

Table 5.3: Selected Deep Learning Literature for Temporal Health Profiling

Year	Author	Data	Context	Model	Strategy
2018	Zhang et al.	EHR (PPMI)	Patient subtyping	TS-GRU	Discount or amplify hidden state by time intervals
2017	Baytas et al.	EHR (PPMI)	Patient subtyping	TLSTM	Discount short-term memory component by time intervals
2017	Ma et al.	EHR	Diagnosis prediction	Bi-GRU	Use ordinal data input (ignore variable intervals)
2017	Nguyen et al.	EHR	Readmission risk prediction	CNN	Code time interval as a feature
2016	Pham et al.	EHR	Diagnosis and readmission prediction	LSTM	Modify forget gate by time intervals

Note: EHR=Electronic Health Record; PPMI=Parkinson's Progression Marker Initiative;

TS-GRU=Time-Sensitive Gated Recurrent Unit; TLSTM=Time-aware Long Short Term Memory;

Bi-GRU=Bidirectional Gated Recurrent Unit; CNN=Convolutional Neural Network; LSTM=Long Short Term Memory

Three key observations are summarized as follows. First, EHR data have been leveraged for predictive analytics including patient subtyping, diagnosis prediction, readmission risk prediction, etc. Second, the majority of the proposed deep learning models are RNN variants, although CNNs are also applied as a minor approach. Third, three general strategies are explored to model time irregularities. The first approach ignores time irregularities in the data, treating them as ordinal to fit an ordinary bidirectional GRU. This approach is less preferred as literature suggest the time irregularity is a key factor for temporal health profiling (Zhang et al. 2018; Baytas et al. 2017). The second approach codes time intervals as an additional special feature to EHR records. While applicable for EHR data, such an approach is not compatible with raw motion sensor data as there are no explicit features coded in it. The third approach modifies a RNN component to account for the effect of elapsed

time. The chosen RNN component is typically the memory, the hidden state, or the forget gate as in LSTM. The Time-aware Long Short Term Memory (TLSTM, Baytas et al. 2017) model is considered to be a milestone in temporal health profiling as it directly adjusts the LSTM memory instead of the forget gate to avoid any alteration of the current input's effect to the current output except for the elapsed time. The model details are illustrated in Figure 5.1 and discussed below.

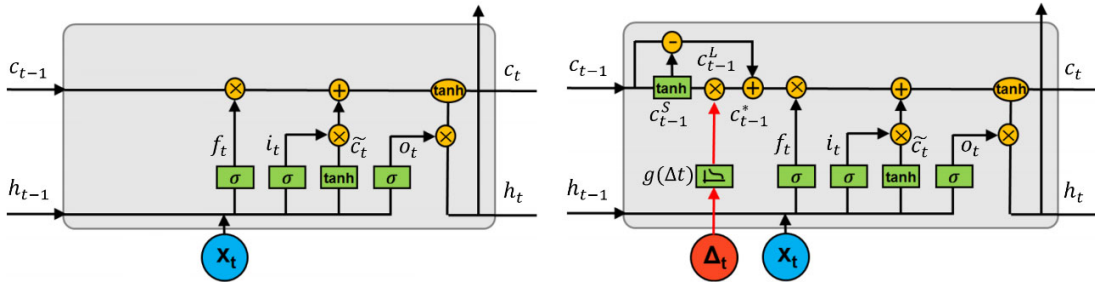


Figure 5.1: (Left) LSTM Cell; (Right) TLSTM Cell

LSTM is a gated variant of RNN. By introducing the memory cell (c_t) in addition to the hidden state (h_t), LSTM is able to store information of the past and pass along time steps with additive linear interactions, which results in an uninterrupted gradient flow. LSTM manipulates input information (x_t) with three gates: forget gate (f_t), input gate (i_t), and output gate (o_t). LSTM is formularized as follows:

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$$

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

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t$$

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

where \tanh denotes the hyperbolic tangent function, σ denotes the sigmoid function, \odot denotes element-wise multiplication, $\{W_f, U_f, W_i, U_i, W_o, U_o, W_c, U_c\}$ denotes corresponding weight matrices, and $\{b_f, b_i, b_o, b_c\}$ denotes corresponding biases.

TLSTM is proposed to incorporate the elapsed time information into the standard LSTM architecture to be able to capture the dynamics of sequential data with time irregularities (Baytas et al. 2017). TLSTM extends LSTM by decomposing the previous memory cell (c_{t-1}) into two components, short-term memory (c_{t-1}^S) and long-term memory (c_{t-1}^L), and discount the short-term memory by a time decay function, $g(\Delta t)$, where Δt is the time interval between two records. The adjusted previous memory cell (c_{t-1}^*) is reconstructed by combining the long-term memory, c_{t-1}^L , and the discounted short-term memory, $g(\Delta t)c_{t-1}^S$, and participates in calculating the current memory cell, c_t . In this way, the model is able to retain long-term effects in the memory cell, while adjust the short-term effects based on the time span between two consecutive test cases. The full formulation of TLSTM is as follows:

$$\begin{aligned}
c_{t-1}^S &= \tanh(W_d c_{t-1} + b_d) \\
c_{t-1}^L &= c_{t-1} - c_{t-1}^S \\
c_{t-1}^* &= c_{t-1}^L + g(\Delta t)c_{t-1}^S \\
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
\tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c) \\
c_t &= f_t \odot c_{t-1}^* + i_t \odot \tilde{c}_t \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}$$

where $g(\Delta t)$ denotes a monotonically non-increasing scalar function such as $g(\Delta t) = 1/\Delta t$ or $g(\Delta t) = 1/\log(e + \Delta t)$ (Pham et al. 2016), $\{W_f, U_f\}$ denotes the weight matrices for the short-term memory decomposition network, $\{b_f\}$ denotes the corresponding bias, and the other notations follow those in LSTM.

Although TLSTM provides an innovative and effective way modeling time irregularities, there are two limitations in it. First, extant TLSTM and other temporal health profiling deep learning models are only applied on EHR data, where semantically meaningful features are readily available. However, this does not naturally applicable to motion sensor data. Motion sensor data is typically collected in a high-frequency manner (e.g., 100 tri-axial acceleration values per second), which translates to 3,000 features for a 10-second mobility test ($= 100 \times 3 \times 10$). Each feature only contains little information, while there are latent temporal and cross-axial patterns lying in the features. Similar to image recognition applications (LeCun et al. 2015), CNNs may be appropriate to identify such patterns. Second, in the current TLSTM formulation, the time decay function $g(\Delta t)$ is a manually selected scalar function, which only allows a uniform discount rate on all components of c_{t-1}^S . However, the key design contribution of deep learning is to let the model learn the transformation functions from the data and decide an adaptive discount rate on each component of c_{t-1}^S . Innovative approaches to address the above limitations are still under exploration.

5.3. Research Gaps and Questions

Several research gaps are identified from existing literature. First, most motion sensor-based health profiling approaches employ manual feature engineering, which is labor-intensive, ad hoc, and may lead to inconclusive results. Second, current deep learning methods on motion sensor data primarily focus on activity recognition. Although temporal modeling techniques

have been explored, they are suboptimal in modeling time irregularities as required in long-term health profiling. Third, existing deep learning models for temporal health profiling mainly concentrate on EHR data, which may not be directly transferrable to motion sensor data. In addition, the current manually selected time delay function as in TLSTM only allows a uniform discount rate on all short-term memory components. It is unclear whether an adaptive time decay function can be learned on each short-term memory component in a data-driven manner. These gaps motivate the following research questions:

1. How can we develop an integrated deep learning framework combining the advantages of CNN and RNN to allow long-term health profiling based on motion sensor data?
2. How can we develop a fully data-driven mechanism for learning the time decay function to avoid manual selection and allow adaptive discount rates on the short-term memory components?
3. How can we leverage the proposed design to model time series data in other related applications?

5.4. Research Design

We carefully design a novel motion sensor-based health profiling framework (Figure 5.2) to address the identified research gaps and the posed research questions. This framework consists of three major components: (1) data collection, (2) data preprocessing, and (3) the core methodological contribution of this research, Adaptive Time-aware Convolutional Long Short Term Memory (ATCLSTM), with benchmark evaluations. Each is summarized in the following subsections.

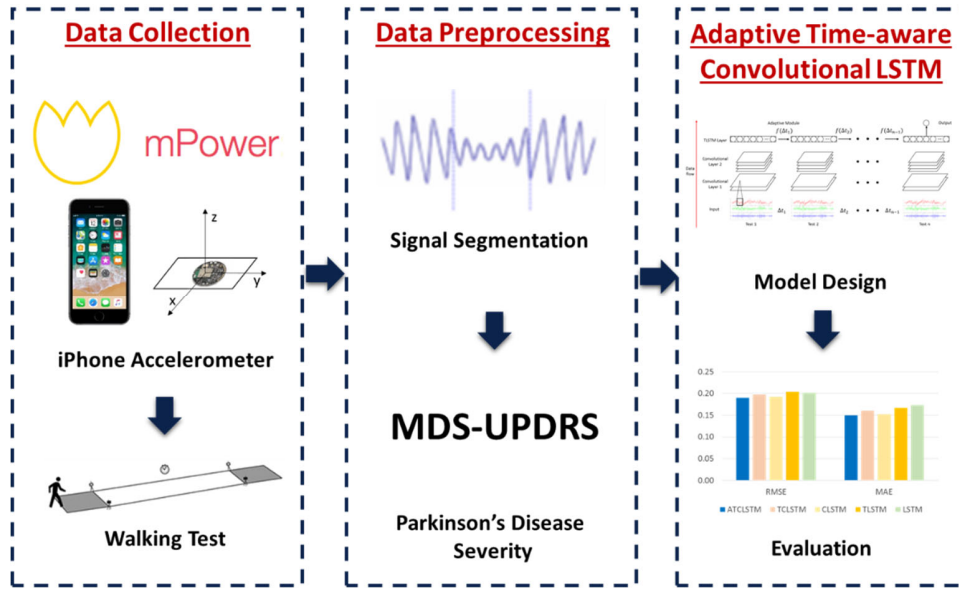


Figure 5.2: Research Design for Motion Sensor-Based Health Profiling

5.4.1. Data Collection

A publicly available dataset, mPower, is used as the research testbed (Bot et al. 2016). mPower is an observational smartphone-based study developed using Apple's ResearchKit library to evaluate the feasibility of remotely collecting sensor data reflecting PD patients' health condition severities. Participants were presented four types of activities, including memory (evaluating short-term spatial memory), tapping (measuring dexterity and speed), voice (recording participants' sustained phonation), and walking (evaluating participants' gait and balance), which they could complete as many as three times a day. All participants were requested (although not required) to respond to standard surveys used for PD assessment, including Parkinson Disease Questionnaire 8 (PDQ-8) and a subset of questions from the Movement Disorder Society Universal Parkinson Disease Rating Scale (MDS-UPDRS) (Bot et al. 2016). The data was collected from March to September 2015, available upon request on Sage Bionetwork's Synapse platform.

We used mPower’s walking data and MDS-UPDRS questionnaire as our dataset. The study protocol instructed participants to walk 20 steps in a straight line at their comfortable speed. The data was recorded by iPhone’s embedded tri-axial accelerometer at a sampling frequency of 100 Hz (100 acceleration values per second). In total, 3,101 participants completed 35,410 walking test cases.

5.4.2. Data Preprocessing

We observed that there are incomplete test cases in the dataset due to participants’ misoperation or other reasons. To filter out such incomplete cases, we dropped sensor signal snippets with a length of less than 512 acceleration values. We chose 512 as the threshold because it is equivalent to around 5 seconds at a sampling frequency of 100 Hz, which is necessary for any human to walk 20 steps in a normal pace. As the proposed model requires a fixed-length input, we subsampled the middle 512 values from each complete case to obtain more stable gait patterns. To provide better temporal health profiling, we only included participants who completed three or more test cases (Baytas et al. 2017) with the time intervals between two consecutive test cases recorded.

In this research, we use the severity of Parkinson’s disease as the health profiling goal. We use the sum of MDS-UPDRS questionnaire scoring items as a proxy for the severity of Parkinson’s disease as a common practice in literature (Hubble et al. 2015). The sum score ranged from 0 to 38. As a result, 825 participants met our selection criteria (those who completed three or more test cases with MDS-UPDRS scores recorded) and their data were used for subsequent analyses.

5.4.3. Adaptive Time-aware Convolutional Long Short Term Memory (ATCLSTM)

As noted in the literature review, extant approaches using motion sensors for health profiling have three critical drawbacks. First, current deep learning models on motion sensor data rarely consider time irregularities between test cases, which is a key concern for long-term health profiling. Second, current deep learning models on temporal health profiling mainly focus on EHR data where semantically meaningful features are readily available, which does not apply to motion sensor data. Third, the current manually selected time delay function as in TLSTM only allows a uniform discount rate on all short-term memory components. We proposed a novel Adaptive Time-aware Convolutional Long Short Term Memory (ATCLSTM) to address these issues. At a high level, ATCLSTM contains two major components: the convolutional module for feature learning from raw sensor data, and the adaptive TLSTM module for modeling temporal sequences and time irregularities. An overview of the proposed model is depicted in Figure 5.3.

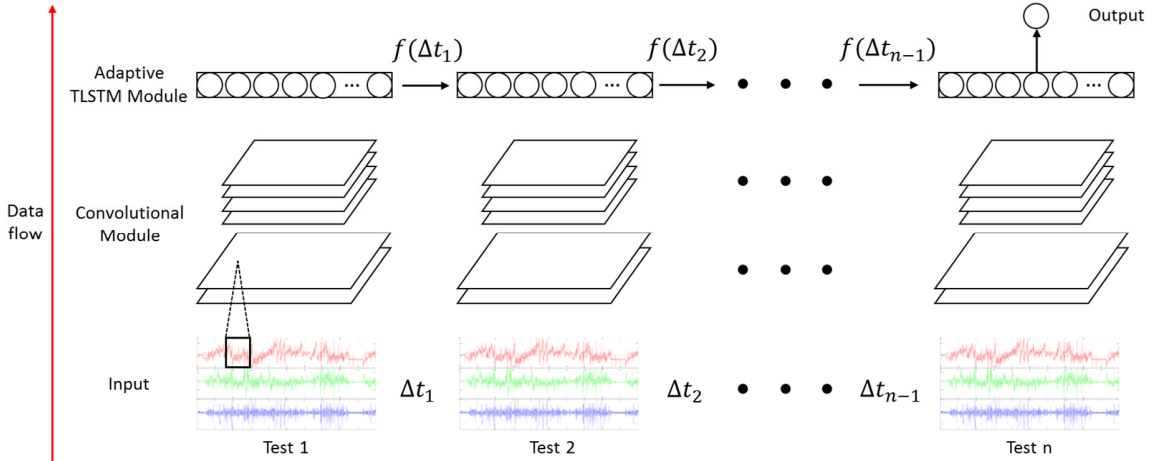


Figure 5.3: Adaptive Time-aware Convolutional Long Short Term Memory (ATCLSTM)

Convolutional Module

CNNs have been widely applied on motion sensor signals to automatically learn feature abstractions from raw data (Kim et al. 2018; Ordóñez and Roggen 2016). The convolutional module of ATCLSTM consists of two sets of convolutional, non-linear, and pooling layers, as guided in Goodfellow et al. (2016). A detailed illustration of the convolutional module is shown in Figure 5.4.

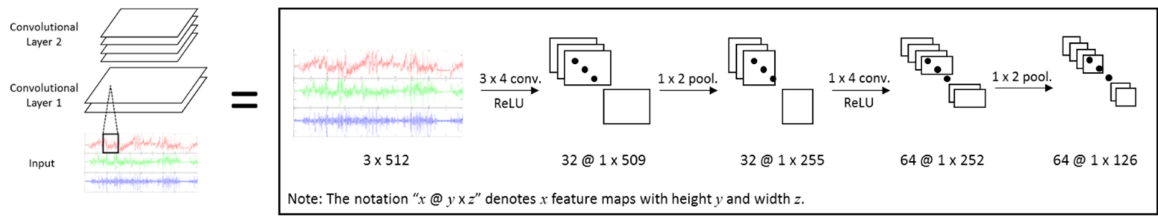


Figure 5.4: Convolutional Module

Convolutional Layers: A convolutional layer convolves input data with a set of convolution kernels and generates feature maps. A convolution kernel is a matrix of weights applicable to small proportions of the input data, which is represented in the following formula:

$$a_{x,y}^{i,j} = b^{i,j} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} w_{p,q}^{i,j,m} z_{x+p,y+q}^{i-1,m},$$

where $a_{x,y}^{i,j}$ is the output value of this layer at row x and column y for feature map j in layer i , $b^{i,j}$ is the bias for feature map j in layer i , m is the number of feature maps in the previous layer, P_i and Q_i are the height and width of the convolution filter in layer i , $z_{x+p,y+q}^{i-1,m}$ is the output from the previous layer (layer $i - 1$) for feature map m at row $x + p$ and column $y + q$, and $w_{p,q}^{i,j,m}$ is the weight for position (p, q) pointing from feature map m in layer $i - 1$ to

feature map j in layer i , which is invariant to the values of x or y and is to be learned through the training process.

Nonlinear Layers: A rectified linear unit (ReLU) layer maps the output of the previous layer to employ model nonlinearity by the following function:

$$a = \max \{z, 0\}.$$

where a is the output and z is the input. The ReLU function is the most widely used nonlinear function for CNNs due to its simplicity and effectiveness (LeCun et al. 2015).

Pooling Layers: In pooling layers, the resolution of feature maps is reduced to keep salient patterns and increase local scale invariance. A maximum pooling layer is added after each nonlinear layer axis by the following function:

$$a_{x,y}^{i,j} = \max_{\substack{0 \leq r \leq R_i-1 \\ 0 \leq s \leq S_i-1}} z_{x+r,y+s}^{i-1,j},$$

where $a_{x,y}^{i,j}$ is the output value of this layer at row x and column y for feature map j in layer i , R_i and S_i are the height and width of the pooling filter in layer i , and $z_{x+r,y+s}^{i-1,j}$ is the output from the previous layer (layer $i - 1$) for feature map j at row $x + r$ and column $y + s$.

After the input data is processed by two sets of convolutional, non-linear, and pooling layers, a flattening layer is attached on the top of the convolutional module such that the feature maps are reorganized to serve as the input for the subsequent adaptive time-aware recurrent module.

Adaptive TLSTM Module

The adaptive TLSTM module is an extension of the ordinary TLSTM proposed by Baytas et al. (2017). A detailed illustration of the adaptive TLSTM module is depicted in

Figure 5.5. The red rectangle highlights our innovation, the adaptive time-aware module, compared to TLSTM.

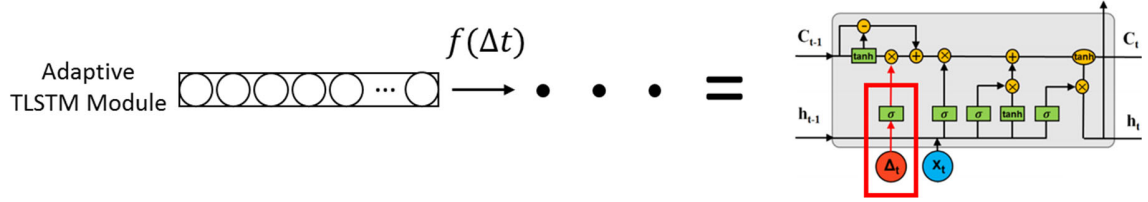


Figure 5.5: Adaptive TLSTM Module

One of the major limitations in TLSTM is that the time decay function, $f(\Delta t)$, is bound to a manually selected scalar function, $g(\Delta t)$, which is suggested to be one of the following two (Pham et al. 2016):

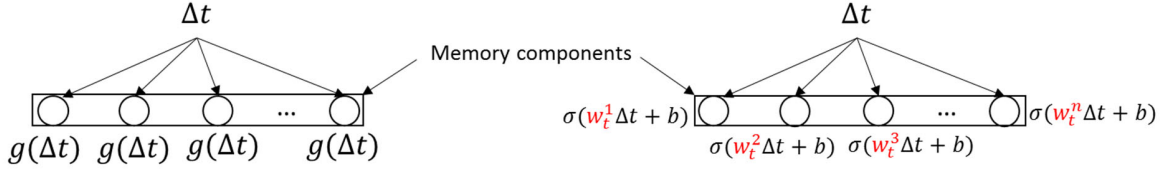
$$g(\Delta t) = \frac{1}{\Delta t}, \text{ or}$$

$$g(\Delta t) = \frac{1}{\log(e + \Delta t)}.$$

This manual, ad hoc selection does not allow adaptive discount rates on the short-term memory components, which may limit model performance. To address this concern, we propose $f(\Delta t)$ to be a neural network layer in the adaptive TLSTM module:

$$f(\Delta t) = \sigma(W_t \Delta t + b_t),$$

where σ is the sigmoid function, W_t and b_t are the weight matrix and bias to be learned by the deep learning model. A head-to-head comparison between the time decay factors of TLSTM and those of adaptive TLSTM is shown in Figure 5.6 assuming the short-term memory contains n components.



Note: Superscripts denote the n -th memory component.

Figure 5.6: (Left) Uniform Time Decay Factors in TLSTM; (Right) Adaptive Time Decay Factors in Adaptive TLSTM

The full formulation of the adaptive TLSTM is summarized in the following formulas with aforementioned notations. The major innovation is highlighted in red:

$$\begin{aligned}
c_{t-1}^S &= \tanh(W_d c_{t-1} + b_d) \\
c_{t-1}^L &= c_{t-1} - c_{t-1}^S \\
c_{t-1}^* &= c_{t-1}^L + \textcolor{red}{c_{t-1}^S \odot \sigma(W_t \Delta t + b_t)} \\
f_t &= \sigma(W_f x_t + U_f h_{t-1} + b_f) \\
i_t &= \sigma(W_i x_t + U_i h_{t-1} + b_i) \\
o_t &= \sigma(W_o x_t + U_o h_{t-1} + b_o) \\
\tilde{c}_t &= \tanh(W_c x_t + U_c h_{t-1} + b_c) \\
c_t &= f_t \odot c_{t-1}^* + i_t \odot \tilde{c}_t \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned}$$

This novel design ($c_{t-1}^S \odot \sigma(W_t \Delta t + b_t)$) allows fully data-driven learning for the time decay function and adaptive discount rates on different components of the short-term memory, c_{t-1}^S , which avoids the manual selection of a uniform time decay function. This non-trivial extension of TLSTM enables the model to better capture the time irregularities in data, thus enhancing model performance.

5.4.4. Evaluation

To validate the technical superiority of our proposed design, we systematically evaluated ATCLSTM against three sets of benchmark experiments. The first experiment evaluates whether ATCLSTM can outperform widely applied single feature-based health profiling models. The second experiment benchmarks the performance of ATCLSTM against classic machine learning models. Finally, the third experiment compares ATCLSTM with alternative deep learning models. The intuition, procedures, and metrics used in each experiment are detailed below.

Single Feature-Based Models

Single feature-based models are the most widely used approach for motion sensor-based disease severity assessment and health profiling (Howcroft et al. 2013; Hubble et al. 2015). We created three benchmark models on three most commonly investigated features: stride-time variability (SVAR), acceleration root mean square (ARMS), and test completion time (TIME).

Stride-time variability (SVAR). SVAR is the ratio of the standard deviation of gait cycle durations over their mean. If there are N gait cycles in the test instance and we denote the duration of gait cycle i as t_i , SVAR is defined as follows:

$$SVAR = \frac{1}{\bar{t}} \sqrt{\frac{1}{N} \sum_{i=1}^N (t_i - \bar{t})^2},$$

where $\bar{t} = \frac{1}{N} \sum_{i=1}^N t_i$. Higher SVAR indicates more variability and less stability in gait patterns.

Acceleration root mean square (ARMS). If there are M acceleration signals in the test instance and we denote the magnitude of signal j as a_j , ARMS is defined as follows:

$$ARMS = \sqrt{\frac{1}{M} \sum_{j=1}^M a_j^2}.$$

A lower ARMS shows smaller changes in acceleration magnitudes, which indicates milder and smaller steps.

Test Completion Time (TIME). For walking tests, the completion time is widely used as a criterion to assess the participant’s disease severity (Shumway-Cook et al. 2000). We recorded the completion time for each 20-step walking test case as instructed by the mPower study protocol.

Classic Machine Learning Models

We compared our ATCLSTM with three classic machine learning models based on combined features extracted from sensor data. The models included support vector machines (SVM), k-nearest neighbors (KNN), and linear regression (LR). We extracted the most widely used features in the fall risk prediction literature, including acceleration mean, acceleration standard deviation, acceleration root mean square (ARMS), jerk mean, jerk RMS, stride time variability (SVAR), and completion time (TIME) (Hubble et al. 2015).

Alternative Deep Learning Models

To demonstrate the advantage of ATCLSTM over prior deep learning models, we set up four LSTM-based benchmark models for comparison: plain LSTM (raw sensor signals as input to LSTM without the time-aware module), TLSTM (raw sensor signals as input to LSTM with the uniform time-aware module as in Baytas et al. (2017)), CLSTM (abstract features generated by the convolutional module as input to LSTM without the time-aware module) (Ordóñez and Roggen 2016), and TCLSTM (abstract features generated by the convolutional

module as input to LSTM with the uniform time-aware module as in Baytas et al. (2017)). For a fair comparison, the other hyperparameters (structure of layers, number of neurons, etc.) are identical to our proposed ATCLSTM.

Evaluation Metrics

As the prediction of the severity of Parkinson’s disease is a regression task, we report root mean squared error (RMSE) and mean absolute error (MAE) as the evaluation metrics.

RMSE and MAE are given by:

$$RMSE = \sqrt{\frac{1}{n} \sum_i (y_i - \hat{y}_i)^2}, \text{ and}$$

$$MAE = \frac{1}{n} \sum_i |y_i - \hat{y}_i|,$$

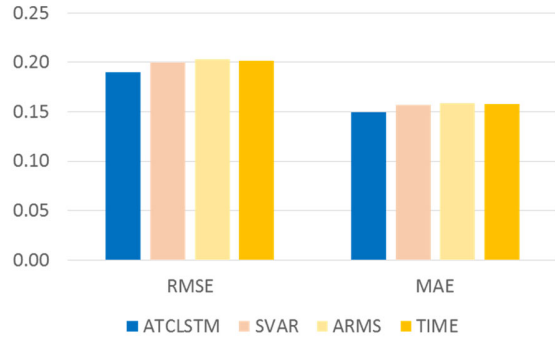
where y_i is the true value for the i -th instance and \hat{y}_i is the predicted value for the i -th instance. As both RMSE and MAE measure the errors of the prediction, a lower score represents a higher predictive power.

Following machine learning practices, we split the 825 participants into a training set (80%) and a test set (20%). We train our ATCLSTM and benchmark models on the training set and report the evaluation metrics on the test set. As deep learning models require random initialization on the parameters, we perform the experiments for five times and report the means of RMSE and MAE, respectively. We also conducted two-sample t-tests to identify whether our ATCLSTM significantly outperformed the benchmark models.

5.5. Experiment Results

5.5.1. Single Feature-Based Models

	RMSE	MAE
ATCLSTM	0.1898	0.1492
SVAR	0.2002***	0.1568***
ARMS	0.2030***	0.1588***
TIME	0.2019***	0.1578***



Note: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$

Figure 5.6: Experiment Results for Single Feature-Based Models

We compared our proposed ATCLSTM with three widely used single feature-based models, SVAR, ARMS, and TIME (Figure 5.6). Our model achieved RMSE of 0.1898 and MAE of 0.1492, significantly outperforming all three benchmark models. SVAR achieved the best performance among the three benchmarks (RMSE 0.2002; MAE 0.1568). Although the single feature-based models are easy to implement and interpret, they focus on only one aspect of sensor signals (e.g., test completion time) and are oversimplified for health profiling. As a result, our model profiles health progression more accurately than the single feature-based benchmark models.

5.5.2. Classic Machine Learning Models

	RMSE	MAE
ATCLSTM	0.1898	0.1492
LR	0.2032***	0.1575**
KNN	0.2232***	0.1727***
SVM	0.2065***	0.1536*

Note: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$

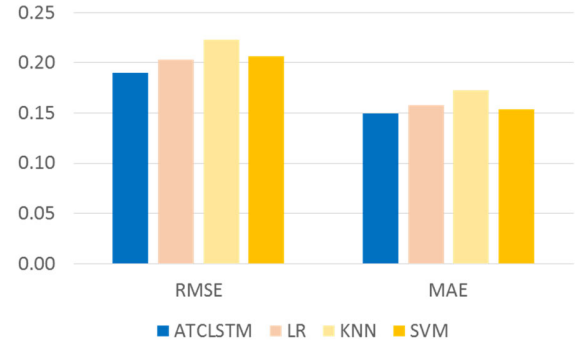


Figure 5.7: Experiment Results for Classic Machine Learning Models

We compared our proposed ATCLSTM with three classic machine learning models, LR, KNN, and SVM (Figure 5.7). Our model achieved RMSE of 0.1898 and MAE of 0.1492, significantly outperforming all three benchmark models. LR and SVM achieved similar results, while KNN's performance was worse than the others by large margins (0.2232 vs. 0.2065 in RMSE; 0.1727 vs. 0.1575 in MAE). One interesting observation is that the machine learning models are not significantly outperforming the single feature-based models. The best-performing machine learning models, LR and SVM (0.2032 in RMSE; 0.1536 in MAE), are merely comparable to the best-performing single feature-based model, SVAR (0.2002 in RMSE; 0.1568 in MAE). This result suggests that adding more manually engineered features into the model does not necessarily increase the predictive power. Furthermore, the information extracted by classic machine learning models is insufficient to handle high-dimensional motion sensor data. In contrast, our ATCLSTM comprehensively utilizes sensor data and automatically learns salient patterns, resulting in improved assessment results compared to the classic machine learning benchmark models.

5.5.3. Alternative Deep Learning Models

	RMSE	MAE
ATCLSTM	0.1898	0.1492
TCLSTM	0.1974**	0.1599*
CLSTM	0.1946*	0.1539*
TLSTM	0.2038**	0.1668**
LSTM	0.2106**	0.1732**

Note: *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$

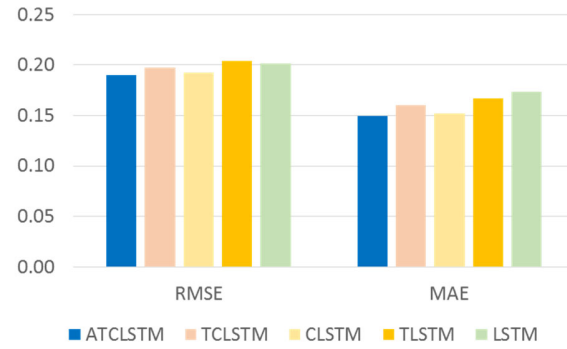


Figure 5.7. Experiment Results for Alternative Deep Learning Models

We compared our proposed ATCLSTM with four alternative deep learning models, TCLSTM, CLSTM, TLSTM, and LSTM (Figure 5.7). Our model achieved RMSE of 0.1898 and MAE of 0.1492, significantly outperforming all four benchmark models. We observed that by introducing the convolutional module, TCLSTM and CLSTM achieved superior performance compared to TLSTM and LSTM. However, although TLSTM outperformed LSTM, TCLSTM did not outperform CLSTM. This result shows that simply adding a manual selected scalar time decay function (e.g., $g(\Delta t) = 1/\log(e + \Delta t)$) to CLSTM may lead to suboptimal performance. By introducing the adaptive time-aware module, ATCLSTM is able to learn the optimal time decay function for each short-term memory component from data, which further enhances the model's ability in modeling time irregularities and improves the accuracies for temporal health profiling.

5.6. Discussion and Conclusions

Providing high-quality healthcare for senior citizens is an increasingly significant societal issue. To better manage chronic conditions such as Parkinson’s disease, long-term health profiling has emerged as an effective tool for prompt therapies and interventions. Motion sensor-based analysis has emerged as the prevailing approach to collect mobility data from senior citizens. Despite their promise, existing health profiling methods often achieve inferior health profiling results due to manual feature engineering (ad hoc, labor intensive, requires significant domain knowledge) and not considering variable time intervals between test cases. These limitations have motivated an alternative approach for long-term health profiling. In this study, we propose a novel ATCLSTM model to alleviate these issues. Grounded in established deep learning literature, ATCLSTM automatically extracts abstract features from raw motion sensor signals, models multiple test cases as time sequences, and adaptively learns an individual time decay function for each short-term memory component. Rigorous evaluations against prevailing feature engineering and deep learning-based health profiling models validated ATCLSTM’s performance on the selected ground-truth dataset.

The contributions of this research are two-fold. First, the ATCLSTM is an advanced analytics model that provides more accurate health profiling for early disease detection and intervention. For instance, for a 75-year-old male senior citizen with a summed MDS-UPDRS score of 15 and another 65-year-old male senior citizen with a score of 13, the ATCLSTM was able to generate assessment scores of 14.4 and 13.5, respectively, accurately detecting the onset of early-stage PD. However, the best performing benchmark model (CLSTM) was only able to generate scores of 5.2 and 2.7 with large errors, failing to recognize the cases. Accurately identifying early-stage PD is of great significance for PD management because health

researchers and practitioners have proposed management strategies aimed at early treatment of PD (Pagan et al. 2012), which alleviates PD progression, improves quality of life with PD, and leads to less financial burden (a 20% reduction in its progression would result in net monetary benefits of \$60,657 per patient (Johnson et al. 2013)). Second, the adaptive time-aware module in ATCLSTM is applicable for motion sensor-based health profiling in other mobility-related health conditions, e.g., frailty, dementia, etc. Future studies can examine the boundaries of ATCLSTM. Furthermore, ATCLSTM is generalizable for other time series data with time irregularity, e.g., EHR, social media timeline, etc., not limited to motion sensor data. It is promising to explore other data sources and potentially fuse multiple sources to provide more advanced and comprehensive healthcare.

6. CONCLUSIONS AND FUTURE DIRECTIONS

Advanced information technologies such as mobile sensing devices are transforming healthcare for both providers and receivers (Chen et al. 2012). Consumer electronics, such as iPhone, Apple Watch, and Fitbit, and specialized mobile sensor-based caregiving platforms proposed by industry and academia have opened up a new and promising multi-disciplinary area of research: mobile health analytics. With the vast amount of personal motion data being collected and stored, mobile health analytics has been receiving increased focus in the IS community. To a large extent, however, we lack specialized IT artifacts for mobile health analytics to better harness the abundant data for advanced business intelligence.

This dissertation takes an important first step in mobile health analytics in the IS discipline. Through the course of four essays, this dissertation explores meaningful uses of high-velocity motion sensor data for senior care, specifically detect adverse events, such as falls, and assess health condition risks and severities, such as Parkinson's disease. Although each essay has a particular concentration, this dissertation is drawn upon a unified data mining and deep learning framework, which also offers practical values for stakeholders including patients, their families, health practitioners, and researchers. The research contributions of each essay is summarized below.

6.1. Research Contributions

The first essay sets the tone of the dissertation by asking the question "how can we promptly detect senior citizens' adverse events, e.g., falls, to alleviate consequences?" This essay presented a novel HMM-CAL model for advanced fall detection. By leveraging the temporal modeling ability of HMM and robustness of the sensor orientation calibration

algorithm, HMM-CAL was able to be deployed in an unsupervised real-life environment to detect falls for senior citizens. The improved sensitivity and reduced false alarm rate compared to benchmark models signaled an increased efficacy in fall detection, which is critical for such a system to be practically used in senior citizens' homes.

Although more severe outcomes could be prevented by timely fall detection, direct injuries such as bone fractures cannot be avoided. This drove the second essay's core research question: "how can we precisely assess senior citizens' health risks, e.g., fall risks, to provide proper interventions," and to prevent falls? By proposing the 2D-hetero CNN model, we were able to predict fall risks by asking senior citizens to perform a 10-meter walking mobility test. In addition, we provided our unique contributions to IS knowledge base by not only proposing effective IT artifacts, but also examining generalizable design principles including data and application characteristics, deep learning method selection, and model building suggestions, which could be helpful for future IS and health researchers.

The third essay moved one step further on the second essay, asking the question "how can we assess senior citizens' health risks in a more holistic manner?" By combining multiple data sources to assess multiple condition risks and severities in a more comprehensive way, a Multisource Multitask Learning for Mobile Health Analytics (DMML-MHA) framework was proposed for transferable health condition risk and severity assessment. Source-specific, general, and task-specific deep learning layers were designed to extract salient features from motion sensor signals and alleviate model overfitting to improve assessment results. An internal clinical experiment dataset and an external publicly available dataset were collected to rigorously evaluate the proposed framework against state-of-the-art benchmark models. The

proposed DMML-MHA framework could contribute to improved life quality for senior citizens and provide design guidelines for future IS research in mobile health analytics.

Finally, the fourth essay expanded the goal of assessing health condition risk and severity to long-term health profiling with disease progression modeling. An advanced deep learning model, ATCLSTM, was proposed to answer the question “how can we profile senior citizens’ long-term health progression for more personalized care?” With its convolutional and adaptive time-aware long short term memory modules, the ATCLSTM was able to automatically extract features from raw motion sensor signals, model multiple test cases as time sequences, and adaptively learn optimal time decay functions for its memory components. Rigorous evaluations against prevailing feature engineering and deep learning-based health profiling models validated ATCLSTM’s performance on the selected mPower dataset.

6.2. Future Research Directions

As with any emerging research, this dissertation leads to promising future directions for IS scholars, which could address more interesting and higher-impact research inquiries in the realm of mobile health analytics. An incomplete list of such possible opportunities can be found below.

(1) *Emotion Detection with Motion Sensors and Social Media.* Emotion plays an important role in human health, especially for those with emotional disorders (such as depression and anxiety). Detecting or predicting human’s emotion may contribute to prevention and treatment of such disorders. Although motion sensors can capture an aspect of such disorders (for instance, people with anxiety may find them more difficult to stay still for a relatively long time), a person’s language expressions are a significant factor that reflects his

or her mental status. Social media is an ideal data source for one's textual language expressions, especially with the wide use of emoji. By simultaneously collecting motion sensor data and social media postings, we may be able to more holistically assess one's motion status. The data collection may be done by smartphone apps, which could access both motion sensors and social media accounts on the same device.

(2) *Healthcare with Ubiquitous Sensors.* Throughout the course of this dissertation, motion sensors are used as the input for senior care, which may be effective for predicting falls, Parkinson's disease, among others. However, there exist numerous diseases that cannot be detected or predicted by motion sensors, with cardiac diseases as an impactful example. Such diseases require the access to additional aspects of health information, including physiological biomarkers (heart rate, blood pressure, blood glucose, etc.), EHR and radiology data, and even genomic data. With the rapid development of sensors (of every type), it is promising that various data sources could be fused (as in Essay III) to support more intelligent healthcare with the minimal need of clinical visits and cumbersome examinations.

(3) *Interpretable Healthcare with Deep Learning.* Essays II, III, and IV in this dissertation were developed in deep learning methodologies. However, deep learning models are generally considered "black box approaches," which confront the controversial interpretability issue. Instead of more accurate but less transparent deep learning systems, many health practitioners are more confident with feature-based (sometimes even single feature-based) systems. To increase its interpretability, it is promising to explore black-box probing techniques to reveal the features learned in deep learning models. Potential directions include adding layers with transparent relationships of variables to control the black box, perturbing inputs to identify what most affects model decision making, and applying a

generator versus classifier structure (e.g., generative adversarial networks) to enhance the model interpretability. The black-box probing techniques can lead to the discovery of clinically meaningful features across multiple data sources (e.g., motion sensor data, audio data, etc.).

7. REFERENCES

- Abbasi, A., & Chen, H. (2008). CyberGate: A Design Framework and System for Text Analysis of Computer-Mediated Communication. *MIS Quarterly*, 32(4), 811–837.
- Abbasi, A., Zhang, Z., Zimbra, D., Chen, H., & Nunamaker Jr., J. F. (2010). Detecting Fake Websites: the Contribution of Statistical Learning Theory. *MIS Quarterly*, 34(3), 435–461. <http://doi.org/Article>
- Abdullah, S., Matthews, M., Frank, E., Doherty, G., & Gay, G. (2016). Automatic detection of social rhythms in bipolar disorder. *Journal of the American Medical Informatics Association*, 23, 538–543. <http://doi.org/10.1093/jamia/ocv200>
- Adipat, B., Zhang, D., & Zhou, L. (2011). The Effects of Tree-View Based Presentation Adaptation on Mobile Web Browsing. *MIS Quarterly*, 35(1), 99–121. Retrieved from <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=2944&context=misq>
- Agarwal, R., Dhar, V. (2014). Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Information Systems Research*, 25(3), 443–448.
- Akhbari, M., Shamsollahi, M. B., Sayadi, O., Armoundas, A. A., & Jutten, C. (2016). ECG segmentation and fiducial point extraction using multi hidden Markov model. *Computers in Biology and Medicine*, 79, 21–29.
- Anderson, C. L., & Agarwal, R. (2011). The Digitization of Healthcare: Boundary Risks, Emotion, and Consumer Willingness to Disclose Personal Health Information. *Information Systems Research*, 22(3), 469–490. <http://doi.org/10.1287/isre.1100.0335>
- Angst, C. M., Wowak, K. D., Handley, S. M., & Kelley, K. (2017). Antecedents of information systems sourcing strategies in U.S. hospitals: A longitudinal study. *MIS Quarterly*, 41(4), 1–18. <http://doi.org/10.25300/MISQ/2017/41.4.06>
- Angst, C. M., & Agarwal, R. (2016). Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion. *MIS Quarterly*, 33(2), 339–370.
- Angst, C. M., Block, E. S., D’Arcy, J., Kelley, K., D’Arcy, J., & Kelley, K. (2017). When Do IT Security Investments Matter? Accounting for the Influence of Institutional Factors in the Context of Healthcare Data Breaches. *MIS Quarterly*, 41(3), 893–916. <http://doi.org/10.25300/MISQ/2017/41.3.10>
- Anthimopoulos, M., Christodoulidis, S., Ebner, L., Christe, A., & Mougiakakou, S. (2016). Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, 35(5), 1207–1216.

- Ayabakan, S., Bardhan, I., Zheng, Z., & Kirksey, K. (2017). The Impact of Health Information Sharing on Duplicate Testing. *MIS Quarterly*, 41(4), 1083–1103.
- Bagalà, F., Becker, C., Cappello, A., Chiari, L., Aminian, K., Hausdorff, J. M., ... Klenk, J. (2012). Evaluation of Accelerometer-Based Fall Detection Algorithms on Real-World Falls. *PloS One*, 7(5), e37062. <http://doi.org/10.1371/journal.pone.0037062>
- Baird, A., Angst, C., & Oborn, E. (2018). MISQ Research Curation on Health Information Technology. *MIS Quarterly*, 1–14. Retrieved from <http://misq.org/research-curations>
- Bakker, B., & Heskes, T. (2003). Task Clustering and Gating for Bayesian Multitask Learning. *Journal of Machine Learning Research*, 4(May), 83–99.
- Bardhan, I., Oh, J. (Cath), Zheng, Z. (Eric), & Kirksey, K. (2015). Predictive Analytics for Readmission of Patients with Congestive Heart Failure. *Information Systems Research*, 26(1), 19–39. <http://doi.org/10.1287/isre.2014.0553>
- Bergen, G., Stevens, M. R., & Burns, E. R. (2016). Falls and Fall Injuries Among Adults Aged ≥ 65 Years — United States, 2014. *Morbidity and Mortality Weekly Report (MMWR)*, 65(37). Retrieved from https://www.cdc.gov/mmwr/volumes/65/wr/mm6537a2.htm?s_cid=mm6537a2_w
- Bot, B. M., Suver, C., Neto, E. C., Kellen, M., Klein, A., Bare, C., ... Trister, A. D. (2016). The mPower Study, Parkinson Disease Mobile Data Collected Using ResearchKit. *Scientific Data*, 3, 1–9. <http://doi.org/10.1038/sdata.2016.11>
- Bourke, A. K., Brien, J. V. O., & Lyons, G. M. (2007). Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm. *Gait & Posture*, 26, 194–199. <http://doi.org/10.1016/j.gaitpost.2006.09.012>
- Bourke, A. K., Klenk, J., Schwickert, L., Aminian, K., Ihlen, E. A. F., Mellone, S., ... Becker, C. (2016). Fall detection algorithms for real - world falls harvested from lumbar sensors in the elderly population : A machine learning approach. In *2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)* (pp. 3712–3715).
- Burns, E. R., Stevens, J. A., & Lee, R. (2016). The Direct Costs of Fatal and Non-fatal Falls among Older Adults — United States. *Journal of Safety Research*, 58, 99–103. <http://doi.org/10.1016/j.jsr.2016.05.001>
- Caby, B., Kieffer, S., Hubert, M. D. Saint, Cremer, G., & Macq, B. (2011). Feature Extraction and Selection for Objective Gait Analysis and Fall Risk Assessment by Accelerometry. *Biomedical Engineering Online*, 10(1), 1–19.

- Cai, F., & Cherkassky, V. (2012). Generalized SMO Algorithm for SVM-based Multitask Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 23(6), 997–1003.
- Caruana, R. (1998). Multitask Learning. In *Learning to Learn* (pp. 95–133).
- Chang, Y.-W., Hsieh, C.-J., Chang, K.-W., Ringgaard, M., & Lin, C.-J. (2010). Training and Testing Low-degree Polynomial Data Mappings via Linear SVM. *Journal of Machine Learning Research*, 11, 1471–1490. Retrieved from http://www.csie.ntu.edu.tw/~cjlin/papers/lowpoly_journal.pdf
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 36(4), 1165–1188.
- Chen, Y., & Xue, Y. (2015). A deep learning approach to human activity recognition based on single accelerometer. In *2015 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 1488–1492).
- Chuang, J., Maimoon, L., Yu, S., Zhu, H., Nybroe, C., Hsiao, O., ... Chen, H. (2015). Silverlink: Smart Home Health Monitoring for Senior Care. In *Proceedings of the International Conference on Smart Health* (pp. 3–14). Phoenix, AZ, USA. http://doi.org/10.1007/978-3-319-29175-8_1
- Cui, L., Xie, X., Shen, Z. M., Lu, R., & Wang, H. (2018). Prediction of the healthcare resource utilization using multi-output regression models. *IJSE Transactions on Healthcare Systems Engineering*, 5579, 1–25. <http://doi.org/10.1080/24725579.2018.1512537>
- Daher, M., El, M., El, B., Ascq, V., Diab, A., & Khalil, M. (2016). Automatic Fall Detection System using Sensing Floors. *International Journal of Computing & Information Sciences*, 12(1), 75–82.
- Dai, X., Wu, M., Davidson, B., Mahoor, M., & Zhang, J. (2013). Image-based fall detection with human posture sequence modeling. In *2013 IEEE International Conference on Healthcare Informatics* (pp. 376–381).
- Dennis, A. R., & Garfield, M. J. (2003). The Adoption and Use of GSS in Project Teams: Toward More Participative Processes and Outcomes. *MIS Quarterly*, 27(2), 289–323. <http://doi.org/10.2307/4132321>
- Ejupi, A., Member, S., Brodie, M., Lord, S. R., Annegarn, J., Redmond, S. J., ... Delbaere, K. (2017). Wavelet-Based Sit-To-Stand Detection and Assessment of Fall Risk in Older People Using a Wearable Pendant Device. *IEEE Transactions on Biomedical Engineering*, 64(7), 1602–1607. <http://doi.org/10.1109/TBME.2016.2614230>

- Fan, X., Zheng, K., Lin, Y., & Wang, S. (2015). Combining Local Appearance and Holistic View: Dual-Source Deep Neural Networks for Human Pose Estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1347–1355). Boston, MA, USA. <http://doi.org/10.1109/CVPR.2015.7298740>
- Florence, C. S., Bergen, G., Atherly, A., Burns, E., Stevens, J., & Drake, C. (2018). Medical Costs of Fatal and Nonfatal Falls in Older Adults. *Journal of the American Geriatrics Society*, 66(4), 693–698. <http://doi.org/10.1111/jgs.15304>
- Fried, L. P., Tangen, C. M., Walston, J., Newman, A. B., Hirsch, C., Gottdiener, J., ... Burke, G. (2001). Frailty in Older Adults: Evidence for a Phenotype. *Journal of Gerontology: Medical Sciences*, 56(3), M146–M157.
- Gao, G. (Gordon), Smith, R. H., Greenwood, B. N., Agarwal, R., & Mccullough, J. S. (2015). Vocal Minority and Silent Majority: How Do Online Ratings Reflect Population Perceptions of Quality. *MIS Quarterly*, 39(3), 565–589. Retrieved from <http://www.fixedeffects.com/wp-content/uploads/2015/02/Gao-et-al-2015.pdf>
- Gao, L., Bourke, A. K., & Nelson, J. (2014). Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems. *Medical Engineering & Physics*, 36(6), 779–785. <http://doi.org/10.1016/j.medengphy.2014.02.012>
- Ghose, A., Goldfarb, A., & Han, S. P. (2013). How Is the Mobile Internet Different? Search Costs and Local Activities. *Information Systems Research*, 24(3), 613–631. <http://doi.org/10.1287/isre.1120.0453>
- Gooch, C. L., Pracht, E., & Borenstein, A. R. (2017). The Burden of Neurological Disease in the United States: A Summary Report and Call to Action. *Annals of Neurology*, 81(4), 479–484. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=28198092
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Greene, B. R., Redmond, S. J., & Caulfield, B. (2017). Fall Risk Assessment Through Automatic Combination of Clinical Fall Risk Factors and Body-Worn Sensor Data. *IEEE Journal of Biomedical and Health Informatics*, 21(3), 725–731. <http://doi.org/10.1109/JBHI.2016.2539098>
- Gregor, S., & Hevner, A. R. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly*, 37(2), 337–355. <http://doi.org/10.2753/MIS0742-1222240302>
- Guo, S., Zoeter, O., & Archambeau, C. (2011). Sparse Bayesian Multi-task Learning. In *Advances in Neural Information Processing Systems* (pp. 1755–1763).

- Hagui, M., Mahjoub, M. A., & Elayeb, F. (2015). Coupled Hidden Markov Model for video fall detection. In *2015 11th International Conference on Natural Computation (ICNC)* (pp. 675–679).
- Hammerla, N. Y., Halloran, S., & Plötz, T. (2016). Deep, Convolutional, and Recurrent Models for Human Activity Recognition Using Wearables. In *International Joint Conference on Artificial Intelligence* (pp. 1533–1540).
<http://doi.org/10.1002/cphy.cp060328>
- He, T., Mao, H., Guo, J., & Yi, Z. (2017). Cell Tracking Using Deep Neural Networks with Multi-task Learning. *Image and Vision Computing*, 60, 142–153.
<http://doi.org/10.1016/j.imavis.2016.11.010>
- Heintzman, N., & Kleinberg, S. (2016). Using uncertain data from body-worn sensors to gain insight into type 1 diabetes. *Journal of Biomedical Informatics*, 63, 259–268.
<http://doi.org/10.1016/j.jbi.2016.08.022>
- Henpraserttae, A., Thiemjarus, S., & Marukatat, S. (2011). Accurate Activity Recognition using a Mobile Phone regardless of Device Orientation and Location. In *2011 International Conference on Body Sensor Networks (BSN)* (pp. 41–46).
<http://doi.org/10.1109/BSN.2011.8>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, 28(1), 75–105.
- Hoehle, H., & Venkatesh, V. (2015). Mobile Application Usability: Conceptualization and Instrument Development. *MIS Quarterly*, 39(2), 435–472.
- Hosseini-asl, E., Keynton, R., & El-baz, A. (2016). Alzheimer’s Disease Diagnostics by Adaptation of 3D Convolutional Network. In *2016 IEEE International Conference on Image Processing (ICIP)* (pp. 126–130).
- Howcroft, J., Kofman, J., & Lemaire, E. D. (2013). Review of Fall Risk Assessment in Geriatric Populations Using Inertial Sensors. *Journal of NeuroEngineering and Rehabilitation*, 10(1), 91. <http://doi.org/10.1186/1743-0003-10-91>
- Huang, J., Breheny, P., & Ma, S. (2012). A Selective Review of Group Selection in High-dimensional Models. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, 27(4).
- Huang, Z., Cheng, G., Wang, H., Li, H., Shi, L., & Pan, C. (2016). Building Extraction from Multi-source Remote Sensing Images via Deep Deconvolution Neural Networks. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (pp. 1835–1838). Beijing, China.
<http://doi.org/10.1109/IGARSS.2016.7729471>

- Hubble, R. P., Naughton, G. A., Silburn, P. A., & Cole, M. H. (2015). Wearable Sensor Use for Assessing Standing Balance and Walking Stability in People with Parkinson's Disease: A Systematic Review. *PloS One*, 10(4), e0123705. <http://doi.org/10.1371/journal.pone.0123705>
- Ibrahim, M. S., Muralidharan, S., Deng, Z., Vahdat, A., & Mori, G. (2016). A hierarchical deep temporal model for group activity recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1971–1980).
- Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D Convolutional Neural Networks for Human Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 221–231.
- Jiang, M., Chen, Y., Zhao, Y., & Cai, A. (2013). A real-time fall detection system based on HMM and RVM. In *2013 Visual Communications and Image Processing (VCIP)* (pp. 1–6).
- Jonas, S. M., Deserno, T. M., Buhimschi, C. S., Makin, J., Choma, M. A., & Buhimschi, I. A. (2016). Smartphone-based diagnostic for preeclampsia: an mHealth solution for administering the Congo Red Dot (CRD) test in settings with limited resources. *Journal of the American Medical Informatics Association*, 23, 166–173. <http://doi.org/10.1093/jamia/ocv015>
- Kale, N., Lee, J., Lotfian, R., & Jafari, R. (2012). Impact of Sensor Misplacement on Dynamic Time Warping Based Human Activity Recognition using Wearable Computers. In *Proceedings of the Conference on Wireless Health* (p. 7).
- Kangas, M., Konttila, A., Lindgren, P., Winblad, I., & Ja, T. (2008). Comparison of low-complexity fall detection algorithms for body attached accelerometers. *Gait & Posture*, 28, 285–291. <http://doi.org/10.1016/j.gaitpost.2008.01.003>
- Kau, L., & Chen, C. (2015). A Smart Phone-Based Pocket Fall Accident Detection, Positioning, and Rescue System. *IEEE Journal of Biomedical and Health Informatics*, 19(1), 44–56.
- Kautz, T., Groh, B. H., Hannink, J., Jensen, U., Strubberg, H., & Eskofier, B. M. (2017). Activity recognition in beach volleyball using a Deep Convolutional Neural Network. *Data Mining and Knowledge Discovery*, 31(6), 1678–1705.
- Kerr, G. K., Worringham, C. J., Cole, M. H., Lacherez, P. F., Wood, J. M., & Silburn, P. A. (2010). Predictors of Future Falls in Parkinson Disease. *Neurology*, 75(2), 116–124.
- Khan, S. S., Karg, M. E., & Kuli, D. (2014). X-Factor HMMs for Detecting Falls in the Absence of Fall-Specific Training Data. In *International Workshop on Ambient Assisted Living* (pp. 1–9).

- Khorasani, A., & Daliri, M. R. (2014). HMM for Classification of Parkinson's Disease Based on the Raw Gait Data. *Journal of Medical Systems*, 38(12), 147. <http://doi.org/10.1007/s10916-014-0147-5>
- Kiranyaz, S., Ince, T., & Gabbouj, M. (2016). Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Transactions on Biomedical Engineering*, 63(3), 664–675.
- Kwon, H. E., So, H., Han, S. P., & Oh, W. (2016). Excessive Dependence on Mobile Social Apps: A Rational Addiction Perspective. *Information Systems Research*, 27(4), 919–939. <http://doi.org/10.1287/isre.2016.0658>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521(7553), 436–444. <http://doi.org/10.1038/nature14539>
- Lee, J. K., Robinovitch, S. N., Park, E. J., & Member, S. (2015). Inertial Sensing-Based Pre-Impact Detection of Falls Involving Near-Fall Scenarios. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 23(2), 258–266.
- Li, Y., Vinzamuri, B., & Reddy, C. K. (2015). Constrained Elastic Net Based Knowledge Transfer for Healthcare Information Exchange. *Data Mining and Knowledge Discovery*, 29(4), 1094–1112.
- Li, Z., Wei, Z., Yue, Y., Wang, H., Jia, W., Burke, L. E., & Baranowski, T. (2015). An Adaptive Hidden Markov Model for Activity Recognition Based on a Wearable Multi-Sensor Device. *Journal of Medical Systems*, 39(5), 57. <http://doi.org/10.1007/s10916-015-0239-x>
- Liang, V. C., Ma, R. T. B., Ng, W. S., Wang, L., Winslett, M., Wu, H., ... Zhang, Z. (2016). Mercury: Metro density prediction with recurrent neural network on streaming CDR data. In *Proceeding of the 2016 IEEE 32nd International Conference on Data Engineering (ICDE)* (pp. 1374–1377).
- Lim, D., Park, C., Kim, N. H., Kim, S., & Yu, Y. S. (2014). Fall-Detection Algorithm Using 3-Axis Acceleration: Combination with Simple Threshold and Hidden Markov Model. *Journal of Applied Mathematics*.
- Lin, Y., Chen, H., & Brown, R. A. (2017). Healthcare Predictive Analytics for Risk Profiling in Chronic Care: A Bayesian Multitask Learning Approach. *MIS Quarterly*, 41(2), 473–495.
- Lipton, Z. C., Kale, D. C., Elkan, C., & Wetzel, R. (2016). Learning to diagnose with LSTM recurrent neural networks. In *Proceeding of the 4th International Conference on Learning Representations (ICLR)* (pp. 1–18).

- Liu, J., & Lockhart, T. E. (2014). Development and Evaluation of a Prior-to-Impact Fall Event Detection Algorithm. *IEEE Transactions on Biomedical Engineering*, 61(7), 2135–2140.
- Lord, S. R., Menz, H. B., & Tiedemann, A. (2003). A Physiological Profile Approach to Falls Risk Assessment and Prevention. *Physical Therapy*, 83(3), 237–252.
<http://doi.org/10.1056/nejm198812293192604>
- Ma, X., Yu, H., Wang, Y., & Wang, Y. (2015). Large-scale transportation network congestion evolution prediction using deep learning theory. *PloS One*, 10(3), e0119044.
- Maimoon, L., Chuang, J., Zhu, H., Yu, S., Peng, K.-S., Prayakarao, R., ... Chen, H. (2016). SilverLink: Developing an International Smart and Connected Home Monitoring System for Senior Care. In *Proceedings of the International Conference on Smart Health* (pp. 65–77). Haikou, China.
- Manic, M., Amarasinghe, K., Rodriguez-Andina, J. J., & Rieger, C. (2016). Intelligent buildings of the future: Cyberaware, deep learning powered, and human interacting. *IEEE Industrial Electronics Magazine*, 10(4), 32–49.
- Mannini, A., Trojaniello, D., Croce, U. Della, & Sabatini, A. M. (2015). Hidden Markov Model-Based Strategy for Gait Segmentation using Inertial Sensors: Application to Elderly, Hemiparetic Patients and Huntington's Disease Patients. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 5179–5182). IEEE.
- Marier, A., Olsho, L. E. W., Rhodes, W., & Spector, W. D. (2016). Improving prediction of fall risk among nursing home residents using electronic medical records. *Journal of the American Medical Informatics Association*, 276–282.
<http://doi.org/10.1093/jamia/ocv061>
- Millor, N., Lecumberri, P., Gómez, M., Martínez, A., Martinikorena, J., Rodríguez-Mañas, L., ... Izquierdo, M. (2017). Gait Velocity and Chair Sit-Stand-Sit Performance Improves Current Frailty-Status Identification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(11), 2018–2025.
<http://doi.org/10.1109/TNSRE.2017.2699124>
- Miscione, G. (2017). Telemedicine in the Upper Amazon: Interplay with Local Health Care Practices. *MIS Quarterly*, 31(2), 403–425.
- Mitchell, V. L. (2006). Knowledge Integration and Information Technology Project Performance. *MIS Quarterly*, 30(4), 919–939.
- Mizell, D. (2003). Using Gravity to Estimate Accelerometer Orientation. In *Proceedings of the 7th IEEE International Symposium on Wearable Computers* (p. 252).

- Mohammadi, M., Al-Fuqaha, A., Sorour, S., & Guizani, M. (2018). Deep learning for IoT big data and streaming analytics: A survey. *IEEE Communications Surveys & Tutorials*, 20(4), 2923–2960.
- Morales, J., Akopian, D., Agaian, S., & Antonio, S. (2014). Human Activity Recognition by Smartphones Regardless of Device Orientation. *SPIE IS&T*, 9030, 90300I. <http://doi.org/10.1117/12.2043180>
- Münzner, S., Schmidt, P., Reiss, A., Hanselmann, M., Stiefelhagen, R., & Dürichen, R. (2017). CNN-based sensor fusion techniques for multimodal human activity recognition. In *Proceedings of the 2017 ACM International Symposium on Wearable Computers* (pp. 158–165).
- Najafi, B., Armstrong, D. G., & Mohler, J. (2013). Novel Wearable Technology for Assessing Spontaneous Daily Physical Activity and Risk of Falling in Older Adults with Diabetes. *Journal of Diabetes Science and Technology*, 7(5), 1147–1160. <http://doi.org/10.1177/193229681300700507>
- Neinstein, A., Wong, J., Look, H., Arbiter, B., Quirk, K., Mccanne, S., ... Adi, S. (2016). A case study in open source innovation: developing the Tidepool Platform for interoperability in type 1 diabetes management. *Journal of the American Medical Informatics Association*, 23, 324–332. <http://doi.org/10.1093/jamia/ocv104>
- Nelson-Wong, E., Appell, R., McKay, M., Nawaz, H., Roth, J., Sigler, R., ... Walker, M. (2012). Increased Fall Risk Is Associated with Elevated Co-contraction about the Ankle during Static Balance Challenges in Older Adults. *European Journal of Applied Physiology*, 112(4), 1379–1389. <http://doi.org/10.1007/s00421-011-2094-x>
- Oliver, D., Britton, M., Seed, P., Martin, F. C., & Hopper, A. H. (1997). Development and Evaluation of Evidence Based Risk Assessment Tool (STRATIFY) to Predict Which Elderly Inpatients Will Fall: Case-Control and Cohort Studies. *BMJ*, 315(7115), 1049–1053.
- Ouyang, W., Chu, X., & Wang, X. (2014). Multi-source Deep Learning for Human Pose Estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2329–2336). Columbus, OH, USA.
- Özdemir, A. T., & Barshan, B. (2014). Detecting Falls with Wearable Sensors Using Machine Learning Techniques. *Sensors*, 14(6), 10691–10708. <http://doi.org/10.3390/s140610691>
- Pagan, F. L. (2012). Improving outcomes through early diagnosis of Parkinson's disease. *American Journal of Managed Care*, 18(7), S176.
- Palermo, E., Rossi, S., Marini, F., Patanè, F., & Cappa, P. (2014). Experimental evaluation of accuracy and repeatability of a novel body-to-sensor calibration procedure for inertial

- sensor-based gait analysis. *Measurement*, 52, 145–155.
<http://doi.org/10.1016/j.measurement.2014.03.004>
- Pereira, C. R., Pereira, D. R., B, J. P. P., & Rosa, G. H. (2016). Convolutional Neural Networks Applied for Parkinson's Disease Identification. In *Machine Learning for Health Informatics* (pp. 377–390). <http://doi.org/10.1007/978-3-319-50478-0>
- Persistence Market Research. (2016). Global Market Study on Fall Detection System: North America to Witness Highest Growth by 2021. Retrieved April 27, 2019, from <http://www.persistencemarketresearch.com/market-research/fall-detection-system-market.asp>
- Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in Neural Information Processing Systems* (pp. 505–513).
- Prathivadi, Y., Wu, J., Bennett, T. R., & Jafari, R. (2014). Robust Activity Recognition using Wearable IMU Sensors, 1–4.
- Rabiner, L. R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2), 257–286.
- Rai, A. (2017). Editor's Comments: Diversity of Design Science Research. *MIS Quarterly*, 41(1), iii–xviii.
- Ray, G., Muhanna, W. A., & Barney, J. B. (2005). Information Technology and the Performance of the Customer Service Process: A Resource-Based Analysis. *MIS Quarterly*, 29(4), 625–652. <http://doi.org/10.2307/4132321>
- Rubenstein, L. Z., & Josephson, K. R. (2002). The epidemiology of falls and syncope. *Clinics in Geriatric Medicine*, 18(2), 141–158.
- Samala, R. K., Chan, H.-P., Hadjiiski, L. M., Helvie, M. A., Cha, K., & Richter, C. (2017). Multi-task Transfer Learning Deep Convolutional Neural Network: Application to Computer-Aided Diagnosis of Breast Cancer on Mammograms. *Physics in Medicine and Biology*, 62(23), 8894. <http://doi.org/10.1088/1361-6560/aa93d4>
- Scheffer, A. C., Schuurmans, M. J., Van Dijk, N., Van Der Hooft, T., & De Rooij, S. E. (2008). Fear of falling: measurement strategy, prevalence, risk factors and consequences among older persons. *Age and Ageing*, 37(1), 19–24.
<http://doi.org/10.1093/ageing/afm169>
- Schwenk, M., Hauer, K., Zieschang, T., Englert, S., Mohler, J., & Najafi, B. (2014). Sensor-Derived Physical Activity Parameters Can Predict Future Falls in People with Dementia. *Gerontology*, 60(6), 483–492. <http://doi.org/10.1159/000363136>

- Schwenk, M., Mohler, J., Wendel, C., Fain, M., Taylor-Piliae, R., & Najafi, B. (2015). Wearable Sensor-Based In-Home Assessment of Gait, Balance, and Physical Activity for Discrimination of Frailty Status: Baseline Results of the Arizona Frailty Cohort Study. *Gerontology*, 61(3), 258–267. <http://doi.org/10.1159/000369095>
- Shen, V. R. L., Lai, H., & Lai, A. (2015). The implementation of a smartphone-based fall detection system using a high-level fuzzy Petri net. *Applied Soft Computing Journal*, 26, 390–400. <http://doi.org/10.1016/j.asoc.2014.10.028>
- Silva, J., & Sousa, I. (2016). Instrumented Timed Up and Go : Fall Risk Assessment based on Inertial Wearable Sensors. In *Proceedings of the IEEE International Symposium on Medical Measurements and Applications (MeMeA)* (pp. 1–6).
- Simm, J., de Abril, I. M., & Sugiyama, M. (2014). Tree-based Ensemble Multi-task Learning Method for Classification and Regression. *IEICE Transactions on Information and Systems*, 97(6), 1677–1681.
- Song, X., Kanasugi, H., & Shibasaki, R. (2016). DeepTransport: Prediction and Simulation of Human Mobility and Transportation Mode at a Citywide Level. In *IJCAI* (Vol. 16, pp. 2618–2624).
- Stone, E. E., & Skubic, M. (2015). Fall Detection in Homes of Older Adults Using the Microsoft Kinect. *IEEE Journal of Biomedical and Health Informatics*, 19(1), 290–301.
- Su, B. Y., Ho, K. C., Rantz, M. J., & Skubic, M. (2015). Using the Wavelet Transform. *IEEE Transactions on Biomedical Engineering*, 62(3), 865–875.
- Suk, H.-I., Lee, S.-W., & Shen, D. (2016). Deep Sparse Multi-task Learning for Feature Selection in Alzheimer’s Disease Diagnosis. *Brain Structure and Function*, 221(5), 2569–2587. <http://doi.org/10.1007/s00429-015-1059-y>
- Sun, Z., Dawande, M., Janakiraman, G., & Mookerjee, V. (2017). Not Just a Fad: Optimal Sequencing in Mobile In-App Advertising. *Information Systems Research*, 28(3), 511–528. <http://doi.org/10.1287/isre.2017.0697>
- Tabak, Y. P., Sun, X., Nunez, C. M., & Johannes, R. S. (2014). Using Electronic Health Record Data to Develop Inpatient Mortality Predictive Model: Acute Laboratory Risk of Mortality Score (ALaRMS). *Journal of the American Medical Informatics Association*, 21(3), 455–463. <http://doi.org/10.1136/amiajnl-2013-001790>
- Tian, Y., & Pan, L. (2015). Predicting short-term traffic flow by long short-term memory recurrent neural network. In *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)* (pp. 153–158).

- Tinetti, M. E. (1986). Performance-Oriented Assessment of Mobility Problems in Elderly Patients. *Journal of the American Geriatrics Society*, 34(2), 119–126. <http://doi.org/10.1111/j.1532-5415.1986.tb05480.x>
- Tinetti, M. E., Liu, W.-L., & Claus, E. B. (1993). Predictors and Prognosis of to Get Up After Falls Among Elderly Persons. *Journal of the American Medical Association*, 269(1), 65–70.
- Tong, L., Song, Q., Ge, Y., & Liu, M. (2013). HMM-Based Human Fall Detection and Prediction Method Using Tri-Axial Accelerometer. *IEEE Sensors Journal*, 13(5), 1849–1856.
- U.S. Census Bureau. (2017). Facts for Features: Older Americans Month: May 2017. Retrieved January 23, 2018, from <https://www.census.gov/newsroom/facts-for-features/2017/cb17-ff08.html>
- Van Schooten, K. S., Pijnappels, M., Rispens, S. M., Elders, P. J. M., Lips, P., & Van Dieën, J. H. (2015). Ambulatory Fall-Risk Assessment: Amount and Quality of Daily-Life Gait Predict Falls in Older Adults. *Journals of Gerontology Series A: Biomedical Sciences and Medical Sciences*, 70(5), 608–615. <http://doi.org/10.1093/gerona/glu225>
- Venkatesh, V., Aloysius, J. A., Hoehle, H., & Burton, S. (2017). Design and Evaluation of Auto-ID Enabled Shopping Assistance Artifacts in Customers' Mobile Phones: Two Retail Store Laboratory Experiments. *MIS Quarterly*, 41(1), 83–113.
- Venkatesh, V., Zhang, X., & Sykes, T. a. (2011). “Doctors Do Too Little Technology”: A Longitudinal Field Study of an Electronic Healthcare System Implementation. *Information Systems Research*, 22(3), 523–546. <http://doi.org/10.1287/isre.1110.0383>
- Vert, J.-P., Tsuda, K., & Schölkopf, B. (2004). A Primer on Kernel Methods. *Kernel Methods in Computational Biology*, 47, 35–70.
- Voosen, P. (2017, July). How AI Detectives Are Cracking Open the Black Box of Deep Learning. *Science*. <http://doi.org/doi:10.1126/science.aan7059>
- Wang, J., Ding, H., Bidgoli, F. A., Zhou, B., Iribarren, C., Molloy, S., & Baldi, P. (2017). Detecting cardiovascular disease from mammograms with deep learning. *IEEE Transactions on Medical Imaging*, 36(5), 1172–1181.
- Wang, K., Delbaere, K., Brodie, M. A. D., Lovell, N. H., Kark, L., Lord, S. R., ... Member, S. (2017). Differences between Gait on Stairs and Flat Surfaces in Relation to Fall Risk and Future Falls. *IEEE Journal of Biomedical and Health Informatics*, 21(6), 1479–1486. <http://doi.org/10.1109/JBHI.2017.2677901>
- Watanabe, A., Noguchi, H., Oe, M., Sanada, H., & Mori, T. (2017). Development of a Plantar Load Estimation Algorithm for Evaluation of Forefoot Load of Diabetic Patients

- during Daily Walks Using a Foot Motion Sensor. *Journal of Diabetes Research*, 2017. <http://doi.org/https://doi.org/10.1155/2017/5350616>
- Weinstein, M., & Booth, J. (2006). Preventing Falls in Older Adults: A Multifactorial Approach. *Home Health Care Management & Practice*, 19(1), 45–50. <http://doi.org/10.1177/1084822306292232>
- Weiss, A., Herman, T., Giladi, N., & Hausdorff, J. M. (2014). Objective Assessment of Fall Risk in Parkinson's Disease Using a Body-Fixed Sensor Worn for 3 days. *PloS One*, 9(5), e96675. <http://doi.org/10.1371/journal.pone.0096675>
- Wimmer, G., Hegenbart, S., Vécsei, A., & Uhl, A. (2017). Convolutional Neural Network Architectures for the Automated Diagnosis of Celiac Disease. In *International Workshop on Computer-Assisted and Robotic Endoscopy* (pp. 104–113).
- World Health Organization. (2016). World Health Statistics 2016: Monitoring Health for the SDGs Annex B: Tables of Health Statistics by Country, WHO Region and Globally. Retrieved January 23, 2018, from http://www.who.int/gho/publications/world_health_statistics/2016/Annex_B/en/
- World Health Organization. (2016). WHO | Falls. Retrieved January 23, 2018, from https://www.who.int/violence_injury_prevention/other_injury/falls/en/
- World Health Organization. (2008). *WHO Global Report on Falls Prevention in Older Age*.
- Wu, J., & Jafari, R. (2014). Zero-Effort Camera-Assisted Calibration Techniques for Wearable Motion Sensors. In *Proceedings of the Wireless Health 2014 on National Institutes of Health* (pp. 1–8).
- Wu, X., Yeoh, H. T., & Lockhart, T. (2013). Fall Risks Assessment and Fall Prediction among Community Dwelling Elderly Using Wearable Wireless Sensors. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 109–113).
- Xiang, S., Yuan, L., Fan, W., Wang, Y., Thompson, P. M., & Ye, J. (2013). Multi-source Learning with Block-wise Missing Data for Alzheimer's Disease Prediction. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '13* (pp. 185–193). <http://doi.org/10.1145/2487575.2487594>
- Xue, Y., Liang, H., & Boulton, W. R. (2008). Information Technology Governance in Information Technology Investment Decision Processes: The Impact of Investment Characteristics, External Environment, and Internal Context. *MIS Quarterly*, 32(1), 67–96.
- Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., & Krishnaswamy, S. (2015). Deep Convolutional Neural Networks On Multichannel Time Series For Human Activity

- Recognition. In *Proceedings of the 2015 International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 3995–4001). Buenos Aires, Argentina.
- Yang, J. (2009). Toward physical activity diary: motion recognition using simple acceleration features with mobile phones. In *Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics* (pp. 1–10). ACM.
- Yang, T.-Y., Brinton, C. G., Joe-Wong, C., & Chiang, M. (2017). Behavior-based grade prediction for MOOCs via time series neural networks. *IEEE Journal of Selected Topics in Signal Processing*, 11(5), 716–728.
- Yao, S., Hu, S., Zhao, Y., Zhang, A., & Abdelzaher, T. (2017). Deepsense: A unified deep learning framework for time-series mobile sensing data processing. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 351–360).
- Yaraghi, N., Du, A. Y., Sharman, R., Gopal, R. D., & Ramesh, R. (2014). Health Information Exchange as a Multisided Platform: Adoption, Usage, and Practice Involvement in Service Co-Production. *Information Systems Research*, 26(1), 1–18.
- Ye, H. (Jonathan), & Kankanhalli, A. (2018). User Service Innovation on Mobile Phone Platforms: Investigating Impacts of Lead Userness, Toolkit Support, and Design Autonomy. *MIS Quarterly*, 42(1), 165–187. <http://doi.org/10.25300/MISQ/2018/12361>
- Yu, J., Zhang, B., Kuang, Z., Lin, D., & Fan, J. (2017). iPrivacy: Image Privacy Protection by Identifying Sensitive Objects via Deep Multi-Task Learning. *IEEE Transactions on Information Forensics and Security*, 12(5), 1005–1016.
- Yu, S., Chen, H., & Brown, R. A. (2018). Hidden Markov model-based fall detection with motion sensor orientation calibration: A case for real-life home monitoring. *IEEE Journal of Biomedical and Health Informatics*, 22(6), 1847–1853.
- Yuan, L., Wang, Y., Thompson, P. M., Narayan, V. A., & Ye, J. (2012). Multi-source Feature Learning for Joint Analysis of Incomplete Multiple Heterogeneous Neuroimaging Data. *NeuroImage*, 61(3), 622–632. <http://doi.org/10.1016/j.neuroimage.2012.03.059>
- Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P., & Zhang, J. (2014). Convolutional Neural Networks for Human Activity Recognition Using Mobile Sensors. In *Proceedings of the 6th International Conference on Mobile Computing, Applications and Services (MobiCASE)* (pp. 197–205).
- Zhang, D., Shen, D., & The Alzheimer's Disease Neuroimaging Initiative. (2012). Multi-modal Multi-task Learning for Joint Prediction of Multiple Regression and Classification Variables in Alzheimer's Disease. *NeuroImage*, 59(2), 895–907. <http://doi.org/10.1016/j.neuroimage.2011.09.069>

Zhang, W., Zhang, Y., Zhai, J., Zhao, D., Xu, L., Zhou, J., ... Yang, S. (2018). Multi-source Data Fusion Using Deep Learning for Smart Refrigerators. *Computers in Industry*, 95, 15–21. <http://doi.org/10.1016/j.compind.2017.09.001>

Zhong, Y., & Deng, Y. (2014). Sensor Orientation Invariant Mobile Gait Biometrics. In *Proceedings of the 2014 IEEE International Joint Conference on Biometrics (IJCB)* (pp. 1–8).