PERSONALIZED FUNCTIONAL HEALTH AND FALL RISK PREDICTION USING ELECTRONIC HEALTH RECORDS AND IN-HOME SENSOR DATA

A Dissertation

presented to

the Faculty of the Graduate School

at the University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

ANUP KUMAR MISHRA

Dr. Marjorie Skubic, Dissertation Supervisor

MAY 2021

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

PERSONALIZED FUNCTIONAL HEALTH AND FALL RISK PREDICTION USING ELECTRONIC HEALTH RECORDS AND IN-HOME SENSOR DATA

presented by Anup Kumar Mishra,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Marjorie Skubic

Curators Professor James Keller

Professor Mihail Popescu

Professor Laurel A. Despins

To my mom Sunanda Misra, my dad Dr. Siva Prasad Mishra, my wife Shradha Shalini,

and all my family and friends.

ACKNOWLEDGMENTS

I would like to gratefully thank Dr. Marjorie Skubic, my adviser, for her guidance, invaluably constructive criticism, and friendly advice during this research work. I am sincerely grateful to her for sharing her truthful and illuminating views on a number of issues related to this work.

I would like to sincerely thank the members of my dissertation committee, Dr. James Keller, Dr. Mihail Popescu, and Dr. Laurel A. Despins for their input and valuable suggestions to further improve my research work.

I would like to thank my wife Shradha Shalini for always being my strength. I also owe the deepest gratitude to my parents and family members. Without their motivation, faith, and continuous support this would not have been possible.

I am very grateful to my friends, especially, Dr. Enayati Moein, Dr. Andrew Buck, and Dr. Chetkar Jha for their insights, support, and help during my studies. I am also thankful to my lab members for their support in the Center for Eldercare and Rehabilitation Technology lab, especially, Brad Harris, Dr. Akshay Jain, Brianna Markway, and Noah Marchal.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure Page

LIST OF TABLES

Table Page

ABSTRACT

Research has shown the importance of Electronic Health Records (EHR) and in-home sensor data for continuous health tracking and health risk predictions. With the increased computational capabilities and advances in machine learning techniques, we have new opportunities to use multi-modal health big data to develop accurate health tracking models. This dissertation describes the development, evaluation, and testing of systems for predicting functional health and fall risks in community-dwelling older adults using health data and machine learning techniques.

In an initial study, we focused on organizing and de-identifying EHR data for analysis using HIPAA regulations. The dataset contained nine years of structured and unstructured EHR data obtained from TigerPlace, a senior living facility at Columbia, MO. The de-identification of this data was done using custom automated algorithms. The de-identified EHR data was used in several studies described in this dissertation. We then developed personalized functional health tracking models using geriatric assessments in the EHR data. Studies show that higher levels of functional health in older adults lead to a higher quality of life and improves the ability to age-in-place. Even though several geriatric assessments capture several aspects of functional health, there is limited research in longitudinally tracking the personalized functional health of older adults using a combination of these assessments. In this study, data from 150 older adult residents were used to develop a composite functional health prediction model using Activities of Daily Living (ADL), Instrumental Activities of Daily Living (IADL), Mini-Mental State Examination (MMSE), Geriatric Depression Scale (GDS), and Short Form 12 (SF12). Tracking functional health objectively could help clinicians to make decisions for

interventions in case of functional health deterioration. We next constructed models for fall risk prediction in older adults using geriatric assessments, demographic data, and GAITRite assessment data. A 6-month fall risk prediction model was developed with data from 93 older adult residents. Explainable AI techniques were used to provide explanations to the model predictions, such as which specific features increased the risk of fall in a particular model prediction. Such explanations to model predictions provide valuable insights for targeted interventions. In another study, we developed deep neural network models to predict fall risk from de-identified nursing notes data from 162 older adult residents from TigerPlace. Clinical nursing notes have been shown to contain valuable information related to fall risk factors. This analysis provides the groundwork for future experiments to predict fall risk in older adults using clinical notes.

In addition to using EHR data to predict functional health and fall risk in older adults, two studies were conducted to predict fall and functional health from in-home sensor data. Models for in-home fall prediction using depth sensor imagery have been successfully used at TigerPlace. However, the model is prone to false fall alarms in several scenarios, such as pillows thrown on the floor and pets jumping from couches. A secondary fall analysis was performed by analyzing fall alert videos to further identify and remove false alarms. In the final study, we used in-home sensor data streaming from depth sensors and bed sensors to predict functional health and absolute geriatric assessment values. These prediction models can be used to predict the functional health of residents in absence of sparse and infrequent geriatric assessments. This can also provide continuous tracking of functional health in older adults using the streaming in-home sensor data.

Chapter 1: Introduction

1.1 Motivation

The number of Americans ages 65 and older is projected to be over 95 million by 2060, sharing about 23 percent of the total population in the USA [1]. Similar demographic shifts are predicted for many other countries, including Japan, South Korea, Hong Kong, Italy, Germany, and China. In addition, studies estimate that in the year 2014, 44.7 million older adults lived independently in housing units, such as homes and apartments (96.8 percent), compared to only 1.5 million in group quarters, and only 1.2 million in nursing homes in the USA [2]. Therefore, creating an environment that can improve quality of life and support independence in the growing population of older adults is currently an important quest in healthcare research.

Creating an environment to support and care for older adults could be done in a variety of ways. Smart home systems such as smart medication reminder systems and in-home sensor systems for continuous health monitoring, advanced care coordination, and clinical interventions can all help in improving the quality of life and independence in older adults. A combination of in-home sensor systems and advanced nursing care coordination has been shown as an effective combination to support aging in place [3, 4]. Senior living facilities like TigerPlace at Columbia, Missouri provides advanced nursing care along with an integrated in-home sensor system that can track gait, predict fall, estimate restlessness, and sleep parameters of older adults [5, 6]. Custom data processing algorithms have been developed to track and recognize patterns in sensor data [7]. In the case of anomalies or

adverse health conditions detected in sensor data patterns, these algorithms generate health alerts and send them to the designated care providers. Health alert generating systems have a critical part of healthcare systems to inform care providers about the critical health conditions of the resident and patients. For example, automated fall detection and alarm system detects falls and sends alerts to the care providers or family members for their immediate attention [8]. Many generated health alerts are usually irrelevant, redundant, or simply false. These false alerts overwhelm the schedule of care providers. This phenomenon is commonly known as alarm fatigue [9]. Sensor-based alerts suggest a significant enough change in sensor data was found, however, they do not necessarily provide sufficient information about the change in the underlying health condition of an older adult. Therefore, understanding the underlying health conditions could help improve the health monitoring systems.

Electronic Health Records (EHR) and in-home sensor data have been shown to contain critical health information including functional health and fall risk factors in older adults [10, 11]. At TigerPlace, care is provided by registered nurse care coordination services and EHR data has been maintained by the nursing staff [12]. However, analysis of this EHR dataset has not been conducted to track or predict the health of residents. This provides a unique opportunity to utilize the TigerPlace EHR data to understand longitudinal underlying health changes of the older adult residents. In this dissertation, we analyzed the TigerPlace EHR data to track personalized functional health in older adults. The current fall risk model used at TigerPlace is based on in-home gait speed only [13]. This provided an opportunity to explore new methods to predict fall risk using the unique

data at TigerPlace. Therefore, in addition to functional health tracking, we have also utilized the EHR data to predict fall risk.

We first de-identified and developed an EHR dataset for predictive analytics combining data from two different EHR databases used at TigerPlace between 2010 and 2019. We developed several health prediction and tracking models using the de-identified EHR data, specifically (1) functional health status prediction using geriatric assessments, (2) 6-month fall risk prediction using geriatric assessments and GAITRite data, (3) fall risks estimation using clinical notes and medications, and (4) functional health status prediction using in-home sensor data.

Functional health is multifactorial comprising of several patterns [14]. We used geriatric assessments in the EHR of 168 residents living at TigerPlace between 2011 – 2019 to develop a composite functional health index. We extended the Rothman Index model developed for critical inpatient care to construct the functional health index using excess risk functions for the geriatric assessments [15]. Additionally, we developed mixed-effect logistic regression models to estimate functional health using the repeated geriatric assessments from the residents. We also developed 6-month fall risk models to accurately predict fall risk from the geriatric assessments and gait parameters from GAITRite data. Furthermore, we used clinical notes and medication information in the EHR to predict fall and hospitalization risks using word embedding models and Long Short-Term Memory neural networks. Finally, we used in-home sensor data including depth sensor and bed-sensor data to predict the functional health of the residents. We used different combinations of sensor data to predict the functional health estimates obtained from the mixed-effect model and the absolute geriatric assessments including ADL, IADL, MMSE, and GDS.

In addition to predicting fall risk, we also improved the fall prediction algorithm by reducing irrelevant or false fall alarms generated because of known issues in processing sensor data. Real-time fall detection system using depth sensors has provided promising results [16]. TigerPlace senior living facility has been using the fall detection system developed by Stone et al. and send real-time alerts to the clinical staff [16]. The algorithm, however, does not perform human detection to confirm that a persona fell in the detected fall video. Hence, along with successfully detecting several hundreds of true falls, the model also detected thousands of non-fall videos as falls. Some examples of these non-falls include pets jumping from couches, linens thrown on the floor, and kids jumping on the floor [17]. We propose a new secondary analysis of the detected fall alert videos to prune out the false alarms. We developed and labeled an alert video dataset containing ~4000 alert videos. We used deep neural networks, specifically convolutional neural networks (CNNs) based architectures for secondary processing of the alert videos to reduce false alarms and alarm fatigue.

1.2 Primary Goals

The primary objectives of this dissertation include:

- Development of an EHR dataset
    - De-identification of the EHR databases used at TigerPlace senior living facility from 2010 to 2018
    - Re-compilation of geriatric assessments data in the de-identified database
    - Organization of EHR data for predictive analysis
- Development of health risk prediction model using geriatric assessments in the electronic health record (EHR) data

    - Development of a Functional Health Index using geriatric assessments to track longitudinal functional health changes in older adults using excess-risk functions
    - Development of mixed effect models using repeated assessment measures to predict functional health
- 6-month fall risk prediction using geriatric assessments, weight change, and GAITRite-assessment data
    - Development of fall risk prediction models using machine learning algorithms from spatiotemporal gait parameters, and assessment data
    - Use of model explainability methods to explain fall risk predictions for personalized interventions
- Retrospective analysis of fall alert videos to prune out false alarms
    - Development and labeling of a fall alert dataset
    - Classification of fall alert videos to identify and prune out false alarms using CNN based deep neural architectures
- Fall and hospitalization predictions using clinical notes and medications
    - Development of deep neural architectures and word embedding models to predict fall and hospitalization risk from clinical notes and medications data
- Functional health predictions using in-home sensor data
    - Development of regression models to predict functional health from in-home sensor data obtained from depth sensors and bed sensors

1.3 Publications

1.3.1 Included in Thesis

- Mishra AK, Buck A, Marchal N, Skubic M (In Preparation). “Multi-modal Health and Sensor Data Storage and Visualization Framework.” GigaScience.
- Mishra AK, Chappell M, Emerson S, Skubic M, Popescu M, Keller J, Rantz M, Miller S (In Preparation). “Fall Risk Prediction in Older Adults using EHR Nursing Notes.” Journal of the American Medical Informatics Association
- Mishra AK, Skubic M, Popescu M, Keller J, Rantz M, Markway B, Tripathi P, Marchal N (In Preparation). “False Fall Alert Reduction Using Deep Neural Networks.” IEEE Journal of Biomedical and Health Informatics
- Mishra AK, Skubic M, Popescu M, Keller J, Rantz M, Abbott C, Enayati M, Miller S (In Preparation). “6-Month Fall Risk Prediction using Geriatric Assessments and GAITRite Data.”
- Mishra AK, Skubic M, Popescu M, Lane K, Rantz M, Despins LA, Abbott C, Keller J, Miller S. “Tracking Personalized Functional Health in Older Adults Using Geriatric Assessments.” BMC Medical Informatics Decision Making 20, 270 (2020). doi: 10.1186/s12911-020-01283-y
- Mishra AK, Skubic M, Popescu M, Keller J, Buck A, Despins LA, Rantz M, Lane K, and Miller S. “Development of a Fuzzy-Functional Health Index for Older Adults using Geriatric Assessments.” In Fuzz-IEEE: International Conference on Fuzzy Systems Late-Breaking Research, New Orleans, USA, June 23-26, 2019
- Mishra AK, Skubic M, Despins LA, Popescu M, Rantz M, Keller J, and Lane K. "Development of a Functional Health Index for Older Adults using the Electronic Health Record." In IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Chicago, IL, USA, May 19-22, 2019

- Mishra AK, Skubic M, Despins LA, Popescu M, Rantz M, Keller J, and Lane K. "Development of a Functional Health Index for Older Adults using the Electronic Health Record." In 29th Annual Caring for the Frail Elderly Conference (CFE), Columbia, MO, USA, August 16-17, 2019
- Mishra AK, Skubic M, Keller J, Popescu M. "Two-Step Evaluation Methodology to Reduce False Alarms in Healthcare Systems." In the US National Science Foundation (NSF) Industry/University Cooperative Research Centers (I/UCRC) program, Baylor College of Medicine, Houston, TX, November 2018

1.3.2 Other

- Calyam P, Mishra AK, Antequera RB, Chemodanov D, Berryman A, Zhu K, Abbott C, Skubic M, "Synchronous big data analytics for personalized and remote physical therapy," Pervasive and Mobile Computing, 2016 Jun 1; 28:3-20. doi: 10.1016/j.pmcj.2015.09.004.
- Calyam P, Jahnke I, Mishra AK, Antequera RB, Chemodanov D, and Skubic M. "Toward an ElderCare Living Lab for Sensor-Based Health Assessment and Physical Therapy." IEEE Cloud Computing 4, no. 3 (2017): 30-39. doi: 10.1109/MCC.2017.46.
- Mishra AK, Skubic M, Willis BW, Guess T, Gray AD, and Sherman SL. "A Novel Depth Image Analysis Method to Calculate the Anterior Reach of the Modified Star Excursion Balance Test." In Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '18). ACM, New York, NY, USA, 2018 (25% acceptance rate). doi: 10.1145/3240925.3240971.
- Mishra AK, Skubic M, Willis BW, Guess T, Razu SS, Abbott C, and Gray AD. "Examining methods to estimate static body sway from the Kinect V2.0 skeletal data: implications for clinical rehabilitation." In Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth '17). ACM, Barcelona, Spain, 2017 (25% acceptance rate). doi:10.1145/3154862.3154874.

- Skubic M, Mishra AK, Harris B, Abbott C, Craver A, Musterman K, and Rantz M. “HCI Challenges for Consumer-Based Aging in Place Technologies.” In International Conference on Human Aspects of IT for the Aged Population (pp. 105-116). Springer, Cham, July 2016. doi:10.1007/978-3-319-39943-0_11.
- Hotrabhavananda B, Mishra AK, Skubic M, Hotrabhavananda N, and Carmen A. "Evaluation of the Microsoft Kinect skeletal versus depth data analysis for timed-up and go and figure of 8 walk tests." In Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the, pp. 2274-2277. IEEE, 2016. doi: 10.1109/EMBC.2016.7591183
- Antequera RB, Calyam P, Chemodanov D, Donato W, Mishra AK, Pescapè A, and Skubic M. "Socio-technical approach to engineer gigabit app performance for physical therapy-as-a-service." In IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), pp. 1-7. IEEE, 2017. doi: 10.1109/HealthCom.2017.8210768

1.4 Organization

Chapter 2 provides an overview of the background associated with this research. It primarily focuses on health monitoring systems, alarm fatigue in health systems, EHR data and nursing assessments, health indices and trajectories in estimating health, and video classification methods using deep neural network models.

Chapter 3 provides an overview of the de-identification and organization of the EHR data.

Chapter 4 describes the development and validation of a functional health status estimation model using geriatric assessment data in the EHR. Also, Chapter 4 includes the

development and validation of mixed-effect logistic regression models to predict health status using the repeated assessment data.

Chapter 5 provides an overview of the development of 6-month fall prediction models using geriatric assessments, change in weight, and GAITRite data.

Chapter 6 provides an overview of the development of a fall alert dataset. Also, it provides an overview of retrospective fall alert video classification using deep neural networks.

Chapter 7 provides an overview of fall prediction strategies using nursing notes and other text-based data in EHR.

Chapter 8 provides an overview of functional health predictions using in-home sensor data obtained from depth sensors and bed sensors.

Chapter 9 provides a brief description of future work and a detailed progress summary of the dissertation research.

Chapter 2: Background

2.1 Health Monitoring Systems

Health monitoring systems continuously monitor the health of an individual as an alternative to traditional healthcare management of patients to reduce healthcare and hospitalization costs, provide disease prevention, detect the change in health status, and provide data for early interventions [18]. Sensor-based health monitoring systems are pervasive and are used in several healthcare facilities, including Intensive Care Units (ICUs), home care, and aging-in-place facilities [3]. Many sensors are used in ICUs, smartphones, and smartwatches for continuous health monitoring [19]. However, studies have shown that older adults prefer non-wearable sensors for health tracking [20]. Aging in place facilities like TigerPlace are equipped with several non-invasive in-home sensors, such as bed sensors and depth sensors to continuously monitor the health of the older residents.

The health monitoring systems have several algorithms to process the sensor data and detect in case of any anomalies. In case of an important health anomaly detection, alerts are generated to the clinicians and other designated care providers. Depending on the number of residents or patients under observation, and the sensitivity of the alert system in a health monitoring system, the number of alerts can significantly vary. If the number of alerts is too high such that it is overwhelming the schedule of the care providers, then the alerts can cause major distractions in the health system and patient safety issues [21].

2.1.1 Alarm Fatigue

Studies show that 72% to 99% of clinical alarms are false [9]. These false health alerts in health monitoring systems overwhelm the schedule of care providers causing the care providers to not take all alerts seriously, thinking they may not be significant. This phenomenon is called alarm fatigue. Alarm fatigue leads to desensitization of alarms which leads to missed true alarms and patient deaths [9, 21, 22]. However, studies show that customization of alarm parameters can reduce the number of false alarms and increase patient safety [9].

The ECRI group has reported that alarm and alert-related issues have been one of the most pervasive medical device hazards reported from 2011-2020 [23]. Funk et al. surveyed to understand the attitude and practice of respondents related to clinical alarms [24]. The study results show that frequent false alarms are the most important alarm-related issue. The survey respondents also reported that patients are experiencing adverse events related to alarms at their institutes. Cvach et al. have conducted an integration review on alarm fatigue that analyzed over seventy-two articles between 1/1/2000 and 10/1/2011 using the John Hopkins Nursing Evidence-Based Practice model [25]. The authors mentioned that alarm fatigue is a national problem and had been the number one medical device hazard in 2012. Cvach et al. also mentioned that several methods including signal filtering, algorithms, and/or artificial intelligence systems are some of the methods researchers have successfully tried to reduce the number of false alarms. In a research study, Graham et al. reported that alarm fatigue could be improved by training nurses to individualize the patients' alarm parameters [26]. Based on the studies by Cvach and Graham et al. personalization of alarm parameters could improve alarm fatigue incidents.

Also, research shows that there is an issue of over-monitoring in health systems, not just in ICU-care but also in non-ICU hospital settings [27]. Therefore, changing sensitivity in health tracking based on patient health status could also play a significant role in improving alarm fatigue.

2.1.2 False and/or Irrelevant Alarms in Eldercare Systems

Alarm fatigue has been reported in eldercare systems due to a significant number of false and irrelevant alerts. Skubic et al. reported that the depth camera-based fall detection system produced a large number of false fall alerts while detecting in-home falls [17]. The authors discussed two longitudinal studies testing fall detection sensor technology for homes of older adults. In the latest fall detection study, 570 fall alerts were generated from 67 apartments over 7 months period, out of which only 67 were actual falls. The source of the false alarms were linens thrown on the floor, pets jumping, visitors, etc. A secondary analysis of these fall alerts could potentially eliminate some of these false alarms.

2.2 Electronic Health Records Data

To reduce false alerts, the alarm system should be personalized based on the health status of the residents or patients. A healthier person with higher functional health should be treated differently as compared to a person with lesser functional health. The functional and overall health status of an individual can be obtained using Electronic Health Records (EHR) [28].

EHR are developed as the digital replacement for the traditional paper-based medical records. The primary benefit of having a digital version of the health records is the ease of storing and retrieving the entire medical history of a patient as necessary. EHRs are

complex and contain both structured and unstructured health data. The structured data in the EHR include physiological measures (vital signs), lab results, health assessments (Activities of Daily Living), medications, and diagnoses. The unstructured data in the EHR include clinical free text in the progress notes and visit notes.

2.2.1 EHR at TigerPlace

In this dissertation, we focus on analyzing the EHR data from a senior living facility, which is different than an EHR in a medical system or a critical care system. An EHR system in the senior living facility contains longitudinal health data compared to the more sparse health data in a hospital EHR with hospitalization-specific data. The longitudinal data in a senior care facility could provide us more insights on the change in health condition over a period that led to hospitalization, fall, or even death. The EHR data at TigerPlace contained geriatric assessments, adverse health events such as hospitalization and fall event information, vital signs, medications, progress notes, medical diagnoses, and lab results.

2.3 Geriatric Assessments, Functional Health, Health Indices, and Health Trajectories for Older Adults

Functional health in older adults is complex and multifactorial [29, 30]. Several geriatric assessments have been developed and validated for measuring different aspects of functional health in older adults, such as Activities of Daily Living (ADL), Instrumental Activities of Daily Living (IADL), Mini-Mental State Examination (MMSE), Geriatric Depression Scale (GDS), and Short Form 12 (SF12) [31-40]. ADLs are defined as activities that are essential for independent living. IADL require a higher level of personal autonomy, referring to tasks that require enough capacity to make decisions through greater interaction

with the environment [38]. MMSE is a widely used test to evaluate the cognitive aspects of mental function. MMSE excludes questions concerning abnormal mental experiences and mood [36]. GDS is a screening tool for measuring depression in older adults [39]. SF-12 is a multipurpose short form that provides a generic measure of health status [40]. All these assessments have good reliability and validity. These individual assessments can track different aspects of functional health in older adults. However, understanding the overall change in the functional health of an older adult using these individual assessments can be difficult, as they are based on different scales and have a wide range of clinical condition focus and importance to overall health. Instead of considering them individually, a combination of these assessments can be used to develop a continuous composite functional health score to predict a more general status of functional health in older adults. Several studies have been conducted in developing health and prognostic indexes to track or predict comorbidity, mortality, frailty, and functional health in older adults [41-46]. Mazzaglia et al. developed two prognostic index models to predict five-month mortality and hospitalization [41]. In the first model, they used a set of 7 questions from ADL and IADL to develop their index. The area under the receiver operating characteristic curves (AUC) to predict mortality and hospitalizations were 0.75 and 0.60, respectively. In the second model, they considered drug use and previous hospitalizations, which increased their hospitalization AUC to 0.67. Gagne et al. developed a single numeric index to predict mortality by combining Charlson and Elixhauser measures [42]. Results show that the combined score performed better in predicting mortality than the individual scores. Carey et al. and Lee et al. developed prognostic models to predict mortality using data from the Program of All-Inclusive Care for the Elderly (PACE) and 1998 wave of the Health and

Retirement Study (HRS), respectively [43, 44]. Schonberg et al. used 39 risk factors, including functional measures, illnesses, behaviors, demographics in a multivariable Cox proportional hazards model to predict 5-year mortality [45]. Giovanni et al. developed a multisource comorbidity score using administrative data, such as diagnostic categories and ICD-9 to measure comorbidity, predict 1-year mortality, and other adverse outcomes [46]. This study did not include functional status as a variable in the predictive model development.

Fried. et al. conducted a study to predict frailty in older adults [47]. They defined frailty as a clinical syndrome in which the older adult has three or more out of five frailty criteria. These five criteria include unintentional weight loss, self-reported exhaustion, weakness, slow walking speed, and low physical activity. This standardized phenotype of frailty detection can identify frail older adults potentially at risk of falls, hospitalizations, disability, and death.

Gait analysis provides critical information in predicting adverse health conditions and functional health in older adults [13, 48]. Nelson et al. developed a Functional Ambulation Performance (FAP) score that captures the gait capacity of an individual using a specific set of spatiotemporal parameters (STPS) [49, 50]. The FAP score is integrated with the GAITRite walkway (CIR System Inc; Clifton, New Jersey), a gold standard system for measuring STPS [51]. FAP has been validated in several independent studies [50].

The mortality-based prognostic models tend to predict future adverse health conditions, specifically death, instead of predicting the overall functional health of an individual at a given time. The frailty phenotyping method developed by Fried et al. can

predict the presence of frailty with minimal granularity as it can only classify an older adult into one of the three frail categories: frail, intermediate frail, and not frail [47]. This may not be able to track the gradual changes in the functional health of an individual. Also, this only considers the physical aspects of functional health; cognitive aspects were excluded [52, 53]. The FAP score is also developed considering only gait ability and does not consider the cognitive aspects of functional health [49]. We argue that a continuous measure of overall functional health can provide critical information about changes in functional health over a period. Interventions based on overall functional health deterioration can help older adults live with higher independence and quality of life [54]. Rothman et al. have developed a continuous measure of patient condition in acute care, called the Rothman Index, using a set of 26 Electronic Medical Record variables [15]. They used one-year post-discharge mortality to develop excess risk functions for each variable. The excess risk functions were then used to calculate the Rothman Index for an individual at a given period. Instead of predicting an adverse health condition, the Rothman Index provides an overall patient condition in inpatient care.

2.3.1 Predicting Functional Health from In-Home Sensor Data

As described in the previous section, several assessments and models have been created to estimate functional health in older adults. However, these functional health assessments dependent on human intervention. Trained nurses and health practitioners conduct these assessments at regular intervals. The two significant issues with the manual assessment of functional health are the subjectivity involved in conducting the assessment and the fact that the assessments are not conducted more frequently to track the change in

health. Therefore, estimating functional health accurately using unobtrusive in-home sensor data could solve these issues.

Several studies have been conducted to predict functional health using technology-enabled platforms and in-home sensors. Cook et al. in a systematic review discussed the use of technology-enabled platforms to predict functional health [55]. In this article, the authors discussed technology-enabled and sensor-based functional health assessments, such as using wearable sensors for physiological, motor function, and mood monitoring; ambient sensors for dementia prediction, analysis of daily activities, and specific state detection including gait, balance, mobility, and mood [56-60]. The in-depth analysis provides evidence that computing, wireless sensors, and machine learning technology have the potential to revolutionize functional health assessment.

Dawadi et al. have presented an automated cognitive assessment method using in-home motion, light, door, and temperature sensor data [61]. In the study, sensor data from 18 community-dwelling seniors from a retirement community with age 73 or older were collected along with biannually administered clinical, cognitive, and motor tests using timed up and go (TUG) and repeatable battery for the assessment of neuropsychological status measure of cognitive status (RBANS). Results show that the automatically labeled activities including ADL, sleep show a statistically significant correlation with RBANS. Also, ADL and sleep combined with other features such as mobility, and out-of-home duration show a statistically significant correlation with TUG.

In another study, Aramendi et al. have presented an automatic assessment of functional health decline in seniors using smart home data [11]. They used a set of 10 day-level activity features to predict Instrumental Activity of Daily Living-Compensation

(IADL-C). The features include time spent per day in cooking, eating, and relaxing; sleep-related features including sleep duration, and frequency, and mobility-related features such as a total number of activated motion sensors, and total distance walked. The study included data obtained from 29 older adults for an average of two years. The authors experimented with different regression algorithms to predict IADL-C using the ten features. Results show that a statistically significant correlation was observed between IADL-C and the predicted IADL-C using different algorithms. However, the correlation values ranged from 0.01-0.29 suggesting a high correlation between the predicted and actual IADL-C was not found. The authors further experimented with the predicted IADL-C values and found that even though the predicted IADL-C was not highly correlated with the actual IADL-C but it was able to successfully suggest if there was a positive or negative change in the actual value.

2.4 Video Classification for Pruning False Fall Alerts

False fall alert reduction is critical in improving the overall alert system at senior living facilities. A longitudinal study conducted by Skubic et al. shows that many false alerts are generated in the depth image-based fall detection system are because of in-home events like linens thrown on the floor and pets jumping off the furniture [17]. A retrospective analysis of these alert videos can provide more context to identify and remove these false fall alerts. Alerts generated using the fall detection system are essentially email or text messages with a link to the fall video [8]. A spatiotemporal video classification could be used as a secondary alert video analysis step (an analysis after the fall detection system detects a fall and before an alert is generated) to predict a non-fall and stop it from becoming an alert.

Video and action classification are fundamental problems in computer vision. Several studies have been conducted to understand and classify videos using hand-crafted spatiotemporal features, including space-time interest points (STIP), SHIFT-3D, HOG3D, Cuboids, and iDT [62-67].

The availability of a large number of images and videos has encouraged the development of new computing techniques and algorithms to analyze these data for various applications. The recent developments in deep neural networks, specifically convolutional neural networks (ConvNets) for image processing have been very successful [68-74]. The ConvNet architectures have been used in several image processing tasks including, image recognition, retrieval, segmentation, and detection [71, 75-79]. Ziller et al. have shown that ConvNets learn powerful interpretable features [80]. Several new techniques such as Batch Normalization, Dropout, Parametrized-RELU, Spatial-Pyramid Pooling have been developed to further improve the classification accuracy [81-84].

Several deep architectures have been proposed to learn spatiotemporal features for video understanding. In [85], Ji et al. have proposed 3D ConvNets to recognize human actions. Restricted Boltzmann Machines and stacked ISA were also used with 3D ConvNets to learn spatiotemporal features [86, 87].

CNN architectures trained on the ImageNet dataset have been used for solving image problems in other domains as well. ImageNet trained network features have been used for the PASCAL VOC classification and detection challenge [78, 88]. Shaoqing et al. have shown that features obtained from a deeper trained network like VGG-16 can provide further improvements in the PASCAL VOC performance [89]. These pioneering papers inspired several experiments to use ImageNet trained deep architectures for transfer

learning or be used as is in completely new tasks, such as video and action classification, depth prediction, segmentation, and pose estimation.

The progress in developing new architectures for video classification and representation learning is slower [90-94]. Tran et al. explain the three important bottlenecks in the development of strong architectures for videos when compared to image-based models:

- *Video ConvNets need higher computation power and memory*

  In [94], Tran et al. have studied that it could take up to 3 to 4 days to train a 3D ConvNet on UCF101 and about two months on h-1M datasets [91, 95].

- *No standard benchmark for architecture search for videos*

  ImageNet is being used as the standard benchmark to test image classification accuracy [69, 70]. Deep architectures can be trained on this dataset at a reasonable expense of time and the trained networks generalize to other tasks, such as segmentation and object detection. The Sports-1M dataset is shown to help learn generic features when trained with a needed architecture [94]. However, the dataset is too large, and it is expensive to perform an architecture search using the Sports-1M dataset. Video data frames in UCF101 are highly correlated. Studies show that models trained on UCF101 suffer from overfitting issues [91, 94]. Tran et al. have reported that training ConvNet models from scratch using UCF101 accuracy of only 41-44% could be obtained. However, finetuning using the Sports-1M dataset, the authors could improve the accuracy to 82% on UCF101 [94].

- <u>Designing a video classification model is non-trivial</u>

The performance of a video classification deep architecture is dependent on many choices, including the pre-processing process, type of convolutions, number of layers, and modeling temporal dimensions. Tran et al. performed an architecture search on UCF101 to solve this issue [90].

Carreira et al. developed a new architecture, Two-Stream Inflated 3D ConvNets, otherwise known as I3D to learn spatiotemporal features [96]. The I3D is based on the idea that 3D ConvNets can benefit from ImageNet 2D ConvNet designs. The architecture inflates 2D ConvNets into 3D. The filters and pooling kernels are inflated endowing them with an additional temporal dimension in a successful 2D classification model to create a 3D ConvNet classification architecture. Square filters are converted into cubic using this design strategy. Results show that this model outperforms the state-of-the-art video classification models.

In another study, Tran et al. have explored and compared convolutional residual block architectures for video classification [97]. The architectures they have included are, 2D convolutions over the entire clip [92], 2D convolutions over frames, 3D convolutions [85, 94, 96], mixed 3D-2D convolutions, and R(2+1)D convolutions. The new spatiotemporal convolutional block R(2+1)D designed by the authors in this study approximates the 3D convolutions. It consists of a 2D convolution followed by a 1D convolution, decomposing spatial and temporal modeling into two separate steps. Results show that the R(2+1)D performs similar or superior to the state-of-the-art architectures like I3D.

Therefore, there is enough evidence in the literature that convolutional neural networks can be used to classify videos effectively. A successful video classification model

will help reduce the number of false fall alarms raised through the depth data-based fall alarm detection model [8]. The depth fall alerts send a message with an RGB video of the fall. The video is segmented [98]. The segmentations generally include are the ground plane and the foreground objects. The foreground objects usually contain the fall. False alarms occur when a foreground is an object other than a human falling on the floor, such as a pillow thrown on the ground, jumping off the furniture, etc. Modeling the spatiotemporal features in these videos and classifying the alert videos can help reduce false alarms.

2.5 Computing with Words to Predict Health Risk

Bjarnadottir et al. have found that nursing notes can contain clinical, organizational, and environmental fall risk indicators that are not explicitly recorded by the providers or other commonly measured fall risk factors [10]. In this exploration study, the authors analyzed 1,046,053 registered nurses' notes (RNs' notes) from the MIMIC-III database. The study results show that the RNs' notes have contained explicit, intrinsic, and extrinsic factors related to the risk of fall. This study inspired us to explore the nursing notes in the EHR at Aging-in-Place facilities to predict fall risk.

Studies show that processing nursing notes to predict health outcomes has many challenges [99, 100]. Hyun et al. conducted a study exploring the ability of NLP to extract data from nursing notes [99]. The authors used the MedLEE library to encode the nursing notes. Hyun et al. have concluded that the RNs' free texts contained several key health information and they could be extracted using NLP techniques. However, the authors observed that MedLEE did not include many Nursing terminologies and abbreviations. MedLEE, like other Medical Language Extraction and/or Encoding systems heavily

depend on a specific lexicon that may not be easily extended to be used in new text domains [99]. Therefore, instead of depending on a predefined Medical Language Extraction System, adaptive and automatic feature engineering of RNs' free texts using advanced NLP models would probably be a step in the right direction. The recent developments in NLP research show promising results in predicting health outcomes from nursing notes [100-103].

More specifically, there is a precedent for predicting health outcomes using machine learning and electronic health records which includes a wide range of uses. A previous study involving the neural network system named DeepCare used machine learning to predict future medical events from electronic health data [104]. In their case studies, testing DeepCare with diabetes and mental health remission, they found that DeepCare performed competitively against current state-of-the-arts [104]. Another study involves predicting psychiatric readmission using statistical models with word patterns identified with natural language processing of the EHR data [105]. Research has also been done regarding the risk of harm a patient may inflict on themselves or others around them, also using natural language processing to analyze EHR data [106].

There have also been previous studies regarding predicting falls in a variety of settings. A study conducted on elderly patients visiting health centers in Maine found that deep learning models were able to successfully improve fall prediction accuracy [107]. In an in-patient setting, a second study used Japanese EMR data to sort patients in the hospital into risk and no-risk groups with relatively high accuracy [108]. However, this study additionally shows that temporal aspects of prediction can affect the success of deep learning models. Specifically, using more recent data provides for more accurate

predictions than long-term data [108]. Despite this, using long-term EMR data is still better for accuracy than excluding EMR data entirely [107, 108]. Predictions are also improved when the entirety of a patient's EHR records is used, including the clinical free-text nursing notes, which are not commonly used in machine learning models due to their unstructured nature [109]. Since clinical text data are often left out of predictive models, we have the opportunity to fill in this gap of research.

Previous research around using natural language processing to decipher EHR data has been around the topic of assigning medical codes to free text EHR data. One study used natural language processing to identify fall risk that needed to be coded within clinical notes to accurately reflect the fall risk in seniors that otherwise would not have been reported [110]. Another study more broadly aimed to identify concepts relating to ICD-9 codes using deep learning methods [111]. While our research does not involve medical codes, NLP identifying concepts related to fall risk suggests that we can identify fall risk for purposes unrelated to medical coding.

When approaching a natural language processing task with deep learning, there are several models and methods used. While convolutional neural networks (CNNs) are used more for text classification tasks, recurrent neural networks (RNNs) are typically better for tasks with a temporal aspect [112]. In a study that compared the performance of RNNs and CNN on several tasks (such as Sentiment Classification, Relations Classification, Textual Entailment, etc.), they found that typically RNN models are better at tasks that require understanding a sentence as a whole and CNN models are better at key-phrase detection [113].

To improve the accuracy of deep learning models, several studies use pre-trained word embeddings from large public datasets [112]. For tasks related to the biomedical field and clinical applications, word-embeddings from those specific areas out-perform other sources [103]. For our research, we started with using a clinical word embedding BioWordVec for this very reason. For clinical note processing specifically, the model that was trained using other notes performed better than the model trained on PubMed Central works [103].

Chapter 3: Eldertech EHR Dataset Development

Abstract

A primary task in developing an efficient alert system is to develop health status prediction models. EHR data contains continuous health data on individuals living in a facility. Therefore, an analysis of EHR data could provide detailed information on a change in the health of the residents to predict the continuous health status of the residents. The EHR data used in this dissertation was obtained from the TigerPlace facility, located at Columbia, Missouri USA [6]. The EHR contained Vital signs, medications, progress notes, medical diagnoses, and lab results. We pre-processed and de-identified the EHR data using HIPAA regulations. We also re-evaluated the geriatric assessment scores data using the answers to the assessment questions found in the de-identified EHR. The final dataset contained 201 subjects with health data collected from 2010 to 2019.

3.1 Background

EHR datasets contain valuable health information in both structured and unstructured formats. Several open-source benchmark EHR datasets have been developed to help the healthcare research community to analyze, conduct research, and solve critical problems concerning healthcare. The Medical Information Mart for Intensive Care (MIMIC – III) database contains deidentified health data associated with ~60,000 intensive care unit admissions from Beth Israel Deaconess Hospital [114]. MIMIC – III database contains sparse samples of time-series health observations recorded during inpatient ICU care, including medications, clinical notes, vital signs, and microbiology reports. MIMIC

– III data, however, does not contain non-emergency everyday health data that is generally collected through the EHR at the senior living facilities. The continuous health monitoring of health data could provide valuable information about what changes in health led to an adverse health event, such as a fall or hospitalization.

In addition to structured data in the EHR, clinical notes, such as progress notes, discharge summaries provide valuable information about patient health [115-117]. i2b2 dataset developed and maintained by Partners Healthcare contains several EHR datasets including free text-based Natural Language Processing (NLP) EHR datasets. However, this EHR data also does not contain the continuous health monitoring EHR data, specifically related to senior care.

The EHR dataset at TigerPlace has been continuously updated for about a decade from 2010 - 2019 for all residents living at the facility making it the ideal dataset for longitudinal data analysis to understand health changes in the residents over a longer period. Therefore, we created a comprehensive de-identified dataset using the TigerPlace EHR. The dataset developed was then used to perform health status estimation for the residents.

3.2 Methods

We followed HIPAA and Institutional Review Board regulations in de-identifying the EHR databases [118]. The de-identified dataset was then used for predictive analytics.

3.2.1 Electronic Health Records Data from TigerPlace

EHR dataset development was a crucial step for this dissertation work. The EHR data developed for the study was a combination of the CyberSense EHR and

PointClickCare EHR database used at TigerPlace [119, 120]. The CyberSense EHR was developed by CyberSense.US and was used as the primary health data record-keeping tool from 2010 to 2017. In 2017, the new PointClickCare EHR was introduced. The PointClickCare EHR is developed by PointClickCare Corp. and is the EHR database currently used at TigerPlace.

3.2.2 De-identifying and Pre-processing EHR Data

Each resident in the database had a unique identifier code in the Resident's table along with their name, date of birth, address, and other personal details. We observed that in several cases the resident's unique ID was composed of their last and first name instead of an all digit-based ID. The unique IDs are used in all tables to identify the resident. Therefore, to remove the trace of resident names through the IDs, we created a new set of IDs for those residents and replaced the older name-based IDs with the new digit-based IDs in all tables. We did not change the preexisting all-digit IDs that did not contain a name. We created a mapping table that contained the old IDs and their corresponding new IDs. We used this map to replace old IDs in all the other tables. The mapping ID table was kept encrypted and used only when another data table containing the older name-based IDs was supposed to be joined with the existing de-identified tables. Updating the name-based IDs was an essential step in de-identification as the older IDs with parts of the resident's first and last name could give away the identity of the residents. We removed all the personal information columns containing their contact and address details. Here is a list of the information we removed according to the de-identification standard of the HIPAA Privacy Rule with the exception of the residents' birth dates [121]:

- Name

- Social Security Numbers
- Admission and discharge dates
- Address including city, county, state
- Zip Code
- Email address
- Phone Number
- Medical Record Numbers
- Any account numbers
- Any images

The age of the residents was important for data analysis. Therefore, instead of deleting the date of births, we modified the date of births by setting the day in the dates to 1. For example, someone with a birthday of 05/14/1953 or 05/02/1953 was converted to 05/01/1953.

Resident's names were repeatedly used in the free text, unstructured data, specifically in the progress notes and visit notes. We identified if the residents had any other common names apart from the given first and last names. We used the resident's first name, last name, and any other common names to replace the names in the text data, in all the tables, in the entire EHR database by their unique IDs. For example, if a resident had their last name, first name as MISHRA, ANUP, here is how the before and after de-identification of text input in the EHR looks like assuming the unique ID for MISHRA, ANUP is 100000044:

Before de-identification:

"i went and checked on mr. mishra and he told me that he didn't have to use the restroom to maybe come back later after dinner."

After de-identification:

"i went and checked on mr. 100000044 and he told me that he didn't have to use the restroom to maybe come back later after dinner."

We used a set of string operations, including regular expressions to search the names in the text and replace them with unique IDs for the de-identification. Finally, we encrypted and removed the mapping table that contained the name-based IDs. The final de-identified dataset contained only digit-based IDs, without any personal identification information about the residents, including names.

3.2.3 Re-evaluating Geriatric Assessments

The de-identified EHR data contained several geriatric assessments. The geriatric assessments include ADL, IADL, MMSE, GDS, and SF-12. There were three different tables maintained for these assessments: questions and the answer options for all the assessments, clinical entries of answers periodically obtained from the residents for each of these assessments, and the scores obtained per individual assessments calculated by predefined algorithms. The nursing staff had concerns regarding the automatic evaluation of the assessments in the Cybersense EHR database. Therefore, all the answers obtained from the assessment were re-evaluated using custom algorithms to re-generate the assessment scores for future analysis.

The geriatric assessments were used in several analysis projects to estimate overall functional health, fall, hospitalization risk in older adults. Therefore, accurate evaluation of these assessments was crucial to the projects based on EHR data. Based on the literature, and clinical guidelines new custom algorithms were developed to re-evaluate the geriatric assessment scores. Specifically, we used the short form of ADL from RAI MDS 2.0, Lawton's IADL, GDS, MMSE, and SF12 using the approach described by Ware et al. to re-evaluate the assessments [36, 39, 53, 122, 123].

3.3 Assessment Data Statistics

Table 3.1 Assessment Statistics

	PCS12	MCS12	ADL	IADL	MMSE	GDS
Count	795	795	917	862	885	920
Mean	37.95	54.23	2.21	3.77	25.01	2.94
Std	11.78	9.10	3.17	1.56	6.63	2.52
Min	-2.00	1.91	0.0	0.00	0.00	0.00
25%	28.43	50.23	0.0	2.00	23.00	1.00
50%	37.11	56.61	1.0	4.00	28.00	2.00
75%	48.59	60.14	3.0	5.00	30.00	4.00
Max	63.43	73.87	15.00	6.00	32.00	15.00

Fig. 3.1 through 3.6 shows the histogram plots of the individual assessments.

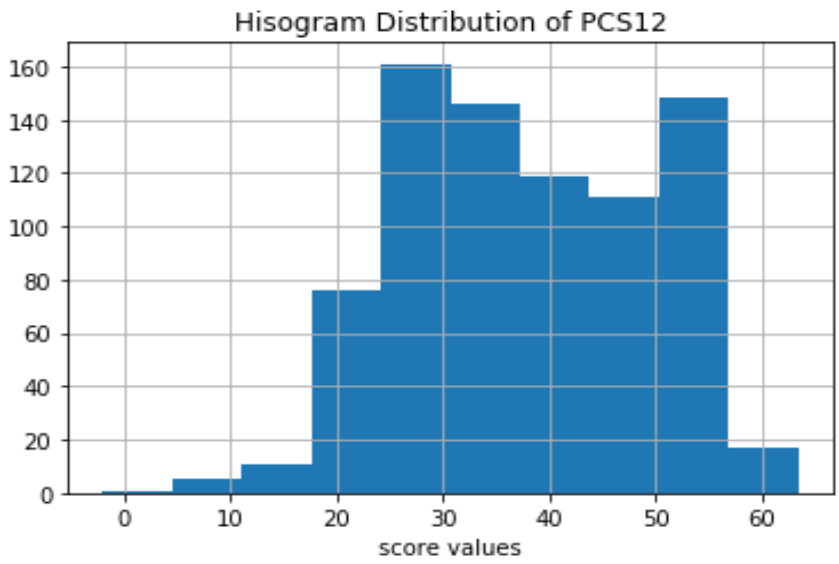


Figure 3.1. Histogram distribution of PCS 12

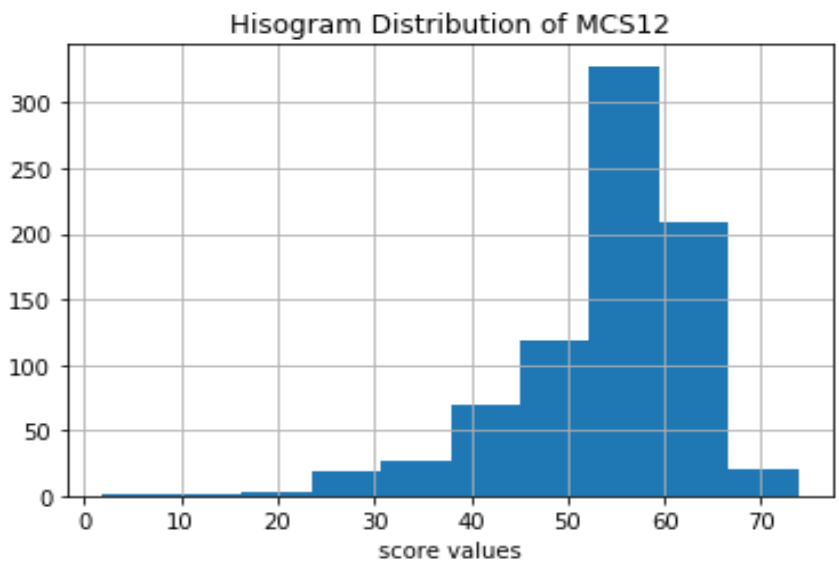


Figure 3.2. Histogram distribution of MCS 12

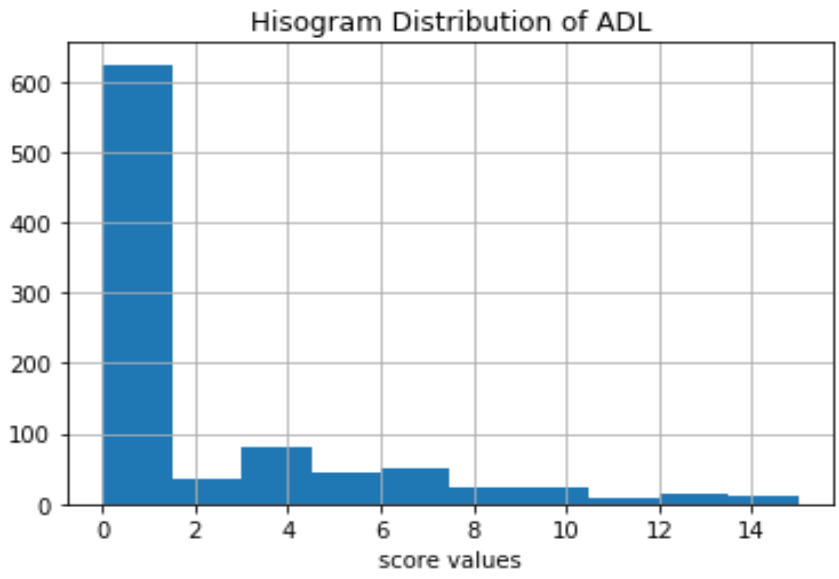


Figure 3.3. Histogram distribution of ADL

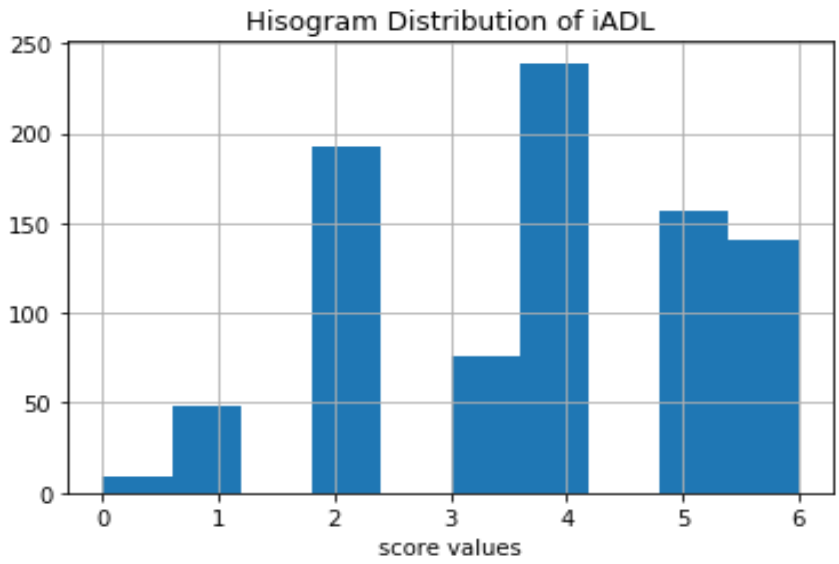


Figure 3.4. Histogram distribution of IADL

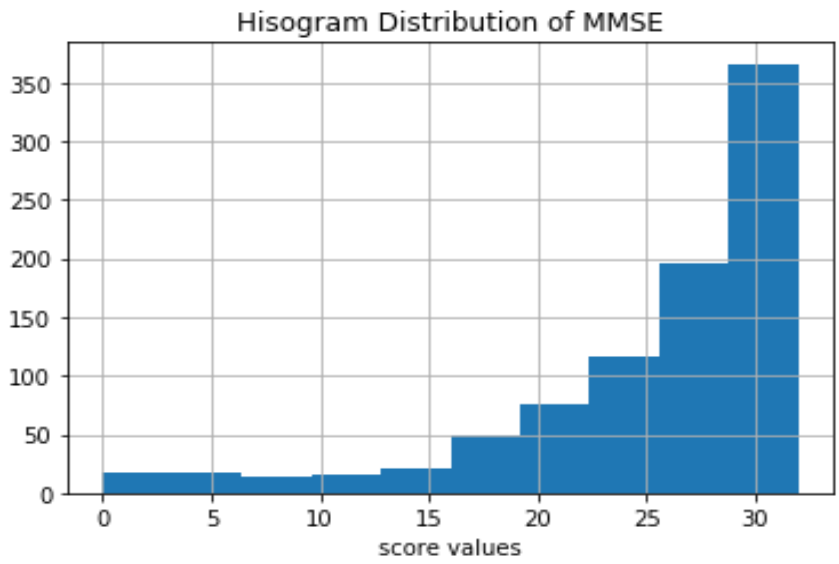


Figure 3.5. Histogram distribution of MMSE

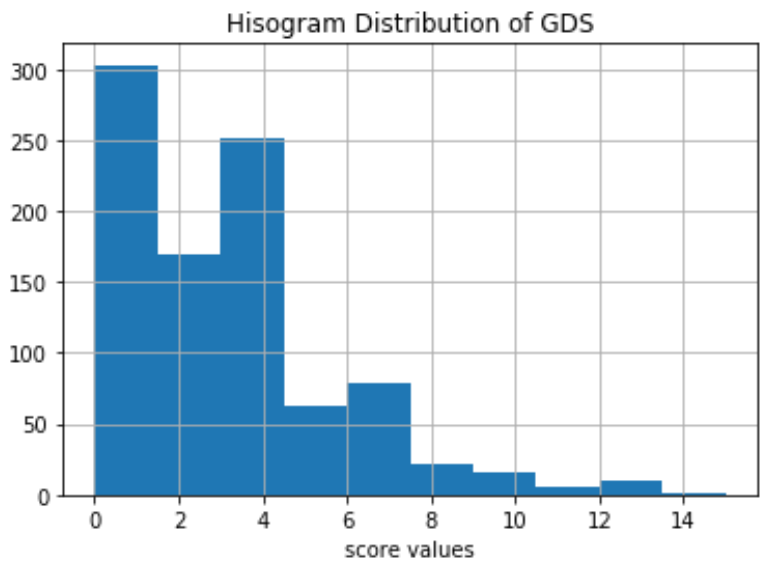


Figure 3.6. Histogram distribution of GDS

We observed that ADL data is extremely skewed as compared to other assessments.

Chapter 4: Development and Validation of Health Risk Prediction Models for Older Adults Using the Geriatric Assessments

4.1 Introduction

Higher levels of functional health in older adults lead to a higher quality of life and improves the ability to age-in-place. Measuring functional health objectively could help clinicians to make decisions for interventions in case of health deterioration. Even though several geriatric assessments capture several aspects of functional health, a composite index representing continuous overall functional health is missing.

We used geriatric assessment data collected from 168 older adults to develop and validate a heuristic functional health index (FHI) model based on risks associated with falls, hospitalizations, emergency visits, and death. The geriatric assessments included were Activities of Daily Living (ADL), Instrumental Activities of Daily Living (IADL), Mini-Mental State Examination (MMSE), Geriatric Depression Scale (GDS), and Short Form 12 (SF12). FHI quantifies the overall functional health risk by summing the risks associated with the individual assessment scores at a given period. Construct validators such as health events vs FHI, six-month fall, six-month mortality, and functional ambulation performance score were used to validate the FHI model.

The FHI model is shown to separate fall or death events from all other health event categories with the area under the receiver operating characteristic curve (AUC) of > 0.73. An AUC of > 0.72 was obtained in predicting a fall within six months, and an AUC of >

0.71 for predicting mortality within six months. Multiple comparisons of means using the Turkey HSD test show that FHI for no health events category versus fall or death was statistically significant ($p < 0.05$). Case studies with FHI trajectories show that changes in FHI over time correspond to critical functional health changes in older adults.

In addition to developing the FHI model using excess risks, we constructed mixed-effect logistic regression models to predict adverse health event outcomes, such as falls and hospitalizations from the health assessments. The repeated assessment measures from an individual resident could be correlated. To overcome the with-in-subject correlations due to repeated measures, we considered mixed-effect models in our analysis. In this analysis, residents are treated as random effects.

The FHI may provide clinicians with a longitudinal view of overall functional health in older adults to help address the early detection of deterioration trends and determine appropriate interventions. It can also help older adults and family members take proactive steps to improve functional health, such as physical therapy, increasing time walking and strength training, or balance exercises to improve balance.

4.2 Background

The number of Americans ages 65 and older is projected to be over 98 million by 2060, which is about 24 percent of the total population in the USA [124]. The aging population is at a higher risk of functional decline than their younger counterparts [125]. Keeping persons over 65 at higher functional levels can lead to a higher quality of life, successful aging-in-place, and reduce healthcare expenditures [126]. To predict the functional health status of the residents using health assessments we evaluated two different methods, (1) excess risk functions and (2) mixed-effect modeling.

4.2.1 Functional Health Index Development Using Excess Risk Functions

FH in older adults is complex and multifactorial [127, 128]. Gordon has defined 11 FH patterns to facilitate nursing diagnoses [127]. The list of FH patterns included health-perception, activities of daily living, cognitive ability, and self-perception. This suggests that FH is not only limited to physical function, but rather is a combination of physical, cognitive, and social function, among other factors. The World Health Organization's 2015 World Report on Aging and Health outlines a framework for Aging-in-Place around the new concept of functional ability [128]. It reinforces that FH is a combination of physical, cognitive, and social functions, and also suggests that the loss of these functions has a detrimental impact on an older adult's health status, quality of life, and independence [128, 129]. Therefore, in this study, we have used a specific set of geriatric assessments that can measure multiple aspects of physical, cognitive, and social function to predict overall FH.

Several geriatric assessments have been developed and validated for measuring different aspects of functional health in older adults, such as Activities of Daily Living (ADL), Instrumental Activities of Daily Living (IADL), Mini-Mental State Examination (MMSE), Geriatric Depression Scale (GDS), and Short Form 12 (SF12) [31-40]. ADLs are defined as activities that are essential for independent living. IADL require a higher level of personal autonomy, referring to tasks that require enough capacity to make decisions through greater interaction with the environment [38]. MMSE is a widely used test to evaluate the cognitive aspects of mental function. MMSE excludes questions concerning abnormal mental experiences and mood [36]. GDS is a screening tool for measuring

depression in older adults [39]. SF-12 is a multipurpose short form that provides a generic measure of health status [40]. All these assessments have good reliability and validity. These individual assessments can track different aspects of functional health in older adults. However, understanding the overall change in the functional health of an older adult using these individual assessments can be difficult, as they are based on different scales and have a wide range of clinical condition focus and importance to overall health. Instead of considering them individually, a combination of these assessments can be used to develop a continuous composite functional health score to predict a more general status of functional health in older adults.

Several studies have been conducted in developing health and prognostic indexes to track or predict comorbidity, mortality, frailty, and functional health in older adults [41-46]. Mazzaglia et al. developed two prognostic index models to predict five-month mortality and hospitalization [41]. In the first model, they used a set of 7 questions from ADL and IADL to develop their index. The area under the receiver operating characteristic curves (AUC) to predict mortality and hospitalizations were 0.75 and 0.60, respectively. In the second model, they considered drug use and previous hospitalizations, which increased their hospitalization AUC to 0.67. Gagne et al. developed a single numeric index to predict mortality by combining Charlson and Elixhauser measures [42]. Results show that the combined score performed better in predicting mortality than the individual scores. Carey et al. and Lee et al. developed prognostic models to predict mortality using data from the Program of All-Inclusive Care for the Elderly (PACE) and 1998 wave of the Health and Retirement Study (HRS), respectively [43, 44]. Schonberg et al. used 39 risk factors, including functional measures, illnesses, behaviors, demographics in a multivariable Cox

proportional hazards model to predict 5-year mortality [45]. Giovanni et al. developed a multisource comorbidity score using administrative data, such as diagnostic categories and ICD-9 to measure comorbidity, predict 1-year mortality, and other adverse outcomes [46]. This study did not include functional status as a variable in the predictive model development.

Fried. et al. conducted a study to predict frailty in older adults [47]. They defined frailty as a clinical syndrome in which the older adult has three or more out of five frailty criteria. These five criteria include unintentional weight loss, self-reported exhaustion, weakness, slow walking speed, and low physical activity. This standardized phenotype of frailty detection can identify frail older adults potentially at risk of falls, hospitalizations, disability, and death. Rockwood et al. developed a 7-point Clinical Frailty scale to predict death or need for institutional care [130]. The Clinical Frailty scale is based on an a-priori selection of features and is intended to predict mortality or the need for institutional care.

Gait analysis provides critical information in predicting adverse health conditions and functional health in older adults [13, 48]. Nelson et al. developed a Functional Ambulation Performance (FAP) score that captures the gait capacity of an individual using a specific set of spatiotemporal parameters (STPS) [49, 50]. The FAP score is integrated with the GAITRite walkway (CIR System Inc; Clifton, New Jersey), a gold standard system for measuring STPS [51]. FAP has been validated in several independent studies [50].

The mortality-based prognostic models tend to predict future adverse health conditions, specifically death, instead of predicting the overall functional health of an individual at a given time. The frailty phenotyping method developed by Fried et al. can

predict the presence of frailty with minimal granularity as it can only classify an older adult into one of the three frail categories: frail, intermediate frail, and not frail [47]. This may not be able to track the gradual changes in the functional health of an individual. Also, this only considers the physical aspects of functional health; cognitive aspects were excluded [52, 53]. The FAP score is also developed considering only gait ability and does not consider the cognitive aspects of functional health [49]. We argue that a continuous measure of overall functional health can provide critical information about changes in functional health over a period. Interventions based on overall functional health deterioration can help older adults live with higher independence and quality of life [54].

Santoni et al. used gait speed, cognitive function, chronic multimorbidity, and disability to predict present and future care need in Swedish older adults [131]. Their model could predict hospitalization with an AUC of 0.78 (95 CI =0.74-0.81) and mortality with an AUC of 0.85 (95% CI=0.83-0.87). The dataset used in the study included older adults with high levels of cognitive and physical function; at least 90% of participants were free of severe disability, and at least 50% were functionally independent despite chronic disorders. In contrast, in our study, the dataset includes older adults with comparatively lower cognitive and physical function. The study did not include falls and emergency visits as outcome measures.

Rothman et al. have developed a continuous measure of patient condition in acute care, called the Rothman Index, using a set of 26 Electronic Medical Record variables [15]. They used one-year post-discharge mortality to develop excess risk functions for each variable. The excess risk functions were then used to calculate the Rothman Index for an individual at a given period. Instead of predicting an adverse health condition, the Rothman

Index provides an overall patient condition in inpatient care. Based on the Rothman Index method, we report the development and validation of a continuous measure of functional health, Functional Health Index (FHI), for chronic care using geriatric assessments and excess risk functions for each assessment type.

We hypothesized that detecting a decline in an individual's FH would represent deterioration in underlying health conditions recorded in the electronic health record. In this study, we develop and validate a method for continuous tracking of personalized FH of older adults using routine geriatric assessments and adverse health outcomes. We use a mixed-effects logistic regression model that allows us to use repeated measurements to build the model and provide personalized health predictions. We hypothesize that these geriatric assessments would provide sufficient information in developing a personalized FH tracking model. We believe that continuous tracking of FH could help early detection of health deteriorations and facilitate earlier interventions by health professionals to improve the health of an older adult.

4.2.2 Mixed Effect Models to Predict Adverse Health Events

A major demerit of the Rothman-style analysis in developing the Functional Health Index is that the model does not account for intra-resident data biases because of the repeated assessment measures. The assessment dataset contains repeated measurements from the residents at TigerPlace obtained over 10 years. These repeated measurements could potentially have high correlations among them. For example, a resident with very low functional health and multiple chronic health conditions will have a significantly different set of assessment values when compared with a healthier resident with no chronic health condition and higher functional health. In such scenarios, the assessments obtained

from these two residents will likely have significantly different distributions. To avoid such bias in the data because generalized linear mixed-effect models could be used to model the assessment data [132-134].

Generalized linear mixed models (GLMMs) extend linear mixed models to allow response variables from different distributions, including binary responses [134, 135]. Mixed effect models consider both fixed-effects and random-effects in a dataset. In our analysis, the geriatric assessments are the fixed effects, and the residents introduce random effects. The general form of the GLMMs is,

$$y = X\beta + Zu + \varepsilon$$

Where y is an $N \times 1$ column vector representing the outcome variable, X is an $N \times p$ matrix of p fixed-effect predictor variables, β is a $p \times 1$ column vector of fixed-effect regression coefficients, Z is an $N \times q$ design matrix for the q random-effects, u is the $q \times 1$ vector of the random effects, and ε is an $N \times 1$ column vector of the residuals. In our analysis, Z depicts the unique residents in our dataset. Z possibly being too large, it only contains 1s and 0s. Each column represents a resident and each row represents an assessment set. If an assessment belongs to a resident in that column the cell value is 1, 0 otherwise.

The vector u is a normal distribution with zero mean and variance G. *u is* generally defined as,

$$u \sim \mathcal{N}(0, G)$$

Where G is the variance-covariance matrix of random effects, defined as (considering we have both random intercepts and slopes),

$$\mathrm{G} = \begin{bmatrix} \sigma^2_{int} & \sigma^2_{int,slope} \\ \sigma^2_{int,slope} & \sigma^2_{slope} \end{bmatrix}$$

Fixed effects are directly estimated, whereas, the random effects are modeled as deviations from the fixed effects with a mean zero. Therefore, random effects are deviations around the value β. In this study, we have only considered random intercepts so the matrix G would just be a 1× 1 matrix. G is a square, symmetric, and positive semi-definite that contains redundant elements. Therefore, for simplification, instead of directly estimating G, θ is estimated (e.g., a triangular Cholesky factorization $\mathrm{G} = \mathrm{LDL}^{\mathrm{T}}$). In a more general form, G can be represented as (a function of θ),

$$\mathrm{G} = \sigma(\theta)$$

The response variables in GLMMs can come from different distributions besides gaussian. The responses can also be modeled using link functions, such as log link. A linear predictor can be defined as,

$$\eta = X\beta + Z\gamma$$

Here η is defined as a linear function that is a combination of fixed and random effects. η excludes the residuals. The link function relates η with the outcome variable *y*. The model for the conditional expectation of *y* is,

$$g(E(\mathrm{y})) = \eta$$

Where g(.) is the link function. The expectation of y can be modeled as,

$$E(\mathrm{y}) = h(\eta) = \mu$$

Where y is,

$$\mathrm{y} = h(\eta) + \varepsilon$$

4.3 Experiments

4.3.1 Data

The proposed model of the functional health index is based on a set of frequently collected geriatric assessment scores in the Electronic Medical Record (EMR), such as ADL (Short Form ADL, RAI MDS 2.0), IADL (Lawton), GDS, MMSE, and SF12 [53, 122, 123]. The SF-12 assessment has two components, a physical component or PCS and a mental component of MCS. We used assessments collected at TigerPlace, an Aging-in-Place facility in Columbia, MO, on 168 independent living older adult residents (females = 106, age = 82.8 ± 8.0) [5]. We included all residents living at TigerPlace for over eight years, from 2011 to 2019. All assessments were collected at a regular interval of six months. Multi-collinearity was determined using the Pearson correlation coefficient for the assessments. None of the included assessment pairs had a Pearson correlation greater than 0.7. The final dataset contained 4,853 individual assessments. The number of assessments in each assessment category was comparable.

Four different health events, including falls, emergency visits, hospitalizations, and death were considered to construct the risk models for the assessments. These health events were assumed to reflect the underlying functional health deterioration of an individual. The dataset contained 2,677 health events, out of which 1,931 were falls. Tables 4.1 (a) and 4.1 (b) show the breakdown of the number of assessments and health events included in the study, respectively. Table 4.1 (C) shows a summary of the characteristics of the assessment data.

For validation purposes, we randomly divided the dataset into two groups of residents: Group-A and Group-B. Group-A contained data from 100 residents and was used to construct the model. Group-B contained data from the remaining 68 residents and was used to validate the model. Individual residents were not considered as samples in our dataset, instead, six months of every individual resident was considered as a sample. Each of these samples typically contained a set of five assessments and health events for that resident during those six months. For example, if a resident lived at TigerPlace for five years, the dataset would have ten samples for that resident, each containing a set of assessments and health events. This study received Institutional Review Board approval at the University of Missouri, Columbia.

Table 4.1 (a) Number of geriatric assessments in each assessment type.

	Geriatric Assessments					
	ADL	IADL	GDS	MMSE	SF-12 (PCS)	SF-12 (MCS)
Count	814	777	878	854	765	765

Table 4.1 (b) Number of health events in each health event type.

	Health Events			
	Falls	Emergency Visits	Hospitalizations	Deaths
Count	1931	396	350	61

Table 4.1 (c) Assessment Data Characteristics.

Assessments (Range) *	Mean (Std)
ADL (0 -16)	2.19 (3.23)

IADL (0 - 8)	3.88 (1.57)
MMSE (0 - 30)	25.09 (6.54)
GDS (0 - 15)	2.88 (2.45)
SF-12, mental score (0 - 100)	54.31 (9.17)
SF-12, physical score (0 - 100)	37.76 (11.85)

* Interpretation of the assessment scores - ADL, higher scores indicate more ADL impairment; IADL, lower scores show low function; MMSE, lower scores show more cognitive impairment; GDS, higher scores indicate depression; SF-12, low scores indicate a low level of mental or physical health

4.3.2 Model Construction using Excess Risk

The assessments had different ranges of scores. All assessment scores were normalized and discretized to integer values ranging from 0-10. After normalization, excess risk functions were computed for the individual assessments. We defined the excess risk functions as the percentage increase in six-month post-assessment health events associated with any value of an assessment, relative to the minimum six-month health events identified for that assessment. Only one health event per resident over six months was included in the excess risk function development. This was done to avoid biases, as some residents had several health events in six months. Figure 4.1 shows the excess risk function for the IADL assessment. The black dot points represent excess risk percentages for normalized IADL scores. The regression fit represented by green diamonds is a polynomial fit to the data points. The excess risk functions were set to a constant where the

data are sparse and were not extrapolated as the underlying function is unknown. We used the six months to determine excess risk functions as new assessments were obtained every six months.

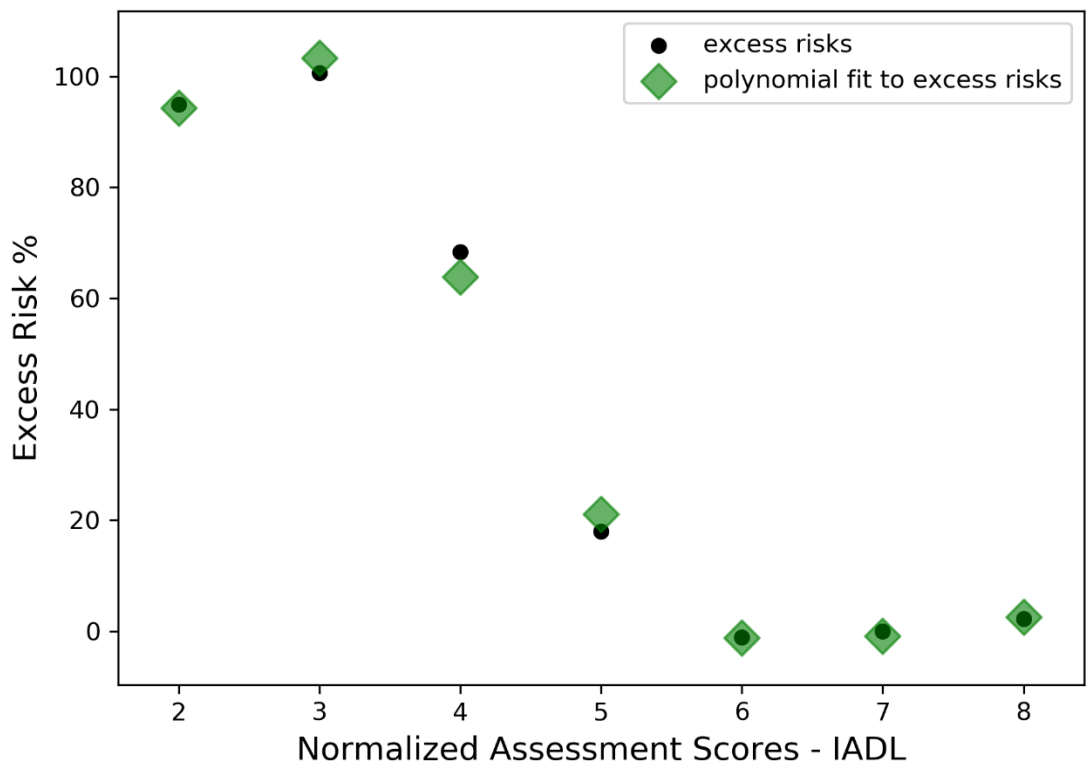


Figure 4.1. Excess six-month health events risk as a function for normalized IADL scores

The excess risk functions were used to determine the Functional Health Index (FHI) for a given six-month period as shown in equation (1).

$$FHIndex = 100 - (ScaleFactor)\sum_{i=1}^{\#Assessments} ExcessRisk_i \qquad (1)$$

A scaling factor is used to ensure that most of the FHI values fall between 0-100. The product of a scaling factor and excess risk sum is subtracted from 100 to make sure that a higher FHI represents a lower functional health risk. To compute continuous FHI values for a resident, missing assessment scores are updated by using the most recent assessment scores. The FHI values for a resident are computed every six months and considered to remain constant over the six months.

4.3.3 Model Assessment and Construct Validators

Similar to the Rothman Index, FHI is a heuristic model and is not designed to predict a specific quantity [15]. The validity and reliability of such a method cannot be exactly quantified [136, 137]. To validate FHI, we followed the construct validity methodology adopted by Rothman et al. and Richardson et al. in developing the Rothman Index and Score for National Acute Physiology, respectively [15, 136]. Boudreaux et al. defined construct validity as ‘‘. . . the degree to which a measure actually assesses the attribute it is purported to measure’’ based on ‘‘whether the measures relate to other variables in expected and predictable ways’’ [138, p. 168]. The relationship of the FHI to a health event independently associated with functional health was examined on the assumption that poorer functional health is expected to correspond to more serious health events.

A set of construct validators was chosen to validate the FHI model. Validation results are reported on both resident groups (Group - A and Group - B) using sensitivity-specificity analysis and area under the receiver operating characteristic curves (AUC) [139, 140].

The correspondence of FHI to different health events was evaluated. The four health event categories considered for this evaluation were: no health event, hospitalization, or an emergency visit, fall, and death. In some instances, multiple types of health events co-occurred in the same six-month period sample. For example, a fall followed by an emergency visit or hospitalization. To avoid such data overlap, samples with more than one category were discarded. The exclusion of overlapped samples was done with one exception. We considered all samples with death events in the death category, irrespective of their overlap with other categories. This was done assuming that death is the worst health outcome and overlap of other health events just indicate the severity of health deterioration at the end of life. However, samples with death events were removed from other categories to make all health event categories independent of each other. After the exclusion of overlapped samples, the remaining 89.10% of samples from Group-A and 90.74% of samples from Group-B were used for the validation.

4.3.3.1 Health Events vs FHI

Average FHI values for all health event categories in both groups were computed. We employed analysis of variance (ANOVA) procedures to determine the statistical significance between FHI values associated with the four categories. The ANOVA analysis was followed with multiple comparisons of means using the Turkey HSD post-hoc test [141]. Also, FHI values were used to separate different health event categories, specifically, no health event versus the rest and no health event and hospitalization and emergency visits versus the rest. The area under the curve (AUC) associated with each separation was calculated.

4.3.3.2 Six-month Fall

To validate the effectiveness of FHI in predicting a fall within six months, the average six-month fall percentages associated with FHI values were computed. Also, FHI values were used to separate the no health event category from the fall category. AUC associated with the separation was computed for each group.

4.3.3.3 Six-month Mortality

The relationship between six-month mortality and FHI was examined. Average six-month mortality percentages associated with FHI values were computed. To examine the sensitivity of the FHI to mortality, AUC associated with the separation of the death category from the rest of the three categories was computed.

4.3.3.4 FAP Score

STPS and the FAP score of TigerPlace residents have been collected every six months using the GAITRite system. The FAP score has been shown to represent underlying impaired functional health [50]. We conducted a correlation test between FHI and FAP scores in both groups to evaluate FHI as a functional health indicator.

4.3.3.5 Case Studies

We explored two case studies to evaluate the correspondence of FHI with actual functional health changes observed in EMR. FHI values were computed for the entire stay of these residents at TigerPlace. A timeline of FHI values was plotted to represent the FHI trajectory of each resident. An investigation of the clinical notes was performed to obtain the actual functional health changes reported in the EMR for these residents. The ground truth on functional health changes was compared with the changes observed in the FHI trajectory.

4.3.4 Mixed Effect Modelling to Predict Fall and Hospitalization Risks

We used mixed-effect modeling to predict fall events that caused hospitalizations using geriatric assessments. We used mixed-effects logistic regression to model binary health event outcomes. The two binary outcomes used for the analysis were, hospitalizations associated with falls vs everything else (including no health events, independent fall events, independent hospitalization events, independent emergency visit events, and independent death events). Apart from the geriatric assessments, we also considered percentage change in weight, falls in the last 6 months, and age during assessment into account for this analysis. The distribution of the continuous variables is shown in Fig. 4.2.

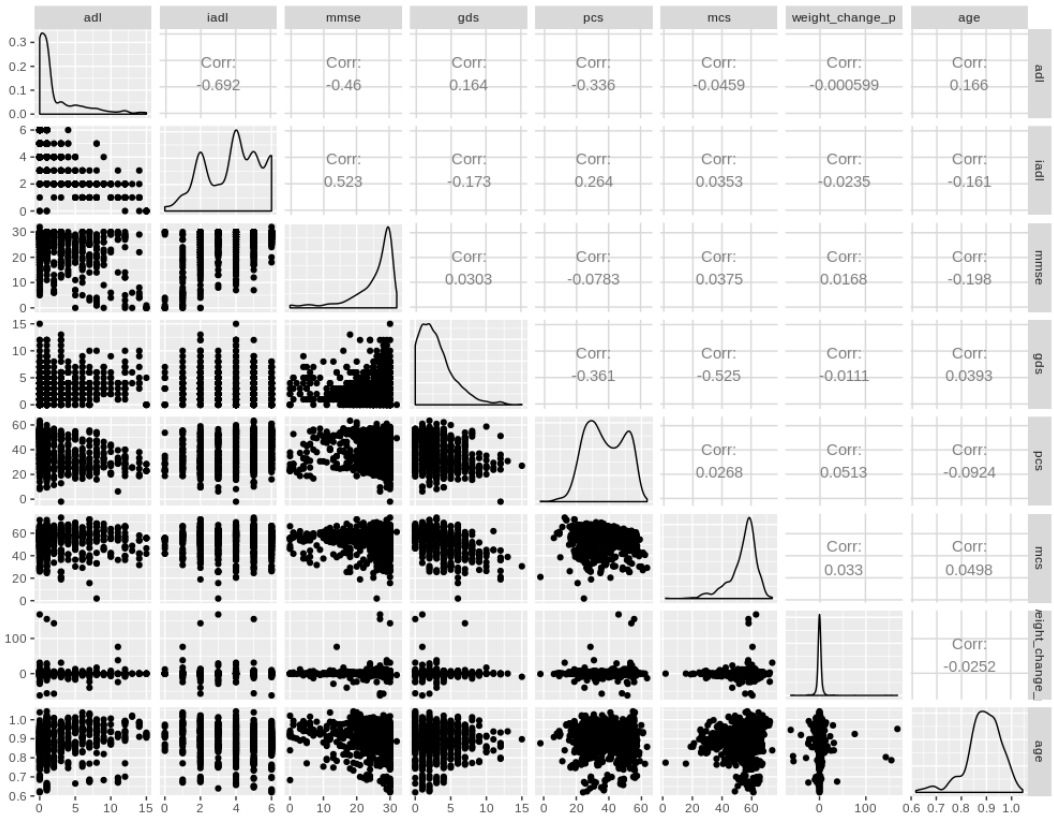


Figure 4.2. Pair plot presenting the distribution of the continuous variables in the analysis

Fig 4.2 shows that ADL and IADL had the highest amount of correlation (-0.692). However, the absolute correlation value was < 0.7 so we considered both variables for the modeling. We used the *glmer* function from the *lme4* R library to perform the mixed effect logistic regression [142-144]. Finally, we performed a multilevel bootstrapping to improve the model [145, 146]. For each resident in our study 2000 samples were resampled with a replacement for the bootstrapping.

The assessments collected over time are nested within the residents. The assessments were considered as fixed effects, and the residents were considered as the random effects because assessment measures collected within the residents may be correlated. Modeling residents as a random effect provides the personalization effect, as the model predictions depend on who the resident is instead of just the assessment scores. The predicted probabilities from the final model were subtracted from 1.0 so that higher values represent better health status and vice versa. We refer to these values as functional health values (FHV) in the rest of the article. We used construct validators such as health event categories vs FHV and six-month fall percentage vs FHV to validate the mixed-effects model.

4.4 Results

In this section, we present the results associated with each construct validator for the excess risk-based FHI model. We also present the results associated with the mixed effect logistic regression model.

4.4.1 Excess Risk Model Validations

4.4.1.1 Health Events vs FHI

Table 2 shows the mean, standard deviation, and percentage of data samples in each health event category for Group-A and Group-B. Mean values for the four categories show that the first two categories (no health event, emergency visit, or hospitalization) have a higher FHI mean compared to the fall and death event categories.

Multiple comparisons of means using the Turkey HSD test show that all pairs of health event categories were statistically significant ($p < 0.05$), except for two pairs: no health event versus emergency visit or hospitalization and fall versus death [147]. Fig. 4.3 shows FHI values for different health event categories. We used the FHI values for separating the first two categories from the last two (fall and death) and obtained an AUC value of 0.733 (95% CI, 0.686-0.78) for Group-A and an AUC value of 0.732 (95% CI, 0.675-0.789) for Group-B. Fig. 4.4 shows the receiver operating curve for the separation. A statistically significant difference was found between FHI values associated with the different health event categories for Group-A, $F(474) = 25.63$, $p < 0.0001$, and Group-B, $F(367) = 16.97$, $p < 0.0001$ [148].

Table 4.2 Mean FHI by Health Event Category for Group-A and Group-B.

Health Event Category	**Group-A (n = 474)**	**Group-B (n = 367)**
No health event	**71.19** (16.17) [a] 56.02%	**69.39** (14.66) 60.96%

Emergency Visit or Hospitalization	**69.92** (17.03) 6.96%	**68.94** (14.23) 6.61%
Fall	**58.10** (14.15) 27.64%	**58.06** (12.71) 27.93%
Death	**53.84** (10.53) 3.38%	**55.32** (12.71) 4.50%

[a] Mean FHI is in bold, with standard deviations in parenthesis, followed by sample size percentages.

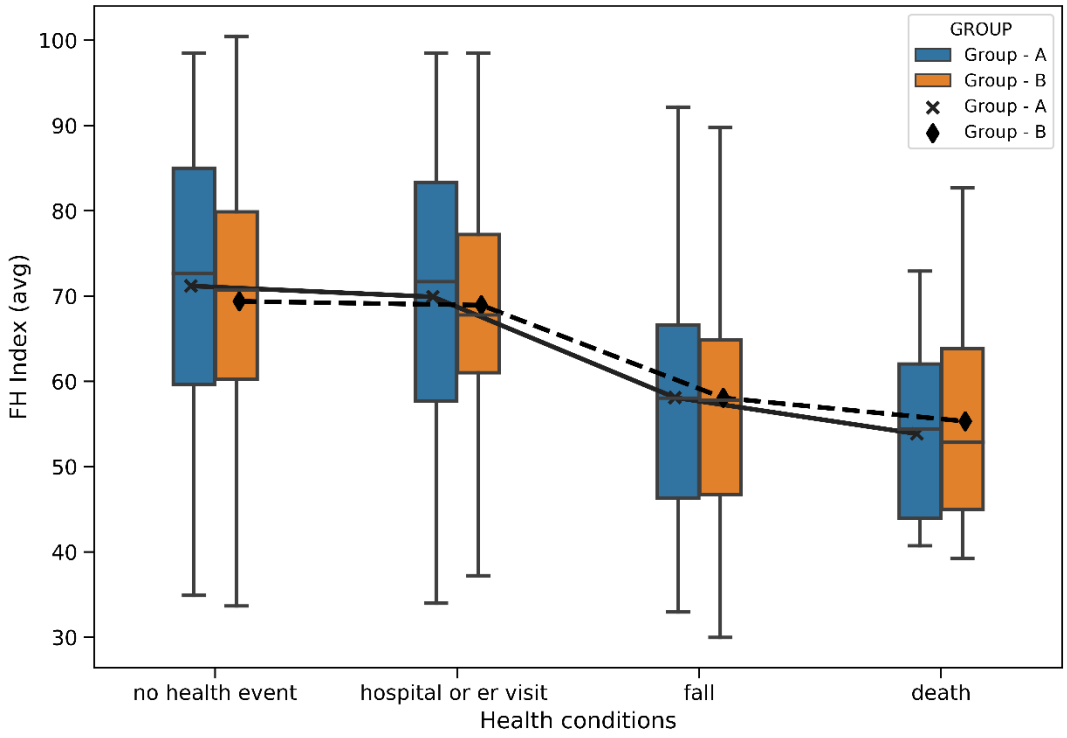


Figure 4.3. FHI vs health event categories for Group-A and Group-B.

The top and bottom of each box represent 75% and 25% percentiles of FHI for that category. Horizontal lines in each box represent the median FHI values for each category. The top of each whisker represents the maximum value of FHI in that category or median plus 1.5 times the interquartile range; the bottom whisker represents the minimum value of FHI in that category or median minus 1.5 times the interquartile range.

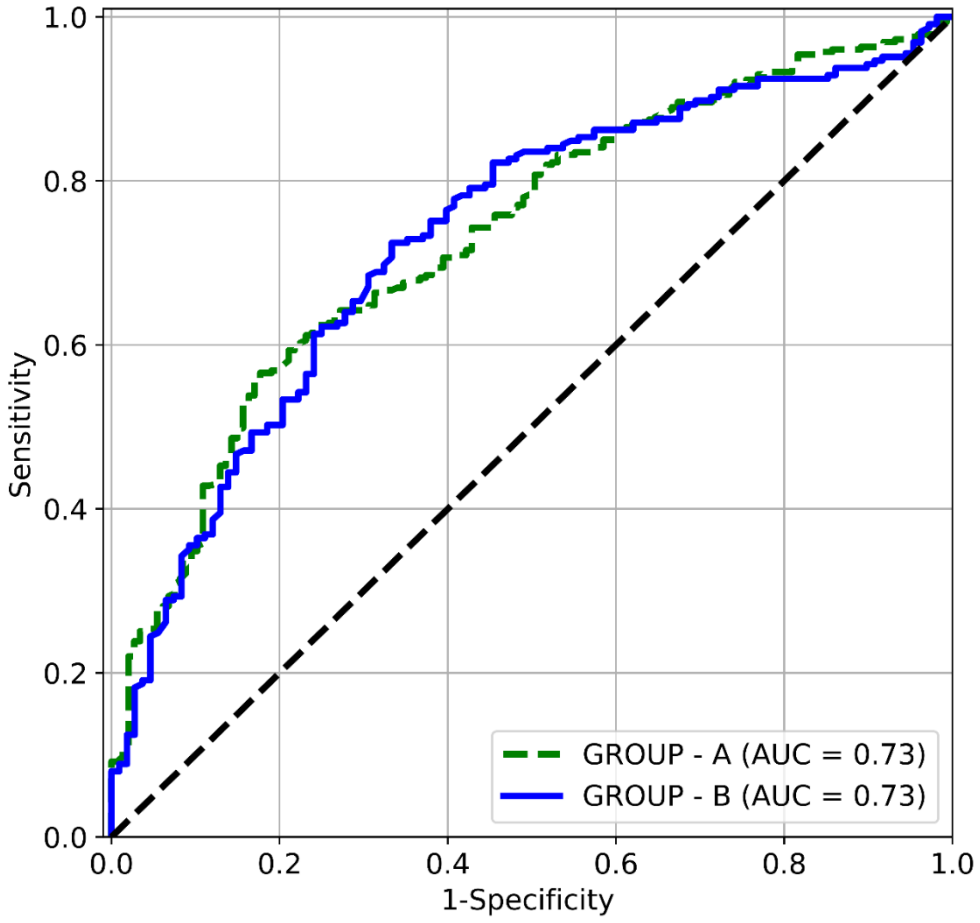


Figure 4.4. ROC showing the separation of the first two health event categories from the last two. Receiver operating curve showing the separation of the first two health event categories (no health event and emergency visit/hospitalization) from the last two health event categories (fall and death).

4.4.1.2 Six-month Fall vs FHI

The average likelihood of fall within six months of computing the FHI is shown in Figure 4.5 The AUCs for separating the no health event category from fall were: 0.726 (95% CI, 0.676-0.776) for Group-A and 0.726 (95% CI, 0.666-0.787) for Group-B. Fig. 4 shows that both groups had similar six-month average fall percentages. Higher FHI values correspond to a lower average fall percentage and vice versa.

4.4.1.3 Six-month Mortality vs FHI

The average likelihood of death within six months of computing the FHI is shown in Fig. 4.6. The AUCs for separating all health event categories from death were: 0.742 (95% CI, 0.649-0.834) for Group-A and 0.712 (95% CI, 0.58-0.844) for Group-B. A lower FHI score corresponds to a higher average death percentage and vice versa. Fig. 5 shows that both groups had similar six-month average death percentages.

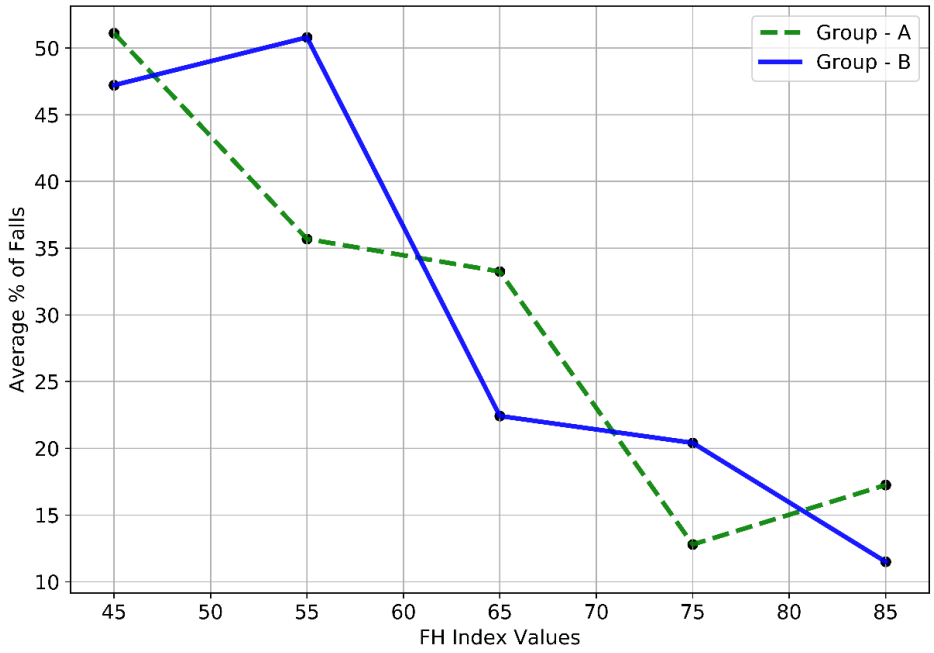


Figure 4.5. Average % of Falls vs FHI.

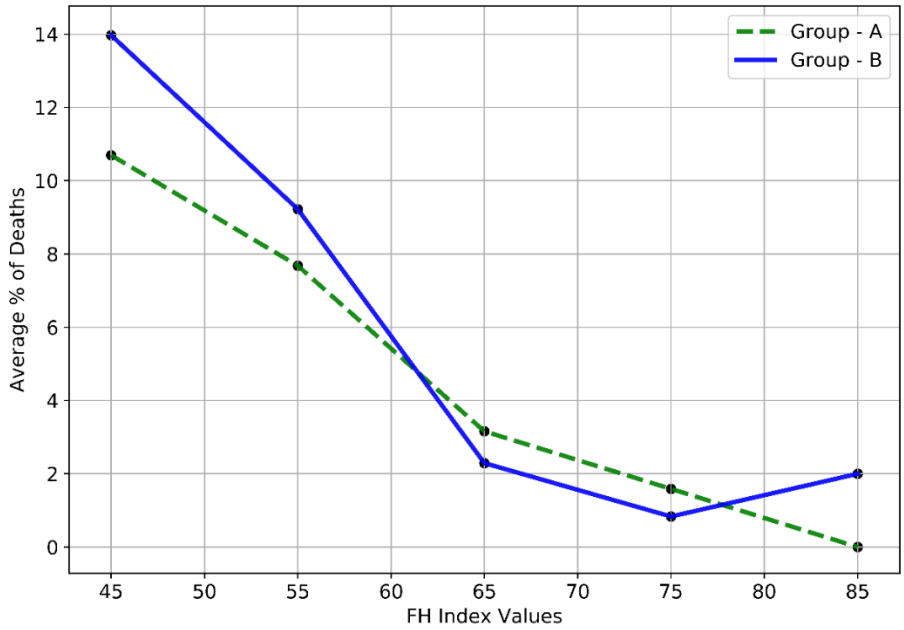


Figure 4.6. Average % of Deaths vs FHI.

4.4.1.4 FAP Score vs FHI

The FAP score from every six months was compared with the respective FHI value. A significant correlation was found (Pearson's correlation = 0.633, $p < 0.0001$). The positive correlation indicates that higher FAP score values correspond to higher FHIs.

4.4.1.5 Case Studies using FHI

Two case studies are presented demonstrating correspondence of changes in FHI with significant health changes reported in EMR clinical notes. See Appendix A for more examples of FHI Trajectories.

4.4.1.5.1 Case Study - 1

Fig. 4.7 shows the FHI trajectory of a TigerPlace resident. Falls, emergency visits, and hospitalizations experienced by the resident are marked on the FHI trajectory timeline. A visual assessment of the plot suggests that lower FHI values correspond to falls, emergency visits, and hospitalizations. The resident did not experience any critical health event during the period between July 2016 to January 2018 when the FHI values were higher.

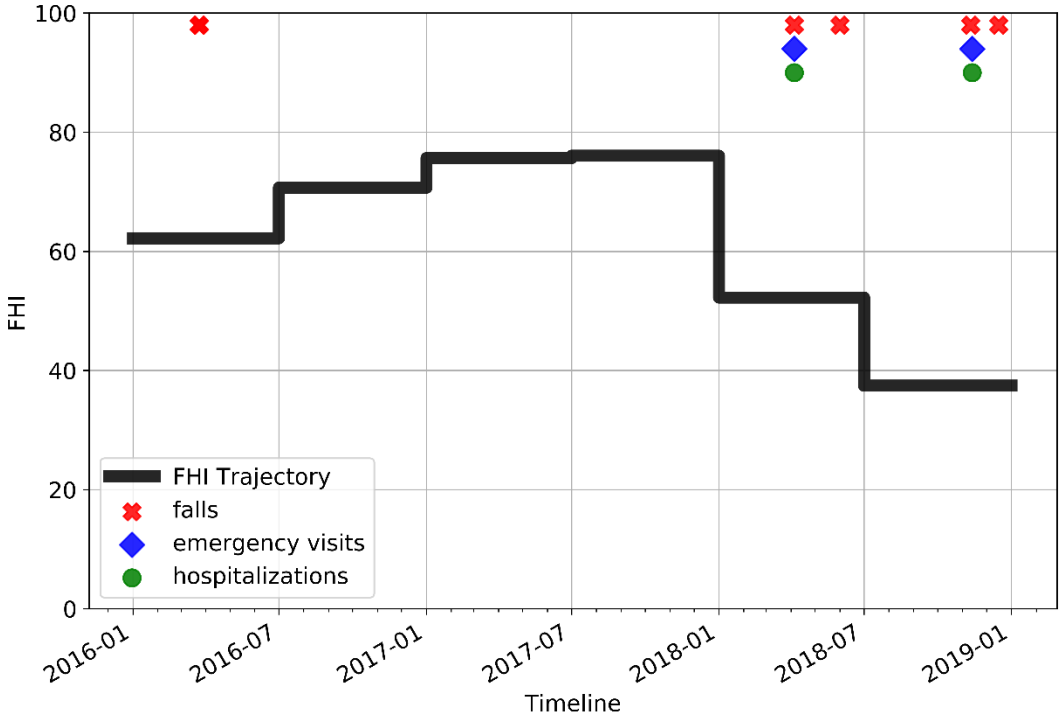


Figure 4.7. FHI Trajectory of a TigerPlace resident from Case Study – 1.

An analysis of EMR clinical notes suggests that the resident was able to walk independently without any support until January of 2018. The resident started walking with a walker and was often in a wheelchair starting from April 2018. A sharp decline in the FHI trajectory at the beginning of 2018, with FHI < 60, confirms that FHI decline may correspond to significant functional health changes.

4.4.1.5.2 Case Study - 2

Fig. 4.8 shows the FHI trajectory of another resident at TigerPlace. An investigation of the resident's EMR clinical notes suggests that the resident had chronic pain and was in a wheelchair for the entire period.

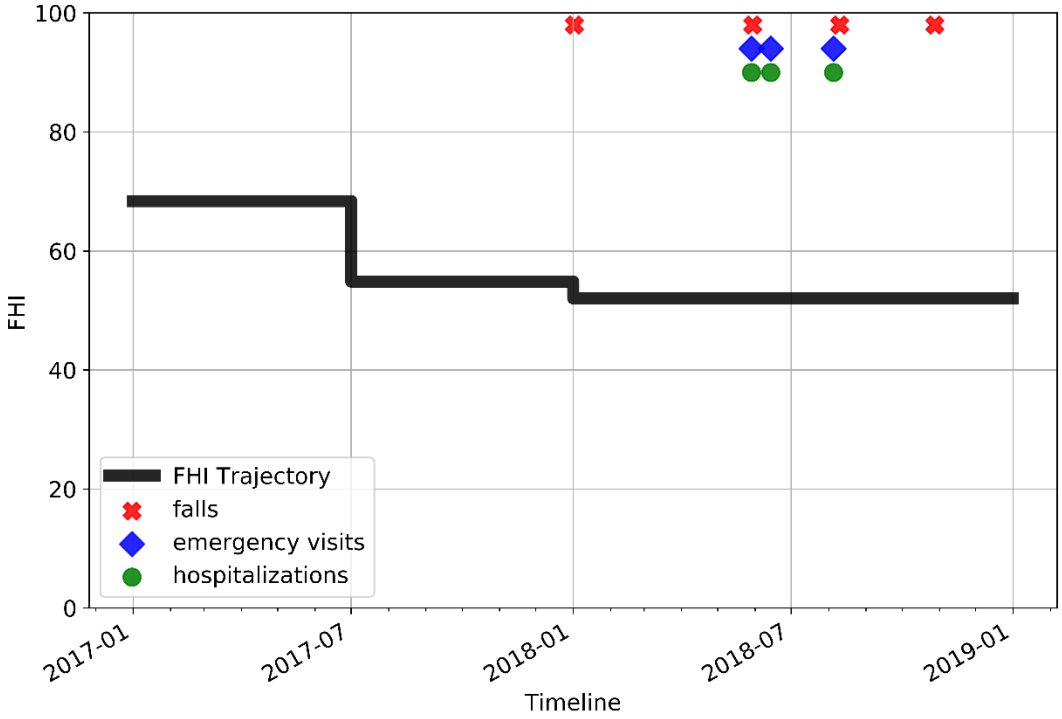


Figure 4.8. FHI Trajectory of a TigerPlace resident from Case Study – 2.

An analysis of EMR notes show that the resident experienced increased back pain during October 2017 and cognitive impairment during November 2017. These health changes correspond to the significant decline in FHI observed in late 2017. FHI for the resident further declined in the later months (FHI ~ 50) corresponding to increased functional health deterioration, leading to several falls, emergency visits, and

hospitalizations between 01-2018 to 01-2019. This shows that a decline in FHI may correspond to cognitive impairments and other functional health deteriorations.

4.4.2 Mixed Effect Logistic Regression for Predicting Adverse Health Events

4.4.2.1 Health Event Categories vs FHV

Table 4.3 shows the mean, standard deviation, and sample size of data in each health event category. Mean values for the five categories show that FHV associated with no health events were higher when compared to the samples associated with hospitalization, emergency visit, fall, and death.

A one-way ANOVA was calculated on FHV values associated with the different health event categories. The analysis was significant ($F=154.99$, $p < .0001$). Multiple comparisons of means using the Turkey HSD test show that all pairs of health event categories were statistically significant ($p < 0.001$), except for two pairs: fall only versus fall with hospitalization and death versus fall with hospitalization [40]. Fig. 4.9 shows FHV for different health event categories. FHV for no health events were well separable from the rest of the health events with an AUC value of 0.85 (95% CI, 0.83 - 0.88). Fig. 4.10 shows the receiver operating curve for the separation.

The model could predict emergency visit or hospitalization with an AUC of 0.72 (95 % CI, 0.65 – 0.79), fall only with an AUC of 0.86 (95 % CI, 0.83 – 0.89), fall with hospitalization with an AUC of 0.89 (95 % CI, 0.85 – 0.92), and death with an AUC of 0.93 (95% CI, 0.88 – 0.97) when separating from no health event category.

Table 4.3 Mean FHV by Health Event Category.

Health Event Category	**FHV (n=899)** **Mean** (Std), Sample size
No health event	**0.69** (0.18), 497
Emergency visit or hospitalization only	**0.54** (0.18), 55
Fall only	**0.38** (0.07), 224
Fall with hospitalization	**0.34** (0.20), 92
Death*	**0.30** (0.16), 31

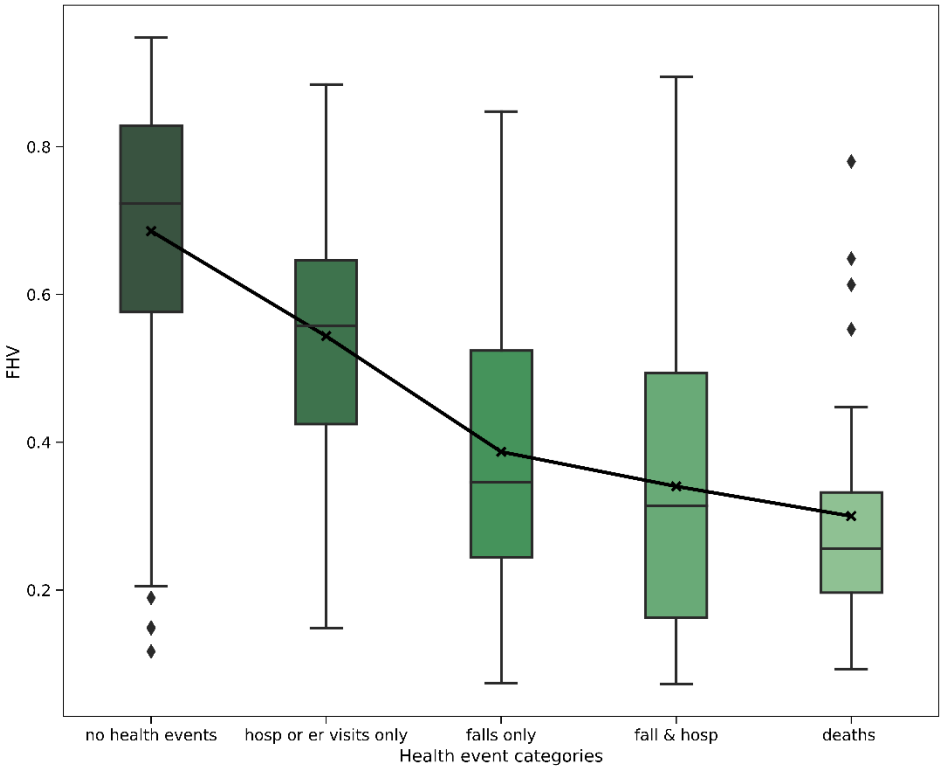


Figure 4.9. FHV vs health event categories.

The top and bottom of each box represent 75% and 25% percentiles of FHV for that category. Horizontal lines in each box represent the median FHV values for each category. The top of each whisker represents the maximum FHV in that category or median plus 1.5 times the interquartile range; the bottom whisker represents the minimum FHV in that category or median minus 1.5 times the interquartile range.

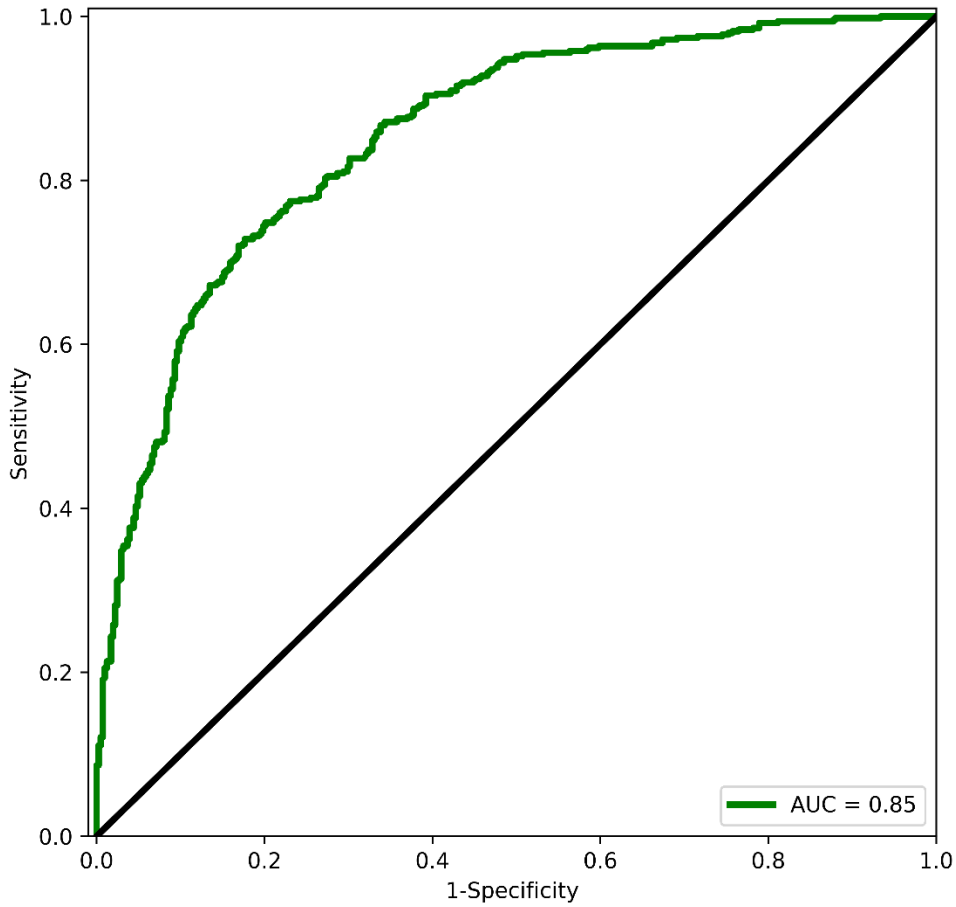


Figure 4.10. Receiver operating curve showing the separation of the no health event category from the rest (emergency visit/hospitalization, fall, fall and hospitalization, and death).

4.4.2.2 Six-month Fall

The average likelihood of fall within six months of computing the FHV is shown in Figure 4.11. We observed that higher FHV correspond to a lower average fall percentage and vice versa. An FHV score of 1.0 corresponds to ~ 0.0 fall percentage. The fall percentage almost linearly increased with a decrease in FHV.

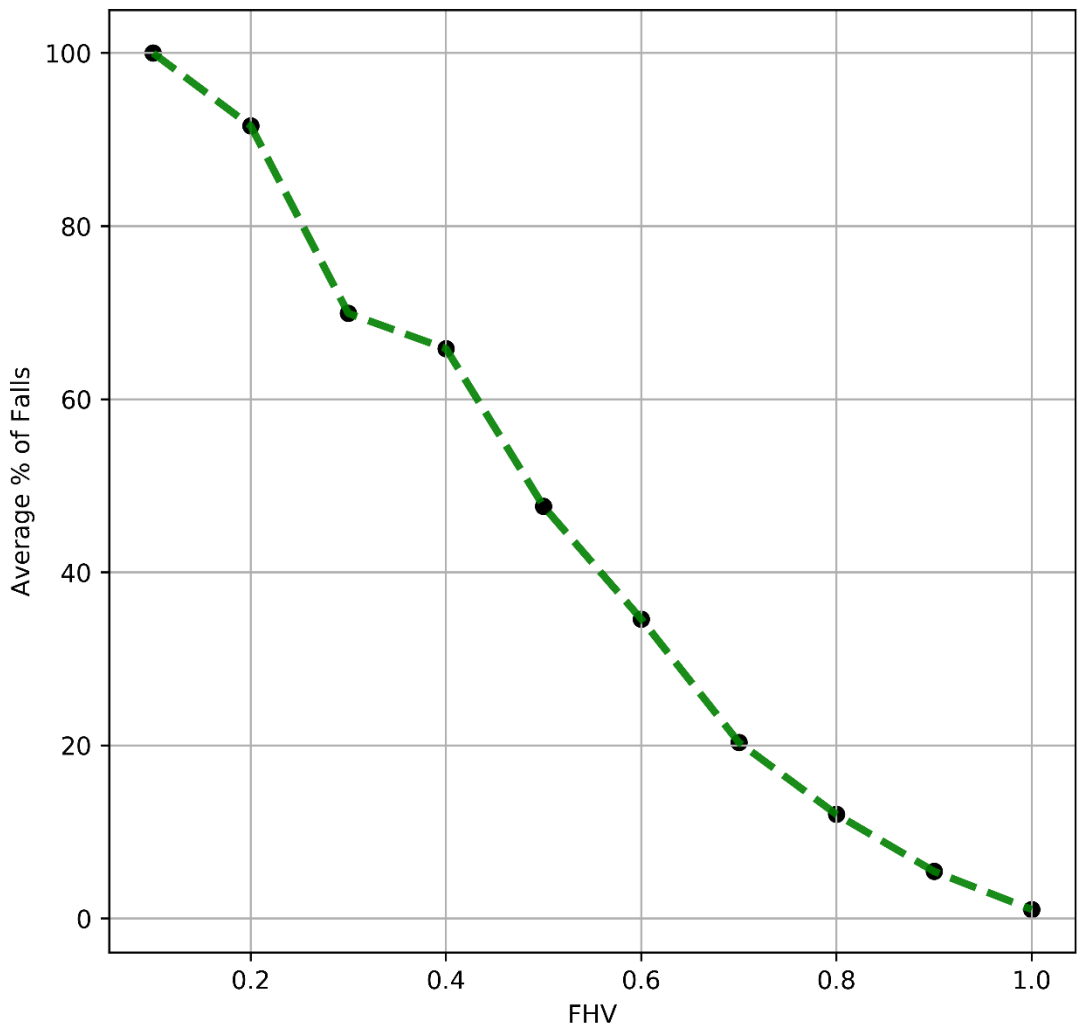


Figure 4.11. Average % of Six-Month Falls vs FHV.

4.4.2.3 Case Studies

We analyzed the same case studies as in the case of the excess risk model to demonstrate correspondence of changes in FHV with significant health changes reported in EMR nursing notes and the trajectories were very similar. We did observe a larger variation in the trajectories in the case of the FHV model. Fig 4.12 and 4.13 show the two trajectories.

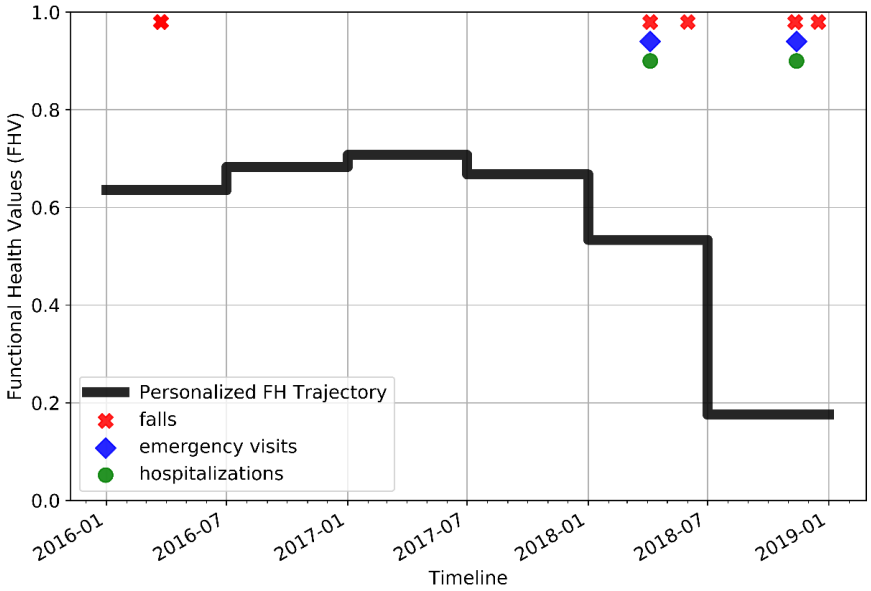


Figure 4.12. FH trajectory of a TigerPlace resident from Case Study – 1 using FHV model.

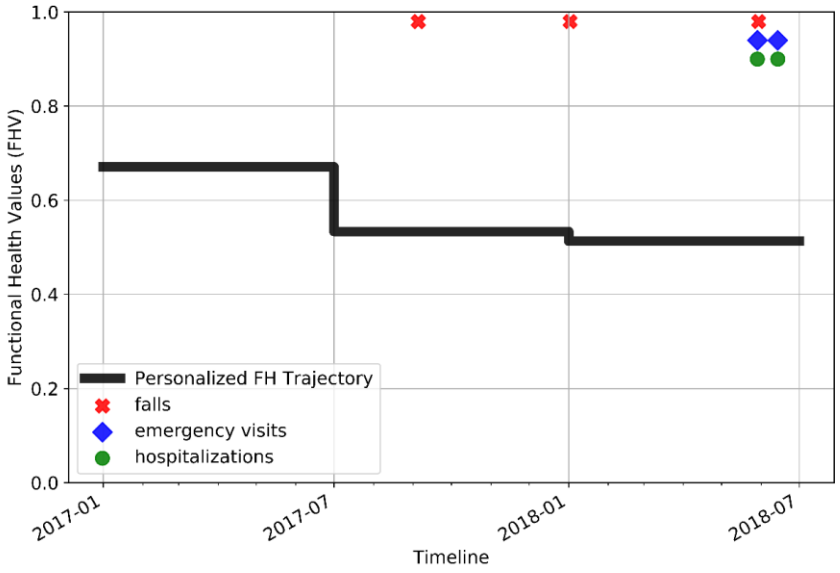


Figure 4.13. FH trajectory of a TigerPlace resident from Case Study – 2 using FHV model.

4.5 Discussion

We have developed two different FH prediction models to track continuous personalized FH using a set of geriatric assessments. However, we observed that the FH tracking model using mixed-effects logistic regression is more efficient when compared with the excess risk-based FH tracking model. Also, the mixed-effects model provides the capabilities to track personalized FH.

An FH trajectory is developed using longitudinal FHV predictions over time. The FH trajectory can be updated for a resident as new assessment scores are available, typically every six months. The model was developed by discriminating geriatric assessment scores associated with adverse health events, such as falls, emergency visits or hospitalizations, and mortality against scores associated with no adverse health events.

Results show that a rank order was observed in mean FHV, when moving from a lower health risk category, such as no health events to a higher health risk category, such as fall with hospitalization and death. This shows the generality of FHV. FHV can be interpreted to the effect that a higher value represents a healthier person. Results show that FHV of >= 0.6 corresponds to < 40% of six-month fall risk. The six-month fall risk percentage almost linearly increases with a decrease in FHV. Results also show that FHV < 0.4 could significantly increase the risk of falls, hospitalizations, and mortality. Case study analyses suggest that changes in FH trajectories are mostly gradual with some sudden drops. Sudden drops in FHV did correspond to significant health changes observed in the EMR. A lower FHV, specifically FHV < 0.4 throughout could suggest a high risk of falls and hospitalizations for the entire stay of the older adult. See Appendix A for more case studies and FH trajectory plots.

The high AUC values of 0.85 obtained for separating samples corresponding to no health events from rest suggests that a higher value of FHV represents a healthier FH state of the resident. FHV is not intended to predict a specific event. Instead, FH trajectories over a period can show the trend of FH changes for an individual. The case studies discussed above show that change in FHV may indicate a change in physical or cognitive FH. A decline in FHV below 0.6 may indicate a severe FH decline and interventions are needed to improve or maintain the FH of the resident. In the case studies, we observed that FHV below 0.6 was associated with an increased number of falls, hospitalizations, and emergency visits. In the case of the second case study, the FH trajectory shows that the FHV of the resident moved below 0.6 in the latter half of 2017. However, the person started experiencing adverse health events in early 2018. We believe that early interventions, specifically in the second case study may have helped the resident to possibly improve overall FH and avoid the following health events.

In [5], Rantz et al. conceptualized that the functional ability tends to decline unless timely interventions are provided. As we studied the FH trajectories for the TigerPlace residents, we found that the FH of an individual can decline as well as improve. We observed that for some residents, as they first start living at TigerPlace, their FH improved over a period. This could be because of the state-of-the-art care coordination provided at TigerPlace and other similar facilities. We observed this effect in the first case study. The predicted FH of the resident improved between 2016 – 2017 before it started to decline in the last half of 2017.

A limitation of this study is the study sample size. We had access to the data of only 150 senior residents from a single aging-in-place facility. We believe that data from a larger population, with more assessments and health events, could improve the generalizability of the model.

A second limitation is that we did not incorporate multimorbidity or age in our model. Previous studies have included chronic morbidity was as a number of chronic conditions to predict health outcomes [131]. While age and multimorbidity are associated with increased adverse health events, we were interested in evaluating the effectiveness of a composite score based on the routinely obtained geriatric assessments reported in the EMR in detecting health changes.

The use of mixed-effects modeling to predict adverse health events from repeated measurements from the residents helped us to use the entire longitudinal data obtained from the residents over the eight years. Also, using residents as random effects in model construction helped to personalize the model predictions. Therefore, even though we had a smaller population to work with, we could use thousands of measurements to build an effective model.

Personalized FH trajectories could equip healthcare providers at TigerPlace with early health risk indications and context about the changes in FH in the residents. In addition to providing a visual representation of the change in FH the model could also provide detailed information about what changes in the new assessments led to the change in FH. This could help care providers to decide on necessary targeted interventions faster.

4.6 Conclusions

FHI and FHV are general measures of functional health in older adults, computed using geriatric assessments electronically available in the EMR.

We developed and validated a model to track personalized functional health in older adults using multiple construct validators. We demonstrated that significant changes in the functional health trajectory could be early indicators of possible adverse health events. The FH trajectories could help caregivers decide appropriate interventions based on trends in overall functional health change. We propose that a larger dataset could be used in future studies to improve the model.

4.7 Code and Data Repository

The analysis code and metadata git repository link for the FHI project is https://vcs.missouri.edu/akmm94/THIL_codes_de_id_db

Chapter 5: 6-month Fall Risk Prediction using Geriatric Assessments and GaitRite Data

Abstract

Older adults age 65 and above are at higher risk of falls. Predicting fall risk early could provide caregivers enough time to provide interventions, which could reduce the risk, potentially avoiding a possible fall. In this paper, we present an analysis of 6-month fall risk prediction in older adults using geriatric assessments, GAITRite measurements, and fall history. The geriatric assessments included were Activities of Daily Living (ADL), Instrumental Activities of Daily Living (IADL), Mini-Mental State Examination (MMSE), Geriatric Depression Scale (GDS), and Short Form 12 (SF12). These geriatric assessments are collected by staff nurses regularly in senior care facilities. From the GAITRite assessments on the residents, we included the Functional Ambulatory Profile (FAP) scores and gait speed to predict fall risk. We used the SHAP (SHapley Additive exPlanations) approach to explain our model predictions to understand which predictor variables contributed to increase or decrease the fall risk for an individual prediction. In case of a high fall risk prediction, predictor variables that contributed the most to elevate the risk could be further examined by the health providers for precise health interventions.

We used LASSO and an ensemble of decision trees to perform a feature selection to understand the importance of the individual assessments, gait parameters, and fall history in predicting fall. We used the geriatric assessments, GAITRite measurements, and fall history data collected from 95 older adult residents (age = 86.04 ± 6.68, female = 59)

to train machine learning models to predict 6-month fall risk. Our models could predict a 6-month fall with an accuracy of 0.80 (95% CI of 0.78-0.83), AUC of 0.72 (95% CI of 0.68-0.75), Sensitivity of 0.79 (95% CI of 0.75-0.83), and Specificity of 0.68 (95% CI of 0.63-0.73). Our early prediction of fall risk could identify residents who are at higher fall-risk, which could potentially help care providers and family members to perform preventive actions.

5.1 Background

The number of Americans ages 65 and older is projected to be over 98 million by 2060, which is about 24 percent of the total population in the USA [124]. Studies show that more than one-third of older adults fall each year [149]. Out of these fallers, 20%-30% of the individuals suffer moderate to severe injuries, which reduces independence and mobility, and increases the risk of premature death [150]. Identifying older adults who are at higher risk of falls requiring interventions is challenging for clinicians [151].

Falls in older adults are multi-factorial [152]. Consequently, several fall risk assessment tools have been developed and validated [152]. Lusardi et al. have presented a systematic review and meta-analysis analyzing fall risk assessment tools [152]. In their analysis, they have included several self-report measures such as the Geriatric Depression Scale (GDS), Medical Outcomes Study Short Form (SF-36), and Mini-Mental State Evaluation (MMSE). Also, the study included medical history questions such as a history of previous falls and requiring any ADL assistance. The analysis shows that no single test/measure demonstrates a strong post-test probability in predicting fall. Deandrea et al. performed another systematic review to provide a comprehensive list of evidence-based risk factors for falls [153]. This analysis did not include SF-12 measures as a risk factor.

Oshiro et al. have used the predictors chosen by Deandrea et al from the Electronic Health Records (EHR) based on psychological and medical factors, medication use, and mobility or sensory factors to predict fall risk [154]. Results show that their final model had a positive predictive value of 8%, a negative predictive value of 98%, and an area under the curve of .74, with a sensitivity of 67% and specificity of 68%. One issue with these analyses is they use a crisp boundary in the range of scores for each assessment instead of using the entire distribution of an assessment to predict fall risk. A common analysis overlap in these two studies suggests that medical history questions, self-reported measures, performance and mobility-based measures are some of the most commonly used predictors to estimate fall risk in the literature. Therefore, in this study we estimated fall risk based on predictors from these three categories.

We hypothesize that a combination of geriatric assessments, containing physical, mental, depression, and fall history questions, along with gait parameters could provide a better prediction of fall risk than the individual assessments by themselves. In this study, we develop a data-driven fall risk prediction model using several different assessments, including Activities of Daily Living (ADL), Instrumental Activities of Daily Living (IADL), Mini-Mental State Examination (MMSE), Geriatric Depression Scale (GDS), and Short Form 12 (SF12) [31, 32, 35, 40, 155]. ADL are defined as activities that are essential for independent living [31]. IADL require a higher level of personal autonomy, referring to tasks that require enough capacity to make decisions through greater interaction with the environment [31]. MMSE is a widely used test to evaluate the cognitive aspects of mental function [155]. MMSE excludes questions concerning abnormal mental experiences and mood. GDS is a screening tool for measuring depression in older adults [32]. SF-12 is a

multipurpose short form that provides a generic measure of health status [35, 40]. SF-12 has a mental (MCS) and a physical (PCS) component. All these assessments have good reliability and validity and they represent different factors of health and wellbeing.

Gait characteristics have been used as fall risk indicators. Functional Ambulation Performance (FAP) score captures the gait capacity of an individual using a specific set of spatiotemporal parameters (STPS) [49, 50]. The FAP score is integrated with the GAITRite walkway (CIR System Inc; Clifton, New Jersey), a gold standard system for measuring STPS [51]. FAP has been validated in several independent studies [50]. GaitRite provides several other spatiotemporal parameters apart from FAP in its gait tests, including gait speed, step length, step time. Performing a pair-wise correlation analysis we found that most gait parameters were highly correlated (Pearson Correlation > 0.8) with either the FAP score or gait speed. Therefore, we chose to include only FAP and gait speed in our study as the only two gait parameters. We have also included falls in the last 6 months as a binary predictor with its value being 1 if there was a fall in the last 6 months, and 0 otherwise.

In addition to constructing a model to predict fall risk, we also used explainable AI techniques, specifically SHAP (SHapley Additive exPlanations) to explain our models and the specific predictions made by the models [156]. SHAP uses game theory to determine the individual contributions of the input features in predicting the outcome by a machine learning model. Lundberg et al. proposed SHAP values as a unified measure of feature importance. SHAP values attribute to each feature the change in the expected model prediction when conditioning on that feature. Considering the model has a base expected value that would be predicted if we did not know any features, SHAP values explain how

to get from the base value to the current output. The additive SHAP values for the individual features will either be positive or negative, hence increasing or decreasing the model prediction value starting from the expected base prediction value. SHAP can be used to provide a global explanation of a model by describing how the individual features have an overall effect on the model's predictions. SHAP can also be used to explain a particular model prediction. For example, fall prediction for an older adult using the model by providing feature importance of the individual features for that prediction. These feature importance values otherwise known as SHAP values could explain a model's prediction by suggesting which amongst the features had a larger contribution in that particular prediction. The fall risk model developed in this study depends on several aspects of functional health and mobility. An explanation to the fall risk predictions of individual older adults, providing more information about the predictors those had a higher contribution in increasing or decreasing the fall risk could provide critical clinical information for targeted interventions.

In this chapter, we present an analysis of fall risk prediction using geriatric assessments, gait parameters, and fall risk. We hypothesize that the assessment scores and fall history can be used to predict fall risk with good reliability and validity. We also hypothesize that a feature selection method, such as LASSO could provide the list of important features amongst all assessments for better classification accuracies. We include a description of the methods along with experiments and results.

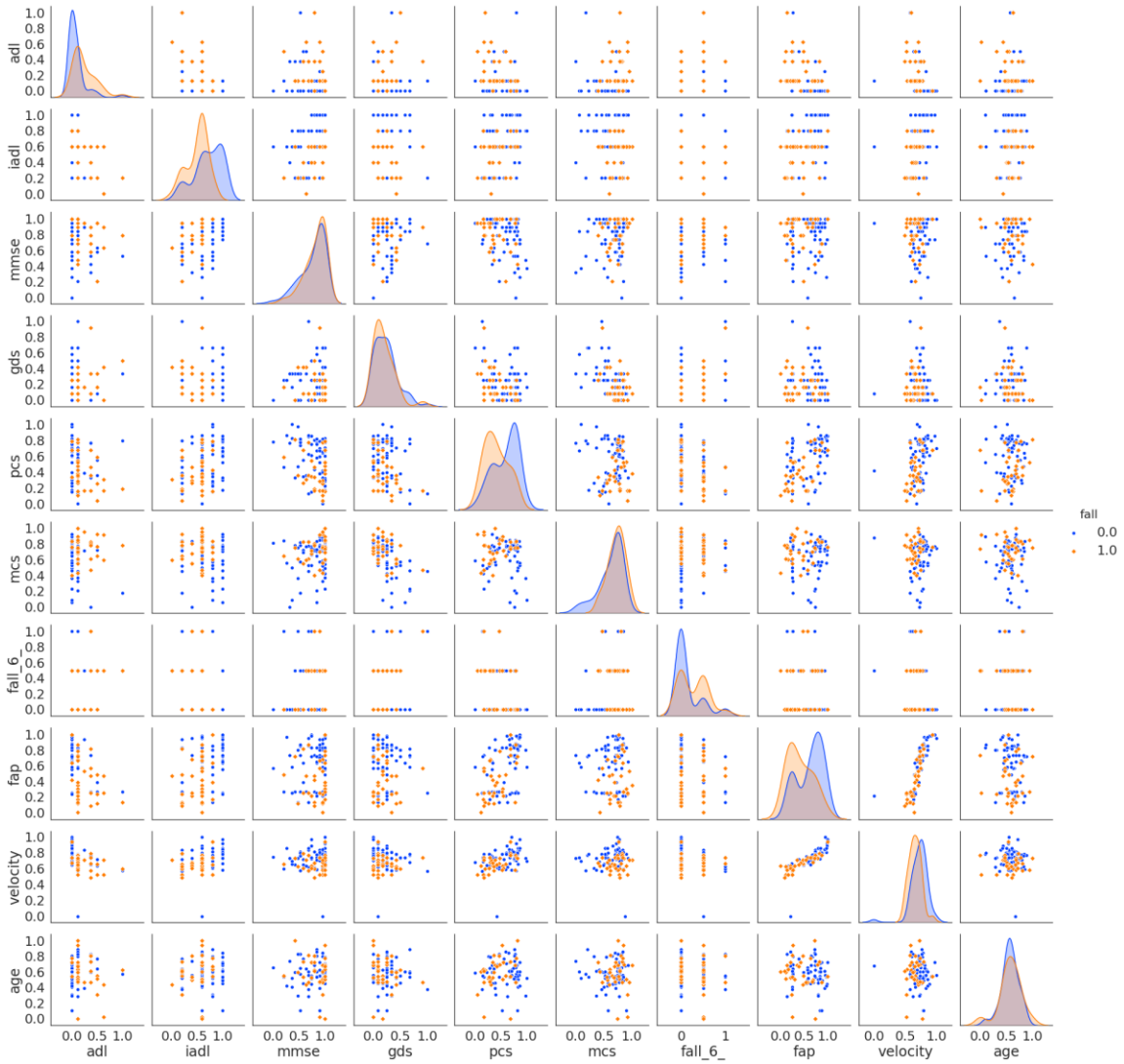


Figure 5.1. Pairwise relationships amongst (a) geriatric assessments and fall history to 6-month fall. Adl = activities of daily living, iadl = instrumental activities of daily living, mmse = mini-mental state examination, gds = geriatric depression scale, pcs = short form 12 physical component, mcs = short form 12 mental component, fall_6_ = fall in the last 6 months, fap = functional ambulation performance, velocity = gait speed, age = age during the assessment period, fall = fall in the next 6-months

5.2 Methods

5.2.1 Data

We used a set of frequently collected geriatric assessment scores in the Electronic Medical Record (EMR), such as ADL (Short Form ADL, RAI MDS 2.0), IADL (Lawton), GDS, MMSE, and SF12 [31, 32, 35, 40, 155]. The SF-12 assessment has two components, a physical component or PCS and a mental component or MCS. We used assessments collected at TigerPlace, an Aging-in-Place facility in Columbia, MO, on 93 independent living older adult residents (female = 58, age = 87.2 ± 7.0). In this study, only the first set of assessments collected on the residents were included to avoid repeated correlated measures from the same resident. We also included the gait speed and the FAP scores of TigerPlace residents collected using the GAITRite system over the same period. All assessments were conducted by the nursing and physical therapy staff at the TigerPlace and University of Missouri in Columbia, MO. Fall events reported by nursing and facility staff were used to develop the 6-month fall outcome and fall history binary predictor data. We only considered falls reported between the date of the first conducted assessment until six months. We did not consider demographic data, including gender into account in this analysis. Table 5.1 shows a summary of the characteristics of the predictor variables. Table 5.2 shows the fall history of the participants. This study received Institutional Review Board approval at the University of Missouri, Columbia.

Table 5.1 Data Characteristics

Variable *	**Non-Fallers (n=57)**	**Fallers (n=36)**
	Mean (Std)	**Mean (Std)**
ADL (0-16)	1.21 (1.94)	2.36 (2.82)
IADL (0-8)	4.29 (1.45)	3.22 (1.2)

MMSE (0-30)	24.50 (5.87)	24.06 (8.25)
GDS (0-15)	2.42 (2.27)	3.06 (2.77)
SF12 - PCS (0-100)	43.81 (11.16)	35.76 (10.91)
SF12 - MCS (0-100)	53.10 (9.11)	54.55 (7.54)
Age	87.27 (6.57)	87.06 (7.93)
FAP (40-100)	75.22 (17.83)	63.66 (15.13)
Gait Speed	71.15 (28.28)	52.22 (26.83)

* Interpretation of the variables - ADL, higher scores indicate more ADL impairment; IADL, lower scores show low function; MMSE, lower scores show more cognitive impairment; GDS, higher scores indicate depression; SF-12, low scores indicate a low level of mental or physical health; Fall History, 1 indicates one or more falls in the past 6 months and 0 indicates no falls in past; FAP, lower scores indicate poorer gait ability; Gait Speed, lower scores indicate poorer gait ability

Table 5.2 Fall History of Study Participants

Fall - Category	Past Falls = 0	Past Falls = 1	Past Falls = 2
Non-Fallers (n=57)	43	10	4
Fallers (n=36)	22	12	2

5.2.2 6-Month Fall Prediction

5.2.2.1 Data Preprocessing

Subjects containing Null values for any assessment were discarded. All assessments were standardized (Center to the mean and component-wise scale to unit variance). Multi-collinearity was determined using the Pearson correlation coefficient for the assessments. Performing a pair-wise correlation analysis we found that most gait parameters were highly correlated (Pearson Correlation > 0.8) with either the FAP score or gait speed. Therefore, we chose to include only FAP and gait speed in our study as the only two gait parameters. We have also included falls in the last 6 months as a binary

predictor with its value being 1 if there was a fall in the last 6 months, and 0 otherwise. None of the included predictor variables had a Pearson correlation greater than 0.7.

5.2.2.2 Feature Selection

We used the Least Absolute Shrinkage and Selection Operator (LASSO) and an ensemble of decision trees to understand the feature importance of our predictor variables [157-159].

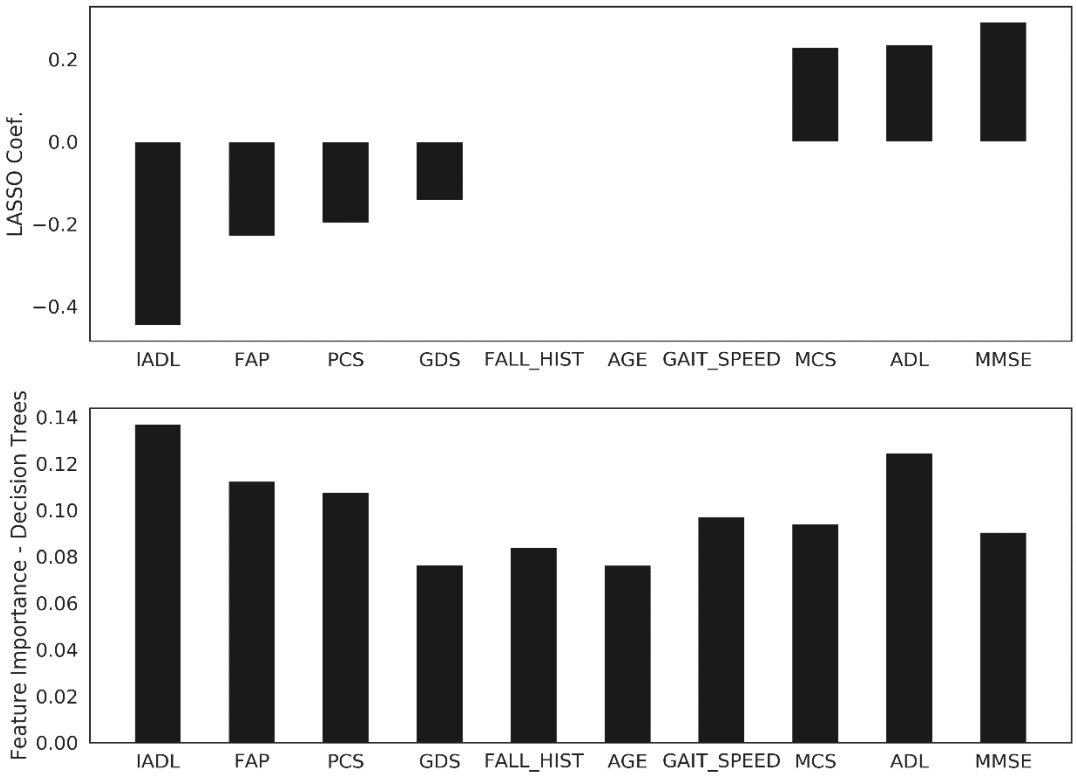


Figure 5.2. Feature Importance using LASSO and Randomized Decision Trees

Fig 5.2 shows the importance of the features and their importance level estimated by LASSO regression and an ensemble of decision trees. An analysis of the feature selection methods shows that age and fall in the previous six months were not estimated as important features in predicting future falls when compared to other features. Also, velocity

was not picked as an important feature by LASSO and gds was one of the least important features estimated by the randomized decision trees. For further evaluation, we ran classification experiments with and without including age and previous falls from our data set as they are chosen as the least important features by both feature selection methods.

5.2.2.3 Classification Experiments

We performed classification experiments to predict 6-month fall risk using all features and just LASSO features to see the performance of the classification models. We explored logistic regression, decision trees, and SVM models for the classification task. For decision trees and SVM we performed a hyperparameter grid search to find optimal parameters for the classification task. We performed five-fold cross-validation for the analysis. Results reported are the mean of the five-fold cross-validation performance measures. The classifiers were evaluated based on Area Under the Curve (AUC), validation accuracy, sensitivity, specificity, and F1 scores.

5.2.2.4 Explaining the Models and Individual Predictions Using SHAP

We used Shapley Additive exPlanations (SHAP) values to explain our models and predictions [156]. SHAP assigns an importance value to each feature for a particular prediction. SHAP values are additive. SHAP values for each feature provide an explanation about which features contributed to either increase or decrease the expected model output. We have included examples to show how SHAP values could show the reason for high fall risk predictions, for example, if one or more key features contributed to increasing fall risk in a particular prediction [156, 160].

5.3 Results

Table 5.3 shows the five-fold cross-validation performance measures of the different classifiers predicting 6-month fall risk. Overall, the SVM classifier using LASSO features performed the best with a sensitivity of 0.79 (95% CI of 0.75-0.83), Specificity of 0.68 (95% CI of 0.63-0.73), F1 score of 0.73 (95% CI of 0.69 – 0.76), and accuracy of 0.72 (95% CI of 0.68-0.75). Logistic regression performed like SVM when modeled with LASSO features. Logistic regression with LASSO features obtained a slightly higher Specificity of 0.70 (0.65 – 0.94) than SVM. We also observed that for all the observed performance matrices, SVM has a smaller 95% CI range as compared to Logistic regression. Decision trees did not perform as well as logistic regression or SVM.

Table 5.3 Classification Results in Predicting 6-month Fall Risk

Classifier	Sn (95% CI)	Sp (95% CI)	F1 (95% CI)	Acc (95% CI)	AUC (95% CI)
Logistic Regression	0.71 (0.62-0.80)	0.68 (0.64-0.73)	0.70 (0.67-0.73)	0.69 (0.66-0.73)	0.75 (0.70-0.81)
Logistic Regression (LASSO Features)	0.74 (0.64-0.84)	0.70 (0.65-0.74)	0.73 (0.69-0.76)	0.72 (0.68-0.76)	0.77 (0.71-0.82)
Decision Tree Classifier	0.62 (0.44-0.80)	0.70 (0.60-0.79)	0.67 (0.61-0.74)	0.67 (0.60-0.73)	0.69 (0.64-0.75)

SVM (kernel=linear)	0.79 (0.68-0.89)	0.67 (0.63-0.72)	0.72 (0.68-0.76)	0.71 (0.67-0.76)	0.78 (0.72-0.84)
SVM (kernel=linear) (LASSO Features)	0.79 (0.75-0.83)	0.68 (0.63-0.73)	0.73 (0.69-0.76)	0.72 (0.68-0.75)	0.80 (0.78-0.83)

5.3.1 Explaining the Models and Individual Predictions Using SHAP

We used SHAP to explain models and model outputs. Fig 5.3 shows the global explanation of a model, an SVM model with a linear kernel in this case. The plot shows how the feature importance values are distributed for each individual feature value range. For example, a lower value for velocity has a higher value of SHAP value and vice versa. Therefore, during a model prediction, a lower value of velocity could increase the fall risk prediction for this model. Which aligns with our general intuition. Figures 5.4 (a) and (b) provide explanations for two different fall risk predictions by the SVM model. Explanations to individual model predictions provide insights about which features increased the fall risk prediction of a resident. For example, a lower iadl increased the fall risk prediction for Resident 1, however, a higher iadl helped reduce the fall risk for Resident 2. The model had an expected base prediction value of 0.6.

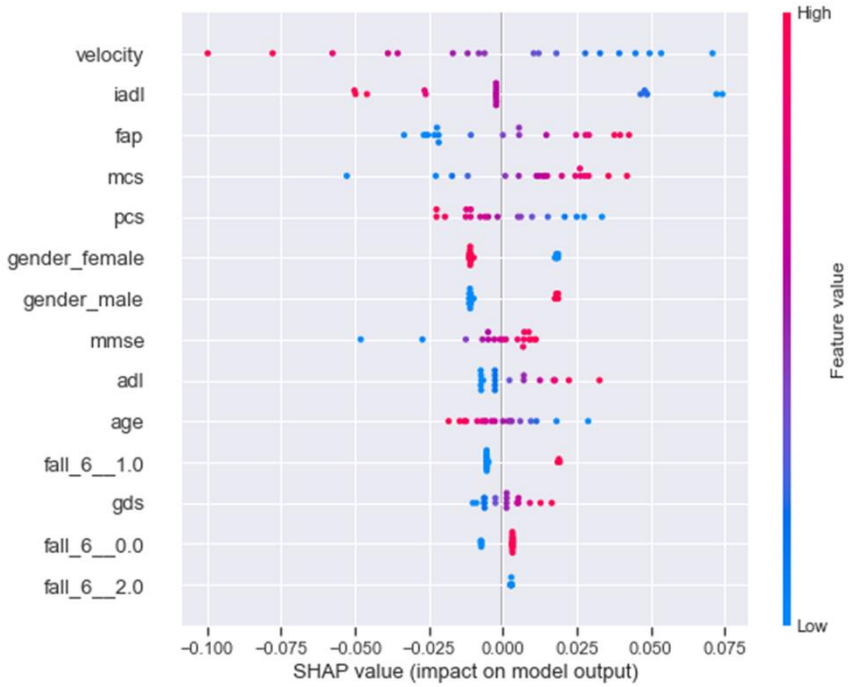
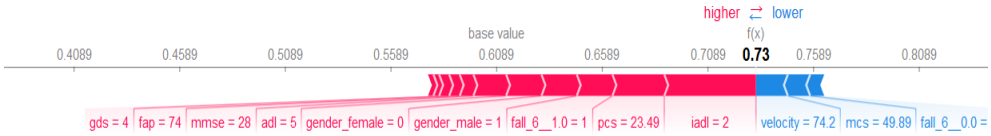
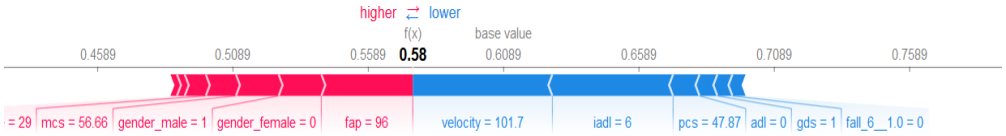


Figure 5.3. Global Explanation of a Model Using SHAP



(a) Model Prediction Explanation for Resident 1



(b) Model Prediction Explanation for Resident 2

Figure 5.4. Explaining an Individual Model Prediction for Two Different Subjects

5.4 Discussion

We observed that SVM performed superior to logistic regression in the classification task of predicting 6-month fall risk. Studies have shown that a history of previous falls significantly increases fall risk [152]. However, through repeated LASSO experiments we observed that in this cohort that was not the case. We further investigated the data to understand the impact of previous falls on new falls. Fig 5.5 shows all residents with their previous fall history and future falls in the next 6 months. If the resident did not have a fall in the past or the next six months a blank space is plotted without any marker.

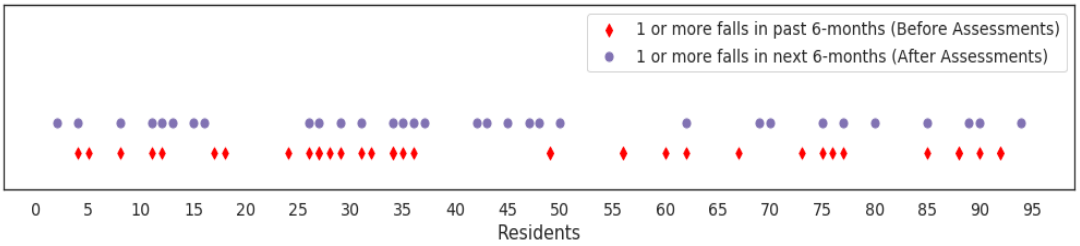


Figure 5.5. Previous and Future falls of the 95 Residents in the Data Set

The plot shows that percentage of new fallers who had a previous fall was 53%. This could be the very reason why previous falls did not have higher importance. Based on the feature importance analysis, iadl, adl, fap, and pcs had significantly higher importance than others. This signifies that physical function and performance played a greater role in predicting fall risk in this population.

We also observed that the age of the residents was not a critical feature in predicting fall risk. The population cohort consists of older adults with a mean age of 88. Which is

significantly larger than 65 that most studies are based on. Therefore, we believe that this analysis fills a literature gap of fall prediction in older adults with age > 80.

Out of the two best performing models, logistic regression would be more explainable than the SVM model. Therefore, in the case of a higher fall risk prediction, the logistic regression model developed using LASSO features can be used to understand which features caused that prediction.

Explanations to the model predictions using SHAP values provide additional insights about which predictor variables have increased fall risk prediction for an individual. Understanding which predictor variables are causing a higher fall risk could help health providers to provide personalized interventions to the residents. Therefore, the fall risk model could provide essential guidance to a health provider to focus on specific factors of fall risk instead of analyzing the individual assessments or predictor variables to understand their effects. For example, in Fig 5.4 (a) we observed that the fall risk for the individual was predicted to be 0.73 suggesting the resident had a relatively higher fall risk. The SHAP explanation to the individual prediction shows that the IADL assessment score, PCS assessment score, and a past fall were the three most prominent features that increased the fall risk. We can also observe that a slightly higher velocity (gait speed) was helping to reduce the fall risk. Similarly, in Fig. 5.4 (b) we can observe that the fall risk predicted by the model is 0.58, indicating the resident had a relatively lower fall risk at the time of assessments. Evaluation of the SHAP values for this individual prediction suggests that a significantly higher gait speed and IADL helped to reduce the fall risk for this resident. These critical and objective explanations could help clinicians save time and provide focused interventions to older adult residents with increased fall risk.

Future analysis with more predictor variables could potentially improve the results. Grip Strength, TUG scores, and medications could potentially be included to improve the prediction. Also, a longitudinal study with repeated measures from the individual subjects for recurrent events (fall and hospitalization) could provide improved and personalized risk assessment scores.

5.5 Conclusions

In summary, we have performed an analysis to predict 6-month fall risk amongst older adults using geriatric assessments, spatiotemporal gait parameters, and fall history. The fall risk models developed are explainable to critical information about which predictor variables were responsible for an increased risk of falls. This could potentially help clinicians to save time from analyzing individual assessments and provide early interventions to avoid a possible future fall.

Chapter 6: Retrospective Depth Video Classification to Reduce False Fall Alerts

6.1 Introduction

Falls remain one of the leading causes of injury in older adults. Millions of older adults age 65 and older fall each year [161]. One out of five falls causes serious injury-causing broken bones or head injury [162]. Each year, 3 million older adults are treated in emergency departments because of fall-related injuries. Also, the fall-related death rates in the U.S. have increased by 30% from 2007 to 2016 for older adults. At this rate, researchers anticipate that there will be 7 fall deaths every hour by 2030. Moreover, $50 billion in medical costs were due to falls in 2015 [163]. Falls are multifactorial [152]. A robust health monitoring system with fall prediction and real-time fall detection capabilities can help to inform family and clinicians, reduce fall-related deaths, serious injuries, and medical costs. Also, a fall monitoring system can help reduce the risk of prolonged periods of lying on the floor because of the inability to get up, especially for those who are living independently [164].

There have been several proposed fall detection systems including accelerometers, push-button systems, acoustic sensors, passive infrared sensors, video-based sensors, and other privacy-preserving techniques such as the use of silhouettes from depth sensors [165, 166]. Stone et al. have developed a fall detection system using depth image data [8]. The fall detection system operates in two stages. The first stage characterizes the foreground

object's vertical state in individual depth image frames and then segments on-ground events from the vertical state time series obtained by tracking the object over time. The second stage uses an ensemble of decision trees to detect if a fall preceded on the ground. The fall detection system is robust; however, it fails in recognizing false alarms in certain cases. The authors have mentioned that objects being dropped on the floor, pets jumping off the furniture, and visitors lying or sitting down on the floor could generate false alarms. The algorithm also does not detect is the foreground is a person or not. Therefore, even though the algorithm successfully detects most real falls it tends to generate a high number of false alerts causing alarm fatigue [17]. A large number of false alarms in healthcare systems are known to increase the burden on clinicians leading to the desensitization of alarms and patient safety issues [9, 21, 22].

In this study, we conducted several experiments to explore possible solutions to reduce false alarms in the depth image analysis-based fall detection system. We annotated several fall alerts, both true fall alerts and false alerts using a set of activities and used deep neural network architectures to perform the classification of the alerts. Alert classes with high accuracy in the classification experiments were pruned out to reduce false alarms. We also experimented with different fall alert thresholds to confirm that we do not miss a true fall alert. Results show that we can reduce the false alerts by 43 – 76 % by choosing different fall alert thresholds. A higher reduction in false alarms by increasing the thresholds also increases the risk of missing true falls.

6.2 Experiments

6.2.1 Fall Alert Dataset Creation, Annotation, and Augmentation

The fall detection system developed at the Center for Eldercare and Rehabilitation Technology, University of Missouri, Columbia has been deployed for production at TigerPlace senior living facility since 2014. To analyze, understand, and ultimately reduce the false fall alerts, we first developed a new fall alert dataset containing over ~4000 alert videos generated between 2016-2019. After collecting the fall alerts, we organized the alerts by annotating them.

We approached the data annotation by describing the activities seen in the alert video in a short statement. For example, instead of creating a generic label for all pet activities, we created 14 labels to describe the unique events in the alert videos with pets. The labels were created organically during the data labeling process. Once the labeling was complete, we ended up with 89 unique video labels for the activities in the entire dataset. Table 6.1 shows the 14 labels our annotators came up with while labeling alert videos containing pets. Along with the new data annotations, we merged the annotation used for alert videos in a previous study by Skubic et al [17]. Also, the annotators have defined a set of condensed descriptions for the 89 classes based on the key subject or activity in the alert videos. The condensed descriptions can be considered as superclasses of the 89 classes. The classes and condensed descriptions are presented in Appendix C.

Table 6.1 Example of Super and Sub-classes in the Dataset

Key Subject/Activity	Annotation Labels for Alert Videos with Pets
Pets	pet jumped off furniture
	dog laying down
	pet walking

	pet running
	pet running, knocked over object
	pet? Moving in corner
	pet walking near resident
	pet walking, non-resident bending over
	pet walking, resident bending over
	pet walking, multiple people in room, lots of activity
	resident pushed object, pet walking near resident
	pet walking, laundry thrown on ground
	pet playing on ground, multiple people in room
	non-resident pushing object on ground, pet walking

Table 6.1 shows the complexity of the classification problem. For instance, the alert videos with pets can contain several variations of pet activities and are significantly overlapped with other resident and non-resident activities, such as laundry thrown on the floor, objects pushed, and non-resident bending over. We have conducted several analyses and observed thousands of videos to understand the variations in these overlapped classes. Several strategies have been employed to effectively separate these classes. We performed several classification experiments by re-grouping and/or removing different sets of class labels from the dataset. The detailed analysis strategies are presented in the Experiments section.

We performed two sets of classification experiments: first with organizing all the data into three classes only, and in the second with 23 distinctive classes. The three-class classification experiment was performed to understand the predictability of neural network architectures in predicting fall versus other activities from the fall alert dataset. The three

classes considered were falls, pets, and all other activities. After a performance evaluation of the neural network models on the three-class classification, we performed an extensive analysis by observing videos from different classes and came up with 23 classes, each having at least 50 alert samples per class.

The original labeled alert dataset with 89 classes had five key fall class categories, including *fall forward, fall backward, fall side, fall chair,* and *other falls.* The fall directions in the fall classes represent the direction of falls with respect to the depth camera. We put all true fall videos in one class category because most fall classes did not have at least 50 samples. 20% of the alert dataset was reserved for testing and another 20% for validation. The rest of the data was used for training the models. The samples in the training set were flipped horizontally to augment the dataset. Also, we replaced each true fall alert sample in the training data with 20 random augmentations to further augment the dataset. We performed the data augmentations to increase the number of training samples, specifically provide more weight to fall samples during classification. Fig 6.1 shows an example of fall video augmentation. The set of frames in Fig 6.3 (a) were directly obtained from the original video. Fig 6.3 (a-1), (a-2), and (a-3) represent three unique augmentations using the frames in 6.3(a). We did not perform any data augmentation for samples in testing and validation datasets.

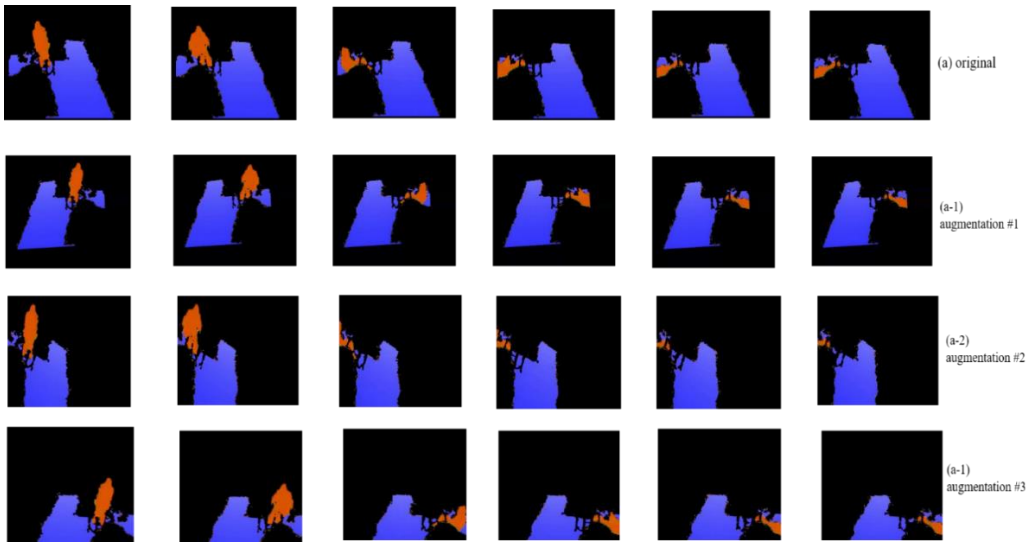


Figure 6.1. Fall Video Augmentation Example; (a) Represents the Actual Fall Frame Sequence

(a-1), (a-2), (a-3) Represent Random Augmentations Made to the Frames in (a)

The final 23-class dataset sample distribution per class is presented in Fig. 6.1.

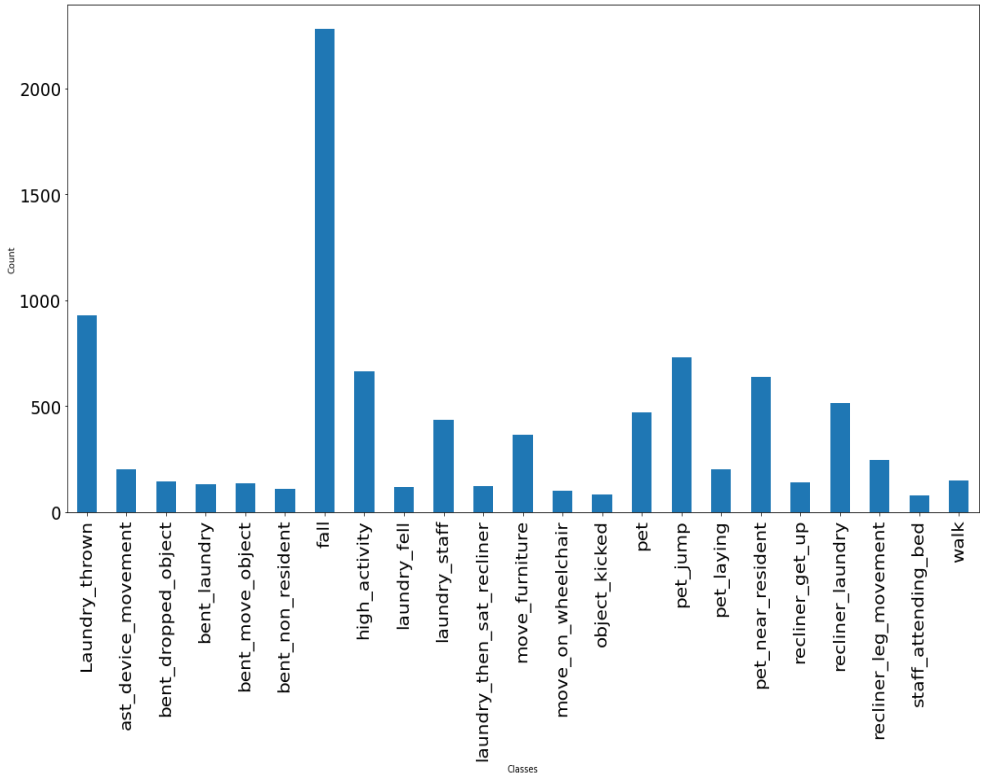


Figure 6.2. 23-Class Dataset Sample Distribution per Class

As the next step, the labeled alert videos were pre-processed to remove the unwanted background section of the frames, reshaped, and normalized the foreground objects for consistency.

6.2.2 Data Pre-processing

The preprocessing of the alert videos was done in three steps.

In the first step, we reshaped all the videos into 320 × 244. The fall detection project has gone through many revisions and hence a set of older alert videos were in different sizes. The most common video size observed was 320 × 244 and the current fall detection system uses this shape of image frames. So, to keep the videos consistent we converted the videos into this shape.

In the second step, we removed the background of the alert videos, which contain furniture, walls, and other non-moving objects. We removed the background using two HSV filters. Because of limited samples in our dataset with repeated alert videos from the same set of apartments, including background images could potentially introduce bias. For example, if an apartment has a pet that often jumps from the couch, by feeding tens of training samples with of the pet jump (in that fixed apartment setting), the neural network might memorize the background to predict pet jumps, instead of generalizing the actual event of pet jump. Therefore, in the case of an actual fall in that apartment setting, the model might favor predicting the fall as a pet jump. Removing the apartment background helps to generalize the model better assuring that the events have no specific connection with that apartment setting. The only video segments colored after the second step of preprocessing were the ground floor and the foreground. Figure 6.3 shows an example of before and after preprocessing alert video frames.

In the third step, we normalized the foreground segments. The depth data-based fall detection algorithm consistently colored the ground floor using shades of blue. The intensity of the blue color was higher closer to the depth camera as shown in Fig. 6.3. The foreground objects in the videos were segmented and colored as well. We observed that a moving object in the foreground may change its color within a small period. For example, Fig 6.3 shows a falling resident's body-colored in shades of green in the early stages of the fall, however, the body color changed to shades of orange in the later frames. We believe this could affect the performance of neural network models, specifically pre-trained models trained on RGB data. Therefore, we normalized the foreground objects such that these objects have a consistent range of colors throughout the video. Fig. 6.3 shows two frames with before and after pre-processing from an alert video of a true resident fall. The two frames show that the color of the segmented resident changed from green in (a-1) to orange in (a-2). After the second step of preprocessing, the foreground image segment of the resident was colored consistently as seen in Fig 6.1 (b-1) and (b-2).

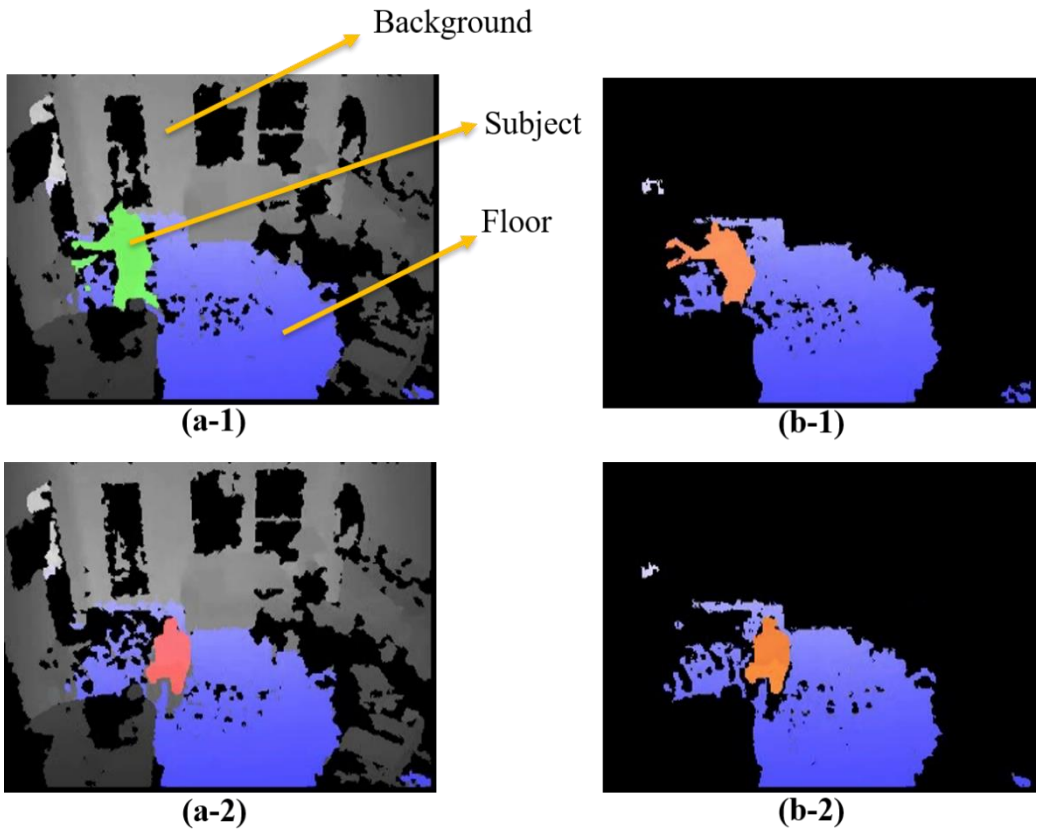


Figure 6.3. Example of fall video preprocessing; (a-1) and (a-2) represent two frames of a fall video before preprocessing, (b-1) and (b-2) represent the two frames of the video after preprocessing.

6.2.3 Classification Experiments

We performed several classification experiments to classify these video classes. To evaluate the effectiveness of deep learning architectures in discriminating falls versus other classes, we performed a classification test using only three classes. We divided the dataset into falls, alert videos involving pets, and all other false alert videos usually containing objects. We used the best performing model in this analysis for a more extensive analysis using alert videos from 23 different classes.

6.2.3.1 Three-class Classification

We performed the three-class classification as a baseline analysis to confirm that deep learning models can discriminate real fall videos from false alerts. The three classes were fall, pets, and objects. This experiment treats all objects equally, so objects such as laundry, pillow, furniture, and recliner were all considered as objects. Similarly, all pet activities described in Table 6.1 were considered in the pet class.

For this classification experiment, we used 3333 labeled alert videos. We considered 80% of all data for training and 20% of the data for testing. The alert videos were typically comprised of 75 to 200 frames. We resampled all the alert videos to 40 frames for the experiment. Also, we flipped all videos in the training data.

We used pre-trained Inception V3 models for feature extraction from each alert video [167]. Pre-trained Inception V3 is a powerful and efficient deep CNN architecture trained on the ImageNet object detection dataset. We used Inception V3 for obtaining the convolutional features from the alert videos as large variations in object location in the images would not matter much in this network. Inception V3 is also a relatively wider network and hence helps not to overfit. Multiple filters at the same level help different levels of filtration. We used the Inception V3 as a feature extractor. So, we removed the last fully connected layer of the Inception V3. The alert videos were reshaped to 299 X 299 X 3 from 320 X 244 X 3 to be able to use them as inputs of the Inception V3 model. Each of the 40 frames of a sample alert video was fed into the Inception V3 model and the outputs obtained from the last but one layer of the model were used to train another custom neural network architecture. The custom neural network models were created either using stacks of dense neural layers or LSTMs accompanied by a softmax layer for classification. Fig.

6.4 shows the classification architecture we used for this experiment. We used Adam optimizer and categorical cross-entropy for the training of the custom models.

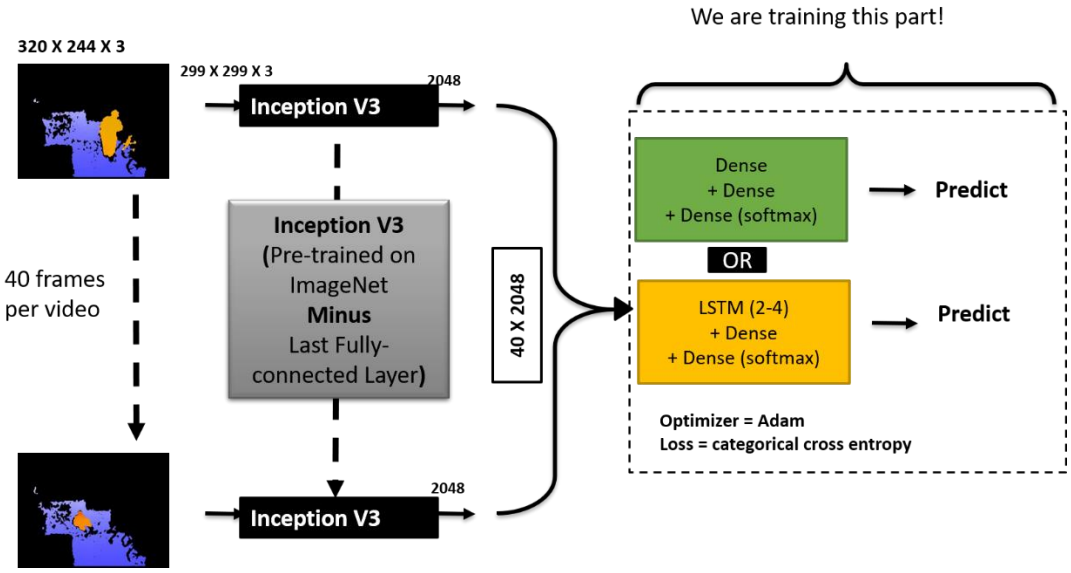


Figure 6.4. Classification Architecture

We received feature vectors of shape 40 × 2048 after passing the 40 frames of alert videos through a pre-trained Inception V3 model as shown in Fig. 6.4. We used these feature vectors to develop custom neural network models consisting of LSTM and CNN layers. We performed experiments with several combinations of LSTM and CNN modules, including two 512 dense modules, 1024 × 512 dense modules, two 1024 dense modules, two 2048 LSTM modules, three 2048 LSTM modules, and four 2048 LSTM modules. Finally, we added a fully connected layer with softmax activation for classification. We used Keras V2.2.0 and Tensorflow V1.9 for the experiments [168, 169].

6.2.3.2 23-class Classification Experiments With Data Augmentation

We ran another set of experiments by re-categorizing the alert dataset into 23 classes. The classes are listed in Table 6.2. The new classes were defined by carefully

observing the videos from the 89 classes. These 23 classes have at least 50 alerts and these classes capture ~98% of all the alert videos.

Table 6.2 List of class labels in the 23-class classification

Class Labels
0: laundry_thrown
1: ast_device_movement
2: bent_dropped_object
3: bent_laundry
4: bent_move_object
5: bent_non_resident
6: fall
7: high_activity
8: laundry_fell
9: laundry_staff
10: laundry_then_sat_recliner
11: move_furniture
12: move_on_wheelchair
13: object_kicked
14: pet
15: pet_jump
16: pet_laying
17: pet_near_resident
18: recliner_get_up
19: recliner_laundry
20: recliner_leg_movement
21: staff_attending_bed
22: walk

The final dataset contained 7010 training samples, 1078 validation samples, and 901 testing samples. From the three-class classification experiment, we observed that the architectures without LSTM modules overfit the training data. We also observed that models with LSTMs were stable, so we considered using LSTM based architectures to train our models in the 23-class classification task.

6.3 Results

6.3.1 Three-Class Classification Results

The results of the three-class classification are summarized in table 6.3. We observed that the test accuracies were comparable across the architectures. However, we found a significantly higher accuracy for training samples for dense models suggesting that they essentially overfit the training data. However, the LSTM models performed consistently and seemed to have a more stable overall performance. For LSTM models, F1 scores ranged from 0.785-0.802. The number of parameters and training time significantly increased with an increase in the number of LSTM layers. The 2-layer LSTM model used with Inception V3 features performed very similarly to other models, therefore, we considered this model for future experiments.

Table 6.3 Three class classification results for different architectures using Inception V3 features

architecture	test accuracy	test F1	test precision	test recall	train accuracy	train F1	train precision	train recall
2X512 Dense	0.762	0.756	0.806	0.76	0.866	0.864	0.893	0.87
1024X512 Dense	0.811	0.81	0.814	0.81	0.936	0.935	0.937	0.94
1024X1024 Dense	0.791	0.793	0.812	0.79	0.901	0.902	0.914	0.9
2LSTM	0.795	0.785	0.802	0.8	0.795	0.784	0.804	0.8
3LSTM	0.79	0.792	0.798	0.79	0.813	0.813	0.816	0.81
4LSTM	0.801	0.802	0.805	0.8	0.818	0.818	0.82	0.82

6.3.2 23-Class Classification Experiment Results

The dataset with 23-classes was trained using the 2 × LSTM models and the results for the validation data are presented in Fig. 6.5. The ROC curves for test data are presented in Fig 6.6.

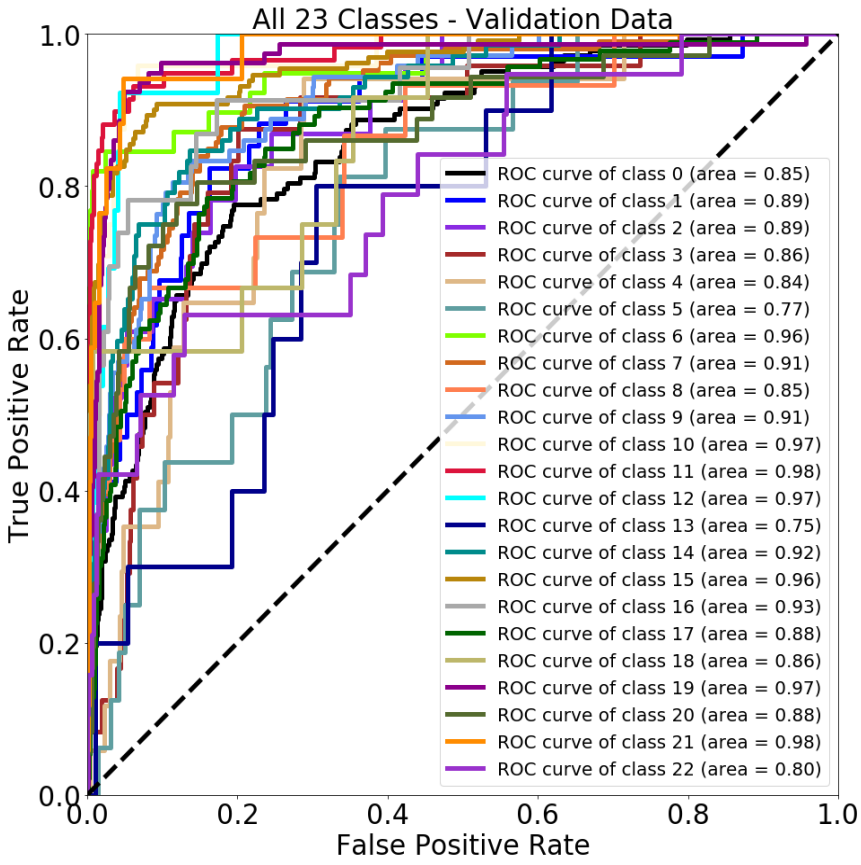


Figure 6.5. ROC Curves from 23-class classification on Validation Data

class labels: 0: Laundry_thrown, 1: ast_device_movement, 2: bent_dropped_object , 3: bent_laundry , 4: , ent_move_object, 5: bent_non_resident, 6: fall, 7: high_activity, 8: laundry_fell , 9: laundry_staff, 10: laundry_then_sat_recliner, 11: move_furniture, 12: move_on_wheelchair, 13: object_kicked, 14: pet, 15: pet_jump, 16: pet_laying, 17: pet_near_resident, 18: recliner_get_up, 19: recliner_laundry, 20: recliner_leg_movement, 21: staff_attending_bed, 22: walk

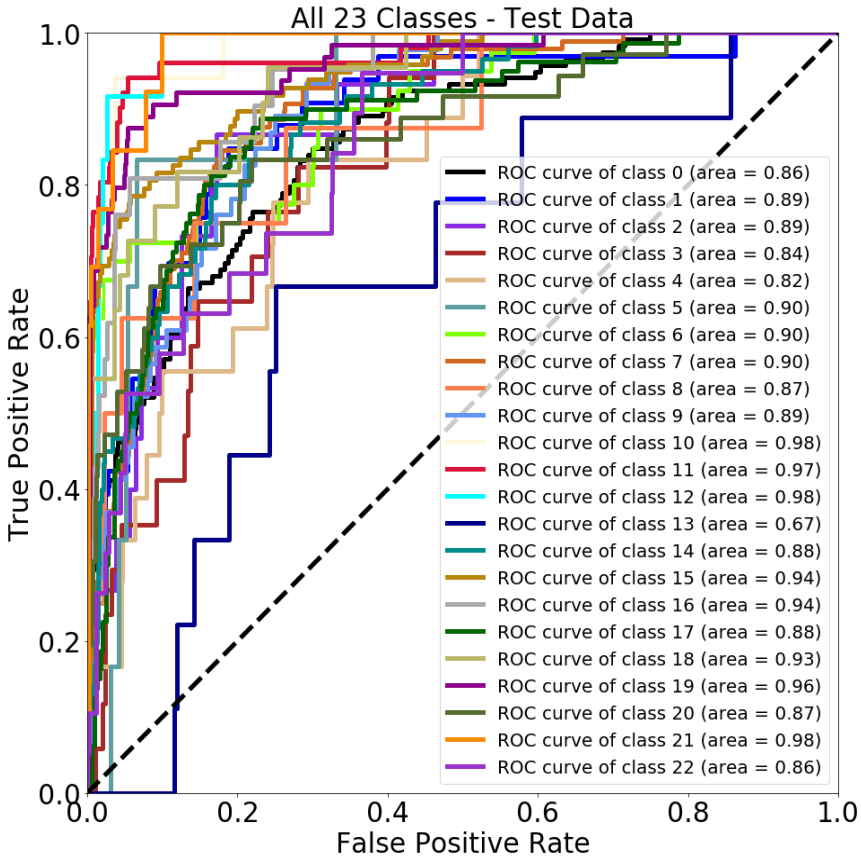


Figure 6.6. ROC Curves from 23-class classification on Test Data

class labels: 0: Laundry_thrown, 1: ast_device_movement, 2: bent_dropped_object , 3: bent_laundry , 4: , ent_move_object, 5: bent_non_resident, 6: fall, 7: high_activity, 8: laundry_fell , 9: laundry_staff, 10: laundry_then_sat_recliner, 11: move_furniture, 12: move_on_wheelchair, 13: object_kicked, 14: pet, 15: pet_jump, 16: pet_laying, 17: pet_near_resident, 18: recliner_get_up, 19: recliner_laundry, 20: recliner_leg_movement, 21: staff_attending_bed, 22: walk

Results show that several classes were well separated. We observed that the fall class had a high AUC in validation and test set, with values 0.96 and 0.90, respectively. Classes with laundry and recliner movement (10, 19), furniture movement (11), moving on a wheelchair (12), staff attending bed (21) had AUC values more than 0.95, suggesting those false alerts are highly separable. We observed that alert videos belonging to objects

kicked (13) had an AUC of 0.67 in the test data suggesting that the class was not very separable.

Fig. 6.9 shows the confusion matrix for the 23-class classification model.

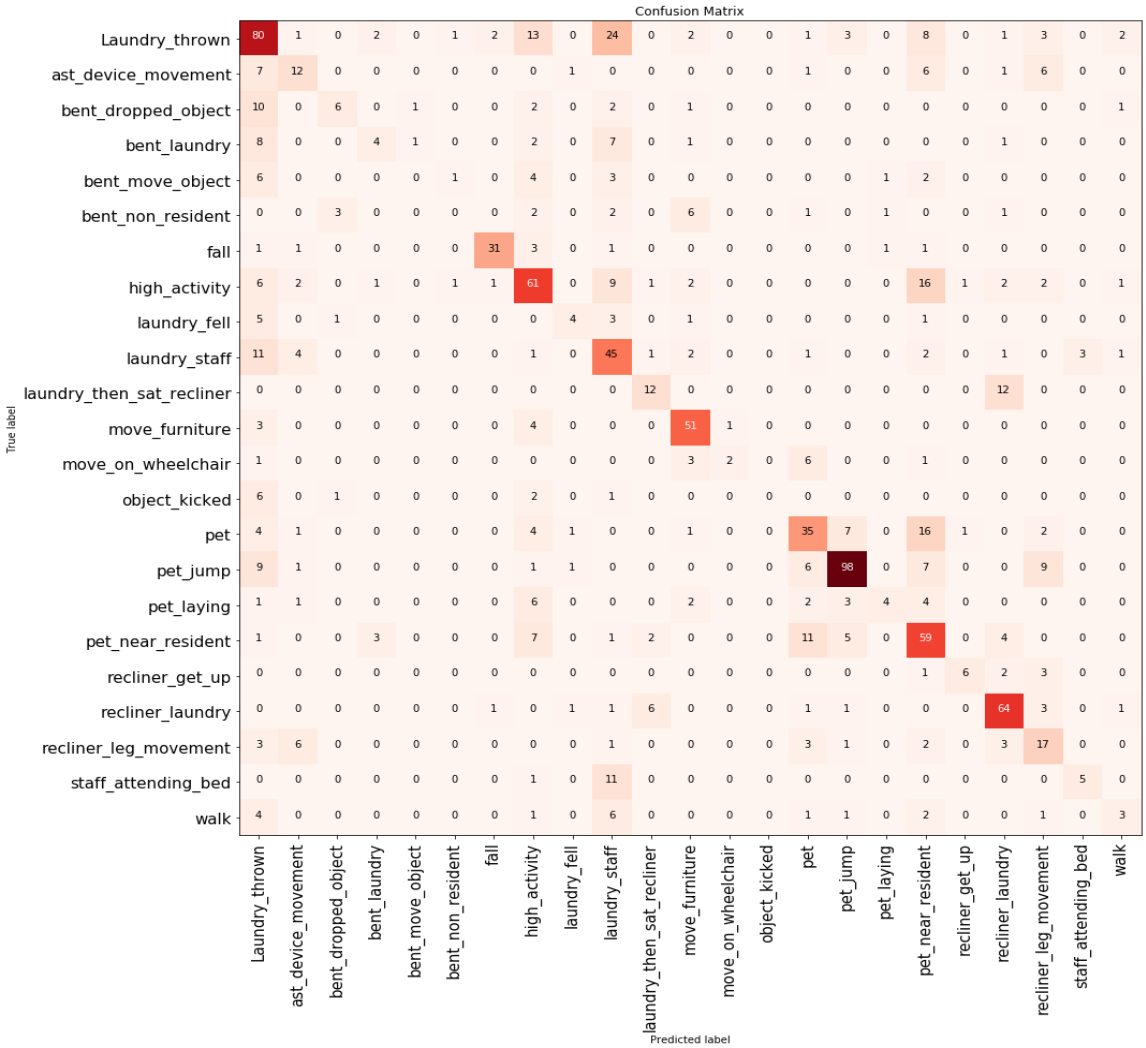


Figure 6.7. Confusion Matrix for 23-Class Classification

6.4 Discussion

The results from the two experiments show that the fall videos were well separated from all other activities with high sensitivity and specificity. We observed that most real

fall alert videos were classified as falls, except for a few being misclassified as *high activity, pet playing, assistive device movement,* or *laundry thrown.*

Several classes had significant overlaps with other classes. For example, *laundry staff* and *laundry thrown* have several overlapped characteristics, such as laundry thrown on the floor, and the possible presence of a staff member. Similarly, *laundry staff* also has overlapped characteristics with *staff attending bed*. We also observed a cluster in the confusion matrix near classes containing pets, such as *pet, pet jump,* and *pet playing.* This suggests that the pet classes were mostly misclassified as other pet classes. The confusion matrix presented in Fig 6.7 confirms that samples from overlapped classes were more likely to be misclassified. A possible solution to this problem is by combining overlapped classes as a single class. However, we kept them separate because of their unique characteristics. Moreover, accurately classifying the classes other than *fall* class is not necessarily a priority. We observed that samples from most false fall alert classes were not classified as *fall,* which is a priority.

We also observed that samples from class *pets near residents* were classified as a *high activity.* The high activity class usually represents a significantly higher number of foreground objects. With the presence of pets and residents or visitors, the amount of foreground activity is assumed to be high, and hence the miss classifications.

Classes with lower sample sizes were shown to have miss classified entirely, for example, *bent nonresident, bent move object,* and *object kicked*. We are constantly adding more samples to our training dataset. With an increased number of samples per class this problem could be resolved.

6.4.1 False Alarm Reduction Using Fall Prediction Thresholds

An alert can be considered as a false alert if the probability of the alert belonging to *fall*, class, is less than a predetermined threshold, θ. In this method, we heuristically defined several fall class probability thresholds such that the model ideally does not miss any actual fall. We analyzed this method using several fall prediction probability thresholds as presented in Table 6.4. The lowest threshold, θ_0 was obtained from the ROC curve separating the *fall* class from the rest of the 22 classes in the test dataset, when True Positive Rate (TPR) was equal to 1. Table 6.4 shows that the % of false alarms decreases with an increase in false alarm thresholds. Using θ_3 as the alert threshold would potentially reduce 76% of the false alarms, however, we might end up missing more true fall alerts with low probabilities.

Table 6.4 False Alarm % using different fall probability thresholds θ0 < θ1 < θ2 < θ3

Threshold	# of missed falls	% of false alarms
θ_0	0	0.57
θ_1	1	0.47
θ_2	1	0.33
θ_3	2	0.24

6.4.1.1 Running the Secondary Analysis in Real-Time to Reduce False Alerts

We made the secondary fall analysis algorithm run live for false fall alert reduction. During the test phase, the algorithm was able to reduce a significant number of false alerts. A reduction in the number of alerts for two different facilities are reported in Figures 6.8 and 6.9.

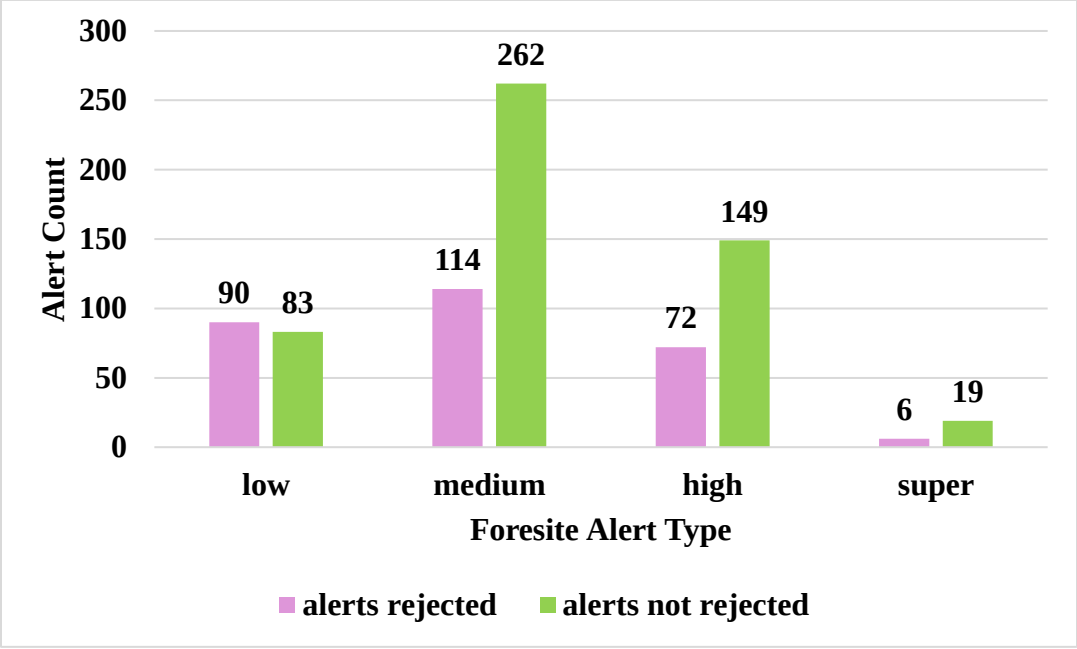


Figure 6.8. Adl THIL Facility - Alerts Summary After Secondary Processing

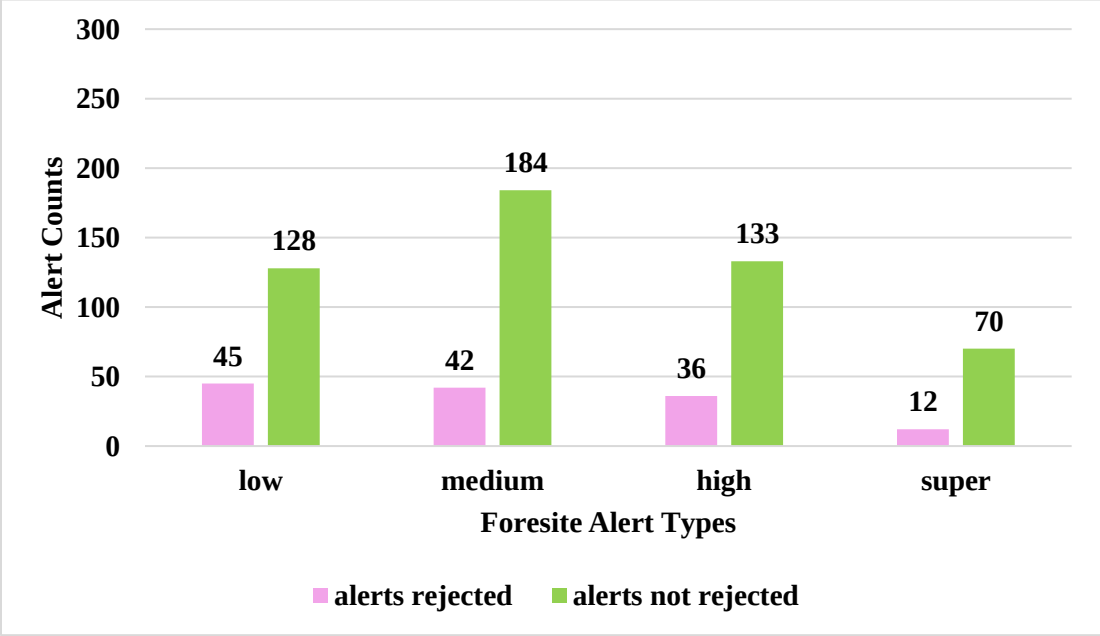


Figure 6.9. Adl TigerPlace Facility - Alerts Summary After Secondary Processing

The results presented in the above two tables are based on θ_0 as the primary goal of this analysis is to not miss any true alarms while reducing as many false alarms as possible. Alerts generated by the already-in-place fall detection system categorize the alerts into four categories: low, medium, high, and super, with low being the fall alerts with the lowest possibility of being a fall and vice-versa. Our model was used to filter high fall alerts. The alerts that were not rejected by the model were sent to the residents. All super alerts were sent to the residents and low or medium alerts were rejected. We observed that the model performed slightly better in the THIL facility by rejecting about 32% of the high alerts, however, it could only reject 21% for the TigerPlace facility. These numbers are still significant as these results are obtained by running the model only between 03/24/2020 till 07/22/20. A larger time range and a larger number of facilities could have an even larger impact in reducing false alerts.

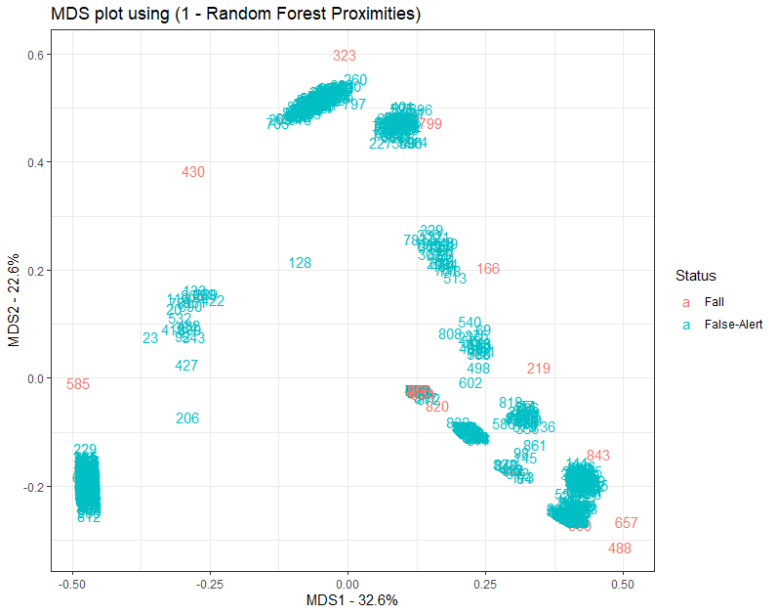
Currently, Foresite Healthcare, our commercial partner who licenses the fall detection system runs their fall detection system in several facilities. In their system, high alerts are monitored by human intervention. The secondary fall analysis could potentially be a step towards reducing human interventions for fall prediction.

6.4.2 False Alarm Identification Using Multi-Dimensional Scaling

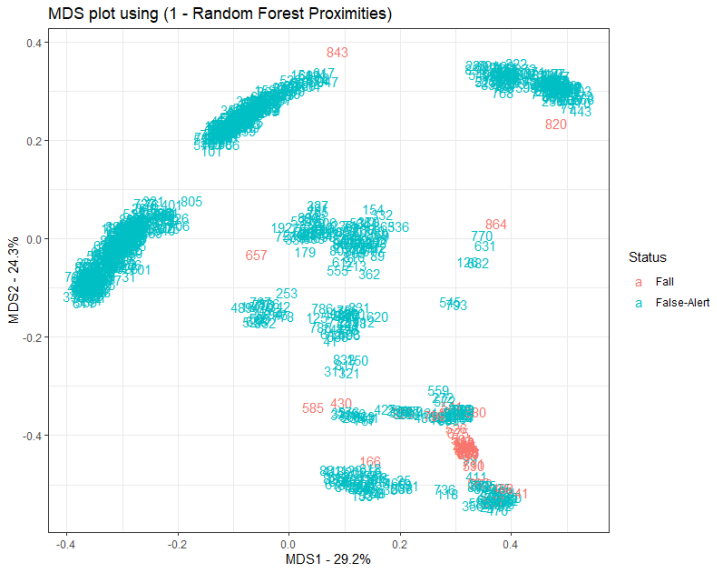
In addition to exploring the thresholding method to reduce false alarms, we also explored the use of top predicted classes to reduce false alarms. We performed this exploration analysis using the test dataset model predictions using the 23-class classification model. In the thresholding method, we only used the fall prediction probabilities to set the thresholds. However, in this study, we analyzed several groups of features to analyze if we could separate false alarms from true falls effectively. The groups

include (a) fall prediction probability only; (b) fall prediction probability with fall confidence (low, medium, high, or super), (c) fall prediction probability, fall confidence, and top predicted class; (d) fall prediction probability, fall confidence, and top three predicted classes; (e) fall prediction probability, fall confidence, and top five predicted classes.

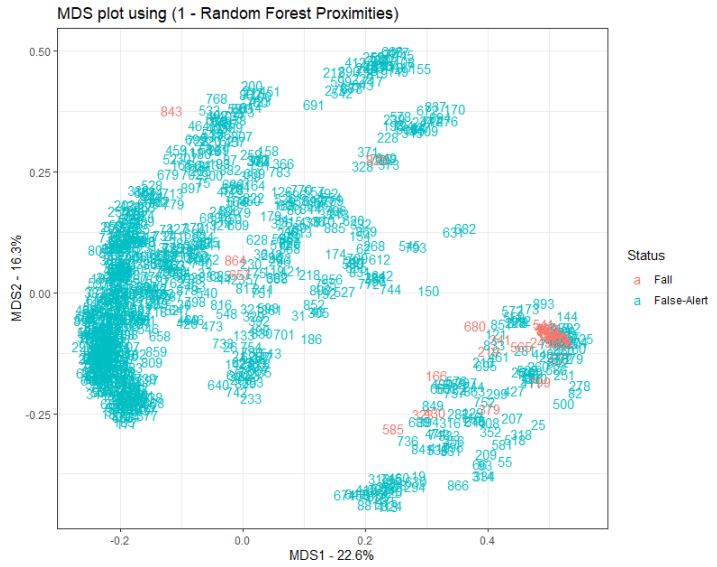
We developed five separate random forest models for each of the feature groups to detect false fall alerts. Proximity measures from the random forest models were then used to draw Multidimensional Scaling (MDS) plots with the test data samples to understand how the samples are related to each other [170]. MDS provides a means to find and visualize patterns or grouping of similar observations in data while preserving the relative distance between observations. In this analysis, we used proximity measures obtained from the trained random forest models to construct MDS plots. Fig 6.10 (a-e) show the MDS plots for the five different feature groups.



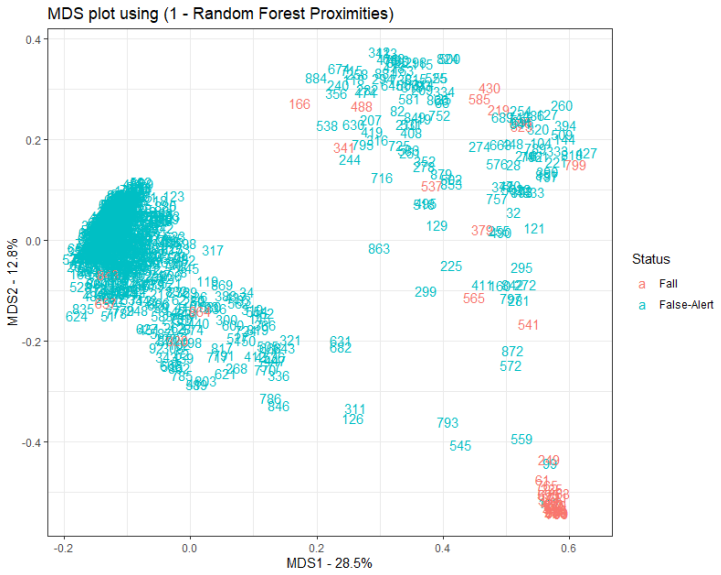
(a) MDS plot using fall prediction probabilities only



(b) MDS using fall prediction probability with fall confidence



(c) MDS plot using fall prediction probability, fall confidence, and top predicted class



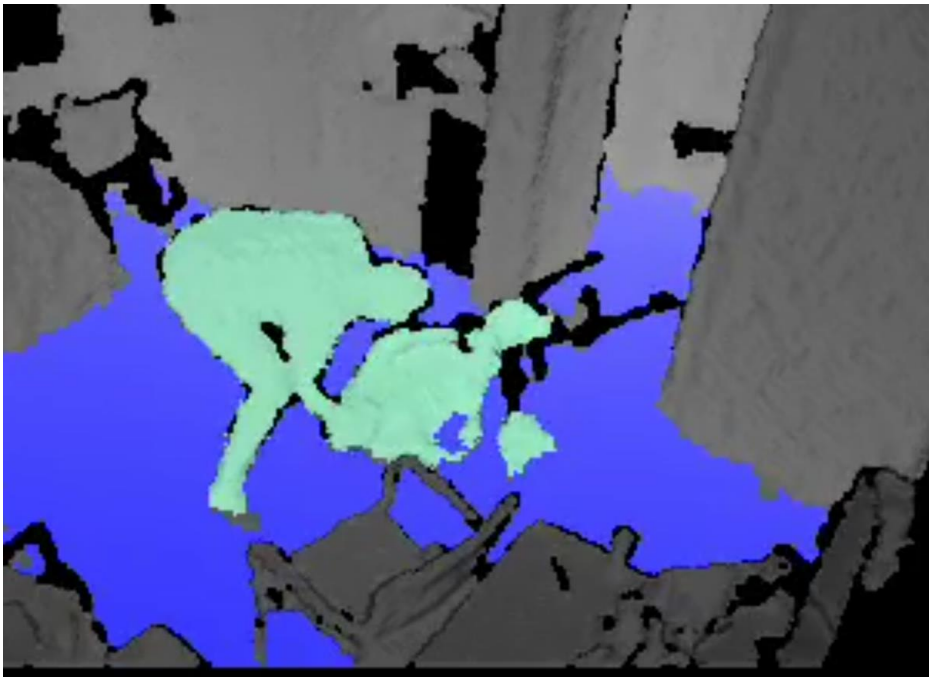
(d) MDS plot using fall prediction probability, fall confidence, and top three predicted classes

(e) MDS plot using fall prediction probability, fall confidence, and top five predicted classes

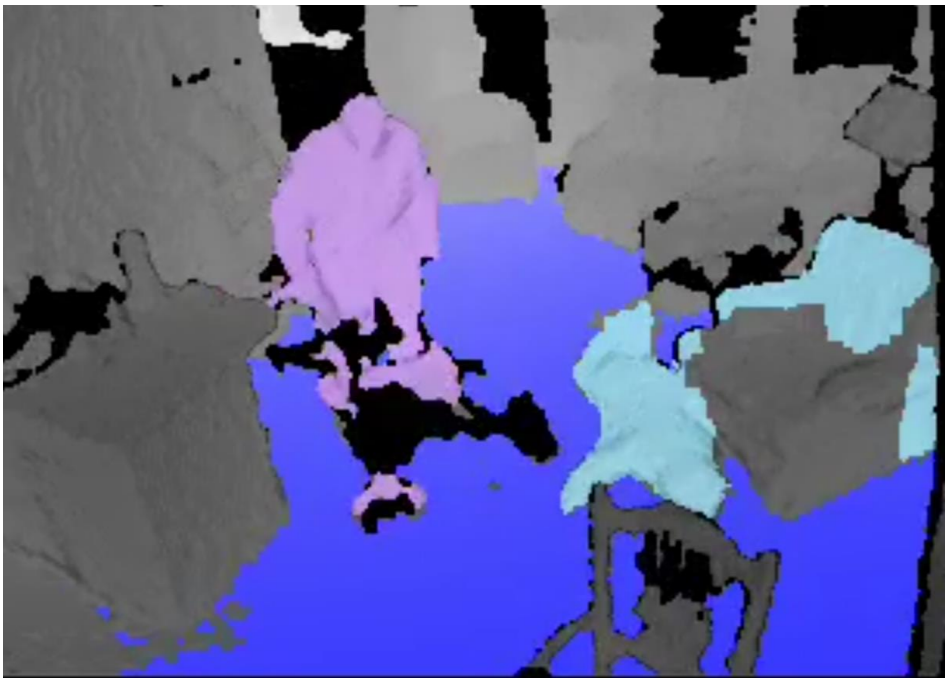
Figure 6.10. MDS plots using random forest proximities for different groups of features

We observed that all features could separate the true falls from false alarms to some degree. However, with the increased number of features, specifically in the group (d) and (e) we could see a significant separation between the true falls and false fall alerts. The final group (e) provided the best separation between true falls and false alarms.

We observed that four true falls in group (e) MDS plots (four falls in the left dense cluster of false alerts in Fig 6.10 (e)) were always grouped along with the false alarms. We further investigated these four alert videos to understand what characteristics in these falls made them closer to the false alarms. In two of these four fall videos, we observed the presence of more than one adult in the field of view of the depth camera. In both alert videos, one of the adults fell and the other adult responded to the fall by moving towards the falling resident. We further analyzed the five closest neighboring observations in the MDS plot to these fall videos. For one of the two of these fall videos containing two adults, the nearest alert video classes were 'bend_laundry', 'laundry_staff', 'pet near resident', and 'pet'. Fig 6.11 (a) shows the final frame in the fall video. In the second video with two adults in view, we observed that the falling adult fell over a couch and a chair. The nearest MDS plot neighbors for this fall belonged to classes: 'recliner_get_up', 'ast_device_movement', 'recliner_leg_movement', 'move_on_wheelchair', and 'pet_jump'. Fig 6.11 (b) shows the final frame of the second fall video with two adults. Observing the scene characteristics in these videos, we can understand why these fall videos were observed close to observations in the above-mentioned class categories in the MDS plot.



(a)



(b)

Figure 6.11. Last frame of the fall videos with two adults in view

Similarly, we also analyzed the other two fall alert videos that remained close to the dense false alarm cluster in 6.10 (d and e). We found that one of the fall videos suffered from significant occlusion. Fig 6.12 shows the last frame of this fall video. We observed that the resident fell in sight of the camera, but the resident was occluded by the bed while falling. Observations closer to this fall video in the MDS plot belonged to classes: 'Laundry_thrown', 'pet_near_resident', 'pet_jump', 'high_activity', and 'walk'.

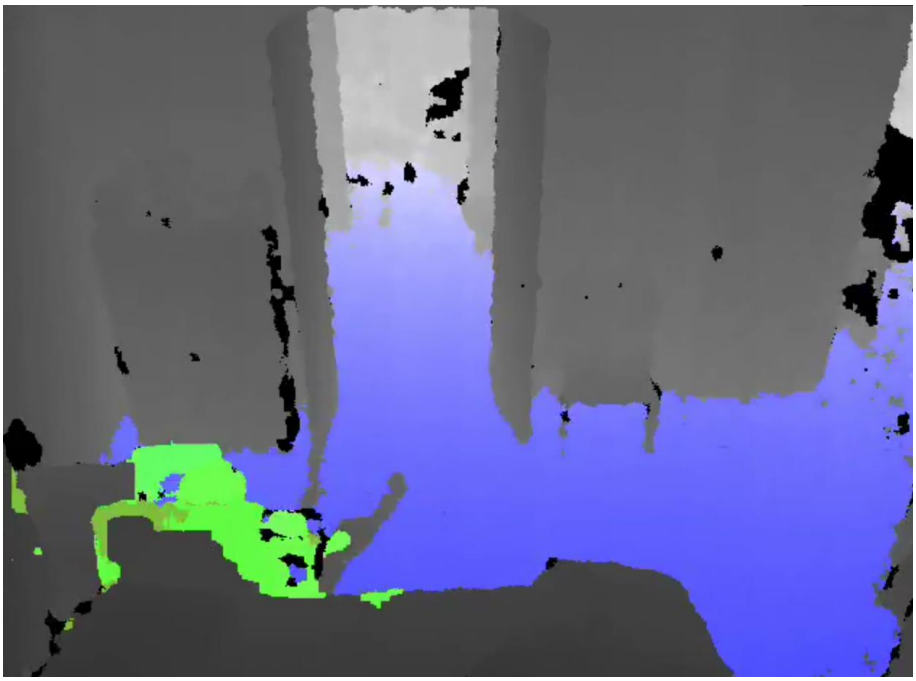


Figure 6.12. Last frame of the fall alert video with significant occlusion

Finally, we observed that in the last video of this category, the resident fell on a recliner, making the recliner a significant foreground. Fig. 6.13 shows the last frame of the fall video. We believe that the significant recliner foreground could have been the reason for the fall to be closer to some of the false fall alerts in the MDS plot. The observations

closer to this fall observation in the MDS plot belonged to classes: ‘bent_move_object’, ‘Laundry_thrown’, ‘recliner_leg_movement’, ‘high_activity’, and ‘bent_laundry’.

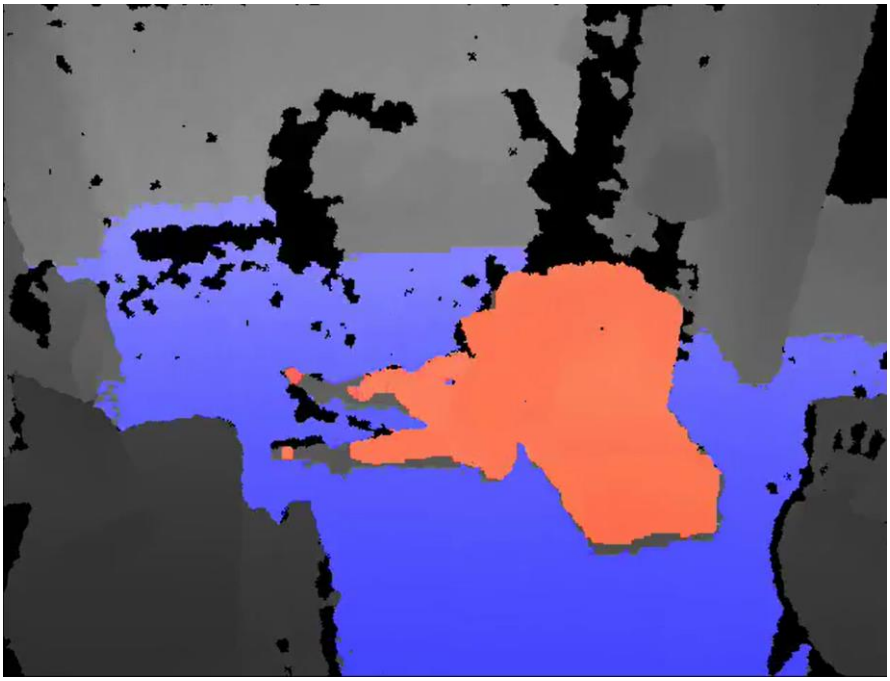


Figure 6.13. Last frame of fall video, resident falling on a recliner

Based on the clusters of the observations in the MDS plots, we believe that features including the top five predicted classes, fall confidence, and fall prediction probability could be used to reduce false alarms. However, we observed that four of the 40 videos in the test dataset were close to the false alarm observations. Analyzing these videos, we understood that these four fall videos had some unique characteristics that made them closer to false fall alerts in the MDS plots.

Using the MDS plot analysis we could potentially reduce false alarms by observing the nearest neighboring observations of the alert video sample in the plot. However, we

could still miss some true fall alerts. For example, if we choose all alerts observed in the left top dense false alarm cluster in Fig 6.10 (e) as false, we will miss four true falls. Therefore, we believe that even though the MDS method of reducing false alarms seems effective, we chose the thresholding method for production. We believe that our conservative thresholding method will less likely miss any true falls as compared to the MDS method.

6.5 Conclusion

Our analysis shows that the secondary fall analysis method could successfully reduce false fall alerts. The model flagged all fall alerts with a fall prediction probability below a fixed threshold as false. Increasing the threshold could reduce false alerts but in the cost of potentially missing true alerts. Reducing false fall alerts using this method could potentially reduce alarm fatigue in senior care facilities using the depth camera-based fall detection system.

Chapter 7: Fall Risk Prediction in Older Adults from Nursing Notes Using BioWordVec Word Embeddings

Abstract

Nursing notes in Electronic Health Records (EHR) contain critical health information, including fall risk factors. However, an exploration of fall risk prediction using nursing notes is not well examined. In this study, we explore deep learning architectures to predict fall risk in older adults using nursing notes in the EHR. We used EHR data including free-text nursing notes, medications, and observed falls from 162 older adults. We used BioWordVec to train multiple recurrent neural network-based models to predict future falls. Our final model predicted falls with a sensitivity of 0.75, a specificity of 0.83, and an F1 score of 0.82. LSTM-based deep neural models were most effective in predicting future falls. Also, the models generally performed better in predicting future falls in a shorter time range as compared to falls in the distant future. This exploratory analysis provides groundwork on the use of word embeddings in predicting fall risk from nursing notes.

7.1 Introduction

The number of Americans with age 65 and above is projected to be over 98 million by 2060, which is about one-fourth of the total population in the USA [124]. Also, older adults with age 65 and above are at higher risk of fall. Studies show that more than one-third of older adults fall each year [149]. Among these fallers, 20%-30% suffer moderate to severe injuries, which reduces their independence and mobility, and increases the risk

of premature death [150]. Identifying older adults who are at higher risk of fall requiring interventions is challenging for clinicians [151].

There have been several studies on predicting fall risk in older adults using history questions, self-report measures, performance-based measures, or a combination of measures [152]. However, there has not been enough exploration in predicting fall risk from free-text nursing notes [10]. Free text nursing notes, unlike structured EHR data, are less standardized, which makes them difficult to use in developing predictive models. However, studies show that these notes have critical information related to health prediction, including fall risk [10]. This provides an opportunity to explore the effectiveness of state-of-the-art natural language processing (NLP) methods to predict fall risk from nursing notes.

In this analysis, we explore word embeddings and recurrent neural network-based deep learning architectures to predict falls from free-text nursing notes, and medications. We hypothesize that free-text clinical notes in the EHR can be used to predict future falls in older adults.

7.2 Materials and Methods

7.2.1 Data

We included the free-text nursing notes and medications in the EHR collected from 162 older adults at TigerPlace, an Aging-in-Place facility in Columbia, MO [6]. The data were collected in the EHR by the nursing staff working at TigerPlace in collaboration with the Sinclair Nursing School at the University of Missouri, Columbia, MO. The free-text nursing notes included staff visit notes and progress notes. We de-identified the nursing notes data based on HIPAA regulations. A detailed description of the data de-identification

process is provided in the supplementary material. The fall events were reported by the TigerPlace staff in the EHR. Table 7.1 shows the number of records available in the EHR dataset.

Table 7.1. Records in TigerPlace EHR data

	# Visit Records*	***# Progress Notes*****	***# Medication Names******	***# Fall Events***
Total	247340	3456	2762	1704
Average Per Resident	1526.79	21.33	17.05	10.52

* Visit Records were descriptions written by nurses every time they visited each resident, which included medication reminders to changing sheets.

**Progress notes were notes from medical visits.

***Medication names were a list of medications the resident was on

7.2.2 Pre-processing

We found several common methods of preprocessing text data for natural language processing from previous studies. We lowercased the text [171, 172], removed stop words [172, 173] (words like ‘a’, ‘an’, ‘so’, ‘of’, which contain little to no contextual information), removed punctuation [171, 174], and tokenized the words. Words that commonly appear together in the clinical text were assigned to one token [171, 172, 174, 175]. Punctuations in the text were removed except for a few specific scenarios, such as ‘/’ in blood pressure measurements, and decimals in floating-point measurements. Finally, we standardized the blood pressure and blood sugar measurements to a format supported by

BioWordVec, which included rounding numbers to the nearest ten. Also, we removed all timestamps as they did not add meaning to our data. We organized all the clinical notes and medications in temporal order: data are organized by resident and by time.

7.2.3 Final Dataset

To avoid using correlated and repeated measurements, we only considered six months of data for each resident when they first started living at TigerPlace. We considered two months starting from the end date of their first six-month stay period to predict fall events. Only 27 of the 162 residents experienced a fall in the two-month prediction period.

Because smaller our dataset was relatively small, we decided to use pre-trained word embeddings rather than developing our own. The table shows the characteristics of the dataset.

This study received Institutional Review Board approval at the University of Missouri, Columbia.

7.3 Model Construction

We evaluated two different pre-trained word embedding models, BioWordVec and GloVe to understand if one of them is more effective in capturing the words found in the dataset [176, 177]. We observed that BioWordVec was able to capture more clinically relevant terms when compared with GloVe. Hence, we considered BioWordVec as our preferred word embedding model to construct the prediction models. We have provided more details about this analysis in our supplementary document. Appendix D provides a more in-depth comparison of the performance of the word-embedding models.

We experimented with several recurrent neural network models with Long Short-Term Memory (LSTM) units and Gated Recurrent Units (GRU). Because of the small

dataset, we constructed simpler models to not overfit our data. The three models we experimented with are shown in Fig 7.1. To test each model, we ran ten repetitions of fivefold cross-validation. We recorded the mean values of each model's accuracy, sensitivity, specificity, F1 score, and area under the receiver operating curve (AUROC).

Because our data set is unbalanced, we also utilized class weights during training so that the fall class would have more effect on models than the non-fall class. After running a parameter search, we determined that the best class weight for the fall class was 1.95 while the non-fall class remained at 1.

7.4 Results

Table 7.2 shows the mean accuracy, sensitivity, specificity, and F1 scores of 10 repeated tests of 5-fold cross-validation for each model. Representations of each model are shown in Figures 4, 5, & 6. With these models, we were able to achieve F1 scores as high as .82, with both an LSTM and GRU architecture, with the LSTM achieving higher sensitivity, despite the GRU having higher accuracy and specificity.

Table 7.2. Classification results

Model	***Fall Class Weight***	***Accuracy***	***Sensitivity***	***Specificity***	***F1-Score***
LSTM 1	1.95	.81	.75	.83	**.82**
LSTM 2	1.95	.72	**.90**	.66	.73
GRU 1	2	.76	.79	.76	.79
GRU 1	1.95	**.83**	.63	**.88**	**.82**

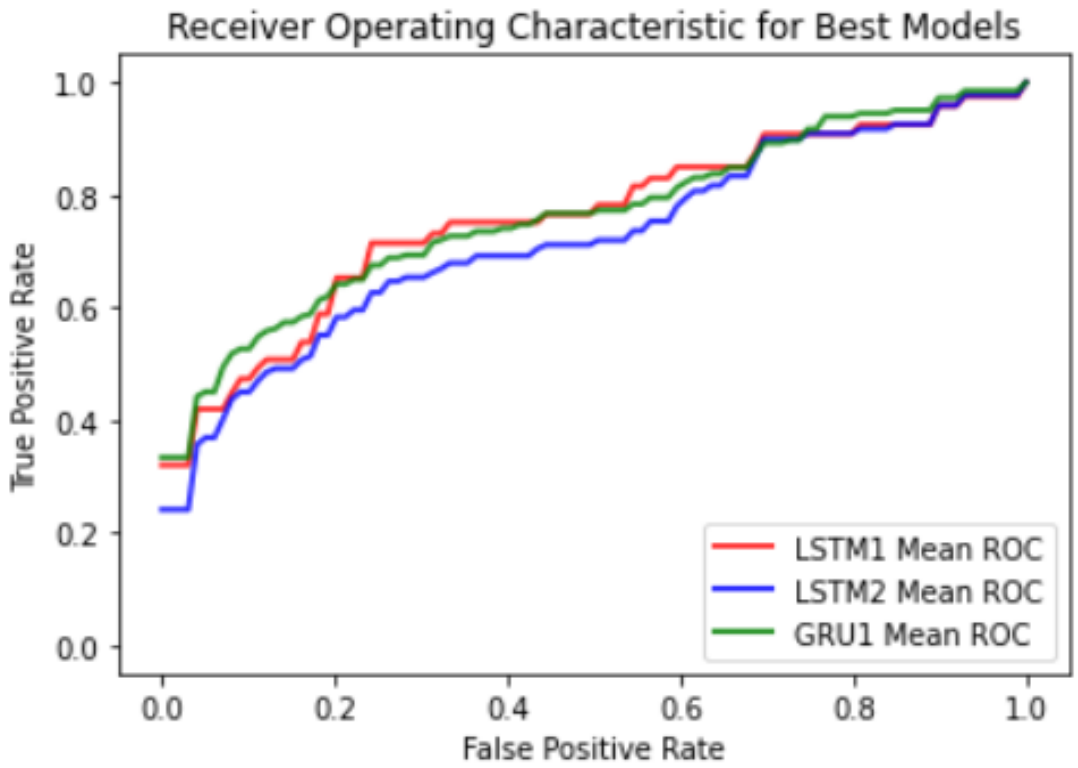
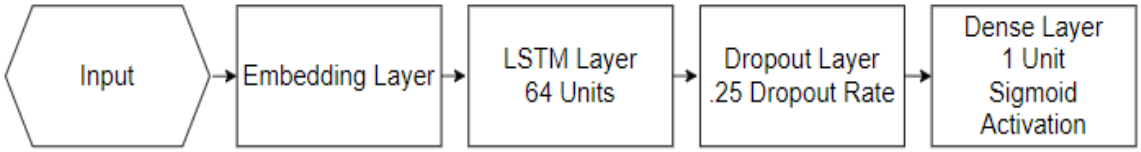
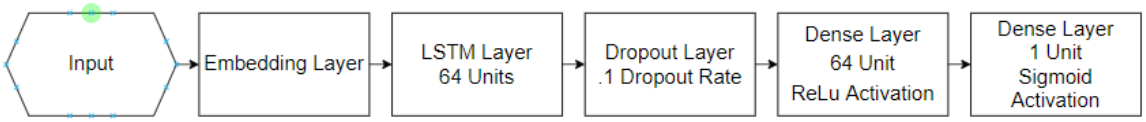


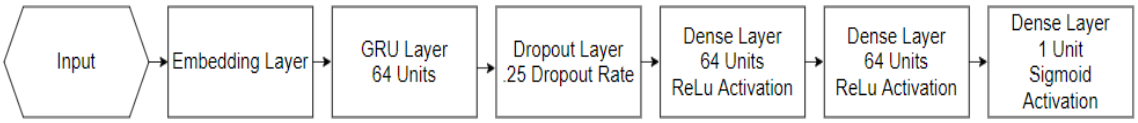
Figure 7.1. ROC curves for classification



(A)



(B)



(C)

Figure 7.2. LSTM and GRU models: (A) LSTM 1, (B) LSTM 2, AND (C) GRU 1

7.5 Discussion

Our results support the idea that fall risk can be determined from free-text clinical notes. With the size of our data set, we were only able to test relatively simple models on our data, however, even with that limitation, we were able to yield results that demonstrate that fall risk can be determined from EHR free-text data.

We consider LSTM 1 to be the best performing model, because while it does not have as high an accuracy or specificity as GRU 1, it has a higher sensitivity, which is important in a situation where you want as many of the true positives identified as possible, additionally, the two models have the same F1-score of .82. This suggests that LSTMs would be appropriate to use in further research, especially when using a larger dataset that allows for more complicated architectures than our simple LSTM models.

The size of our dataset may be a limitation in this study. We try and mitigate this by using 5-Fold cross-validation and using early stoppers within our training to prevent overfitting. Future research into free text EHR data should include larger sets than our 162-person set. Another aspect that may have limited us was the fact that BioWordVec did not catch all the words in our datasets, therefore it may be more useful to create a word embedding that could catch the entire vocabulary.

7.6 Conclusion

We have demonstrated that fall risk can be determined from nursing notes. Hopefully, we have set the groundwork for research into this topic, and that similar research involving EHR data will be conducted in the future, especially with larger datasets. Future research should look into what machine learning architectures are best for analyzing nursing notes, and work should be done in incorporating fall risk predictions into health alert systems to decrease false positives and help mitigate alert fatigue in health care workers.

Chapter 8: Functional Health Prediction Using In-Home Sensor Data

Abstract

With the increasing computational capabilities and advances in internet of things technologies, unobtrusive in-home sensors have improved continuous health monitoring. These health monitoring systems have been shown to predict fall, track gait, sleep, and overall activity successfully in the homes of community-dwelling older adults. In addition to tracking these objective and clinically relevant health measurements, functional health tracking could provide information about the overall well-being of older adults. Higher functional health has been shown to increase the quality of life and successful aging in place. Functional health is traditionally measured using questionaries and self-evaluation surveys that generally involve subjectivity and are conducted less frequently. Predicting functional health from the in-home sensor data could provide continuous functional health monitoring and hence early detection of functional health deterioration. In this study, we explore developing accurate functional health prediction models using in-home sensor data.

8.1 Background

Research has shown that higher functional health leads to a higher quality of life and improved aging in place [126]. Functional health is complex and multi-factorial. Tracking functional health has been traditionally conducted in clinical settings under the

supervision of trained clinicians. Based on the 2015 WHO world report on aging and health, functional health consists of physical, cognitive, and social function [128]. Several screening tests such as ADL, IADL, MMSE, and GDS provide different aspects of functional health. In Chapter 4, we used these manually collected assessment measures to develop a composite functional health score.

With the technological advancements, functional health assessments have been automatized in several ways, such as computer-assisted assessments, gamification, and virtual reality-based assessments, wearable sensor-based, and unobtrusive in-home sensor-based assessments [55]. Non-wearable unobtrusive in-home sensors, such as motion sensors, depth sensors, and bed sensors provide high autonomy and continuous flow of daily activity data. Studies have shown that the use of these sensor data could be used to predict different aspects of functional health. Arcelus et al. have successfully measured sit-to-stand timing with a mean error of 0.11 seconds and symmetry with an accuracy of 93.0% from bed pressure sensors [178]. Authors suggest that the sit-to-stand timing estimation and symmetry of bed departures could be studied for early indication of fall risk and other function health deteriorations. In another study, Greene et al. have developed a cognitive decline prediction model from Timed up and go tests performed using body-worn inertial sensors [179]. Their study included 189 community-dwelling older adults. The authors have included several quantitative movement parameters obtained from the sensors including temporal gait parameters, spatial gait parameters, turn parameters, and angular velocity parameters. Results show that baseline TUG parameters and change from baseline TUG parameters were the strongest predictors of cognitive decline. The authors used MMSE as the key assessment to predict cognitive decline in this study. In similar studies,

Verghese et al. and Lord et al. have successfully demonstrated that gait impairment could be used to predict cognitive decline and mild depressive symptoms in early Parkinson's disease, respectively [180, 181].

Stone et al. have mapped in-home sensor Kinect-based gait speed measurements to TUG time. TUG time has been a successful predictor of fall risk; hence they used the automated TUG estimation to predict fall risk in older adults. Alberdi et al. have used longitudinal smart home data collected from 29 older adult residents over an average duration of two years to detect early stages of age-related disorders such as Alzheimer's disease [182]. They obtained 10-time series behavioral features from the sensor data, such as duration of specific activities including time spent for cooking, eating, and relaxing; mobility-related features including total number of activated sensors, and total distance covered walking inside the home per day; sleep-related features including sleep duration, and frequency. Results show that the behavioral features could be successfully used to predict TUG with a statistically significant regression correlation of ~0.55. The models however struggled to predict all other absolute test scores including GDS, and Prospective and Retrospective Memory Questionnaire (PRMQ).

In-home sensors, such as depth sensors, bed sensors, and motion sensors have been successfully deployed at TigerPlace, a senior living facility in Columbia, MO for continuous health monitoring of older adults. The data obtained from the sensors have been used to estimate several clinically relevant matrices including gait speed, estimated TUG, sleep restlessness, and in-home activities. In this analysis, we explore methods to predict functional health from these matrices. In our earlier studies (Chapter 4), we have estimated overall functional health in older adults using ADL, IADL, MMSE, GDS, and SF-12

matrices from the electronic health records data. Our task in this study is to estimate the overall functional health estimation we obtained in the previous study using health matrices obtained from the in-home sensors. Also, we will estimate absolute assessment scores of the geriatric assessments from in-home sensor data. We have used the linear regression method to solve this regression problem. We have also constructed functional health tracking plots using predicted functional health from the sensor data and estimated overall functional health from EHR data. These plots help us understand if the trends observed in the composite functional health values are reflected in the predicted functional health values.

8.2 Methods

8.2.1 Data

We used a set of continuously measured data obtained from depth sensors and bed sensors embedded at TigerPlace. The features obtained from depth sensors were stride length, stride time, gait speed, walks per day, and estimated TUG time. The features obtained from bed sensors were seconds in bed, seconds restless, restlessness percentage, average respiration rate, and average heart rate. We used these features to predict the absolute geriatric assessments and the overall functional health estimated using the mixed-effect models as described in chapter 4 [183].

The dataset contained 150 (females =97, age = 87.2 ± 7.2) older adult residents who lived at TigerPlace from 2011-2019 were included in the analysis. Because of a large number of missing values in the dataset, a subset of this dataset was used for each regression model construction. Functional health values were estimated biannually as

geriatric assessments were collected once every six months at TigerPlace. The mean date of all assessments was obtained and considered as the date of functional health assessment. Based on the date of functional health assessment, mean values of the sensor features were obtained using sensor feature measurement 15 days before and after the date of functional health assessment. We then used these average values as our features for regression models. We excluded functional health values that did not accompany sensor data. A correlation plot of the bed sensor and depth sensor features in Fig 8.1 shows that stride length and estimated TUG were highly correlated to gait speed. Therefore, stride length and estimated TUG were discarded from the analysis. Seconds in bed and seconds restless were also highly correlated, therefore, feature seconds in bed was discarded from the analysis.

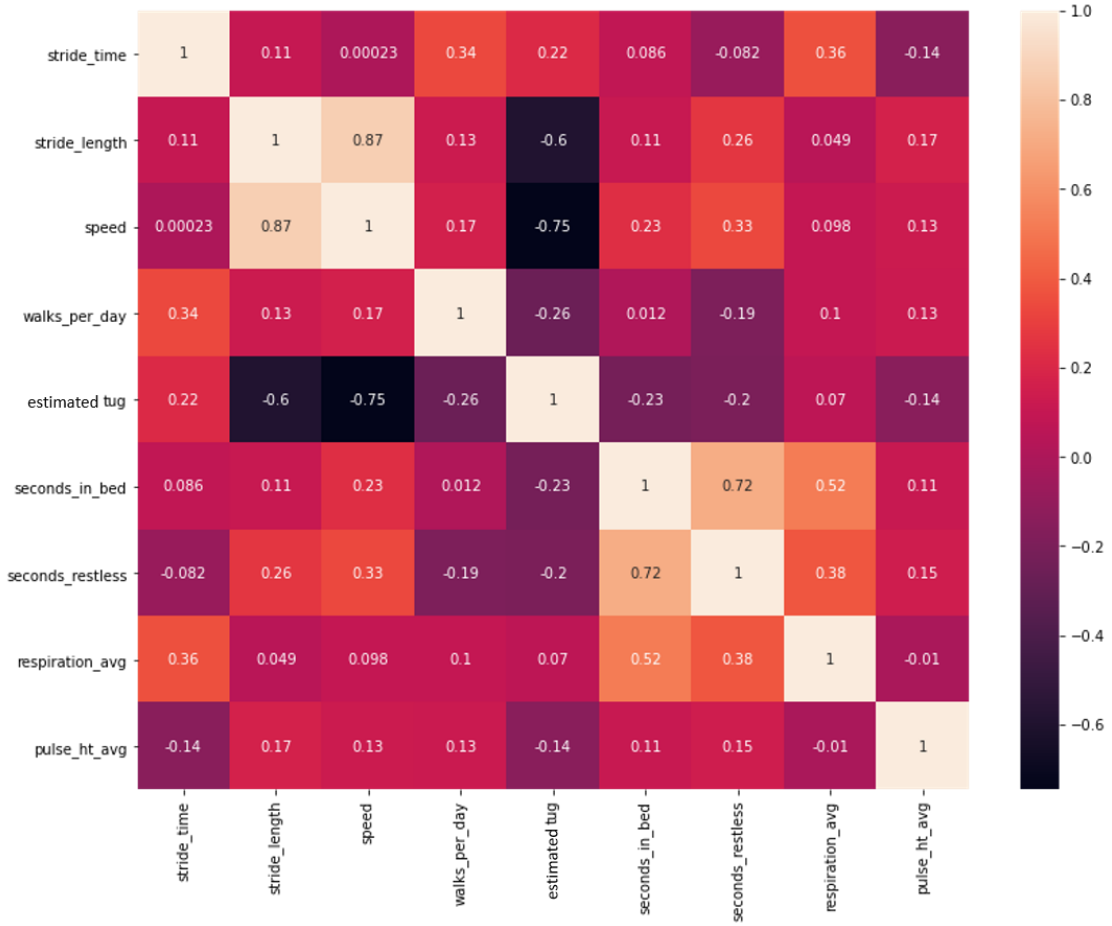


Figure 8.1. Correlation plot of depth and bed sensor features

For each geriatric assessment, we have also performed a feature selection to obtain the best features to predict that particular assessment. We used all features including the correlated features for the feature selections. This study received Institutional Review Board approval at the University of Missouri, Columbia.

8.2.2 Functional Health Estimation from In-Home Sensor Features

The dataset obtained from the previous steps was used to construct regression models to estimate the functional health of an individual using the in-home sensor data.

We not only estimated the composite functional health values developed from the geriatric assessments but also the geriatric assessments themselves. We finally constructed functional health trajectories from the cross-validation regression predictions and compared them with the true functional health estimates based on methods described in chapter four.

8.2.2.1 Regression Analysis

We used linear regression to construct the prediction models. We used the set of depth sensor-based features and bed sensor-based features separately to construct the regression models. Also, we have used a set of selected features obtained using univariate F-test statistics feature selection to construct the regression models to estimate each outcome variable. We used the leave-one-out method (LOOCV) of cross-validation because of limited data availability.

8.2.2.2 Evaluation

For the regression models, we evaluated the models using correlation coefficients (r) and means square errors (MSE).

8.3 Results

8.3.1 Regression Analysis

We performed regression analysis to predict the geriatric assessments, including ADL, IADL, MMSE, GDS, and the composite functional health values obtained in Chapter 4. Table 8.1 shows the regression output correlation and MSE values for predicting the outcome variables from depth sensor-based features, bed sensor-based features, depth

sensor-based features with seconds restlessness obtained from bed sensors, and selected features for each outcome variable obtained using feature selection.

Table 8.1 Predicting geriatric assessments and FHV from in-home sensor features

	Depth sensor-based Features (n=37, F=26)		**Bed sensor-based features (n=30, F=21)**		**Depth sensor-based Features + seconds restless (n=16, F=9)**		**Features obtained from feature selection**	
	r (p-value)	**mse**	**r (p-value)**	**mse**	**r (p-value)**	**mse**	**r (p-value)**	**mse**
ADL	0.54 ‡ (<0.0001)	7.4	0.35 ∗ (0.002)	9.41	0.72 ‡ (<0.0001)	18.21	0.78 ‡ (<0.0001)	30.4
IADL	0.37 † (0.0002)	11.99	0.16 (0.17)	10.23	0.39 ∗ (0.02)	9.91	0.33 ∗ (0.07)	9.59
MMSE	0.063 (0.55)	621.3	0.304 ∗ (0.009)	585.0	0.097 (0.590)	587.21	0.40 ∗ (0.023)	599.01
GDS	0.24 ∗ (0.019)	6.59	-0.20 (0.079)	8.29	0.37 (0.03)	8.96	0.73 ‡ (<0.0001)	11.94
FHV	0.32 ∗ (0.002)	0.04	-0.10 (0.37)	0.04	0.309 (0.079)	0.025	0.36 ∗ (0.04)	0.027

∗ p<0.05

† p<0.001

‡ p<0.0001

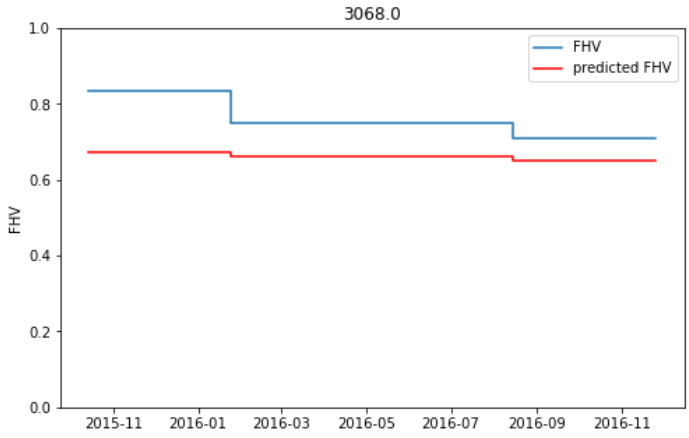
Results show that the constructed linear regression models could estimate the geriatric outcome variables with statistical significance. Features obtained using the feature selection method could predict all individual assessments including composite functional health values with statistical significance. Out of the observed correlation in Table 8.1,

GDS and ADL were predicted with the highest correlation values with $p<0.0001$. The set of features selected for the individual outcome variables using the feature selection method are shown in Table 8.2.

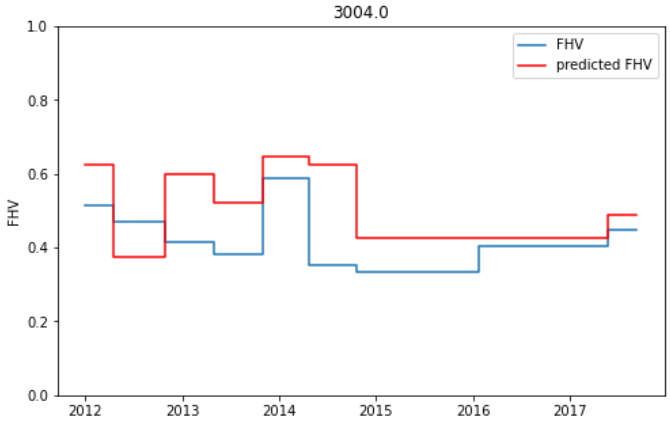
Table 8.2 Selected features using univariate F-test statistics

Geriatrics assessments / FHV	**Selected features**
ADL	Stride time, stride length, speed, walks per day, estimated tug, seconds restless, restless percentage
IADL	Stride time, speed, walks per day, estimated tug, restless percentage
MMSE	Stride time, stride length, speed, walks per day, average respiration
GDS	Stride time, speed, walks per day, seconds restless
FHV	Stride time, stride length, estimated tug, average respiration

In addition to the correlation tests, we compare predictions obtained from this model to the actual functional health estimates from mixed-effect models, we plotted the time series functional health measures for a subset of residents. Figures 8.2 (a) and (b) show trends of actual and predicted functional health values for two TigerPlace residents.



(a)



(b)

Figure 8.2. Predicted and expected functional health values of two TigerPlace residents

8.4 Discussions

Results show that correlation values and mean square errors in predicting the geriatric assessments and functional health values are comparable or better than results observed in the literature [11]. Health trends observed from the functional health value estimations show that predicted functional health values generally follow the actual

functional health trends even though the model overestimated the functional health values in both case studies.

To understand the poor relatively lower correlation values in the prediction, we constructed pair plots of the features suggest that none of the individual features or pairs of the features are well separated for the distribution of values in IADL. Figures 8.3 and 8.4 shows that gait and sleep feature distributions for different ranges of IADL were significantly overlapped.

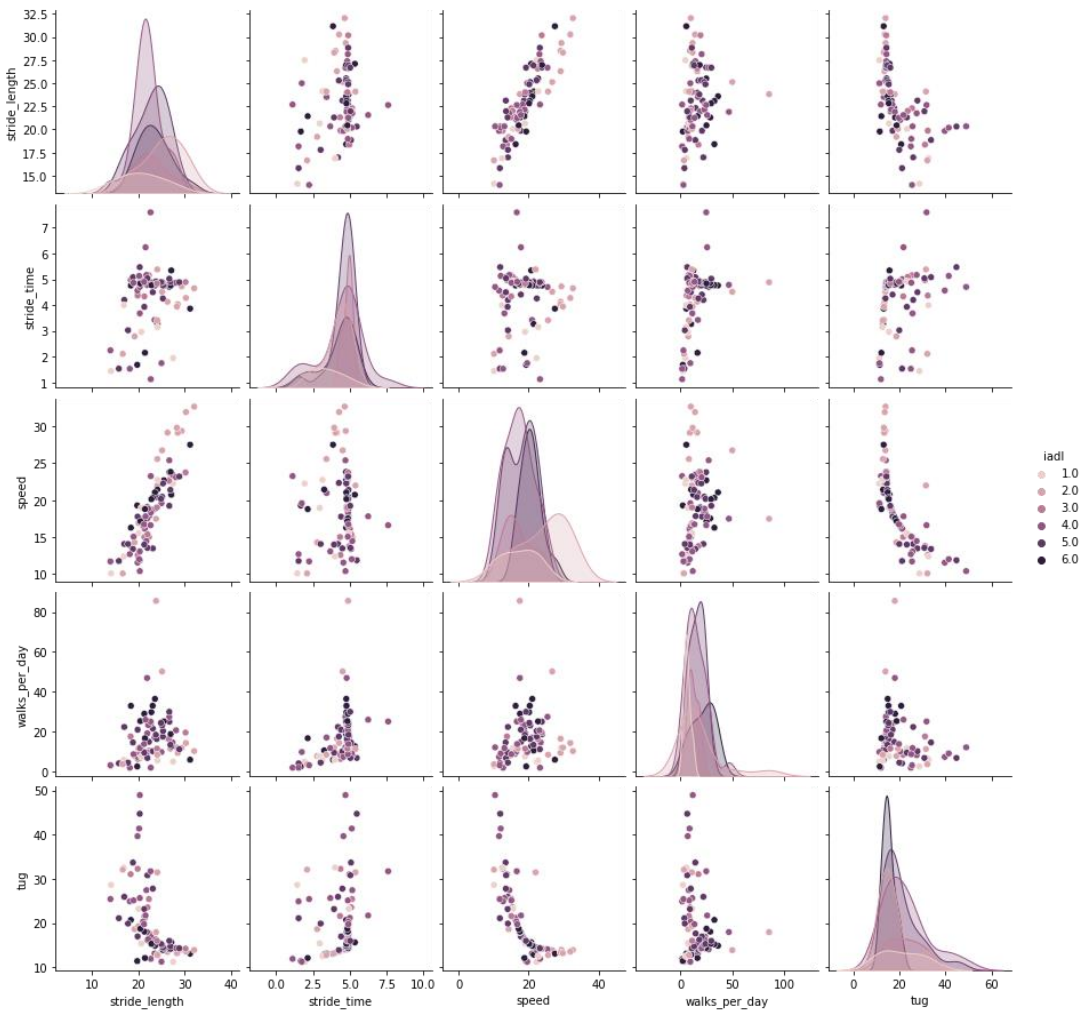


Figure 8.3. Pair plots of gait features with color hues based on IADL range

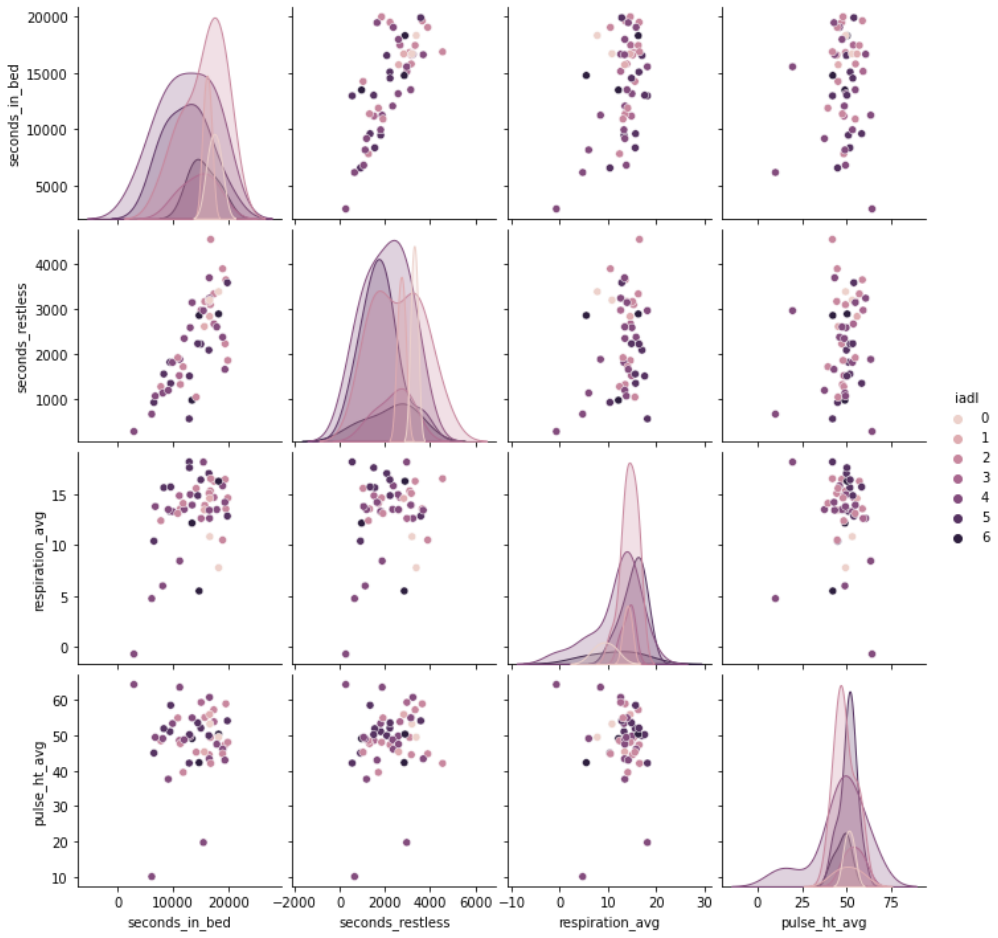


Figure 8.4. Pair plots of sleep-related features with color hues based on IADL range

The pair plots provide evidence that the data used to develop regression models to predict IADL were significantly overlapped. This could explain the relatively lower correlation values in Table 8.1.

Results show that model performances were typically better when the models were constructed from depth sensor-based features as compared to bed sensor-based features. This could be because functional health values may depend more on gait features than bed sensor-based features. The feature selection method also provides further evidence that

gait-based features were more successful in predicting functional health than features obtained from bed sensors.

8.5 Conclusions

We performed an analysis to predict geriatric assessments and composite functional health estimates from in-home sensor data from bed and depth sensors. We demonstrated that the outcome variables can be estimated from sensor data. We also demonstrated that predicted composite functional health values do follow the trend of true functional health values, suggesting positive or negative changes in functional health values could be successfully estimated.

Chapter 9: Conclusion and Future Work

9.1 Summary

The work described in this dissertation lays the groundwork for using de-identified EHR and in-home sensor data to track personalized functional health, predict fall risk using unstructured clinical nursing notes and geriatric assessments, explaining fall risk models to understand which features contributed to increased fall risk to provide precise interventions, and performing a secondary fall analysis to reduce false alarms. The models use nine years of longitudinal EHR, and in-home sensor data obtained from depth sensors and bed sensors at TigerPlace to predict functional health and fall risk.

9.2 Major Contributions

The major contributions of this dissertation study include:

- Development of an electronic health records (EHR) dataset for predictive analytics
    - De-identified the EHR databases used at TigerPlace senior living facility from 2010 to 2018
    - Re-compiled geriatric assessments data in the de-identified database
    - Organized EHR data and used it for predictive analytics in constructing personalized functional health tracking and fall risk prediction models
- Development of health risk prediction models using geriatric assessments in the EHR data

    - Developed a Functional Health Index using geriatric assessments to track longitudinal functional health changes in older adults using excess-risk functions
    - Developed mixed effect models using repeated assessment measures to predict composite personalized functional health values
    - Developed and tested the models on 150 TigerPlace residents and results show higher functional health corresponds to lower fall risk, hospitalizations, and mortality
    - Analyzed case studies that show that the predicted functional health correlate with underlying health changes as observed in the EHR
- 6-month fall risk prediction using geriatric assessments and GAITRite-assessment data
    - Developed fall risk prediction models using machine learning algorithms from spatiotemporal gait parameters and geriatric assessment data
    - Used model explainability methods to explain fall risk predictions for personalized interventions
    - Evaluated models on 93 independent living residents at TigerPlace
- Retrospective analysis of fall alert videos to prune out false alarms
    - Developed and labeled a fall alert dataset containing more than 4000 individual fall alert videos
    - Developed classification models to accurately identify true falls reducing false fall alerts by 21-33% using deep neural architectures

    - Deployed the secondary fall analysis models successfully in production environments after testing the model for six months
- Fall risk prediction using clinical notes and medications
    - Developed deep neural architectures and word embedding models to predict fall and hospitalization risk from clinical notes and medications data
    - Utilized EHR data, specifically nursing notes and medication data from 162 TigerPlace residents to develop the models with state-of-the-art results in predicting fall risk in older adults
- Functional health prediction using in-home sensor data
    - Developed regression models to predict functional health from in-home sensor data obtained from depth sensors and bed sensors
    - Used model prediction to understand functional health trends in older adults

9.3 Future Work

Although the models developed in this dissertation have significant capabilities, avenues exist for further research and improvement. The first of these is using the predictions obtained from the personalized functional health model to improve the alert system at TigerPlace. Studies show that a significant number of alerts generated at TigerPlace are irrelevant and redundant. The personalized functional health predictions could be used to personalize the alert system. For example, the alert threshold could be adjusted such that an individual with higher functional health received fewer alerts as compared to an induvial with lower functional health.

Second, the personalized functional health predictions could be further evaluated using more case studies and in-depth analysis of corresponding nursing notes. The general

notion of functional health in older adults is the overall functional health of an individual decreases over time. However, based on our model predictions we have observed increased functional health in a few scenarios for some residents. Understanding what caused these increases in functional health could provide valuable information about intervention strategies to help older adults maintain higher functional health through their stay at senior living facilities like TigerPlace.

Lastly, there are several aspects of the TigerPlace EHR data that are not being used to predict fall risk or track functional health. For example, vital signs and lab results. Developing a fall risk model that uses other structured data in the EHR to predict fall risk could result in developing a more robust fall prediction model. In the current models, only geriatrics assessments, nursing notes, visit notes, and medications were used to predict fall risk.

Appendix A FHI Trajectories of TigerPlace Residents

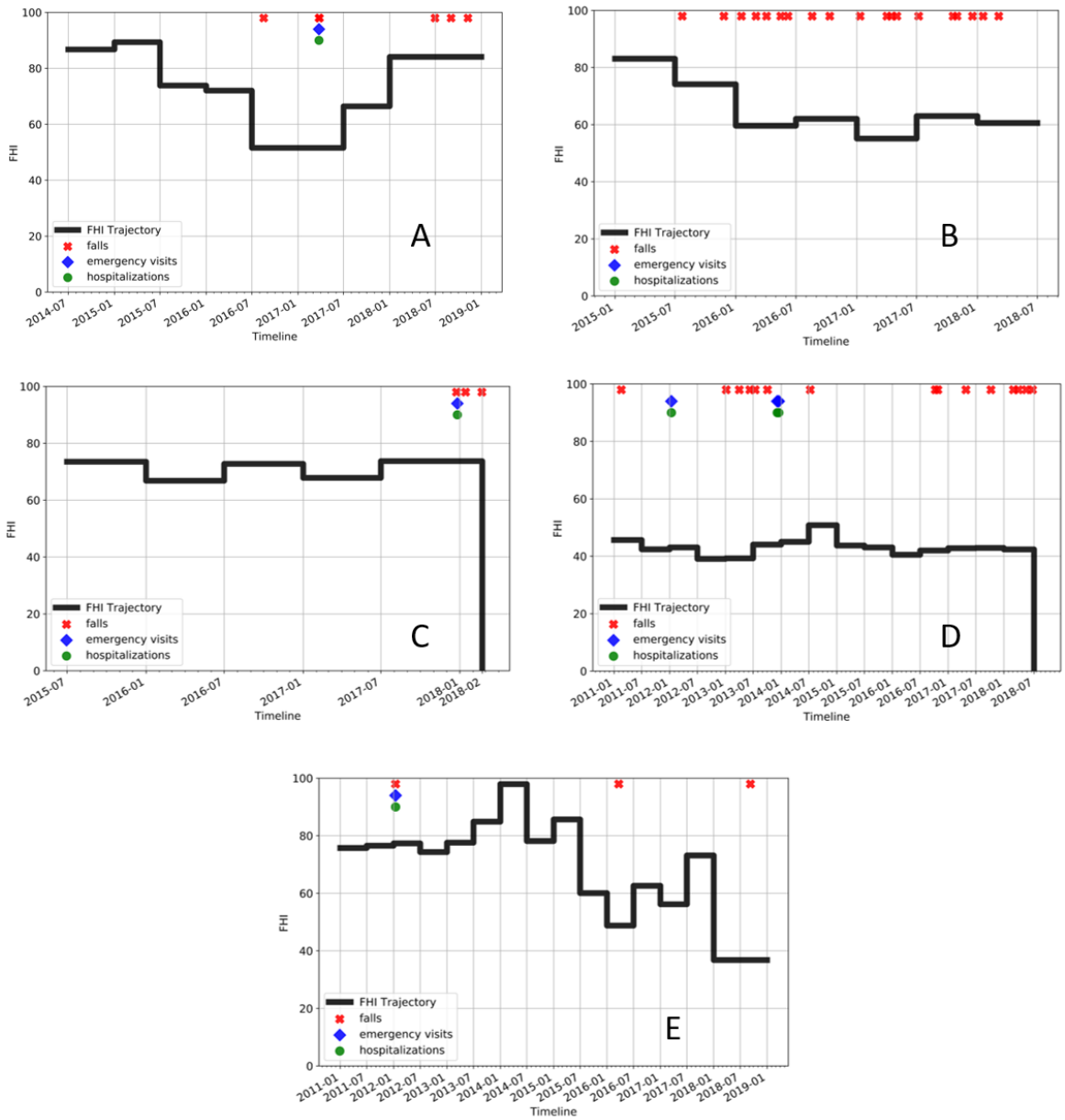
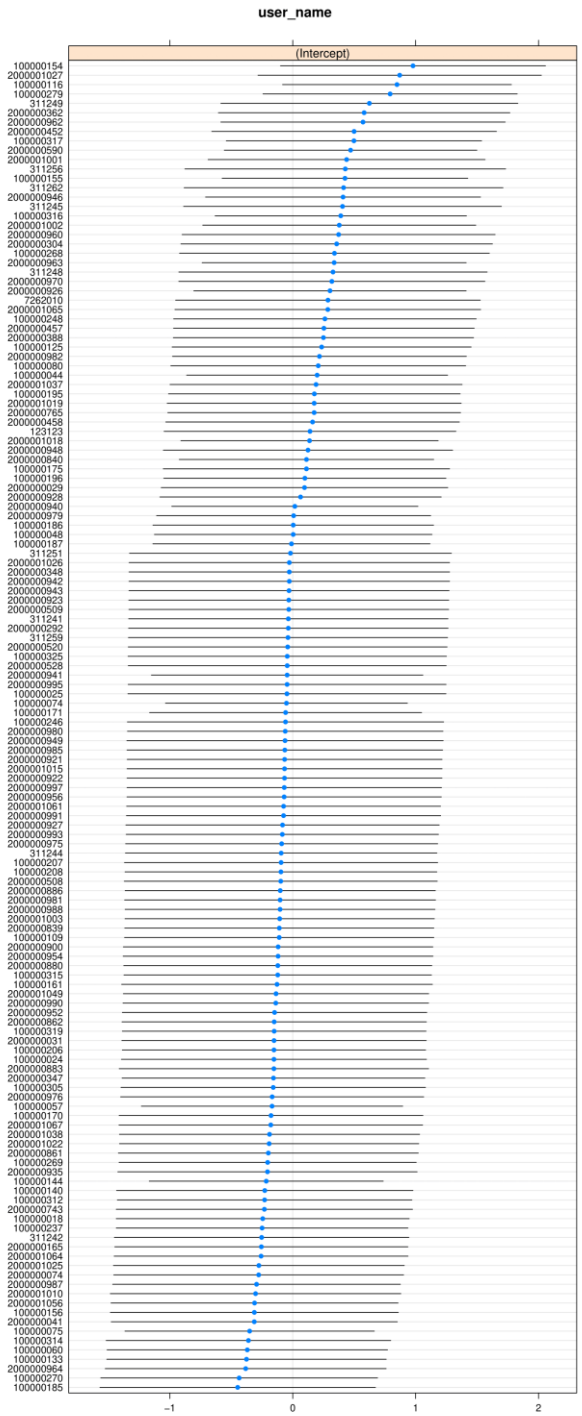


Fig. A1 FHI Trajectories of TigerPlace Residents. A: A significant drop in FHI between 2016-2018 corresponds to hospitalizations and falls. B: FHI continuously declined until Jan 2016 (~60) and never improved, followed by several falls. C: An example of consistent functional health with no significant FHI declines. Several health events were observed

close to death. D: Lower FHI (< 60), corresponding to a significant number of falls and other health events throughout the stay. E: The initial fall and other health events cannot be explained (FHI~70). FHI improved over time until June 2014. Significant drops in FHI during July-Dec 2015 and January-June 2018 correspond to falls.

Appendix B Random Effects for the final Mixed Effect Model Development to Estimate Functional Health in Older Adults Using Geriatric Assessments



Appendix C Fall Alert Labelling Categories

Common Descriptions	Condensed Descriptions	Class 1	Class 2	Class 3	fall
staff making bed	laundry being done	laundry			no
clothing fell to floor	laundry being done	laundry			no
resident threw off pillows/sheets while in bed	laundry being done	laundry			no
laundry thrown on ground	laundry being done	laundry			no
laundry being thrown on ground, non-resident bending over	laundry being done	laundry	bend		no
non-resident bending over, folding laundry	laundry being done	laundry	bend		no
bedding/pillows thrown on ground	laundry being done	laundry			no
pet jumped off furniture	pet jumped off furniture	pet			no
dog laying down	pet moving around apartment	pet			no
pet walking	pet moving around apartment	pet			no
pet running	pet moving around apartment	pet			no
pet running, knocked over object	pet moving around apartment	pet	object		no
pet? Moving in corner	pet moving around apartment	pet			no
pet walking near resident	pet moving around apartment	pet			no
pet walking, non-resident bending over	pet moving around apartment	pet	bend		no
pet walking, resident bending over	pet moving around apartment	pet	bend		no

pet walking, multiple people in room, lots of activity	pet moving around apartment	pet	high_activity		no
resident pushed object, pet walking near resident	pet moving around apartment	pet	object		no
pet walking, laundry thrown on ground	pet moving around apartment	pet	laundry		no
pet playing on ground, multiple people in room	pet moving around apartment	pet	high_activity		no
non-resident pushing object on ground, pet walking	pet moving around apartment	pet	object		no
resident moving object around on ground	object being pushed around apartment	object			no
non-resident bent over pushing object	object being pushed around apartment	object	bend		no
non-resident bending over, moving objects	object being pushed around apartment	object	bend		no
non-resident bent over, dropped object	object being pushed around apartment	object	bend		no
non-resident bending over, tossed object across floor	object being pushed around apartment	object	bend		no
non-resident pushing object, multiple people in room	object being pushed around apartment	object	high_activity		no
object kicked	object being pushed around apartment	object			no
object fell to floor	object being pushed around apartment	object			no
resident rocking in recliner	resident in recliner (rocking, using footrest, adjusting blanket)	recliner			no
resident unfolding blanket while sitting in recliner	resident in recliner (rocking, using footrest,	recliner	laundry		no

	adjusting blanket)				
resident unfolded blanket then sat in recliner	resident in recliner (rocking, using footrest, adjusting blanket)	recliner	laundry		no
resident put up footrest on recliner	resident in recliner (rocking, using footrest, adjusting blanket)	recliner			no
resident moved legs while sitting in recliner	resident in recliner (rocking, using footrest, adjusting blanket)	recliner			no
resident leaned forward in chair, dropped laundry on floor, multiple people in room	resident in recliner (rocking, using footrest, adjusting blanket)	recliner	laundry	high_activity	no
resident sat in recliner then crossed legs	resident in recliner (rocking, using footrest, adjusting blanket)	recliner			no
resident crossed feet while sitting in recliner	resident in recliner (rocking, using footrest, adjusting blanket)	recliner			no
resident dropped recliner footrest and removed blanket	resident getting out of chair	recliner	laundry	object	no
resident attempting to get out of chair	resident getting out of chair	recliner			no
resident got out of chair	resident getting out of chair	recliner			no
resident walking with walker	resident moving with assistive device in apartment	ast_device			no

resident walking with walker and being assisted, multiple people in room	resident moving with assistive device in apartment	ast_device	high_activity		no
resident using walker to back into chair	resident moving with assistive device in apartment	ast_device			no
resident backed up in wheelchair then bent over	resident moving with assistive device in apartment	ast_device	bend		no
resident using arms to move in wheelchair	resident moving with assistive device in apartment	ast_device			no
resident backed up in wheelchair	resident moving with assistive device in apartment	ast_device			no
resident being pushed by staff in wheelchair	resident moving with assistive device in apartment	ast_device			no
resident moved while sitting in chair	resident moving with assistive device in apartment	ast_device			no
resident moving slowly	gait change, bending over	walk			no
resident bent over	gait change, bending over	bend			no
resident walking fast	gait change, bending over	walk			no
non-resident exercising	gait change, bending over	exercise			no
non-resident walking fast	gait change, bending over	walk			no
stretcher brought into room	Multiple people in room, activity level high	high_activity	object		no
non-resident moved furniture	Multiple people in room, activity level high	high_activity	object		no

non-resident brought in dolly, then bent over	Multiple people in room, activity level high	high_activity	bent	object	no
non-resident folding blankets, non-resident and child on floor, multiple people in room, lots of activity	Multiple people in room, activity level high	high_activity			no
CERT staff working on depth computer	Multiple people in room, activity level high	high_activity	fall		other
child sat on floor, multiple people in room	Multiple people in room, activity level high	high_activity			no
multiple people in room	Multiple people in room, activity level high	high_activity			no
non-resident and child on floor, multiple people in room, lots of activity	Multiple people in room, activity level high	high_activity			no
staff leaning towards resident	staff/family assisting resident	other			no
non-resident bent over to assist with recliner	staff/family assisting resident	bend			no
staff attending to resident in bed	staff/family assisting resident	other			no
non-resident kneeling by resident on floor, multiple people in room	staff/family assisting resident	high_activity			no
unsure, no good image	unsure, no good image	other			no
resident fell backwards with walker	resident fell backwards	fall_back			yes
resident fell backwards into chair	resident fell backwards	fall_back			yes
resident fell backwards into recliner	resident fell backwards	fall_back			yes
resident fell backwards into bed	resident fell backwards	fall_back			yes

resident leaned back to sit in chair and fell	resident fell backwards	fall_back			yes
resident bent over and fell backwards	resident fell backwards	fall_back			yes
resident bent over to move object	resident/non-resident bending over	bend			no
non-resident bending over	resident/non-resident bending over	bend			no
resident fell forward	resident/non-resident fell forward	fall_forward			yes
resident fell forward while walking with cane	resident/non-resident fell forward	fall_forward			yes
non-resident fell to knees	resident/non-resident fell forward	fall_forward			yes
resident fell forward into chair while walking, more than one person in room	resident/non-resident fell forward	fall_forward			yes
non-resident sat on floor	resident/non-resident sitting on floor, recumbent position	sit			no
non-resident sat on floor, laid on back, pet walking	resident/non-resident sitting on floor, recumbent position	sit			no
non-resident laying on ground	resident/non-resident sitting on floor, recumbent position	sit			no
staff sitting and scooting on ground	resident/non-resident sitting on floor, recumbent position	sit			no
resident fell to side into bed while using walker	resident fell sideways	fall_side			yes

resident fell sideways into chair	resident fell sideways	fall_side			yes
resident fell to side while walking	resident fell sideways	fall_side			yes
resident fell to side into chair while using walker	resident fell sideways	fall_side			yes
resident slid off of chair and slumped to floor	resident fell out of chair	fall_chair			yes
resident fell out of wheelchair	resident fell out of chair	fall_chair			yes
screen partition fell near resident	screen partition fell near resident	other			no

Appendix D Word Embedding Model Comparison

We compared BioWordVec, which is trained on biomedical text, and GloVe Embeddings, which is trained on Wikipedia. These comparisons were made over the first 6 months of notes each resident had (the notes that would become the dataset). When applied to our data, BioWordVec found 9,435 unique words. The Maximum amount of words found within someone's text data was 18,242, and the mean for each resident was 1,889.93. Similarly, GloVe Embeddings had 8,037 unique tokens and an average per resident of 1,844.20 (Table III). Additionally, the distribution of tokens per resident for each pre-trained embedding is shown in Figures 1 & 2.

We decided to do our testing with BioWordVec because of its specialization with medical terms. Although GloVe Embeddings had a similar average, there were 1,500 words that BioWordVec had embeddings for that GloVe did not. These words included blood pressure measurements, blood sugar measurements, medication names, etc. that would contribute to fall prediction.

TABLE D1. Comparison of Word Embedding Models First Six-Month Statistics

Embedding	***Total Tokens***	***Unique Tokens***	***Mean Tokens per Resident***	***Tokens Unique to This Embedding***
BioWordVec	306,169	9,435	1,889.93	1,435
GloVe	298,760	8,037	1,844.20	37

Appendix E Code Repositories

Chapter 3, Chapter 4, Chapter 5, Chapter 8:

https://vcs.missouri.edu/akmm94/THIL_codes_de_id_db.git

Chapter 3: THIL_codes_de_id_db/source/create_deid_db

Chapter4:

THIL_codes_de_id_db/source/data_analysis/journal_data_analysis_JBHI_2019

Chapter 5: THIL_codes_de_id_db/source/data_analysis/PervasiveHealth2020

Chapter 8: THIL_codes_de_id_db/source/regress_sensor_to_ehr

Chapter 6:

Preprocessing: https://vcs.missouri.edu/akmm94/sfa_pre_processing_dev.git

Development: https://vcs.missouri.edu/akmm94/secondary_fall_analysis.git

Production:

https://vcs.missouri.edu/CERT-Students/studentProjects/secondary_fall_analysis.git

Chapter 7:

https://vcs.missouri.edu/akmm94/reu2020_nlp.git

Bibliography

[1] M. Mather, P. Scommegna, and L. Kilduff. "Fact sheet: aging in the United States." https://www.prb.org/aging-unitedstates-fact-sheet/ (accessed 6/5/2020.

[2] J. C. f. H. S. o. H. University, "Projections and implications for housing a growing population: Older households 2015–2035," ed: Harvard University Cambridge, MA, 2016.

[3] M. J. Rantz *et al.*, "Sensor technology to support Aging in Place," (in eng), *J. Am. Med. Dir. Assoc.*, vol. 14, no. 6, pp. 386-91, Jun 2013, doi: 10.1016/j.jamda.2013.02.018.

[4] M. Rantz *et al.*, "The continued success of registered nurse care coordination in a state evaluation of aging in place in senior housing," (in eng), *Nurs. Outlook*, vol. 62, no. 4, pp. 237-46, Jul-Aug 2014, doi: 10.1016/j.outlook.2014.02.005.

[5] M. J. Rantz *et al.*, "A technology and nursing collaboration to help older adults age in place," *Nurs. Outlook*, vol. 53, no. 1, pp. 40-5, Jan-Feb 2005, doi: 10.1016/j.outlook.2004.05.004.

[6] "TigerPlace." https://www.americareusa.net/senior-living/mo/columbia/tiger-place/ (accessed 06/26, 2020).

[7] M. Rantz *et al.*, "Randomized Trial of Intelligent Sensor System for Early Illness Alerts in Senior Housing," (in eng), *J. Am. Med. Dir. Assoc.*, vol. 18, no. 10, pp. 860-870, Oct 1 2017, doi: 10.1016/j.jamda.2017.05.012.

[8] E. E. Stone and M. Skubic, "Testing real-time in-home fall alerts with embedded depth video hyperlink," in *International Conference on Smart Homes and Health Telematics*, 2014: Springer, pp. 41-48.

[9] S. Sendelbach and M. Funk, "Alarm fatigue: a patient safety concern," *AACN Adv. Crit. Care*, vol. 24, no. 4, pp. 378-386, 2013.

[10] R. I. Bjarnadottir and R. J. Lucero, "What Can We Learn about Fall Risk Factors from EHR Nursing Notes? A Text Mining Study," (in eng), *EGEMS (Wash DC)*, vol. 6, no. 1, p. 21, Sep 20 2018, doi: 10.5334/egems.237.

[11] A. Alberdi Aramendi, A. Weakley, A. Aztiria Goenaga, M. Schmitter-Edgecombe, and D. J. Cook, "Automatic assessment of functional health decline in older adults based on smart home data," *J. Biomed. Inform.*, vol. 81, pp. 119-130, 2018/05/01/ 2018, doi: https://doi.org/10.1016/j.jbi.2018.03.009.

[12] M. J. Rantz *et al.*, "TigerPlace, A State-Academic-Private Project to Revolutionize Traditional Long-Term Care," (in eng), *J Hous Elderly*, vol. 22, no. 1-2, pp. 66-85, 2008, doi: 10.1080/02763890802097045.

[13] E. Stone, M. Skubic, M. Rantz, C. Abbott, and S. Miller, "Average in-home gait speed: investigation of a new metric for mobility and fall risk assessment of elders," (in eng), *Gait Posture*, vol. 41, no. 1, pp. 57-62, Jan 2015, doi: 10.1016/j.gaitpost.2014.08.019.

[14] M. Gordon, "Functional health patterns," *Nursing Diagnosis Process and Application. New York: Mc. Graw Hill Book Comp*, pp. 685-702, 1982.

[15] M. J. Rothman, S. I. Rothman, and J. t. Beals, "Development and validation of a continuous measure of patient condition using the Electronic Medical Record," *J. Biomed. Inform.*, vol. 46, no. 5, pp. 837-48, Oct 2013, doi: 10.1016/j.jbi.2013.06.011.

[16] E. E. Stone and M. Skubic, "Fall detection in homes of older adults using the Microsoft Kinect," (in eng), *IEEE J Biomed Health Inform,* vol. 19, no. 1, pp. 290-301, Jan 2015, doi: 10.1109/jbhi.2014.2312180.

[17] M. Skubic, B. H. Harris, E. Stone, K. C. Ho, S. Bo-Yu, and M. Rantz, "Testing non-wearable fall detection methods in the homes of older adults," (in eng), *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference,* vol. 2016, pp. 557-560, 2016/08// 2016, doi: 10.1109/embc.2016.7590763.

[18] M. Rantz *et al.,* "Enhanced registered nurse care coordination with sensor technology: Impact on length of stay and cost in aging in place housing," *Nurs. Outlook,* vol. 63, no. 6, pp. 650-655, 2015.

[19] B. Reeder and A. David, "Health at hand: a systematic review of smart watch uses for health and wellness," *J. Biomed. Inform.,* vol. 63, pp. 269-276, 2016.

[20] G. Demiris *et al.,* "Older adults' attitudes towards and perceptions of 'smart home'technologies: a pilot study," *Med. Inform. Internet Med.,* vol. 29, no. 2, pp. 87-94, 2004.

[21] A. M. Horkan. (2014, 2014 January-February) Alarm fatigue and patient safety. *Nephrology Nursing Journal* [Article]. 83+.

[22] B. J. Drew *et al.,* "Insights into the Problem of Alarm Fatigue with Physiologic Monitor Devices: A Comprehensive Observational Study of Consecutive Intensive Care Unit Patients," *PLoS One,* vol. 9, no. 10, p. e110274, 2014, doi: 10.1371/journal.pone.0110274.

[23] ECRI. "2020 Top 10 Health Technology Hazards Executive Brief." https://www.ecri.org/landing-2020-top-ten-health-technology-hazards (accessed.

[24] M. Funk, J. T. Clark, T. J. Bauld, J. C. Ott, and P. Coss, "Attitudes and Practices Related to Clinical Alarms," *Am. J. Crit. Care,* vol. 23, no. 3, pp. e9-e18, 2014, doi: 10.4037/ajcc2014315.

[25] M. Cvach, "Monitor alarm fatigue: an integrative review," *Biomed. Instrum. Technol.,* vol. 46, no. 4, pp. 268-277, 2012.

[26] K. C. Graham and M. Cvach, "Monitor Alarm Fatigue: Standardizing Use of Physiological Monitors and Decreasing Nuisance Alarms," *Am. J. Crit. Care,* vol. 19, no. 1, pp. 28-34, 2010, doi: 10.4037/ajcc2010651.

[27] S. Feder and M. Funk, "Over-monitoring and alarm fatigue: for whom do the bells toll?," (in eng), *Heart Lung,* vol. 42, no. 6, pp. 395-6, Nov-Dec 2013, doi: 10.1016/j.hrtlng.2013.09.001.

[28] B. A. Goldstein, A. M. Navar, M. J. Pencina, and J. Ioannidis, "Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review," *J. Am. Med. Inform. Assoc.,* vol. 24, no. 1, pp. 198-208, 2017.

[29] L. Ferrucci *et al.,* "Designing randomized, controlled trials aimed at preventing or delaying functional decline and disability in frail, older persons: a consensus report," *J. Am. Geriatr. Soc.,* vol. 52, no. 4, pp. 625-634, 2004.

[30] R. J. Johnson and F. D. Wolinsky, "The Structure of Health Status Among Older Adults: Disease, Disability, Functional Limitation, and Perceived Health," *J. Health Soc. Behav.,* vol. 34, no. 2, pp. 105-121, 1993, doi: 10.2307/2137238.

[31] J. C. Millán-Calenti *et al.*, "Prevalence of functional disability in activities of daily living (ADL), instrumental activities of daily living (IADL) and associated factors, as predictors of morbidity and mortality," *Arch. Gerontol. Geriatr.*, vol. 50, no. 3, pp. 306-310, 2010.

[32] J. A. Yesavage *et al.*, "Development and validation of a geriatric depression screening scale: a preliminary report," *J. Psychiatr. Res.*, vol. 17, no. 1, pp. 37-49, 1982.

[33] W. J. Burke, R. L. Nitcher, W. H. Roccaforte, and S. P. Wengel, "A prospective evaluation of the Geriatric Depression Scale in an outpatient geriatric assessment center," *J. Am. Geriatr. Soc.*, vol. 40, no. 12, pp. 1227-1230, 1992.

[34] B. Gandek *et al.*, "Cross-validation of item selection and scoring for the SF-12 Health Survey in nine countries: results from the IQOLA Project," *J. Clin. Epidemiol.*, vol. 51, no. 11, pp. 1171-1178, 1998.

[35] C. Jenkinson *et al.*, "A shorter form health survey: can the SF-12 replicate results from the SF-36 in longitudinal studies?," *Journal of Public Health*, vol. 19, no. 2, pp. 179-186, 1997.

[36] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "“Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician," *Journal of Psychiatric Research*, vol. 12, no. 3, pp. 189-198, 1975/11/01/ 1975, doi: https://doi.org/10.1016/0022-3956(75)90026-6.

[37] M. Folstein, J. C. Anthony, I. Parhad, B. Duffy, and E. M. Gruenberg, "The meaning of cognitive impairment in the elderly," (in eng), *J Am Geriatr Soc*, vol. 33, no. 4, pp. 228-35, Apr 1985, doi: 10.1111/j.1532-5415.1985.tb07109.x.

[38] W. H. Organization, *International classification of functioning, disability and health: ICF*. Geneva: World Health Organization, 2001.

[39] J. I. Sheikh, Yesavage, J.A., "Geriatric Depression Scale (GDS). Recent evidence and development of a shorter version. ," in *Clinical Gerontology: A Guide to Assessment and Intervention*, T. L. Brink Ed. NY: The Haworth Press, Inc., 1986, pp. 165-173.

[40] J. E. Ware, M. Kosinski, and S. D. Keller, "A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity," *Med. Care*, vol. 34, no. 3, pp. 220-233, 1996. [Online]. Available: http://www.jstor.org/stable/3766749.

[41] G. Mazzaglia *et al.*, "Screening of older community-dwelling people at risk for death and hospitalization: the Assistenza Socio-Sanitaria in Italia project," (in eng), *J. Am. Geriatr. Soc.*, vol. 55, no. 12, pp. 1955-60, Dec 2007, doi: 10.1111/j.1532-5415.2007.01446.x.

[42] J. J. Gagne, R. J. Glynn, J. Avorn, R. Levin, and S. Schneeweiss, "A combined comorbidity score predicted mortality in elderly patients better than existing scores," (in eng), *J. Clin. Epidemiol.*, vol. 64, no. 7, pp. 749-759, 2011, doi: 10.1016/j.jclinepi.2010.10.004.

[43] S. J. Lee, K. Lindquist, M. R. Segal, and K. E. Covinsky, "Development and validation of a prognostic index for 4-year mortality in older adults," (in eng), *JAMA*, vol. 295, no. 7, pp. 801-8, Feb 15 2006, doi: 10.1001/jama.295.7.801.

[44] E. C. Carey, K. E. Covinsky, L. Y. Lui, C. Eng, L. P. Sands, and L. C. Walter, "Prediction of mortality in community-living frail elderly people with long-term care needs," *J. Am. Geriatr. Soc.*, vol. 56, no. 1, pp. 68-75, 2008.

[45] M. A. Schonberg, R. B. Davis, E. P. McCarthy, and E. R. Marcantonio, "Index to predict 5-year mortality of community-dwelling adults aged 65 and older using data from the National Health Interview Survey," (in eng), *J. Gen. Intern. Med.*, vol. 24, no. 10, pp. 1115-22, Oct 2009, doi: 10.1007/s11606-009-1073-y.

[46] G. Corrao *et al.*, "Developing and validating a novel multisource comorbidity score from administrative data: a large population-based cohort study from Italy," *BMJ Open*, vol. 7, no. 12, p. e019503, 2017, doi: 10.1136/bmjopen-2017-019503.

[47] L. P. Fried *et al.*, "Frailty in older adults: evidence for a phenotype," *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, vol. 56, no. 3, pp. M146-M157, 2001.

[48] H. H. Atkinson *et al.*, "Cognitive Function, Gait Speed Decline, and Comorbidities: The Health, Aging and Body Composition Study," *The Journals of Gerontology: Series A*, vol. 62, no. 8, pp. 844-850, 2007, doi: 10.1093/gerona/62.8.844 %J The Journals of Gerontology: Series A.

[49] A. J. Nelson, "Functional ambulation profile," (in eng), *Phys. Ther.*, vol. 54, no. 10, pp. 1059-65, Oct 1974, doi: 10.1093/ptj/54.10.1059.

[50] A. Gouelle, "Use of functional ambulation performance score as measurement of gait ability," *J. Rehabil. Res. Dev.*, vol. 51, no. 5, p. 665, 2014.

[51] B. Bilney, M. Morris, and K. Webster, "Concurrent related validity of the GAITRite® walkway system for quantification of the spatial and temporal parameters of gait," *Gait Posture*, vol. 17, no. 1, pp. 68-74, 2003/02/01/ 2003, doi: https://doi.org/10.1016/S0966-6362(02)00053-X.

[52] F. Panza *et al.*, "Different Cognitive Frailty Models and Health- and Cognitive-related Outcomes in Older Age: From Epidemiology to Prevention," (in eng), *J Alzheimers Dis*, vol. 62, no. 3, pp. 993-1012, 2018, doi: 10.3233/JAD-170963.

[53] M. P. Lawton and E. M. Brody, "Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living," *The Gerontologist*, vol. 9, no. 3_Part_1, pp. 179-186, 1969, doi: 10.1093/geront/9.3_Part_1.179 %J The Gerontologist.

[54] A. D. Beswick *et al.*, "Complex interventions to improve physical function and maintain independent living in elderly people: a systematic review and meta-analysis," *The Lancet*, vol. 371, no. 9614, pp. 725-735, 2008/03/01/ 2008, doi: https://doi.org/10.1016/S0140-6736(08)60342-6.

[55] D. J. Cook, M. Schmitter-Edgecombe, L. Jonsson, and A. V. Morant, "Technology-Enabled Assessment of Functional Health," (in eng), *IEEE Rev. Biomed. Eng.*, vol. 12, pp. 319-332, 2019, doi: 10.1109/rbme.2018.2851500.

[56] J. C. Ayena, L. D. C. T, M. J. D. Otis, and B. A. J. Menelas, "An efficient home-based risk of falling assessment test based on Smartphone and instrumented insole," in *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*, 7-9 May 2015 2015, pp. 416-421, doi: 10.1109/MeMeA.2015.7145239.

[57] J. Wang, Z. Liu, Y. Wu, and J. Yuan, "Learning Actionlet Ensemble for 3D Human Action Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 5, pp. 914-927, 2014, doi: 10.1109/TPAMI.2013.198.

[58] M. Mancini *et al.*, "Continuous Monitoring of Turning Mobility and Its Association to Falls and Cognitive Function: A Pilot Study," *The Journals of Gerontology: Series A*, vol. 71, no. 8, pp. 1102-1108, 2016, doi: 10.1093/gerona/glw019.

[59] A. Godfrey, S. D. Din, G. Barry, J. C. Mathers, and L. Rochester, "Within trial validation and reliability of a single tri-axial accelerometer for gait assessment," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 26-30 Aug. 2014 2014, pp. 5892-5895, doi: 10.1109/EMBC.2014.6944969.

[60] M. Boukhechba, Y. Huang, P. Chow, K. Fua, B. A. Teachman, and L. E. Barnes, "Monitoring social anxiety from mobility and communication patterns," presented at the Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, Maui, Hawaii, 2017. [Online]. Available: https://doi.org/10.1145/3123024.3125607.

[61] P. N. Dawadi, D. J. Cook, and M. Schmitter-Edgecombe, "Automated Cognitive Health Assessment From Smart Home-Based Behavior Data," *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 4, pp. 1188-1194, 2016, doi: 10.1109/JBHI.2015.2445754.

[62] T. Lindeberg and I. Laptev, "Space-time interest points," in *Proceedings of IEEE International Conference on Computer Vision*, 2003, pp. 432-439.

[63] P. Scovanner, S. Ali, and M. Shah, "A 3-dimensional sift descriptor and its application to action recognition," in *Proceedings of the 15th ACM international conference on Multimedia*, 2007: ACM, pp. 357-360.

[64] A. Klaser, M. Marszałek, and C. Schmid, "A spatio-temporal descriptor based on 3d-gradients," 2008.

[65] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie, "Behavior recognition via sparse spatio-temporal features," 2005: VS-PETS Beijing, China.

[66] S. Sadanand and J. J. Corso, "Action bank: A high-level representation of activity in video," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012: IEEE, pp. 1234-1241.

[67] H. Wang and C. Schmid, "Action recognition with improved trajectories," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 3551-3558.

[68] Y. LeCun, L. Bottou, Y. Bengio, and P. J. P. o. t. I. Haffner, "Gradient-based learning applied to document recognition," vol. 86, no. 11, pp. 2278-2324, 1998.

[69] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*, 2009: Ieee, pp. 248-255.

[70] O. Russakovsky *et al.*, "Imagenet large scale visual recognition challenge," vol. 115, no. 3, pp. 211-252, 2015.

[71] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[72] K. Simonyan and A. J. a. p. a. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014.
[73] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.
[74] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
[75] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in neural information processing systems*, 2012, pp. 2843-2851.
[76] C. Farabet, C. Couprie, L. Najman, Y. J. I. t. o. p. a. LeCun, and m. intelligence, "Learning hierarchical features for scene labeling," vol. 35, no. 8, pp. 1915-1929, 2012.
[77] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2014, pp. 806-813.
[78] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580-587.
[79] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. J. a. p. a. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," 2013.
[80] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision*, 2014: Springer, pp. 818-833.
[81] S. Ioffe and C. J. a. p. a. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015.
[82] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929-1958, 2014.
[83] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026-1034.
[84] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904-1916, 2015.
[85] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 221-231, 2012.
[86] G. W. Taylor, R. Fergus, Y. LeCun, and C. Bregler, "Convolutional learning of spatio-temporal features," in *European conference on computer vision*, 2010: Springer, pp. 140-153.
[87] Q. V. Le, W. Y. Zou, S. Y. Yeung, and A. Y. Ng, "Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis," in *CVPR 2011*, 2011: IEEE, pp. 3361-3368.

[88] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1717-1724.

[89] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91-99.

[90] D. Tran, J. Ray, Z. Shou, S.-F. Chang, and M. Paluri, "Convnet architecture search for spatiotemporal feature learning," *arXiv preprint arXiv:1708.05038*, 2017.

[91] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-Scale Video Classification with Convolutional Neural Networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 23-28 June 2014 2014, pp. 1725-1732, doi: 10.1109/CVPR.2014.223.

[92] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in neural information processing systems*, 2014, pp. 568-576.

[93] J. Yue-Hei Ng, M. Hausknecht, S. Vijayanarasimhan, O. Vinyals, R. Monga, and G. Toderici, "Beyond short snippets: Deep networks for video classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4694-4702.

[94] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 4489-4497.

[95] K. Soomro, A. R. Zamir, and M. Shah, "UCF101: A dataset of 101 human actions classes from videos in the wild," *arXiv preprint arXiv:1212.0402*, 2012.

[96] J. Carreira and A. Zisserman, "Quo vadis, action recognition? a new model and the kinetics dataset," in *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 6299-6308.

[97] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri, "A closer look at spatiotemporal convolutions for action recognition," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018, pp. 6450-6459.

[98] ElderTech. "ElderTech Fall Detection Examples." YouTube. https://www.youtube.com/watch?v=TFB7YOUmHho (accessed 06/15, 2020).

[99] S. Hyun, S. B. Johnson, and S. Bakken, "Exploring the ability of natural language processing to extract data from nursing narratives," (in eng), *Comput. Inform. Nurs.*, vol. 27, no. 4, pp. 215-23; quiz 224-5, Jul-Aug 2009, doi: 10.1097/NCN.0b013e3181a91b58.

[100] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis," (in eng), *IEEE J Biomed Health Inform*, vol. 22, no. 5, pp. 1589-1604, Sep 2018, doi: 10.1109/jbhi.2017.2767063.

[101] A. N. Jagannatha and H. Yu, "Structured prediction models for RNN based sequence labeling in clinical text," (in eng), *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, vol. 2016, pp. 856-865, 2016, doi: 10.18653/v1/d16-1082.

[102] S. Sheikhalishahi, R. Miotto, J. T. Dudley, A. Lavelli, F. Rinaldi, and V. Osmani, "Natural Language Processing of Clinical Notes on Chronic Diseases: Systematic Review," (in eng), *JMIR Med Inform*, vol. 7, no. 2, p. e12239, Apr 27 2019, doi: 10.2196/12239.

[103] Y. Wang *et al.*, "A comparison of word embeddings for the biomedical natural language processing," (in eng), *J. Biomed. Inform.*, vol. 87, pp. 12-20, Nov 2018, doi: 10.1016/j.jbi.2018.09.008.

[104] T. Pham, T. Tran, D. Phung, and S. Venkatesh, "Predicting healthcare trajectories from medical records: A deep learning approach," *J. Biomed. Inform.*, vol. 69, pp. 218-229, 2017.

[105] A. Rumshisky *et al.*, "Predicting early psychiatric readmission with natural language processing of narrative discharge summaries," *Translational psychiatry*, vol. 6, no. 10, pp. e921-e921, 2016.

[106] D. Van Le, J. Montgomery, K. C. Kirkby, and J. Scanlan, "Risk prediction using natural language processing of electronic mental health records in an inpatient forensic psychiatry setting," *J. Biomed. Inform.*, vol. 86, pp. 49-58, 2018.

[107] C. Ye *et al.*, "Identification of elders at higher risk for fall with statewide electronic health records and a machine learning algorithm," *Int. J. Med. Inform.*, p. 104105, 2020.

[108] H. Nakatani, M. Nakao, H. Uchiyama, H. Toyoshiba, and C. Ochiai, "Predicting Inpatient Falls Using Natural Language Processing of Nursing Records Obtained From Japanese Electronic Medical Records: Case-Control Study," *JMIR Medical Informatics*, vol. 8, no. 4, p. e16970, 2020.

[109] A. Rajkomar *et al.*, "Scalable and accurate deep learning with electronic health records," *npj Digital Medicine*, vol. 1, no. 1, p. 18, 2018/05/08 2018, doi: 10.1038/s41746-018-0029-1.

[110] V. J. Zhu, T. D. Walker, R. W. Warren, P. B. Jenny, S. Meystre, and L. A. Lenert, "Identifying falls risk screenings not documented with administrative codes using natural language processing," in *AMIA annual symposium proceedings*, 2017, vol. 2017: American Medical Informatics Association, p. 1923.

[111] L. V. Ho, D. Ledbetter, M. Aczon, and R. Wetzel, "The Dependence of Machine Learning on Electronic Medical Record Quality," *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2017, pp. 883-891. [Online]. Available: http://europepmc.org/abstract/MED/29854155
https://europepmc.org/articles/PMC5977633
https://europepmc.org/articles/PMC5977633?pdf=render

[112] S. Wu *et al.*, "Deep learning in clinical natural language processing: a methodical review," *J. Am. Med. Inform. Assoc.*, vol. 27, no. 3, pp. 457-470, 2020.

[113] W. Yin, K. Kann, M. Yu, and H. Schütze, "Comparative study of cnn and rnn for natural language processing," *arXiv preprint arXiv:1702.01923*, 2017.

[114] A. E. Johnson *et al.*, "MIMIC-III, a freely accessible critical care database," (in eng), *Sci Data*, vol. 3, p. 160035, May 24 2016, doi: 10.1038/sdata.2016.35.

[115] K. L. C. Barajas and R. Akella, "Dynamically Modeling Patient's Health State from Electronic Medical Records: A Time Series Approach," presented at the Proceedings of the 21th ACM SIGKDD International Conference on Knowledge

Discovery and Data Mining, Sydney, NSW, Australia, 2015. [Online]. Available: https://doi.org/10.1145/2783258.2783289.

[116] M. Ghassemi *et al.*, "Unfolding physiological state: mortality modelling in intensive care units," presented at the Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, New York, New York, USA, 2014. [Online]. Available: https://doi.org/10.1145/2623330.2623742.

[117] E. Ford, J. A. Carroll, H. E. Smith, D. Scott, and J. A. Cassell, "Extracting information from the text of electronic medical records to improve case detection: a systematic review," *J. Am. Med. Inform. Assoc.*, vol. 23, no. 5, pp. 1007-1015, 2016, doi: 10.1093/jamia/ocv180.

[118] U. S. D. o. H. H. Services. "Health Information Privacy." https://www.hhs.gov/hipaa/index.html (accessed 06/20, 2020).

[119] "CyberSense.US." http://cybersense.us/CyberSense/Welcome.html (accessed 06/19, 2020).

[120] "PointClickCare Corp." https://pointclickcare.com/ (accessed 06/19, 2020).

[121] HHS.gov. "Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule: The De-identification Standard." https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#standard (accessed 07/07, 2021).

[122] C. I. f. H. Information. "Describing Outcome Scales for Resident Assessment Instrument 2.0 " https://www.cihi.ca/sites/default/files/document/outcome_rai-mds_2.0_en.pdf (accessed 09/19/2019, 2019).

[123] J. Ware, M. Kosinski, and S. Keller, "SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales," 01/01 1998.

[124] S. L. Colby and J. M. Ortman, "Projections of the size and composition of the US population: 2014 to 2060: Population estimates and projections," 2017.

[125] Y. Ostchega, T. B. Harris, R. Hirsch, V. L. Parsons, and R. Kington, "The prevalence of functional limitations and disability in older persons in the US: data from the National Health and Nutrition Examination Survey III," *J. Am. Geriatr. Soc.*, vol. 48, no. 9, pp. 1132-1135, 2000.

[126] T. R. Fried, E. H. Bradley, C. S. Williams, and M. E. Tinetti, "Functional disability and health care expenditures for older persons," *Arch. Intern. Med.*, vol. 161, no. 21, pp. 2602-2607, 2001.

[127] M. Gordon, *Manual of nursing diagnosis*. Jones & Bartlett Publishers, 2014.

[128] W. H. Organization, *World report on ageing and health*. World Health Organization, 2015.

[129] C. Milner. "Focus on Function for Healthy Aging." https://health.gov/news-archive/blog/2015/10/focus-on-function-for-healthy-aging/index.html (accessed 15 September, 2020).

[130] K. Rockwood *et al.*, "A global clinical measure of fitness and frailty in elderly people," (in eng), *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*, vol. 173, no. 5, pp. 489-495, 2005, doi: 10.1503/cmaj.050051.

[131] G. Santoni *et al.*, "Defining Health Trajectories in Older Adults With Five Clinical Indicators," (in eng), *J. Gerontol. A Biol. Sci. Med. Sci.*, vol. 72, no. 8, pp. 1123-1129, Aug 1 2017, doi: 10.1093/gerona/glw204.
[132] P. Mccllagh and J. Nelder, "Generalized linear models 2nd ed," ed: Chapman and Hall: New York, 1989.
[133] A. Agresti, *Categorical data analysis*. John Wiley & Sons, 2003.
[134] N. E. Breslow and D. G. Clayton, "Approximate inference in generalized linear mixed models," *Journal of the American statistical Association*, vol. 88, no. 421, pp. 9-25, 1993.
[135] "Introduction to Generalized Linear Mixed Models." https://stats.idre.ucla.edu/other/mult-pkg/introduction-to-generalized-linear-mixed-models/ (accessed 15 September, 2020).
[136] D. K. Richardson, J. E. Gray, M. C. McCormick, K. Workman, and D. A. Goldmann, "Score for Neonatal Acute Physiology: a physiologic severity index for neonatal intensive care," (in eng), *Pediatrics*, vol. 91, no. 3, pp. 617-23, Mar 1993.
[137] S. T. Adams and S. H. Leveson, "Clinical prediction rules," *BMJ*, vol. 344, p. d8312, 2012, doi: 10.1136/bmj.d8312.
[138] E. D. Boudreaux, J. Friedman, M. E. Chansky, and B. M. Baumann, "Emergency department patient satisfaction: examining the role of acuity," (in eng), *Acad. Emerg. Med.*, vol. 11, no. 2, pp. 162-8, Feb 2004.
[139] B. H. Cuthbertson and G. B. Smith, "A warning on early-warning scores!," *BJA: British Journal of Anaesthesia*, vol. 98, no. 6, pp. 704-706, 2007, doi: 10.1093/bja/aem121.
[140] W. A. Grobman and D. M. Stamilio, "Methods of clinical prediction," (in eng), *Am. J. Obstet. Gynecol.*, vol. 194, no. 3, pp. 888-94, Mar 2006, doi: 10.1016/j.ajog.2005.09.002.
[141] R. Lowry, *Concepts & Applications of Inferential Statistics, Chapter 14*. 2012.
[142] D. Bates, D. Sarkar, M. D. Bates, and L. Matrix, "The lme4 package," *R package version*, vol. 2, no. 1, p. 74, 2007.
[143] J. Allaire, "RStudio: integrated development environment for R," *Boston, MA*, vol. 537, p. 538, 2012.
[144] R. C. Team, "R: A language and environment for statistical computing," 2013.
[145] "Mixed Effects Logistic Regression | R Data Analysis Examples." https://stats.idre.ucla.edu/r/dae/mixed-effects-logistic-regression/ (accessed 15 September, 2020).
[146] "Applied Nonparametric Bootstrap with Hierarchical and Correlated Data " http://biostat.mc.vanderbilt.edu/wiki/Main/HowToBootstrapCorrelatedData (accessed 06/17, 2020).
[147] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in *Proceedings of the 9th Python in Science Conference*, 2010, vol. 57: Scipy, p. 61.
[148] E. Jones, T. Oliphant, and P. Peterson, "SciPy: Open source scientific tools for Python," no. 2019-10-01, 2001. [Online]. Available: http://www.scipy.org/
[149] J. M. Hausdorff, D. A. Rios, and H. K. Edelberg, "Gait variability and fall risk in community-living older adults: A 1-year prospective study," *Arch. Phys. Med. Rehabil.*, vol. 82, 2001, doi: 10.1053/apmr.2001.24893.

[150] B. H. Alexander, F. P. Rivara, and M. E. Wolf, "The cost and frequency of hospitalization for fall-related injuries in older adults," (in eng), *Am. J. Public Health*, vol. 82, no. 7, pp. 1020-3, Jul 1992, doi: 10.2105/ajph.82.7.1020.

[151] A. G. Society and B. G. Society, "Prevention of Falls in Older Persons: AGS/BGS Clinical Practice Guideline," 2009.

[152] M. M. Lusardi *et al.*, "Determining Risk of Falls in Community Dwelling Older Adults: A Systematic Review and Meta-analysis Using Posttest Probability," (in eng), *Journal of geriatric physical therapy (2001)*, vol. 40, no. 1, pp. 1-36, Jan/Mar 2017, doi: 10.1519/JPT.0000000000000099.

[153] S. Deandrea, E. Lucenteforte, F. Bravi, R. Foschi, C. La Vecchia, and E. Negri, "Risk Factors for Falls in Community-dwelling Older People: A Systematic Review and Meta-analysis," *Epidemiology*, vol. 21, no. 5, pp. 658-668, 2010, doi: 10.1097/EDE.0b013e3181e89905.

[154] C. E. Oshiro *et al.*, "Fall ascertainment and development of a risk prediction model using electronic medical records," *J. Am. Geriatr. Soc.*, vol. 67, no. 7, pp. 1417-1422, 2019.

[155] M. F. Folstein, S. E. Folstein, and P. R. McHugh, ""Mini-mental state". A practical method for grading the cognitive state of patients for the clinician," (in eng), *J. Psychiatr. Res.*, vol. 12, no. 3, pp. 189-98, Nov 1975, doi: 10.1016/0022-3956(75)90026-6.

[156] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," *Adv. Neural Inf. Process. Syst.*, vol. 30, pp. 4765-4774, 2017.

[157] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267-288, 1996.

[158] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Machine learning*, vol. 63, no. 1, pp. 3-42, 2006.

[159] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research*, vol. 12, pp. 2825-2830, 2011.

[160] S. M. Lundberg *et al.*, "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery," *Nature Biomedical Engineering*, vol. 2, no. 10, pp. 749-760, 2018/10/01 2018, doi: 10.1038/s41551-018-0304-0.

[161] C. f. D. C. a. Prevention. "Important Facts about Falls." https://www.cdc.gov/homeandrecreationalsafety/Falls/adultfalls.html (accessed 06/22, 2020).

[162] D. A. Sterling, J. A. O'connor, and J. Bonadies, "Geriatric falls: injury severity is high and disproportionate to mechanism," *Journal of Trauma and Acute Care Surgery*, vol. 50, no. 1, pp. 116-119, 2001.

[163] C. S. Florence, G. Bergen, A. Atherly, E. Burns, J. Stevens, and C. Drake, "Medical costs of fatal and nonfatal falls in older adults," *J. Am. Geriatr. Soc.*, vol. 66, no. 4, pp. 693-698, 2018.

[164] M. E. Tinetti, W. L. Liu, and E. B. Claus, "Predictors and prognosis of inability to get up after falls among elderly persons," (in eng), *JAMA*, vol. 269, no. 1, pp. 65-70, 1993/01// 1993, doi: 10.1001/jama.1993.03500010075035.

[165] N. Noury *et al.*, "Fall detection-principles and methods," in *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2007: IEEE, pp. 1663-1666.

[166] A. Bourke, J. O'brien, and G. Lyons, "Evaluation of a threshold-based tri-axial accelerometer fall detection algorithm," *Gait Posture*, vol. 26, no. 2, pp. 194-199, 2007.

[167] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818-2826.

[168] F. Chollet, "keras," *GitHub repository*, 2015. [Online]. Available: https://github.com/fchollet/keras.

[169] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, 2016, pp. 265-283.

[170] M. C. Hout, M. H. Papesh, and S. D. Goldinger, "Multidimensional scaling," *Wiley Interdiscip. Rev. Cogn. Sci.*, vol. 4, no. 1, pp. 93-103, 2013.

[171] M. Topaz *et al.*, "Mining fall-related information in clinical notes: Comparison of rule-based and novel word embedding-based machine learning approaches," *J. Biomed. Inform.*, vol. 90, p. 103103, 2019.

[172] I. Banerjee, S. Madhavan, R. E. Goldman, and D. L. Rubin, "Intelligent word embeddings of free-text radiology reports," in *AMIA Annual Symposium Proceedings*, 2017, vol. 2017: American Medical Informatics Association, p. 411.

[173] E. Loper and S. Bird, "NLTK: the natural language toolkit," *arXiv preprint cs/0205028*, 2002.

[174] A. L. Beam *et al.*, "Clinical concept embeddings learned from massive sources of multimodal medical data," *arXiv preprint arXiv:1804.01486*, 2018.

[175] J. P. Carvalho and S. Curto, "Fuzzy preprocessing of medical text annotations of intensive care units patients," in *2014 IEEE Conference on Norbert Wiener in the 21st Century (21CW)*, 2014: IEEE, pp. 1-7.

[176] Y. Zhang, Q. Chen, Z. Yang, H. Lin, and Z. Lu, "BioWordVec, improving biomedical word embeddings with subword information and MeSH," *Scientific data*, vol. 6, no. 1, pp. 1-9, 2019.

[177] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532-1543.

[178] A. Arcelus, I. Veledar, R. Goubran, F. Knoefel, H. Sveistrup, and M. Bilodeau, "Measurements of Sit-to-Stand Timing and Symmetry From Bed Pressure Sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 5, pp. 1732-1740, 2011, doi: 10.1109/TIM.2010.2089171.

[179] B. R. Greene and R. A. Kenny, "Assessment of Cognitive Decline Through Quantitative Analysis of the Timed Up and Go Test," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 4, pp. 988-995, 2012, doi: 10.1109/TBME.2011.2181844.

[180] J. Verghese, C. Wang, R. B. Lipton, R. Holtzer, and X. Xue, "Quantitative gait dysfunction and risk of cognitive decline and dementia," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 78, no. 9, pp. 929-935, 2007, doi: 10.1136/jnnp.2006.106914.

[181] S. Lord, B. Galna, S. Coleman, D. Burn, and L. Rochester, "Mild depressive symptoms are associated with gait impairment in early Parkinson's disease," *Mov. Disord.*, vol. 28, no. 5, pp. 634-639, 5 2013, doi: 10.1002/mds.25338.

[182] A. Alberdi *et al.*, "Smart Home-Based Prediction of Multidomain Symptoms Related to Alzheimer's Disease," (in eng), *IEEE J Biomed Health Inform*, vol. 22, no. 6, pp. 1720-1731, Nov 2018, doi: 10.1109/jbhi.2018.2798062.

[183] A. K. Mishra *et al.*, "Tracking personalized functional health in older adults using geriatric assessments," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, p. 270, 2020/10/20 2020, doi: 10.1186/s12911-020-01283-y.

VITA

Anup Kumar Mishra received the Bachelor of Technology degree in Electronics and Communication Engineering from Biju Patnaik University of Technology, Odisha, India in 2011 and the M.S. degree in Computer Engineering from the University of Missouri, Columbia in 2015. He worked as a graduate research assistant at the Center for Eldercare and Rehabilitation Technology at the University of Missouri, Columbia, MO U.S.A. from 2013 to 2020. His research interests include machine learning, pattern recognition, medical informatics, and computer vision.