Practical Learning Modeling Techniques with Personalized Actionable Intervention for In-the-field Prediction

A

Dissertation

Presented to

the faculty of the School of Engineering and Applied Science
University of Virginia

in partial fulfillment
of the requirements for the degree

Doctor of Philosophy

by

Nutta Homdee

May 2021

APPROVAL SHEET

This

Dissertation

is submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Author: Nutta Homdee

This Dissertation has been read and approved by the examing committee:

Advisor: John Lach

Advisor:

Committee Member: John A. Stankovic

Committee Member: Farzad Farnoud

Committee Member: Mehdi Boukhechba

Committee Member: Virginia LeBaron

Committee Member:

Committee Member:

Accepted for the School of Engineering and Applied Science:

Craig H. Benson, School of Engineering and Applied Science

May 2021

This page intentionally left blank

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincerest gratitude to my advisor Prof. John Lach for his guidance during my Ph.D. study. I'm very grateful for Prof. Lach's support and all of the opportunities he has provided. Without him, I would not have had the chance to collaborate with many clinical experts and experience engineering in medicine allowing my work to have real-world impact. Throughout this journey, Prof. Lach encouraged me to become more confident in doing research and taught me to be an independent researcher. He continues to be an inspiration and a role model as a professor for me. I am so fortunate to have him as my advisor.

I would also like to thank professors John Stankovic, Farzad Farnoud, Mehdi Boukhechba, and Virginia LeBaron for being my dissertation committee and spending the time to help me shape my research. I am grateful to have such a great thesis committee. Your feedback, comments, and suggestions are valuable in completing this dissertation.

I am so grateful for the opportunity to work with clinical researchers. I want to express my appreciation to Dr. Azziza Bankole and Dr. Martha Anderson who showed me many things about clinical research and clinical perspective especially about working with elderly persons, person-with-dementia, and in-home caregivers. I'd also like to thank Dr. Virginia LeBaron again for her positivity and motivation. She is a great collaborator who led me to see the potential impact of my work on real-world symptom management such as managing in-home, cancer pain.

I would also like to thank the graduate students and undergraduate students in the INERTIA group. I appreciate the research discussions that we had which helped me with my work. I enjoyed the time I spent with students in this group with whom I shared many laughs and lessons.

Many thanks to the Thai Royal Embassy who awarded me with a full scholarship giving me the opportunity to complete my Ph.D. research.

I am also grateful to my parents and my brother for the support through many international phone calls and video calls. I thank my family for being supportive in every step and decision throughout my life. I'd also like to thank the Thai students at the University of Virginia as well, for the frequent hangouts and conversations which helped me through graduate school. Lastly, I'd like to thank my girlfriend for all her support over the course of my Ph.D. Thank you for *literally* being by my side for every part of the Ph.D. journey, from the qualifying exam to the defense. You've helped me through the tough times. Thank you for always seeing the best in me and trusting me.

Thank you to everyone. Your support made this possible!

ABSTRACT

Many chronic diseases, if not well-managed, are major healthcare problems globally with two-third of the total US healthcare spending going toward chronic-related conditions in 2015. To avoid solely relying on expensive treatments, sensing systems and learning modeling techniques have been implemented in healthcare fields to focus on the prevention, early detection, and minimally invasive management of diseases in-the-field. As opposed to the old-fashioned method where a patient only sees a single point of data when he/she meets the clinician, the sensing system can continuously collect data to identify signs of disease anywhere and anytime. However, symptom management, especially informing the patient of symptom prevention and mitigation suggestions, using sensing and modeling systems faces many real-world problems such as weak symptom labeling and providing users actionable information. Thus, this work proposes three practical approaches addressing fundamental data challenges found in real-world sensing and learning modeling systems. The proposed approaches are also implemented in two real-world use cases: dementia caregiver empowerment and improving cancer pain management.

First, the implementation of machine learning models to time-series data collected in the real-world is presented. This work evaluates and implements different learning algorithms, as well as suggests techniques that can address real-world data problems such as imprecise user annotations, small training samples, and imbalanced label distributions. The results suggest that a machine learning algorithm can be applied to predict health events, dementia-related agitations, and cancer-pain episodes from ambient environmental stimuli.

To create appropriate prevention and mitigation suggestions (e.g. symptom intervention) based on the learning model's prediction, some interpretability of the model is needed. Unfortunately, many real-world applications use complex learning models such as deep learning for their prediction making model interpretation difficult. In this dissertation, a black-box model interpretation technique is presented using the already learned model, predictor importance analysis, and cross-correlation to learn and suggest actionable information to users in real-time. Real-world actionable suggestions can be

extracted such as notifying the in-home dementia caregiver to intervene before dementia-related agitation escalates by turning the lights on because the ambient light level is decreasing, which has triggered agitation episodes in the past.

User engagement and domain expert contribution is also important for many in-the-field applications, especially those that involve interventions. Here, a personalized intervention suggestion selection technique is presented that involves user engagement from surveys, self-reports, and sensing systems in combination with domain expert assessments to create scalable and automated prevention and mitigation intervention suggestions for real-world applications. This approach has been used to notify dementia caregivers with personalized agitation interventions. The results suggest that the interventions may help caregivers minimize the stress associated with dementia caregiving tasks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FREQUENTLY USED ABBREVIATIONS

BESI	Behavioral and Environmental Sensing and Intervention
CG	Caregiver
CNN	Convolutional Neural Network
FPR	False Positive Rate
IoT	Internet of Things
LSTM	Long-Short-Term-Memory
ML	Machine Learning
NN	Neural Network
PWD	Person with Dementia
TPR	True Positive Rate

1. INTRODUCTION

1.1 Motivations and objectives

Recently, one of the major healthcare challenges is the escalation in costs and incidences related to chronic diseases, such as cancer and cardiovascular diseases. A study in 2015 shows that two-thirds of the total healthcare spending in the US is from adults with chronic conditions [1]. People with three or more chronic conditions also tend to have higher healthcare costs due to more hospital stays. For example, people with chronic atrial fibrillation, a type of long-term abnormal heart rhythm, often are required to stay in clinics or hospitals setting to monitor symptoms and receive health care [2]. This higher cost of healthcare coupled with a rapidly aging society will lead to one of every three dollars spent in America paying for healthcare by 2040 [3]. Instead of solely relying on expensive chronic symptom treatments, to sustain the global healthcare system, many technologies have been developed to focus on prevention, early detection, and minimally invasive management of diseases [4]. Sensing technologies' advances can continuously monitor health in the real world, making them applicable as monitoring systems for chronic symptom occurrences.

Sensing technologies and systems such as body sensors networks and internet-of-things have been implemented in many health care areas such as remote health monitoring, chronic disease management, and elderly care. These technologies have the potential to reduce monitoring downtime, due to the continuous operation of these devices. The continuous sensing systems can constantly monitor for relevant health events compared to traditional methods which the health monitoring only happens when patients interact with the physicians [5]–[8]. However, the continuous monitoring of the sensing systems will also generate a large amount of time-series data. This masses amount of data needs to be studied and interpreted before we can extract useful information that would benefit the healthcare field [9]–[11]. The amount of generated data also makes it difficult for clinical experts to reliably search through all health-related data to manually extract the relevant health information. Fortunately, with advances in machine learning and modeling approaches, these techniques have been utilized with sensing technologies for detection and early diagnosis of symptoms and medical emergencies. This combination of sensing technology and learning model approach has been implemented in many healthcare applications such as atrial fibrillation detection, sports injury classification, and in-home elderly care [12]–[14].

For many healthcare applications, labeled data needs to be obtained and used together with learning models to assess, manage, or diagnose symptoms. Unfortunately, it could be difficult to obtain well-labeled training data in the real-world. Examples of problems found in real-world data are weakly labeled data due to human-annotated errors, small training samples, and unbalanced label distribution. Furthermore, to effectively and noninvasively manage chronic diseases in the real-world, personalized healthcare approaches have received increasing attention in the last decade [15]. Personalized healthcare can be achieved by involving patients and clinicians in the planning, development, and analysis of symptoms. Data-driven and continuous sensing technology can play a major role in the emergence of personalized healthcare such as explaining and monitoring for risk factors for disease symptoms. However, the interpretation of symptom's risks or triggers can be so complex to analyze in real-time (in time to interact with patients). Also, many real-world personalized healthcare technologies often ignore the obtrusiveness and privacy invasiveness aspect. For example, Dem@care (Dementia Ambient Care) uses video cameras and wearables to record the daily activities of elderly persons living at home [16]. The records are used to assess the cognitive impairment of the user to provide personalized healthcare. The usage of video cameras raises privacy concerns which contribute to users' reluctance to use the technology. For sensing technology-based healthcare, research evidence has shown that patient and clinician engagement is still limited [10].

The objective of this work is to bring together the monitoring system, the data, and the people (e.g. patients and clinical experts). A monitoring system can continuously collect data which helps with capturing signs of disease anywhere and anytime instead of the traditional method where only one data point is generated when patient and clinician meet. Accordingly, the monitoring system will generate masses amount of data. Especially, with the rise of smart devices and the internet-of-things, more and more devices will be able to generate data [17]–[20]. To take advantage of all these data, this work utilizes data processing data managing and machine learning techniques to extract useful and relevant information. Challenges found in real-world data will also be discussed such as the temporally imprecise data labels due to self-report data, a small number of observations to train machine learning models, and unbalanced label distribution. This work will provide practical solutions to these challenges. Apart from the processing and modeling techniques for real-world data, interpretation is also an important part. This work aims to create a model and data interpretation technique to learn input triggers that impact the prediction and help the user make a decision and take action based on the interpretation. The model interpretation technique is built to work with complex models (e.g. black-box learning models). This actionable interpretation is particularly useful in applications that benefit from an early intervention such as disease symptom management. This work also aims to enhance the intervention aspect by incorporate patients and clinicians in the development and delivery of personalized symptom intervention strategies.

Despite the increased usage of smart devices and sensing devices (e.g. mobile phones, smart watches, and smart-home gadgets) in the last decade, the potential of using these devices to capturing signs of health events, anywhere and anytime, has not been utilized [21]–[23]. A study shows that one of the reasons users abandon their smart devices is that the data collected by the devices were perceived to not be useful [24]. The work done in this dissertation is aimed to encourage the adoption of sensing devices/systems and data analytic techniques to make sense of the data and assist in the healthcare and health monitoring field. Following is the thesis statement of this dissertation: *Smart health systems, machine learning, and data interpretation techniques can help predict symptoms of in-home health events and create automated, actionable, and scalable intervention suggestions for real-world applications.*

1.2 Dissertation outline

The remaining chapters of this dissertation are organized as follows. Chapter 2 introduces the real-world health monitoring system called behavioral and environmental sensing and intervention (BESI) and two application case-studies. The first case-study, BESI for dementia caregiver empowerment, aims to assist dementia caregivers, who live with person-with-dementia, handling dementia-related agitation events. Dementia agitation is inappropriate verbal or physical behavior expressed by the person-with-dementia [25]. The BESI system learns to predict those behaviors and informs the caregivers of early signs of agitation before it escalates. BESI is implemented in another case-study that aims to improve cancer pain management. Most cancer pain symptom management occurs in homes of the cancer-patients. Usually, family caregivers are the ones helping with the management of cancer pain. The BESI system is deployed at homes of the cancer-patients to learn information about cancer pain at home and help to manage pain such as predicting upcoming cancer-pain episodes based on the ambient environment. The BESI system is deployed in actual patients' houses collecting time-series data related to the health events.

Chapter 3 presents the implementation of machine learning models to time-series data collected in the real-world. This chapter demonstrates techniques for extracting data segments from real-world time-series data and techniques for transforming the data for machine learning modeling. Machine learning basics and algorithms for time-series data are presented in this chapter as well as relevance evaluation metrics for choosing appropriate learning models. Lastly, the result of using a machine learning model to predict health events in the two case-studies is presented.

Chapter 4 discusses data challenges associated with the healthcare approach which combines sensing technology in real-world, personalized healthcare, and real-world learning modeling. To minimize the invasiveness associate with video camera usage, a real-world symptom sensing system can involve patient engagement by utilizing the self-annotation approach. For example, patient with chronic cancer pain can self-report their symptom occurrences which can be used as training data labels for cancer pain prediction applications [26]. One of the complications with using this self-annotation approach could be the weak training labels (e.g. timestamp of cancer pain event is temporally imprecise). The imprecision can occur due to the real-world difficulty of the patient dealing with the occurring symptom (e.g. cancer patient taking pain medication) and cannot report the symptom event immediately. Here, a boosting convolutional neural network is implemented with regularization and weighted sample technique to address the following in-the-field data challenges: temporally imprecise training labels, small training samples, and unbalanced label distribution. The goal of this learning framework approach is to create a robust model that can overcome mentioned real-world challenges such as model overfitting [27], [28].

Chapter 5 presents a model interpretation for real-time actionable intervention. Sensing systems and prediction models should also be leveraged for disease prevention strategies for better real-world symptom management [29]. The predictive model interpretation techniques can be used to learn symptom triggers, and inform users to avoid certain triggers. However, many machine learning models used to detect/diagnose symptoms lack interpretability due to the complexity of the learning models. This can make it difficult for the user to take action or make a decision based on the model result, especially in healthcare applications where decisions can come with severe consequences. In this chapter, a black-box model interpretation method to learn symptom triggers based on the already learned model, predictor importance analysis, and cross-correlation technique is presented. This approach is designed to be used in the real-time scenario to match its need for just-in-time intervention suggestions.

Chapter 6 shows a personalized symptom intervention strategy selection that involves both patient engagement and clinical expert assessments. The method combines clinical assessment tools, self-report surveys from the user (for user engagement), and a rule-based algorithm to customizes a list of personalized symptom intervention strategies. This method is used in a real-world application, in-home dementia caregiver empowerment, where personalized intervention lists are delivered to dementia caregivers via the sensing and intervention system presented in this dissertation.

Lastly, this dissertation closes with a summary, thoughts, and discussions of the future of the monitoring systems and data interpretations in Chapter 7.

1.3 Dissertation Contributions

This dissertation work includes implementations of data processing, data transformations, and data modeling techniques on time-series data. These techniques extract relevant information from the sensor data, based in part on the users' self-reported labels, using machine learning models to detect/predict health events and suggesting actionable strategies to users to prevent unwanted outcomes. The work is performed within two case-studies – the detection and prediction of dementia-related agitation and cancer-related pain from wearable and in-home ambient environment data. Procedures for implementing predictive machine learning models to real-world time-series data are shown. This consists of the time-series data filtering, usage of user-report data labels and data segmentation, data transformation, and relevance feature extractions, and structuring of machine learning models for time-series data.

While this work is implemented in healthcare-related studies, it can be generalized for many other real-world applications that use sensor systems to continuously collect data. These applications will also benefit from the presented data modeling and interpretation techniques to extract useful and relevant information to the users who can take action or make a decision based on the provided information. The presented implementations and evaluations of machine learning models to make predictions from time-series data can be used to help researchers build a learning model for other applications. Examples of these applications are the smart-devices for real-world infrastructure monitoring and the sensors for fault detection and prevention [30], [31].

The following are the contributions of this dissertation:

1. The implementation and evaluation of machine learning models for the prediction of health events from time-series data obtained from real-world sensor systems. Time-series data from sensors is often paired with temporal machine learning models. However, the complex models are not robust to real-world data problems, including imprecise labels from user annotations, small training samples, and imbalanced datasets. This work evaluates and implements different learning algorithms, as well as suggests techniques that can enhance real-world learning tasks. This information can aid researchers in the selection and implementation of appropriate machine learning algorithms for their studies.

2. A novel black-box model and data interpretation that uncovers actionable information from the previously learned model, predictor importance analysis, and cross-correlation to suggest actions to the user in real-time. An example of an actionable item, in the dementia case study, could be to notify the in-home dementia caregiver to intervene before agitation escalates by turning the lights on because the ambient light level is decreasing (e.g. the sun is setting), which has triggered agitation episodes in the past.
3. A rule-based algorithm for generating an appropriate intervention plan based on domain experts' assessments and user engagement via self-report questionnaires. The algorithm utilizes experts' assessments to identify problematic areas that required intervention and uses users' reports to personalize the intervention plan. This algorithm has been used to create personalized dementia-related agitation intervention suggestions for in-home dementia caregivers.

The techniques presented in this dissertation are implemented in real-world healthcare applications, dementia-caregiver empowerment and cancer-pain management, to show the contribution in the healthcare field such as prevention, early detection, and minimally managing symptoms. The clinical contributions of this work are:

1. An in-home symptom prediction using sensing system and predictive modeling approach. Here, cancer-pain episodes are predicted by learning from the patients' pain reports and the ambient environment surrounding the patients.
2. A dementia-related agitation prediction based on the in-home ambient environment.
3. A real-time dementia-related agitation intervention suggestion based on the ambient environment triggers using the model interpretation which examines the environmental time-series that impact the agitation prediction.
4. An automated and personalized dementia-related agitation intervention suggestions based on dementia symptom severity and patient/caregivers' reports. The intervention suggestions have been used to assist in-home dementia caregivers living with person-with-dementia. The approach can minimize or

eliminates the need for on-call clinical experts for scalable real-world symptom sensing and intervention systems.

2. CASE STUDY: BEHAVIORAL AND ENVIRONMENTAL SENSING AND INTERVENTION (BESI)

This chapter presents the Behavioral and Environmental Sensing and Intervention (BESI) system. BESI is an integrative sensing, analytics, modeling, and intervention system that learns/detects the early symptom signs and notifies the user of intervention suggestions [16]. The development of the BESI system is a collaborative work by graduate students of the Department of Electrical and Computer Engineering, University of Virginia. My main contributions here are on-node processing of the sensing nodes, data pipeline for real-time data modeling and intervention suggestion, notification delivery system, cloud services, intervention suggestion systems, system fault-tolerant programs, and real-world data collection deployments. In this section, the BESI system and real-world data are described alongside two real-world applications that the system is deployed on. The BESI system is designed to remotely sense behavioral and environmental context in a residential area, detects potential early stages of health symptoms such as dementia-related agitations and cancer-pain episodes, and sends personalized intervention suggestions to the users (e.g. in-home dementia caregivers) in a timely manner. The BESI system is deployed in actual homes of dementia-patients and cancer-patients living with in-home caregivers, who often are the family member of the patients. During each sensing system deployment, BESI continuously monitors for ambient environmental stimuli that correlate with the occurrences of the health events (dementia-related agitations, cancer-pain episodes). Figure 2.1 shows an overview of the system which consists of remote sensing nodes, a processing hub, and user interfacing devices.

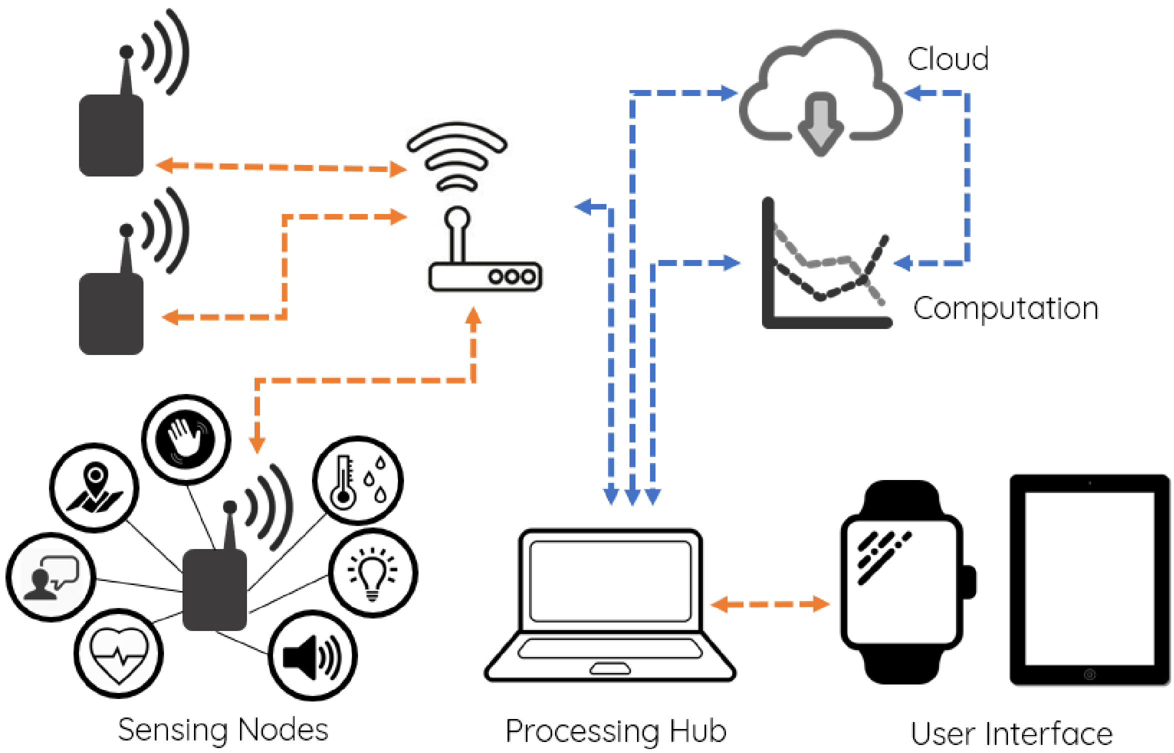

Figure 2.1. BESI system architecture consists of the sensing nodes, processing hub, and user interfacing device(s)

The sensing nodes are mounted with room-level environmental sensors that collect light level, in-home temperature, humidity, air pressure, and ambient noise level. All environmental data is sequentially sampled once per second (1 Hz) except for the ambient noise level which is sampled at 10 Hz then averaged down to 8 Hz to remove any distinguishable conversation. BESI system also deploys smartwatches to sense the physical motion of the users. The in-home room-level location of the user is monitored using doorway sensors and a monitoring algorithm [32]. The sensor data are pre-processed and transferred to the central processing hub via socket programming [33]. The processing hub, also called the base-station, is a laptop deployed with the rest of the components. The tasks performed by the hub are: (1) monitors the sensing nodes; (2) runs symptom predictive models; and (3) communicates with cloud services for intervention delivery. The processing station connects with all deployed sensing nodes via a local wireless network provided by a wireless router and runs a program that ensures all sensing nodes are running as intended. This network also used by the processing station to receive behavioral and environmental context from the sensing nodes. Based on the data, predictive models are used to monitor for upcoming symptoms. The base-station also notifies the user and provides intervention strategies in real-time. The base-station has internet access for communicating with user interfacing devices and cloud services. Lastly, the BESI system acquires the non-sensor information regarding the in-home symptom events and symptom-related information via the user inputs. The system relies on these user inputs (e.g. symptom event timestamp, patient behavior) as training labels to learn symptom prediction models. The interfacing devices are also utilized for notifying and delivering personalized symptom intervention to the users.

2.1 BESI for dementia caregiver empowerment

The proposed approaches, which aim to predict patients' symptoms, learn symptom triggers, and suggest personalized interventions, are implemented to learn in-home ambient environment triggers for dementia-related agitation. Persons with dementia (PWD) often suffer from cognitive impairments which often lead to difficult daily lives. The difficulty related to the impairments can cause the PWD to express agitated behavior such as verbal outbursts, or aggressive motor behaviors which can be called dementia-related agitation [19]. The occurrence of agitation can be unpredictable or be affected by the ambient environment around PWD [20]. Many studies have shown the causation between the ambient environment and PWD. The College of Nursing, University of Wisconsin-Milwaukee, suggests that dementia agitation has a significant correlation with the level of sound in the environment (in this case, nursing homes) [21]. Another study by Van Hoof observed that high-intensity bright light can affect restlessness behavior and increase the frequency of agitation occurrences of institutionalized older adults with dementia [22]. The BESI system is used to collect ambient environmental data in the homes of PWD using our developed integrative sensing system; through a collaboration with dementia experts from the Virginia Tech Carilion School of Medicine [23]. The system has been deployed to collect data in real dementia patient and caregiver (CG) homes. Each takes two months during which the system passively collects the environmental data. To gather information about the occurrences of the in-home dementia-related agitation, CGs use a tablet survey application to label each agitation episode. In the dementia case-study, the BESI system has been deployed in five 2-month-long deployments. On average, we received 6 agitation labels per week or around 48 agitation labels for each deployment. Table 2.1 shows characteristics of participants and deployments in the dementia case-study. In this table, the Cohen-Mansfield Agitation Inventory (CMAI-C) is shown [25]. The CMAI-C is a clinical assessment tool that measures dementia-related behavior frequency and occurrence. The higher the score refers to more agitated behavior occurrence as well as more frequency. This assessment tool is a standard measurement for assessing agitation in dementia.

Table 2.1. Participant and deployment characteristics of the dementia case-study

Deployment #	Age, Gender	CMAI-C score (1-232)	Days of sensor deployment	Number of reported agitations	Average agitation severity reported (1-10)
1	83, Male	132	74 days	81	7.18
2	77, Female	65	78 days	16	6.00
3	71, Female	52.5	59 days	48	2.16
4	85, Male	57	68 days	24	6.19
5	81, Female	117	57 days	73	4.12

In the dementia case-study, a smart watch and a tablet are used as caregiver interfacing devices. CGs are encouraged to use a tablet survey application to record information about each PWD agitation episode. The application can record the time of agitation, agitation intensity, PWD behavior during the agitation, recent PWD activity, and PWD mood. This agitation information will be based on the CG inputs which enable the system to learn agitated behavior and contextual triggers specifically of each PWD-CG dyad. The tablet application also allows CGs to input their mood and stress which we use to assess CGs' burden level. The information is securely stored in the cloud server. We also designed a second method of marking an agitation event to increase the reliability of the agitation event ground truth. In cases that PWD agitation requires the full attention of the CG, the agitation timestamp can be easily marked with one click on the CG watch. This allows CGs to handle the agitation, then fill in a more detailed agitation report via the tablet application later without the burden to remember the exact time of the episode. Once the predictive model on the base-station detects a potential agitation behavior, the base-station sends a notification to the CG watch, which buzzes and alerts the CG. At the same time, the base-station uses the Firebase cloud service to pull personalized agitation interventions from the cloud server and display them on the tablet. The intervention suggestions are chosen, by clinical experts, specifically for each CG-PWD dyad based on the agitation pattern of the PWD. Figure 2.2 shows the dementia caregivers interfacing devices and examples of the dementia-related agitation notifications.

Figure 2.2. User interfacing devices for dementia caregivers

2.2 BESI for improving cancer pain management

The sensing system, BESI, is deployed in the homes of cancer patients to measure and describe variables relevant to cancer pain in the home setting. In the management of cancer disease, 60%-90% of cancer patients experience moderate to severe pain [24]. One type of cancer pain is breakthrough pain which is sudden, unpredictable, and increases in pain [25]. This difficult pain, if not early managed, can escalate and increase distress for both patients and caregivers. Furthermore, most cancer symptom management happens in home settings [34]. Normally, the family caregivers assist the cancer patients in managing the pain symptom. In-home cancer symptom management is the main factor in helping and assisting patients with cancer. However, the family caregivers often have limited information and supporting tools which cause emotional distress in caregiving tasks [35]. The smart health systems, such as mobile technology and sensing system, can be utilized to improve symptom monitoring and management [36]. But despite the increased adoption of the smart sensing system and smart devices which generate a many types of data, the relationship between the data and in-home cancer-pain has not been studied, such as ambient environmental triggers of cancer-pain episodes [37].

In the cancer pain management case-study, the BESI system is deployed in five 2-week-long deployments in homes of cancer patients from an academic medical center outpatient palliative care clinic [26]. The system continuously collects the patient's physiological data (e.g. motion, heart rate) using smartwatch devices and room-level ambient environmental data (light, temperature, humidity, air pressure, noise level) using the sensor nodes. Cancer pain events can be marked by the patient and caregiver via the smartwatch device. Marking a pain event will also allow users to report information regarding the events such as describing the pain such as the level of pain. The collected system data and self-report pain markers will be used to infer and detect cancer pain events. Figure 2.3 below shows the sensing station used for collecting in-home ambient environmental data surrounding the patients and the smartwatches with user interfacing questionnaires for labeling occurrences of cancer-pain episodes.

(a.) (b.)

Figure 2.3. (a.) Sensing station for in-home ambient environmental data collection. (b.) Smartwatches with example user interfacing questionnaires.

In the cancer management case-study, 5 dyads (cancer patient and their family caregiver pair) participated in the BESI deployments. Participant characteristics are as follow: One patient is between 45–54-year-old, two patients are between 55-64, and another two patients are 65-74 years of age; Three patients are male and two are female; Three patients have cancer diagnosis of head and neck cancer, one has diagnosed with colorectal cancer, and the other patient has diagnosed with lung cancer. On average, each sensor system deployment has 46 cancer-pain episodes reported with average pain severity of 5.29 out of 10.

2.3 Real-world data

In this section, figures of the data collected in real-world are shown. Figure 2.4 shows an example period of the collected ambient environmental data from the dementia case-study and the CG-provided agitation label. In these deployments, once the sensor patterns associated with PWD agitation are learned, just-in-time intervention suggestions are sent to CG via user interfacing devices.

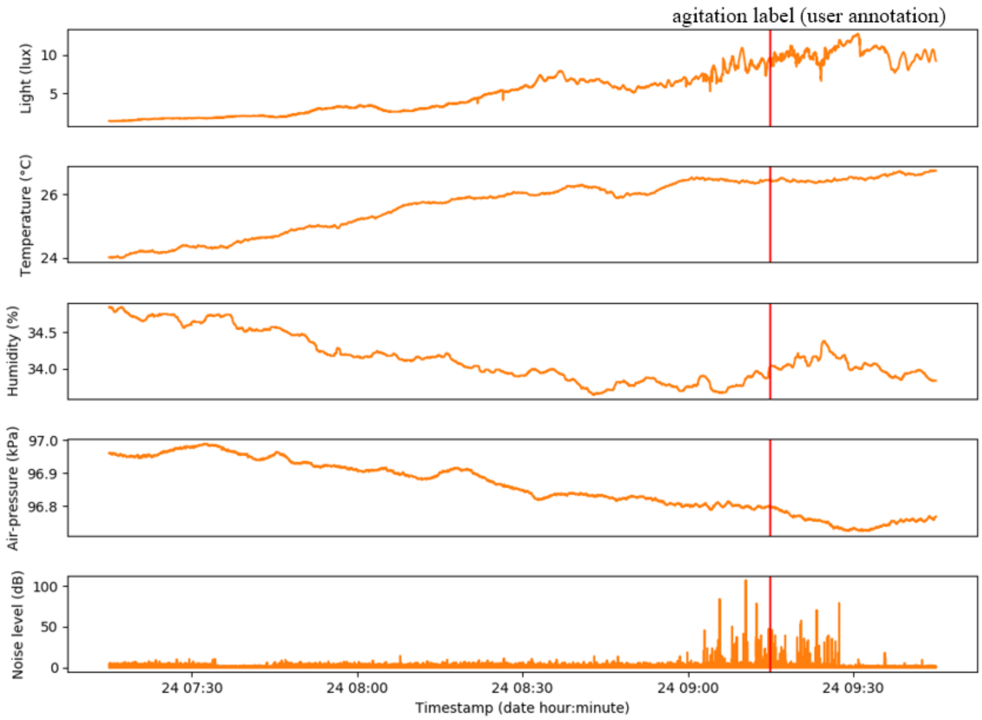

Figure 2.4. A period of ambient environmental data collected by the sensor node at the PWD home. The ambient environmental data, from top to bottom, are (from top to bottom) light level (lux), ambient temperature (°C), in-door humidity level (% relative humidity), air-pressure (kPa), and ambient noise level (dB). The red vertical line shows a dementia-related agitation event label provided by the PWD's CG.

These ambient environmental time-series data are used with machine learning models to predict upcoming health events based on the observation label from users. For example, if certain behavior of the ambient light level, such as decreasing light level in the evening, has shown to be followed by dementia-related agitation episodes, the learning model will learn to predict future agitation episodes based on these triggers. To show the relationship between the in-home ambient environment and the occurrences of health events (dementia-related agitation, cancer-pain episodes), Figure 2.5 shows the comparison between collected ambient environmental data and the number of cancer pain reported from the patient each day. From the data of the deployment shown in figure 2.5, the daily air-pressure value is correlated with the number of cancer-pain episode reports per day, with the Pearson correlation coefficient of 0.60 [38].

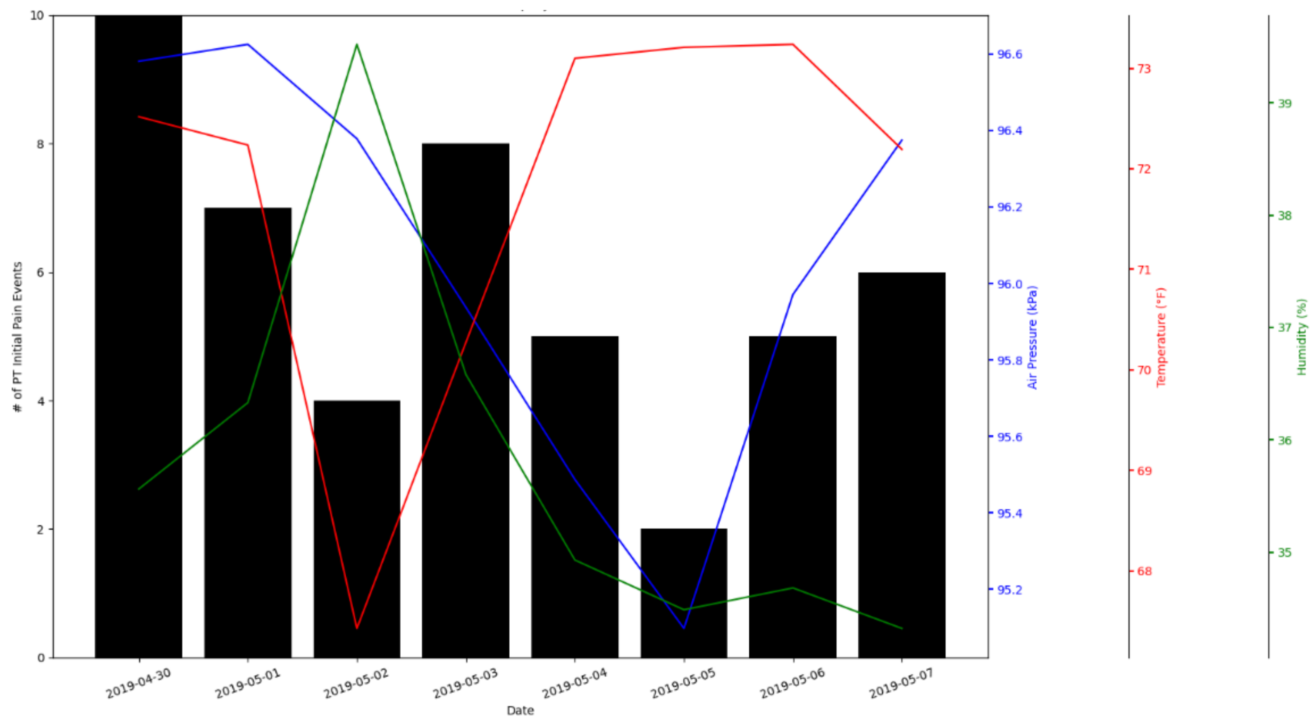

Figure 2.5. Number of the daily reported cancer-pain episodes and the daily value of the in-home ambient environmental data

The time-of-the-day is also used as one of the health event predictors. For dementia-related agitation, many patients can frequently display restlessness behaviors during a similar time-of-day. This symptom is called Sundowning syndrome [39]. The time-of-day also shows a correlation relationship with the reported cancer-pain episodes. Figure 2.6 illustrates all cancer-pain reports in one deployment with their timestamps. In this figure, the time-of-day is plotted at the vertical axis from the beginning of the day on the top of the axis to the end of the day at the bottom of the axis. The dotted lines show common time-of-day which cancer-pain has been frequently reported. The common timestamps are computed using k-mean clustering of all cancer-pain timestamps [40]. In figure 2.6, 75% of all cancer-pain reports occur within one standard deviation window of the three common timestamps.

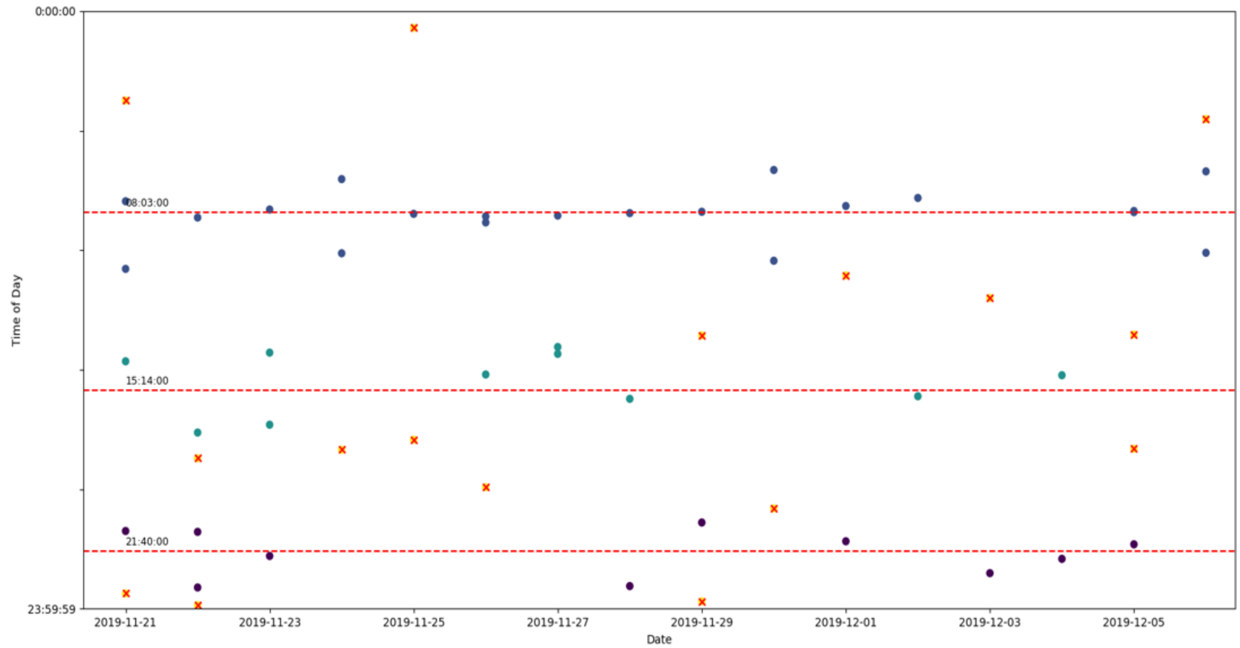

Figure 2.6. Reported cancer-pain episodes throughout one sensor deployment and their timestamps. The vertical axis shows the time of the day, from mid-night (top) to the next day (bottom). The dotted lines show the common time-of-day that cancer-pain episodes have been reported. The x-marked episodes are the outliner timestamps (more than one standard deviation away from the three common times).

3. TIME-SERIES DATA AND HEALTH EVENT PREDICTION

Over the past two decades, machine learning has progressed significantly. Machine learning has shown successes in many fields of technology and science, such as finances, speech processing, image classification, and health care [41]. Also, there is a breakthrough in deep learning systems that outperform not only classical methods but also human benchmarks in many tasks [42]. A machine learning algorithm is an algorithm that learns from data; and is defined as *“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E* [43]*.”* This chapter presents the procedures of implementing predictive machine learning models from time-series data. Predictive machine learning is a learning algorithm that focuses on classification tasks in which the program is asked to specify which categories some inputs belong to. Accordingly, this chapter will focus on supervised learning algorithms. Supervised learning algorithms need labeled data (also called dependent variable or response in statistical terminology) for training [44]. Figure 3.1 shows an overview of the procedures for predictive machine learning tasks (e.g. health event prediction) with time-series data input. Usually, sensor systems or smart devices generate time-series data. The time-series data can be labeled with the event of interest. For example, in the dementia case study, ambient environmental time-series data is labeled by the dementia-caregivers when they observed agitated behavior from the person-with-dementia. The data labels are used to segment periods of relevant data. Then, the segmented data is feature extracted and used as inputs to machine learning models. The models learn to make predictions from the training data and compared the model’s predictions to the actual labels. Finally, the results of the comparison are used to improve the model.

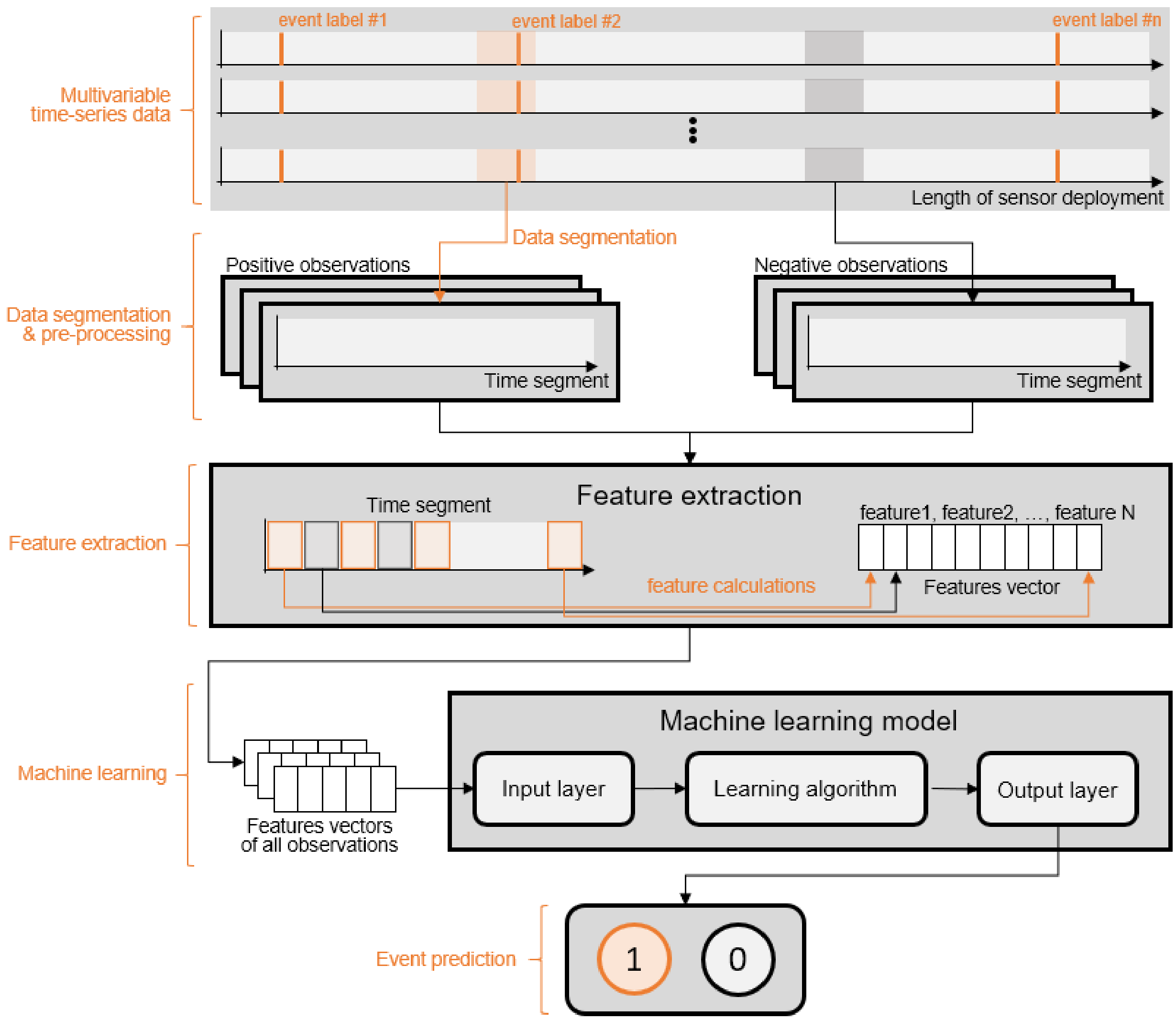

Figure 3.1. An overview of machine learning structure for event prediction of time-series data.

3.1 Related works

Many machine learning models have been implemented to solve tasks with time-series data [45]. In this section, types of machine learning algorithms that have been used with time-series classification/prediction are explored. A study uses a neural network (NN) machine learning model to predict wind speed for renewable energy management [46]. Since the neural network model takes numerical data as inputs, the study extracts wavelet decomposition features from the time-series wind-speed data. The features are used as inputs to the neural network model to predict hourly wind speed. One disadvantage of using NN models with time-series prediction is that NN models do not learn temporal information in the time-series inputs. In many applications dealing with time-series data, the temporal relationship can be very important such as speech recognitions, natural language processing, or activity recognition from human motion [47]–[49]. To include the temporal dependency in the time-series data, a long-short-term-memory (LSTM) neural network model has been introduced [50]. The LSTM neural network models have been implemented and show success in prediction/classification tasks with time-series data [51]–[53]. LSTM models take the temporal information in the time-series data into account by implementing memory states which retain temporal information in adjacent inputs if they are relevant [50]. Recently, boosting models have gained increasing usage and success for real-world prediction tasks [54]–[56]. The boosting model works by converting multiple weak learners to a strong model which creates a more robust learning model [57]. However, similar to the neural network model, the boosting models do not consider the temporal relationship in time-series data. Lastly, convolutional neural network (CNN) models have been implemented for time-series classification and prediction tasks [58]. CNN models are widely used for image processing learning problems [59], [60]. Thus, many studies have taken the well-defined CNN parameters and implemented them to time-series classification tasks [61], [62]. CNN models capture the temporal behavior of time-series data at the convolutional layer which slides convolutional filters along the time-series data to find relevance patterns [42]. While LSTM and CNN take in temporal information in time-series data into their learning algorithm, both models are complex models that required a vast amount of training parameters which could cause performance issues such as model overfitting. This chapter will explore each machine learning model and compare their performance on the case-studies tasks.

Sensor data collected in the real-world is rarely in a ready form for machine learning models. Thus, data pre-processing, data transformation, and data filtering tasks are often needed. Traditional time-series prediction models use segmentation technique [63]. The technique chooses a time-series period with training labels (e.g. 20-minutes around the label), divides the time-series period into segments, then extracts features from each segment. These features, such as statistical mean, variance, are used as inputs for a learning model. When the traditional time-series prediction models are trained with temporally imprecise training labels, it can mislead the learning model and result in lower prediction performance. Moreover, using the time-series segmentation and feature extraction required a significant amount of task-specific knowledge, e.g. choosing the input features related to the task.

3.2 Machine Learning and Time-series Data

This section describes the basics of machine learning models and their implementations. These models can be used in classification and prediction tasks with time-series data.

3.2.1 Neural network

Artificial neural networks are machine learning models inspired by the human nervous system which consists of electrochemical stimulation receivers (dendrites), responsive cells which release neurotransmitters, and output terminals called axon terminals which fire neurotransmitters to the connected neurons [64]. Similarly, artificial neural networks imitate human interconnected neuron networks. One of the simplest forms of artificial neural networks is shown in figure 3.2 which consists of an input layer, a hidden layer, and an output layer. A neural network unit, called a perceptron, consists of the input signal $x_0, x_1, \dots, x_j$ that are combined with weights $w_0, w_1, \dots, w_j$ as shown in figure 3.2.

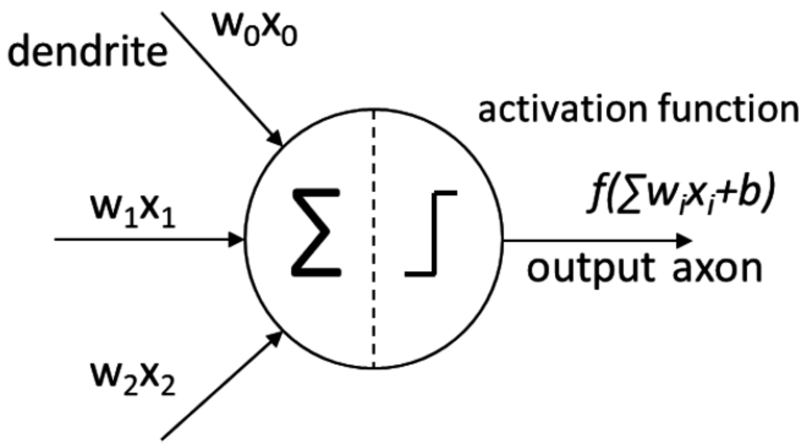

Figure 3.2. Artificial neural network and Perceptron [44].

These inputs can be observations or data or values from connected perceptrons (in the case of a network with multiple hidden layers). The goal of an artificial neural network is to learn an output function f which maps an input to output by adjusting the weights of each input $w_0x_0, w_1x_1, \ldots, w_jx_j$ in the weight matrix. The weighted inputs are summed up, add a bias constant b, and passed to an activation function which represents the action potential of a real neuron. There are different types of activation functions [65]. For example, the simplest activation function could be a step function that either fires or not fires depending on the value of the summed weighted inputs and the bias. The output of a perceptron with activation function f is defined by equation 3.1.

$$\hat{y}(x_i) = f\left(w_0x_0, w_1x_1, \ldots, w_jx_j + b\right) \tag{3.1}$$

where:

$\hat{y}(x_i)$ = Output of i^{th} perceptron

w_j = Weight of the j^{th} input of the i^{th} perceptron

b = Bias

Each input signal has its weight that is individually adjusted. The weights are adjusted during training until the desired result is obtained (e.g. least error at the output). The weight matrix, which represents the relationship between inputs and outputs, is learned during training by using a weight update formula shown below.

$$w_j^{k+1} = w_j^k + \lambda(y_i - \hat{y}_i^k)x_{ij} \tag{3.2}$$

where:

w^k	=	Weight for i^{th} input after k^{th} iteration
λ	=	Learning rate
$\hat{y}_i^k$	=	Predicted output
x_{ij}	=	Value of j^{th} attribute of the training instance

The weights are usually initially set with random values and adjusted based on the prediction error (calculated as the differences between actual output and predicted output $(y - \hat{y})$. The updated weight is done according to equation 3.2. The learning rate λ is a value between 0 and 1. The learning rate determines how much of the weights will be adjusted for each learning iteration. The learning algorithm keeps learning for many iterations until the output is converging (i.e. the output can classify training instances with no or little errors).

3.2.2 LSTM

Long-short-term-memory model (LSTM) is a neural network that is designed to learn problems with long-term dependencies (e.g. Natural language processing, where the next dialog depends on previous texts or conversations). In 1997, Hochreiter and Schmidhuber designed the LSTM neural network which consists of a set of connected neural networks known as memory blocks as shown in figure 3.3 [50]. An LSTM block involves the current block's input and output signal $x^{(t)}$ and $y^{(t)}$, the previous block's output $y^{(t-1)}$, and the gates (a neural network with activation function).

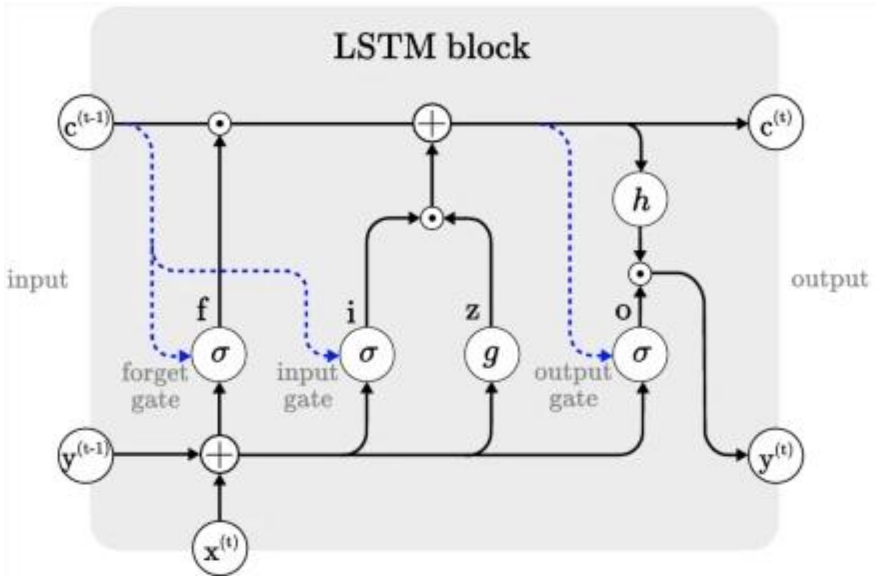

Figure 3.3. Long-short-term-memory (LSTM) block [66]

The key difference between regular neural networks and the LSTM gates are neural networks that combine current input $x^{(t)}$ and the output of the last LSTM block iteration $y^{(t-1)}$. This encourages the learning network to take any sequential behavior of the input data into account. This is defined in the equation below:

$$z^{(t)} = g\left(W^k x^{(t)} + R^k y^{(t-1)} + b\right) \tag{3.3}$$

where:

$z^{(t)}$	=	output of the gate
g	=	activation function
W^k, R^k	=	weights associated the input x(t)and the previous output y(t-1)
b	=	bias

Another component of an LSTM block is a cell state $c^{(t)}$ (and the last LSTM iteration cell state $c^{(t-1)}$). In the previous LSTM layer, the LSTM determines which information should be retained in the next LSTM block and stores the information in the cell state. The LSTM network will learn to add, remove, or modify the cell state depending on the training and the learning problem. This cell state also has an impact on the LSTM gates as shown in the input gate equation below Where $\odot$ represents point-wise multiplication of two vectors, and W^k, R^k, p^k are the weights associated with $x^{(t)}$, $y^{(t-1)}$, and the previous cell state $c^{(t-1)}$:

$$i^{(t)} = \sigma\left(W^k x^{(t)} + R^k y^{(t-1)} + p^k \odot c^{(t-1)} + b\right) \tag{3.4}$$

The LSTM gates, neural network layer(s) with an activation function, are input gate, forget gate, and output gate. The input gate combines the current input $x^{(t)}$ with the last iteration output $y^{(t-1)}$ and cell value $c^{(t-1)}$. The forget gate determines which information should be removed from the previous cell state $c^{(t-1)}$. Lastly, the output gate calculates the output of the LSTM block by combining the current input, output from the previous LSTM block, and cell value.

The LSTM neural networks have been implemented and shown successes in time-series prediction problems due to the temporal sequences in data [67].

3.2.3 Convolutional neural networks

The convolutional neural networks (CNN) have become very popular for learning problems associated with image processing or image classification tasks where there are patterns in many areas of the inputs [59]. CNN is inspired by how humans take in visual information by learning patterns. CNN learns patterns in the input data by creating convolutional filters. Let the input data be denoted by f and the convolutional filters (or kernel) by g, and let f has length n, and g has length m. The filter performs convolution operation as shown in the following formula:

$$(f * g)(i) = \sum_{j=1}^{m} g(j) \cdot f\left(i - j + \frac{m}{2}\right) \tag{3.5}$$

Convolution operation reverses and shifts the convolutional filters through the input data f [68]. In the CNN context, this creates feature sets based on the product between the convolutional filters and the input data. After the convolutional layers, CNN takes the feature sets through a pooling layer which reduces the size of the feature sets (also called downsampling). Next, the pooled layer is passed through an activation function layer (often ReLU). Lastly, the output layer, multilayer perceptron or neural network, calculates the output vector (e.g. Classification, prediction, regression). An example of CNN layers is shown below in figure 3.4.

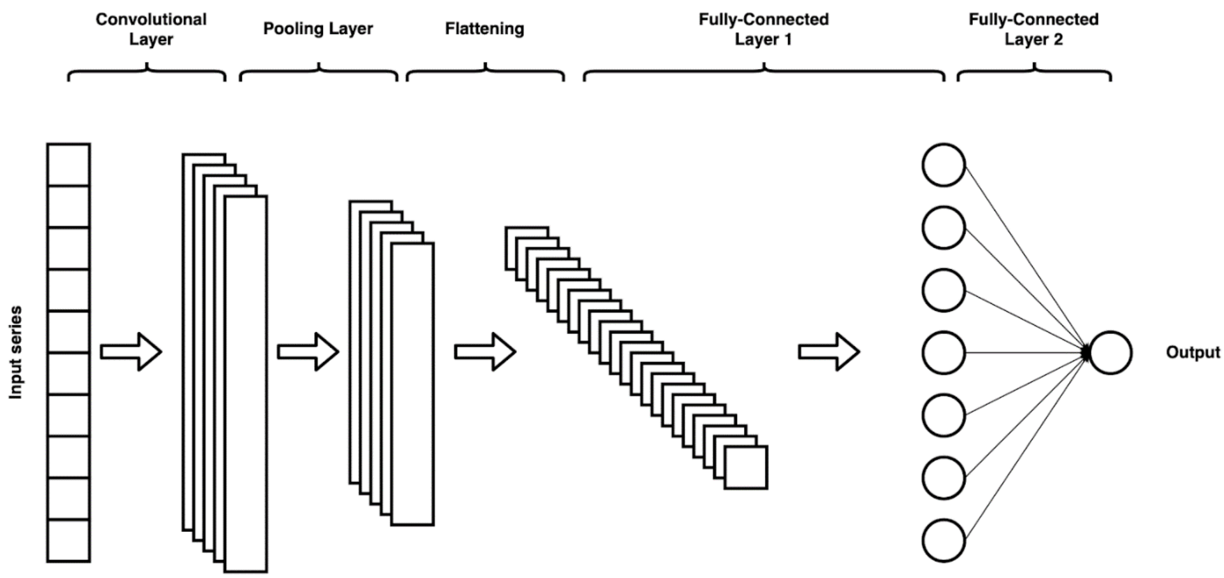

Figure 3.4. One-dimensional convolutional neural network structure and its layers [69]

To summarize the CNN components and layers; First, the convolutional layer performs several convolutions to produce feature sets. Then, the pooling function modifies the feature sets by producing summary statistics of nearby feature set information. For example, the max-pooling produces the maximum value of the feature set within a certain window size. The pooling helps approximate the feature sets. When the approximation is correct, it can greatly improve the statistical efficiency of the CNN such as lower parameters to train which leads to less likelihood of overfitting [70]. Finally, the fully connected layer, multi-perceptron layers, connects weights and values from all prior pooling layers and calculates the class score (in the case of classification). The class score is outputted as a vector at the end of the CNN model.

3.2.4 Boosting model

The idea of boosting algorithms came from using multiple weak learners and combining them to create a better learner without relying on complex models with numerous parameters to train. This is beneficial in many learning problems where model overfitting is an issue such as problems with small real-world training data. For example, we can use three weak predictors to vote for the final prediction. If each predictor is a weak predictor with only 60% accuracy but their error never intersects, each observation can only be false-predicted by one of the three predictors, the voted final prediction will always be correct. Thus, boosting algorithm trains the weak learners sequentially with each preceding learner learning from the errors of the prior learners as shown in figure 3.5.

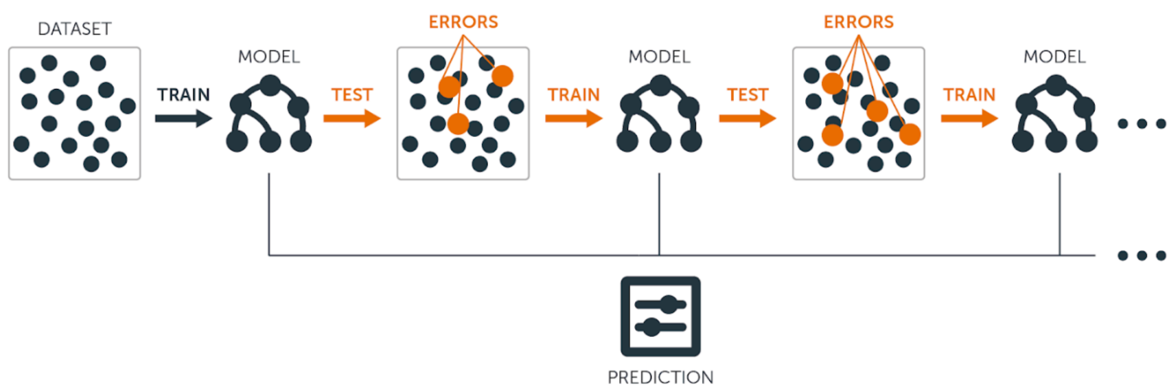

Figure 3.5. Boosting algorithms procedures. Each preceding learner is trained from the errors of the prior learner [71].

First, boosting algorithm training the first weak learner. Let the inputs $x_0, x_1, \dots, x_j$ are associated with the outputs $y_0, y_1, \dots, y_j$. The prediction $\hat{y}$ of the learner is defined in equation 3.6 below, where $w_0, w_1, \dots, w_j$ are the weights of the inputs:

$$\hat{y}(x_i) = f\left(w_0 x_0, w_1 x_1, \dots, w_0 x_j + b\right) \tag{3.6}$$

Then, boosting algorithm calculates the learner's error ε_i by comparing the actual outputs $y(x_i)$to the predicted outputs $\hat{y}(x_i)$. After that, the next learner will be trained with a lower weight on the input $x_0, x_1, \ldots, x_j$ with the correct prediction on the prior learner. An example of the weight penalization is shown in equation 3.7 below:

$$w_i = \frac{1}{2} \ln \left(\frac{1 - \varepsilon_i}{\varepsilon_i} \right) \tag{3.7}$$

Where ε_i are the errors from the prior learner.

This training of a weak learner, calculate errors, and updating/lowering the weights of the preceding learner are repeated for learner $i = 0, 1, \ldots, m$. Finally, the final prediction will be from the voting of multiple weak learners which are boosted from the errors of the prior learners.

3.3 Pre-processing of time-series data for prediction models

Real-world data is often noisy and often contains irrelevant information. The raw real-world data is rarely in a suitable form for machine learning algorithms. For example, data collection in the real world can generate a lot of redundant data, which will increase the cost of training machine learning on irrelevant data. To ensure that the learning algorithms can be efficiently trained, data pre-processing techniques are required. Using pre-processing techniques, we can extract useful information, features, from raw data or transform raw data into appropriate formats like converting texts into numbers. In the case-studies, behavioral and environmental sensing and intervention, the data are continuously collected via ambient environmental sensing nodes. Sensors can produce signals with noises, so data filtering is needed to reduce the presence of the sensor noises. Continuous data collection also produces a lot of time-series data. A two-month-long deployment can generate over 100Gb of data. However, we mainly focus on the period when health events occur, as shown in figure 3.1 (dementia-related agitations, cancer breakthrough pain events), which sometimes only account for 10% of the whole collected data.

In this subsection, relevant data pre-processing techniques are presented which include data filtering, data normalization, data segmentation, and feature extraction.

3.3.1 Filtering and normalization

Data Filtering: Raw sensor data can contain random, spike-like noises [68]. Figure 3.6 shows raw environmental data from one of the real-world deployments with the noises. In the figure, the spike noise, that the sensor produced, can be seen especially in the humidity data. These random noises on the sensor are not relevant to the data and can inflate the value of the actual data.

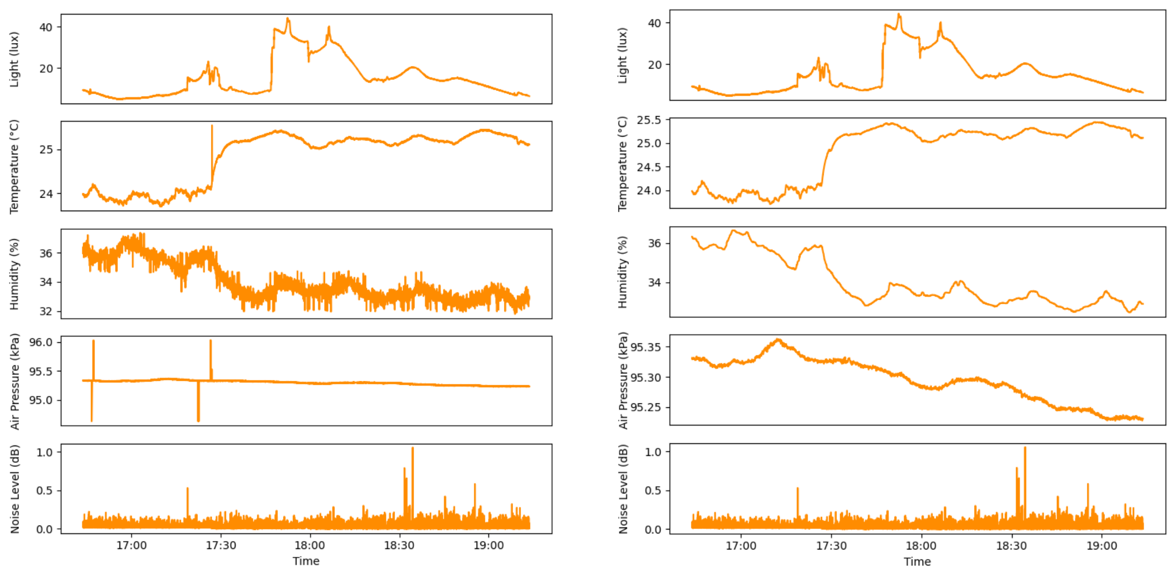

Figure 3.6. (Left) raw environmental data with noises from a sensing station. (Right) The environmental data after 5-seconds median filtering.

A one-dimensional (1D) median filter can be used to remove the spike-noises [72]. Let the raw sensor data be represented by the data sequence $X_j = x_0, x_1, \dots, x_n, \dots, x_j$ where n is the temporal sample index of the data and j is the sequence length. The 1D window-median-filter operation, with a filter-window size C, can be defined in the algorithm below:

ALGORITHM 3.1

Input: data sequence X_j

Output: median-filtered data sequence $median(X_j)$

LOOP Process

1: Setting the filter-window center point at sample n^{th}, and getting all adjacent samples within the window size C (C is usually an odd number, so the number of samples on both sides of the center are equal.)

2: Sorting all samples in C from smaller to larger

3: After sequencing, pick the center value as an output at the n^{th} point

4: Move the filter-window N steps and repeat the process until all j samples of the data sequence are median-filtered

---

Figure 3.6 shows the raw time-series environmental data with noises from the sensors and the data after the median filtering with window size $C = 5$ seconds.

Normalization: Normalization in machine learning often means transforming data values, that have been measured at different scales, into a common scale. For example, in the case-studies with collected environmental data, the sensor collecting air-pressure, also known as barometric pressure or Atmospheric pressure, can measure approximately 100kPa. While the ambient temperature sensor can produce a sensor value of 70°F. To use these different scale values in machine learning algorithms, data normalization is needed for the value to receive similar learning weights. In the works with the case-studies, all inputs of machine learning models, ambient environmental data, and time-of-day are normalized into a range from 0 to 1.

3.3.2 Data transformations

Machine learning often required data transformations to transform raw data into a format that can be used as inputs for learning models. In the case-studies, ambient environmental time-series data is collected around person-with-chronic-disease (e.g. dementia, cancer). These time-series data are transformed from continuous values into discrete values to be used as models' inputs. In this subsection, data segmentation and feature extraction techniques are presented.

Data segmentation: To transform time-series data into inputs of predictive machine learning models, data segmentation is needed to define the temporal length of the input. In the dementia case-study, the data segmentation can be viewed as the length of environmental data, which may have patterns or specific values such as bright light level, that can be used to predict upcoming dementia-related agitations. To learn ambient environmental data patterns that correlate to the occurrences of dementia-related agitation, the data segment immediately before each agitation label is used. Different data segment ranges are tried. Environmental data segment length of 90-minutes, 60-minutes, 42-minutes, 30-minutes, and 18-minutes are tested for performance comparison. Ambient environmental data around the time of the health event label has not been used; it is observed that there are certain activities in the data during most event episodes. For example, we often see high changes in the ambient noise level during the time of agitation; which can be explained as verbally agitated behavior or verbal activities between the patient and caregiver. Since we are interested in the effect of the ambient environment on the patient which leads to the occurrences of health events, we excluded the data at the time of the event labels (specifically, 12-minutes before the event labels). The time-segment period of the ambient environment that the predictive model used for training is shown in figure 3.7.

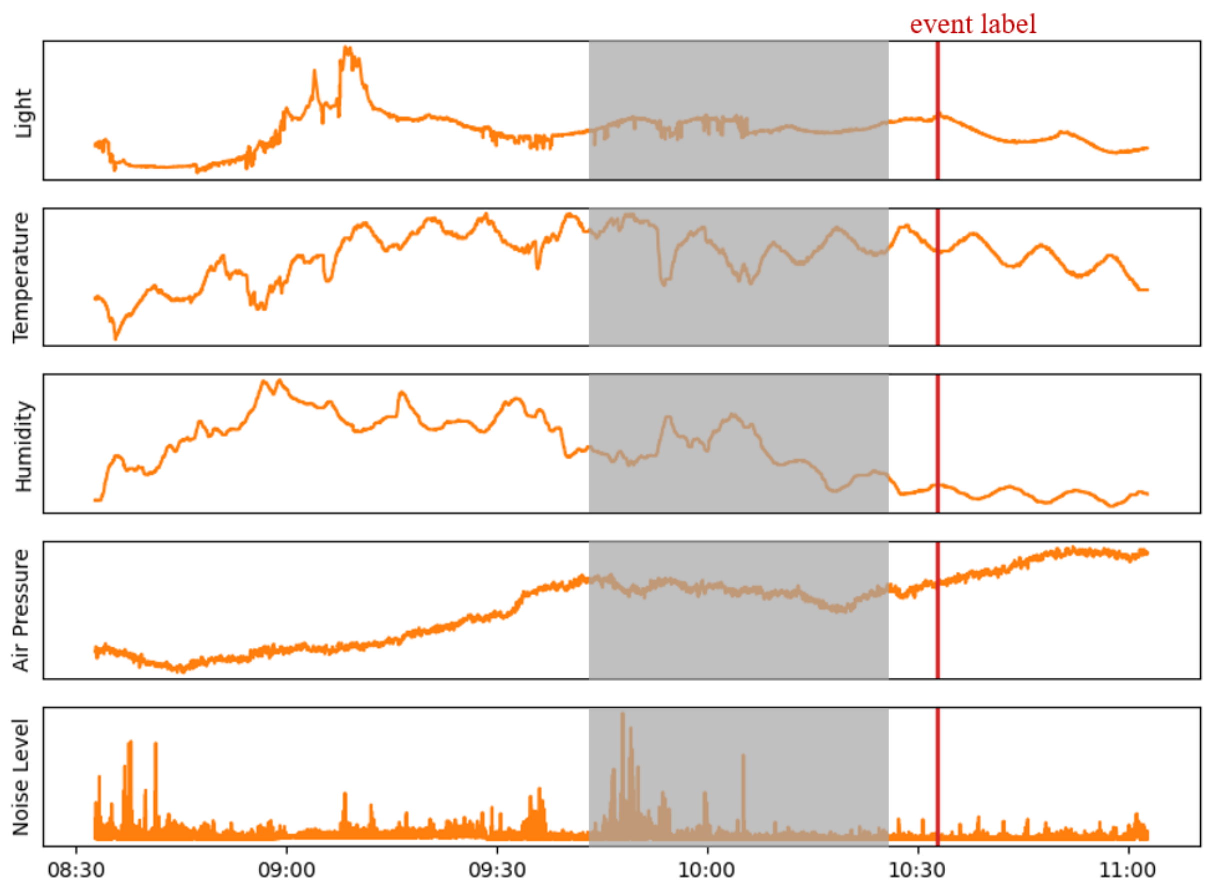

Figure 3.7. Ambient environmental time-segment and the event label used for feature extraction

Feature extraction: Feature extraction is one of the most important tasks to extract relevant information from raw data. Raw dataset is rarely used as input for a learning algorithm. Instead, the data is usually transformed into more useful information for the subsequent machine learning tasks. Here, the feature extraction technique for transforming time-series data into learning model inputs is explained.

Features extraction for neural network models and boosted models: For perceptron models (NN, LSTM) and boosted models, number-vector(s) is required at the input layer. To transform time-series data, such as continuous ambient environmental sensor data, into number-vector features by the process below:

ALGORITHM 3.2

Input: time-series data segment X_{raw}

Output: number vector (input for the learning models) X_{input}

1: Divide the time-series data segment into smaller sub-segments, each sub-segment cover C period of time in the raw time-series data

2: Calculate statistical value(s) of each sub-segment (e.g. statistical mean), the computed statistical values are the feature representing the sub-segment

3: Create number vector(s) with the computed features

For the case-studies data, the collected environmental time-series data is segmented (e.g. 60-minute window) and divided into 6-minutes sub-segments for feature extraction. Features are extracted from the ambient environmental data to represent the current value of the data (e.g. if bright light level affects/causes dementia agitation). Many studies also show that the changes in the ambient environment around the patient can trigger the health events (e.g. if the rapid changes of light intensity affect/cause dementia agitation) [73]. Thus, we extract another set of features that represent the changes in the environmental data. Finally, the time-of-day feature is also extracted to be used as one of the model inputs. The time-of-day is shown to be correlated to certain health events such as dementia-related agitation; studies show that dementia agitation often occurs at a similar time of the day, this phenomenon is called the sundowning syndrome [9]. The summary of all extracted features used for NN, LSTM, and boosted models are shown in table 3.1, assuming time-series data in each 6-minute sub-segment is $X = x_0, x_1, x_2, \ldots, x_i, \ldots, x_{n-1}, x_n$ where n is number of data sample in the sub-segment and x_i is the data at sample i^{th}.

Table 3.1. List of features extracted from each six-minutes window of the ambient environmental data

Feature Type	Description	Statistical Features	
		Feature Name	Statistical Computation
Data value	Features representing the data value of the time-series within the six-minutes window	Mean	$Mean\ \bar{x} = \sum \frac{x}{n}$
		Median	$Median = \begin{cases} \frac{(n+1)^{th}}{2} sample;\ n\ is\ odd \\ \frac{\frac{n^{th}}{2} + (\frac{n}{2}+1)^{th}}{2} sample;\ n\ is\ even \end{cases}$
		Maximum	$x_{max} =$ Sample data with the highest value in X
Data deviation	Features representing the changes in the environmental data within the six-minute window	Variance	$Variance\ \sigma^2 = \frac{\sum(x_i - \bar{x})^2}{n}$
		Mean of differential	$Mean\ of\ \Delta X;$ $\Delta X\ is\ X\ where\ x_i = x_i - x_{i-1}$
		Max of differential	$Max(\Delta X) =$ Sample data with the highest value in ΔX
Time-of-day	A number representing the time of the day	Rescale the time to the ranges of 0 (0:00) to 100 (23:59)	

Feature extraction for convolutional neural network model: The deep learning model CNN is trained to learn patterns in the whole data segment (e.g. the whole 60-minutes of all five environmental data). Thus, the data transformation from time-series to number-vectors is not required as opposed to the perceptron and boosted models' requirements. However, for the case-studies purposes, many clinical studies show that the changes in the ambient environment around that patient with chronic diseases can affect and trigger health events such as dementia-related agitation or cancer pain episodes. Therefore, a rate-of-change feature is computed from each ambient environmental data. The absolute of a gradient of time-series data is used to represent the rate-of-change. Let a time-series segment $X = X_0, X_1, X_2, \dots, X_i, \dots, X_{n-1}, X_n$. the gradient of a time-series array X is defined as:

$$\nabla X_i = \begin{cases} X_1 - X_0\,; \quad i = 0 \\ (X_{i+1} - X_{i-1})/2; \quad 1 \le i \le \; n-1 \\ X_n - X_{n-1}\,; \quad i = n \end{cases} \tag{3.8}$$

Thus, the ROC of a time-series data X is defined as shown in the equation below:

$$\mathrm{ROC}_X = |\nabla X| \tag{3.9}$$

Time-series data with high spike-noise can result in a high fluctuation in the rate-of-change calculation. Therefore, a noise reduction filter is needed before the calculation. Note that the requirement of a noise reduction filter depends on the application and its time-series data. In the case-studies, a five-second median filter is used to reduce the spike-noises which come from the sensors used to collect ambient environmental data. A graphical example of time-series data and its rate-of-change is presented in figure 3.8.

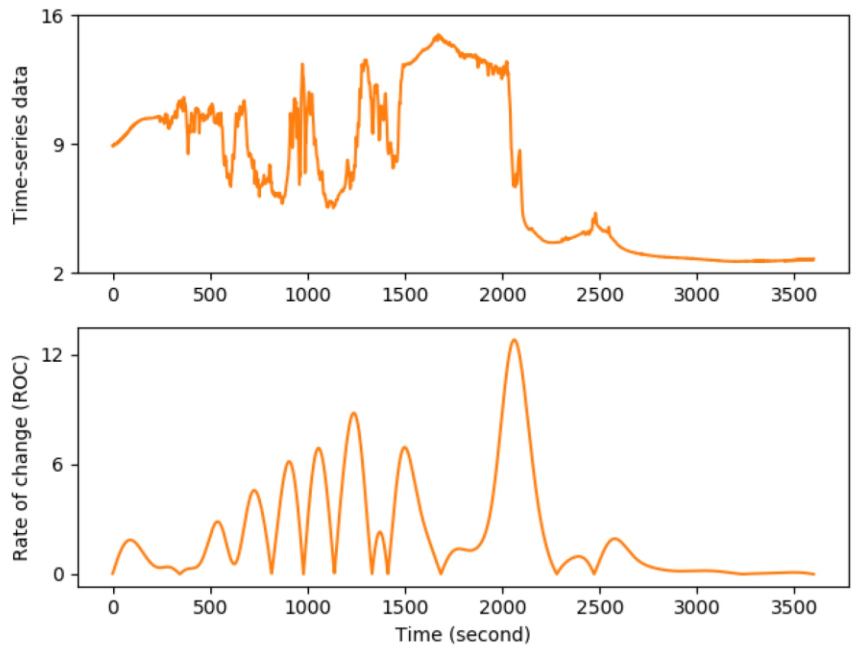

Figure 3.8. Time series plot of the original time-series data segment and its rate-of-change

3.4 Evaluation Metrics

Once a machine learning model is built, model evaluation is one of the most important tasks in the learning problem which determines the performance of the model output (e.g. predictions, regressions). In this subsection, several evaluation metrics and performance measures related to the predictive model are explained.

A classification/prediction model can be evaluated by testing the model with a testing dataset. A testing dataset is a dataset that is independent of the training dataset but still follows the same probability distribution as the training dataset [44]. Let actual class be the known actual values in the testing dataset, and predicted class is the prediction model output, the results of comparing prediction model result to the testing datasets can be divided into four parameters, assuming binary prediction problem (e.g. only predicting yes/no or true/false), as shown in the table below. These four parameters are (1) true positives, (2) true negatives, (3) false positives, and (4) false negatives. The table used to describe the four parameters related to the performance of a prediction model on a test data set is called a confusion matrix [44].

Table 3.2. Confusion matrix

	Predicted Class: Positive	**Predicted Class: Negative**
Actual Class: Positive	True Positive (TP)	False Positive (FP)
Actual Class: Negative	False Negative (FN)	True Negative (TN)

True positives and true negatives are the observations that are correctly predicted. On the other hand, false positives and false negatives are the values that occur when the actual class contradicts the predicted class. From the parameters in the confusion matrix, performance evaluation measurements of a machine learning model can be calculated. Here, four evaluation metrics, accuracy, precision, recall, and F1 score, will be described.

Accuracy: Accuracy indicates how well the machine learning model correctly predicted observations and can be calculated using the equation below:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{3.10}$$

Precision: Precision is a metric that quantifies the number of correct positive predictions made. In another word, precision looks at all positive predictions and tells you the probability of the correct predictions. Thus, a high precision rate is related to a low false-positive rate. Precision can be computed with the equation below:

$$Precision = \frac{TP}{TP + FP} \tag{3.11}$$

Recall: Recall shows the number of correct positive predictions made, out of all actual positive samples. This is different from precision which focuses on the correctly predicted values out of all positive predictions. Hence, recall provides information about the missed positive predictions. High recall means the model rarely classifies an actual positive sample as a negative prediction.

$$Recall = \frac{TP}{TP + FN} \quad (3.12)$$

F1 Score: F1 score is the average precision and recall. Therefore, the F1 score uses both false positives and false negatives to calculates. Thus, the F1 score can be a useful performance evaluation metric when the dataset has an unbalanced or uneven class distribution. Since the case-studies presented in this work have highly unbalanced label distributions (e.g. dementia-related agitation only occurred 48 events during a 2-months-long deployment), the F1 score here will be weighted. The weighted F1 score takes precision and recall of both positive and negative class, as well as the number of instances in both classes into account. This makes the weighted F1 score a better performance evaluation metric in learning problems with highly imbalanced data distribution [74]. The weighted F1 score is shown in the following equations:

$$weighted\ F1\ score\ (F1_w) = 2 \times \frac{precision_{weighted} \times recall_{weighted}}{precision_{weighted} + recall_{weighted}} \quad (3.13)$$

Where:

$$precision_{weighted} = \frac{precision_p \times N_p + precision_n \times N_n}{N_p + N_n}$$

$$recall_{weighted} = \frac{recall_p \times N_p + recall_n \times N_n}{N_p + N_n}$$

The $precision_p$ and $precision_n$ are the precision metric for the positive and the negative class and N_p and N_n are the number of instances in positive and negative class (the same applied for the $recall_p$ and $recall_n$).

The machine learning models' hyperparameters, the points of choice or configuration that allow a machine learning model to be customized for a specific task or dataset, are tuned using grid search approach [42]. The grid search method defines a search space as a grid of hyperparameter values [75]. Then, the machine learning model is trained and evaluated at every position in the grid. In other words, every defined hyperparameter values are used to configure the machine learning model. The hyperparameter with the best evaluation performance is chosen to configure the final model.

3.5 Case-studies results

In this work, the environmental time-series data is used to predict upcoming health events, dementia-related agitation, and cancer pain episodes. Due to the limited data samples, all performance results are done using the 5-folds cross-validation approach [76]. The cross-validation approach randomly divided the datasets into five groups. Four groups will be used as training datasets and the other group is used as a testing dataset. Once each unique group has been used as the testing dataset, the evaluation scores are summarized. The dementia dataset consists of five 2-months-long deployments where actual environmental sensors are deployed at the person-with-dementia and their caregiver houses. The agitation event labels are marked by the dementia caregivers and used for machine learning models to learn features in ambient environmental data which correlate to the occurrences of agitation episodes.

As described in this chapter, the ambient environmental time-series is segmented as inputs to learning models, which are the neural network (NN), boosting model, LSTM neural network, and CNN. For NN, boosting, and LSTM-NN models. The time-series segment is 42-minutes long, from 6-minutes before the CG-provided agitation label to 48-minutes before. For NN, boosting, LSTM-NN models, statistical features are extracted from the 42-minutes long segment of ambient environmental time-series. The 42-minute segment is divided into seven subsegments, each subsegment is 6-minutes long. Each of the subsegments is used to calculate statistical features mean, median, maximum, variance, mean of differentiate, max of differentiate. Time-of-day is also used as one of the input features. For the CNN model, the 42-minutes environmental time-series is used to compute a rate-of-change feature. Both the rate-of-change and the actual environmental time-series (both 42-minutes long) are input to the CNN learning model.

All models are trained and evaluated using a five-fold cross-validation procedure. The results of the cross-validation are summarized with the mean and variance of each evaluation metrics, accuracy, precision, recall, and weighted F1 score. The performance metrics of each model across all five deployments are shown in figure 3.9 below. All models are also compared with a baseline model or a no-skill model which randomly predict an upcoming agitation. The dark stripe on top of each bar represents the cross-validation variance of each performance metric.

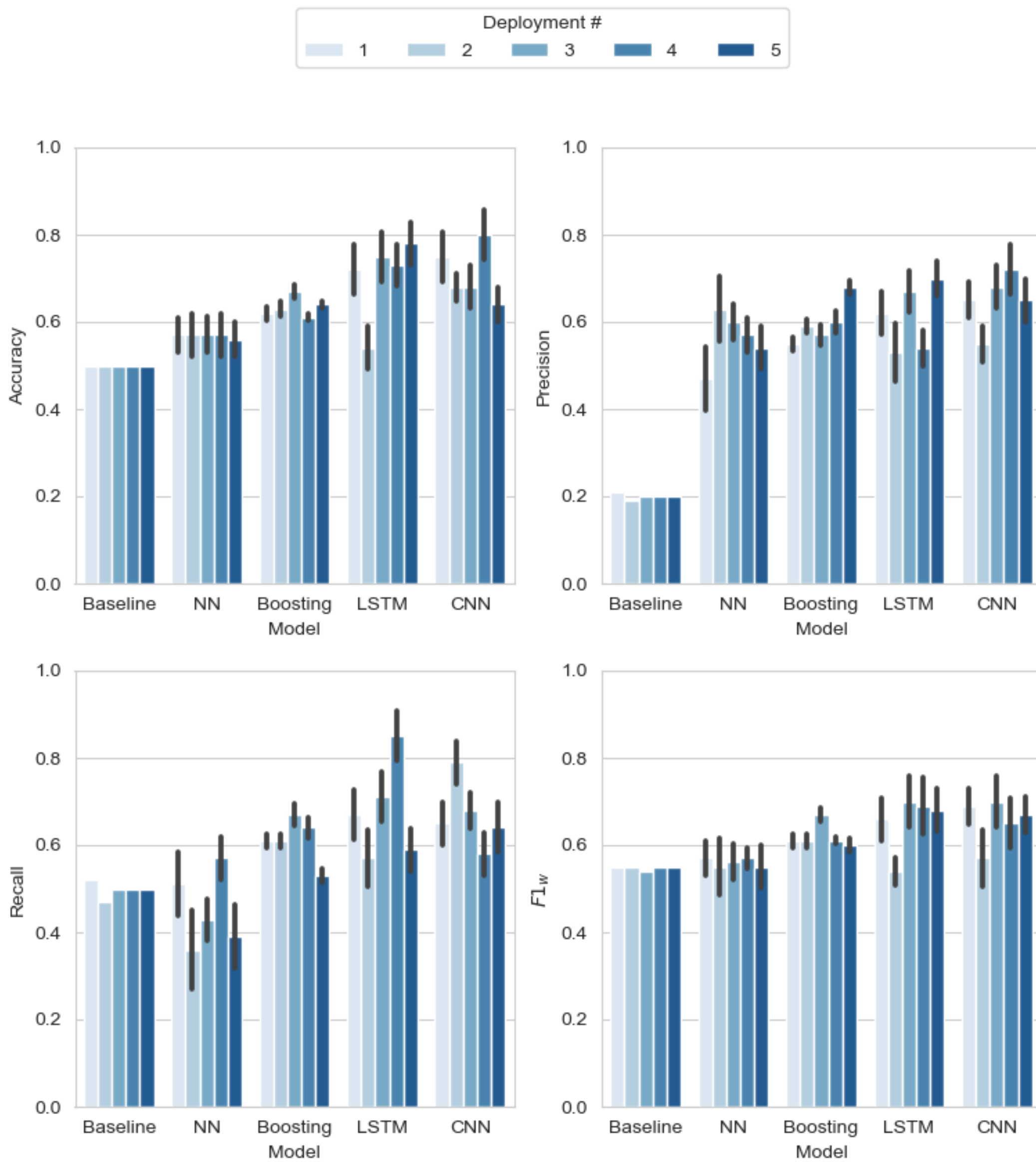

Figure 3.9. Machine learning model performance comparison for the prediction of upcoming dementia-related agitation

Because the dementia-related agitation dataset has unbalanced distribution, the ratio between agitation event label (environmental time-series that followed by an agitation event) and non-agitation label is around 1 to 4, the performance metrics for the baseline model is as follow: 50% accuracy, 20% precision, 50% recall, and 55% weighted F1 score. Table 3.3. also shows a summary of all models' performance which are the average values across all five deployments.

Table 3.3. Summary of machine learning model performance

Model	Performance Metrics (Value ± Variance)			
	Accuracy	Precision	Recall	F1
Baseline	0.50±0.00	0.20±0.00	0.50±0.00	0.55±0.00
NN	0.57±0.05	0.56±0.07	0.45±0.08	0.56±0.05
Boosting	0.63±0.02	0.60±0.02	0.61±0.02	0.62±0.02
LSTM	0.70±0.06	0.61±0.06	0.68±0.07	0.65±0.06
CNN	0.71±0.06	0.65±0.06	0.67±0.06	0.66±0.06

From the result in figure 3.9. and table 3.3., all machine learning models perform better than a baseline model which randomly predicts between agitation and non-agitation. The neural network model has the worst performance with a lower average recall compared to the baseline model (ability to correctly predict upcoming agitation out of all actual agitation data). Overall, the long-short-term-memory neural network and the convolutional neural network have the best performance values. The weighted F1 score, which is a good evaluation metric for imbalanced learning problem [77], of the LSTM is 10%, 9%, 3% higher than the baseline, NN, and boosting model; and only 1% lower than the CNN model.

In another case-study, cancer pain management, ambient environmental data are also collected from five 2-weeks-long deployments in homes of cancer patients. The data label, timestamp of cancer breakthrough pain events, are marked by the patient using Android smartwatches. Then, the environmental time-series data prior to the cancer pain labels are segmented, feature extracted and used as inputs to the same set of machine learning models. In this case study, ambient environment and cancer pain management, the training dataset suffer less from the unbalance label distribution (ambient environment before pain events vs. non-pain events). This is because there are more reported cancer-pain episodes (average of 3.82 episodes per day) compared to the reported dementia-related agitations (0.8 agitation events per day on average). Figure 3.10 below shows the cross-validation performance of all models. All performance metrics are also summarized in table 3.4. Similar to the dementia case study, all machine learning models outperform the baseline model which randomly predicts the cancer-pain episodes. LSTM and CNN models are the best models with a 0.67 and 0.66 weighted F1 score. However, both LSTM and CNN performance scores have higher cross-validation variance which is presented as the dark-stripes at the top of each performance metric bar in figure 3.10. The high cross-validation variances of LSTM and CNN models are due to the mass number of model parameters in both models. A learning model that trains many parameters with a small number of observations (e.g. dementia-agitation events or cancer-pain episodes) can cause the model to overfit the training data which results in a higher cross-validation variance [78]. On the contrary, the boosting model, which uses multiple weak learning instances and required fewer learning parameters, has a significantly lower cross-validation variance. In this case-study, the average recall variance of the boosting model is 0.01 compared to 0.04 and 0.05 average variance of LSTM and CNN model.

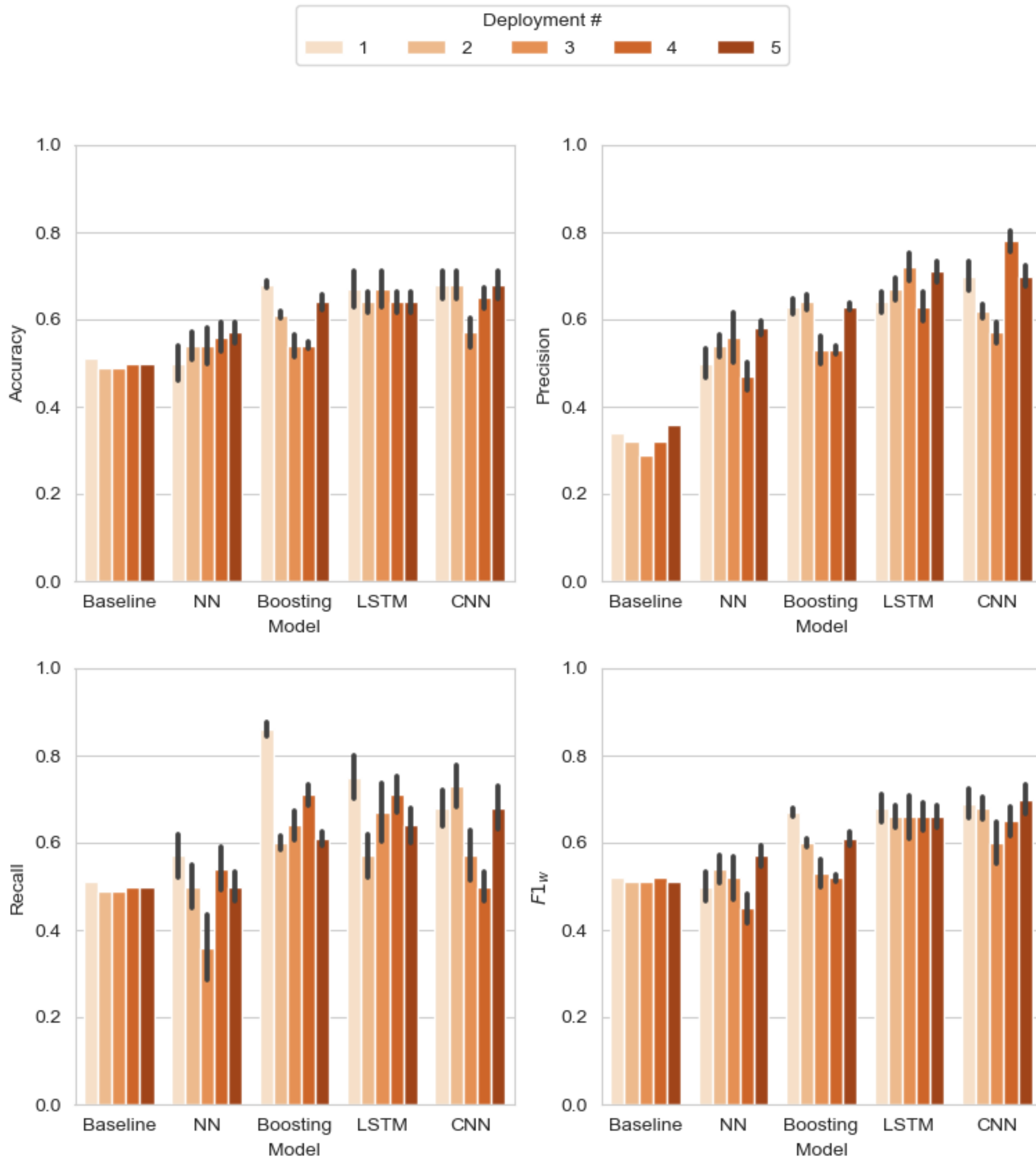

Figure 3.10. Machine learning model performance comparison for the prediction of upcoming cancer pain episodes

Table 3.4. Summary of machine learning model performance for the prediction of upcoming cancer pain episodes

Model	Performance Metrics (Value ± Variance)			
	Accuracy	Precision	Recall	F1
Baseline	0.50±0.00	0.33±0.00	0.50±0.00	0.51±0.00

NN	0.54±0.04	0.53±0.04	0.49±0.05	0.51±0.04
Boosting	0.60±0.02	0.59±0.02	0.69±0.01	0.59±0.02
LSTM	0.65±0.04	0.67±0.03	0.67±0.04	0.67±0.05
CNN	0.65±0.04	0.67±0.03	0.63±0.05	0.66±0.04

3.6 Conclusion and Discussions

In both case studies of dementia-related agitation and cancer pain management, the results show that machine learning models and data processing techniques such as data transformation and segmentation can be utilized to predict health events from ambient environmental time-series data collected near the patient. In this chapter, methods to transform continuously collected environmental data into inputs for machine learning models are described. Various machine learning models and their implementations with time-series data are also shown with relevant performance evaluation metrics. The model evaluation result suggests that LSTM and CNN models are suitable for learning problems in which inputs are time-series data.

The results in figure 3.9 and figure 3.10 show that the LSTM and the CNN model have the best performance in predicting health events from ambient environmental time-series data. The better performances of LSTM and CNN are possibly due to the model's ability to take temporal and sequential information into account (e.g. the ramping up of environment data in each subsegment). The LSTM achieves this by having memory cells that pass on relevant information from the previous feature or input window to the next [50]. On the other hand, the CNN model is using convolution operations which slide a convolution filter across the time-series input, learning any relevant patterns in the data. However, both LSTM and CNN models required a vast amount of learning parameters which is not suitable for a learning problem with a small amount of training data. This makes the cross-validation variance in both LSTM and CNN higher compared to a simpler model such as the boosting model which required fewer learning parameters.

This leads to challenges presented in learning problems in the real world. In the case study examples, the data analysis shows that each patient reacts differently to their ambient environment, some could be sensitive to ambient light level, some are sensitive to the noise around them. Thus, a personalized learning model is required. Each learning model has to be trained on specific patients and their ambient surroundings. This makes the training observations for each model become even smaller and is a challenge for training deep learning models that required a lot of training data/observations. In the next chapter, these real-world challenges will be discussed as well as approaches to create a robust machine learning model that can overcome common real-world learning challenges such as small training data and unbalanced label distribution.

4. PRACTICAL HEALTH EVENT PREDICTION FOR IN-THE-FIELD APPLICATION

Deep learning system advances and achievements show the potential for the implementation of deep learning in real-world problems. Deep learning models usually involve millions of parameters. This required a very large amount of training data for deep learning to outperforms other traditional learning models [79]. In the academic world, obtaining a large amount of well-labeled data to train the learning algorithm may not be a problem if the data collection process is not focused. However, in real-world scenarios, datasets often come with challenges such as imprecise data labels, a small number of training data, and unbalanced data labels. Consider one of the use-cases presented in this work, the dementia-related agitation and in-home caregiver empowerment. The training labels, dementia agitation events, are used to train the prediction model to predict the occurrences of dementia-related agitation symptoms. The labels are provided by in-home caregivers. Thus, they can be temporally imprecise depending on when the caregiver noticed an agitation event. Furthermore, the average number of agitations reported in two-month deployment is 48 events, which is a small number of labels to train complex learning models such as deep learning.

This chapter discusses data challenges associated with real-world learning modeling approaches. For example, a weak training label is a common problem in the real-world dataset (e.g. timestamp of the event of interest is temporally imprecise). To minimize the invasiveness associate with video camera usage, a real-world symptom sensing system can involve patient engagement by utilizing the self-annotation approach. Patient with chronic cancer pain can self-report their symptom occurrences which can be used as training data labels for cancer pain prediction applications [26]. The imprecision can occur due to the real-world difficulty of the patient dealing with the occurring symptom (e.g. cancer patient taking pain medication) and cannot report the symptom event immediately. Here, the implementations and evaluations of a boosted convolutional neural network with regularization and weighted sample technique is presented to addresses the following in-the-field data challenges: temporally imprecise training labels, small training samples, and unbalanced label distribution. The goal of this is to evaluate different machine learning models to address the mentioned real-world challenges such as model overfitting [27]. The work presented in this chapter can assist researchers in implementing machine learning models to solve similar real-world problems.

4.1 Related works

Many studies in the field of machine learning often use the standard procedure to measure, compare, and evaluate model performance on a standard dataset [80]–[82]. These methods are used to compare learning techniques with other approaches to make sure that one method outperforms others. However, in real-world learning problems, the dataset often comes with complications such as imbalanced label distribution or a small number of training observations. Thus, building a robust learning model is a higher priority than modifying and improving a model to gain an extra 1% in performance [83].

Imbalanced label distribution is one of the common problems associated with real-world data acquisition [66], [67]. In one of the case-study presented in the dissertation, on average, the dementia-related agitation events are reported 48 times per 2-months long deployment. The predictive model used to predict dementia-related agitation is trained to recognize those time-periods as agitation events and other time-periods as the non-agitation event. Thus, the number of agitation time-periods are much less than the non-agitation time-periods which creates an imbalanced dataset. The imbalanced dataset can misguide the learning process of a machine learning model. A learning model that focuses on increasing the model's accuracy may classify every prediction as a negative prediction if the training dataset is imbalanced with a lot more samples of negative class. Many studies use a weighted sample technique to solve the imbalanced label distribution problem [86]. The weighted sample gives different learning weights to each training class to modified the minority class's importance in the model's learning process.

Another common challenge with real-world machine learning problems is the small amount of training data. For many real-world problems, it can be rather expensive to get well-labeled training data. A study estimates that the cost for acquiring the Computed Tomography (CT) images with annotations from domain experts could be as high as 250k US dollars for 10,000 labeled images. Another cause that affects the number of training labels is the requirement of a personalized learning model. Recently, precision medicine has been adapted and implemented to provide healthcare that takes the variability between persons into account [87], [88]. Unfortunately, the requirement of a personalized learning model also reduces the number of training datasets available. The small number of training observations is usually associated with an overfitting problem in machine learning [89]. Complex machine learning models required more training instances to outperform traditional learning models due to the mass amount of model parameters [90]. Solutions to the small amount of training data problems could be to utilize boosting learning model approach which combines multiple simple-but-weak learners to a better model [91], [92]. However, in the previous chapter, the results show that machine learning models that consider the temporal relationship in the time-series data, such as LSTM and CNN model, perform better than other models.

In this chapter, a boosted convolutional neural network is implemented for health event prediction from time-series data. The boosted CNN is a hybrid machine learning model between the boosting model, which incorporates the boosting weights with multiple learners, and the convolutional neural network model, which uses a convolutional filter to extract temporal features in time-series data. The boosted CNN is compared with a standard boosting model and a regular CNN to evaluate the model's robustness with real-world datasets.

4.2 Practical techniques for in-the-field learning algorithms

Learning problem that uses data continuously collected in real-world can come with many challenges such as small training labels. In the case studies, prediction of dementia-related agitation and cancer-pain episodes, to minimize the invasiveness associate with video camera usage as shown in many other studies [93], [94], our sensing system involves patient engagement by utilizing the self-annotation approach. During the real deployments at actual patients' houses, on average, we collect 48 dementia-related agitation events per deployment and 46 cancer-pain episodes per deployment. This small number of training datasets can cause a more complex learning model to overfit the data [27]. However, as shown in the results of the previous chapter, complex learning algorithms such as the long-short-term-memory neural network (LSTM) and convolutional neural network (CNN) are needed to achieve good performance with predicting health events from time-series data (i.e. ambient environmental time-series data). In this section, approaches to overcome in-the-field challenges associated with implementing a learning algorithm for real deployment data are presented. The regularization technique is presented to help with model overfitting problems. Next, as presented in the dementia case study where only 48 agitation events occur throughout 2-months data collection deployment, the unbalanced label distribution is resolved with the weighted sample technique. Lastly, the boosting model is modified to be more suitable for learning problem with time-series data. The boosting model is shown, in the previous chapter, to be able to produce the best cross-validation performance variances; but the model still has a downside that it does not take temporal information of the time-series input into account.

4.2.1 Regularization

Complex learning tasks trained with a small number of training samples can be overfitted due to a large number of model parameters are trying to fit on little observation points [27]. Figure 4.1 illustrates the overfitting behavior of machine learning models. Normally, when training a machine learning model, we want the model to fit the training data, i.e. the model is trying to minimize the training error. After the model is trained, the model is evaluated with another dataset called the testing dataset. In an ideal situation, the testing dataset will come from the same population and share the same probability distribution as the training dataset; if the model has a low training error, it should fit well with the testing dataset and has a low testing error as well. However, when using a more complex model, such as LSTM and CNN models shown in the previous chapter, the model may have low generalization error where the model fits well with the training data but has a high testing error as shown in figure 4.1. This is called model overfitting. The overfitting problem occurs when there are a lot more model parameters compared to the number of training observations [95], [96].

In the case studies, we have a small amount of training data or reported health events (48 dementia-related agitation events per deployment and 46 cancer-pain episodes per deployment), but we need the more complex machine learning models that utilized the temporal information in the ambient environmental time-series collected around the patient. And the result in the previous chapter showed that both complex models are prone to the overfitting problem, i.e. high cross-validation variances presented in both models. A solution to the overfitting problem due to the complexity of the model is the implementation of regularization. Machine learning regularization minimizes model complexity by penalizing the model's loss function. In this section, the L1 (lasso) regularization technique is utilized to improve the overfitting problem.

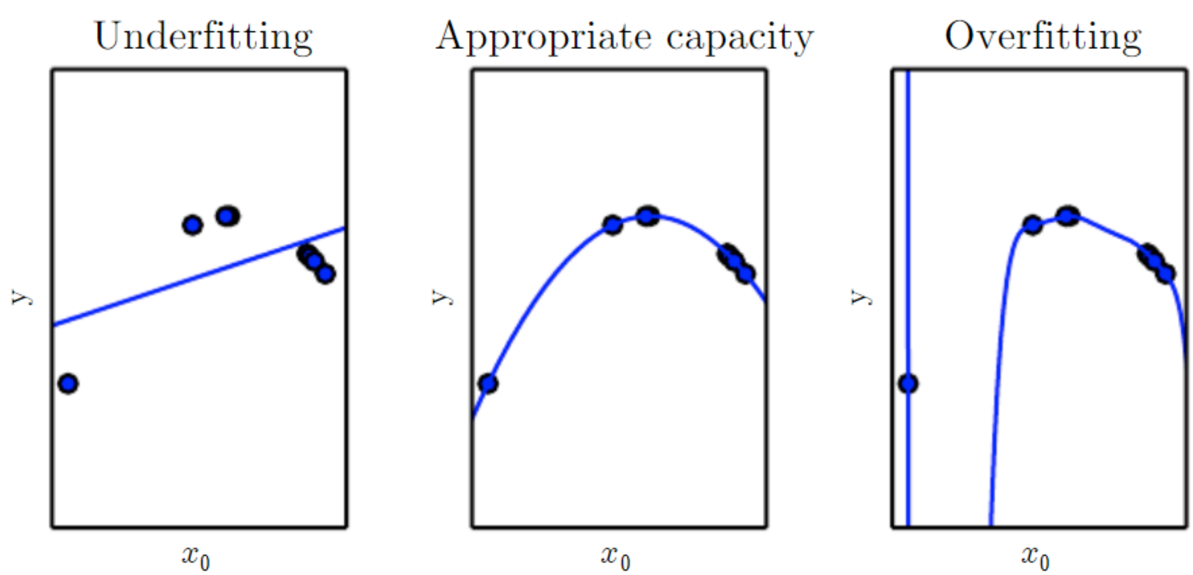

Figure 4.1. Illustrations of model overfitting behavior. (Left) A linear function fit to the data suffers from underfitting—it cannot capture the curvature that is present in the data. (Center) A quadratic function fit to the data generalizes well to unseen points. It does not suffer from a significant amount of overfitting or underfitting. (Right) A polynomial of degree 9 fit to the data suffers from overfitting [42].

Regularization is a technique that helps reduce model overfitting due to the complexity of the model. It achieves this by penalizing the model's cost function [42]. Normally, a learning model tries to minimize its cost function. If x is the input with m

samples, $h_\theta(x)$ is the classification prediction output with θ as parameters (e.g. weights), and y is the target output, the following formula shows a cost function example of a logistic loss:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta(x_i), y_i)$$

$$where: Cost(h_\theta(x), y) = \begin{cases} -\log\big(h_\theta(x)\big) & if\ y = 1 \\ -\log\big(1 - h_\theta(x)\big) & if\ y = 0 \end{cases} \quad (4.1)$$

$$and: h_\theta x_i = \theta_0 + \theta_1 x_1 + \theta_2 x_2^2 + \theta_3 x_3^3 + \cdots + \theta_m x_m^m$$

From the formula (2), the degree of the input features will increase model complexity (more θ), making the model tries to fit all training data points which can result in an overfitted model, especially if some of the input features have a small contribution to the prediction [27]. Lasso regularization (L1) remedies this by adding the regularization term to penalize all the weights as shown in equation (3):

$$L1: \quad J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta(x_i), y_i) + \frac{\lambda}{m}\sum_{j=1}^{n}\left|\theta_j\right| \quad (4.2)$$

The variable λ helps regularize the original fit term in (2) by penalizing the weights θ. This can make the learning model less overfitted.

4.2.2 Weighted Samples

Another difference between academia and the real-world is the distribution of data labels. In academia, datasets usually have balanced label distribution. Thus, supervised learning tasks usually have an equal number of samples per class. However, in the real-world, balanced label distribution is rare. In the use case of dementia-related agitation, 48 agitation events were labeled in a 2-month deployment. If 1-hour of data is label as agitation period, that means only 48 hours of time-series data are labeled as agitation periods out of the total of 1,460 hours long deployment. This makes the label distribution highly unbalanced. A prediction model trained on highly unbalanced training labels can learn to predict only the majority class and still maintain high prediction accuracy. To address this problem, the weighted samples technique is utilized. This technique modifies the model's learning rate to provide more weight to the minority sample (e.g. symptom event labels) [84]. The weight ratio between majority samples and minority samples can also be tuned depending on the consequences. For example, diagnosing somebody healthy as sick may be acceptable if the doctor double checks; but does not recognize a sick patient, and letting the person go without treatment can have dire consequences.

4.2.3 Hybrid Learning Framework

In the previous result, it is shown that machine learning models that take temporal information in the time-series data into account outperform models that do not. In the analysis of machine learning models with the case studies data, using ambient environment surrounding patient living at home to predict health events, the long-short-term-memory neural network (LSTM) and convolutional neural network model (CNN) can learn to predict health event from the environmental time-series data better than the neural network and boosting model. However, the result showed that the LSTM and CNN are complex models that have many learning parameters which, when implemented to a real-world learning problem that often has a small number of training observations, can overfit the training data. The data overfitting can be seen from the higher cross-validation performance variances. Simpler machine learning models, such as the boosting model, show potential in the case studies results with the least cross-validation performance variance as shown in figure 3.9 and 3.10. Still, the boosting model is not built to consider the temporal information in the time-series data.

In this section, a hybrid machine learning model is shown. This hybrid learning model takes advantage of the simpler boosting model to reduce the overfitting problem and the advantage of the CNN model that takes temporal information in the predictors into account [28]. In other words, the hybrid model combines the boosting algorithm and parts of the CNN model (hence, it is a *hybrid* model). Figure 4.2 shows the framework of the hybrid learning model, called boosted CNN. The structure of the hybrid learning framework is as follows. First, the time-series dataset is pre-processed (e.g. data segmentation, noise filtering) and feature extracted. In both dementia and cancer case-studies, rate-of-change of the environmental data is extracted as features. Next, extracted features are input into the CNN model to learn patterns in the training dataset. Here, instead of normally using the result of the CNN fully-connected layer as an output, the product of the CNN pooling layer is extracted as shown in figure 4.2. As explained in the previous chapter, the CNN convolutional layer extracts temporal features from each small region of the time-series data, then the pooling layer connects and downsamples the features which create a list of relevant temporal features from the CNN model. Then, this list of CNN features is used as an input vector for the boosting model. The boosting model is implemented with regularization and appropriate weighted sample technique to help with overfitting and unbalance label distribution problems.

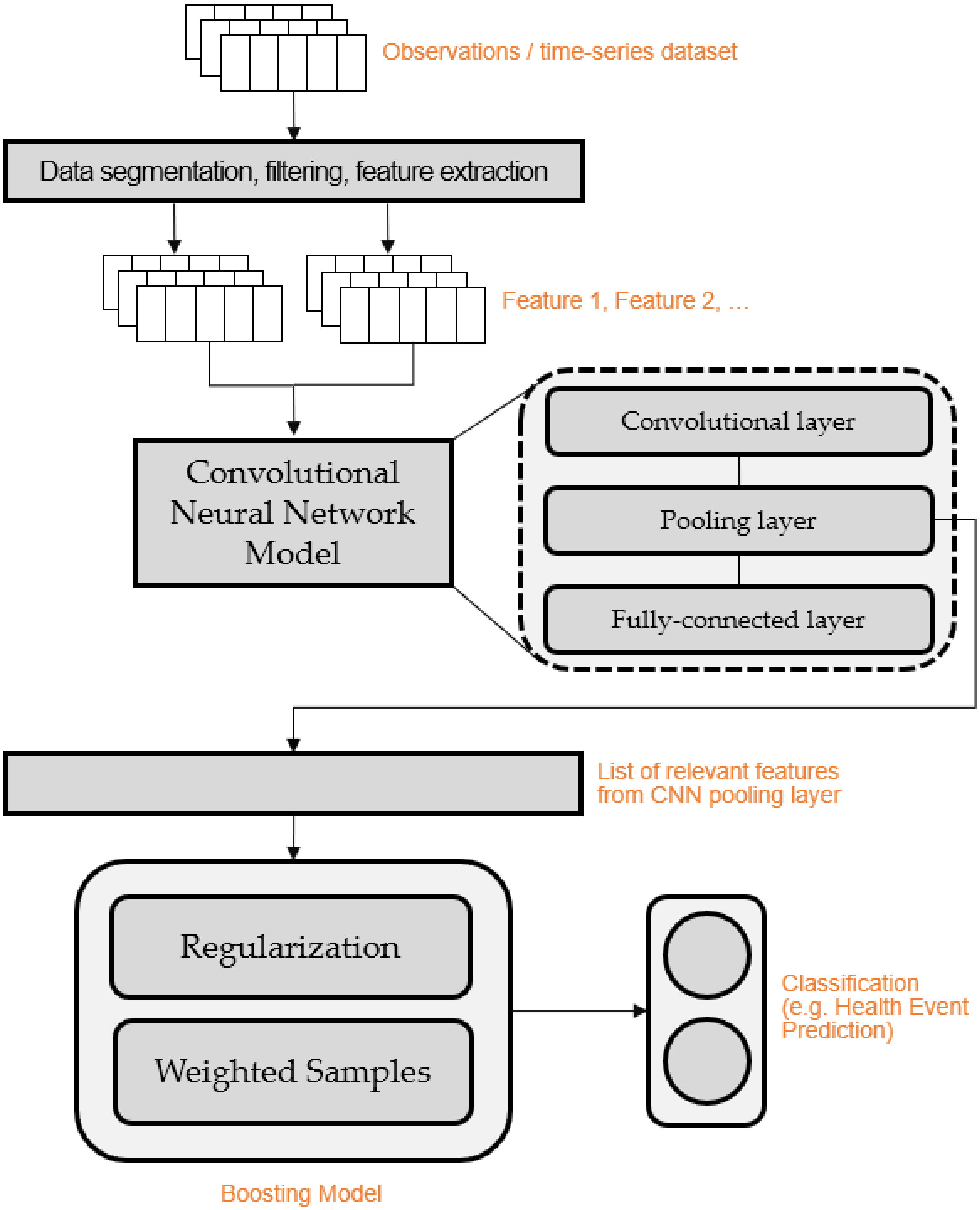

Figure 4.2. Hybrid learning framework with practical performance improvement techniques for real-world learning problems.

4.3 Case-study Results

In this section, practical techniques are shown to address common challenges found in real-world sensing and learning applications. In the case-studies, the sensor system is deployed at dementia and cancer patients' homes to collect ambient environmental data to study the effect of the surrounding ambient environment on the patients [73], [97]–[99]. The environmental data (i.e. light level, ambient temperature, humidity, air-pressure, noise level, and time-of-day) is continuously monitored and used to predict health events (dementia-related agitations and cancer-pain episodes). In both case-studies, the collected health events observations are limited. Only 48 dementia-related agitation events per deployment (on average, out of five 2-months long deployments) were observed; and 46 cancer-pain episodes were marked per deployment (averaged from five 2-weeks long deployments). The observation data from a machine learning model is even smaller because each user reacts to their ambient environment differently which required a personalized learning model. Furthermore, the observation data is unbalanced with 1:4 agitation observations to non-agitation observations ratio and 1:2 pain episode to non-pain episode. To address these real-world challenges, a lasso regularization and a weighted sample technique have been implemented with a hybrid learning framework which combines the ability to extract temporal features of the CNN model and the simplicity of the boosting model.

Here, health event observations from both case studies are predicted using ambient environmental sensor data prior to the event timestamps. Four learning models are implemented to predict health events. The models are the baseline model which randomly predicts health events, boosting model, CNN model, and the presented hybrid model. The models are trained and evaluated using five-fold cross-validation. The performance evaluation metrics are accuracy, precision, recall, and weighted F1 score. Accuracy shows how well each model classifies the ambient environment that can trigger health events and the one that does not. Precision indicates the probability that the predicted health events are correct (e.g. if the model predicts that there will be an upcoming dementia-related agitation, what is the probability that it is correct). The recall represents the sensitivity of the probability that the model will detect health events given that the input is known to trigger upcoming health events. Lastly, the weighted F1 score is used as an overall evaluation metric that considers the precision, recall, and label distribution of the data.

4.3.1 Dementia-related agitation prediction results

The result of predicting upcoming dementia-related agitation events from the ambient environment is shown in figure 4.3 and table 4.1 below. In figure 4.3, the boosting mode, CNN model, and hybrid model are shown to outperform the baseline model which randomly predicts dementia-related agitation events from ambient environmental data. Comparing the simpler boosting model and the more complex CNN model, CNN has higher scores in all evaluation metrics which is likely due to its ability to extract temporal features from time-series data. However, the boosting model has significantly lower performance variances as shown in the dark-stripes in figure 4.3. This lower scores' variance is due to less model overfitting. The hybrid learning model, which can extract temporal features like the CNN model and less likely to overfit as the boosting model, is shown to have the best performance compared to other models. The hybrid model's performance variances are comparable to the boosting model, as shown in table 4.1, and the performance scores are higher than the CNN model with a 0.77 weighted F1 score compared to CNN's 0.66 F1 scores.

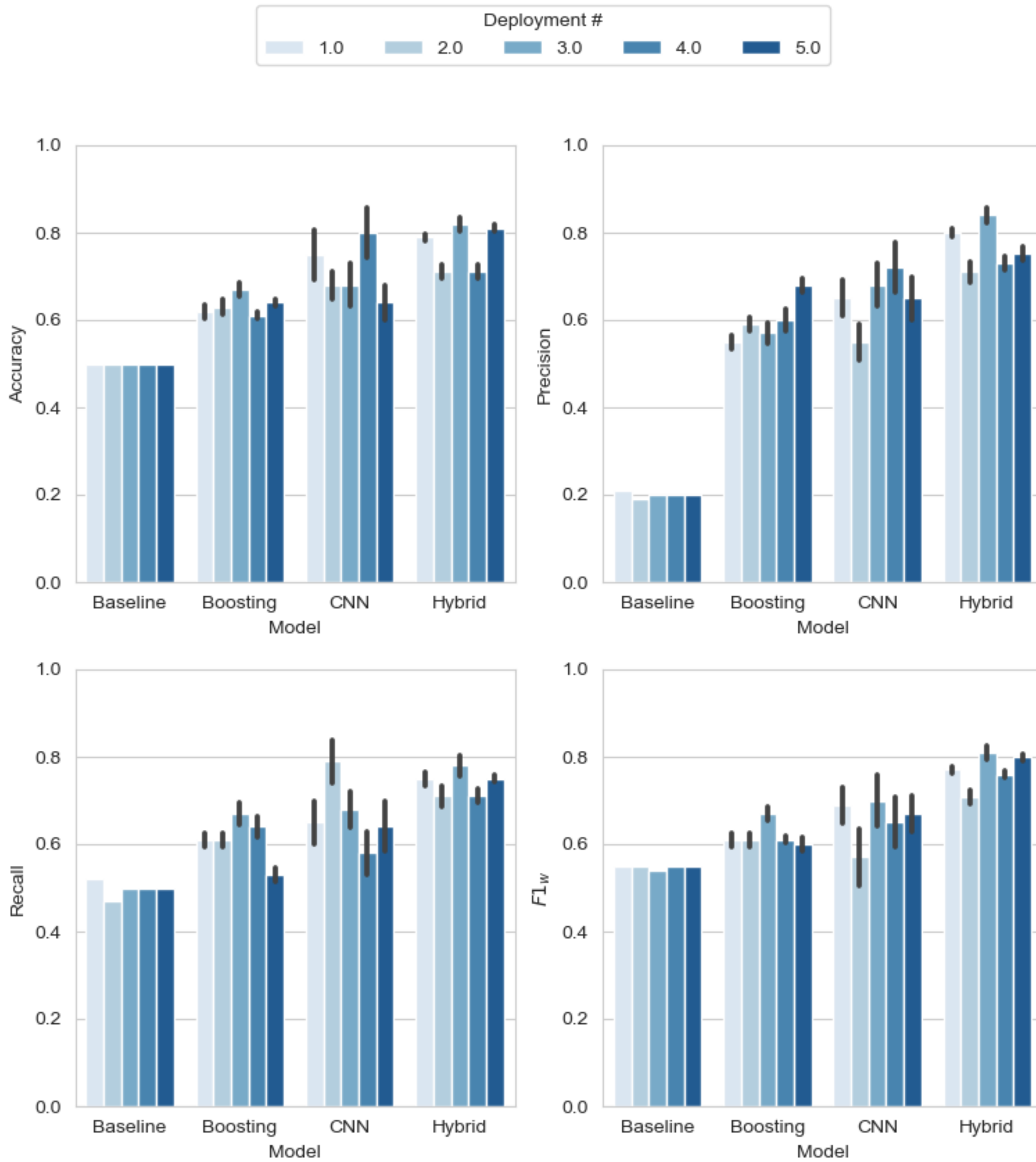

Figure 4.3. The hybrid learning model comparison to the baseline model (random prediction), boosting model, and CNN model in the prediction of dementia-related agitation

Table 4.1. Summary of machine learning model performance for the prediction of dementia-related agitation events

Model	Performance Metrics (Value ± Variance)			
	Accuracy	Precision	Recall	F1
Baseline	0.50±0.00	0.20±0.00	0.50±0.00	0.55±0.00
Boosting	0.63±0.02	0.60±0.02	0.61±0.02	0.62±0.02

CNN	0.71±0.06	0.65±0.06	0.67±0.06	0.66±0.06
Hybrid	0.77±0.02	0.77±0.02	0.74±0.02	0.77±0.01

Figure 4.4 below also show the dementia-related agitation prediction comparison between (1) within-deployment prediction where individual learning model learns to predict agitation from environmental data and agitation labels within each deployment, and (2) cross-deployment prediction where a global model is trained from the combined dataset, a dataset which combines all agitation observations from all five deployments. Here, the shown within-deployment prediction performance is the average of the individual learning models from each deployment. The result in figure 4.4 shows that the within-deployment approach outperforms the global learning model approach. The accuracy score of the hybrid model using the within-deployment approach, which is an average from the accuracy scores of all five deployments, is 0.77 which is a much better performance compared to the 0.53 accuracy score of the cross-deployment or the one global model approach. The result, which the individual learning modeling approach has much better agitation prediction performance compared to the one global model, suggests that each person-with-dementia react differently to their ambient environmental surrounding. The learning model needs to be configured and trained with individual PWD to be able to detect dementia-related agitation events reliably. The need for personalized learning also emphasizes that the small number of available real-world observations and datasets which is one of the challenges that the work in this chapter is aimed to solve.

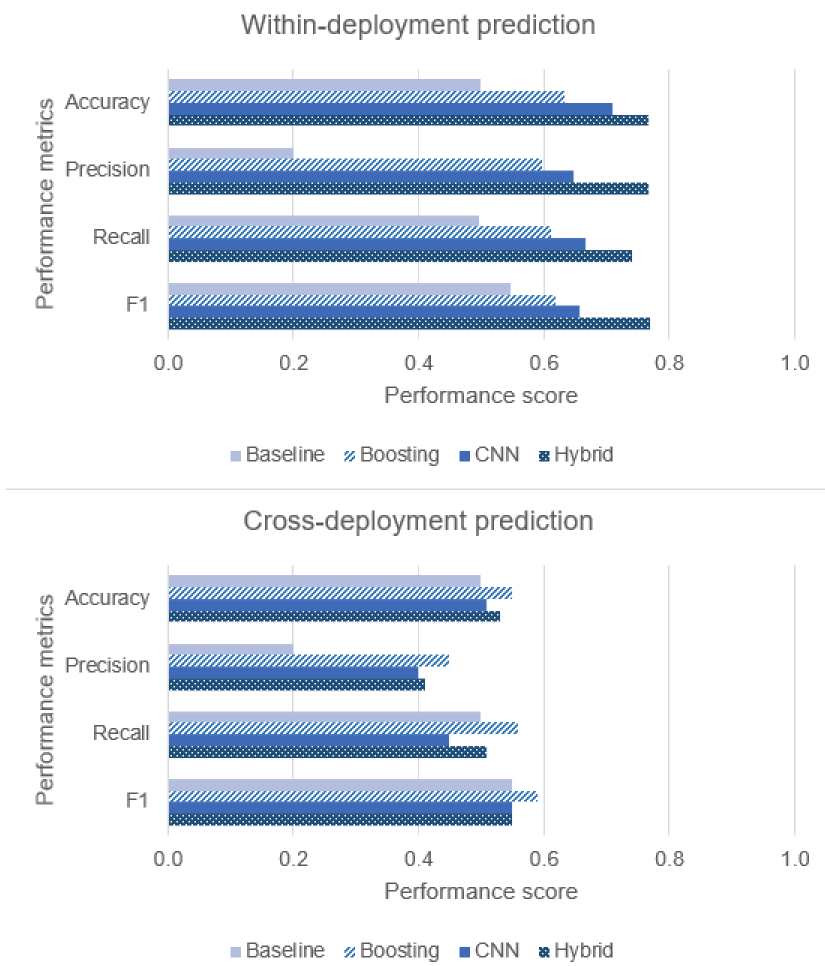

Figure 4.4. Dementia-related prediction performance comparison between the within-deployment prediction, average across all five deployments and the cross-deployment prediction where the data from all five deployments are combined to one dataset

In actual sensor deployments at PWD/CG houses, the dementia-caregivers observe for signs of dementia-related agitation and mark the agitation labels using the user interfacing devices. This self-labeling approach can help to collect ground truth data about the occurrences of agitation without the involvement of privacy-invasive devices such as video cameras. However, the agitation labels from the caregiver can be temporally imprecise which happened due to the need for immediate CG attention when there is an agitation event or the CG observed the agitation event at the later stages causing the label-marker to be later than when the agitation event started to happen [100]. The machine learning models implemented to predict upcoming dementia-related agitation use data segmentation technique which the ambient environmental data before the agitation labels is segmented and used as inputs to the learning models. Different segment sizes have been used and compared. The result of the effect of the segment size on the prediction of dementia-related agitation is shown in figure 4.5 below. Here, different observation segment sizes, 18 minutes, 30 minutes, 42 minutes, 60 minutes, and 90 minutes are used and evaluated with the weighted F1 score which shows the balanced precision and recall of both positive and negative observations. The larger segment size will make the learning model more robust to the temporally imprecise training labels but will include data that might not be useful to the prediction. The result in figure 4.5 shows that time segmentation around 30 minutes to 42 minutes is the most suitable segment size for the prediction of dementia-related agitation.

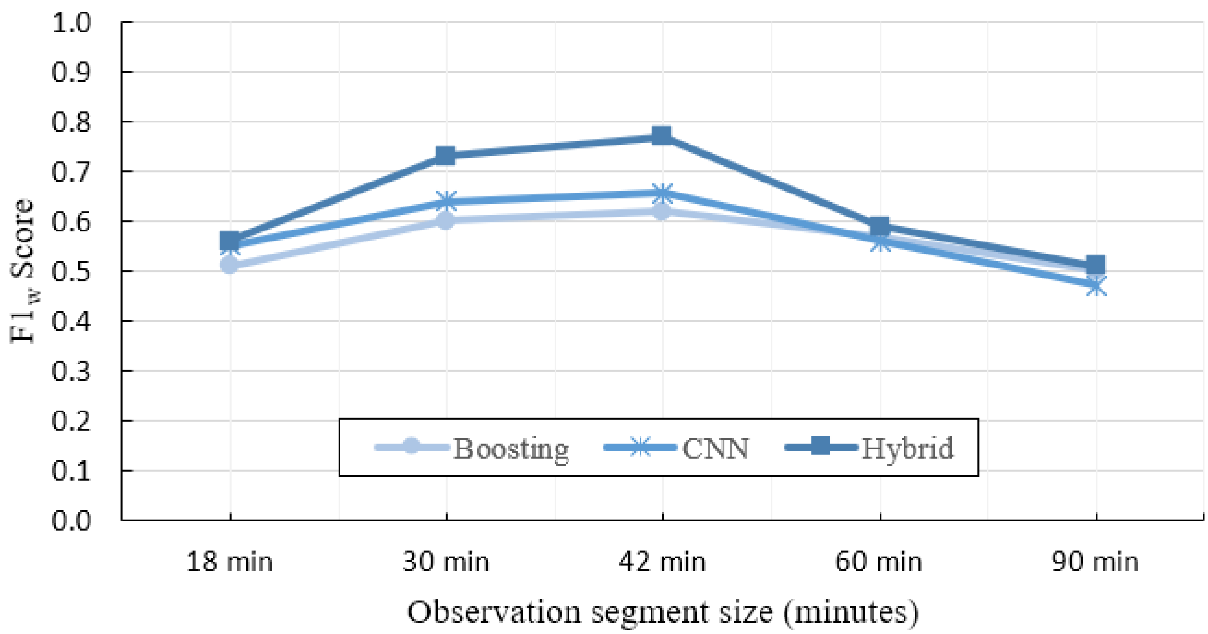

Figure 4.5. Effect of different data segmentation size on the weighted F1 score of the dementia-related agitation prediction

4.3.2 Cancer-pain episodes prediction results

Similar to the dementia case study, observations of cancer-pain episodes collected from the deployments at cancer patient houses are limited (46 cancer-pain episodes per deployments on average). This small number of observations can cause complex learning models to overfit. Thus, the practical techniques presented in this chapter are required for real-world learning problems such as this case study. Below, figure 4.6 and table 4.2 show the learning model results of predicting cancer-pain episodes from ambient environmental data around the cancer patients living at home.

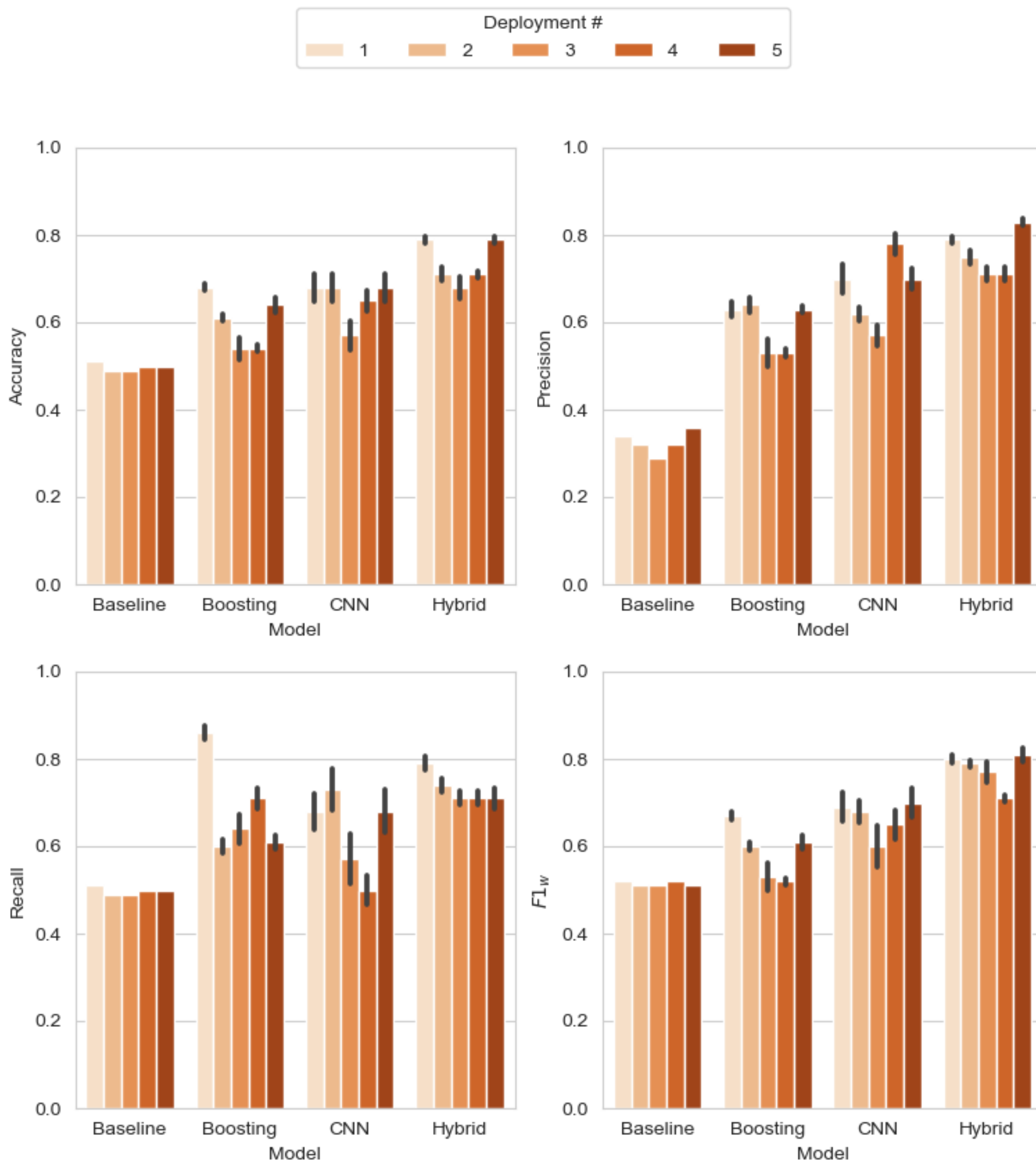

Figure 4.6. The hybrid learning model comparison to the baseline model (random prediction), boosting model, and CNN model in the prediction of in-home cancer-pain episodes

Here, the label distribution (ambient environmental time-series before pain event timestamps vs. ambient environment at other times) is not as severely imbalance as the observation labels of the dementia case-study, the ratio of cancer-pain observations and non-cancer-pain observations are about one to two (the observation labels ratio of the dementia case study is one to four). This label imbalance can be seen at the precision score of the baseline model that is 0.32 on average (from table 4.2) compared to the 0.20 precision score of the dementia case-study in table 4.1. The result here also shows that the presented hybrid model has the best performance scores for all five cancer-case-study deployments as shown in figure 4.6.

Table 4.2. Summary of machine learning model performance for the prediction of cancer-pain episodes

Model	Performance Metrics (Value ± Variance)			
	Accuracy	Precision	Recall	F1
Baseline	0.50±0.00	0.32±0.00	0.50±0.00	0.51±0.00
Boosting	0.60±0.02	0.59±0.02	0.69±0.03	0.59±0.02
CNN	0.65±0.04	0.67±0.03	0.63±0.05	0.66±0.04
Hybrid	0.73±0.02	0.75±0.02	0.71±0.02	0.77±0.02

Next, results of the cancer-pain episode prediction from both the within and cross-deployments are presented. First, the cancer-pain prediction performances of the learning models trained on each individual deployment, or the within-deployment performances, are shown in figure 4.7 (top). The cross-deployment performances are also illustrated. Cross-deployment refers to performance scores from having one global learning model trained on the combined dataset, which combines all cancer-pain observations regardless of the deployments. The cross-deployment approach shows weak cancer-pain prediction scores compared to the individual modeling approach. The recall score of the hybrid model or the probability that the model will predict a cancer-pain event if the patient is actually experiencing a pain episode, in predicting cancer-pain is 0.71 compared to the cross-deployment recall score of 0.49. This result shows the need for an individual or personalized learning model which is trained on the behavior of the individual patient.

This will make the observation data size even smaller, but the methods presented in this chapter, such as the hybrid modeling and model regularization will help increase the model's robustness and tolerance to real-world data problems.

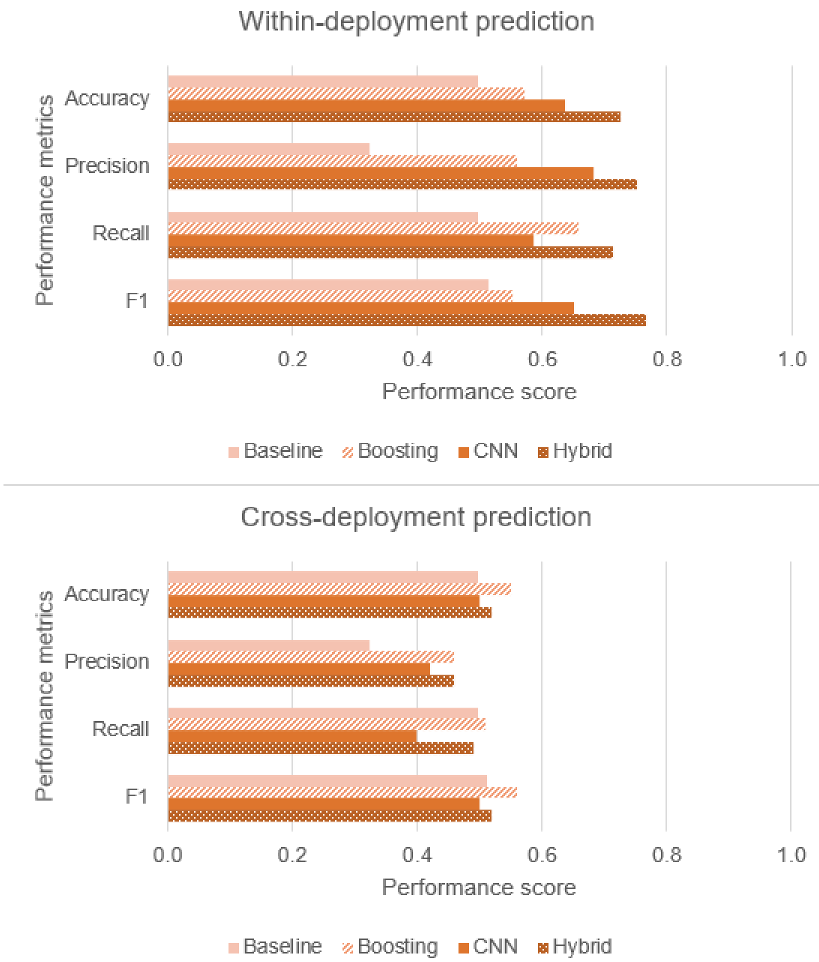

Figure 4.7. In-home cancer-pain episode prediction performance comparison between the within-deployment prediction, average across all five deployments, and the cross-deployment prediction where the data from all five deployments are combined to one dataset

The sensor deployment and the data collection of the cancer-pain management case-study also use the self-reporting approach for cancer-pain episode labels. During the sensor deployment at houses of cancer patients and caregivers, the patients help to mark their cancer-pain episode using smartwatch devices. Since the pain episode can require full and immediate attention from both the patient and caregiver to manage such as taking the pain medication, the self-report labels may be temporally imprecise. In figure

4.8, different data segmentation sizes are compared with the models' performance to predict in-home cancer-pain episodes. The result shows that using time segment of 30 minutes to 60 minutes before the cancer-pain labels yield the best weighted F1 score.

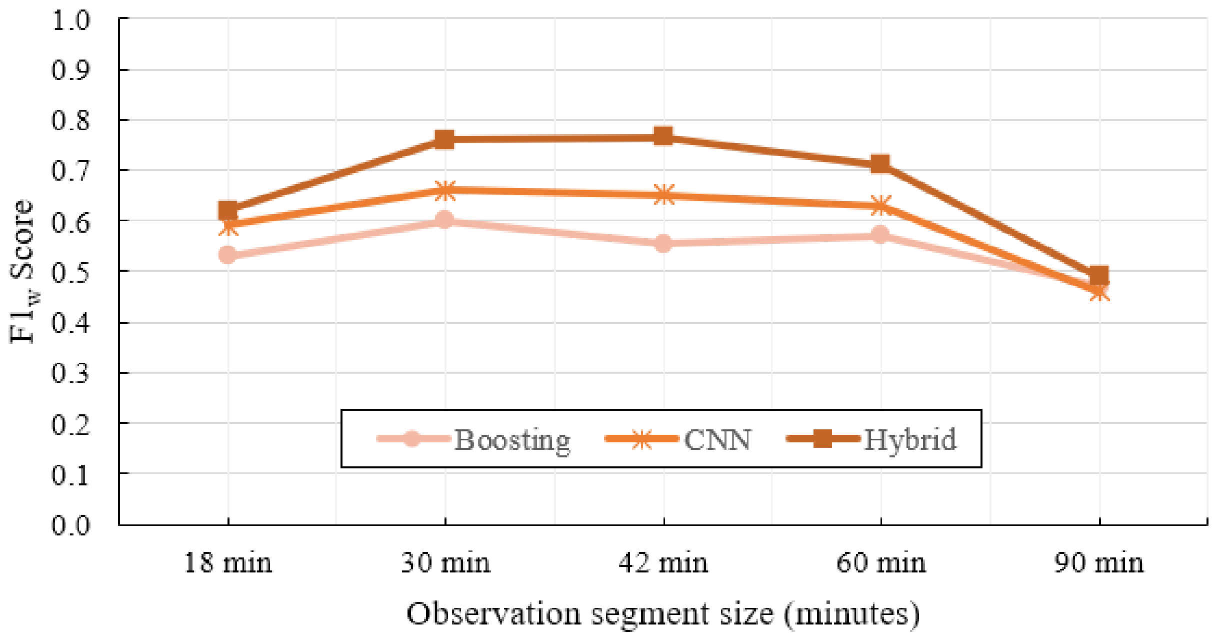

Figure 4.8. Effect of different data segmentation size on the weighted F1 score of the in-home cancer-pain episode prediction

4.3.3 Weighted sample and prediction sensitivity

In this chapter, the weighted sample technique is explained and implemented to help learning models deal with imbalanced label distribution. In a real-world application with a prediction machine learning model, such as the case-studies that predict health events from environmental time-series data, the sample-weight parameter can be tuned to make the predictive model more or less sensitive to the prediction. The sample-weight parameter represents the learning priority associate with each observation class. For example, a sample-weight of 4 means the negative observations only have one-fourth impact on the learning model during training compared to the positive observation. Figure 4.9 shows the receiver operating characteristic plot of the hybrid model predicting upcoming dementia-related agitation from environmental time-series. The receiver operating characteristic plot (ROC) illustrates the predictive ability of a binary classifier model. The ROC is plotted with the true positive rate (TPR) against the false-positive rate (FPR). The true positive rate, also called recall or sensitivity, represents the probability that the model will predict a positive class when the actual positive class is given. The false-positive rate or false alarm rate shows the probability that the model falsely predicts an observation to be positive. For this dementia-related agitation dataset, the positive observations or the time period that agitation has been reported is less than the negative observations with a ratio of one to four. In figure 4.9, the different sample-weight parameters can affect the model's ROC plot. The higher sample-weight gives higher priority to positive observations making the learning model more likely to predict a positive class but also more likely to produce a false-positive. Table 4.3 shows results of the effect of different sample-weights on the true positive rates and false negative rates of the prediction of dementia-related agitation with one to four ratio of positive to negative observations. As shown in table 4.3, in learning problems with an imbalanced dataset, there is a tradeoff between the TPR and the FPR. When the sample-weight is lower, the predictive model tends to have lower TPR or the probability that the model will predict a positive class when the actual positive class is given, but the FPR or the false-alarm rate is lower as well. This can be seen when the sample-weight is 1, the model has 0.5 TPR and only 0.09 FPR, given that the model will predict a positive class when the model's prediction probability is above 0.5. On the contrary, higher sample-weight such as the sample-weight of 10, the model has 80% probability to predict an agitation event if a known agitation time-period is given; but the model will also falsely predict a non-agitation period as agitation with 38% probability. Thus, the tuning of the sample-weight will depend on the application. For applications that being cautious is beneficial (such as coronavirus saliva tests [101]–[103]), a higher sample-weight is recommended. For other applications that frequent false alarms may be harmful such as household smoke detectors, lower the sample-weight can reduce the false positive rate but the cause of lowering the true positive rate should still be considered.

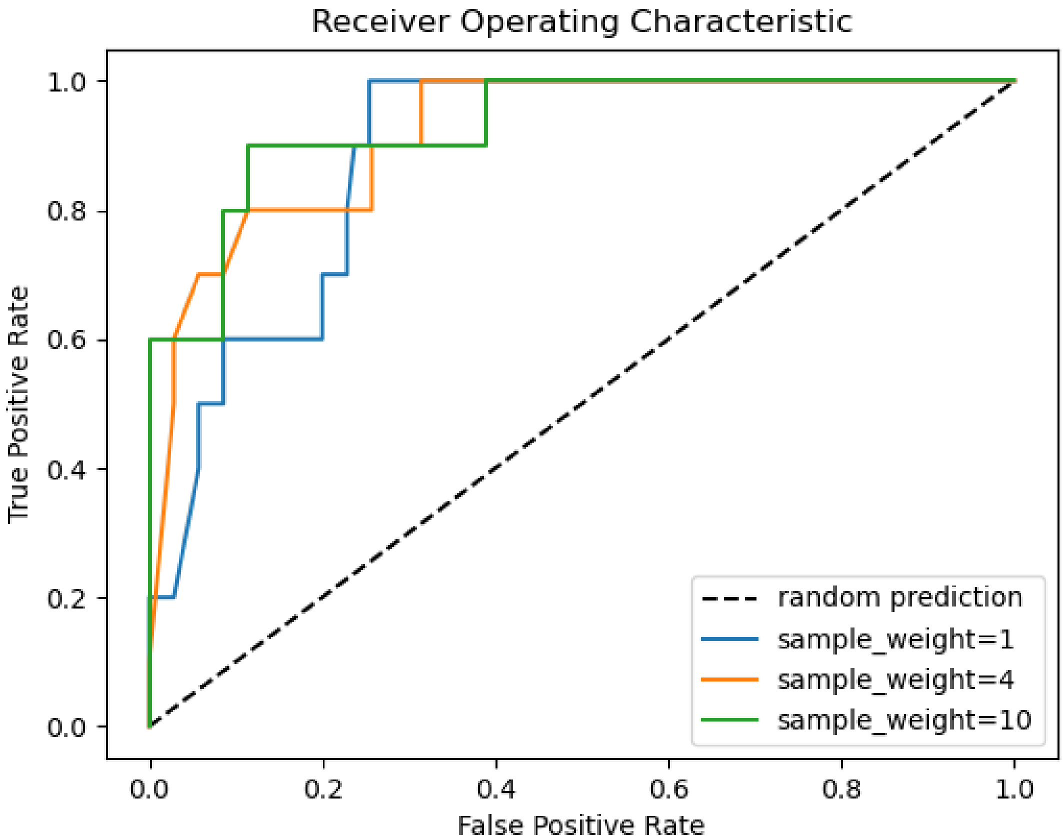

Figure 4.9. Different sample-weights and their effects on the true positive rate and false positive rate of the dementia-related agitation prediction model.

Table 4.3. True positive rate (TPR) and false positive rate (FPR) comparison at different sample-weights with prediction probability threshold = 0.5

	TPR and FPR at prediction probability threshold = 0.5	
	True positive rate (TPR)	**False positive rate (FPR)**
Sample-weight = 1	0.5	0.09
Sample-weight = 4	0.7	0.25
Sample-weight =10	0.8	0.38

4.3.4 Machine learning models and effects on early event prediction

When using a machine learning model for prediction tasks from time-series data inputs, the ability to predict the event of interest early can be crucial, especially in preventive applications such as informing patients of symptom interventions. The time-series data collected from in-the-field applications that use self-annotating labels, for training machine learning models, can show patterns regarding upcoming events before the user observes and labels. These early patterns, after the model is trained, can gradually increase the model's likelihood to predict an event of interest even before the label's timestamp that is used for training, especially for machine learning models that consider temporal information such as LSTM and CNN [67], [104], [105]. An example of the early event prediction is shown in figure 4.10 below. In the figure, the ambient environmental time-series on the left is used to predict whether there will be an upcoming dementia-related agitation episode. The red line on the left environment figure indicates the timestamp of a known agitation episode. Normally, the environmental time-series will be segmented from 6-minutes before the label to 48-minutes before (total of 42-minute segment window). The normal 42-minute time-series segment is feature extracted (rate-of-change) and input into a trained CNN model which output that there will be an upcoming agitation event with 0.68 prediction probability which is shown in figure 4.10 on the right at position 0 on the x-axis (Time window at 0 minutes). The model prediction probability is extracted from the weighted sum layer of a machine learning model. The weighted sum layer is a layer before the model uses the final activation function or threshold to decide the prediction class or output of the model [106]. Traditionally, a positive class threshold is set at 0.5 prediction probability, meaning if the probability is above 0.5, the model will predict a positive class. For a machine learning model with prediction tasks with time-series inputs, such as the prediction of cancer-pain episodes from ambient environmental time-series data, the time-series input(s) are continuously streamed into the predictive model. Learning models that sensitive to the temporal information in the data *may* be able to pick up patterns in the data which increases the prediction probability even before the known event marker, as shown in figure 4.10.

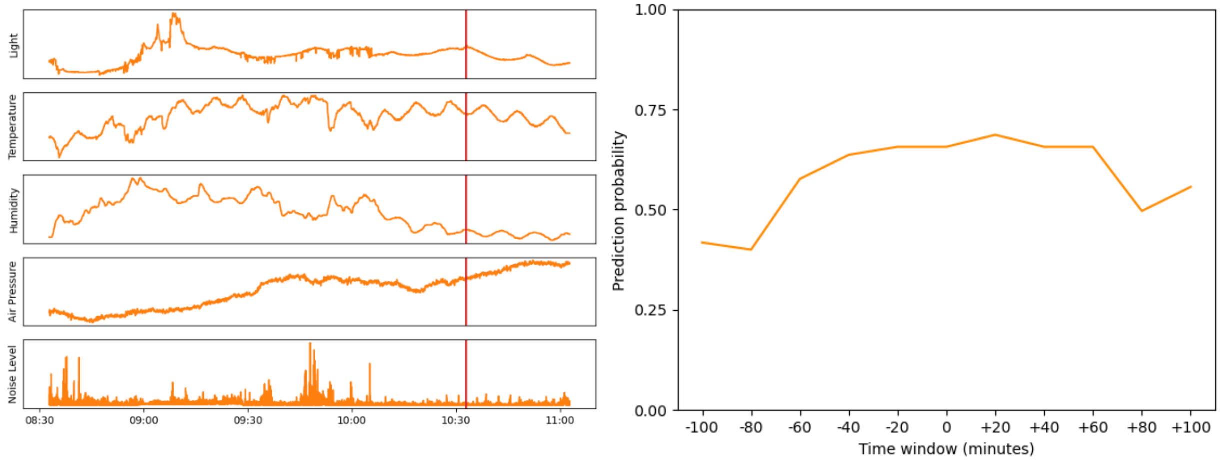

Figure 4.10. Early prediction of a dementia-related agitation event. The plot on the right shows that the prediction probability goes above the 0.5 threshold before the known agitation timestamp at the point of time-window 0.

Table 4.4 below shows the summary of the effect of the machine learning model and the early prediction. The early prediction refers to the period that the model's prediction probability goes above the positive-class threshold (threshold set at 0.5) before the known event markers. The early prediction duration in table 4.4 is the average duration across all five deployments in the dementia case-study and all five deployments for the cancer-pain management study. Here, it is shown that machine learning models that take temporal information in the time-series data into account such as CNN can predict health events earlier than the boosting model. The result also shows that the presented hybrid model can also early predict health events similar to the CNN. This is because the hybrid model uses part of the CNN model structure that reads temporal information in time-series data. Another observation is that the early prediction durations of the dementia study are longer than the durations in the cancer-pain management study. This could happen due to the differences in the self-reporting approach in the two case-studies in which the cancer patients are the ones marking the pain episodes but the dementia-caregivers are the ones observing agitation behaviors then mark the agitation events. Nevertheless, this section encourages the consideration of the early prediction when implementing predictive machine learning models in applications where event intervention or prevention is important such as symptom prevention and mitigation.

Table 4.4. Machine learning models and their effect on the early health event predictions.

Learning model	Early prediction (minutes)	
	Dementia-related agitation prediction	Cancer-pain episode prediction
Boosting model	8.6 minutes	3.0 minutes
Convolutional neural network (CNN)	25 minutes	13.3 minutes
Hybrid model	23.6 minutes	11.2 minutes

4.4 Conclusion and discussion

This chapter's results show that machine learning models can be implemented for health event prediction from time-series data which is generated from a continuous sensing system. In this chapter, the fundamental challenges associated with learning problems in the real world such as imbalanced label distribution, a small number of observations, and temporally imprecise label marker due to the self-labeling approach are addressed. The presented hybrid learning model approach, as shown in figure 4.2, takes advantage of the CNN model which can take temporal information in time-series data into account to mitigate the temporally imprecise label problem. The hybrid model also incorporates the regularization technique and the simpler boosting model to help with the model overfitting problem which comes from having a small number of event observations and the required personalized learning model. Metrics that can be helpful to considered when deploying a machine learning model in real-world applications are discussed. The weighted sample technique is utilized to address the imbalanced label distribution problem. Additionally, the sample-weight parameter can be tuned so that the learning model can focus more on the positive or negative label. Lastly, predictive learning models with time-series predictors may be able to early predict events of interest such as the early dementia-related agitation prediction shown in figure 4.10. The early prediction can occur due to certain time-series behavior such as the ramping up in the time-series data. For example, if the increase in ambient noise level is learned to triggered dementia-related agitation in the past, a learning model that considered temporal behavior in the input data should be able to gain increasing prediction probability to the prediction as well.

5. MODEL INTERPRETATION FOR REAL-TIME ACTIONABLE INTERVENTION

Machine learning has progressed significantly over the past two decades. Many fields of technology and science, such as finances, speech processing, image classification, and health care, have successes in applying machine learning [107]. Machine-learning algorithms have been developed to suit several types of data and learning problems. Generally, these algorithms can be viewed as searching through a large set of possible parameters, guided by training samples, to find a set of parameters that best optimize performance (e.g., highest classification accuracy, least regression error, etc.). Many of the widely-used machine-learning algorithms are considered black-box; Rudin explains that a black-box model could be a function that is too complicated for a human to understand or a proprietary function [108]. Examples of black-box models are the predictive deep learning models that can involve millions of parameters. Deep learning has been implemented in applications such as medicine, finance, and other fields that often require a more complicated model due to the complexity of the data and its learning problems, such as multivariable input data or image/time-series classifications that consider the temporal relationship of the data [109].

However, these implementations of the black-box model can be problematic. The lack of interpretability of the machine learning model can make it difficult for the user to take action or make a decision based on the model result, especially in healthcare applications where decisions can come with severe consequences. Recent research is trying to develop techniques to explain/interpret machine learning algorithms, especially the black-box models. Došilović defines two categories of approaches to model interpretability: (1) integrated interpretation, which traces the logic behind the model, and

(2) post-hoc, which extracts information from an already learned model [110]. Samek describes that, currently, machine-learning model interpretation methods often focus on understanding how the model works and why it produces a certain result [111]. This can increase the user's trust in the machine-learning algorithm and can be used to improve the model's performance. However, it does not always provide actionable measures, which are the important component of many applications. For example, if a black box model predicts a certain outcome (the outcome is usually known to the user), what action/decision should the user take to get the desired outcome?

In this work, we focus on the actionable interpretation of black-box machine learning models. Actionable interpretation is designed to inform users on what action/decision should the user take to get the desired outcome (or avoid undesired outcome). We also focus on learning problems with sequential/time-series data types that use black-box predictive models such as convolutional neural networks or deep neural networks. Our proposed technique uses a combination of information from the model outcome; the already learned model; the used training data; and the newly observed data to extract actionable items, as shown in Figure 5.1. The actionable items are aimed to help the user make decisions or take actionable measures based on the interpretation; as opposed to many black-box model interpretation studies that focus on understanding the model's decision (e.g. via model's logic, model parameters, etc.) to improve the performance of the model [111]. The extracted actionable items are related to the time-series behavior of the sequential input data that impact the model result. Examples of these time-series behaviors can be the abnormally low value of the data, the increasing of the data, or sudden changes in the data. The proposed actionable interpretation method is post-hoc, which uses information from the already-learned model without modifying the algorithm of the black-box model. This makes the proposed method robust and can be implemented for most traditional black-box predictive models with sequential/time-series data inputs.

As a case study, we implement our proposed actionable interpretation method with a real-world application of dementia-related agitation and ambient environment causation. Dementia patients express agitated behavior in various ways such as verbal outbursts or aggressive motor behavior [25]. The occurrence of dementia agitation can be affected by the ambient environment, such as the patient's long exposure to high ambient sound levels or bright light [73], [97]. We use a black box prediction model to predict upcoming dementia agitation symptoms (or behaviors) based on the in-home ambient environment around the dementia patient. Then, we implement our actionable interpretation method to determine the potential ambient environment that causes the patient's agitation and extract actionable items related to the ambient environment's behavior. An example of an actionable item, in this use case, could be to notify the in-home dementia caregiver to turn the lights on because the ambient light level is decreasing (e.g. the sun is setting), which has triggered agitation episodes in the past. The main contributions of this work are:

1. Actionable model interpretation technique that uses the prediction result, the already learned model, the used training data, and the newly observed data without modifying the black-box model.
2. Time-series analysis techniques, using permutation importance and cross-correlation, for the model's predictors importance and time-series behavior extraction.
3. Per-observation interpretation does not require the whole testing dataset to interpret models, which is suitable for applications that require real-time action/decision.
4. Actionable interpretation in a use case: ambient environmental causation for dementia-related agitation.

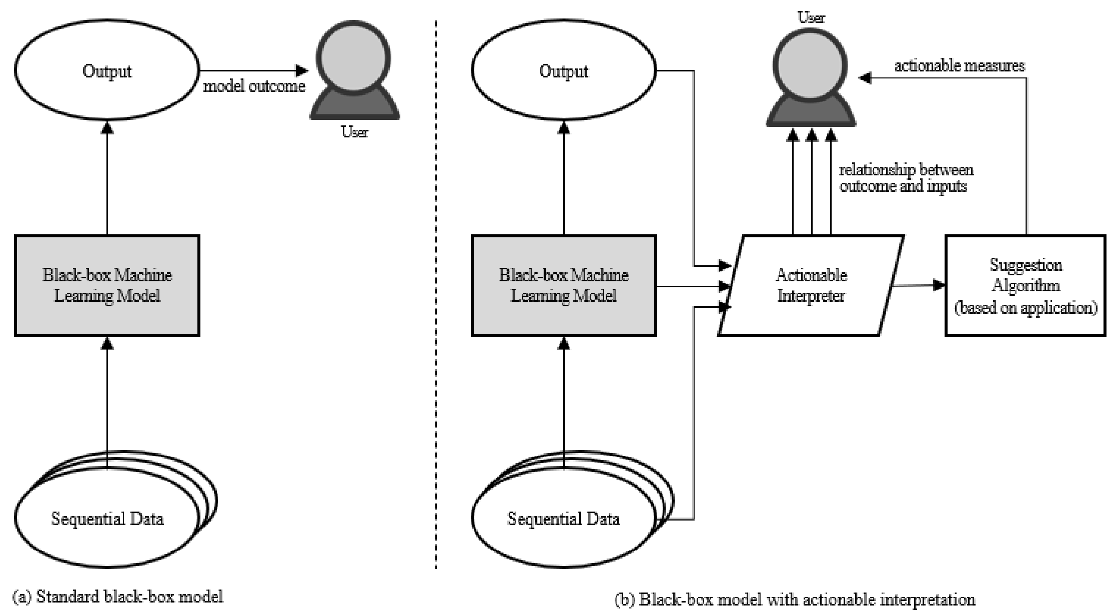

Figure 5.1. Standard black-box machine learning model vs. model with actionable interpretation. (a) Standard machine learning model: users only receive a model outcome, difficult to make a decision or take action based on the outcome only. (b) A model with the proposed actionable interpretation: users receive suggested actions based on the already learned model, the used training dataset, and the model outcome. No modification is made to the black-box model.

5.1 Related works

Recently, many works have been conducted in the model interpretation area. In this section, we discuss the related work regarding machine learning interpretability and our approach, actionable interpretation. The proposed approach focuses on interpreting black-box machine learning models to extract actionable items (e.g., the ambient environment's correlation to the prediction of an upcoming dementia-related agitation event). This actionable interpretation technique is aimed to help the user make decisions or take actions from machine learning models rather than interpreting models to validate outcomes or improve model performance.

Approaches to machine learning interpretation focus on explaining the model's logic by tracing through parts of the model [112]. Došilović categorized this type of model interpretation as an integrated interpretation in which the interpretability comes from the transparency of the model [113]. This interpretation approach can be limited to simpler models such as linear models and decision trees. In the decision tree models, the logic in the model such as tree split gains and sample weights can be used to explain the model [114]. Other works make sense of complex models by modifying the model to extract interpretable information. Martens extracts rules from support vector machines (SVM) to make the model interpretable for credit scoring [115]. Choi modified the recurrent neural network (RNN) to generate coefficients at certain stages of the network [116]. The generated coefficient can be used to calculate input contribution to the model's output. These works often rely on models with low complexity or require certain modifications to learning models that make them specific to a certain problem. Also, the integrated interpretations often give insights that can be used for the model's verification and improvement; but not necessarily appropriate in the extraction of actionable items. A work by Selvaraju et al., Grad-CAM, provides users with visual explanations from deep learning models by computing the gradient of a prediction class with respect to Convolutional Neural Network (CNN) feature map [117]. This informs users of the part of the input image that contributes to the feature map (contributes to the model's outcome). These integrated interpretations help improve user's trust in the model's outcome. However, they require model modifications, which may not be available in black-box models.

Another category of machine learning model interpretation is post-hoc interpretation, which extracts the model's information from the already learned model [113]. Krause validates predictors to the diabetes risk prediction by implementing partial dependence plots which show the changes in the model's output when the predictor's value changes [118]. Colubri evaluated the impact of the predictors to the prediction outcome of Ebola patients by analyzing the already trained model to rank the predictors [119]. In both of the mentioned studies, the predictor interpretation is done using all data samples – training samples, testing samples - to see which predictors contribute a high impact on the prediction outcome. This method may not be suitable to be implemented in a real-time situation, especially in applications that required real-time action/decision. In this work, we also address the real-time implementation by a predictor interpretation technique per-observation. Still, an advantage of these post-hoc model interpretation approaches is the robustness because they can be implemented in most traditional machine-learning models. The proposed actionable interpretation uses the post-hoc predictor interpretation to inform users on what action/decision the user should take to get the desired outcome.

Recent research on model interpretation in deep learning models is described as *“the analysis for human to comprehend why certain decisions or predictions have been made* [110].*”* A post-hoc model interpretation, LIME, explains model predictions to users such as identifying the area of importance in an image that corresponds to a certain prediction to improve user’s trust [120]. Another work, Axiomatic Attribution for Deep Networks, interprets models using the integrated gradients method [121]. The integrated gradient is the tracking of the model prediction, from a baseline input (informationless input such as a blank image) to a prediction of the actual input. The prediction score should increase with more information given to the baseline until saturation when it is identical to the actual input. The place where the prediction score is increasing is the gradient of interest which lets the users know the important parts of the input (e.g. area of an input image that impact the prediction score). These recent works on model interpretations can be used to inspect individual predictions that are suitable for real-world applications. However, they are used to increase the user’s trust in the learning algorithm, not suggesting users with potential actions to get the desired model’s outcome. Our proposed technique interprets the model’s predictions and predictor importance. Furthermore, the actionable interpretation also analyzes the predictors to extract information that helps users to take actionable measures. Similar to the Axiomatic Attribution for Deep Networks, this work also inspects individual predictions to help users make decisions. This is particularly useful in applications that benefit from real-time prevention and mitigation suggestions.

Black-box models have been implemented for sequential data such as video records and time-series data. In this work, we focus on the actionable interpretation of such models. The interpretation of models with sequential data can be complicated, as traditional model interpretation may lack insights on the temporal relationship in the sequential data. Norgeot conducted ambulatory outcome forecasting from time-series electronic health record features [122]. In the study, the features are extracted from window sequences in the time-series data. Then, model interpretation is done using the permutation importance scoring technique to rank the important features. This interpretation technique still only provides the user with information regarding the importance of the predictors. It does not give insight into the temporal behavior of the time-series predictors. Information on temporal behavior (e.g., increasing, decreasing) of time-series inputs could be directly used to take actionable measures such as disrupting unwanted time-series behavior. In this work, as part of the actionable model interpretation, we also provide techniques to analyze the behavior of the time-series/sequential predictors, which is crucial in the extraction of actionable measures for the models' users.

5.2 Model interpretation for actionable intervention

In this section, the model interpretation for real-time actionable intervention methods is explained. The interpretation for actionable intervention focuses on the extraction of items that the model's users can take action upon. The extracted actionable items are the behavior of the model's input data (sequential/time-series data). As an example, in our use case, dementia-related agitation prediction from the ambient environment, the extracted actionable items are the ambient environmental conditions and dynamics that are likely to trigger an agitation event. Such behaviors could be low-light-level, sudden changes in ambient sound level, or the decreasing in-home temperature/humidity. By learning the potential trigger of an upcoming agitation, the user (e.g. in-home dementia caregiver) can intervene or de-escalate agitation episodes.

An overview architecture of the actionable interpretation methods is shown in Figure 5.2. The methods consist of: (1) rate-of-change (ROC) calculation, (2) predictors ranking for sequential/time-series predictors, and (3) actionable items extraction. The ROC of a time-series input is utilized to help the actionable interpretation identify appropriate time-series behavior (e.g. stationery, changing). The ROC calculation is implemented both during the model training stage and when using the model for predictions. When there is new observation data, the already trained model uses both the observation data and its ROC as inputs. The outcome of the prediction model – model prediction (e.g. classification output) and prediction probability – is used in the predictor ranking. The predictor ranking is used to find impactful predictors; at this stage, the predictor ranking also provides information on whether the predictor's value of its ROC has more impact on the prediction outcome. This information is used for actionable item extraction which extracts behaviors of the impactful predictors. All of the mentioned techniques can be used to extract actionable items without any modification to the model, making them implementable to most black-box machine learning algorithms.

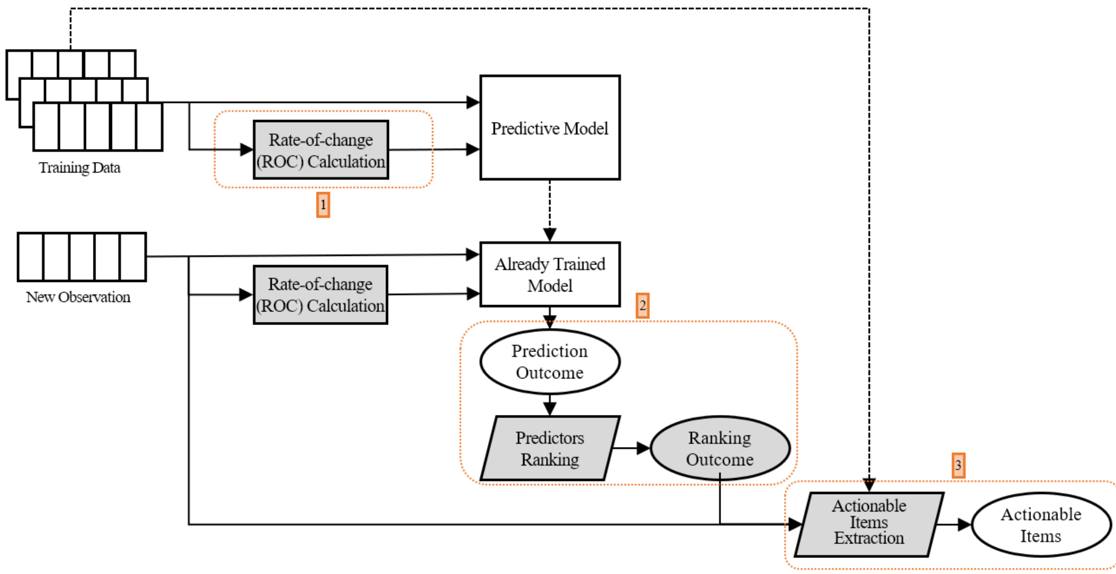

Figure 5.2. Actionable interpretation overview architecture: The interpretation uses an already trained model, used training data, and the model's outcome to extract actionable items and provide an actionable measure to users. Components of the actionable interpretation for black-box models are (1) rate-of-change calculation, (2) predictors ranking, and (3) extraction of actionable items.

5.2.1 Calculating rate-of-change (ROC)

The goal for the proposed actionable interpretation is to find the time-series behavior of the impactful predictors. The predictor's behavior can be statistically stationary or trend-differencing [123]. Thus, the calculation of the rate-of-change (ROC) in the time-series data is used to better understand whether it is the stationary behavior of the data (e.g. value too low) or the changes in the data (e.g. increasing, decreasing) that impact the prediction outcome the most. Before the model is trained, the ROC of the input/training data needs to be calculated. Then, both the training data and their ROC are used as inputs to train the model, as shown in Figure 5.2. When the already trained model is used, e.g. for prediction, the calculation of ROC is also required.

In this work, we use the absolute of a gradient of time-series data to calculate the ROC. Let a time-series array $X = X_0, X_1, X_2, \dots, X_i, \dots, X_{n-1}, X_n$. The gradient of a time-series array X is defined as:

$$\nabla X_i = \begin{cases} X_1 - X_0\,; & i = 0 \\ (X_{i+1} - X_{i-1})/2; & 1 \leq i \leq n-1 \\ X_n - X_{n-1}\,; & i = n \end{cases} \tag{5.1}$$

Thus, the ROC of a time-series data X is defined as shown in the equation below:

$$\mathrm{ROC}_X = |\nabla X| \tag{5.2}$$

In equation (4), time-series data with high noise can result in a high fluctuation in the ROC. Therefore, a noise reduction filter is recommended before the calculation of the ROC. Note that the requirement of a noise reduction filter depends on the application and its time-series data. In our use case, dementia-related agitation prediction from the ambient environment, a median filter is used to reduce the noises which come from the sensors used to collect ambient environmental data. More information is explained in the use case section. A graphical example of time-series data and its ROC is presented in Figure 5.3.

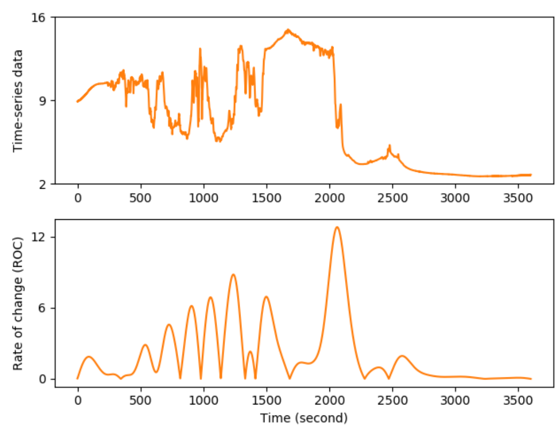

Figure 5.3. Time series plot of (a) original time-series data and (b) rate-of-changes

5.2.2 Predictors ranking

Predictor ranking is used to determine impactful predictors or predictors which highly contribute to the outcome/decision/predictions of the model. From the outcome of the prediction model, we record the prediction probability. The prediction probability can be acquired using methods depending on the machine-learning algorithm such as the output layer's value before the classification activation function (Neural Network-based algorithm), or the value of the last node before leaf nodes (Tree-based algorithm).

After the prediction probability is acquired, the predictor ranking uses the idea of permutation importance on time-series data [124]. The permutation importance is a post-hoc interpretation method in which the calculation is done after a model is trained, and does not require any changes in the model. In this work, we use permutation importance to determines the predictors that have a high impact on predictions by disarranging one predictor at a time (note that predictors can be a normal time-series data or its ROC); The disarranged predictor will lose any time-series pattern which contributes to the prediction of the model. Then, the already trained model is used for prediction again, but with a dataset with one disarranged predictor. A new prediction probability, from data with a disarranged predictor, is recorded and compared with the original prediction probability. This process of disarranging one predictor at a time, recording new prediction probability, and compare with the original probability is repeated until every permutation of one disarranged predictor is achieved. The predictor which, when disarranged, has a high impact on the prediction probability (compared to the original) is determined to be the impactful predictors. This predictor ranking technique uses the prediction probability of an observation to rank predictor importance, as opposed to the traditional approach which uses a set of observations [124]; Making this ranking method implementable for real-time application where the model only sees one observation at a time.

5.2.3 Extraction of actionable items

Actionable items are extracted from predictors which are determined to have a high impact on the model's predictions. In this work, we learn the time-series behavior of the high-impact predictors and extract actionable items from the behavior. For example, the proposed actionable interpretation can learn the model's prediction is impacted by the decreasing of a predictor. Then, based on the current application, we can produce an appropriate actionable measure to prevent/stop the decrease of the predictor. In this work, the detection of time-series behaviors can be divided into two categories: (1) statistically stationary, and (2) differencing or the changing in the time-series data [123].

During the predictor ranking step, if the predictor with a high impact on the predictions is the original time-series (not its ROC), stationary behavior will be extracted. First, the action interpretation learns basic statistics from the used training data which are the means and standard deviations of each predictor. The learned statistics from the used training data are compared with the average value of the high impact predictor. The comparison provides information on the predictor's stationary behaviors. Possible stationary time-series behaviors are the abnormally low value and abnormally high value. If the average value of the high impact predictor is significantly low compared to the mean and standard deviation of the same predictor, from the used training data, the behavior is inferred as an abnormally low value.

On the contrary, if the predictor ranking informs that the ROC of a predictor has a high impact, the actionable items will be extracted from the time-series differencing behavior instead. In this work, we focus on time-series differencing behaviors which are: increasing, decreasing, or sudden changes in the data. To learn the differencing behaviors, a cross-correlation between the original time-series predictor and our cross-correlation filters is computed. Let a time-series array $X = X_0, X_1, \dots, X_i, \dots, X_{n-1}, X_n$ and a cross-correlation filter $g = g_0, g_1, \dots, g_i, \dots, g_{n-1}, g_n$ For a better cross-correlation comparison, we normalized X with its mean $\bar{X}$ and standard deviation σ_x [125]. The cross-correlation filters are designed to have the same size as the time-series data and also normalized. The normalized cross-correlation between X and g is defined as:

$$(X * g)_i = \sum_{m=-\infty}^{\infty} \frac{(X_{m-i} - \bar{X})}{\sigma_x} \times \frac{(g_m - \bar{g})}{\sigma_g} \tag{5.3}$$

Cross-correlation represents the similarity between the predictor's pattern and the filters which are: increasing filter, decreasing filter, and sudden changes filters (positive and negative sudden changes). The cross-correlation filters are shown in Figure 5.4. After the cross-correlation is computed, we extract the maximum value. The point of maximum value represents when the two data arrays (time-series data and the filter), are most similar. The differencing time-series behavior with the highest cross-correlation maximum value is identified to be the impactful behavior.

The extraction of actionable items can provide insights into the action/decision that the user can take which will impact the predictions. The exact course of action is usually depending on the application which is discussed in the use case application section.

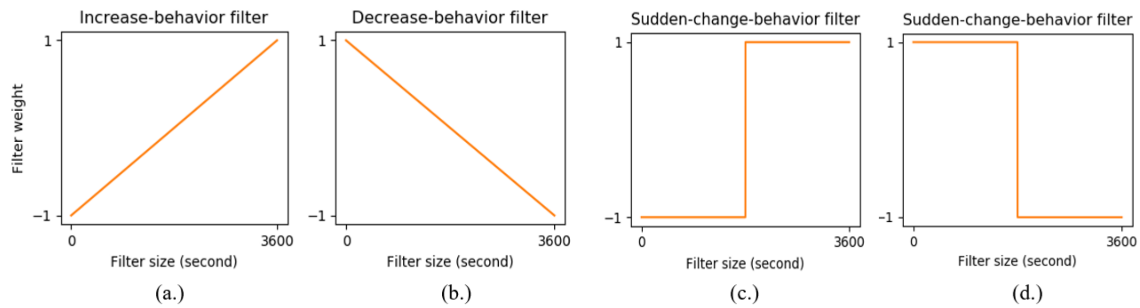

Figure 5.4. Cross-correlation filters: (a) increasing behavior filter, (b) decreasing behavior filter, and (c, d) sudden changes behavior filters.

5.3 Results

5.3.1 Comparison to related model interpretation techniques

The presented work, actionable interpretation, is compared with two other popular black-box machine learning model interpretation techniques, SHapley and LIME [120], [126]. SHapley interprets the impact of a feature in a certain value in comparison to the prediction the model would make if that feature is at a baseline value. Local interpretable model-agnostic explanations or LIME explains an individual instance by generating new data points based on the instance and train an easy-to-explain model to explain the black-box model behavior on the generated data. Lastly, the actionable model interpretation measures the importance of a feature by shuffling its values, then interpret the changes in the important input that will impact the prediction using cross-correlation filters. The comparison of these three black-box model interpretation techniques is shown in table 5.1.

Table 5.1 Comparison of black-box machine learning model interpretation techniques

	SHapley [126]	**LIME** [120]	**Actionable model interpretation** [127]
Works with black-box machine learning model	Yes	Yes	Yes
Supported data types	Numerical	Various. Numerical, categorical, texts, images [120]	Time-series
Ease of implementation	Easy to use	Complex. LIME required the user to manually define the explanation model	Easy to use
Flexibility/customizability	No customizability	Highly customizable. The explanation ML model can be modified	Somewhat. The cross-correlation filters can be modified based on applications
Explain individual prediction	No, required access to the full training dataset	Yes	Yes
Tolerance to unrealistic data instances	No. Shapley only has information from the model's training dataset	Yes. LIME generates new data points to interpret models. The explanation model can learn from generated data.[128]	No. The actionable interpretation only uses the distribution of the training data to interpret models
Interpretation stability	Yes. Shapley uses permutation importance as part of the interpretation [124]	Somewhat. The stability depends on how well the data generation and explanation model are defined. [129]	Yes. The presented work uses permutation importance as part of the interpretation
Interpret the changes in the input that impact the prediction (or change to the desired outcome)	No	No	Yes. The actionable interpretation uses cross-correlation filters to explain the inputs changes

*ML=Machine learning

The three black-box machine learning model interpretations are compared in multiple aspects. The ease of implementation represents the complexity when using the interpretation techniques to your black-box model. For SHapley and the presented work, only minor configurations are required such as giving the interpretation access to the training data. LIME, on the other hand, trains another simpler machine learning model to explain a black-box model. The explanation model has to be defined in advance which added complexity for the users. The flexibility of the interpretation technique shows the customizability of the technique to better suit the interpretation application. SHapley will always interpret the importance of all features – has no customizable options. For LIME, the user can manually define the explanation model which provides high flexibility. The actionable model interpretation can be used right out of the box or can be modified to suit various applications by modifying the cross-correlation filters. The next comparison metric is the algorithm's ability to interpret or explain individual observation which is desirable for real-world predictive model interpretations. LIME explains a new and unknown observation by generating similar data points to the new observation, and interpret how the model reacts to the data points. The actionable model interpretation finds the behavior of the new observation that is important to the prediction. On the other hand, SHapley analyzes all training datasets to explains black-box model behavior. Thus, Shapley cannot be used to explain an individual observation. The algorithm's ability to tolerate unknown data points is also important to ensure that the interpretation is working as intended. LIME generates data points around the data that it tried to explain. This data generation can create unknown data points from which the explanation model will learn [128]. The presented work uses the data distribution of the training dataset as part of the interpretation. Thus, there may be unexpected model behavior if unknown data instances were input into the model and interpretation algorithm. For interpretation stability, similar interpretations of the data should be offered if they are interpreting similar data sets. A study shows that LIME can explain two close data points very differently depending on the manually-defined data generation and explanation model [129]. The SHapley and the actionable interpretation uses the permutation importance technique to interpret data which has been proven to be a stable technique [124]. Lastly, the presented work, actionable interpretation, is designed to inform the user on what action/decision should the user take to get the desired model prediction. It learns the behavior of the input that has a high impact on the current model prediction. Thus, the user can decide on a cause of action to take or intervene undesired outcome.

5.3.2 Permutation importance in time-series data

Here, the evaluation of the predictor ranking approach is presented. A common way to perform predictor ranking using the permutation importance approach which relies on evaluating the model's outcomes when a certain predictor is absent [124]. The input data with the absence predictor is called the baseline input. In many applications, the baseline input is informationless such as a blank image assigning zero-value to all pixels or blank sentence when doing natural language processing [117], [121]. Ideally, predictor ranking using this baseline approach wants to compare the model's prediction probability at the baseline to the prediction when the model sees an actual input. However, for time-series data, using the informationless baseline might not provide the desired result comparison. For example, in the dementia-related agitation case study, ranking the light-level as a predictor means assigning all zero-value to the light-level data. The model may pick up the low light level (especially in the late afternoon) as the trigger for dementia-related agitation which is not the purpose of doing predictor ranking.

Figure 5.5 shows the comparison of the traditional baseline, assigning zero-value to the time-series input, and the presented baseline approach which shuffles the time-series input based on the means and standard deviation of the training data. This result uses the dementia-related agitation dataset of a deployment that light-level has been determined as the high-impact predictor. In this result, the baseline input is the input with the absence of the light-level data. The horizontal axis in figure 5.5 represents the ratio of baseline data. The prediction probability associate with the baseline data is represented at 0 horizontal value. As the horizontal value increases, the input data is modified to match the actual input. Thus, the horizontal value of 1 represents the prediction probability when the actual data is input to the prediction model. In an ideal scenario, if the predictor is impactful to the prediction, the prediction probability should be low at the baseline data and high at the actual data. Figure 5.5 shows that, when using the informationless approach as baseline data, the prediction probability path from baseline to an actual input misbehave in which the prediction probability decreases as the data becoming more similar to actual data. The presented approach which generates the baseline input based on the means and standard deviation of the training dataset shows a much more resemblance to the ideal scenario.

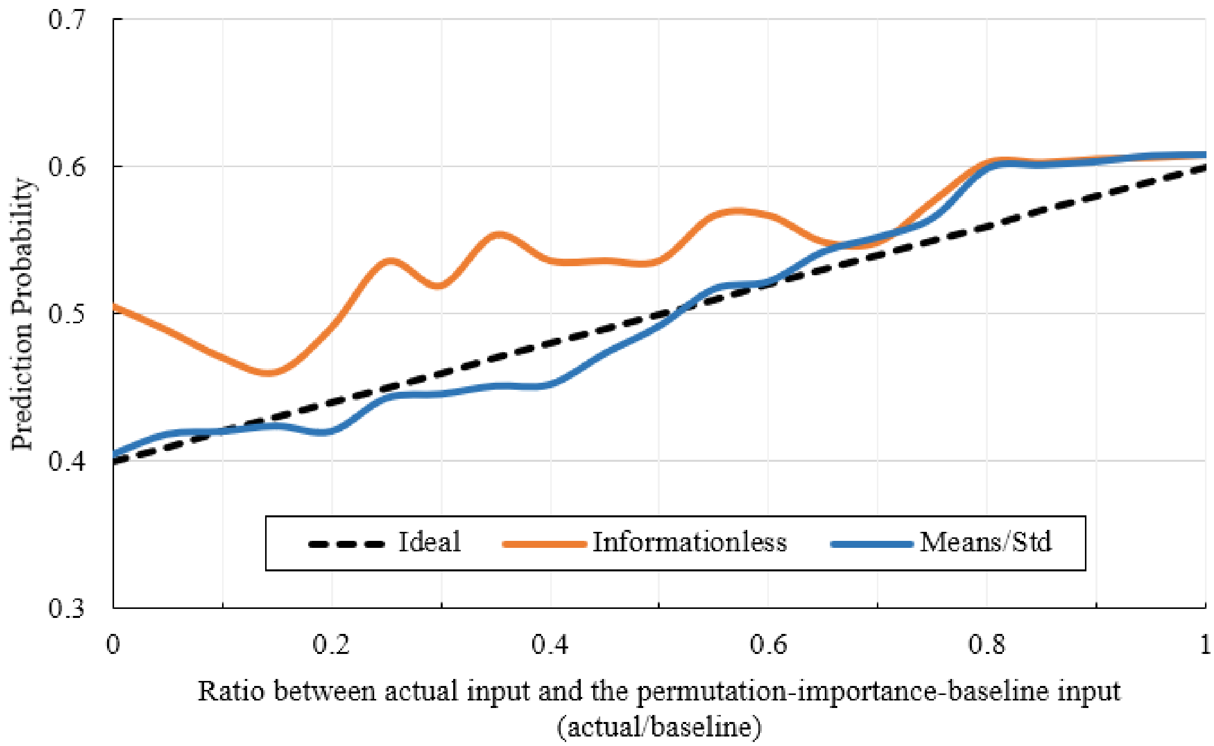

Figure 5.5. Prediction probability gradient between the baseline input and the actual time-series input. The baseline input is the modified time-series data that is inserted in the place of the actual input data to evaluate the impact on the prediction probability. The 'informationless' is the baseline input with only zero value; the 'means/std' is the baseline input generated by shuffling time-series data based on the means and standard deviation of the training dataset.

5.3.3 Case-study results

The proposed actionable interpretation method is used with the dementia-related agitation and ambient environmental use case. The actionable interpretation includes the calculation of ROC, predictor ranking techniques for time-series predictors, and actionable items extraction. By implementing the proposed technique, we extract ambient environment behaviors that correlate to the occurrences of dementia-related agitation in the homes of PWD.

Figure 5.6 shows the implementation of the actionable interpretation of one of the observations. The ambient environmental data, as shown in Figure 5.6.a, is predicted to cause an upcoming agitation by the predictive model. The predictor ranking technique determines important predictors by disarranging one predictor at a time and comparing the prediction probabilities. The predictor importance scores (shown in Figure 5.6.b) are computed by subtracting the prediction probability of a disarranged predictor from the original prediction probability. If a predictor has a high impact on the agitation prediction, when that predictor is disarranged, the prediction probability will alter drastically. This results in a high predictor importance score. Figure 5.6.b shows that, in this observation case, the ROC of light has the highest impact on the prediction of dementia-related agitation. Since the ROC of light is the important predictor, the actionable item is extracted from the differencing time-series behavior of the light data. Targeted differencing time-series behaviors are increasing, decreasing, and sudden changes in the data. Figure 5.6.c shows the result of the most likely behavior which is computed by cross-correlating the light data with cross-correlation filters. The similarity scores are then normalized mean value of the cross-correlation result. The result shows that the sudden changes in the light data are the most likely cause of the upcoming agitation episode. Thus, the actionable measure could be to close the window (if during daytime) or stop activities that cause fluctuations of light (e.g. CG walks in-out of bedroom).

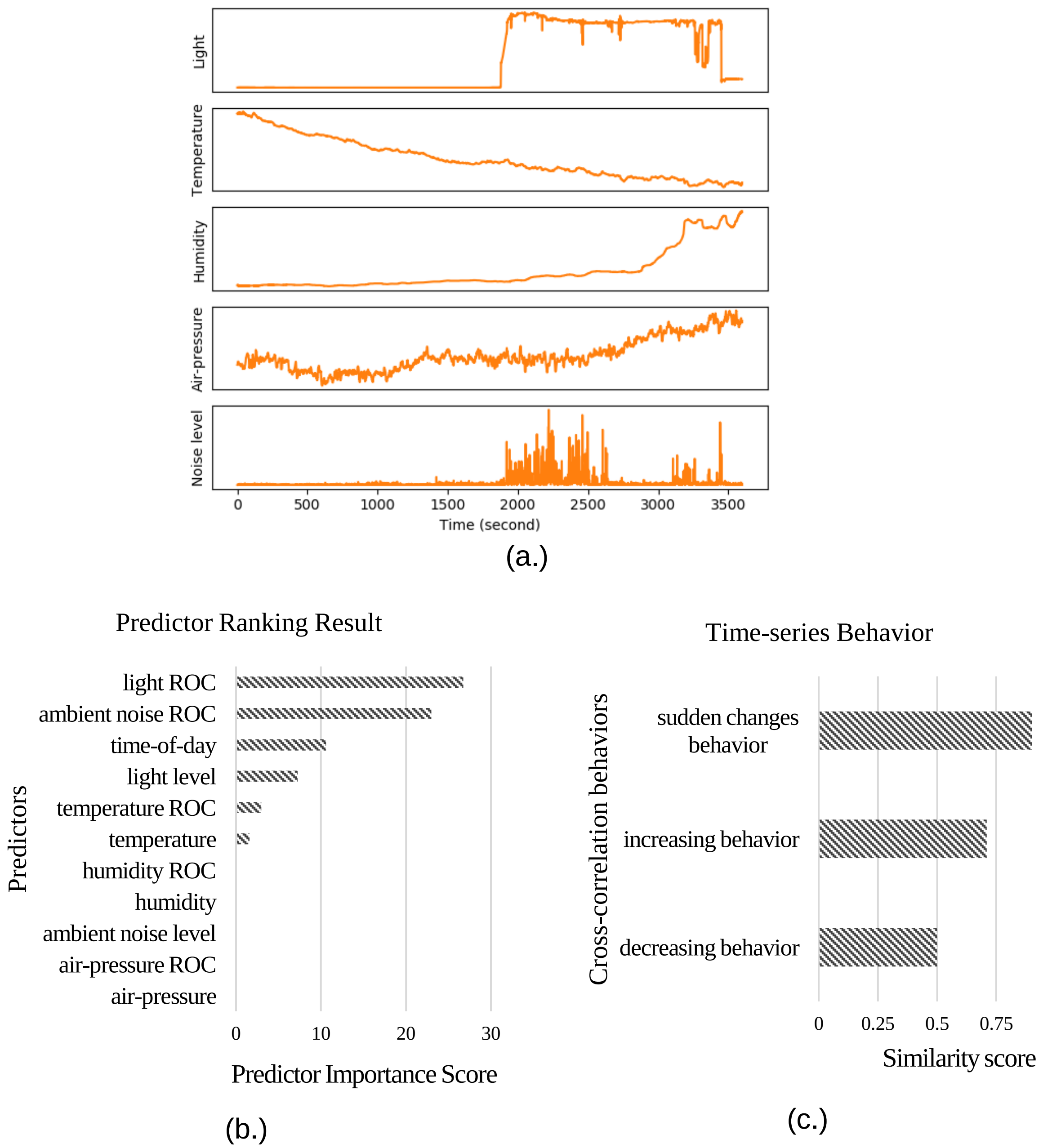

Figure 5.6. Actionable interpretation of dementia-related agitation on one observation. (a.) An ambient environment observation (one-hour period) that results in a positive prediction classification (upcoming agitation is predicted). (b.) Predictors ranking result showing rate-of-change (ROC) in the light as the most impactful predictor. Predictor importance score is the subtraction of disarranged prediction probability from the original probability. (c.) Cross-correlation results which present that the light ROC is showing sudden changes behavior.

The actionable interpretation technique is implemented on the use case of dementia-related agitation prediction from ambient environmental time-series data. Here, the interpretation can enable the user (i.e., in-home dementia CG) to be notified of upcoming PWD agitation episodes and to be provided with actionable measures to mitigate or potentially prevent the episode from occurring. Table 5.2 shows the summary result of the actionable model interpretation's find potential causes of dementia-related agitation and the ability to extract actionable items. For example, during the sensor deployment number 5 at the dementia patient's house, the result shows that the high ambient light level around the PWD can be observed during most of the reported agitations (65%). With this information, the CG can take action such as *"When hyperactivity is an issue, bright fluorescent lights should be turned down."* The results in table 5.2 also show that each PWD reacts differently to his/her surrounding ambient environment. Here, the most common ambient environmental triggers of dementia-related agitation are the light level which is ranked as the predictor, with the most impact on the agitation prediction in 3 out of 5 deployments, but the behavior of the light level that causes agitation varies. Other common agitation triggers are noise-level and time-of-day. The variety of potential agitation triggers, shown in table 5.2, shows that we need a personalized learning model to learn ambient environmental triggers for each PWD. This is supported by the results from the analysis of the predictive machine learning model and the dementia-related agitation prediction in the previous chapters which the individual learning model outperforms the global model.

Table 5.2 Actionable interpretation of the prediction of dementia-related agitation

Deployment No.	**Number of agitation observations**	**Agitation prediction performance**	**Actionable interpretation**	
			Common high-ranked predictor (% of all testing observations)	**Common actionable item**
1	81	*Acc: 0.79 **$F1_w$: 0.77	Time-of-day: 42% Noise level: 33%	High noise level

2	16	Acc: 0.71 $F1_w$: 0.71	Changes in noise level: 60%	Sudden changes in noise level
3	48	Acc: 0.82 $F1_w$: 0.81	Changes in light level: 52% Changes in noise level: 39%	Light level increasing Sudden changes in noise level
4	24	Acc: 0.71 $F1_w$: 0.76	Changes in light level: 47% Changes in ***temp.: 44%	Light level decreasing Temp. decreasing
5	73	Acc: 0.81 $F1_w$: 0.80	Changes in light level: 65% Time-of-day: 29%	High light level

*Acc: Accuracy, **$F1_w$: Weighted F1 score, ***temp.: Ambient temperature

The actionable items extracted using the presented model interpretation method are paired with agitation intervention suggestions. Table 5.3 shows examples of the actionable measures that are generated based on the extracted actionable items and are developed by our team's dementia experts (Anderson and Bankole). The intervention/prevention of in-home PWD agitations is essential, as dementia-related agitation can cause high burden and stress to the CG; and the CG burden associated with dementia agitation has been reported to be one of the principal factors prompting the institutionalization of community-dwelling PWD [114].

Table 5.3 Examples of actionable measures for dementia-related agitation intervention

Ambient environmental behaviors*	**Agitation intervention suggestions for CG**
High ambient noise level	Turn off the television/radio before you start talking to minimize background noise and distraction.
Sudden changes in ambient noise level	Consider lowering sound in the room to decrease agitation.

Low ambient noise level (no stimulation)	Provide appropriate levels of stimulation e.g. talking in a conversational tone, calming music.
Decreasing light level	To prevent sundowning, increase lights at least one hour before typical onset.
High light level	When hyperactivity is an issue, bright fluorescent lights should be turned down.
Sudden changes in ambient light level	Consider lowering or increasing light in the room to decrease agitation.
Decreasing in-home temperature	Make sure basic needs are met: temperature, comfort. Close windows and turn-on AC/heat if needed.
Time-of-day or daily schedule related	Offer distractions. Turn the action into an activity (pacing the house=hand them a broom and thank them for help)

*Time-series behavior that the actionable interpretation determines as high impact to the agitation prediction
**This table is a collaborative development with team of clinical experts from Virginia Tech Carilion School of Medicine

5.4 Conclusions and Discussion

In this chapter, an actionable interpretation of black-box machine learning models for time-series predictors is presented. The actionable interpretation informs users on what action/decision should the user take to get the desired outcome which is different from the interpretability methods used in other works that help to improve user's trust or to validate and optimize models. The presented technique aims to assist the model's user in taking actionable steps or to make a decision by extracting actionable items. The actionable items are based on the time-series behavior of the predictors that have a high contribution to the predictions. The actionable interpretation is robust and can be implemented in most traditional black-box machine learning models. The proposed technique achieves this because it is a post-hoc interpretation that utilizes an already learned model, used training data, and post-trained computations without any modification to the model itself. Time-series data analysis techniques, segmentation, feature extractions, and cross-correlation, are also utilized to acquire actionable items from a black-box model. This work shows that this technique can interpret each prediction. This makes the proposed technique appropriate in real-time applications, especially ones that benefit from prevention and mitigation suggestions.

The actionable model interpretation has been implemented to learn potential ambient environmental triggers to dementia-related agitations. The presented model interpretation was also used to extract actionable items, which can be paired with agitation intervention suggestions, to notified dementia caregivers to take informative action that may intervene or prevent the agitation episode. The result of agitation triggers also shows that each PWD reacts differently to their ambient environmental surrounding. This encourages the need for a personalized health event intervention approach which the works in this dissertation are trying to accomplish.

The proposed actionable interpretation also contributes to the studies of causality in machine learning. Usually, a machine learning model with good accuracy represents a high correlation between the input(s) and the outcome/target. It does not validate the causality between the inputs and the model outcome. The nature of proving causality in machine learning algorithms is still difficult in a lot of cases, and common sense or domain experts' opinion is used to infer the causality. Accordingly, Das states that "The general agreement in the statistics community is that you cannot prove a causal effect at least without performing an experiment" [130]. The actionable interpretation can be a tool that provides users with actionable measures. Then, actionable measures can be directly used in experiments to prove the causality effect of the input(s) and the outcome.

6. PERSONALIZED AND ADAPTIVE INTERVENTION SELECTION

In this section, a personalized, in-the-field symptom intervention selection method is explained. The proposed method uses a rule-based intervention selection technique and personalized the approach with user involvement. User involvement (or patient 'engagement') is the planning or development of health care with an emphasis on an individualized approach [29]. Patient engagement has increasingly received attention recently. However, research evidence still shows that the implementation of patient engagement in health care is limited [15].

For effective in-the-field healthcare, the clinician-in-the-loop approach is also required. In the last decade, technologies for in-the-field health care applications have emerged [131]. These technologies often tried to support in-the-field patients with their decease's symptoms which will benefit from clinician involvement. But with the raises of health care technologies (e.g. internet of things) and the required clinician involvement, these approaches may not scale well as clinical experts must be called for at all times. Thus, this personalized intervention strategy selection is proposed. The proposed work aims to automated intervention recommendations based on assessments from video-free sensing data and user reports. The clinical assessments are collaborative work with Virginia Tech Carilion School of Medicine. This combination of the user-involved clinical assessments and the proposed rule-based intervention selection enables the recommendations to be personalized to the user in the field.

6.1 Related works

Telemedicine has recently been shown to be effective at assessing and providing healthcare in the home, with remote clinical experts providing patients/caregivers with customized non-pharmacological intervention recommendations based on their video-assisted observations. For example, FamTechCare allows dementia caregivers to video record a person with dementia's agitated situations which are reviewed by a team of dementia experts who provide individualized interventions and feedback [94]. However, telehealth does not scale well as clinical experts must be ‘on-call’ at all times, or a professional team of experts must gather to review video recordings of home care incidences. The usage of video cameras can also raise concerns regarding privacy which contributes to the reluctance of using the technology. Furthermore, while the at-home user is interacting with telehealth devices and videos, they cannot give their full attention to the ongoing symptom or health crisis event. For instance, a troublesome dementia-related agitation episode crucially needs the full attention of the caregiver who will not be able to operate the technology such as record a video of the agitation event. Lastly, telehealth systems, such as FamTechCare, are reactive systems and not real-time. The system requires in-home users to actively record deceases’ symptoms to be reviewed by clinicians. The intervention, in this case, is provided to the CG after the health event which may not be effective. Many health crises, such as dementia-related agitation, are best managed at early stages before they escalate and proactively to prevent them from even beginning [93].

In this section, the issues with traditional telehealth are addressed with personalized and automated intervention recommendations based on assessments from video-free sensing data and reports from users. The symptom intervention automation can minimize or even eliminate the need for on-call clinical experts, team review by experts, video-related privacy concerns, user attention diversion to monitors or phones during periods of health crisis events, and purely reactive approaches to in-home healthcare.

6.2 Personalized intervention selection

The proposed approach utilizes clinical expert assessments and rule-based algorithms to personalized intervention strategies selection. The goal of this approach is to find disease severities (e.g. symptom severity, patient needs, caregiver burdens) and pair them with appropriate interventions. For example, if a dementia assessment battery finding indicates intellectual stimulation needs, a generated intervention such as “Pouring, squeezing, folding, scooping procedural movements are useful activities for stimulation” will be delivered to the in-home dementia caregiver. The overview framework of this approach is shown in the figure. 6.1.

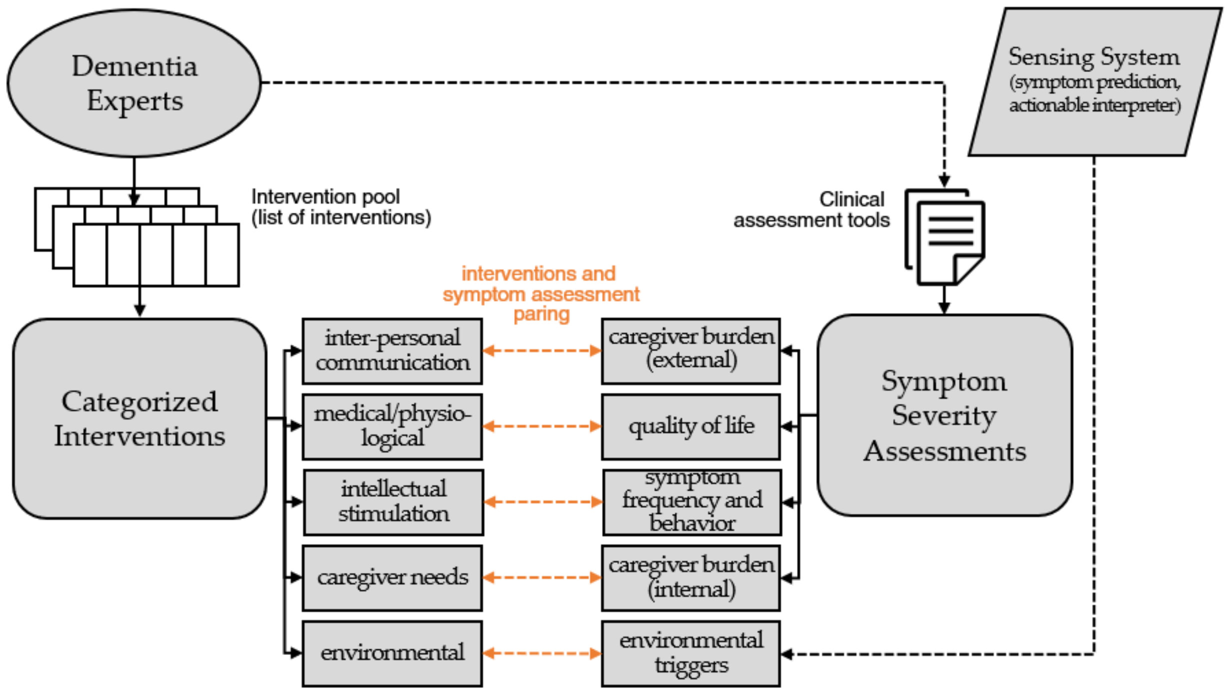

Figure 6.1. Personalized dementia-related agitation intervention selection which selects appropriate intervention strategies based on knowledge from clinical assessments and sensing data during real-world deployments.

To create personalized symptom severity assessment, the algorithm learns disease severities from clinical assessment tools. For the dementia-related agitation intervention case study, the clinical assessment tools are provided by dementia experts from Virginia Tech Carilion School of Medicine. The clinical assessment process involves an initial home visit before a deployment, formalized assessment, and clinical evaluation of the assessment tools. The mentioned clinical assessment tools for dementia are Center for Epidemiologic Studies Depression Scale (CES-D), Clinical Dementia Rating (CDR), Cornell Scale for Depression in Dementia (Cornell), Patient/Caregiver Demographics (Demographic), Pittsburgh Sleep Quality Index (PSQI), Neuropsychiatric Inventory Questionnaire (NPI-Q), Revised Scale for Caregiving Self-Efficacy (RSSE), and

Zarit Burden Interview (ZBI) [132]–[139]. CES-D is used to assess depression in caregivers. CDR is a clinical dementia rating for the assessment of cognitive and functional impairment. Cornell rates symptoms of depression in persons with dementia. Demographic is an assessment of PWD behaviors and emotions. PSQI assesses factors relating to sleep quality. NPI-Q is an inventory that assesses types of neuropsychiatric disturbances. RSSE assesses caregiver's level of confidence in their ability to perform activities. And ZBI scores caregiver burden which reflects how caregivers feel about taking care of another person. These assessment tools are interviewed from an in-home patient with dementia and his/her caregiver by dementia experts. The assessment tools produced hundreds of questionnaires. Then, the proposed rule-based technique is implemented with dementia experts' opinion to categorized the questionnaires into four groups: (1) symptom frequency and behavior, (2) caregiver burden due to external sources (from person-with-dementia), (3) caregiver burden internally (e.g. stress, ability to obtain help if needed), and (4) quality of life. Another group of symptom severity assessment is the environmental triggers of agitation which is obtained from the mentioned sensing system and data analysis which is shown in the previous chapter on the actionable model interpretation [127].

The personalized assessment of symptom severity is paired with appropriate intervention suggestions. First, dementia experts evaluate a pool of hundreds of dementia intervention suggestions and produce related dementia agitation interventions [140], [141]. These interventions are categorized by the rule-based algorithm into five groups: (1) intellectual stimulation, (2) inter-personal communication, (3) caregiver needs, (4) medical/ physiological, and (5) environmental-related interventions. Each intervention group is consisted of associated intervention suggestions and paired with symptom severity assessment categories to create intervention suggestions that are personalized to the patient/caregiver due to their involvement in the symptom assessments. Table 6.1 shows the clinical assessment tools, the selected questionnaires, and their associated intervention suggestions. Answers to the selected questionnaires will be determined by the algorithm to measure the severity associated with the question. For example, the rule-based algorithm considers PWD has severe sensory deficits if the answers to PWD sensory deficits are either "has limited hearing" or "is using hearing aid" or "deaf-related", Then, the personalized intervention selection algorithm will deploy the appropriate intervention suggestions to the user (e.g. dementia caregivers). The paired interventions are also shown in table 6.1.

Table 6.1. List of assessment tools, selected assessment questionnaires, and their corresponding intervention suggestions

Clinical tool	Selected questionnaires	Associated intervention suggestions
CES-D	I was bothered by things that usually don't bother me.	Take care of your needs. Try meditation, deep breathing, spiritual resources.
	I did not feel like eating; my appetite was poor.	Try not to postpone your hunger, thirst, toileting, rest in favor of theirs. It may take just a few moments to help yourself so you are better able to help them.
	I felt that I could not shake off the blues even with help from my family or friends.	Reward yourself for doing a great job as a caregiver.
	I felt depressed.	Have a list of friends, family, neighbors, church family etc. for you to call if you need someone to relieve you.
	I felt that everything I did was an effort.	Reward yourself for doing a great job as a caregiver.
	I thought my life had been a failure.	Remember that feelings of sadness, stress, anxiety, anger, guilt, grief and frustration are normal feelings for caregivers of someone with dementia. It helps to share those feelings with other caregivers.
	I felt fearful.	Remember that feelings of sadness, stress, anxiety, anger, guilt, grief and frustration are normal feelings for caregivers of someone with dementia. It helps to share those feelings with other caregivers.
	My sleep was restless.	Consider soft music, a sound machine to calm you
	I talked less than usual.	Regularly assess your own mood.
	I felt lonely.	Have a list of friends, family, neighbors, church family etc... for you to call if you need someone to relieve you.
	People were unfriendly.	People often don't understand dementia and the load of caregiving. Try to change your response to their lack of understanding
	I had crying spells.	Take care of your needs.

	I felt sad.	Remember that feelings of sadness, stress, anxiety, anger, guilt, grief and frustration are normal feelings for caregivers of someone with dementia. It helps to share those feelings with other caregivers.
	I felt that people dislike me.	Keep a list of caregiver support groups in your area and set aside time at least once per week to spend time with other caregivers.
	I could not get *going*	See your own family doctor. Ask for help with caring for PWD
CDR	Has there been some decline in memory during the past year?	1. Make a memory scrapbook or look through old photo albums together. 2. Label things like the key bowl or glasses tray to help them keep track of easily lost items.
	What was his/her main occupation/job (or spouse's job if patient was not employed)?	Don't hesitate to replay a favorite sports event, movie, concert, positive news event, home movie.
	What was his/her major job (or spouse's job if patient was not employed)?	Don't hesitate to replay a favorite sports event, movie, concert, positive news event, home movie.
	What hobbies did the PWD enjoy prior to illness?	Don't hesitate to replay a favorite TV shows or movies
	What can he/she still do well?	Don't hesitate to replay a favorite sports event, movie, concert, positive news event, home movie.
	What is your estimate of his/her mental ability in: Eating habits?	Place finger snacks and drinks within eye sight throughout the day.
	What is your estimate of his/her mental ability in: Sphincter control?	Consider bowel status: Could constipation be a problem? Is diarrhea present? Be sure to keep hydrated by drinking water or Gatorade-type drinks.
Cornell	Anxiety	To diminish worry and fretting, minimize information shared about upcoming activities.
	Sadness	Allow them to express themselves, if possible. Offer comfort.
	Lack of reaction	Apathy is a common symptom in dementia and is hard to change
	Irritability	Check for possible pain. Make sure basic needs are met: temperature, hunger, thirst, comfort, rest.
	Agitation	1. Create a calm and predictable environment by limiting distractions and removing stressors. 2. Play favorite music to set the mood for the activity or time of day. Like calming music in the evening or lively music during the day.
	Retardation	Allow for plenty of time to answer any questions.

	Multiple physical complaints	Offer distracting activity during personal hygiene care e.g. Hand them a washcloth to clean face
	Loss of interest	Relax expectations on ability to participate in tasks even if they were able to do it yesterday. Encourage gently.
	Appetite loss	Place finger snacks and drinks within eye sight throughout the day.
	Weight loss	Make sure basic needs are met: temperature, hunger, thirst, comfort, rest.
	Lack of energy	Consider talking to their physician about any medications that could be making them tired
	Variation of mood	Reduce expectations in the morning. It may take time to get motivated
	Difficulty falling sleep	Discourage excessive daytime sleeping.
	Multiple awakenings during sleep	Naps before mid-afternoon may be helpful but later may disturb nighttime sleep.
	Early morning awakening	Create a bedtime routine of not drinking liquids before bedtime.
	Suicidal	1. Ask what they mean. 2. Talk to your medical provider
	Poor self-esteem	Offer reassurance. Reminiscence. Review family pictures
	Pessimism	Controlling emotions is difficult for some PWD.
	Mood congruent delusions	See the medical provider.
Demographic	Bowel habits	1. Consider bowel status: Could constipation be a problem? 2. Consider bowel status: Is diarrhea present? Be sure to keep hydrated by drinking water or Gatorade-type drinks.
	Urinary Control Issues	Check to see if toileting is needed. Dry garments are more comfortable.
	Appetite	Offer small meals or healthy snacks often. Finger foods may be more appealing.
	Sleep	1. Discourage excessive daytime sleeping. 2. Naps before mid-afternoon may be helpful but later may disturb nighttime sleep.
	Energy/alertness during the day	1. Use energy of restlessness with an activity suitable to their ability. 2. Prepare for anxious times by planning pleasant activities before typical time of anxiety.
	Interventions	1. Play familiar/favorite music to set the mood for the activity or time of

		day: calming music in the evening or lively music during the day. 2. Turn the action into an activity (e.g. pacing the house=ask for help with laundry)
	Pain	Check for possible pain. Make sure basic needs are met: temperature, hunger, thirst, comfort, rest.
	Sensory Deficits	1. Turn off the television/radio before you start talking to minimize background noise and distraction. 2. Be sure to accommodate for hearing impairment. Pronounce words clearly. Check hearing aid(s).
	Speech, Volume & Tone	1. Try and determine the goal they are pursuing and support that goal in a positive way. 2. Allow for plenty of time to answer any questions. 3. Ask "yes" or "no" questions if possible.
	Mood	Provide reassurance with responses like: "You are safe." "I am here to help you."
PSQI	Cannot get to sleep within 30 minutes	Naps before mid-afternoon may be helpful but later may disturb nighttime sleep.
	Wake up in the middle of the night or early morning	Avoid stimulants such as caffeinated or sugary foods/beverages and also avoid ""downers"" like alcohol and Benadryl (diphenhydramine).
	Have to get up to use the bathroom	Discourage excessive daytime sleeping.
	Cannot breathe comfortably	Encourage daily exercise. Any movement is better than none.
	Have bad dreams	Try out 'new' types of being active such as exercising to a tape or TV program, marching in place, dancing to the radio, etc.
	Have pain during sleep	Check for possible pain. Make sure basic needs are met: temperature, hunger, thirst, comfort, rest.
NPI-Q	Delusion	Actions for delusions - redirect. Do not ignore but change the topic and keep patient safe
	Hallucinations	Actions for delusions - redirect. Do not ignore but change the topic and keep patient safe
	Agitation/Aggression	Remove the source of agitation if possible
	Dysphoria/Depression	Ensure safety. Offer comfort such as pets, hand massage, calming music, etc.
	Anxiety	Acknowledge feelings rather than words. Allow the safe expression of

		fear and frustration. E.g. "I can see that you're anxious."
	Euphoria/Elation	If the symptoms do not bother the PWD, then ignore
	Apathy/Indifference	Apathy occurs often in dementia and is hard to motivate. Try not to be bothered by it. Encourage mobility and bathroom breaks to keep PWD up and moving
	Disinhibition	Communicate in a calm, soothing manner. Do not raise your voice.
	Irritability/Lability	To diminish worry and fretting, minimize information shared about upcoming activities.
	Aberrant Motor	If not a safety issue, try to minimize effects such as offering quieter options for motor behaviors such as tapping. If pacing, clear environment so they can pace safely.
	Nighttime Behavior	1. Naps before mid-afternoon may be helpful but later may disturb nighttime sleep. 2. Offer less liquids before bedtime and be sure to toilet before bedtime
	Appetite/Eating	1. Consume nutritious foods and beverages, limit refined sugar and caffeine. 2. In severe dementia, may want to eat only a few favorite foods. May be ok to avoid losing weight.
RSSE	How confident are you that you can ask a friend/family member to stay with PWD for a day when you need to see the doctor yourself?	When things are calm, review your available resources for respite, entertainment, social activities, etc.
	How confident are you that you can ask a friend/family member to stay with PWD for a day when you have errands to be done?	When things are calm, review your available resources for respite, entertainment, social activities, etc.
	How confident are you that you can ask a friend or family member to do errands for you?	Have a list of helpful ways your friends, family, etc. can help regularly. If someone offers their time, consult your list.
	How confident are you that you can ask a friend/family member to stay with PWD for a day when you feel the need for a break?	If family lives out of town, they can still help. Consider asking them to: Plan on coming to town and watching over PWD once a month for a day/weekend to give you relief; Call each week to tell PWD "I love you" or "How are you doing?"
	How confident are you that you can ask a friend/family member to stay with PWD for a week when you need the time for yourself?	1. Remember that feelings of sadness, stress, anxiety, anger, guilt, grief and frustration are normal feelings for caregivers.

		2. Keep a list of caregiver support groups in your area and set aside time at least once per week to spend time with other caregivers.
	When PWD forgets your daily routine and asks when lunch is right after you've eaten, how confident are you that you can answer him/her without raising your voice? (Clarify that "answer" can be direct or a distraction.)	1. Call the Alzheimer's Association national help-line any time of day at 1-800-272-3900. 2.Do not argue, ignore or raise your voice.
	When you get angry because PWD repeats the same question over and over, how confident are you that you can say things to yourself that calm you down?	1. Try not to become frustrated when answering the same question over and over. 2. Remember that repetition is a symptom of dementia. They can't recall having just asked that question or performed that behavior.
	When PWD complains to you about how you're treating him/her, how confident are you that you can respond without arguing back? (e.g., reassure or distract him/her?)	1. Provide reassurance with responses like: "You are safe." "I am here to help you." 2. Offer distracting activity during personal hygiene care e.g. hand them a washcloth to clean face"
	When PWD asks you 4 times in the first one hour after lunch when lunch is, how confident are you that you can answer him/her without raising your voice?	1. Remember that repetition is a symptom of dementia. They can't recall having just asked that question or performed that behavior. 2. Always have a calm approach
	When PWD interrupts you for the fourth time while you're making dinner, how confident are you that you can respond without raising your voice?	Redirection is a useful tool for many behaviors in dementia. Try offering something else to do/talk about/ focus on.
	How confident are you that you can control thinking about unpleasant aspects of taking care of PWD?	Remember that feelings of sadness, stress, anxiety, anger, guilt, grief and frustration are normal feelings for caregivers of someone with dementia. It helps to share those feelings with other caregivers.
	How confident are you that you can control thinking about what a good life you had before PWD's illness and how much you've lost?	Consider scheduling a regular weekly respite to meet your social needs (outside of your monthly caregiver support group)."
ZBI	Do you feel that because of the time you spend with your relative that you don't have enough time for yourself?	1. Make a habit of setting aside time for yourself every day. CB 2. Consider adding to your list of ways friends/family can help: Family spending time with the PWD; Shopping and cooking with PWD; Taking PWD to doctor visits, get haircut, errands; cleaning your home.
	Do you feel stressed between caring for your relative and trying to meet	Take care of your needs.

	other responsibilities for your family or work?	
	Do you feel angry when you are around the relative?	Remember that feelings of sadness, stress, anxiety, anger, guilt, grief and frustration are normal feelings for caregivers of someone with dementia. It helps to share those feelings with other caregivers.
	Do you feel that your relative currently affects your relationships with other family members or friends in a negative way?	If family lives out of town, they can still help. Consider asking them to: Plan on coming to town and watching over PWD once a month for a day/weekend to give you relief; Call each week to tell PWD "I love you" or "How are you doing?"
	Do you feel strained when you are around your relative?	Do not use restraints
	Do you feel that your health has suffered because of your involvement with your relative?	1. Do not try to be the caregiver alone. Share what you are learning about dementia and caregiving with your family. It may help them think how they can help you or help them understand why you need planned time off. CB 2.Put your health at the top of your list today. Is there a neglected area you could check off (e.g., MD appt., exercise, dentist)?
	Do you feel that you don't have as much privacy as you would like because of your relative?	Make a habit of setting aside time for yourself every day.
	Do you feel that your social life has suffered because you are caring for your relative?	Consider scheduling a regular weekly respite to meet your social needs (outside of your monthly caregiver support group).
	Do you feel that you have lost control of your life since your relative's illness?	When things are calm, review your available resources for respite, entertainment, social activities, etc.
	Do you feel uncertain about what to do about your relative?	Call the Alzheimer's Association national help-line any time of day at 1-800-272-3900.
	Do you feel you should be doing more for your relative?	Call the Alzheimer's Association national help-line any time of day at 1-800-272-3900.
	Do you feel you could do a better job in caring for your relative?	Call the Alzheimer's Association national help-line any time of day at 1-800-272-3900.

*This is a collaborative work with clinical experts from Virginia Tech Carilion School of Medicine

In an actual deployment of the intervention suggestions, many of the questionnaires may be scored as severe. As a result, too many intervention suggestions are generated. Thus, the selection algorithm limits the intervention suggestions to four interventions at a time. In other words, only the four interventions with the highest priority will be sent to the user at a time (see priority algorithm below). To summarize, the personalized intervention selection algorithm is shown below.

ALGORITHM 6.1

Input: Clinical assessment tool scores/answers

Output: Personalized intervention suggestions

1: Use a rule-based approach to determine the symptom severity level of each assessment tool questionnaire (e.g. a symptom is considered severe if the score is above a certain threshold)

2: Select all intervention suggestions paired with the severe symptoms

3: Rank the top four intervention suggestions

- 3.1 Intervention related to the NPI-Q caregiver distress level has the highest priority (if there’s any).
- 3.2 Repeat interventions rank second on the priority list. (Some questionnaires can associate with the same intervention suggestions).
- 3.3 Third-priority interventions are from categories frequently selected by *Step 2.* The categories are intellectual stimulation, inter-personal communication, caregiver needs, and medical/ physiological.

6.3 Case-study Result

The personalized intervention strategy selection has been implemented, deployed, and delivered to real-world person-with-dementia and caregivers. The real-world deployments are from the dementia case-study presented in chapter 2. All five participants from the case-study participated and receives automated and personalized dementia-related intervention suggestions. While each system deployment lasted for 2-months, the real-time intervention suggestions are active for the last month of the deployments. For each real-world deployment, the assessments of the patient/caregiver dyad are done at the beginning by dementia experts. Then, the assessments are used as inputs to the rule-based intervention selection algorithm which provides lists of suggestions personalized to the dyad. The customized intervention suggestions are stored on a cloud service, for an automated intervention delivery system. Whenever the BESI system detects upcoming dementia-related agitation episodes, it notifies and delivers the suggestions to the caregiver via a wearable (for just-in-time notification) and tablet application (for a more-detailed intervention suggestion). The findings of this result section are from the dementia caregivers' feedbacks and qualitative scores measured at the end of each deployment.

From post-deployment surveys regarding caregiver acceptance of the intervention system, caregiver responses were largely positive. These responses indicate that the system is relatively easy to use with the score ranges from 1-3 where 1 indicates that the system is very easy to use and 6 indicates that the system is very difficult to use. The caregivers also show positive feedbacks to system usability questions and most found it was not intrusive, saying *"They were not a bother to me"*.

The results regarding the intervention suggestions are summarized and shown in table 6.2. Here, the quantitative scores such as the number of agitation reports, average agitation severity level, intervention suggestions rating, and clinical tool scores are shown. Table 6.2 compares the scores between before and after personalized suggestions and agitation notification are delivered to caregivers. Each period (before intervention and after intervention) lasts for 1-month. All caregivers gave positive survey responses to the individualized caregiver suggested interventions. The categories of 'caregiver needs' and 'interpersonal communication' were both received positively. Agitation scores, caregiver distress, caregiver depression, and stage of dementia in the PWD were measured. In general, 2 of the 5 caregivers measured decreased agitation in their PWD from before intervention to the end of deployment. The pilot intervention suggestions were positively noted in post-deployment interviews with caregivers. Furthermore, 3 out of 5 caregivers measured decreased distress from the NPI-Q caregiver distress score, and one caregiver also show more confidence in obtaining respite (increased RSSE score).

Table 6.2. Scores comparison between before and after personalized suggestions and agitation notification are delivered to caregivers

Dept.	**Agitation reports**	**Intervention Scores**	**Clinical tool score**	**Additional comments**
1	Averaged agitation level decreased by 35% (from	-	-	CG says that *"The system let me know to help the PWD when a social worker was triggering his agitation."* (post-deployment surveys)

	averaged 7.18->5.29)			
2	minimal changes in the number of agitations and agitation level	CG Needs suggestions received a higher rating	1. CG has much better confidence in obtaining respite (18% increase in RSSE score) 2. PWD get slightly better sleeping quality (23.81% better PSQI)	(high rating intervention) Have a list of friends, family, neighbors, church family, etc... for you to call if you need someone to relieve you.
3	Averaged agitation level increase by 21%	CG Needs suggestions received a high rating (always rated 10/10)	1. Symptoms gradually became more severe (13.89% increase in NPI-Q symptom severity) 2. More agitation behaviors occurred (13.79% increase in CMAI-C behavior score) 3. CG Distress level lower despite the symptoms becoming more severe (5.83% decrease in NPI-Q CG distress)	The CG confirms that the symptoms are becoming more severe over the deployment period. (post-deployment interview)
4	minimal changes in the number of agitations and agitation level	Inter-personal communication suggestions received high rating and have been used a lot	CG distress due to PWD decrease a lot (51.67% decrease in NPI-Q CG distress)	High rating interventions: - (Environment) Turn off the television/radio before you start talking to minimize background noise and distraction. - (Inter-Personal Comm) Allow for plenty of time to answer any questions. CG says that distraction works for the PWD. *"I tried getting him to think of something else, to get his mind off of something he was agitated about. You don't want to call him out, it makes it worse. "*
5	Much fewer agitations (from 4.22 to 1.12/day), decreased by 73.56%	Inter-personal communication suggestions received high rating and have been used a lot	1. Agitation behavior increases slightly (13.79% increase in CMAI-C behavior score) 2. CG distress decreased (23.33% decrease in NPI-Q CG distress)	(high rating intervention) Avoid arguing or speaking loudly as persons with dementia may misinterpret and become more agitated.

6.4 Conclusion

In this chapter, a personalized intervention selection and suggestion techniques are presented. The results show that the system that suggests agitation intervention strategies based on deployment data and clinical scores is relatively easy to use and not bothersome. The helpfulness of the system was demonstrated by the caregivers' answers post-deployments such as *“The system let me know to help the PWD when a social worker was triggering his agitation”* and *"I tried getting him to think of something else, to get his mind off of something he was agitated about. You don’t want to call him out, it makes it worse. "* The clinical tool score comparison also shows that some CG has less caregiving distress and better ability to obtaining respite. Though this work is a pilot study with five participants (each participated in the study for 2-months), the result shows that the intervention system has potential in automated symptom-related suggestions delivery. The personalized intervention selection technique involves user engagement and clinical expertise in the form of self-reports and questionnaires, assessment tool scores, and sensing/intervention systems. This creates scalable and automated prevention and mitigation intervention suggestions for real-world applications.

7. CLOSING REMARKS

Despite the availability of mobile sensing technology or *smart health*, the technology still has not been fully utilized for remote monitoring purposes. This dissertation is aimed to encourage the adoption of smart health technology to real-world applications. Here, I have discussed the implementation of smart health technology in two healthcare applications, the dementia-caregiving empowerment and the management of in-home cancer-pain. In both case-studies, the sensing system called BESI system is deployed at houses of person-with-chronic-disease living with their caregiver. The system continuously monitors for signs of health symptom such as dementia-related agitation or cancer breakthrough pain episodes. The system utilized machine learning models to learn the health symptom stimuli using the collected ambient environmental data and event labels which are provided by either the patients or the caregivers. By using predictive machine learning models to predict health events based on sensor data, the health event monitoring can be done anytime and anywhere compared to traditional method where only one health-related data point is generated when patients and doctors meet. Thus, this creates a much more scalable health monitoring system.

To fully utilized predictive machine learning approach with the monitored sensor data, real-world data challenges need to be addresses. To avoid privacy invasiveness of the usage of video camera for symptom ground truth, a self-reporting approach has been used to collect observations regarding health symptom such as the occurrences of dementia-related agitation and cancer-pain episodes which the caregiver or the patient help with marking the event labels. This self-reporting approach can be solved by using data segmentation techniques with learning models that read temporal information. Real-world challenges such as imbalanced observations and small training labels are also addressed. The result in this dissertation shows that machine learning models can be utilized to predict in-the-filed health events. Furthermore, the result also shows that each patient, both dementia-patients and cancer-patients, reacts differently to different environmental stimuli. This shows the need for personalized learning approach which the work in this dissertation addresses.

The interpretation of machine learning models is also important for user to make decision or take action based on the model's outcome. In chapter 5, an actionable model interpretation for real-time intervention suggestion is presented. This approach differs from other machine learning interpretations; Most model interpretation techniques aim to understand how and why the learning model behave in such a way or to gain trusts in the learning models. The presented actionable interpretation focuses on the extraction of actionable items from the data instant and the model's prediction. If a model predicts an unwanted outcome, the presented interpretation will provide the user with specific input behavior that causes the model to output certain prediction. This actionable interpretation has been used in the dementia case study. For example, the system could notify the in-home dementia caregiver to intervene before agitation escalates by turning the lights on because the ambient light level is decreasing (e.g. the sun is setting), which has triggered agitation episodes in the past. This enhances the adoption of predictive machine learning models in real-world. Now, not only the model predictions are provided to users, but the potential cause of action as well.

Personalized health symptom intervention suggestions are also studied in this dissertation. Results of machine learning predictions in chapter 4 and dementia-related agitation triggers in chapter 5 show that patients react differently to different ambient stimuli. Thus, it is advantageous to personalized symptom intervention to specific patients as well. In chapter 6, a personalized algorithm used to calculate disease severity and produce appropriate symptom intervention is presented. Normally, intervention suggestions are done when patients or caregivers meet the doctors; Or, in the telehealth case, clinicians still need to actively communicate with patients and caregivers through phone-calls or video-calls which is not scalable. Thus, the work here aims to create personalized symptom intervention suggestions in a more automate and scalable ways by pairing symptom severity form clinical assessment tools to the appropriate intervention suggestion. This method has been implemented to send dementia-related agitation intervention suggestions to caregivers in real-time when agitation is detected. This result shows potential benefit to dementia caregivers.

7.1 Publications

During my time as a Ph.D. student at the University of Virginia, I have had the opportunity to work with and conduct researches with experts from many fields of study. Below are the lists of my publications during the Ph.D. study:

[1] R. Alam, J. Dugan, **N. Homdee**, N. Gandhi, B. Ghaemmaghami, H. Meda, A. Bankole, M. Anderson, J. Gong, T. Smith-Jackson, and J. Lach. 2017. BESI: Reliable and Heterogeneous Sensing and Intervention for In-home Health Applications. In *2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 147–156.

[2] **Nutta Homdee**, Ridwan Alam, Sean Wolfe, and John Lach. 2018. Mind the Gaps: Edge-based Monitoring for Continuous Room-level Location tracking. *Smart Health* 9–10, (December 2018), 297–306.

[3] Ridwan Alam, **Nutta Homdee**, Sean Wolfe, James Hayes, and John Lach. 2019. Besi: behavior learning and tracking with wearable and in-home sensors - a dementia case-study: poster abstract. In *Proceedings of the International Conference on Internet of Things Design and Implementation* (IoTDI '19), Association for Computing Machinery, New York, NY, USA, 281–282.

[4] V. Mandalapu, **N. Homdee**, J. M. Hart, J. Lach, S. Bodkin, and J. Gong. 2019. Developing Computational Models for Personalized ACL Injury Classification. In *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, 1–4.

[5] D. Kim, V. Mandalapu, J. M. Hart, S. Bodkin, **N. Homdee**, J. Lach, and J. Gong. 2019. Examining Cross-Validation Strategies for Predictive Modeling of Anterior Cruciate Ligament Reinjury. In *2019 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, 27–28.

[6] J.M. Hart, V. Mandalapu, S. Bodkin, J. Lach, **N. Homdee**, and J. Gong. 2019. Wearable Sensor-based Classification of ACL Reconstructed Limbs During Exercise in Male and Female Patients. *Medicine & Science in Sports & Exercise* 51, 6S (June 2019), 460.

[7] **N. Homdee**, R. Alam, J. A. Hayes, T. Hamid, J. Park, S. Wolfe, H. Goins, N. Fyffe, T. Newbold, T. Smith-Jackson, A. Bankole, M. S. Anderson, and J. Lach. 2019. Agitation Monitoring and Prevention System for Dementia Caregiver Empowerment. *Computer* 52, 11 (November 2019), 30–39.

[8] **Nutta Homdee**, Mehdi Boukhechba, Yixue W. Feng, Natalie Kramer, John Lach, and Laura E. Barnes. 2019. Enabling Smartphone-based Estimation of Heart Rate. *arXiv:1912.08910v1 [eess]* (December 2019).

[9] Virginia LeBaron, James Hayes, Kate Gordon, Ridwan Alam, **Nutta Homdee**, Yudel Martinez, Emmanuel Ogunjirin, Tanya Thomas, Randy Jones, Leslie Blackhall, and John Lach. 2019. Leveraging Smart Health Technology to Empower Patients and Family Caregivers in Managing Cancer Pain: Protocol for a Feasibility Study. *JMIR Research Protocols* 8, 12 (December 2019), e16178.

[10] **Nutta Homdee**, Hilda Goins, Azziza Bankole, Martha S. Anderson, and John Lach. 2019. Prediction of Dementia-related Agitation Using Multivariate Ambient Environmental Time-series Data. *arXiv:2009.05097 [eess]* (2019).

[11] Azziza Bankole, Martha S. Anderson, **Nutta Homdee**, Ridwan Alam, Ashley Lofton, Nykesha Fyffe, Hilda Goins, Temple Newbold, Tonya Smith-Jackson, and John Lach. 2020. BESI: Behavioral and Environmental Sensing and Intervention for Dementia Caregiver Empowerment—Phases 1 and 2. *American Journal of Alzheimer's Disease & Other Dementias* 35, (January 2020), 1533317520906686.

[12] Martha S. Anderson, **Nutta Homdee**, Azziza Bankole, Ridwan Alam, Brook Mitchell, James Hayes, Grace Byfield, and John Lach. 2020. Behavioral Interventions for Alzheimer’s Management Using Technology: Home-Based Monitoring. *Curr Geri Rep* 9, 2 (June 2020), 90–100.

[13] Virginia LeBaron, Rachel Bennett, Ridwan Alam, Leslie Blackhall, Kate Gordon, James Hayes, **Nutta Homdee**, Randy Jones, Yudel Martinez, Emmanuel Ogunjirin, Tanya Thomas, and John Lach. 2020. Understanding the Experience of Cancer Pain from the Perspective of Patients and Family Caregivers to Inform Design of an In-Home Smart Health System: Multimethod Approach. *JMIR Formative Research* 4, 8 (August 2020), e20836.

[14] **Nutta Homdee**, Martha S. Anderson, Azziza Bankole, Hilda Goins, and John Lach. 2020. Actionable Interpretation of Machine Learning Models for Sequential Data: Dementia-related Agitation Use Case. *ACM Trans. Comput. Healthcare* (Under Review).

[15] Martha S. Anderson, Azziza Bankole, **Nutta Homdee**, Brook Mitchell, Grace Byfield, and John Lach. 2021. Dementia Caregiver Experiences and Recommendations using the BESI System in the Home Setting: A Pilot Study with Descriptive Qualitative Design. *JMIR Aging* (Under Review).

REFERENCES

[1] S. Ahn, M. L. Smith, M. Altpeter, L. Post, and M. G. Ory, “Healthcare Cost Savings Estimator Tool for Chronic Disease Self-Management Program: A New Tool for Program Administrators and Decision Makers,” *Front. Public Health*, vol. 3, 2015, doi: 10.3389/fpubh.2015.00042.

[2] I. Björholt, S. Andersson, G. H. Nilsson, and I. Krakau, “The cost of monitoring warfarin in patients with chronic atrial fibrillation in primary care in Sweden,” *BMC Fam Pract*, vol. 8, no. 1, p. 6, Feb. 2007, doi: 10.1186/1471-2296-8-6.

[3] E. J. Emanuel, “Where Are the Health Care Cost Savings?,” *JAMA*, vol. 307, no. 1, pp. 39–40, Jan. 2012, doi: 10.1001/jama.2011.1927.

[4] J. Andreu-Perez, D. R. Leff, H. M. D. Ip, and G.-Z. Yang, “From Wearable Sensors to Smart Implants-–Toward Pervasive and Personalized Healthcare,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 12, pp. 2750–2762, Dec. 2015, doi: 10.1109/TBME.2015.2422751.

[5] P. Sundaravadivel, E. Kougianos, S. P. Mohanty, and M. K. Ganapathiraju, “Everything You Wanted to Know about Smart Health Care: Evaluating the Different Technologies and Components of the Internet of Things for Better Health,” *IEEE Consumer Electronics Magazine*, vol. 7, no. 1, pp. 18–28, Jan. 2018, doi: 10.1109/MCE.2017.2755378.

[6] M. M. Baig and H. Gholamhosseini, “Smart Health Monitoring Systems: An Overview of Design and Modeling,” *J Med Syst*, vol. 37, no. 2, p. 9898, Jan. 2013, doi: 10.1007/s10916-012-9898-z.

[7] K. Ullah, M. A. Shah, and S. Zhang, “Effective ways to use Internet of Things in the field of medical and smart health care,” in *2016 International Conference on Intelligent Systems Engineering (ICISE)*, Jan. 2016, pp. 372–379, doi: 10.1109/INTELSE.2016.7475151.

[8] S. Warren, R. L. Craft, and J. T. Bosma, “Designing Smart Health Care Technology into the Home of the Future,” Sandia National Labs., Albuquerque, NM (US); Sandia National Labs., Livermore, CA (US), SAND99-0863C, Apr. 1999. Accessed: Mar. 22, 2021. [Online]. Available: https://www.osti.gov/biblio/5696.

[9] M. I. Pramanik, R. Y. K. Lau, H. Demirkan, and Md. A. K. Azad, “Smart health: Big data enabled health paradigm within smart cities,” *Expert Systems with Applications*, vol. 87, pp. 370–383, Nov. 2017, doi: 10.1016/j.eswa.2017.06.027.

[10] N. V. Chawla and D. A. Davis, “Bringing Big Data to Personalized Healthcare: A Patient-Centered Framework,” *J GEN INTERN MED*, vol. 28, no. S3, pp. 660–665, Sep. 2013, doi: 10.1007/s11606-013-2455-8.

[11] A. Holzinger, C. Röcker, and M. Ziefle, “From Smart Health to Smart Hospitals,” in *Smart Health: Open Problems and Future Challenges*, A. Holzinger, C. Röcker, and M. Ziefle, Eds. Cham: Springer International Publishing, 2015, pp. 1–20.

[12] A. Aliamiri and Y. Shen, “Deep learning based atrial fibrillation detection using wearable photoplethysmography sensor,” in *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, Mar. 2018, pp. 442–445, doi: 10.1109/BHI.2018.8333463.

[13] V. Mandalapu, N. Homdee, J. M. Hart, J. Lach, S. Bodkin, and J. Gong, “Developing Computational Models for Personalized ACL Injury Classification,” in *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, May 2019, pp. 1–4, doi: 10.1109/BSN.2019.8771078.

[14] M. Z. Uddin, W. Khaksar, and J. Torresen, “Ambient Sensors for Elderly Care and Independent Living: A Survey,” *Sensors*, vol. 18, no. 7, p. 2027, Jul. 2018, doi: 10.3390/s18072027.

[15] N. Armstrong, G. Herbert, E.-L. Aveling, M. Dixon-Woods, and G. Martin, “Optimizing patient involvement in quality improvement,” *Health Expectations*, vol. 16, no. 3, pp. e36–e47, 2013, doi: 10.1111/hex.12039.

[16] T. G. Stavropoulos, G. Meditskos, E. Kontopoulos, and I. Kompatsiaris, “Multi-sensing Monitoring and Knowledge-driven Analysis for Dementia Assessment,” *IJEHMC*, vol. 6, no. 4, pp. 77–92, Oct. 2015, doi: 10.4018/IJEHMC.2015100106.

[17] S. Madakam, “Internet of Things: Smart Things,” *International Journal of Future Computer and Communication*, vol. 4, pp. 250–253, May 2015, doi: 10.7763/IJFCC.2015.V4.395.

[18] M. N. Kamel Boulos and N. M. Al-Shorbaji, “On the Internet of Things, smart cities and the WHO Healthy Cities,” *International Journal of Health Geographics*, vol. 13, no. 1, p. 10, Mar. 2014, doi: 10.1186/1476-072X-13-10.

[19] G. Chong, L. Zhihao, and Y. Yifeng, “The research and implement of smart home system based on Internet of Things,” in *2011 International Conference on Electronics, Communications and Control (ICECC)*, Sep. 2011, pp. 2944–2947, doi: 10.1109/ICECC.2011.6066672.

[20] B. L. Risteska Stojkoska and K. V. Trivodaliev, “A review of Internet of Things for smart home: Challenges and solutions,” *Journal of Cleaner Production*, vol. 140, pp. 1454–1464, Jan. 2017, doi: 10.1016/j.jclepro.2016.10.006.

[21] A. van Heerden, M. Tomlinson, and L. Swartz, “Point of care in your pocket: a research agenda for the field of m-health,” *Bull World Health Organ*, vol. 90, pp. 393–394, May 2012, doi: 10.2471/BLT.11.099788.

[22] S. Alter, “Making Sense of Smartness in the Context of Smart Devices and Smart Systems,” *Inf Syst Front*, vol. 22, no. 2, pp. 381–393, Apr. 2020, doi: 10.1007/s10796-019-09919-9.

[23] WHO Global Observatory for eHealth and World Health Organization, *MHealth: new horizons for health through mobile technologies.* Geneva: World Health Organization, 2011.

[24] A. Lazar, C. Koehler, T. J. Tanenbaum, and D. H. Nguyen, “Why we use and abandon smart devices,” in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, New York, NY, USA, Sep. 2015, pp. 635–646, doi: 10.1145/2750858.2804288.

[25] J. Cohen-mansfield, M. S. Marx, and A. S. Rosenthal, “A Description of Agitation in a Nursing Home,” *J Gerontol*, vol. 44, no. 3, pp. M77–M84, May 1989, doi: 10.1093/geronj/44.3.M77.

[26] V. LeBaron *et al.*, “Leveraging Smart Health Technology to Empower Patients and Family Caregivers in Managing Cancer Pain: Protocol for a Feasibility Study,” *JMIR Res Protoc*, vol. 8, no. 12, p. e16178, Dec. 2019, doi: 10.2196/16178.

[27] J. Subramanian and R. Simon, “Overfitting in prediction models – Is it a problem only in high dimensions?,” *Contemporary Clinical Trials*, vol. 36, no. 2, pp. 636–641, Nov. 2013, doi: 10.1016/j.cct.2013.06.011.
[28] M. Moghimi, S. Belongie, M. Saberian, J. Yang, N. Vasconcelos, and L.-J. Li, “Boosted Convolutional Neural Networks,” in *Procedings of the British Machine Vision Conference 2016*, York, UK, 2016, p. 24.1-24.13, doi: 10.5244/C.30.24.
[29] W. Hardyman, K. L. Daunt, and M. Kitchener, “Value Co-Creation through Patient Engagement in Health Care: A micro-level approach and research agenda,” *Public Management Review*, vol. 17, no. 1, pp. 90–107, Jan. 2015, doi: 10.1080/14719037.2014.881539.
[30] A. H. Alavi, P. Jiao, W. G. Buttlar, and N. Lajnef, “Internet of Things-enabled smart cities: State-of-the-art and future trends,” *Measurement*, vol. 129, pp. 589–606, Dec. 2018, doi: 10.1016/j.measurement.2018.07.067.
[31] J. Lee, B. Bagheri, and H.-A. Kao, “A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems,” *Manufacturing Letters*, vol. 3, pp. 18–23, Jan. 2015, doi: 10.1016/j.mfglet.2014.12.001.
[32] N. Homdee, R. Alam, S. Wolfe, and J. Lach, “Mind the gaps: Edge-based monitoring for continuous room-level location tracking,” *Smart Health*, Jul. 2018, doi: 10.1016/j.smhl.2018.07.020.
[33] M. Xue and C. Zhu, “The Socket Programming and Software Design for Communication Based on Client/Server,” in *2009 Pacific-Asia Conference on Circuits, Communications and Systems*, May 2009, pp. 775–777, doi: 10.1109/PACCS.2009.89.
[34] “Experience and management of cancer pain at home : Cancer Nursing.” https://journals.lww.com/cancernursingonline/abstract/1988/04000/experience_and_management_of_cancer_pain_at_home.2.aspx (accessed Mar. 30, 2021).
[35] J. A. Smyth, M. Dempster, I. Warwick, P. Wilkinson, and N. K. McCorry, “A Systematic Review of the Patient- and Carer-Related Factors Affecting the Experience of Pain for Advanced Cancer Patients Cared for at Home,” *Journal of Pain and Symptom Management*, vol. 55, no. 2, pp. 496–507, Feb. 2018, doi: 10.1016/j.jpainsymman.2017.08.012.
[36] E. Hernandez Silva, S. Lawler, and D. Langbecker, “The effectiveness of mHealth for self-management in improving pain, psychological distress, fatigue, and sleep in cancer survivors: a systematic review,” *J Cancer Surviv*, vol. 13, no. 1, pp. 97–107, Feb. 2019, doi: 10.1007/s11764-018-0730-8.
[37] “Journal of Medical Internet Research - Understanding Monitoring Technologies for Adults With Pain: Systematic Literature Review.” https://www.jmir.org/2017/10/e364 (accessed Mar. 30, 2021).
[38] “Pearson Correlation Coefficient | SpringerLink.” https://link.springer.com/chapter/10.1007%2F978-3-642-00296-0_5 (accessed Mar. 30, 2021).
[39] N. Khachiyants, D. Trinkle, S. J. Son, and K. Y. Kim, “Sundown Syndrome in Persons with Dementia: An Update,” *Psychiatry Investig*, vol. 8, no. 4, pp. 275–287, Dec. 2011, doi: 10.4306/pi.2011.8.4.275.

[40] "Data clustering: 50 years beyond K-means - ScienceDirect." https://www.sciencedirect.com/science/article/pii/S0167865509002323 (accessed Aug. 28, 2018).
[41] J. H. Friedman, "Recent Advances in Predictive (Machine) Learning," *Journal of Classification*, vol. 23, no. 2, pp. 175–197, Sep. 2006, doi: 10.1007/s00357-006-0012-4.
[42] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," 2016. https://www.deeplearningbook.org/ (accessed Jul. 29, 2020).
[43] T. M. Mitchell, *Machine Learning*. New York: McGraw-Hill, 1997.
[44] *Machine Learning and its Applications*. CRC Press, 2019.
[45] A. Nielsen, *Practical Time Series Analysis: Prediction with Statistics and Machine Learning*. O'Reilly Media, Inc., 2019.
[46] B. Doucoure, K. Agbossou, and A. Cardenas, "Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data," *Renewable Energy*, vol. 92, pp. 202–211, Jul. 2016, doi: 10.1016/j.renene.2016.02.003.
[47] A. Graves, N. Jaitly, and A. Mohamed, "Hybrid speech recognition with Deep Bidirectional LSTM," in *2013 IEEE Workshop on Automatic Speech Recognition and Understanding*, Dec. 2013, pp. 273–278, doi: 10.1109/ASRU.2013.6707742.
[48] B. O'Connor, R. Balasubramanyan, B. Routledge, and N. Smith, "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series," *ICWSM*, vol. 4, no. 1, Art. no. 1, May 2010, Accessed: Mar. 26, 2021. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14031.
[49] F. Zhou and F. D. la Torre, "Generalized time warping for multi-modal alignment of human motion," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2012, pp. 1282–1289, doi: 10.1109/CVPR.2012.6247812.
[50] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[51] J. Liu, T. Zhang, G. Han, and Y. Gou, "TD-LSTM: Temporal Dependence-Based LSTM Networks for Marine Temperature Prediction," *Sensors*, vol. 18, no. 11, Art. no. 11, Nov. 2018, doi: 10.3390/s18113797.
[52] F. A. Gers, D. Eck, and J. Schmidhuber, "Applying LSTM to Time Series Predictable Through Time-Window Approaches," in *Neural Nets WIRN Vietri-01*, London, 2002, pp. 193–200, doi: 10.1007/978-1-4471-0219-9_20.
[53] L. Yang, K. Ting, and M. B. Srivastava, "Inferring occupancy from opportunistically available sensor data," in *2014 IEEE International Conference on Pervasive Computing and Communications (PerCom)*, Mar. 2014, pp. 60–68, doi: 10.1109/PerCom.2014.6813945.
[54] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, Jul. 2012, doi: 10.1109/TSMCC.2011.2161285.
[55] M. Praveena and V. Jaiganesh, "A Literature Review on Supervised Machine Learning Algorithms and Boosting Process," *IJCA*, vol. 169, no. 8, pp. 32–35, Jul. 2017, doi: 10.5120/ijca2017914816.

[56] J. Tanha, Y. Abdi, N. Samadi, N. Razzaghi, and M. Asadpour, “Boosting methods for multi-class imbalanced data classification: an experimental review,” *J Big Data*, vol. 7, no. 1, p. 70, Sep. 2020, doi: 10.1186/s40537-020-00349-y.
[57] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. CRC Press, 2012.
[58] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *2017 International Conference on Engineering and Technology (ICET)*, Aug. 2017, pp. 1–6, doi: 10.1109/ICEngTechnol.2017.8308186.
[59] N. Tajbakhsh *et al.*, “Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, May 2016, doi: 10.1109/TMI.2016.2535302.
[60] G. Meditskos, P.-M. Plans, T. G. Stavropoulos, J. Benois-Pineau, V. Buso, and I. Kompatsiaris, “Multi-modal activity recognition from egocentric vision, semantic enrichment and lifelogging applications for the care of dementia,” *Journal of Visual Communication and Image Representation*, vol. 51, pp. 169–190, Feb. 2018, doi: 10.1016/j.jvcir.2018.01.009.
[61] S. Abdoli, P. Cardinal, and A. Lameiras Koerich, “End-to-end environmental sound classification using a 1D convolutional neural network,” *Expert Systems with Applications*, vol. 136, pp. 252–263, Dec. 2019, doi: 10.1016/j.eswa.2019.06.040.
[62] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, “1D convolutional neural networks and applications: A survey,” *Mechanical Systems and Signal Processing*, vol. 151, p. 107398, Apr. 2021, doi: 10.1016/j.ymssp.2020.107398.
[63] T. Fu, “A review on time series data mining,” *Engineering Applications of Artificial Intelligence*, vol. 24, no. 1, pp. 164–181, Feb. 2011, doi: 10.1016/j.engappai.2010.09.007.
[64] J. J. Hopfield, “Artificial neural networks,” *IEEE Circuits and Devices Magazine*, vol. 4, no. 5, pp. 3–10, Sep. 1988, doi: 10.1109/101.8118.
[65] P. Ramachandran, B. Zoph, and Q. V. Le, “Searching for Activation Functions,” *arXiv:1710.05941 [cs]*, Oct. 2017, Accessed: Mar. 26, 2021. [Online]. Available: http://arxiv.org/abs/1710.05941.
[66] G. Van Houdt, C. Mosquera, and G. Nápoles, “A review on the long short-term memory model,” *Artif Intell Rev*, vol. 53, no. 8, pp. 5929–5955, Dec. 2020, doi: 10.1007/s10462-020-09838-1.
[67] C. B. Şahín and B. Dírí, “Robust Feature Selection With LSTM Recurrent Neural Networks for Artificial Immune Recognition System,” *IEEE Access*, vol. 7, pp. 24165–24178, 2019, doi: 10.1109/ACCESS.2019.2900118.
[68] S. V. Vaseghi, *Advanced Digital Signal Processing and Noise Reduction*. John Wiley & Sons, 2008.
[69] *Convolutional neural networks for time series forecasting - Python for Finance Cookbook*. 2020.
[70] H. Phan, L. Hertel, M. Maass, and A. Mertins, “Robust Audio Event Recognition with 1-Max Pooling Convolutional Neural Networks,” *arXiv:1604.06338 [cs]*, Jun. 2016, Accessed: Mar. 26, 2021. [Online]. Available: http://arxiv.org/abs/1604.06338.
[71] “Understand different types of Boosting Algorithms.” https://iq.opengenus.org/types-of-boosting-algorithms/ (accessed Mar. 26, 2021).

[72] Y. Liu, C. Liu, and D. Wang, “A 1D time-varying median filter for seismic random, spike-like noise elimination,” *GEOPHYSICS*, vol. 74, no. 1, pp. V17–V24, Dec. 2008, doi: 10.1190/1.3043446.

[73] A. L. Barrick *et al.*, “Impact of ambient bright light on agitation in dementia,” *International Journal of Geriatric Psychiatry*, vol. 25, no. 10, pp. 1013–1021, 2010, doi: 10.1002/gps.2453.

[74] B. Behera, G. Kumaravelan, and P. K. Borugadda, *Performance Evaluation of Deep Learning Algorithms in Biomedical Document Classification*. 2019, p. 224.

[75] B. K.-S. Cheung, A. Langevin, and H. Delmaire, “Coupling genetic algorithm with a grid search method to solve mixed integer nonlinear programming problems,” *Computers & Mathematics with Applications*, vol. 34, no. 12, pp. 13–23, Dec. 1997, doi: 10.1016/S0898-1221(97)00229-0.

[76] M. W. Browne, “Cross-Validation Methods,” *Journal of Mathematical Psychology*, vol. 44, no. 1, pp. 108–132, Mar. 2000, doi: 10.1006/jmps.1999.1279.

[77] P. Branco, L. Torgo, and R. Ribeiro, “A Survey of Predictive Modelling under Imbalanced Distributions,” *arXiv:1505.01658 [cs]*, May 2015, Accessed: Mar. 17, 2021. [Online]. Available: http://arxiv.org/abs/1505.01658.

[78] O. Sharma, “Deep Challenges Associated with Deep Learning,” in *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, Feb. 2019, pp. 72–75, doi: 10.1109/COMITCon.2019.8862453.

[79] Z. Qiu, J. Chen, Y. Zhao, S. Zhu, Y. He, and C. Zhang, “Variety Identification of Single Rice Seed Using Hyperspectral Imaging Combined with Convolutional Neural Network,” *Applied Sciences*, vol. 8, no. 2, Art. no. 2, Feb. 2018, doi: 10.3390/app8020212.

[80] D. A. Gredell *et al.*, “Comparison of Machine Learning Algorithms for Predictive Modeling of Beef Attributes Using Rapid Evaporative Ionization Mass Spectrometry (REIMS) Data,” *Scientific Reports*, vol. 9, no. 1, Art. no. 1, Apr. 2019, doi: 10.1038/s41598-019-40927-6.

[81] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. Ch. Chatzisavvas, “A comparison of machine learning techniques for customer churn prediction,” *Simulation Modelling Practice and Theory*, vol. 55, pp. 1–9, Jun. 2015, doi: 10.1016/j.simpat.2015.03.003.

[82] N. Williams, S. Zander, and G. Armitage, “A preliminary performance comparison of five machine learning algorithms for practical IP traffic flow classification,” *SIGCOMM Comput. Commun. Rev.*, vol. 36, no. 5, pp. 5–16, Oct. 2006, doi: 10.1145/1163593.1163596.

[83] J. Z. Li, “Principled approaches to robust machine learning and beyond,” Thesis, Massachusetts Institute of Technology, 2018.

[84] Q. Dong, S. Gong, and X. Zhu, “Imbalanced Deep Learning by Minority Class Incremental Rectification,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 6, pp. 1367–1381, Jun. 2019, doi: 10.1109/TPAMI.2018.2832629.

[85] H. He and E. A. Garcia, “Learning from Imbalanced Data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, Sep. 2009, doi: 10.1109/TKDE.2008.239.

[86] J. Byrd and Z. C. Lipton, "What is the Effect of Importance Weighting in Deep Learning?," p. 10.
[87] R. Hodson, "Precision medicine," *Nature*, vol. 537, no. 7619, Art. no. 7619, Sep. 2016, doi: 10.1038/537S49a.
[88] G. S. Ginsburg and K. A. Phillips, "Precision Medicine: From Science To Value," *Health Affairs*, vol. 37, no. 5, pp. 694–701, May 2018, doi: 10.1377/hlthaff.2017.1624.
[89] S. Yeom, I. Giacomelli, M. Fredrikson, and S. Jha, "Privacy Risk in Machine Learning: Analyzing the Connection to Overfitting," in *2018 IEEE 31st Computer Security Foundations Symposium (CSF)*, Jul. 2018, pp. 268–282, doi: 10.1109/CSF.2018.00027.
[90] A. Alwosheel, S. van Cranenburgh, and C. G. Chorus, "Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis," *Journal of Choice Modelling*, vol. 28, pp. 167–182, Sep. 2018, doi: 10.1016/j.jocm.2018.07.002.
[91] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.
[92] "Explaining AdaBoost | SpringerLink." https://link.springer.com/chapter/10.1007/978-3-642-41136-6_5 (accessed Mar. 26, 2021).
[93] M. E. Kunik *et al.*, "Consequences of Aggressive Behavior in Patients With Dementia," *JNP*, vol. 22, no. 1, pp. 40–47, Jan. 2010, doi: 10.1176/jnp.2010.22.1.40.
[94] K. Williams *et al.*, "A randomized trial using telehealth technology to link caregivers with dementia care experts for in-home caregiving support: FamTechCare protocol," *Research in Nursing & Health*, vol. 41, no. 3, pp. 219–227, 2018, doi: 10.1002/nur.21869.
[95] S. Arora, R. Ge, B. Neyshabur, and Y. Zhang, "Stronger generalization bounds for deep nets via a compression approach," *arXiv:1802.05296 [cs]*, Nov. 2018, Accessed: Mar. 19, 2021. [Online]. Available: http://arxiv.org/abs/1802.05296.
[96] G. K. Dziugaite and D. M. Roy, "Computing Nonvacuous Generalization Bounds for Deep (Stochastic) Neural Networks with Many More Parameters than Training Data," *arXiv:1703.11008 [cs]*, Oct. 2017, Accessed: Mar. 19, 2021. [Online]. Available: http://arxiv.org/abs/1703.11008.
[97] L. L. Joosse, "Do Sound Levels and Space Contribute to Agitation in Nursing Home Residents with Dementia?," *Research in Gerontological Nursing*, vol. 5, no. 3, pp. 174–184, Jul. 2012, doi: 10.3928/19404921-20120605-02.
[98] F.-C. Hsieh *et al.*, "Effect of home-based music intervention versus ambient music on breast cancer survivors in the community: A feasibility study in Taiwan," *European Journal of Cancer Care*, vol. 28, no. 4, p. e13064, 2019, doi: https://doi.org/10.1111/ecc.13064.
[99] A. Keenan and J. K. Keithley, "Integrative Review: Effects of Music on Cancer Pain in Adults," *Oncology Nursing Forum*, vol. 42, no. 6, pp. E368–E375, Nov. 2015, doi: 10.1188/15.ONF.E368-E375.

[100] N. Homdee, "Prediction of Dementia-related Agitation Using Multivariate Ambient Environmental Time-series Data," *arXiv:2002.07237 [eess]*, Feb. 2020, Accessed: Mar. 31, 2021. [Online]. Available: http://arxiv.org/abs/2002.07237.
[101] J. Alizargar, M. Etemadi Sh, M. Aghamohammadi, and S. Hatefi, "Saliva samples as an alternative for novel coronavirus (COVID-19) diagnosis," *J Formos Med Assoc*, vol. 119, no. 7, pp. 1234–1235, Jul. 2020, doi: 10.1016/j.jfma.2020.04.030.
[102] E. Pasomsub *et al.*, "Saliva sample as a non-invasive specimen for the diagnosis of coronavirus disease 2019: a cross-sectional study," *Clinical Microbiology and Infection*, vol. 27, no. 2, p. 285.e1-285.e4, Feb. 2021, doi: 10.1016/j.cmi.2020.05.001.
[103] D. Sapkota, S. B. Thapa, B. Hasséus, and J. L. Jensen, "Saliva testing for COVID-19?," *British Dental Journal*, vol. 228, no. 9, Art. no. 9, May 2020, doi: 10.1038/s41415-020-1594-7.
[104] 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 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, Inc., 2015, pp. 91–99.
[105] S. P. Shashikumar, Q. Li, G. D. Clifford, and S. Nemati, "Multiscale network representation of physiological time series for early prediction of sepsis," *Physiol. Meas.*, vol. 38, no. 12, pp. 2235–2248, Nov. 2017, doi: 10.1088/1361-6579/aa9772.
[106] M. Antkowiak, "Artificial Neural Networks vs. Support Vector Machines for Skin Diseases Recognition," Mar. 2021.
[107] M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255–260, Jul. 2015, doi: 10.1126/science.aaa8415.
[108] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, Art. no. 5, May 2019, doi: 10.1038/s42256-019-0048-x.
[109] K. R. Varshney and H. Alemzadeh, "On the Safety of Machine Learning: Cyber-Physical Systems, Decision Sciences, and Data Products," *Big Data*, vol. 5, no. 3, pp. 246–255, Sep. 2017, doi: 10.1089/big.2016.0051.
[110] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artificial Intelligence*, vol. 267, pp. 1–38, Feb. 2019, doi: 10.1016/j.artint.2018.07.007.
[111] W. Samek, T. Wiegand, and K.-R. Müller, "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models," Aug. 2017, Accessed: Mar. 25, 2020. [Online]. Available: https://arxiv.org/abs/1708.08296v1.
[112] Z. C. Lipton, "The mythos of model interpretability," *Commun. ACM*, vol. 61, no. 10, pp. 36–43, Sep. 2018, doi: 10.1145/3233231.
[113] F. K. Došilović, M. Brčić, and N. Hlupić, "Explainable artificial intelligence: A survey," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, May 2018, pp. 0210–0215, doi: 10.23919/MIPRO.2018.8400040.
[114] V. Schetinin *et al.*, "Confident Interpretation of Bayesian Decision Tree Ensembles for Clinical Applications," *IEEE Transactions on Information Technology in*

Biomedicine, vol. 11, no. 3, pp. 312–319, May 2007, doi: 10.1109/TITB.2006.880553.

[115] D. Martens, B. Baesens, T. Van Gestel, and J. Vanthienen, “Comprehensible credit scoring models using rule extraction from support vector machines,” *European Journal of Operational Research*, vol. 183, no. 3, pp. 1466–1476, Dec. 2007, doi: 10.1016/j.ejor.2006.04.051.

[116] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart, “RETAIN: An Interpretable Predictive Model for Healthcare using Reverse Time Attention Mechanism,” in *Advances in Neural Information Processing Systems 29*, D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, Eds. Curran Associates, Inc., 2016, pp. 3504–3512.

[117] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization,” 2017, pp. 618–626, Accessed: Oct. 15, 2020. [Online]. Available: https://openaccess.thecvf.com/content_iccv_2017/html/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.html.

[118] J. Krause, A. Perer, and K. Ng, “Interacting with Predictions: Visual Inspection of Black-box Machine Learning Models,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, San Jose California USA, May 2016, pp. 5686–5697, doi: 10.1145/2858036.2858529.

[119] A. Colubri, T. Silver, T. Fradet, K. Retzepi, B. Fry, and P. Sabeti, “Transforming Clinical Data into Actionable Prognosis Models: Machine-Learning Framework and Field-Deployable App to Predict Outcome of Ebola Patients,” *PLOS Neglected Tropical Diseases*, vol. 10, no. 3, p. e0004549, Mar. 2016, doi: 10.1371/journal.pntd.0004549.

[120] M. T. Ribeiro, S. Singh, and C. Guestrin, “‘Why Should I Trust You?’: Explaining the Predictions of Any Classifier,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, Aug. 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[121] M. Sundararajan, A. Taly, and Q. Yan, “Axiomatic Attribution for Deep Networks,” *arXiv:1703.01365 [cs]*, Jun. 2017, Accessed: Oct. 15, 2020. [Online]. Available: http://arxiv.org/abs/1703.01365.

[122] B. Norgeot, D. Lituiev, B. S. Glicksberg, and A. J. Butte, “Time Aggregation and Model Interpretation for Deep Multivariate Longitudinal Patient Outcome Forecasting Systems in Chronic Ambulatory Care,” *arXiv:1811.12589 [cs, stat]*, Nov. 2018, Accessed: Mar. 12, 2020. [Online]. Available: http://arxiv.org/abs/1811.12589.

[123] “Statistical forecasting: notes on regression and time series analysis.” https://people.duke.edu/~rnau/411home.htm (accessed Mar. 10, 2020).

[124] A. Altmann, L. Toloşi, O. Sander, and T. Lengauer, “Permutation importance: a corrected feature importance measure,” *Bioinformatics*, vol. 26, no. 10, pp. 1340–1347, May 2010, doi: 10.1093/bioinformatics/btq134.

[125] R. W. Schafer and L. R. Rabiner, “Digital representations of speech signals,” *Proceedings of the IEEE*, vol. 63, no. 4, pp. 662–677, Apr. 1975, doi: 10.1109/PROC.1975.9799.

[126] S. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” *arXiv:1705.07874 [cs, stat]*, Nov. 2017, Accessed: Mar. 22, 2021. [Online]. Available: http://arxiv.org/abs/1705.07874.

[127] N. Homdee and J. Lach, “Actionable Interpretation of Machine Learning Models for Sequential Data: Dementia-related Agitation Use Case,” *arXiv:2009.05097 [cs, eess]*, Sep. 2020, Accessed: Mar. 22, 2021. [Online]. Available: http://arxiv.org/abs/2009.05097.

[128] D. Slack, S. Hilgard, E. Jia, S. Singh, and H. Lakkaraju, “Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods,” in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, New York, NY, USA, Feb. 2020, pp. 180–186, doi: 10.1145/3375627.3375830.

[129] D. Alvarez-Melis and T. S. Jaakkola, “On the Robustness of Interpretability Methods,” *arXiv:1806.08049 [cs, stat]*, Jun. 2018, Accessed: Mar. 22, 2021. [Online]. Available: http://arxiv.org/abs/1806.08049.

[130] S. M. Das, “Causality in Machine Learning 101 for Dummies like Me,” *Medium*, Nov. 29, 2019. https://towardsdatascience.com/causality-in-machine-learning-101-for-dummies-like-me-f7f161e7383e (accessed Mar. 25, 2020).

[131] S. M. R. Islam, D. Kwak, MD. H. Kabir, M. Hossain, and K.-S. Kwak, “The Internet of Things for Health Care: A Comprehensive Survey,” *IEEE Access*, vol. 3, pp. 678–708, 2015, doi: 10.1109/ACCESS.2015.2437951.

[132] A. Bankole *et al.*, “BESI: Behavioral and Environmental Sensing and Intervention for Dementia Caregiver Empowerment—Phases 1 and 2:,” *American Journal of Alzheimer’s Disease & Other Dementias®*, Mar. 2020, doi: 10.1177/1533317520906686.

[133] W. W. Eaton, C. Smith, M. Ybarra, C. Muntaner, and A. Tien, “Center for Epidemiologic Studies Depression Scale: Review and Revision (CESD and CESD-R),” in *The use of psychological testing for treatment planning and outcomes assessment: Instruments for adults, Volume 3, 3rd ed*, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2004, pp. 363–377.

[134] J. C. Morris, “The Clinical Dementia Rating (CDR): Current version and scoring rules,” *Neurology*, vol. 43, no. 11, pp. 2412–2412, Nov. 1993, doi: 10.1212/WNL.43.11.2412-a.

[135] G. S. Alexopoulos, R. C. Abrams, R. C. Young, and C. A. Shamoian, “Cornell scale for depression in dementia,” *Biological Psychiatry*, vol. 23, no. 3, pp. 271–284, Feb. 1988, doi: 10.1016/0006-3223(88)90038-8.

[136] D. J. Buysse, C. F. Reynolds, T. H. Monk, S. R. Berman, and D. J. Kupfer, “The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research,” *Psychiatry Research*, vol. 28, no. 2, pp. 193–213, May 1989, doi: 10.1016/0165-1781(89)90047-4.

[137] D. I. Kaufer *et al.*, “Validation of the NPI-Q, a Brief Clinical Form of the Neuropsychiatric Inventory,” *JNP*, vol. 12, no. 2, pp. 233–239, May 2000, doi: 10.1176/jnp.12.2.233.

[138] A. M. Steffen, C. McKibbin, A. M. Zeiss, D. Gallagher-Thompson, and A. Bandura, “The Revised Scale for Caregiving Self-Efficacy: Reliability and Validity Studies,” *The Journals of Gerontology: Series B*, vol. 57, no. 1, pp. P74–P86, Jan. 2002, doi: 10.1093/geronb/57.1.P74.

[139] M. Bédard, D. W. Molloy, L. Squire, S. Dubois, J. A. Lever, and M. O'Donnell, "The Zarit Burden Interview: A New Short Version and Screening Version," *The Gerontologist*, vol. 41, no. 5, pp. 652–657, Oct. 2001, doi: 10.1093/geront/41.5.652.
[140] J. F. Broyles and University of Arkansas for Medical Sciences, *Coach Broyles' playbook for Alzheimer's caregivers: a practical tips guide*. Fayetteville, AR: University of Arkansas] : Distributed in Arkansas by Barbara Broyles Legacy, 2006.
[141] "Alzheimer's Association," *Alzheimer's Disease and Dementia*. https://alz.org/homepage-modal (accessed Jul. 29, 2020).