UNIVERSITY OF CALIFORNIA
Los Angeles

Efficient Representation Learning for Longitudinal Data in Healthcare Applications

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Computer Science

by

Shayan Fazeli

2023

ABSTRACT OF THE DISSERTATION

Efficient Representation Learning for Longitudinal Data in Healthcare Applications

by

Shayan Fazeli
Doctor of Philosophy in Computer Science
University of California, Los Angeles, 2023
Professor Majid Sarrafzadeh, Chair

Efficient utilization of longitudinal observations is a crucial component in proposing machine learning solutions to problems in healthcare. The temporal nature of numerous problems in this domain, such as understanding fluctuations in physiological signals through time pertinent to health status, renders this avenue of research particularly important for the intersection of Health Analytics and Artificial Intelligence (AI). In the healthcare domain, compared to other fields such as Computer Vision or Natural Language Processing, the data is often available in limited quantities. Additionally, reliable supervision signals for training inference pipelines are scarce. Furthermore, some data modalities and domains are critical to health applications which are, at the same time, considerably less investigated in machine learning research. These challenges are essential bottlenecks to address in improving the efficacy and usability of machine learning-based healthcare solutions.

In this dissertation, we investigate the role of longitudinal data in medical and health applications in various related domains. Namely, we consider the domains of 1) Physical Health: Representation learning for monitoring the physical health of an individual useful for in-patient and out-patient setups, with examples being physiological signals, activity

data, and posture tracking. 2) Electronic Health Records: The multi-modal and temporal reports in different time resolutions on patients' health trajectories 3) Mental Health: Efficient multi-resolution monitoring of stress and anxiety as an example use-case with important applications, and 3) Public Health: Pandemic analytics and representation of population-level spatio-temporal health data. We suggest novel techniques to address the primary challenges in each task efficiently. In our solutions, we use approaches such as optimizing self-supervised contrastive objectives, knowledge transfer, and adversarial training so as to minimize the reliance on accurate and large-scale supervision signals. We discuss the empirical validation of our suggested solutions and shed light on some of the key future research directions.

The dissertation of Shayan Fazeli is approved.

Yizhou Sun

Wei Wang

Baharan Mirzasoleimanbarzi

Majid Sarrafzadeh, Committee Chair

University of California, Los Angeles

2023

To my cherished circle of love . . .

Much like the constellations that light up the desert's night sky,

your unwavering love and support have been my guiding stars throughout this journey.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I'd like to express my heartfelt gratitude to my advisor, Prof. Majid Sarrafzadeh. His unwavering support and wise counsel were instrumental in navigating the complexities of graduate school. Equally, I appreciate Prof. Baharan Mirzasoleiman, Prof. Wei Wang, and Prof. Yizhou Sun from my committee for their invaluable feedback and constructive insights.

Lastly, a profound thank you goes out to my family and friends. Whether close in the valleys of Los Angeles or miles away in my home country, their enduring support was my anchor throughout this journey.

VITA

2012–2017 B.Sc. of Electrical Engineering, Sharif University of Technology

2012–2017 B.Sc. of Computer Science, Sharif University of Technology

2017–2020 M.Sc. of Computer Science, University of California, Los Angeles

PUBLICATIONS

- Fazeli S, Levine L, Beikzadeh M, Mirzasoleiman B, Zadeh B, Peris T, Sarrafzadeh M. A Self-supervised Framework for Improved Data-Driven Monitoring of Stress via Multi-modal Passive Sensing. arXiv preprint arXiv:2303.14267. 2023 Mar 24.
- Fazeli S, Levine L, Beikzadeh M, Mirzasoleiman B, Zadeh B, Peris T, Sarrafzadeh M. Passive monitoring of physiological precursors of stress leveraging smartwatch data. In2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2022 Dec 6 (pp. 2893-2899). IEEE.
- Fazeli S, Moatamed B, Sarrafzadeh M. Statistical analytics and regional representation learning for covid-19 pandemic understanding. In2021 IEEE 9th International Conference on Healthcare Informatics (ICHI) 2021 Aug 9 (pp. 248-257). IEEE.
- Fazeli S, Zamanzadeh D, Ovalle A, Nguyen T, Gee G, Sarrafzadeh M. COVID-19 and Big Data: Multi-faceted Analysis for Spatio-temporal Understanding of the Pandemic with Social Media Conversations. arXiv preprint arXiv:2104.10807. 2021 Apr 22.

- Ford CL, Amani B, Harawa NT, Akee R, Gee GC, Sarrafzadeh M, Abotsi-Kowu C, Fazeli S, Le C, Nwankwo E, Zamanzadeh D. Adequacy of existing surveillance systems to monitor racism, social stigma and COVID inequities: a detailed assessment and recommendations. International journal of environmental research and public health. 2021 Dec 12;18(24):13099.

- Darabi S, Kachuee M, Fazeli S, Sarrafzadeh M. Taper: Time-aware patient ehr representation. IEEE journal of biomedical and health informatics. 2020 Apr 3;24(11):3268-75.

- Levine LM, Gwak M, Kärkkäinen K, Fazeli S, Zadeh B, Peris T, Young AS, Sarrafzadeh M. Anxiety detection leveraging mobile passive sensing. InBody Area Networks. Smart IoT and Big Data for Intelligent Health: 15th EAI International Conference, BODYNETS 2020, Tallinn, Estonia, October 21, 2020, Proceedings 15 2020 (pp. 212-225). Springer International Publishing.

- Gwak M, Fazeli S, Ershadi G, Sarrafzadeh M, Ghodsi M, Aminian A, Schlechter JA. Extra: exercise tracking and analysis platform for remote-monitoring of knee rehabilitation. In2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN) 2019 May 19 (pp. 1-4). IEEE.

- Hosseini A, Fazeli S, van Vliet E, Valencia L, Habre R, Sarrafzadeh M, Bui A. Children activity recognition: Challenges and strategies. In2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2018 Jul 18 (pp. 4331-4334). IEEE.

- Kachuee M, Fazeli S, Sarrafzadeh M. Ecg heartbeat classification: A deep transferable representation. In2018 IEEE international conference on healthcare informatics (ICHI) 2018 Jun 4 (pp. 443-444). IEEE.

CHAPTER 1

Introduction

The computing and digital design world has come a long way since the early 20th century. The earliest digital computers were introduced in the 1940s and were capable of running thousands of operations. In contrast, these days, almost everyone around the globe owns a smartphone and thus, on a day-to-day basis, uses a portable computer capable of running trillions of operations. The role of digital computing in healthcare is no exception to this rapid progress.

In the late 20th century, and the rapid progress in technology and digital devices, the healthcare system and hospitals started a transition in patient record keeping, going from paper-based written records to digitally kept information. The increased access to effortless and efficient communication with the introduction of the Internet was another driving factor in this matter. The quick and efficient data transmission opportunities further highlighted the importance of moving towards more digital healthcare systems. This was a considerable change in healthcare, and while it introduced novel challenges related to patient privacy, it far surpassed the old approach in efficiency and reliability. Electronic Health Records (EHRs), defined as "the systematized collection of patient and population electronically-stored health information in a digital format," helped by providing a much easier way to keep track of the comprehensive longitudinal health information for patients and the services they received at the hospitals. Enhanced access to information and reduced medical errors due to more efficient and less error-prone representation of data were but a few of what involving EHRs and technological advancements in health offered.

Parallel to the use of technology in in-patient settings, in the early 2000s, the use of mobile phones and other personal digital devices started to become prevalent. With digital systems becoming more advanced and following the reduction of prices and increased affordability of access to them, researchers started to pay more attention to the potential that the interconnection of digital computing and health monitoring offered for out-patient healthcare. Early forms of connection, such as Short Message Service (SMS), were a popular choice in many of the proposed solutions relying on transmitting health-related notifications to individuals.

Beyond these movements, public health experts have also relied on computer-assisted modeling and spatiotemporal data acquisition in dealing with health crises and disease management. The unfortunate occurrence of the COVID-19 pandemic in 2020 is a clear example in this domain, showcasing the importance of data-driven monitoring and surveillance in understanding the spread of the virus.

In summary, one can categorize the impacts of technological advancements on health as follows:

- **Hardware**: As the computational power and energy efficiency of digital systems are improved, it becomes easier to integrate them into healthcare solutions. In the context of mHealth, for example, nowadays, it is possible to produce small wearable components to monitor patients or rely on commercially available smartphones, smartwatches, and other devices such as smart rings. Additionally, the enhanced and customized hardware allows for running more computationally intensive algorithms and artificial intelligence pipelines that are an inseparable component of many modern solutions.

- **Communication**: Transmitting information plays a crucial role in digital health. In the in-patient setup and electronic health records, it enables crucial functionalities such as easy access to the detailed medical history of patients for medical experts from anywhere. In the out-patient setup, it helps cut the costs of physically attending a health

center or a hospital via solutions such as telehealth and remote health monitoring.

- **Software**: At the heart of digital health, software on the back-end, front-end, and service layers help drive the system forward. The improvements in the back-end systems, such as efficient databases, real-time record keeping, and file storage, allow the solutions to leverage the Internet and rely more on cloud services in record-keeping. The front-end systems play an essential role in engaging users and help create a user-friendly experience that incentivizes further usage of the systems. The services and functionalities that run in the background, either on the cloud services or on the device itself, rely on proper algorithm design and model selection to prepare the result of analyses that the system needs.

- **Intelligent Systems**: Artificial Intelligence (AI) has seen exponential progress in the past decade, and machine learning algorithms are becoming more common in healthcare systems as time passes. Such advancements have been a game-changer in the eHealth domain, and the research on improving the performance, explainability, and reliability of AI-powered models is ongoing.

AI-powered models have been achieving impressive performance across a variety of tasks, with state-of-the-art models in the fields of Computer Vision and Natural Language Processing pushing the limits of machine performance further and further away every year. Nonetheless, the intersection of AI with Health Analytics is a niche field constrained by the challenges specific to health-related data and tasks. The advancements in technology allows efficient capturing and storage of high-quality health data, both in in-patient and out-patient settings and population level. Nonetheless, compared to images and text, the modalities involved in health are considerably less investigated. Additionally, going beyond the analyses of static information and understanding health data from the lense of *temporal* changes is crucial to many health-related tasks, and active research tries to shed more light on how this data can be of good use.

This dissertation focuses on our work in the utilization and understanding of temporal and longitudinal health data for modernized healthcare solutions. We demonstrate this as a critical problem in digital health with applications across healthcare problems, namely, out-patient remote health monitoring and health trajectories, in-patient record keeping and inference-making based on Electronic Health Records (EHRs), and finally, in public-health and population-level health analyses.

The rest of this manuscript is structured as follows: Chapter 2 categorizes the research subfields pertinent to this dissertation and reviews the related work and proposed solutions in each domain. Chapters 3 and 4 discuss our works in the domain of personal physical health. Chapter 3 focuses on transferrable representation learning for physiological signals and introduces an approach for learning universal electrocardiogram embeddings via transfer learning. Chapter 4 discusses the applications of tracking bodily activity in ehealth, defining problems and proposing automated solutions in three critical subproblems in this domain. Moving on to in-patient records, Chapter 5 proposes a longitudinal prediction pipeline leveraging multi-modal electronic health records as its input. Chapter 6 discusses Pandemic Analytics and presents an approach for inference-making in public health crises leveraging spatiotemporal data. Chapter 7 focuses on remote monitoring of the indicators of mental well-being and presents a solution for the recognition and tracking of stress as an example of a short-term episodic mental health problem affecting the general public. It leverages advanced techniques such as adversarial regularization and self-supervised pre-training to optimize performance. Finally, Chapter 8 concludes this dissertation.

CHAPTER 2

Related Work

This dissertation discusses our proposed solutions to efficiently model data from various data modalities and targetting optimized performance on numerous domains of longitudinal health tasks. In what follows, we provide the related literature prior to each of the presented solutions in their corresponding categories.

2.1 Personal Health

2.1.1 Physical Health: Tracking and Recognition

2.1.1.1 Representation Learning for Electrocardiogram

Electrocardiogram (ECG or EKG) is one of the key vital signals in domain of health analytics, comprising an important source of information across various applications. Reviewing ECG readings by cardiologists and medical professionals is essential in patient diagnosis. To address the problems raised with the manual analysis of ECG signals, many studies in the literature explored using machine learning techniques to accurately detect the anomalies in the signal [EKS17, DKS17]. Most of these approaches involve a preprocessing phase for preparing the signal (e.g., passing it through band-pass filters). Afterward, the handcrafted features, primarily statistical summarizations of signal windows, are extracted from these signals and used in further analysis for the final classification task. As for the inference engine, conventional machine learning approaches for ECG analysis include Support Vector Machines, multi-layer perceptrons, decision trees, and so on [IGK06, SSC10, KKM17].

These handcrafted features provide us with an acceptable representation of the signal. Based on recent machine learning studies, automated feature extraction and representation methods are proven to be more scalable and have the potential to make accurate predictions. An end-to-end deep learning framework allows the machine to learn the features best suited to the specific task it is dedicated to carrying out [AOH17b, KIG16, JD17]. This approach provides us with a more accurate representation of the ECG signal, using which the machine can compete with a human cardiologist in analyzing the signal [RHH17]. Deep learning approaches, however, contain a tremendously large amount of variables that often require massive amounts of data to be trained. This renders it critical to make the most of the available training data to enhance the semantic interpretability of the learned representations.

2.1.2 Mental Health

Stress and anxiety-related disorders are common mental health challenges. Such disorders can have significant negative impacts on people's lives, including higher chances of depression and suicide as well as associated comorbidities with physical health issues [Cli]. Unfortunately, these issues remain inadequately treated in many cases due to challenges ranging from lack of viable access to therapeutic services to associated stigmas with utilization [Men, APA]. However, even when an individual decides to seek psychotherapeutic help to alleviate these problems, challenges persist in the diagnosis and effective treatment of their disorder. At the inception of care, the steps to diagnose and monitor often include clinical evaluation and comparing personalized symptoms to standardized criteria, for example, the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), which is commonly used for this matter. Researchers continue to study and improve the practicality and accuracy of guidelines such as DSM-5[BSG12, ZMC17]. Still, there are challenges in converting aggregated and generalized diagnostic criteria down to episodic-level incidents of stress and anxiety.

The most obvious approach to doing so leverages biometric data extracted from wearable

sensors embedded in smart devices that measure a physiological stress response. However, while such data is incredibly valuable, and notably, sensing devices have become increasingly sophisticated at monitoring physiological stress; the resulting analyses are incomplete at best. This owes to the fact that from the standpoint of straightforward correlative analytics, it is known that there is not a direct monotonic correlation between the emotional perception of stress an individual may feel and the manifestation of the underlying physiological stress response. For instance, a meta-analysis in the social stress domains has recently shown that merely 25% of studies in the field demonstrated a significant correlation between physiological stress and perceived emotional stress [CE12]. Given that self-reports of perceived stress often do not contain information on the physiological stress response, understanding the complex relationship between the two becomes a crucial matter[CE12]. It is also plausible to assume that such complexity also arises from various other confounding factors (e.g., demographics, occupation, and other mental health disorders such as attention-deficit hyperactivity disorder (ADHD) can influence how prone someone is to stress). This discrepancy has meaningful impacts on the utility of passive stress detection based largely on physiological indicators. While sensors may be returning accurate readings on physiological stress, if they do not align with the user's own perceptions of stress, notably if they fail to properly account for moments when a user feels acute emotional distress, then it will demotivate further engagement with a mental health platform. This hindrance comes in spite of considerable progress that has been made in recent decades regarding the capabilities and efficacy of personal digital devices, including smartwatches, smartphones, and wearable devices. This fact has made such devices attract much research and commercial attention, employing them for various monitoring objectives [A 18, Tex17]. These monitoring approaches focus primarily on fitness and health-related aspects, resulting in extensive research and countless commercialized applications. Examples include tracking athletes' training, detecting falls for older adults, tracking post-surgery therapeutic and rehabilitation exercises, and posture correction [DMH17, VFB15, GFE19, WRK14].

While the central focus of health monitoring applications has undoubtedly been on physical health, a wide range of research works has focused on understanding the relationship between observations obtained leveraging digital devices and some aspects of individuals' mental health status. Notably, a primary goal in designing smart and automated mental health monitoring approaches involves proposing meaningful passive-sensing tools so that informative observations regarding health status can be made by eliminating or diminishing the need to interfere with users' daily activities or request repeated active interactions. Considering the prevalence of smartphones, there have been efforts exploring whether pervasive mental health monitoring could be feasible through a smartphone and the embedded sensors, such as motion sensors, ambient light, microphone, camera, Global Positioning System (GPS), proximity, and touch screen[CFL17][OPC14][BSW15][NGT20]. These efforts showed the promise of this approach in successfully tying behavioral monitoring to mental health; however, such approaches have not translated into fully mature frameworks for monitoring of anxiety, and have focused almost exclusively on depression-related conditions, which while often spoken in conjunction with anxiety, manifest in distinct ways[FIO19]. As a remote mental health monitoring task, social anxiety was studied in the previous literature, and it was shown that analyzing trajectories obtained via smartphone location services can paint a comprehensive picture concerning individuals' proneness to it [BCF18, CFH17, HXL16, GHC19]. The advantages of leveraging a smartphone-based platform are that the continuous collection of quantitative data potentially provides a more reliable indicator of an individual's risk at any given time, as well as offering a mechanism for just-in-time intervention should a mental health episode occur[BSW15]. Conversely, smartphone-derived data present several challenges, some of which have already been noted, which can result in limited accuracy owing to differences in behavioral patterns across users, and the indirect manner of detection[FIO19]. Another choice of hardware for gathering data pertinent to health data is application-specific wearable sensors. For instance, wearable electrocardiogram (ECG) sensors were used to recognize perceived anxiety via pattern recognition [KMW18].

Smartwatches have a unique position amongst the wide range of various commonly used digital devices. They are in close contact with the skin and, given their attachment user's wrist, which is a distal point of a major appendage, make it possible to obtain most measurements (e.g., activity) at higher accuracy, as well as enabling additional measurements such as heart-rate or pulse oximeter. In case of the need for brief questions, interactions, or Ecological Momentary Assessments (EMAs), smartwatches can also be used to issue messages and acquire responses and entries by the user [A 18, Tex17, J 17]. Additionally, smartwatches are prevalent, and relying on them as the hardware for health applications provides a better alternative in most cases to application-specific wearable devices in terms of cost, comfort, and user-friendliness. Data-driven analyses leveraging smartwatches' sensory readings have been successful at the problem of patient classification for bipolar disorder, schizoaffective disorder, and depression [HZY22].

In the health analytics domain, data and human annotations are often limited. Therefore, dealing with overfitting and memorization is a crucial matter. Additionally, it is beneficial to go beyond the limited number of human annotations available in training efficient inference pipelines. Less reliance on annotations by focusing on unsupervised and self-supervised approaches has received a lot of research attention in recent years [CKN20, CMM20, CH21, GSA20]. The core idea in most works in this area is that comparing and contrasting the latent representations of examples that are expected to share certain similarities (e.g., augmented versions of the same image) can benefit the trained weights and help with regularizing the learned decision boundaries [ZCD17].

2.2 Body Activity and Movement Analyses

2.2.1 Activity Recognition

A wealth of research has been dedicated to using accelerometer and gyroscope data for activity recognition [CPR11, QMX10]. A large amount of this work has gone into finding the

optimal classifier for activity detection. Classic models, such as random forest, have shown some success, achieving between 90% and 94% accuracy in identifying activities such as walking, sitting, and taking the stairs [CPR11]. Similarly, simple models based on Support Vector Machines (SVM) have achieved reasonable performance on this task [QMX10]. Recently, deep learning models, namely, convolutional neural networks (CNN) and recurrent neural networks (RNN) have been applied to activity recognition [YWC08]. However, both of these studies have used multiple inertial sensors that are placed in different parts of the body, which is not feasible on a daily basis, especially when it comes to conducting activity recognition for children. Moreover, none of the above-mentioned studies have focused on activity recognition for children.

2.2.2 Rehabilitation Exercises

Part of this dissertation discusses our work on knee-angle recognition to track the performance of post-surgery rehabilitation exercises. On this front, there have been numerous works regarding monitoring knee movements and the effect of different physical activities on its health. In [JG93], authors tried to determine the effect of different amounts of effusion on knee intraarticular pressure (IAP) during knee movement. In [SMA87], authors used Electromyographic analysis to compare the effects of knee exercises in healthy subjects and subjects with knee pathologies. Gyroscope, flex sensors reading, and knee movement have been the main sources of information for measuring knee joint angle in the previous research [MIH13, LKB98, PKK01].

2.2.3 Posture Tracking

Remote monitoring of the human body, postures, and activities is a crucial area of mobile health research. Many researchers and companies have worked on developing utilities and analytical solutions to monitor, understand, track, and leverage information related to body

posture. These solutions assist subjects in improving physical activities, ranging from general posture to particular therapeutic exercises [Dev, GFE19].

Numerous research works have focused on monitoring sitting as an important daily activity. These works were mostly focused on analyzing the sitting posture to better fathom its transitions and variations throughout the process of sitting [MHS16, EV16]. The system in [ZHL19] is designed for sitting behavior tracking and analysis. This work has utilized many stretchable sensors and pressure sensors to prepare a thorough sensor-driven system, with their main contribution being their sitting behavior recognition using neural networks and dynamic time warping. The system proposed in [WCY19] with the objective of helping mitigate the impacts of poor sitting posture, such as pain and discomfort in the back and neck. Their system is based on multiple motion sensors, and it is close to our solution discussed in this manuscript in terms of low cost and affordability. Their system differentiates between sitting, lying, walking, and standing postures. Their framework's main components are cardboard, a test pole, and dynamic measure units (DMU), including accelerometer, gyroscope, and magnetometer. The framework proposed in [SGH19] follows a different objective of monitoring floor sitting postures. This work proposes the use of a number of pressure sensors in a system to achieve this objective. A personalized transportable folding device is utilized in [WCC20] to assist with maintaining a better sitting posture while eating. Their results led to the empirical validation of the hypothesis that improving the sitting posture while eating can help mitigate the adverse effects of dysphagia. The use of sensor-augmented and specialized chairs is also investigated in the literature [Ami12]. In [SGG19], the authors have proposed a system based on a chair embedded with pressure sensors to differentiate between the three main sitting postures of leaning forward, reclining backward, or neutral sitting. In terms of the sitting postures that this system attempts to recognize, this work bears considerable similarity to ours. Nevertheless, there is no need for any other circuitry besides a single motion sensor in our approach. A similar grouping of sitting postures can also be found in [LBA18]. Another work in which a chair is augmented with a large number

of electrical sensors is [FDN19]. In a different approach, [SKN16] uses a pressure mat for sitting posture recognition. A sensor-augmented cushion is also used in [IS18] and [Hwa19] for sitting posture detection, which serves as an approach similar to the previously mentioned works. Feedback mechanism has also been used for posture correction. The system in [BAS17] is composed of a Kinect device for extracting body landmarks and helping the user by providing feedback, urging them to maintain proper posture. Several other Kinect-based approaches focused on finding unhealthy sitting postures [MP18, XCY19, He18, YMC17]. In [TMB19], authors combine the information obtained from Kinect with smartwatches to improve the resulting detections. As sitting posture recognition is especially important for remote health monitoring of older cohorts, a similar work focused on Kinect-based posture recognition for older adults is presented in [TLQ18].

2.3 Electronic Health Records

In the context of EHR representation learning, many works have attempted to perform this task by employing auto-encoding (AE) schemes [RHW88, Bal12]. This includes a pre-training sequence optimizing a data-reconstruction objective followed by supervised fine-tuning on downstream tasks [HS06, HOT06, VLB08, MMC11, KW13].

The utilization of unsupervised and self-supervised learning has led to significant improvements in the domain of natural language processing, specifically, the use of context windows and skip-grams, as well as learning language models for language generation [MSC13, SVL14, BCB14a]. Deep unsupervised representation learning has been successfully applied to various health settings such as EHRs (clinical concepts), medical text, and imaging data for phenotyping or classification. The general approach is to extract features using an unsupervised model and stack it with a classifier for downstream tasks. For instance, in [MLK16], a stack of denoising autoencoder multi-layer perceptron was used to predict future disease codes. They run their method on an extensive database of approximately 700,000 patients

and show improved predictive performance.

Similarly, in [CBS16, SSG17], a multi-layer perceptron using the skip-gram model was trained with an added regularization for the co-occurrence of codes within a visit to embed the clinical concepts on MIMIC III [JPS16]. As EHR data might have limited occurrences of specific codes and hence hinder the learning process of deep models, later work suggested adding an attention mechanism on external ontology's [CBS17]. More recent work on clinical concept embedding focused on leveraging temporal context using attention [CGN18a]. From the results, they suggested learning time-aware representations is critical to improving the performance for a clinical decision task.

Similar to the language modeling task, and considering the electronic health records as a language of its own ordered by the information reflecting on the temporal order of the records, one could argue that the information (e.g., codes in a visit) should be helpful to in predictive modeling of the surrounding context [CBS16]. There have been works focusing on the use of convolution and leveraging external ontology for EHR representation learning as well [NTW16, CBS17, ZKH18].

When representing medical codes, their hierarchical structure can be leveraged for efficient learning. For example, there have been works on embedding codes at the treatment level (medication and procedure codes to predict diagnosis codes, as well as going from diagnosis codes to predict next visit's codes) [CXS18]. Although such models exhibit reasonable performance and outperform earlier solutions which are often validated on proprietary datasets, they do not emphasize leveraging the temporal nature of these records. Training a CBOW model with temporal attention on the code representations has also been investigated as a helpful solution [CGN18b]. These methods primarily focus on one domain (e.g., medical code embedding) and disregard other critical modalities (e.g., doctor notes).

The medical notes in the EHR corpora often comprise large amounts of valuable information in the form of unstructured text data, which due to the challenges associated with their representation, have been studied less extensively, with works focusing on clinical concept

embedding instead. These notes provide vast amounts of information on a patient's progress trajectory and health status; nonetheless, given the use of medical terminology and the domain shift compared to the general English text, their representation could be challenging [GOM18].

The use of unstructured EHR data in supervised and semi-supervised fashion for learning patient representation has been the focus of some works [ROC18, DM18]. Traditionally, the use of CNN and RNN (e.g., in LSTM setting) has been shown to be effective in embedding text data in the EHRs [LZR18]. The introduction of Transformers has been a game changer in the natural language processing domain [DCL18a]. There are variations of Transformer-based language models focused on embedding medical text [HAR19].

Few works have considered multi-modal EHR representations, for instance, jointly representing text data and clinical codes [BCE18, MWD18]. In providing efficient representation learning pipelines for the longitudinal records present in large-scale EHR corpora, it is essential to consider the temporal nature of the data as well as try to alleviate the need for expert annotations by leveraging unsupervised and self-supervised training. Taking advantage of multiple modalities is also beneficial, as focusing on a single modality can be limiting.

2.4 Public Health

The advances in computer systems have also had a transformative impact on the Public Health domain. One of the prominent examples of this transformation is how health education and promoting healthier lifestyles have changed through time with the introduction of novel ways to communicate via the internet and social media. Advances in computing, communication, and AI have also provided researchers with more fine-grained means to model health trends and optimize the decision-making process for policy-makers. A clear example is the case of the COVID-19 pandemic, in which the task of determining the symptoms, long-term effects (long-covid), and modeling the spread given the limited information available

at the time, were of considerable interest to the researchers.

As for modeling the spread of the pandemic, the classic approach has been to define a system of ordinary differential equations (ODEs) such as SIR and its various modifications and parameterize it to predict the patterns [KM27, DH00]. For example, in the SIR system, you have the number of *susceptible*, *recovered*, and *infected* individuals, as well as effective transmission and recovery rates. With the advancements in deep learning, using neural networks trained via gradient-based optimization in pandemic analytics has become another interesting topic in public health. Given the multi-faceted nature of the problem of modeling the spread of a pandemic, it is plausible to assume that many factors reflecting on the context of the situation could potentially provide further information to improve the statistical relevance of the observations to the objective targets.

Since the beginning of the COVID-19 pandemic, there have been efforts to utilize computerized advancements in controlling and understanding this disease. An example is the applications developed to monitor the patients' locations and routes of movement. A notable work in this area is MIT's SafePaths application [saf], which contains interview and profiling capability for places and paths. It is worth mentioning that these platforms have also caused worries regarding maintaining patients' privacy [RSB20]. To provide researchers and government agencies with frequently updated monitoring information regarding the coronavirus, the 1point3acres team provided an API allowing access to the daily updated numbers of coronavirus cases [covb, YSH20]. Datasets such as [kagd] were also released to the public. A large corpus of scientific articles on coronaviruses was also released, resulting from a collaboration between AllenAI Institute, Microsoft Research, Chan-Zuckerberg Initiative, NIH, and the White House [kaga]. Projects such as a work at John Hopkins University focused on providing US county-level summaries of COVID-19 pandemic information and important attributes [KWS20, WGL20]. The information in social networks has also been used in predicting the number of COVID-19 cases in mainland China [SCL20]. The work in [PS20b] focused on an AI-based approach for predicting mortality risk in COVID-19 patients. There

have been numerous approaches to modeling the pandemic using AI that have the historical outbreak information at the core of their analyses, such as the modified versions of the SEIR model and ARIMA-based analysis [WGL20, WWG20, cova, SP20, SXP20, Kuf20]. The works in [WWG20] and [PS20a] were the main attempts in county-level modeling of the disease dynamics. In [WWG20], authors proposed a non-parametric model for epidemic data that incorporates area-level characteristics in the SIR model. The work in [PS20a] used a combination of iterated filtering and the Ensemble Adjustment Kalman filter for tuning their model, and their approach was based on a county-level SEIR model.

CHAPTER 3

Universal Representation Learning for Electrocardiogram Readings[1]

3.1 Introduction

ECG is widely used by cardiologists and medical practitioners for monitoring the cardiac health. The main problem with manual analysis of ECG signals, similar to many other time-series data, lies in difficulty of detecting and categorizing different waveforms and morphologies in the signal. For a human, this task is both extensively time-consuming and prone to errors. Note that the proper diagnosis of cardiovascular diseases is of paramount importance since these are the cause of death for about one-third of all deaths around the globe [Org]. For instance, millions of people experience irregular heartbeats which can be lethal in some cases. Therefore, accurate and low-cost diagnosis of arrhythmic heartbeats is highly desirable [Soc18].

In this part, we discuss our framework for ECG analysis that is able to represent the signal in a way that is transferable between different tasks. For this to happen, we describe a deep neural network architecture which offers a considerable capacity for learning such representations. This network has been trained on the task of arrhythmia detection for learning which it is plausible to assume that the model needs to learn most of the shape-related features of the ECG signal. Also, we have a large amount of labeled data for this task,

[1]**This chapter is based on "Kachuee M, Fazeli S, Sarrafzadeh M. Ecg heartbeat classification: A deep transferable representation. In2018 IEEE international conference on healthcare informatics (ICHI) 2018 Jun 4 (pp. 443-444). IEEE." [KFS18]**

which makes it easy to train a network with a large amount of parameters. Furthermore, we show that the signal representation learned from this task is successfully transferable to the task Myocardial infarction (MI) prediction using ECG signals. This method allows us to use these deep representations to share knowledge between ECG recognition tasks for which enough information may not be available for training a deep architecture.

3.2 Data

we use PhysioNet MIT-BIH Arrhythmia and PTB Diagnostic ECG Databases as data source for labeled ECG records [GAG00, MM01, BKS95]. Furthermore, we demonstrate that the knowledge learned from the former database can be successfully transferred for training inference models for the latter. In all of our experiments, we have used ECG lead II re-sampled to the sampling frequency of $125Hz$ as the input.

The MIT-BIH dataset consists of ECG recordings from 47 different subjects recorded at the sampling rate of $360Hz$. Each beat is annotated by at least two cardiologists. We use annotations in this dataset to create five different beat categories in accordance with Association for the Advancement of Medical Instrumentation (AAMI) EC57 standard [Med98]. See Table 3.1 for a summary of mappings between beat annotations in each category.

The PTB Diagnostics dataset consists of ECG records from 290 subjects: 148 diagnosed as MI , 52 healthy control, and the rest are diagnosed with 7 different disease. Each record contains ECG signals from 12 leads sampled at the frequency of 1000Hz. In this study we have only used ECG lead II, and worked with MI and healthy control categories in our analyses.

Table 3.1: Summary of mappings between beat annotations and AAMI EC57 [Med98] categories.

Category	Annotations
N	• Normal • Left/Right bundle branch block • Atrial escape • Nodal escape
S	• Atrial premature • Aberrant atrial premature • Nodal premature • Supra-ventricular premature
V	• Premature ventricular contraction • Ventricular escape
F	• Fusion of ventricular and normal
Q	• Paced • Fusion of paced and normal • Unclassifiable

3.3 Methods

3.3.1 Preprocessing

As ECG beats are inputs of the proposed method we suggest a simple and yet effective method for preprocessing ECG signals and extracting beats. The steps used for extracting beats from an ECG signal are as follows (see Fig. 3.1):

1. Splitting the continuous ECG signal to $10s$ windows and select a $10s$ window from an ECG signal.
2. Normalizing the amplitude values to the range of between zero and one.
3. Finding the set of all local maximums based on zero-crossings of the first derivative.
4. Finding the set of ECG R-peak candidates by applying a threshold of 0.9 on the normalized value of the local maximums.
5. Finding the median of R-R time intervals as the nominal heartbeat period of that window (T).
6. For each R-peak, selecting a signal part with the length equal to $1.2T$.
7. Padding each selected part with zeros to make its length equal to a predefined fixed length.

It is worth mentioning that the suggested beat extraction method is simple and effective in extracting R-R intervals from signals with different morphologies. For example, we have not used any form of filtering or any processing that makes any assumption about the signal morphology or spectrum. Additionally, all the extracted beats have identical lengths which is essential for being used as inputs to the subsequent processing parts.

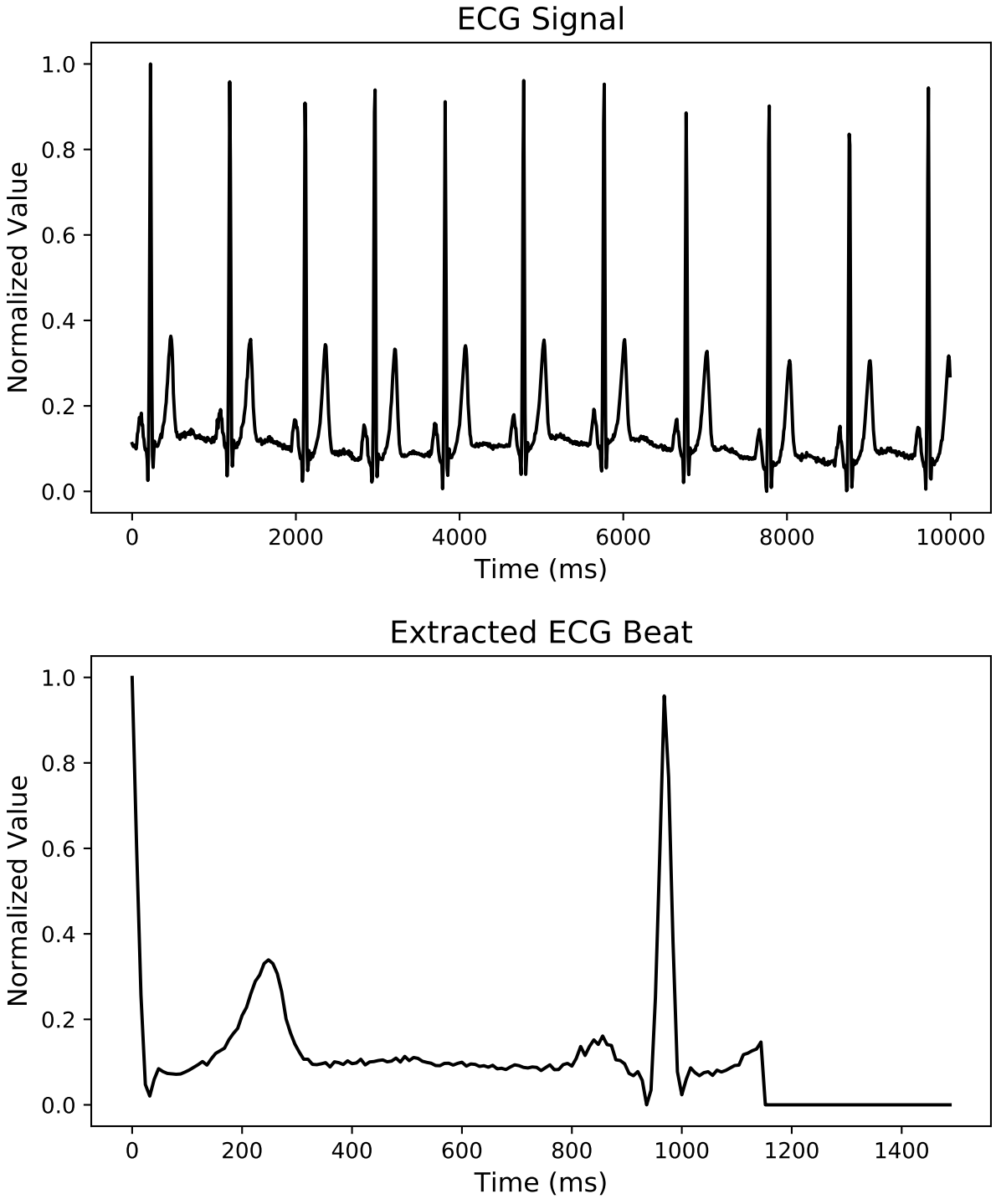


Figure 3.1: An example of a 10s ECG window and an extracted beat from it.

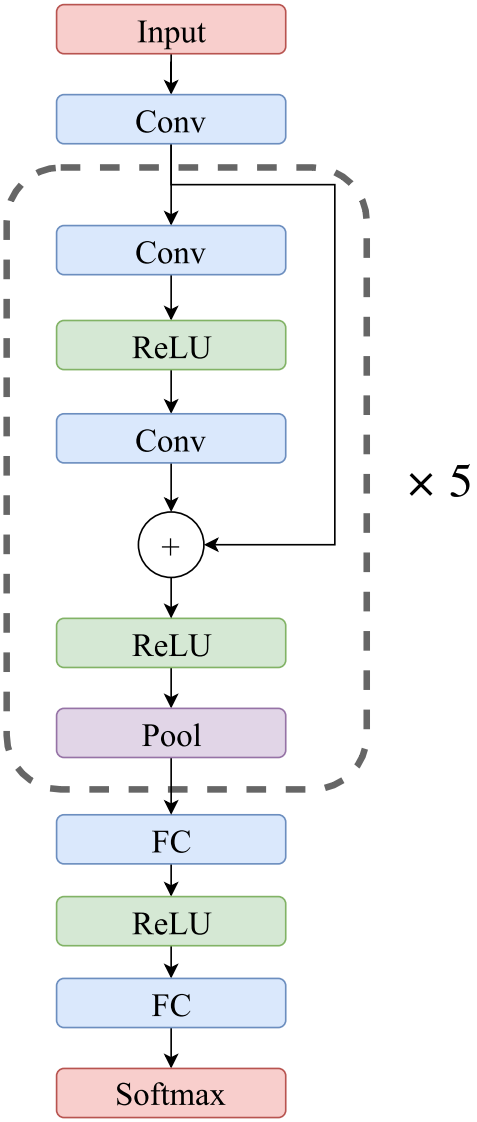


Figure 3.2: Architecture of the proposed network.

3.3.2 Training the Arrhythmia Classifier

In this paper we suggest training a convolutional neural network for classification of ECG beat types on the MIT-BIH dataset. The trained network not only can be used for the purpose of beat classification, but also in the next section we show that it can be used as an informative representation of heartbeats.

Fig. 3.2 illustrates the network architecture proposed for the beat classification task.

Extracted beats, as explained in Section 3.3.1, are used as inputs. Here, all convolution layers are applying 1-D convolution through time and each have 32 kernels of size 5. We also use max pooling of size 5 and stride 2 in all pooling layers. The predictor network consists of five residual blocks followed by two fully-connected layers with 32 neurons each and a softmax layer to predict output class probabilities. Each residual block contains two convolutional layers, two ReLU nonlinearities [NH10], a residual skip connection [HZR16], and a pooling layer. In total, the resulting network is a deep network consisting of 13 weight layers.

3.3.3 Training the MI Predictor

After training the network suggested in Section 3.3.2, we use the output activations of the very last convolution layer as a representation of input beats. Here, we use this representation as input to a two layer fully-connected network with 32 neurons at each layer to predict MI. It is noteworthy to mention that during the training for the MI prediction task, we freeze the weights for all other layers aside from the last two. In other words, we only train the last two network layers and use the learned representation of Section 3.3.2.

3.3.4 Implementation Details

In all experiments, TensorFlow computational library [AAB16] is used for model training and evaluation. Cross entropy loss on the softmax outputs is used as the loss function. For training the networks, we used Adam optimization method [KB14] with the learning rate, beta-1, and beta-2 of 0.001, 0.9, and 0.999, respectively. Learning rate is decayed exponentially with the decay factor of 0.75 every 10000 iterations. Training all the networks took less than two hours on a GeForce GTX 1080Ti processor.

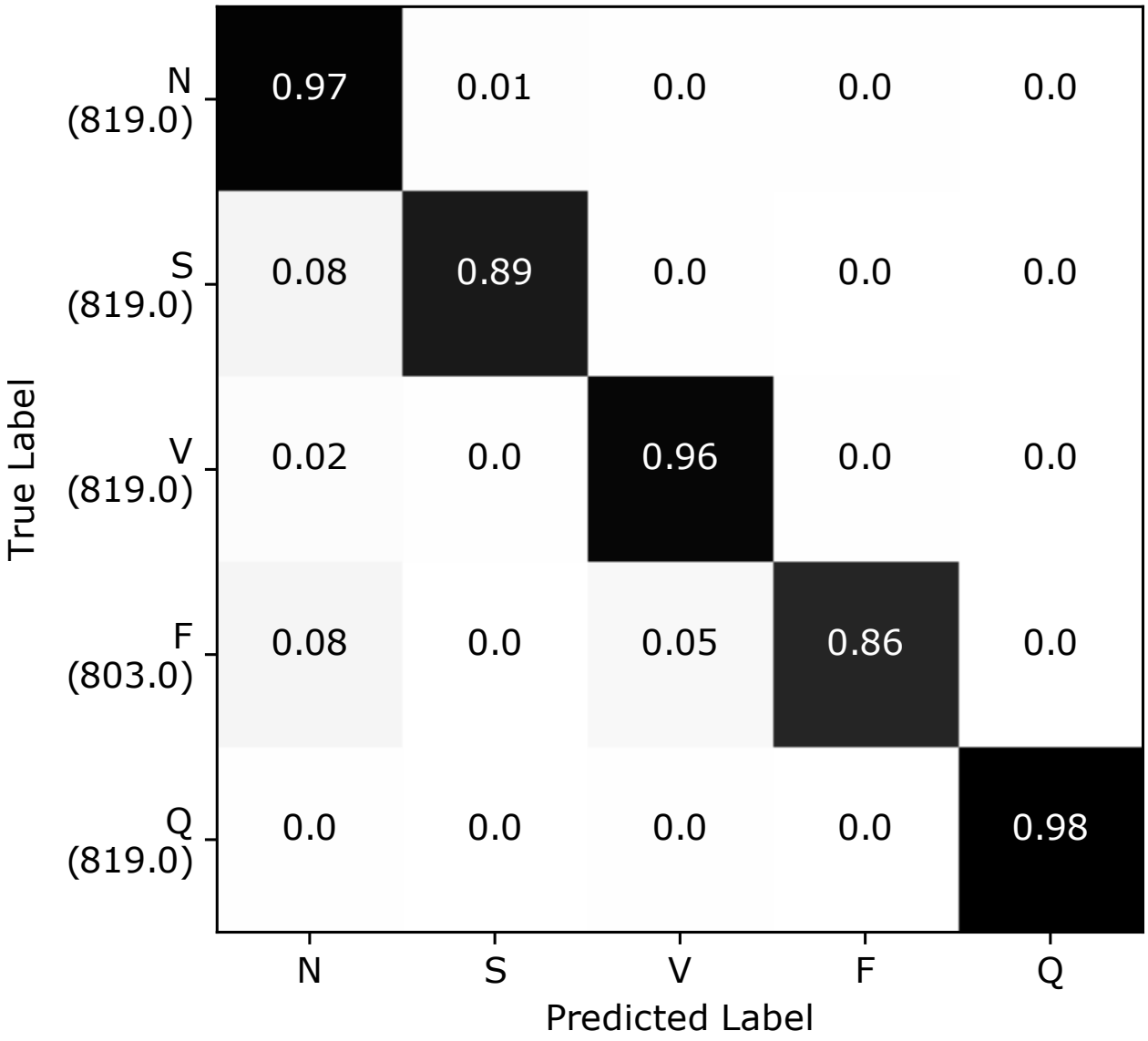


Figure 3.3: Confusion matrix for heartbeat classification on the test set. Total number of samples in each class is indicated inside parenthesis. Numbers inside blocks are number of samples classified in each category normalized by the total number of samples and rounded to two digits.

Table 3.2: Comparison of heartbeat classification results.

Work	Approach	Average Accuracy (%)
This Paper	**Deep residual CNN**	**93.4**
Acharya *et al.* [AOH17a]	Augmentation + CNN	93.5
Martis *et al.* [MAL13]	DWT + SVM	93.8
Li *et al.* [LZ16]	DWT + random forest	94.6

3.4 Experiments

3.4.1 Arrhythmia Classification and learning the representation

We evaluated the arrhythmia classifier of Section 3.3.2 on 4079 heartbeats (about 819 from each class) that are not used in the network training phase. Fig. 3.3 presents the confusion matrix of applying the classifier on the test set. As it can be seen from this figure, the model is able to make accurate predictions and distinguish different classes.

Table 3.2 presents the average accuracy of the proposed method and compares it with other relevant methods in the literature. While suggesting a predictor for MIT-BIH is not the sole purpose of this study, according to the results, the accuracies achieved in this paper are competitive to the state of the art methods. The main reason behind this might be the fact that we have used residual connections in our network architecture which allows us to train deeper networks compared to using traditional convolutional architectures.

3.4.2 MI Classification using the learned representation

We have trained our MI predictor using the learned representations, and took 80% of the PTB dataset as our training set. We have used the remaining 20% to test our model.

Table 3.3: Comparison of MI classification results.

Work	**Accuracy (%)**	**Precision (%)**	**Recall (%)**
This Paper[1]	**95.9**	**95.2**	**95.1**
Acharya *et al.* [AFO17][1]	93.5	92.8	93.7
Safdarian *et al.* [SDA14][1]	94.7	–	–
Kojuri *et al.* [KBD15][2]	95.6	97.9	93.3
Sun *et al.* [SLY12][3]	–	82.4	92.6
Liu *et al.* [LLW15][3]	94.4	–	–
Sharma *et al.* [STD15][3]	96	99	93

[1]: PTB dataset, ECG lead II

[2]: dataset collected by authors, 12-lead ECG

[3]: PTB dataset, 12-lead ECG

Table 3.3 presents a comparison between the average accuracy, precision, and recall of the proposed method for MI classification and other work in the literature. The performance of the proposed method is better than all other works except the method suggested by Sharma *et al.* [STD15] that reports higher accuracy and precision values. However, it noteworthy to mention that Sharma *et al.* use 12-lead ECG as opposed to us using only the lead II.

3.4.3 Visualization of the learned representation

In order to visualize the learned representation, we have used t-SNE visualization method [MH08] to map high-dimensional vector created by the last convolutional layer to the 2D space. In a nutshell, t-SNE creates a mapping such that the joint probability of data-points appearing close to each other in the high-dimensional space is similar to the same probability

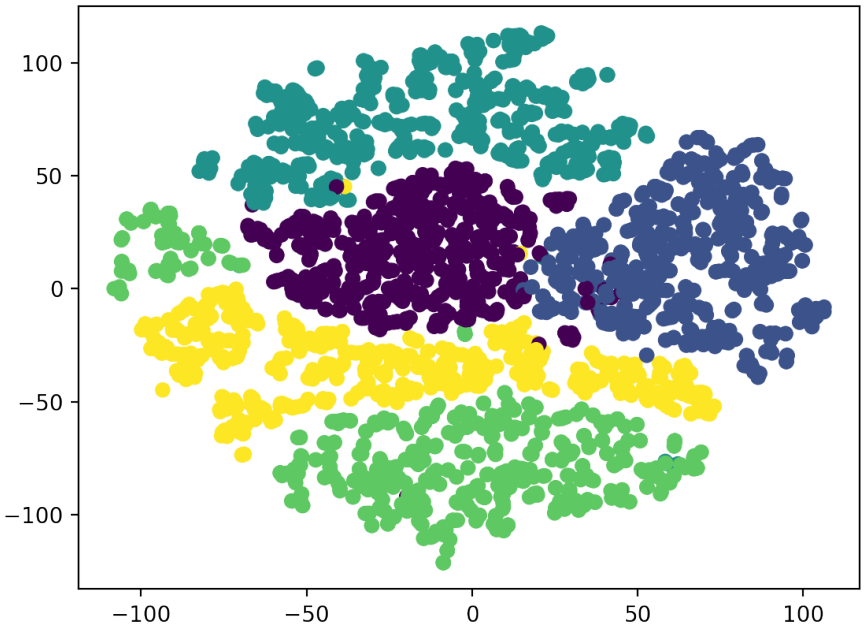
distribution in the low-dimensional mapped points.

Fig. 3.4a illustrates the visualization of the learned representation on the MIT-BIH dataset samples. As it can be seen from this figure, data-points from different classes are easily separable using the learned representation. Fig. 3.4b shows the visualization of the MI classification task on the PTB samples using the representation trained on MIT-BIH. It can be inferred from this figure that the transferred representation for the beat classification task is able to provide a reasonable separation for the MI classification task. It should be noted that here we only use class labels to colorize the plots and other than this we do not use sample labels in the visualizations.

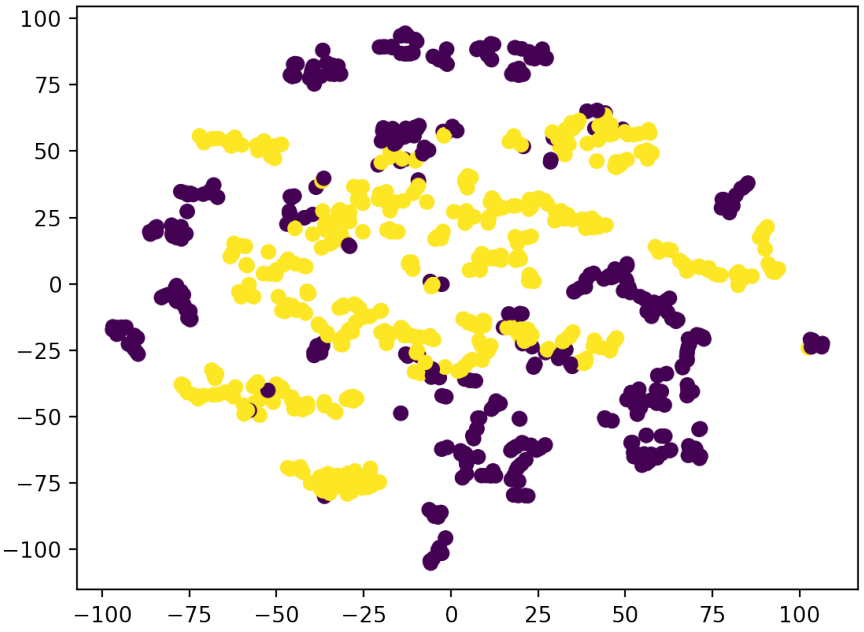
3.5 Discussion

3.5.1 Conclusion

In this study we have presented a method for ECG heartbeat classification based on a transferable representation. Specifically, we have trained a deep convolutional neural network with residual connections for the arrhythmia classification task and shown that the representation learned for this task can be used as a base to train accurate classifiers for the classification of MI. According to the results, the suggested method is able to make predictions on both tasks with accuracies comparable to the state of the art methods in the literature. Furthermore, we visualized the learned representation using t-SNE method and illustrated the effectiveness of the proposed approach.



(a)



(b)

Figure 3.4: t-SNE visualization of the learned representation: (a) samples from MIT-BIH for ECG beat classification (b) samples from PTB dataset for MI classification. Labels for each task are indicated with colors (best viewed in color).

CHAPTER 4

Body Activity and Movement Analyses[1]

4.1 Introduction

In the HealthAI domain, a crucial form of longitudinal digital observation involves modeling bodily movements through time. From the short-term tasks, such as posture or activity recognition, to the longer-term tasks, such as the trajectories related to the dynamic range of body anchors through time, determining the rehabilitation progress, having to deal with the challenges of modeling longitudinal data is an inevitable component of these healthcare solutions.

In this chapter, we discuss three problems in healthcare to which we have offered solutions, namely:

[1]**This chapter is based on the following papers [GFE19, FKS, HFV18]:**

- **Fazeli S, Kachuee M, Sarrafzadeh M, Aminian A. WatChair: AI-Powered Real-time Monitoring of Sitting Posture and Corrective Suggestions using Wearable Motion Sensor System.**
- **Hosseini A, Fazeli S, van Vliet E, Valencia L, Habre R, Sarrafzadeh M, Bui A. Children activity recognition: Challenges and strategies. In2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) 2018 Jul 18 (pp. 4331-4334). IEEE.**
- **Gwak M, Fazeli S, Ershadi G, Sarrafzadeh M, Ghodsi M, Aminian A, Schlechter JA. Extra: exercise tracking and analysis platform for remote-monitoring of knee rehabilitation. In2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN) 2019 May 19 (pp. 1-4). IEEE.**

- In one of our projects[2], a cohort of pediatric patients who have asthma were selected to be monitored, and the study reflected on the importance of proper activity recognition for improving the patient outcomes. In terms of the applications, for example, to provide proper interventions in terms of alarms and health notifications to these individuals, we aimed to develop an activity recognition system specifically targeting the cohort of pediatric asthma patients.

- The prevalence of a sedentary lifestyle has led to various health problems, especially those related to individuals maintaining bad postures throughout the day. In this work, we aimed to provide a sitting posture monitoring system so as to leverage a recognition core and a cell app to monitor and track the posture through time using a wearable component, as well as provide users with the breakdown of their activities and other health data to raise their awareness.

- Knee replacement surgery is a common operation requiring a long time for the patient to recover the range of motion in their knee joints. In this work, we leveraged the readings from a wearable component we created using flex sensors to create statistically significant observations of the patient's knee movement. The goal is to provide an inexpensive system to measure the angles with reasonable accuracy. Such a system can be used to create longitudinal records for the medical experts to review and further personalize the treatments for the individuals.

4.1.1 Activity Recognition for Children with Pediatric Asthma

With the recent prevalence of smartwatches and smartphone technologies, wireless health systems and mobile health (mHealth) applications are increasingly adopting these technologies for healthcare applications. As the popularity of smartwatches for health monitoring

[2]BREATHE Project at UCLA, focusing on improved healthcare for children with pediatric asthma [HBH17]

grows, so do the challenges that come with finding meaning in the newly available physiological data. One goal of many existing studies is to use data such as accelerometer and gyroscope readings for activity recognition. While significant advances have been made in this area [CPR11, QMX10], it remains an active area of research, partly because of the complexity and diversity of human movement based on age, health condition, and behavioral patterns. Following our study aimed at the prediction and prevention of asthma attacks in children [HBH17] and clinical studies showing the impact of activity level on asthma exacerbation [KKE17], an activity recognition model for children seemed necessary. However, the major focus of most activity recognition studies has been on adults [QMX10, YWC08], and the classic machine learning models used for adult activity recognition often do not translate well to children [SMI12]. This is partly because of significant differences in the way children and adults perform basic activities such as running or climbing stairs.
In general, activity recognition imposes two main challenges specific to children:

1. Collecting large labeled datasets for children is difficult, partly due to the fact that children tend to change activities more frequently and listen less well to the experimenter's commands. In most activity recognition studies, class labels are obtained in constrained laboratory settings. [BI04].

2. There tend to be significant variations between children when performing the same activities, which translates to more variation in the signal data being processed for activity recognition.

Smartwatches show two main advantages over smartphones when studying health monitoring solutions for children. Firstly, unlike smartphones that are bulky to handle during different activities, they allow for continuous data monitoring throughout the day and can easily be worn even in high levels of activity. This is especially important for applications such as asthma management. Secondly, they can collect additional data such as heart rate and externally connected sensor data for health monitoring purposes which are particularly

promising in applications that require multiple sensor data for remote health monitoring and management.

Our study aims to tackle the smartwatch-based activity recognition problem specifically for children. For this aim, we collected labeled activity data from 25 children aged 8-14 and compared the performance of deep neural network models to the widely used classic activity recognition models.

We show that mentioned challenges and limitations of smartwatches introduce major errors to basic models and demonstrate that using a bidirectional recurrent neural network (BRNN)[GS05] can improve the results compared to other baseline models while not adding too much complexity as fully connected deep neural network models. We then show the results of our proposed models when used to capture the intensity of activity level with high accuracy, which is essential for remote monitoring of children and our efforts in asthma attack prevention. To the best of our knowledge, this is the first study with a systematic focus on smartwatch-based activity recognition for children. The results of this study can pave the way for other children's health monitoring platforms that are dependent on children activity recognition.

4.1.2 Sitting Posture Tracking via Wearable Sensors

The sedentary lifestyle has led many people to spend a considerable portion of their lives sitting. The act of sitting has the potential to cause severe short-term and long-term health problems if not done correctly, such as the feeling of pain and discomfort in the neck and back area [NKB18] [SSG18] [KKH18]. Numerous studies indicate a strong connection between the sitting posture in performing different activities with outcomes such as health status and eating, claiming that controlling and maintaining proper sitting posture can help [MYP20] [TBK19]. As simple as it might seem, this activity has also been known for its impacts on certain types of decision-making, such as online grocery shopping [KF19].

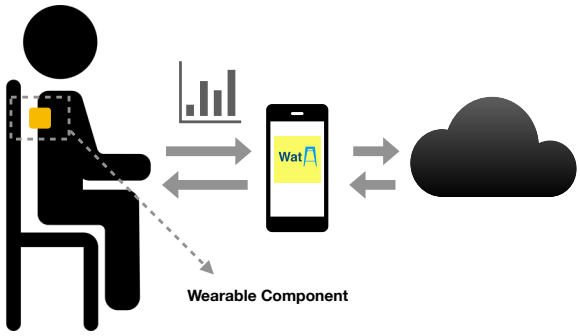


Figure 4.1: The system architecture of our posture monitoring platform

Due to the reasons mentioned above, active monitoring and interfering with sitting postures is an exciting area of research. This stems from the fact that a convenient and thorough solution for this problem has the potential to change many lives and is, therefore, invaluable.

In our study, we introduce a platform for continuous monitoring of sitting posture that aids subjects with corrective suggestions. This work attempts to bring a smooth remote monitoring experience to the subjects while attempting to keep the cost and intrusion associated with the data acquisition system as low as possible. This is done by focusing on providing more efficient solutions to the software-related aspects of the problem. Our system will monitor the sitting posture, recognize different sitting posture patterns, and help the subject with a thorough report of the statistics of their sitting posture throughout the day. It also provides them with their history of sitting information, assisting them with easy progress tracking.

4.1.3 Tracking Post-surgery Rehabilitation Knee Exercises

Knee reconstructive surgery is one of the most common orthopedic procedures in the United States. Orthopedic surgeons decide on the necessity of knee reconstructive surgery due to a variety of reasons on the knee, such as severe pain, stiffness, or swollen knees. After surgery, therapeutic exercises are essential in keeping down swelling and enhancing muscle

strength [web]. Therefore, patients must perform a set of exercises correctly and regularly to experience an early and safe recovery.

Current orthopedic treatments and rehabilitation exercises mostly do not support quantitative ways to measure knee activity or to provide biofeedback. The intelligent ways to track patients' rehabilitation progress are under active research. The time and financial limitations of visiting orthopedists or physical therapists may resist early recovery from knee reconstructive surgery, which further motivates proposing solutions that can help in this area.

We propose EXercise TRacking and Analysis Platform (EXTRA) for remote monitoring of knee rehabilitation. Our system uses an embedded flex sensor in a commercial hinged knee brace to measure the knee flexion-extension angle. The connected Android application receives and transmits the sensor values to the cloud-based database. Real-Time monitoring of the knee angle and an exercise progress chart support improving the quality of the therapeutic exercises and knee rehabilitation.

4.2 Data

4.2.1 Activity Recognition for Children with Pediatric Asthma

To collect children's activity data, an Android smartwatch app was designed to record accelerometer and gyroscope signals in real time and transfer the anonymized data to a web server for activity prediction. Figure 4.2 illustrates the overview of our activity recognition system. In this study, 25 children (10 girls and 15 boys) aged 8 to 14 were recruited to participate. Each child was asked to wear the smartwatch and perform six different activities as instructed. Activities included running, walking, standing, sitting, lying down, and stair climbing, and each was recorded for a duration of 10 minutes, and sensor data was collected with a frequency of 10 Hz. After instructing the children before each 10-minute time span, they were left free to perform activities to obtain real-world data. Data collection

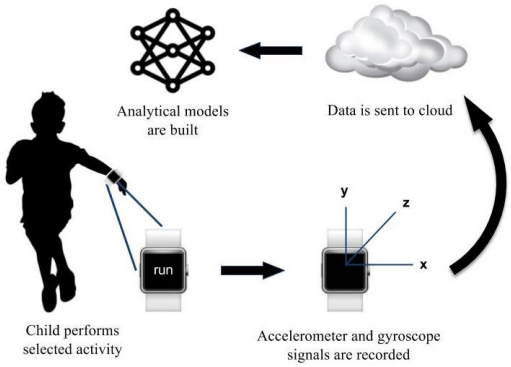


Figure 4.2: The overall system architecture for a wearable activity monitoring device.

was stopped during time spans when children stopped doing the required activity.

4.2.2 Sitting Posture Tracking via Wearable Sensors

Our cohort comprises six subjects in the age range of 24-27. The central and only sensory part of this platform's data acquisition component is composed of affordable, portable, and widely available MetaTracker boards [mbi19]. These boards are widely used when motion sensors are utilized for evaluating physical readiness, an example being training pilots [KVH19]. The single wearable component of this platform is placed on the subject's back and between the arms, which can be easily done using a strap or cord. The main component for the effective tracking of sitting habits is posture recognition. The main three postures that our platform focuses on recognizing are depicted in Figure 4.3 [anR], and empirical results indicate that the data acquired from this location is primarily sufficient for making such a determination. In addition, the sensor alignment can be automatically determined, and the system is flexible in terms of small displacements in using the wearable component.

The wearable component uses a Lithium battery, and in our framework, there is no need for frequent battery replacement. This is due to the fact that our framework is focused on seated posture and is compatible with low-frequency sensor readings. This property and being equipped with 2.4 GHz Bluetooth Low Energy chips enables us to maintain a smooth

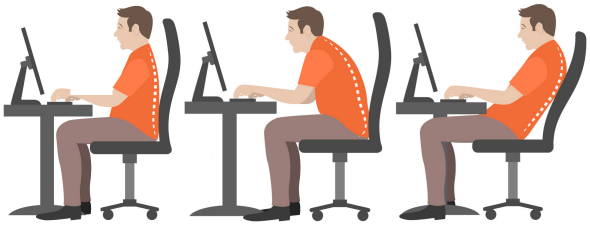


Figure 4.3: Three main postures for sitting habit monitoring [fre]

connection route between the wearable component and the app.

The mobile application then handles the data-related routines, including cloud storage, analytics, and transmission.

4.2.3 Tracking Post-surgery Rehabilitation Knee Exercises

We have utilized flex sensors to measure the knee joint angle and leverage their reading as our main observed signal. According to the degree of bending, the resistivity of the flex sensor changes and results in different values for the voltage on the two ends of the sensor. The voltage is continuously read and quantized through the microcontroller.

We selected Adafruit Feather [Ada] for the microcontroller board, which is light, thin, Arduino compatible, and widely available at a considerably low price. It has up to 6 analog reading pins, which can support flex sensors for measurement. Feather features a JST jack to which a 4.2/3.7V Lithium Polymer (Lipo/Lipoly) or Lithium Ion (LiIon) battery can be connected and a USB port through which the board can be programmed and debugged. The USB port also allows the board to run on a rechargeable battery. Feather automatically switches to USB for power when the USB cable is connected and begins charging the attached battery at 100mA. An LED on the board starts blinking while the power is on. We added an on/off switch to minimize power consumption and make it more convenient for patients to power on/off the board as well. Feather also includes Bluetooth Low Energy (BLE) development that makes the board-to-phone transmission take place with a new low-power, 2.4GHz spectrum wireless protocol. The analog readings on the board are transferred to the

Figure 4.4: The system before placing on the knee brace (left) and the final version of the system in the knee brace (right).

connected Android application via BLE in a real-time manner.

We placed the compact system on the standard commercial hinged knee brace commonly used as an orthopedic treatment (Figure 4.4) so that the knee angle can be measured precisely and patients can easily employ it. One end of the flex sensor was affixed on a metal hinge of the knee brace, and we covered the circuit with fabric. It is noteworthy that any commercial hinged knee brace should be able to contain the flex sensor. Our data acquisition component is economically efficient because of the low-cost electronics and an easily purchasable knee brace. Our goal in designing the system was minimizing expense and power consumption, besides ease of use for patients.

4.3 Methods

4.3.1 Activity Recognition for Children with Pediatric Asthma

4.3.1.1 Data Representation

Representing raw signals received from sensors in an informative way is necessary for achieving better data interpretations. Although deep machine learning models allow for automated representation learning [ASN16], [YHZ17], they ask for a tremendously large amount of labeled data, which is not available for children activities. Time-series signals can be described nicely by their time-domain and frequency-domain features. This enables us to efficiently work with small datasets for training high-performance inference models without the need for heavy automatic representation learning. Therefore, we first employed a time-based windowing technique as it has shown superior performance compared to other methods [BGD14] to segment the signal. Next, informative statistical features were extracted for each time window to form its representation. We extracted widely used features for time-series analysis, studied in the literature [FDF10]. Table 4.1 lists the features extracted in this study from each or a couple of axes of accelerometer and gyroscope signals.

Table 4.1: Features extracted for each window in pre-processing stage

Feature Type	**Axis**
Every Single Axis	Mean, Median, Range, Min, Max, std, 25 and 75 percental RMS, zero-crossing, fft-entropy Range, Integration
Every Two Axes	Correlation, delta
All Axes	Signal Vector Magnitude

4.3.1.2 Deep Models

Deep neural network models have shown outstanding performance in information discovery and have outperformed classic and shallow models in learning hidden relations in different areas [Sch15]. We study and compare two deep models, one with a sequential processing design (RNN) and one with a deep, fully connected structure.

Fully Connected Design: Once our signal is segmented into feature representations (windows), each window coupled with a label can be viewed as a data sample to be fed into the prediction models. To study the performance of deeper models over shallow ones, we designed a multi-layer neural network model in which layers were decoupled from each other by ReLU nonlinear activation functions. As mentioned earlier, activity recognition datasets for children are usually small, and more complex deep models cannot be trained on them. The best model found was a three-layer neural network with 100 neurons in each hidden layer. The model was trained using the Adam optimization scheme to help with the convergence.

Sequential Design: In another approach, instead of working with one single window at a time, we can view extracted representations as a consecutive window sequence and learn from activities happening close to a time window. A Bi-Directional Recurrent Neural Network, specifically a Bi-directional LSTM network, was employed for this aim. Such a model can receive a sequence of input activity windows of any length and, in each stage, make the prediction based on the input in that stage and the memory state that is formed by processing the previous windows. LSTM [CGC14] configuration was employed to tackle the problem of vanishing gradient and enable the network to better recognize what information is worth remembering and what is not. Moreover, Bi-directional architecture enabled the model to process the sequence of inputs in both forward and backward directions to better capture time dependencies for each activity. The structure of the designed BiLSTM network is shown in Figure 4.5. In our setup, an LSTM layer in which each unit has 32 neurons is

followed by two layers of the fully connected network with 50 and 25 neurons, and prediction is done through a final Softmax layer.

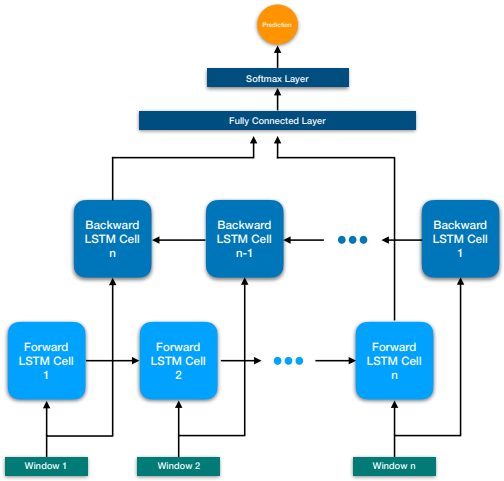


Figure 4.5: Bi-Directional LSTM, along with fully connected and softmax layer to predict the activity based on the sequence of time-window representations

4.3.2 Sitting Posture Tracking via Wearable Sensors

The overall architecture of WatChair is depicted in Figure 4.1. The system's first part is a simple wearable component, which includes the motion sensors we are interested in monitoring. This component would be located on the subjects' backs and between the arms and used for sampling data points and transmission. The system obtains its electrical power from a coin battery inside it, and given the frequency of reading, there is no need for frequent battery replacement.

This wearable component interacts with an Android device using its mobile app. Upon running the application, the connection is maintained if the device is in range, and the data acquisition, recognition, and transmission process to our cloud backend begins.

The application provides a user-friendly interface between the board, the cloud backend, and the subject. The user can observe long-term and short-term information about their

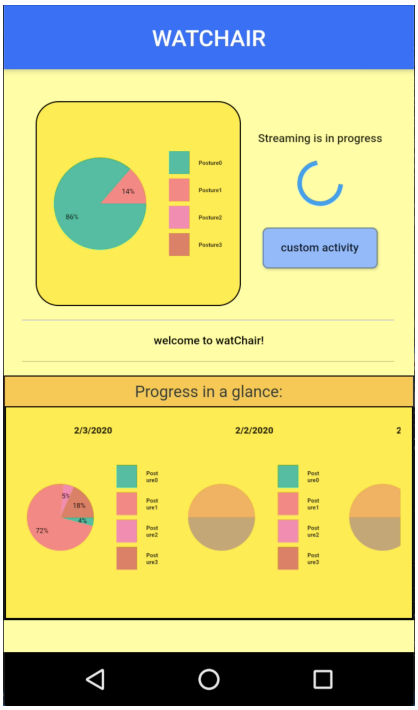


Figure 4.6: The main interface of our application is shown in this figure.

pattern of sitting and score their seated behavior accordingly. This framework renders it possible to fine-tune the models as more data is gathered in the cloud database. In this work, we attempted to evaluate the performance of a general model with no need for per-person calibration. Nevertheless, this framework can be used with ease to provide each user with separate user-calibrated models as well.

4.3.3 Application

Another essential component of our system is the mobile application. Upon launching the application, it writes the configuration necessary for the motion sensors, such as the sampling frequency, and issues the start command for the wearable component. The information then can be monitored continuously using the app's interface, as depicted in Figure 4.6.

Afterward, the information is continuously retrieved, filtered, and pushed to our cloud database. We use Google Cloud Firestore as a NoSQL database composed of documents

and collections for the backend database. The application builds a communication channel to the Firebase instance in order to efficiently perform the transactions. These transactions include pushing the gathered data to the cloud database, an instance of Cloud Firestore, and reading the historical information to update the app's visualizations.

Please note that the related intervention information regarding how to correct the posture is also presented in the application, providing the subject with a more in-depth understanding of how to proceed regarding the given personalized posture correction information.

The quality of the application was surveyed through a usability questionnaire developed according to the Usefulness, Satisfaction, and Ease-of-Use questionnaires in [Lew95a, Lun01a]. This questionnaire was given to the subjects in our study, with the idea of receiving feedback on the application as the main component in this framework. Via that questionnaire, it was concluded that the users found getting familiar with the framework and using it a smooth and straightforward process.

Recognition

The machine learning component of our platform is composed of an inference pipeline that directly links sensor readings to different sitting postures and characteristics. The trained model weights are used in the application to enable efficient use and effortless alterations in the future. The relevant sensor readings that come from motion sensors (mainly, the three-axis sensors of accelerometer and gyroscope) are buffered and transmitted to the application. In our experiments, we considered the time window of 60 seconds; therefore, the seated behavior of the subject for every minute is represented and used by the model for evaluation. The findings are then stored with the timestamps to help with the progress review.

The recognition pipeline is designed using Support Vector Classification, which enables differentiating between the labels using one-vs-all classifiers. The training objective uses hinge loss and is as follows:

$$\min_{\theta,b} \frac{1}{2}\theta^T\theta + C\sum_{i=1}^{N} \max(0, y_i(\theta^T\phi(x_i) + b)) \tag{4.1}$$

In the above formula, θ and b are the model parameters that are trained, and ϕ is the linear kernel. The Support Vector classification with a linear kernel is chosen so the model complexity could fit the problem well, and the trained model could be easily implemented and used in the mobile application as well.

4.3.4 Tracking Post-surgery Rehabilitation Knee Exercises

4.3.4.1 Mobile Application

A connected Android application converts the sensor readings to the angle of flexion and extension in degrees. The knee brace sends analog readings of the flex sensor to the connected app. We observed that the sensor readings and the linear motion of the knee joint are linearly correlated.

$$b = y_1 + (x - x_1)\frac{y_2 - y_1}{x_2 - x_1} \tag{4.2}$$

$$A = sx + b \tag{4.3}$$

For the sensor reading calibration, we used a linear interpolation (4.2) where (x_1, y_1) and (x_2, y_2) are two different angle observations. The app uses (4.3) to convert the received sensor readings (s) to the knee joint angle (A).

We also estimated the effort put into the knee activity using the standard deviation of the samples. The standard deviation is an accepted measure denoting the intensity of signal fluctuations, which works as an average of power instead of amplitude. This measure successfully distinguishes active knee movement from staying still. In each exercise session, we would compute: S_x and S_{x^2} which are defined as follows:

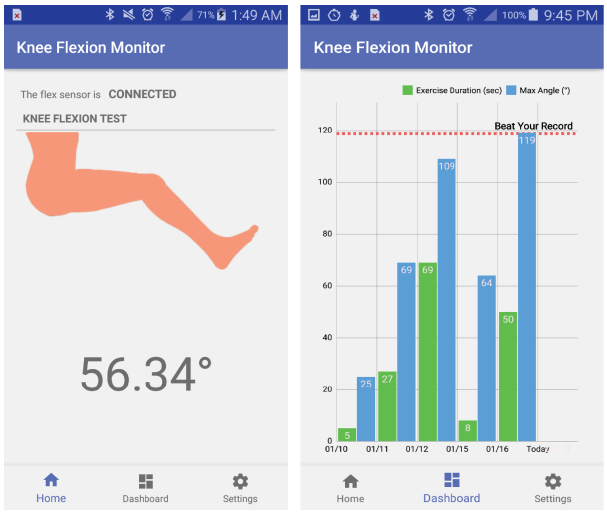


Figure 4.7: Android application of EXTRA: (left) a real-time monitor of the knee joint angle and (right) an exercise progress chart

$$S_x = \sum_{i=1}^{n} x \tag{4.4}$$

$$S_{x^2} = \sum_{i=1}^{n} x^2 \tag{4.5}$$

The values of x are normalized quantized values read from the flex sensor. With the number of samples, the standard deviation is computed using the following formula:

$$\sigma = \sqrt{E(x^2) - E(x)^2} \tag{4.6}$$

The users do not require extra inputs than turning on the knee brace and the app. The app automatically connects the knee brace and saves helpful information about knee activities. The app searches the microcontroller board by the unique board name and connects via BLE when the power of the board on the knee brace is on. After each exercise session, the app also transmits the received data to the cloud-based database.

The Android app provides a real-time monitor of knee flexion and a daily summary of

the activities. Our user-friendly app increases the motivation for required exercises. The left screenshot of Figure 4.7 shows the charts of the knee joint angle, and the right screenshot shows the knee activity progress chart. The users can easily understand the connection status with the knee brace. The users can also catch the flexion angles during the knee exercise through the numerical value and the illustration. The exercise progress chart shows the daily achievement of exercise duration and the maximum angle of flexion. When the active knee movements are detected using the standard deviation, the exercise duration is increased. The *Beat Your Record* line encourages the user to perform better than the previous maximum angle of flexion through the exercise. For knee-injured patients, our approach of knee activity monitoring and the progress chart is less intrusive and supports therapeutic exercises effectively to speed up rehabilitation.

4.3.4.2 Data Storage

The Android app saves valuable information about knee activity to the Google Cloud Firestore, a cloud-based NoSQL document database. The Firestore allows the app to access the database directly and to store the records as much as necessary in real-time. The Firestore also provides an analytics library to perform machine learning and data science-related analyses on the gathered data and real-time visualization to display the data. The app stores the knee joint angles, the maximum angle during the knee movement, in the database. The app also loads the recorded data from the database to create an exercise progress chart.

4.4 Experiments

4.4.1 Activity Recognition for Children with Pediatric Asthma

To validate the performance of our proposed deep models, we chose two widely used models in the area, a random forest (RF) and a shallow one-layer neural network (FF1), as our base-

Table 4.2: Mean Absolute Error value for experimental angles

Angle	**MAE**
0	0.19
15	8.74
30	24.95
45	12.13
60	15.06
75	13.07
90	17.68
105	14.66

lines. Both models were tested over validation sets to find the best configuration: training a random forest with 50 decision trees and a neural network with 100 ReLU nodes. Figure 4.8 shows the F1-score of all models for predicting each activity. It can be easily inferred that the RNN model shows higher performance than the other deep model (FF3) and baseline models. Although FF1 shows strong results in detecting walking activity, it cannot efficiently detect sitting or standing activities. Random Forest also shows competitive results with deep models in walking and running tasks; however, it performs poorly on other tasks compared to deeper models. We can also see that the deep network (FF3) achieves competitive results to RF. However, it cannot beat the RNN model in most cases. One crucial observation is that the obtained results from children show generally lower accuracy than reported results for adults [QMX10] due to high variance in their activity shape.

To study the source of accuracy loss in previous results, we analyze the confusion matrix of activities in Figure 4.9. We can infer that the model has difficulty distinguishing between sitting and standing. This confusion is due to the fact that movement of the wrist in these

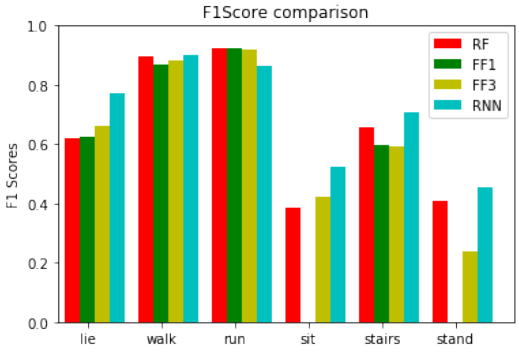


Figure 4.8: F1-score results of activity prediction comparing deep models to baseline models

two activities can be very similar to each other especially when the hand is left free to the side. The same confusion arises for lying and standing activities. These challenges are inevitable when smartwatches are used as the activity recognition tracker.

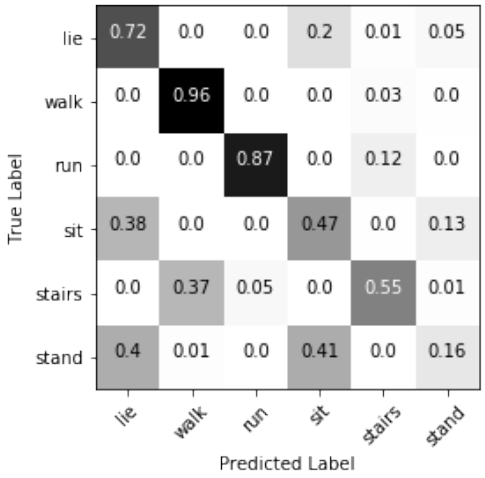


Figure 4.9: Confusion matrix for six detected activities

As discussed earlier, in many applications of activity recognition on children, such as asthma exacerbation prevention, the intensity level of activity is of greater importance than the activity itself. Figure 4.10 demonstrates the performance of our model when classifying activities into three levels of low, medium, and high intensity. The RNN model allows us to detect the intensity levels of activities in children with more than 80 percent of the average

F1 score.

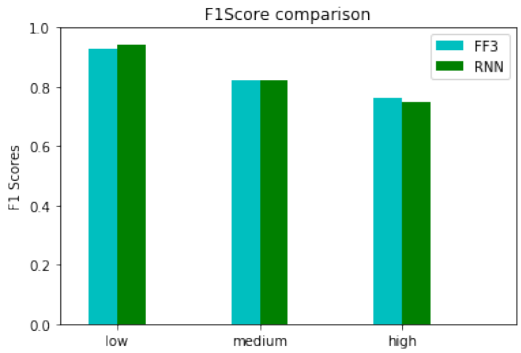


Figure 4.10: F1-score performance on activity intensity level prediction

Overall, the results of our experiments demonstrate the superiority of recurrent neural networks in detecting children's activity.

4.4.2 Sitting Posture Tracking via Wearable Sensors

Our analytical platform is implemented in Python3.6, and using Google Cloud SDK and Sci-Kit learn machine learning library enables us to retrieve the information from our cloud database and perform the corresponding analytical investigations. For evaluation purposes, several experiments were done, and in each experiment, the data from one subject was used as the test set while the model was trained on the rest.

The empirical results indicate that our model is able to recognize and distinguish between our three main pre-defined sitting postures accurately. The micro and macro average F1-Score for our current model over the three sitting postures are 72.12% and 75.09%, respectively.

The normalized confusion matrix for the predictions is shown in Table 4.4. These results suggest that the system exhibits accurate performance, even though the flexibility of the system in terms of the positioning of the wearable component can potentially add to the error level. As another example, "Slouching" data, which can be argued as the primary

posture of interest, has been efficiently captured.

This work aimed to propose an accurate and efficient pipeline to perform monitoring without needing per-person calibration. However, it is understandable that in specific use cases (e.g., certain disabilities or the cohort of older adults), adding per-person calibration might be necessary to improve the performance further. In such cases, labeled data can be easily obtained and analyzed using our platform's interface.

To use the model in the mobile application as well, this inference engine was implemented in Java and is included in the application to perform continuous recognition and update the cloud database accordingly.

4.4.3 Tracking Post-surgery Rehabilitation Knee Exercises

The accuracy of measuring the knee joint angle and the usability of our system was evaluated by eight healthy adults (3 female, 5 male, age range: 23-27). We asked the participants to wear the knee brace and to bend the knee with different angles while sitting down on a chair. We considered a wide range of angles from 0.0 to 105.0 degrees (with increments of 15.0 degrees) and confirmed the validity of the knee angle with the use of a goniometer. We collected 60 samples for each participant corresponding to 30 seconds of angle recording. After assessing different angles, the participant performed two exercises: quad lift while sitting and standing knee lift. These exercises helped with experiencing the mobile app's interface and wearing the knee brace. After the exercises, the participant filled out a questionnaire regarding the usability of our system. The questionnaire was developed based on the Usefulness, Satisfaction, and Ease-of-Use (USE) questionnaires on [Lew95b] and [Lun01b].

Table 4.2 shows the mean absolute error (MAE) for each angle measurement. The average MAE was 13.31 degrees, but this value indicates sufficient accuracy for monitoring therapeutic exercises. Our knee joint angle measurement system is quite sensitive to the body postures and different shapes of the leg. We observed that bending the spine or mov-

ing arms caused changes in the measurement of the knee joint angle. A participant with lateral pelvic tilt had a higher MAE about the angle measurement than others. Further calibration is required to support reliable knee flexion-extension angle measurement.

The results of the usability questionnaire are shown in table 4.3. The high average score of the questionnaire indicates that our software components effectively deliver information about knee activity. The mobile application does not require complicated instructions. Any individual can install the system and interact with the app easily.

4.5 Discussion

4.5.1 Activity Recognition for Children with Pediatric Asthma

The solution we proposed focused on the problem of activity recognition for children, which is of great importance in the domain of remote health monitoring for children. We showed that variance in activity and limitations of smartwatches are significant challenges for this task. We also demonstrated that RNN-based models can be a good choice for children activity recognition because of their ability to capture more information while being simple enough to be trained on small datasets. Future work in this area should focus on personalized learning of activities for children. Learning a model for a group of children cannot translate well to each individual. However, if models adapt to each child's activities, results can be improved.

4.5.2 Sitting Posture Tracking via Wearable Sensors

We presented an AI-powered remote monitoring framework for short-term and long-term tracking of sitting habits and proposing corrective suggestions. The effectiveness of this system in performing this task is then empirically validated. Given the critical health impacts of improper sitting habits, this low-cost and affordable system, combined with accurate machine learning inference, can improve user behavior while seated and help prevent medical

complications associated with improper sitting.

4.5.3 Tracking Post-surgery Rehabilitation Knee Exercises

We presented a remote-monitoring system that supports exercise tracking and analysis for knee rehabilitation. Our platform provides a portable, wearable, low-cost, reliable system to measure the knee angle. Recording the rehabilitation progress encourages users to exercise. We validated our solution's angle measurement and usability, which may benefit orthopedic professionals and patients going through knee reconstructive surgery by assisting with effective rehabilitation.

Table 4.3: Usability Questionnaire and the average score: Participants scored between 1 (strongly disagree) to 10 (strongly agree) to respond each question.

Question	**AVG**
I am comfortable with using this system again.	9.89
The system provided me with helpful information regarding my knee movements.	9.78
The app was responsive and showed the values in a real-time manner.	8.56
I could effectively perform the requested exercises using the system.	9.67
The knee pad was very comfortable to wear and did not affect the quality of my motions.	8.89
It was simple to learn to use the application (UI).	10.0
Proper messages in the application were provided to inform me of any unexpected behavior.	9.0
I believe this system provides the user with incentives for improving the knee exercises and staying on track.	9.67
I believe that using this system will positively influence the quality of therapeutic knee exercises.	9.89
The visualizations and the information provided by the app were very clear.	9.89
The interface of this system was pleasant.	9.89
The system has all the capabilities I expected it to have.	9.67
Switching the system on and off was easy for me to do.	10.0
It requires the fewest steps possible to accomplish what I want to do with it.	9.89
I was able to use the system without written instruction.	9.33
I did not notice any inconsistencies as I use it.	9.0
Setting up the system for an exercise session, including wearing the knee pad and starting the app was very simple.	9.33
I easily remember the steps of using this system.	10.0
I would recommend this system to others.	10.0
Overall, I am satisfied with using the system.	9.67

Table 4.4: Micro-averaged confusion matrix for evaluating our model using leave-one-subject-out scheme - Each row corresponds to the predictions for a ground-truth label

	Upright	**Leaning Back**	**Slouching**
Upright	0.65	0.22	0.14
Leaning Back	0.05	0.64	0.31
Slouching	0.02	0.0	0.98

CHAPTER 5

Transferable Representation Learning for Electronic Health Records[1]

5.1 Introduction

Electronic health records (EHR) are commonly adopted in hospitals to improve patient care. In an intensive care unit (ICU), various data sources are collected on a daily basis as preempted by medical staff as the patient undergoes care in the unit. The collected data consists of data from different modalities: medical codes such as diagnosis, which are standardized by well-organized ontology's like the International Classification of Disease (ICD)[2] and medication codes standardized using National Drug Codes (NDC)[3]. Similarly, at various stages of the patient's care, physicians input text noting relevant events to the patient's prognosis. Additionally, lab tests and bedside monitoring devices are used to collect signals, each of which is collected at varying frequencies for a quantitative measure of patient care. There is a wealth of information contained within EHRs that has a significant potential to be used to improve care. Examples of inference tasks using such data include estimating the length of stay, mortality, and readmission of patients [TLA87, CCA08].

The traditional approach to healthcare analysis has mainly focused on classical methods

[1]**This chapter is based on "Darabi S, Kachuee M, Fazeli S, Sarrafzadeh M. Taper: Time-aware patient ehr representation. IEEE journal of biomedical and health informatics. 2020 Apr 3;24(11):3268-75." [DKF20]**

[2]http://www.who.int/classification/icd/en

[3]http://www.fda.gov

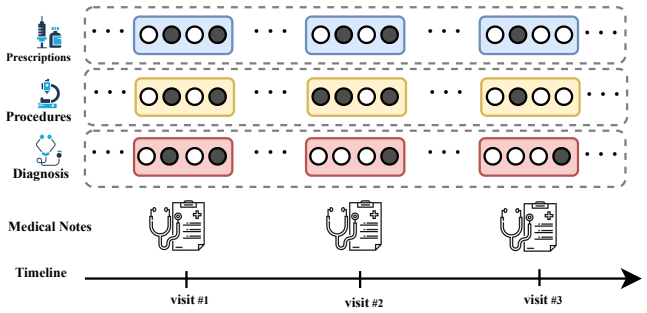


Figure 5.1: Patient timeline during an ICU visit where different data points are collected. These include prescriptions, diagnosis codes, procedure codes, and medical notes.

for extracting hand-engineered features and designing rule-based systems. More recently, deep learning has demonstrated state-of-the-art results on various tasks, in which learning intermediate representation is at the heart of all these analyses [BCV13]. This representation can be obtained without domain-specific expertise by leveraging available EHR data. Although such methods have demonstrated remarkable performance on image and audio datasets, leveraging deep learning techniques on healthcare data presents new challenges as the data entered are sparse and contain different modalities.

As is common in natural language processing tasks, the typical method for embedding medical codes and text could be through the use of one-hot vectors. However, these are naturally high-dimensional and sparse resulting in poor performance. To alleviate this, the idea of learning distributed representations as applied to natural language processing [MSC13] has also been applied to medical data [CGN18b]. Such methods share the intuition that similar medical codes should share a similar context. Additionally, codes have varying temporal contexts; as such, patients may have multiple visits with a similar set of codes. As an example, flu is short-lived, whereas a diagnosis code for a more terminal disease such as cancer has a more extended scope and hence will be present on all of the patient's visits. Due to the varying temporal context, it is also essential to take into account the temporal scope of codes and text assigned [CGN18b]. This demands a model which takes the sequential

dependencies of the patient's visits into account.

To capture the sequential dependencies present in medical data, recurrent neural networks (RNNs) and especially RNNs configured as Long Short-Term Memory (LSTM) have been considered as powerful models. RNN auto-encoder models are commonly augmented with attention mechanisms allowing the model to attend to specific time steps either through soft/hard attention resulting in improvement and interpretability in the final representation obtained [BCB14b, LPM15]. In NLP tasks, such attention mechanisms are not required to be causal in time. They hence can attend to both past representations as well as future representations to generate the current representation. However, in a healthcare setting, it is desirable to have the representation be causal in time as clinical decisions are made sequentially. Recently, transformer models [VSP17] were proposed for natural language processing tasks and have shown impressive results. It uses self-attention, and as the model creates intermediate representations of the input, it attends to its representation at previous and future timesteps when considering the present representation.

The majority of patient representation work has solely focused on embedding medical codes or text as a patient representation for downstream tasks but not both. To address this, we study the use of transformer networks to embed structured medical code data as well as a language model to embed the text portion of visits. In this chapter, we discuss our proposed EHR representation learning solution to combine the medical representation from text and medical codes into a unified embedding which can then be used for downstream prediction tasks. Lastly, the presented study takes into account the temporal context of a visit and embeds subsequent visits given the patient's history.

5.2 Data

We evaluate our model on the publicly available MIMIC-III Clinical Database [JPS16]. It consists of EHR records of 58,976 hospital admissions of 38,597 ICU patients from 2001 to

2012. On average, each patient has 1.26 visits. The database contains tables associated with different data, where we extracted demographics, medical codes, and medical notes.

5.2.1 Readmission task

The first task is to predict 30-day unplanned readmission to the ICU after being discharged. In this task, we formulate it as a binary problem, that is, to predict whether a patient will be readmitted within 30 days after being discharged. Text entered into the MIMIC database contains different reports, such as nurse notes, lab results, discharge summaries, and so on. We limit the text for each visit to contain the discharge summaries or text entered within the last 48h before the patient is discharged in the absence of discharge summaries.

5.2.2 Mortality task

The second task is to predict the mortality of patients, whether they passed away after being discharged or within the ICU. Similar to readmission, it is formulated as a binary task. In this task, mortality-related codes are discarded from the dataset, and patients admitted for organ donation are removed. Additionally, the input text for each visit is limited to the first 24h of the admission.

5.2.3 Length of stay task

The third task is to forecast patients' length of stay (LOS). In this task, longer LOS indicates more severe illness and complex conditions. We formulate this problem as a multi-class classification problem by bucketing the length of stay into 9 classes: 1-7 correspond to one to seven days, respectively, 8 corresponds to more than 1 week but less than 2, and 9 corresponds to more than 2 weeks. The model is tasked to predict $P(Y = L|Z_t)$ where $L \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ denoting the previously defined time intervals. In this classification task, we limit the medical text to the first 24 hours of the current patient visit in which the

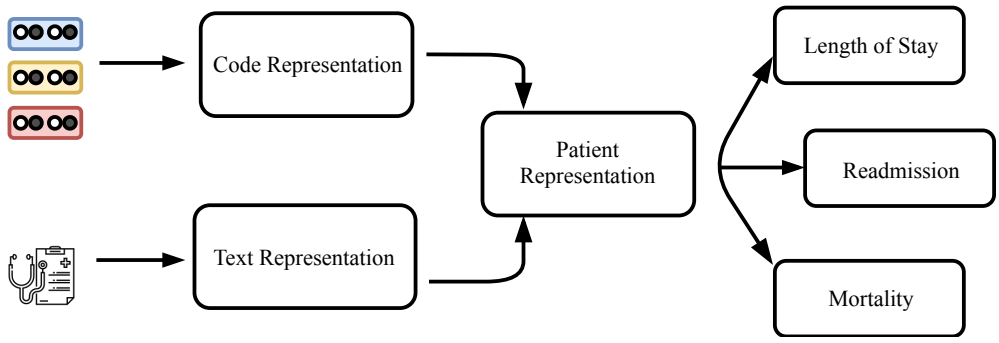


Figure 5.2: Overview of the method used to obtain patient visit representation.

length of stay is predicted. Limiting the note context window is done as medical text could include information on the date patient has been discharged.

5.2.4 Code prediction task

In this task, clinical codes are predicted for new admissions of patients given past clinical codes and historical patient data. The predicted vector is high-dimensional equal to the size of unique codes.

5.3 Methods

Our work focused on patient embedding leveraging unstructured text as well as medical codes. This representation is then fed to the classifier for predictive analytics tasks such as mortality, length of stay, and readmission (Figure 5.2). We split the training into two main components, (1) Skip-gram model using transformer networks to learn medical code representation, (2) a BERT model is trained on medical notes, and the resulting representations at a time step are summarized using auto-encoder architectures [DCL18b, VLB08, KDM19]. The final representation of a patient results from a concatenation of these two. We discuss the approach in more detail in the following subsections.

To present the problem setting, the sequence of EHR data under consideration consists of

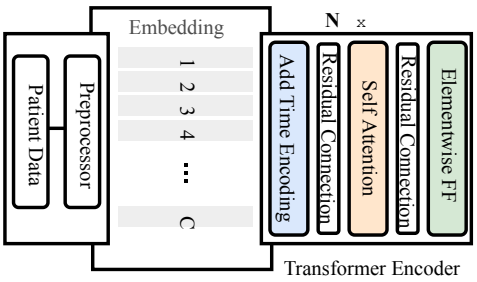


Figure 5.3: The code representation module is a transformer encoder, which takes as input patient clinical codes. The embedding matrix is a $\Re^{d \times C}$ matrix. Clinical codes are embedded using the embedding matrix, which is then passed to the transformer encoder block.

a finite set of medical concepts $\mathcal{C} = \mathcal{M} \cup \mathcal{D} \cup \mathcal{P}$, where $\mathcal{M}$ is the set of medication codes, $\mathcal{D}$ is the set of diagnosis codes, and $\mathcal{P}$ is the set of procedure codes. The codes for a visit are simply concatenated as multi-hot vectors to form the visit code vector. Accompanying the codes are medical notes $\mathcal{T}$. We denote a patient's longitudinal data as $\mathcal{D}^T = \{(c_0, t_0), \cdots (c_T, t_T)\}$ with T visits where c_i and t_i correspond to the codes and text assigned respectively within the same visit window.

5.3.1 Medical Code Embedding

We use the skip-gram model to learn code embeddings as it is able to capture relationships and co-occurrence between codes. We briefly review the skip-gram model presented by *Mikolov et al.*[MSC13]. Given a sequence of codes $\{c_1, c_2, \ldots, c_T\}$, where each code vector is a binary vector $c_t \in \{0,1\}^{|\mathcal{C}|}$, the model is tasked to predict the neighboring codes given a code c_t. The objective can be written as

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{-w \leq j \leq w, j \neq 0} c_{t+j}^{\top}\log(\hat{c}_t) + (1 - c_{t+j}^{\top})\log(1 - \hat{c}_t) \tag{5.1}$$

Here w is the context window defined as patient visits in our setting, and the softmax function is used to model the distribution $p(c_{t+j}|c_t)$. We use multiple transformer encoder

layers with self-attention mechanism as the model for $\hat{c}_t$. This model takes as input a set of medical code sequences $S = \{c_0, \cdots c_T\}$ by stacking code vectors into a matrix $K \in \{0,1\}^{T \times |\mathcal{C}|}$. The resulting set of codes is then converted into a set of embedding codes $e_{c_t} \in \Re^d$ using an embedding matrix $W \in \Re^{|\mathcal{C}| \times d}$. The embedding for the set of codes c_t at visit t is obtained as $e_{c_t} = W^T c_t$.

As the model does not contain any recurrence or convolution, we need to inject information about the relative positioning of each embedding to enable the model to make use of the ordering. This is done by adding to each embedding position a sinusoidal with frequency as a function of its timestamp t as suggested by the original transformer network. This signal acts as positional-dependent information that the model could use to incorporate time. The model is summarized in Fig. 5.3. We stack multiple transformer layers following on top of the embedding matrix. By the transformer layer, we mean a block containing the multi-head self-attention sub-layer followed by feed-forward and residual connections. For more details on this, refer to [VSP17] and the *tensor2tensor* library. As multi-head attention can attend to future time steps, to ensure that the model's predictions are only conditioned on past visits, that is, embedding at time step t can only attend to previous time steps $t-1, t-2 \ldots$, we mask the attention layers with a causal triangular mask. This is the same "masked attention" in the decoder component of the original transformer network. This mask is applied to the set of embedding in the encoder block to ensure causality. The self-attention is formulated as

$$E = \{e_{c_1}, e_{c_2}, \ldots e_{c_T}\} \tag{5.2}$$

$$Attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d}})\boldsymbol{V} \tag{5.3}$$

The query Q, key K, and value V are set to the sequence of embeddings E, and d is the embedding dimension. To obtain the final code representation for timestep t, the t^{th} output of the self-attention output is used. We call this code representation E_{c_t}.

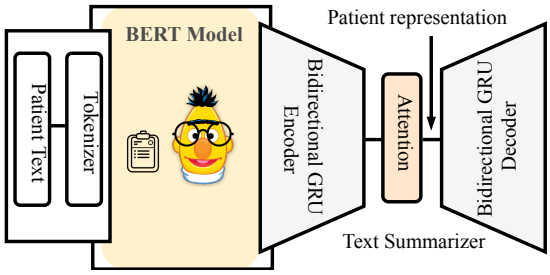


Figure 5.4: The text representation module takes as input patient text and pre-processes them into tokens, which are embedded using BERT. Subsequently, it is fed to a text summarizer autoencoder network. This work uses an LSTM-AE to learn the intermediate patient representation.

5.3.2 Medical Text Embedding

Figure 5.4 shows an overview of the text module. The embedding for the medical text sequence of a patient for time sequence $\{t_1, t_2, \ldots, t_T\}$ is obtained by using a pre-trained BERT model initialized from BioBERT [LYK19] followed by a bidirectional GRU as a text summarizer. This is done as the pre-trained BERT model has a fixed maximum sequence length of n, limiting the sequential scope of the text. Further, medical notes could get very lengthy during a visit, and they contain different types of notes, such as nurse notes, pharmacy notes, discharge notes, and so on. The aggregate length of these at a time step t could surpass the fixed-length size of n. As the aim is to obtain a single visit representation, the aggregate notes at until time step T are batched into a set of sentences $(u_1, u_2, \cdots, u_m)$, where $u_i \in \mathcal{Z}^n$ and m is the maximum occurring length in the corpus after batching each visits text into sentences of n words. The resulting set of sentences for a visit is embedded

using the BERT model, resulting in a matrix $U \in \Re^{m \times d_{BERT}}$.

$$h_{1:m} = GRU_{enc}(U_m, h_0) \tag{5.4}$$

$$E_{U_t} = \text{softmax}(\frac{\boldsymbol{h_{1:m}} \boldsymbol{h_{1:m}^T}}{\sqrt{d^{text}}}) \boldsymbol{h_{1:m}} \tag{5.5}$$

$$\hat{U} = GRU_{dec}(U_m, h_m) \tag{5.6}$$

$$L(U_{1:m,t}, \hat{U}_{1:m,t}) = \sum_{i}^{m} (u_i - \hat{u_i})^T (u_i - \hat{u_i}) \tag{5.7}$$

Subsequently, the set of sentence representations is summarized into a single patient text representation using a text-summarizer module. This module follows an auto-encoder architecture with GRUs as the building block. The input is a set of sentence representations $\in \Re^{d_{text}}$ obtained by the BERT module applied on the aggregate text until time t followed by a self-attention head on the hidden representations, where the decoder is tasked to output sentence representations $\in \Re^{d_{text}}$ following the bottleneck. The objective of the summarize, as in (5.7) is to reduce the MSE loss between the input sequence of the text embeddings and the model's predicted representation at the corresponding time steps. Finally, the patients visit text representation is obtained by summarizing the set of sentence representations $\{U_1, U_2, \cdots, U_m\}$, where the output of the attention head applied on the encoder's hidden representations is used as the text representation.

5.3.3 Patient Representation

The final patient representation Z^t at time t is obtained by concatenating the code and text representations. Additionally, the demographics d_t of the patient recorded in the visit at time t is concatenated to the resulting vector. Patient demographics contain information such as age, gender, and race, where categorical values are coded as one-hot vectors. The final representation is denoted as $Z^t = [E_{c_t}; E_{U_t}; d_t]$, where the size of this vector is the sum of the components $d_{embedding} + d_{enc} + d_{demographics}$. This representation is used for downstream tasks.

We provide specific values for each component's dimensions in our implementation details in the experiments section.

5.4 Experiments

5.4.1 Pre-processing

We extracted procedure and diagnosis codes for each patient visit. These codes are defined by the International Classification of Disease (ICD9) and medications using the National Drug Code (NDC) standard. The total number of ICD9 codes in MIMIC-III is 6984, the number of drug codes is 3389, and the number of procedure codes is 1783. Codes whose frequency is less than 5 are removed. We used the Clinical Classification Software for ICD9-CM[4] to group the ICD9 diagnosis codes into 231 categories. The Clinical Classification Software for Services and Procedures[5] was used to group the procedure codes into 704 categories. Additionally, patients of age under 18 were removed from the cohort. As medical notes contain many errors, we correct grammatical errors and remove non-alphanumerical characters. The text pre-processing processing closely follows [HAR19]. After pre-processing, 31766 patients have single visits, 6636 patients have multiple visits, the average recorded number of codes per visit is 20.52, the average number of words in medical notes is 7898, and the average number of visits per patient is 1.29. The statistics of the compiled cohort are depicted in Fig.5.5

5.4.2 Modeling and Training

The training follows the method discussed in Section 5.3. A medical concept model is trained independently on clinical codes in an unsupervised manner, and similarly, the text summarizer is trained on the text portion.

[4]https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp

[5]https://www.hcup-us.ahrq.gov/toolssoftware /ccs_svcsproc/ccssvcproc.jsp

To train the transformer encoder, we explored different values for hyperparameters, namely the number of layers, number of multi-head attention heads n_{head}, dimension for each head d_{head}, and the final model representations d_{code}. We found 2 layers perform well. We set the model's representation dimension (d_{code}) to 128. Further, the self-attention module contains 8 heads (n_{head}), each with dimension 64 (d_{head}), which is a common configuration used in transformer networks. The network is trained using Adam [KB14] with a cosine annealing schedule and with a period of 50 epochs. The initial learning rate is set to 0.00025. The window size for the skip-gram objective is set to 2.

We initialize the BERT model with the pre-trained weights on medical nodes as presented in [AMB19]. In this work, the BERT language model is initialized with BioBERT, a model trained on a large corpus of public medical data such as PubMed, medical abstracts, and so on. Then the model was fine-tuned on the MIMIC-III clinical notes.

To train the text summarizer, we use a 2-layer bidirectional GRU autoencoder with the intermediate representation set to $d_{enc} = 128$. A teacher-forcing ratio of 0.5 is used with a step learning rate schedule decay of 0.1 every 50 epochs with initial lr set to 10^{-3} [LGZ16]. We have also tried using a cosine annealing schedule, though this did not result in improvements.

Lastly, the classifier for downstream tasks is a simple 2-layer fully connected network. The first layer contains $\frac{d_{code}+d_{text}+d_{demographics}}{2}$ neurons with ReLU activation followed by a layer that maps to the number of classes in the downstream task. When training on downstream tasks, only the classifier weights are trained for 30 epochs with a step learning rate schedule decay of 0.1 every 10 epochs. This setting is used for all downstream tasks.

5.4.2.1 Implementation details

We implemented all the models with Pytorch 1.0 [PGC17]. For training the models, we use the Adam optimizer [KB14]. In all experiments, the batch size is set to 32 on a machine

equipped with 1 NVIDIA 1080TI CUDA 9.0, 32GB Memory & 8 CPU cores.

5.4.3 Evaluation Metrics

The compiled cohort consists of unique keys for each patient used to create the test/train split. This is done using k-fold with $k = 7$ on the patient keys, in which the unsupervised models are trained on the train portion and validated on the test ($\approx 15\%$ of total data). The output for a particular time step is evaluated using the patient representation Z_t at time t.

5.4.3.1 Area under the precision-recall (AU-PR)

this metric is the cumulative area under the curve by plotting precision and recall while varying the outputs $P(y_t = 1|Z_t)$ true/false threshold from 0 to 1.

5.4.3.2 Receiver operating characteristic curve (AU-ROC)

this metric is the area under the plot of the true positive rate against false positive rate while varying outputs $P(y_t = 1|Z_t)$ true/false threshold from 0 to 1.

5.4.4 Baselines

We compare our model with the following baselines

- **Med2Vec** [CBS16]:

  A multi-layer perceptron is trained on medical codes using the skip-gram objective function on a visit basis. An additional loss term is used for the co-occurrence of codes within the same visit as a regularization. The resulting output is a set of code representations in $\Re^d$.

- **ClinicalBERT** [HAR19]:

  A BERT model is pre-trained on public medical data, which is then fine-tuned on

clinical text. Following this pre-training, a BERT classifier is initialized with pre-trained weights and further fine-tuned on downstream tasks. The input to this network is text.

- **Time-aware Embedding (MCE)** [CGN18b]:

  A multi-layer perceptron is trained on medical codes with an additional attention layer to take into account the temporal context of medical codes. The resulting model is trained using either skip-gram/CBOW.

- **Patient2Vec** [ZKH18]:

  In this work, a sequence of medical codes is embedded using the word2vec model. The sequence of visits with irregular time intervals is then binned into a set of subsequences with standard intervals. Subsequently, the embedded vectors are stacked into a matrix where convolution stacked with GRU and attention models are applied to obtain the final patient representation.

- **Joint-Skipgram**[BCE18]:

  The embeddings are trained using both text and code as the vocabulary. In addition to the traditional skip-gram loss, i.e., codes in the same visit predicting surrounding codes or text predicting surrounding text, the skip-gram objective is modified such that text in a visit predicts codes in the same visit and vice versa.

- **Deepr**[NTW16]:

  A set of clinical codes are embedded using the skip-gram model. As visits contain multiple codes, the vectors corresponding to each code are stacked into a matrix, then the set of matrices for each visit is fed to a convolutional neural network and max-pooling layers to extract the final patient representation.

- $Sg_{code} + Sg_{text}$:

  Embeddings for both code and text are learned using the skip-gram objective inde-

pendently. Subsequently, a patient representation is obtained for downstream tasks by concatenating the code and text embeddings.

- **Supervised**:
  The BERT model and summarizer take as input raw text and transformer model raw codes, which are trained jointly on downstream tasks without pre-training.

We do not compare with more traditional text embeddings, such as the bag of words (BOW), as other works have shown the benefits of using BERT as text representation in NLP tasks.

We study the effect of different components presented by adding/removing text/code/demographics representation to our final patient visit representation.

Table 5.1: Diagnosis & Procedure code recall comparison with other methods.

Method	**Diagnosis Recall@k**				**Procedure Recall@k**			
	k=10	k=20	k=30	k=40	k=10	k=20	k=30	k=40
Ours	$47.92\%_{\pm(0.6)}$	$65.11\%_{\pm(0.4)}$	$75.25\%_{\pm(0.3)}$	$82.00\%_{\pm(0.4)}$	$50.99\%_{\pm(1.2)}$	$62.25\%_{\pm(0.9)}$	$67.97\%_{\pm(1.1)}$	$71.52\%_{\pm(1.0)}$
Joint-Skipgram [BCE18]	$41.16\%_{\pm(0.6)}$	$59.49\%_{\pm(0.7)}$	$71.26\%_{\pm(0.7)}$	$79.73\%_{\pm(0.5)}$	$50.88\%_{\pm(0.9)}$	$61.33\%_{\pm(0.7)}$	$67.32\%_{\pm(0.9)}$	$70.95\%_{\pm(0.7)}$
Med2Vec [CBS16]	$42.24\%_{\pm(0.2)}$	$60.18\%_{\pm(0.3)}$	$70.89\%_{\pm(0.4)}$	$78.42\%_{\pm(0.4)}$	$47.44\%_{\pm(0.7)}$	$58.10\%_{\pm(0.6)}$	$64.19\%_{\pm(0.6)}$	$68.36\%_{\pm(0.7)}$
MCE [CGN18b]	$45.70\%_{\pm(0.4)}$	$63.43\%_{\pm(0.4)}$	$74.44\%_{\pm(0.4)}$	$81.14\%_{\pm(0.4)}$	$51.87\%_{\pm(0.8)}$	$61.62\%_{\pm(1.0)}$	$68.42\%_{\pm(0.8)}$	$71.97\%_{\pm(0.8)}$
Deepr/P2V [NTW16, ZKH18]	$35.30\%_{\pm(0.3)}$	$52.40\%_{\pm(0.4)}$	$65.04\%_{\pm(0.4)}$	$74.14\%_{\pm(0.3)}$	$42.47\%_{\pm(0.7)}$	$55.59\%_{\pm(1.1)}$	$62.42\%_{\pm(0.8)}$	$66.49\%_{\pm(0.9)}$
Skipgram	$35.23\%_{\pm(0.4)}$	$52.30\%_{\pm(0.4)}$	$65.08\%_{\pm(0.4)}$	$74.13\%_{\pm(0.2)}$	$42.56\%_{\pm(0.7)}$	$55.81\%_{\pm(1.0)}$	$62.61\%_{\pm(0.9)}$	$66.70\%_{\pm(0.9)}$

To show the expressiveness of our representation, we evaluate its performance compared to baseline methods on downstream tasks and unsupervised learning tasks presented in the experiment section. We present the results obtained by embedding both text and code as patient representation on three downstream tasks: (1) 30-day readmission, (2) mortality, and (3) length of stay (LOS). Note that in MIMIC-III database clinical codes are entered into the database upon discharge of a patient, as consequently the presented results may

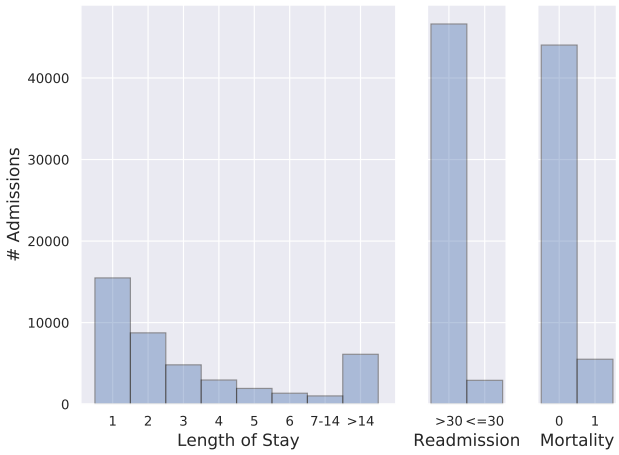


Figure 5.5: Statistics of the compiled cohort for both length of stay and readmission downstream tasks.

not be immediately clinically actionable in this case. Although, this may not be the case in other datasets where codes are updated throughout a visit. To this end the codes on the same visit are not fed to the model, but rather the previous set of codes are used.

5.4.5 Code pre-training

The encoder network is trained on clinical concepts and as most patients have 3 hospital visits after filtering to atleast 2 visits the window size for the loss term is set to 2. The performance of our network is compared with baseline methods using recall@k. This metric is evaluated by computing

$$\text{recall@}k = \frac{\text{\# of relevant codes in top k}}{\text{\# number of relevant code}}$$

This metric mimics a practitioner's method of arriving at a diagnosis or prescribing medications where they generally have several sets of candidates as a presumed cause for the underlying condition of the patient. We provide recall performance on different portions of

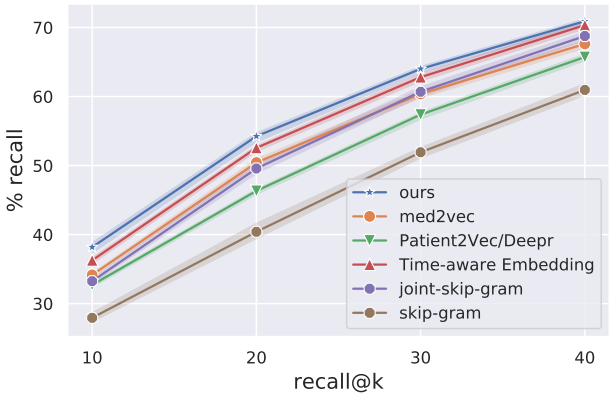


Figure 5.6: Performance of different embedding schemes on next visit code prediction using recall@k.

the code vector for both diagnosis and procedure codes in table 5.1, as well as the overall recall performance evaluated on the complete set of codes in Fig. 5.6. All baselines are fine-tuned on the same corpora by exploring different architectural hyper-parameters, except the embedding size which is fixed to $d_{code} = 128$ for all models. Note that the skip-gram baseline depicted in these figures is the same model used for the $Sg_{code} + Sg_{text}$ baseline. Additionally, the supervised baseline has the same capacity as "ours" in the figures as a result this is not included as well as it would result in the same performance. Lastly, ClinicalBERT is not evaluated on this task as it does not use codes as inputs. From the results, the time-aware code representations outperform other baselines.

5.4.6 Ablation Study

To evaluate the different components of the proposed method we conduct an ablation study on the inclusion/exclusion of the components in the final representation for downstream tasks. The complete representation using demographics, text representation and clinical code representation is the concatenation of these on a visit $Z^t = [E_{c_t}; E_{U_t}; d_t]$.

5.4.6.1 Readmission

To better evaluate the effect of the different embedding components, we run an ablation study. The results are reported in table 5.2.

Table 5.2: Downstream Tasks: Readmission

Method	AUC-ROC	PR-AUC
Text+Code+Demo	**67.42%**	**68.03%**
Text+Code	65.74%	65.43%
Text+Demo	61.44%	62.81%
Code+Demo	64.53%	67.68%
Text	56.44%	56.55%
Code	60.74%	57.89%
Demo	54.76%	59.01%

From the results, it can be seen both text and code are informative for classifying readmission. The complete combination of text, code, demographics outperforms others.

5.4.6.2 Mortality

Mortality task is concerned with predicting whether a patient will pass away within a predefined window. We predict mortality on a visit basis i.e. does the patient pass away in the current visit to the ICU. An ablation is done in table 5.3. Similar, to previous ablations, the combination of text, code, and demographics outperforms other combinations.

Table 5.3: Downstream Tasks: Mortality

Model	AUC-ROC	PR-AUC
Text+Code+Demo	**63.42%**	**65.65%**
Text+Code	59.75%	60.33%
Text+Demo	61.54%	64.07%
Code+Demo	60.41%	64.41%
Text	57.14%	57.72%
Code	56.91%	53.71%
Demo	60.10%	62.02%

5.4.6.3 Length of Stay

In general length of stay is a much more challenging task compared to readmission binary task. In this task the network is trained on balanced class data split and tested on imbalanced data, this is done as the majority of classes are discharged within 24h-48h. As shown in table 5.4, the combination of text, code and demo outperforms others in top-1 prediction accuracy. We conclude in this set of experiments, solely using text and diagnosis codes is not predictive enough to forecast the length of stay of patients.

5.4.6.4 Comparison with Other Work

We ran all baseline models on three downstream tasks using the same pre-processing steps. The final results are reported in table 5.5. To compare our method, we use the combination of text, code, and demographics. As demographics have proven to have predictive power for

Table 5.4: Downstream Tasks: Length of Stay Top-1 prediction accuracy

Method	Top-1
Text+Code+Demo	**25.57%**
Text+Code	23.38%
Text+Demo	21.10%
Code+Demo	22.22%
Text	18.82%
Code	20.54%
Demo	20.13%

the downstream tasks, other methods are augmented to use this as input/concatenated to the output representation before prediction. From Table. 5.5 the presented method outperforms others by a 1-2% margin, on all tasks demonstrating the usefulness of unsupervised pre-training.

5.4.7 Calibration

The reliability of a model's confidence is critical in health settings and calibration plots are a common measure for the model's reliability [AJH18, CAT16, JOK11]. Generally, as deep networks become complex they may not be calibrated. To this end, we provide calibration plots of our model for both readmission and mortality tasks in Fig. 5.7. To generate this plot the fraction of positives are binned according to classifiers mean predicted value for the positives. From this figure, the model is slightly over confident at lower prediction values

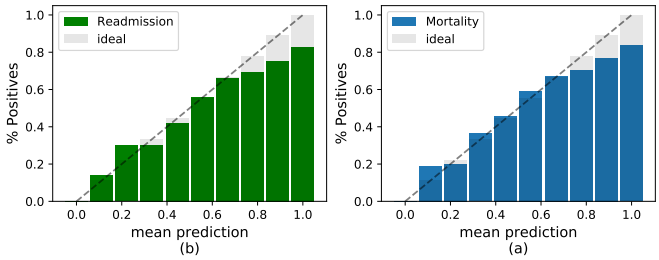


Figure 5.7: Calibration plots for mortality (a) and readmission (b) binary tasks.

and under confident at higher values.

5.5 Conclusion

Effective representation learning for EHR data is an essential step to improving care. We study embedding medical codes and notes into a unified vector representation for downstream task prediction. The presented method effectively takes the temporal context of these two data streams and provides a patient visit representation. The proposed method was evaluated on three tasks, namely, readmission, mortality, and length of stay, outperforming other methods. An ablation study was also done, showing the usefulness of both text and code when modeling patient visits. Future work could focus on adding additional data streams to the pipeline by taking into account real-time vitals and measurements taken from a patient as they undergo care.

Table 5.5: Comparison with other work on downstream tasks.

Method	**Readmisison**		**Mortality**		**LOS**
	ROC	PR-AUC	ROC	PR-AUC	Top-1
Ours	67.42%±(0.021)	68.03%±(0.015)	63.42%±(0.006)	65.65%±(0.009)	25.57%±(0.027)
ClinicalBert [HAR19]	64.41%±(0.023)	67.71%±(0.013)	61.23%±(0.019)	64.83%±(0.011)	22.65%±(0.016)
Joint-Skipgram* [BCE18]	64.59%±(0.029)	65.27%±(0.027)	61.73%±(0.016)	65.03%±(0.011)	23.27%±(0.016)
SG_{code} + SG_{text}*	65.79%±(0.019)	66.72%±(0.006)	62.23%±(0.006)	64.77%±(0.011)	24.54%±(0.004)
MCE* [CGN18b]	64.61%±(0.031)	63.96%±(0.021)	62.72%±(0.031)	64.78%±(0.022)	22.47%±(0.023)
Deepr* [NTW16]	62.63%±(0.019)	62.37%±(0.015)	60.49%±(0.012)	61.45%±(0.015)	22.81%±(0.005)
Patient2Vec * [ZKH18]	56.13%±(0.033)	65.12%±(0.011)	61.19%±(0.014)	62.51%±(0.016)	21.46%±(0.012)
Med2Vec [CBS16]	63.44%±(0.028)	62.77%±(0.022)	61.67%±(0.016)	63.22%±(0.08)	24.15%±(0.004)
Ours - Supervised	65.38%±(0.034)	66.07%±(0.024)	62.26%±(0.017)	64.32%±(0.023)	25.12%±(0.032)

1. Methods marked * is our best-effort re-implementation due to unavailability of source.

CHAPTER 6

Public Health Crises and Pandemic Analytics[1]

6.1 Introduction

In the early days of 2020, the world faced another widespread pandemic, this time the COVID-19 strand, otherwise known as the novel coronavirus. The family of Coronaviruses to which this RNA virus belongs can cause respiratory tract infections of various severities. These infections range from cases of the common cold to more lethal degrees. Many confirmed cases and deaths reported due to COVID-19 showed evidence of severe forms of the aforementioned infections [KAC12, FZS19, LCC19].

The rapid spread of this virus led to many lives being lost and extremely overwhelmed healthcare providers. It also led to worldwide difficulties and had considerable negative economic impacts. In addition, the negative impacts that it likely had on mental health due to prolonged shutdowns and quarantines is another important matter [DP].

---

[1]**This chapter is based on the following papers [FMS21, FZO21, FAH21]:**

- **Fazeli S, Moatamed B, Sarrafzadeh M. Statistical analytics and regional representation learning for covid-19 pandemic understanding. In2021 IEEE 9th International Conference on Healthcare Informatics (ICHI) 2021 Aug 9 (pp. 248-257). IEEE.**
- **Fazeli S, Zamanzadeh D, Ovalle A, Nguyen T, Gee G, Sarrafzadeh M. COVID-19 and Big Data: Multi-faceted Analysis for Spatio-temporal Understanding of the Pandemic with Social Media Conversations. arXiv preprint arXiv:2104.10807. 2021 Apr 22.**
- **Ford CL, Amani B, Harawa NT, Akee R, Gee GC, Sarrafzadeh M, Abotsi-Kowu C, Fazeli S, Le C, Nwankwo E, Zamanzadeh D. Adequacy of existing surveillance systems to monitor racism, social stigma and COVID inequities: a detailed assessment and recommendations. International journal of environmental research and public health. 2021 Dec 12;18(24):13099.**

In this work, we have gathered, processed, and combined a wide variety of well-known publicly available datasets on the COVID-19 outbreak in the United States. The idea is to provide a reliable source of statistically crucial information derived from a wide range of sources on important features describing a region and its population from various perspectives. These features primarily concern demographics, socio-economic, and public health aspects of the US regions. They are chosen in this manner because it is plausible to assume they can be potential indicators of commonalities between the affected areas. Even though demonstrating causality is not the objective of the work presented in this chapter, our analyses attempt to shed light on these possible commonalities that allow public health researchers to obtain a better perspective on the nature of this pandemic and the potential factors contributing to a slower outbreak. This is vitally important as the critical role of proper policies enforced at the proper time is evident after the COVID outbreak more than ever.

There has been widespread attention to designing and utilizing Artificial Intelligence-based tools to understand this pandemic better. Accordingly, we present a neural architecture with recurrent neural networks in its core to allow the machine to learn to predict pandemic events in the near future, given a short window of historical information on static and dynamic regional features. The main assumption that this work attempts to empirically validate is that the concise pandemic-related region-based representations can be learned and leveraged to obtain accurate outbreak event prediction with only minimal use of the historical information related to the outbreak. Aside from the theoretical importance, an essential application of this framework is when the reported historical pandemic information, e.g., the number of cases, is not reliable. An example of this is when a region discovers a problem in its reporting scheme that makes the historical information on the pandemic inaccurate due to overestimation or underestimation. Such unreliability will severely affect the models which rely on long-term historical information on the pandemic outcomes at the core of their analysis.

In summary, the contributions of this work are as follows:

- Gathering and providing a thorough collection of datasets for the fine-grained representation of US counties as sub-regions. This collection includes data from various US bureaus, health organizations, the Center for Disease Control and Prevention, and COVID-19 epidemic information.
- Evaluation of the informativeness of individual features in distinguishing between different regions
- Correlation analyses and investigating monotonic and non-monotonic relationships between several key features and the pandemic outcomes
- Proposing a neural architecture for accurate short-term predictive modeling of the COVID-19 pandemic with minimal use of historical data by leveraging the automatically learned region representations

We also designed an interactive platform with various monitoring and analytics functionalities aligned with the main components of this work. It allowed both expert researchers as well as users with little or no scientific background to study outbreak events and explore regional characteristics[2].

6.2 Data

This study focuses on analyzing the regions of the United States with statistical and AI-based approaches to obtain results and representations associated with their pandemic-related behavior. A primary and essential step in doing so is to prepare a dataset covering a wide range of information topics, from socio-economic to regional mobility reports. More details

[2]The codes, platform, data, and further analytics will be available at https://github.com/shayanfazeli/olivia.

regarding the primary data sources from which we have obtained information for this work's dataset are elaborated upon hereunder.

6.2.1 COVID-19 Daily Information per County

Our first step towards the mentioned objective is to gather the daily COVID-19 outbreak data. This data should include the number of cases that are confirmed to be caused by the novel coronavirus and its associated death toll. We are using the publicly accessible dataset API in [covb, YSH20] to fetch the relevant data records. The table of data obtained using this API contains the numerical information along with dates corresponding to each record, and each document includes the number of confirmed cases and the number of deaths that occurred due to COVID-19 on that date. It also includes the number of recoveries from COVID-19 in the same format. This dataset's significance is that it provides us with a detailed and high-resolution temporal trajectory of the COVID-19 outbreak in different urban regions across the United States. Using the dates, one can constitute a set of time-series for every county and monitor the outbreak along with the other metadata to make relevant inferences.

6.2.2 US Census Demographic Data

The US Census Demographic Data gathered by the US Census Bureau [kagf] plays a critical role in our analysis by providing us with necessary information on each region's population. Additionally, this information includes specific features such as the types of work people in that region mainly take part in, their income levels, and other invaluable demographical and social information.

6.2.3 US County-level Mortality

The fluctuations in the mortality rate of a region is also a potential critical feature in pandemic analytics. The US county-level mortality dataset was incorporated into our collection to add the high-resolution mortality rate time-series throughout the years [kagj, kagg]. The age-standardized mortality rates provide us with information on variables, the values of which can be considered as the effects of specific causes. It is crucial since some of these causes might have contributed to the faster spread of COVID-19 in different regions [DBS16].

6.2.4 US County-Level Diversity Index

Another dataset that offers a race-based breakdown of the county populations is available at [kagb] with the diversity index values corresponding to the notion of ecological entropy. For a particular region, if K races comprise its population, the value of diversity index can be computed using the following formula:

$$d_i = 1 - \sum_{i=1}^{K} (\frac{n_i}{N})^2$$

In the above formula, N is the total population and n_i is the number of people from race i. This formula represents the probability p, which means that if we randomly pick two persons from this cohort, they are of different races with probability p. In addition to that, we have the percentages of different races in the regional population as well.

6.2.5 US Droughts by County

Another source of valuable information regarding the land area and water resources per county is the data gathered by the US drought monitor [kagh, kage]. This data is incorporated into our collection as well.

6.2.6 Election

Based on the 2016 US Presidential Election, a breakdown of county populations' tendencies to vote for the main political parties is available [ele]. These records are added to our collection as the democratic-republican breakdown of regional voters can reflect socio-economic and demographical features that form the underlying reasons for the regional voting tendencies.

6.2.7 ICU Beds

Since COVID-19 imposes significant problems in terms of the extensive use of ICU beds and medical resources such as mechanical ventilators, having access to the number of ICU beds in each county is helpful. This information offers a glance at the medical care capacity of each region and its potential to provide care for the patients in ICUs [kagc]. It could be argued that having knowledge of the ICU-related capacity of regional healthcare providers can, to some extent, represent the amount of their COVID-19 related resources, such as ventilators and other needed resources.

6.2.8 US Household Income Statistics

The aggregate dataset on central statistical values on the US household income per county (including average, median, and standard deviation) is used to provide information on the financial well-being of the affected regions' occupants [kagi].

6.2.9 COVID-19 Hospitalizations and Influenza Activity Level

Aside from the socio-economical and demographical features of a region, the number of active and potential COVID-19 cases is a critical factor. This information can be leveraged to provide a possible threat level for the region. These records are made available by CDC for specific areas and are incorporated into our collection as well [CDCb, CDCa].

6.2.10 Google Mobility Reports

The COVID-19 virus is highly contagious. Therefore, the self-quarantine and social distancing measures are principal effective methodologies in bolstering the prevention efforts. Our collection includes Google's mobility reports obtained from [KWS20]. These records elaborate on the mobility levels across US regions, which are broken down into the following categories of mobility:

1. Retail and Recreation
2. Grocery and Pharmacy
3. Parks
4. Transit Stations
5. Workplaces
6. Residential

In addition, we have computed a compliance measure that has to do with the overall compliance with the shelter at home criteria:

$$\text{compliance} = -1 - \frac{(1/6)\sum_{i=1}^{6} m_i - 100}{100.0}$$

In the above formula, m_i is the mobility report for the ith mobility category. This value is computed through time to provide an overall measure of mobility through time. The compliance measures of $+1$ and -1 mean $+100\%$ and -100% changes from the baseline mobility behavior, respectively.

6.2.11 Food Businesses

Restaurants and food businesses are affected severely by the economic impacts of this outbreak. At the same time, they have not ceased to provide services that are essential and required by many. To reach a proper perspective of the food business in each region, we have prepared another dataset based on records in [Ass] to provide statistics on regional restaurant revenue and employment. Analysis of restaurants' status is important in the sense that they are mostly public places that host large gatherings, and in the time of a pandemic, their role is critical.

6.2.12 Physical Activity and Life Expectancy

Various features have been selected from the dataset in [hea] to reflect on the obesity and physical activity representation for different US regions. These features include the last prevalence survey and the changes in patterns. Also, Life Expectancy related features are valuable information for representing each region. They are included as well in our analyses.

6.2.13 Diabetes

Different features to represent a region according to the diabetes-related characteristics were selected from the data in [hea]. These include age-standardized features and clusters that have to do with diabetes-related diagnoses.

6.2.14 Drinking Habits

Information on regional drinking habits from 2005-2012 has also been used in this work [hea]. This information includes the proportions of different categories of drinkers clustered by sex and age. The categories are as follows:

- "Any": a minimum of one drink of any alcoholic beverage per 30 days

- "Heavy": a minimum average of one drink per day for women and two drinks for men per 30 days
- "Binge": a minimum of four drinks for women and five drinks for men on a single occasion at least once per 30 days

6.3 Methods

In what follows, the analytical techniques that we have designed and used in this work are explained. To draw meaning from the data that we have at hand, we have designed and utilized a variety of techniques. These methodologies range from traditional statistical methodologies to the design and testing of deep learning inference pipelines for event prediction. We select a set of representative features to use in our analytics from the gathered collection of datasets. More details on the nature of these features are shown in Table 6.1.[3]

6.3.0.1 Feature Informativeness for Sub-region Representation

An important question that is raised in analyzing a dataset with well-defined categories of features is how important these features are in describing the entities associated with them. From the particular perspective of enabling the differentiation between two regions, it can be said that a measure of importance is the contribution of each one of these selected features to the overall variation in datapoints. The boundary case is that if a feature always has the same value, it is not informative as there is no entropy value associated with its distribution. To begin with, we associate a mathematical vector with each data point, which contains the values of all its dynamic and static features associated with a specific date and location. Since we are mainly targeting US counties in this study, each record would be associated with

[3]Please note that the focus of this work, in terms of regions, is on US counties. Nevertheless, if some features are only available in the state-level resolution, they will also be used in representing the region as they are associated with county characteristics as well.

Table 6.1: Overview of the Features Characterizing Regions

Category	**Description**
Food Businesses (static)	Food and Beverage Locations Restaurant Employments Sale and Economy
Gender (static)	Percentage of Male and Female
Race (static)	Ratio of different races
Election (static)	Ratio of Democratic, Republican, and other voters
Income (static)	Wage Statistics Poverty Information
Commute (static)	Statistics of Methods of Commute to Work and Their Ratio
Hospitals and Mortality (static)	Information on ICU Capacity and Statistics on Region's Mortality
Obesity and Physical Activity (static)	Information on the Statistics of Obesity and Physical Activity and the Changes in Patterns
Life Expectancy (static)	Regional Life Expectancy Values in Years
Drinking (static)	Alcohol Consumption Patterns and Changes
Diabetes (static)	Patterns of Different Types of Diabetes Diagnoses and Changes in Them
Land and Water (static)	Information on Land and Water Resources of Regions
Employment (static)	Ratio of Different Job Types and Other Statistics
CDC Hospitalizations and Surveys (dynamic)	Number of Hospitalizations due to COVID-19 and Influenza Activity Surveys
Google Mobility Reports (dynamic)	Breakdown of Regional Mobility in Different Categories Based on Which Our Compliance Score Is Computed

a US county at a specific date. We then use Linear Principal Component Analysis (PCA) [WEG87] to reduce the dimensionality of these data points and to evaluate the importance of the selected features in terms of their contribution to the overall variation. Results show that in order to retain over 98% of the original variance, a minimum of 55 principal components should be considered. Each one of these components is found as a linear combination of the original set of features, and that along with the percentage of variance along the axis of that component can be used as a measure of performance. To be more specific, considering n features and m data points that result in p PCA components to retain 98% of the variation, we will have:

$$\vec{c_i} = \langle v_1, v_2, \cdots, v_n \rangle \in \mathbb{R}^n$$

And u_i is the total variance along the axis of ith PCA component. This can be thought of as a measure of importance for the PCA components, and the absolute value of v_is magnitudes can be considered as the importance of original feature i's contribution to its making. Therefore, we will have the following measure of informativeness defined for our features:

$$\vec{I} \in \mathbb{R}^n$$
$$\vec{I} = \sum_{j=1}^{p} u_j \cdot \vec{c_j}$$

The features can be sorted according to these values, and the categories can also be considered in their relevant importance. Note that this is just one definition of informativeness; for example, certain features might not vary a lot, but when they do, they are potentially associated with severe changes in the COVID-19 events. Therefore, the importance score that has been captured here merely has to do with how better we are able to distinguish between locations based on a feature.

6.3.0.2 Statistical Analytics

In order to better understand the co-occurrences of the features in our input dataset and their corresponding COVID-19 related events, we have performed an in-depth correlation analysis on them. We have considered four principal measures of correlation, namely: Pearson, Kendall, Histogram Intersection, and Spearman, as described in Table 6.2. We have used the Pearson correlation coefficient along with the p-values to shed light on the presence or absence of a significant relationship between the values of each specific feature and each category of pandemic outcome. We have also computed nonparametric Spearman rank correlation coefficients between any two of our random variables. This value would be computed as the Pearson measure of the raw values converted to their ranks. The formulation is shown in Table 6.2 in which d_i is the difference in paired ranks. We also employed mutual information as an additional measure and confirmed the presence of relationships between the key region representation features pointed out by the correlation analyses and the pandemic outcomes. This coefficient measures the strength of the association between the values of these random variables in terms of their ranks. Since many of the relationships in our dataset can be intuitively thought of as monotonic, these values are particularly important. To better understand the concordance and discordance, Kendall correlation is computed as well. The main correlation analysis formulations used to evaluate the monotonic and general relationship between region representation variables and pandemic outcomes are shown in Table 6.2. In these formulations, m_1 and m_2 are the numbers of concordant and discordant pairs of values, respectively. Additionally, Normalized Histogram Intersection is another methodology directly targeting the distributions of these variables. The degree of their overlap represents how closely x's distribution follows the distribution of y. It has also been utilized in finding the results of this section.

Table 6.2: Correlation Analysis Formulas

Correlation	**Formula**
Pearson	$r_{x,y} = \frac{\sum_{i=1}^{m}(x_i - \mu_x) \cdot (y_i - \mu_y)}{\sqrt{(\sum_{i=1}^{m}(x_i - \mu_x)^2) \cdot (\sum_{i=1}^{m}(y_i - \mu_y)^2)}}$
Spearman	$s_{x,y} = 1 - \frac{6 \sum d_i^2}{m(m^2 - 1)}$
Kendall	$k_{x,y} = \frac{m_1 - m_2}{\binom{m}{2}}$

6.3.1 Neural Event Prediction

In continuation of our statistical analyses on COVID-19 event distributions, we have designed a neural inference pipeline to help with the effective utilization of both learned deep representations and the embedded sequential information in the dataset.

In this work, we introduce a neural architecture, which is trained and used for COVID-19 event prediction across the US regions. The Double Window Long Short Term Memory COVID-19 Predictor (DWLSTM-CP) is comprised of multiple components for domain mapping and deep processing. First, using its dynamic projection which is a fully connected layer, the dynamic feature vectors which reflect on temporal dynamics will be mapped to a new space and represented with a further concise mathematical vector.

$$\langle \tilde{x}_1^{\text{dynamic}}, \tilde{x}_2^{\text{dynamic}}, \cdots, \tilde{x}_T^{\text{dynamic}} \rangle = F_{\text{projection}}^{\text{dynamic}}(\langle x_1^{\text{dynamic}}, x_2^{\text{dynamic}}, \cdots, x_T^{\text{dynamic}} \rangle) \tag{6.1}$$

This step is essential due to the fact that an optimal deep inference pipeline is the one that retains only the information required by each level and minimizes redundancies [TZ15]. The projections are designed to help the network achieve this objective. These are then fed to the LSTM core for processing.

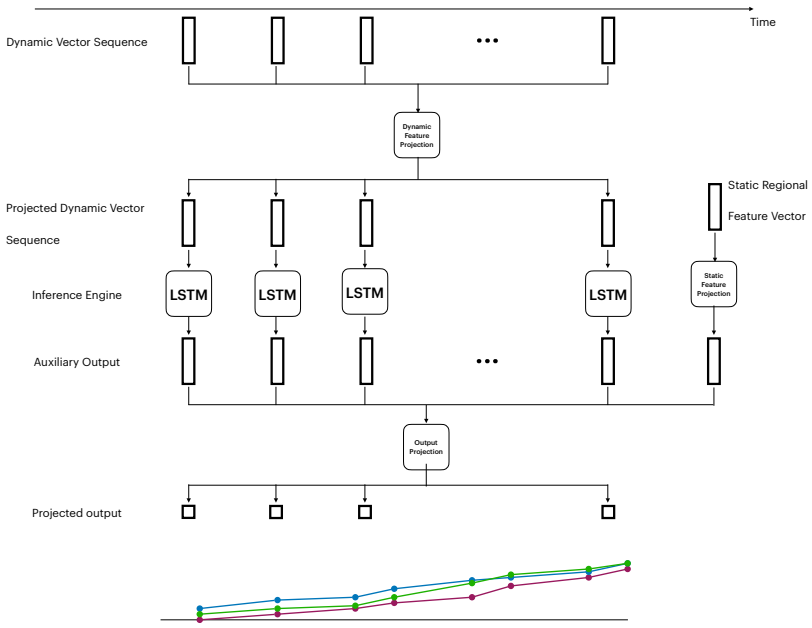


Figure 6.1: The full inference pipeline of our Double Window LSTM-based COVID-19 event prediction is shown in this block diagram.

$$A^{\text{dynamic}} = \text{LSTM}(\tilde{X})$$

Each one of these outputs is concatenated with the projected version of static features, $F^{\text{static}}_{\text{projection}}(x^{\text{static}})$, and fed to the output regression unit. The outputs are compared with the ground truth time-series, and a weighted Mean Squared Error loss along with Norm-based regularization is used to guide the training process while encouraging more focus on the points with large values. The overall pipeline is shown in Figure 6.1.

It is worth mentioning that this approach leverages and utilizes all of the features discussed in the previous sections in representation of regions, as well as a minimal historical window of pandemic outcomes. It learns representations that take various factors, from different categories of mobility and activities to socio-economic information, to make accurate short-term predictions while reducing the need for lengthy historical data on the pandemic

outcomes. There are many occasions in which accurate and reliable historical data on the pandemic is not available due to a variety of reasons (e.g., a problem in reporting scheme), which motivates approaches with less dependency on it.

6.4 Experiments

The results on our regional dataset in terms of feature importance from the principal component analysis indicate the following features contribute to the overall concise representation of a region significantly:

- Restaurant businesses, namely the contribution to the state economy and the count of food and beverage locations. Even though we only have access to state-level data, its importance can be intuitively argued as it reflects on the counties that the state includes. It is also worth noting that restaurants and food services play an important role in the pandemic dynamics, as the restaurant employees are considered essential workers.

- The influenza activity level is another critical feature in the analysis. Given the similarity of symptoms between Influenza and COVID-19 infection, monitoring Influenza activity is very helpful for COVID-19 pandemic understanding.

- Diversity index, which signifies the probability of two randomly selected persons belonging to different races from a population, also plays a crucial role in representing the regions.

- The changes in the mortality rate that is not associated with COVID-19 are beneficial as well. This is also intuitively arguable as it can be thought of as a measure of mortality related sensitivity for the regions.

Figure 6.2 shows how the projected points scatter after the PCA as well. The results indicate that 55 PCA components are required to retain over 98% of the variance of the

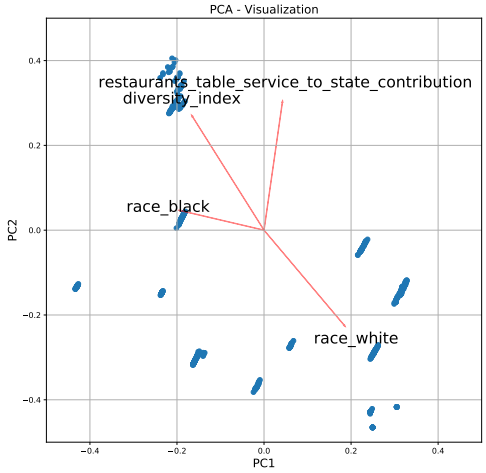


Figure 6.2: The plot in this figure is a PCA BiPlot which shows the variations of the first two PCA components and axes of some of the selected features.

dataset, and Figure 6.3 shows the progress of covering the variance by adding each one of them sorted by their importance. Table 6.3 provides the values of the aforementioned importance metric computed for sample features in different importance levels.

6.4.1 Statistical Analytics

The results of correlation analyses help empirically and quantitatively validate many of the relationships mentioned in the known hypotheses regarding the COVID-19 outbreak. The Pearson correlation of -0.287 with the p-value of 0.046 indicates a significant relationship between the percentage of food businesses in the state economy, and the average cumulative death count in its counties. Another example is the value of the Spearman correlation coefficients between the different types of commute to work associated with each county and the values of the pandemic-related events. From Table 6.4, it is apparent that there is a positive relationship between the proportion of public transit as a method of commute to work and the spread of COVID-19 in the region. It indicates that the more the percentage

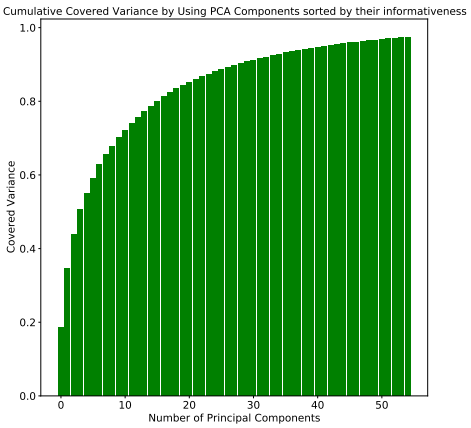


Figure 6.3: The cumulative amount of variance covered by using up to a certain number of PCA components. This is assuming that they are sorted by their corresponding eigenvalue, meaning that the first component contributes more to variance coverage than the ones selected after it.

Table 6.3: Sample Features of High and Low Informativeness Score

Level	**Feature**	**Score**
	Diversity Index	0.148
High	Contribution of Restaurants' Table Service to State Economy	0.130
	African American Ratio	0.109
	Percentage of Men	0.020
Low	Pacific Islanders Ratio	0.013
	Percentage of Family Jobs	0.006

Table 6.4: The Spearman correlation coefficients between the share of different methods of commute in county transportation and the cumulative pandemic outcomes

	Cumulative Deaths	Cumulative Cases	Cumulative Recoveries
Drive	0.22	0.20	−0.03
Carpool	−0.04	0.04	0.04
Transit	0.20	0.12	−0.06
Walk	−0.29	−0.35	0.05

of public transit is for the method of commute to work in a US county, the more the number of potential cases is expected to be as the Spearman correlation coefficient is an indicator of a monotonic relationship between the variables. Another example is the Pearson correlation between the ratio of different races in regions and the pandemic outcomes. It is known that COVID-19 is affecting the African American community disproportionately [Sco]. Accordingly, Table 6.5 shows a higher correlation between the ratio of African Americans and the severity of COVID-19 outcomes. It also indicates that the more diverse regions were impacted the most, which is in accordance with the findings of the feature importance section listing diversity-related features as critical parameters.

6.4.2 Neural Event Prediction

The collected set of datasets in this work provide a sufficient number of records for enabling the efficient use of Artificial Intelligence for Spatio-temporal representation learning. We show this by training instances of our proposed DoubleWindowLSTM architecture on the two main short-term tasks regarding epidemic modeling; namely, new daily death and case count. In our dataset, we considered the US COVID-19 information from March 1st, 2020 to July 22nd, 2020, in which the July data is used for our evaluations, and the rest are leveraged for training and cross-validation. The objective using which the proposed architecture was

Table 6.5: Pearson correlation between the race percentages per county and COVID-19 variables

	Cumulative Deaths	Cumulative Cases	Cumulative Recoveries
White	−0.30	−0.48	−0.15
African-American	0.34	0.42	−0.01
Hispanic	0.04	0.23	0.2
Native American	0.03	0.02	0.03
Asian	0.14	0.11	0.04
Pacific Islander	−0.03	−0.02	0.03

trained is a multi-step weighted Mean Squared Error (MSE) loss, which helps to minimize a notion of distance between the predictions and the target ground-truth while encouraging (by assigning larger weights) to the windows that exhibit larger values. These thresholds are empirically tuned and set prior to the training procedure. The learning curves for both experiments indicate clear convergence in Figure 6.4.

To quantitatively evaluate the performance, we have reported the Root Mean Square Error (RMSE) for the prediction of new daily deaths and cases due to COVID-19 in Table 6.6. For comparison, we have used the ARIMA model as well with the parameters set according to the work in [Kuf20] that have fine-tuned this scheme for forecasting the dynamics of COVID-19 cases in Europe. We have also found the empirically optimal ARIMA model in each scenario according to Augmented Dickey-Fuller (ADF) tests and based on Akaike information criterion (AIC) and reported the results denoted by ARIMA*. The evaluation is conducted on the test set, which includes the data from the end of June 2020 to July 22nd, 2020.

To compare with other predictive modeling works in this area, which are mostly focusing on making inference for *US states*, we aggregated our county-level findings to form esti-

Table 6.6: Daily Average RMSE Evaluations

Objective and Timeframe	**New Daily Deaths**		**New Daily Cases**	
	10-day	15-day	10-day	15-day
DWLSTM	4.4347	3.0435	81.4205	92.4027
ARIMA*	29.9813	12.7631	233.3008	235.3828
ARIMA(1,2,0)	57.1886	22.4285	394.3747	566.5686

mators for state-level prediction. From the results reported in Table 6.7, it is interesting to observe that the aggregated estimator based on our model achieves strong evaluation result comparable to the models that achieve highest scores, while clearly outperforming the other two models that are inherently county-level, namely, the works in [WWG20] and [PS20a]. The baseline methodologies are evaluated on predicting the next 14 days and until June 28th, 2020. The groundtruth pandemic outcomes come from different datasets on pandemic events, namely, Johns Hopkins University (JHU) [YSH20, covb], New York Times dataset (NYT) [nyt], and the US Facts dataset (USF) [usf]. It should also be noted that even though the objective for the DWLSTM model was to predict county-level information, the provided state-level errors which are obtained by aggregation fall in the range of the dominant COVID-19 predictor models that rely heavily on the accuracy of the historical epidemic data. The evaluation for our model has been conducted on the next 14 days during July 2020, which is also particularly challenging due to drastic changes in the pattern of outbreak trajectories in states such as California and Texas.

The predictions for several regions exhibiting different levels of severity are shown in Figure 6.5. These results can help the reader in a qualitative assessment of the model performances, in which the outputs of our approach demonstrate high stability and follow the trajectory of the ground-truth with precision.

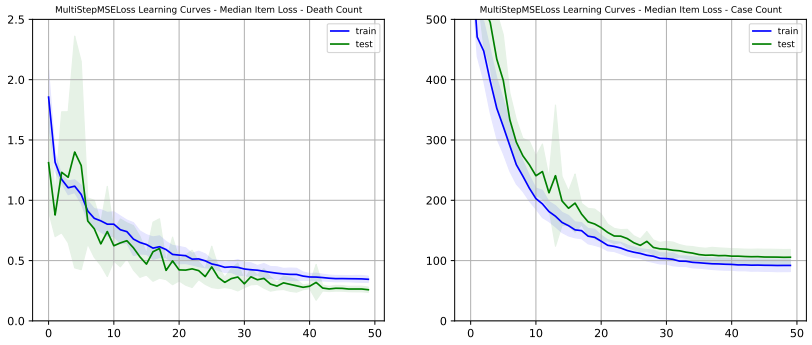


Figure 6.4: Sample Test Prediction of Cumulative Death Count per 100k Population - Four regions exhibiting different severity levels are chosen to show the efficacy of the model. The 95% confidence intervals for ARIMA* and DWLSTM models are shown and clearly indicate the stability in training our model and the predictions made by it.

Table 6.7: Comparison with Published Models

	Prediction Window (days)	**Average Daily RMSE**	**Ground-truth Source**
DWLSTM	**14**	**24.94**	**JHU**
SIkJa	14	23.63	JHU
UCLA SuEIR	14	22.97	NYT
CovidActNow SEIR CAN	14	27.78	NYT
IowaStateLW. STEM	14	26.67	JHU
Covid19Sim Simulator	14	27.82	JHU
JHU IDD CovidSP	14	48.97	USF
CU Select	14	32.36	USF

6.5 Discussion

6.5.1 Principal Findings

The primary objective of this work is focused on leveraging regional representations for accurate short-term predictive modeling of the epidemic with minimal use of historical data. It is plausible to assume that the features chosen in this work, which reflect on different characteristics of a region, include valuable information for efficient prediction of pandemic events. The static features include various socio-economic and demographical properties associated with a region and its population. Combined with the dynamic set of features such as influenza activity level and mobility patterns, this information was leveraged along with a short track of pandemic time-series for predictive modeling. We do not claim that the data points coming from this domain are statistically sufficient for the pandemic event prediction tasks; however, empirical results indicate that they can be effectively utilized for these objectives. There are occurrences outside of this domain that can impact the outcomes (e.g., the initial impact of a large number of infected people arriving in a specific location is not initially captured by our scheme). Nevertheless, the results indicate that the data points coming solely from this work's domain can help in the effective knowledge extraction regarding the current and future values of pandemic-related time-series. The result section elaborated on the statistical findings and introduced a measure of feature importance. In addition, a neural network architecture that has a long short-term memory configured recurrent neural network in its core was introduced to serve as a new baseline for COVID-19 event prediction.

6.5.2 Comparison with Previous Studies

Since the beginning of the COVID-19 outbreak, there have been works focusing on gathering information or performing statistical analysis related to this epidemic. This work is focused on learning and analysis of the high-resolution spatiotemporal representation of urban areas.

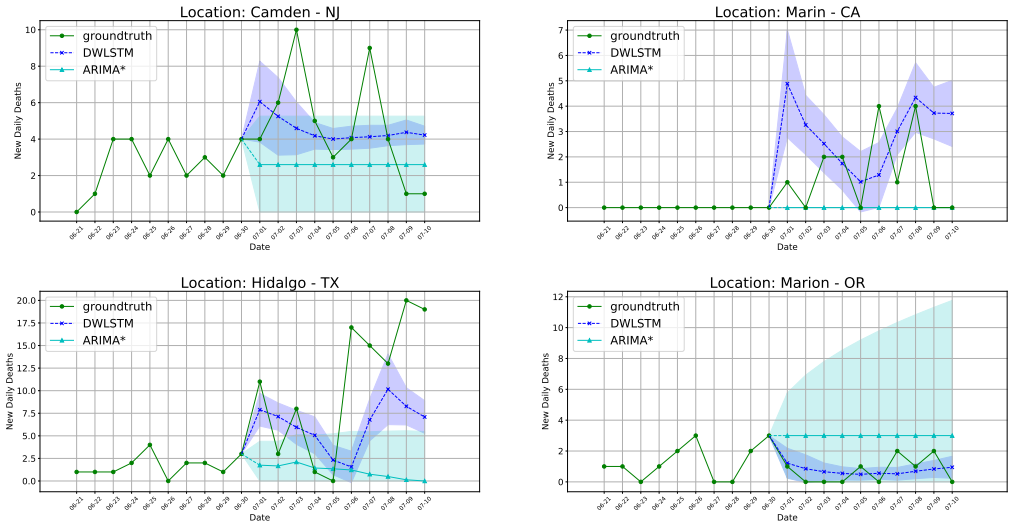


Figure 6.5: Sample Test Prediction of Cumulative Death Count per 100k Population - Four regions exhibiting different severity levels are chosen to show the efficacy of the model. The 95% confidence intervals for ARIMA* and DWLSTM models are shown and clearly indicate the stability in training our model and the predictions made by it.

We provide a collection of datasets and select a large number of features to reflect on various demographics, socio-economics, mobility, and pandemic information. We have used statistical analysis techniques to investigate the relationships between individual features and the epidemic, while also considering the contribution of such features to the overall representation power. We have also proposed a deep learning framework to validate this idea that such region-based representations can be leveraged to obtain accurate predictions of the epidemic trajectories while using but a minimal amount of historical data on the outbreak events (e.g., number of cases). Even though our model is trained with the objective of providing county-level predictions, we have aggregated these county-level predictions and used these aggregate values as state-level estimators to evaluate the loss on the most recent data. In Table 6, we have compared these results with the information on the similar performance measure of the eight COVID-19 prediction works that perform state-level inference making. It can be seen that our framework provides a simple solution which outperforms the other county-level methodologies (namely, [WWG20] and [PS20a]) on this task.

6.5.3 Applications

The importance of clearly defined policies enforced at the proper time on alleviating the adverse impacts of a pandemic in different areas is crystal clear. One of the important applications of this work is in providing researchers and agencies with a more in-depth understanding of the co-occurrence of idiosyncratic patterns associated with regions and the predicted pattern of the outbreak. This information can be used to assist policymakers, for example, to render the details of their decisions such as lockdowns, more fine-grained and attuned to the regional needs. These include the intensity and length of enforcing such measures. The ability to predict pandemic-related occurrences (e.g., number of deaths, cases, and recoveries) is another valuable application of this work. This knowledge will provide hospitals and healthcare facilities with targeted information to help with the efficient allocation of their resources. Another important application of this work is when there is a

lack of availability for accurate and reliable historical data on the epidemic events, which is also important for developing countries. For example, when it is realized that the previous reports on the number of cases and deaths due to the pandemic were not reliable, such finding will not affect our solution due to its less degree of dependence on the historical data on the epidemic than other models which base their analysis on them at the core of their analyses.

6.5.4 Limitations

This study has several limitations that should be discussed. The initial notion of feature informativeness which was discussed in the earlier sections of this article mainly has to do with the contribution of features to the variance in representing regions and areas. Given the nature of this study, combining this and the relationship between them and the pandemic and providing more in-depth prior domain knowledge can help with a better definition of feature importance. Our methodology provides a means to use region-based representations to obtain predictions with less reliance on the historical epidemic data. Nevertheless, generalizing the network architecture in this work and providing access to more extended and reliable historical data, if possible, can be an improvement and is worthwhile as a potential future direction. Utilizing attention-based methodologies and other interpretation techniques with the pre-trained weights is also a well-suited future direction to better understand what the models learn.

6.5.5 Conclusions

In this study, we gathered a collection of datasets on a wide range of features associated with US regions. Our approach then used various statistical techniques and machine learning to measure the relationship between these regional representations and the pandemic time-series events and perform predictive modeling with minimal use of historical data on the epidemic.

Both quantitative and qualitative evaluations were used in assessing the efficacy of our design, which renders it suitable for applications in various areas related to pandemic understanding and control. This is crucial since the information on the patterns and predictions related to an outbreak play a critical role in elaborate preparations for the pandemic, such as improving the allocation of resources in healthcare systems that will otherwise be overwhelmed by an unexpected number of cases.

CHAPTER 7

Non-intrusive Monitoring of Mental Well-being[1]

7.1 Introduction

Within the spectrum of mental health disorders, disorders related to stress and anxiety are some of the most common groups of psychiatric problems affecting both children and adults [BHG18, Car06, MHB10], with up to one in three people in the US meeting full diagnostic criteria by early adulthood [TEW17, HHP08]. This manifests in the form of roughly to 7 to 9% of the population in the US suffering from a specific phobia, 7% from social anxiety disorder, and 2 to 3% each from panic disorder, agoraphobia, generalized anxiety disorder, and separation anxiety disorder [APA19]. In the United States alone, estimates are that 20% of the population suffers from at least one mental health condition [Leo21]. Around 60 million adults are affected by at least one anxiety-related disorder (e.g., generalized anxiety disorder (GAD) and social anxiety), with less than 40% of them seeking treatment [Men, APA]. In the United States, about 73% of people suffer from acute bouts of stress to a degree of magnitude that impacts their mental well-being. Incidents of Anxiety

[1]**This chapter is based on the following papers [FLB23, FLB22]:**

- **Fazeli S, Levine L, Beikzadeh M, Mirzasoleiman B, Zadeh B, Peris T, Sarrafzadeh M. A Self-supervised Framework for Improved Data-Driven Monitoring of Stress via Multimodal Passive Sensing. arXiv preprint arXiv:2303.14267. 2023 Mar 24. (To be published in IEEE International Conference on Digital Health (ICDH) 2023)**
- **Fazeli S, Levine L, Beikzadeh M, Mirzasoleiman B, Zadeh B, Peris T, Sarrafzadeh M. Passive monitoring of physiological precursors of stress leveraging smartwatch data. In2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) 2022 Dec 6 (pp. 2893-2899). IEEE.**

often manifest similarly to stress. However, it is notable that it is not always immediately tied to a specific triggering or inciting event and may take longer to resolve. All told, both stress and anxiety problems are very common, to the extent that most adults have been affected by at least one anxiety-related disorder [APA]. Individuals with anxiety disorders contend with substantial distress and impairment. They are at heightened risk for a host of adverse long-term outcomes, including depression, substance abuse, educational underachievement, and poor physical health [NMW14, BMC98, WF01, Cli]. Effectively managing mental health-related disorders is a growing challenge to the healthcare community. In addition to the significant costs and considerable resources required to provide proper help for patients, limited access to care renders seeking treatment further challenging for the individuals in need of it [JJD01, CWD04, SWM06, SW09]. The optimal method for the prevention or care of mental illness is early identification, diagnosis, and proactive treatment[WKW16]. Time-sensitive intervention is, therefore, crucial for preventing conditions from becoming chronic and debilitating. However, traditional methods of psychiatric assessment, including clinical interviews and self-reports, are limited in their ability to provide just-in-time interventions as well as early identification. They depend heavily on retrospective summaries collected in clinical settings, conditions often resulting in reporting biases, inaccurate recall, or late and ineffectual treatment.

Additionally, stress and anxiety disorders are, for the most part, vastly overlooked and under-treated in the community; only about 15-30% of anxious individuals receive treatment of any kind. Recent research has found strikingly high levels of anxiety among college-age youth. Indeed, 58.4% of college-aged youth report feeling "overwhelmed by anxiety" [Ass16]. Several other recent studies document the high proportion of college students meeting full diagnostic criteria for an anxiety disorder [BMK18]. At the same time, young adults are mainly overlooked within the health care system, with screening, identification, and referral rates falling below those of children or adults [WF01]. Given this landscape, there remains a pressing need for tools that improve early identification of anxiety symptoms, provide users

with platforms to monitor their activities, and raise awareness of factors impacting their well-being.

The growing ubiquity of consumer devices, among them smartphones, smartwatches, and in-home sensors, all equipped with an array of sensors and user logs, have resulted in an unprecedented opportunity to catalog and quantify the daily aspects of an individual's life, creating repositories of personalized information [Swa12]. If harnessed and utilized by the individuals themselves, there is significant potential for such monitoring to improve their healthcare outcomes dramatically. This potential has long been recognized in physical behavior and physiological health, as both are extensively tracked. In contrast, the mental health domain is considerably less investigated.

7.1.1 Long-term Patterns of Anxiety

There have been research efforts that explored whether pervasive mental health monitoring could be feasible through a smartphone and the embedded sensors, such as motion sensors, ambient light, microphone, camera, Global Positioning System (GPS), proximity, and touch screen[CFL17][OPC14][BSW15][NGT20]. These efforts have shown the promise of this approach in successfully tying behavioral monitoring to mental health; however, such approaches have not translated into fully mature frameworks and have focused almost exclusively on depression-related conditions, which, while often spoken in conjunction with anxiety, manifest in distinct ways[FIO19].

The advantages of leveraging a smartphone-based platform are that the continuous collection of quantitative data potentially provides a more reliable indicator of an individual's risk at any given time, as well as offering a mechanism for just-in-time intervention should a mental health episode occur[BSW15]. Conversely, smartphone-derived data present several challenges, some of which have already been noted, which can result in limited accuracy owing to differences in behavioral patterns across users and the indirect manner of detection[FIO19].

Towards the recognition and monitoring of anxiety, we propose a smartphone-centered platform named eWellness. The eWellness framework is designed to capture a broad spectrum of remote monitoring, survey data acquisition, secure data transmission and management, data analytics, and visualization. The primary component of eWellness is a mobile application that facilitates data collection and transmission harvested from an array of sensors and usage logs from a user's smartphone. The data is collected passively, pre-processed, and transmitted through a secure gateway to the cloud, where it is securely stored and indexed using a scalable database.

Concurrently the eWellness application includes an active querying component where users can be prompted with Ecological Momentary Assessments (EMA) of their mental health status. A back-end analytic engine complements this architecture, capable of mapping observed metrics and exogenous data sources to a user's mental health state based on adaptive statistical models and machine learning algorithms.

7.1.2 Short-term Patterns of Stress

Within the spectrum of personal electronics, smartwatches offer unique potential in the realm of remote health monitoring [A 18]. While smartphones are the platform most typically associated with mobile healthcare applications, smartwatches are worn continuously and directly on the skin, enabling a host of physiological sensing modalities that smartphones cannot emulate. The watch's location, worn at a distal point of a major appendage (the user's wrist), also typically results in a more accurate reading of a user's activities than smartphones, whose location in relation to the user's body may vary throughout the day [Tex17]. Embedded smartwatch sensors increasingly rival the capabilities and sophistication of dedicated wearable sensors such as chest straps. Additionally, smartwatches have advantages over dedicated sensors in that their interactive features provide a means for engaging the user with additional queries and therapeutic responses [J 17]. Finally, the generalized nature of smartwatches increases the likelihood that users will utilize their healthcare fea-

tures in addition to the broader set of desirable functions that users find in smartwatches. This is much like how a user is likely to be more inclined to use a healthcare application on a smartphone rather than purchase and use a custom device solely for the same intended purpose.

As previously noted, smartwatch devices provide near real-time data measurements, allow users to provide feedback via interactive queries, and enable intervention mechanisms in ordinary day-to-day situations outside clinical environments. Furthermore, this platform aggregates and analyzes patient data that can be communicated to healthcare providers and family members to improve their treatment experience. This study, therefore, focuses on presenting a smartwatch-based system for recognizing and monitoring physiological patterns preceding episodes of short-term stressful responses.

While smart devices, like smartwatches, have demonstrated the technical capacity to monitor physiological stress to varying degrees of accuracy, they demonstrate only a partial picture of a wearer's overall mental health.

This stems from the intrinsic disconnect between emotional perceptions of stress and the underlying physiological stress response. There is growing evidence of notable differences in how individuals perceive their stress levels and the physiological manifestation of a stress response. For instance, a meta-analysis found that a significant correlation between physiological stress and perceived emotional stress was found in only 25% of social stress studies[CE12].

Several factors are presumed to contribute to this low correlation. Foremost among them is that, somewhat counter-intuitively, a direct linear relationship between these two stress profiles is not present but instead is influenced by many additional factors. A detailed study concluded that self-reports of perceived stress did not provide useful information about physiological stress responses[OOB11]. Differing populations may also exhibit different correspondence patterns that population-level models fail to account for (e.g., individuals with ADHD, those who experience chronic stress, and family histories of substance abuse).[CE12]

It is also theorized that differing response patterns may impact how stress manifests physiologically. For instance, differing degrees of cognitive regulation of negative emotions may result in individuals being able to adapt to stress differently[KMC15]. Previous work demonstrating a higher correspondence between self-reported anxiety and physiological arousal in women when compared to men is hypothesized to be explained by men showing emotion-suppressing coping strategies to a greater extent than women[KMC15]. Finally, it has been observed that positive emotions are more strongly associated with physiological responses than negative emotions. This may be due to their social acceptability and disinclination to control or dampen a physiological response when compared to negative ones[KMC15].

Perhaps the most challenging effort in effectively mapping these two stress profiles stems from the fact that accurate and consistent measurement of emotional stress experiences is currently considered an intractable challenge for the academic community[OOB11]. While multiple scales have been developed to assess differing quantifiable mental health measures, there is no clinically validated gold standard for assessing emotional well-being accurately and consistently, both across a population and individuals over time. This stems from issues with an individual's recall and perception of stress, particularly if data is collected after the fact, their evolving willingness to be forthright and transparent regarding emotional well-being, and differences in the subjective assessment of stress across individuals (what constitutes a *moderate* level of stress may differ dramatically across individuals)[OOB11].

In this work, we present an end-to-end wearable stress recognition and monitoring framework that provides the opportunity for continuous, scalable, and low-cost monitoring of relevant mental health indicators over time. Our contributions are as follows:

- Focusing on users' usual anxiety responses during everyday life events, we propose a smartwatch-based system for non-invasive and efficient monitoring of key physiological attributes, including heart rate, heart rate variability, respiration, and pulse oximetry readings. We then aim to leverage this information to make accurate predictions on the occurrence of short-term episodes that individuals would perceive as stressful.

- In a 10-day anxiety study approved by the Internal Review Board (IRB)[2], we created a corpus of primary physiological data from monitoring 14 college students with an active-duty or ex-military affiliation. This helps us study the data obtained from a critical cohort regarding susceptibility to stress and anxiety.

- We designed a survey-based self-assessment method for the subjects to receive feedback regarding *time* and *intensity* of the moments and episodes they perceived as stressful throughout the day. These assessments correspond to the stressful episodes in their everyday lives and their perceived levels of experienced anxiety. Labeling these self-reported episodes was done with the help of domain experts, which rendered our corpus an informative benchmark for empirical performance evaluations.

- We designed deep inference pipelines for the fusion and processing of our smartwatch data [FL]. Our framework enables the efficient use of artificial intelligence for stress-focused deep representation learning. This is done by pre-processing preparation and fusion of the multi-modal physiological time-series data recorded via smartwatch sensors as well as leveraging the details of self-reported stressful events.

- We experimented with fusing sensory data at various pipeline stages and presented comparative results.

- To make the most of our data which is limited in size and number of annotations, we leveraged label smoothing and augmentation in windowing and observation formation, as well as specific techniques of adversarial information removal and self-supervised pre-training and contrastive regularization so as to improve the performance.

- The proposed late-fusion method is inherently modular with regards to the different modalities of data. Therefore proper data-layer transforms allows leveraging various

[2]The Internal Review Board approved our study at the University of California, Los Angeles.

devices (e.g., smartwatches and wearable sensors different from ours) to learn efficient representations for health monitoring.

- We employed an attention-based approach providing a diagnostic view of the system, allowing the researchers to look into the empirical connection between various modes of data for specific monitoring tasks, counteracting the masking effect of many deep-learning frameworks on interpretability.
- In developing this framework, we conducted experiments on real-world data collected on perceived stress and demonstrated the efficacy of our approach.

7.2 Data

7.2.1 Long-term Anxiety

7.2.1.1 Privacy

Data from the study, both sensor feeds and usage logs, along with user-generated EMA responses, are first encrypted, cached locally on the user's device, and then transmitted to a secure remote server, where it is stored in an encrypted scalable MySQL database.

7.2.1.2 eWellness Data Collection

The eWellness mobile application, developed for Android devices, collects passive behavioral data derived from communications logs, embedded sensors, and user logs capturing the following metrics:

- **Communication**: is monitored by incoming and outgoing phone calls and text messages, including the duration of phone calls, the number of texts and phone calls, and unique individuals contacted. This does not assess the content of communications or the recipient of the communication beyond establishing a unique contact.

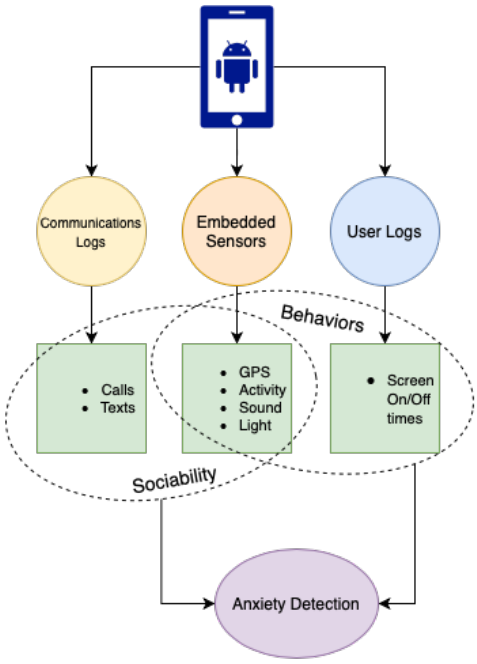


Figure 7.1: eWellness Data Collection Hierarchy

- **Location**: is periodically sampled using GPS, network, and Wi-Fi detection. Prompts for a new location after moving 5 meters, up to once a minute. This metric leverages the Google Fused Location API. The application does not track specific locations; instead, it keeps a total distance traveled using the vectorized haversine distance function.

- **Ambient Sound**: is a numeric measure designed to detect speech and communication above 50 decibels using the phone's microphone. It samples every 5 minutes for 5 seconds. This metric does not capture the audio files of communications and merely documents the sound frequency and decibel level as numeric values.

- **Activity and Movements**: leverage the device's accelerometer, gyroscope, and GPS tracking. Activity is sampled every 60 seconds. In order to determine stationary and moving activity-type, the application leverages Google's Activity Recognition API.

- **Light**: detects light level associated with possibly being in an outdoor or indoor location. This sensor is sampled every 6 seconds.

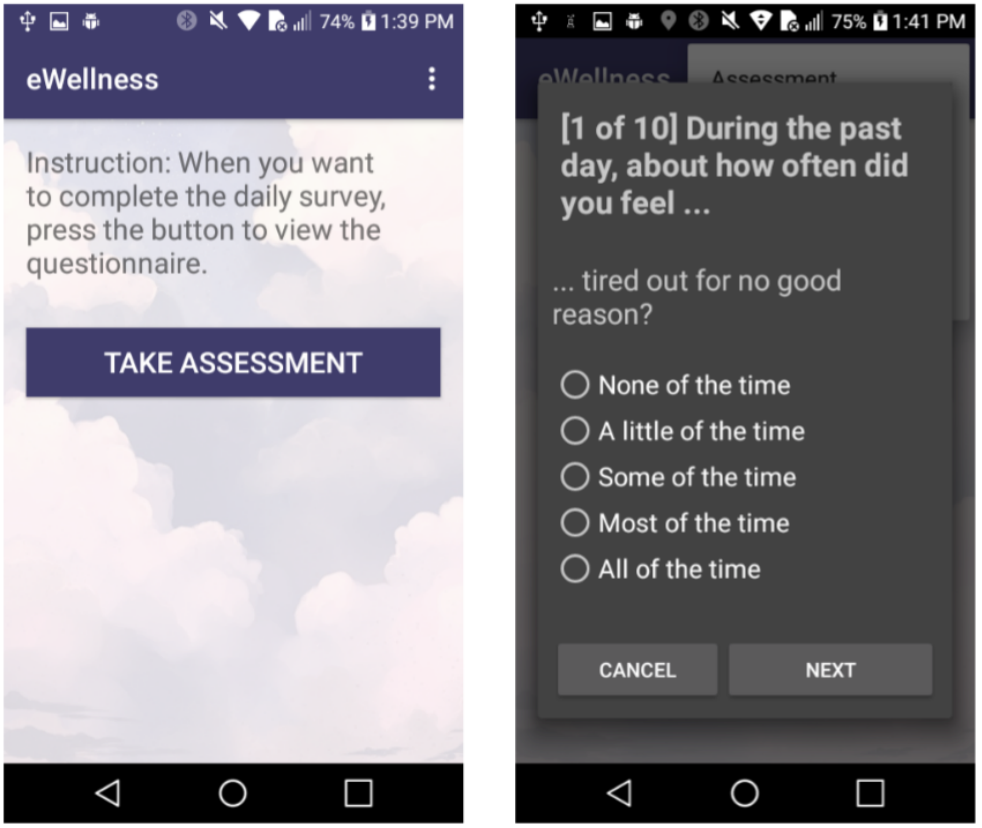
- **Phone use**: is user-log monitoring the device's screen on time.

We derived daily aggregated features from these metrics from these raw values to infer a user's sociability and behavioral patterns. These were then used to learn a model for predicting anxiety symptom severity. We obtained statistical characteristics, such as minimum, maximum, mean, standard deviation, and the 25th, 50th, and 75th percentiles, of the numeric values of noise exposure and ambient luminescence. The number of activity transitions and duration of each physical activity per day also became a significant metric for identifying mentally distressed days.

Recognizing the potentially invasive nature of applications like this, data collection was carefully scoped to avoid collecting Personally Identifiable Information (PII) that could link a particular user to a particular dataset. For example, when attempting to gauge sociability, the application logs the total number of phone calls made, total time on the phone, and the number of unique contacts called; the identities of specific callers were not tracked. This has the consequence of introducing a degree of obscurity into an observed finding (e.g., as the application cannot differentiate between calls to friends and calls to a customer-service hotline). At the same time, in the interest of both respecting privacy and ensuring the acceptability of the app, these efforts were felt to be necessary constraints on data collection.

An IRB-approved pilot study was conducted on a dozen individuals using the Android version 5.0 and above from the university community, including students and staff. Study participants did not have a reported history of mental illness. Participants were asked to download and install the eWellness application (Figure 7.2) and then run it on their phones for a month. Passive data was collected continuously by the application throughout the month. Participants were asked to answer EMA daily through the eWellness app but did not provide any other personal information, such as name, gender, or age, during participation.

The Kessler Psychological Distress Scale (K10) [AS01] is a validated measure of psychological distress over the past 30 days, used for clinical and epidemiological purposes. It has



(a) Landing page. (b) Daily EMA questionnaire.

Figure 7.2: eWellness - User Interface

notable success in measuring feelings of anxiety along with depression. For this pilot, the K10 was modified to assess criteria over the previous 24-hour period. The modified K10 prompted the users as daily EMA to measure their feelings of anxiety and depression. The K10 is composed of ten questions, structured on the following standardized template, "Over the past 24 hours, how often have you...", to which users can provide one of five standardized responses: All of the time, Most of the time, Some of the time, A little of the time, and None of the time). These responses are scored on a range from five (All of the time) through one (None of the time). The minimum possible score of K10 is 10, and the maximum possible score is 50. K10 results are categorized into four levels of psychological distress: low distress, moderate distress, high distress, and very high distress (Table 7.1).

Table 7.1: Categorization of K10 Scores [aus]

K10 Score	Level	Samples (N=146)
10-15	Low distress	91
16-21	Moderate distress	29
22-29	High distress	21
30-50	Very high distress	5

7.2.2 Short-term Stress

Considering the ultimate objective of our study, which has been to present a working solution that the general population can use for continuous and passive monitoring of bouts of stress, we concentrated on the intersection of two groups of people particularly at risk for stress, namely, college students and military members. To create our cohort of data on short-term stress, we recruited former and current members of the military within the University of California - Los Angeles (UCLA) community for participation in this study. We published our study's advertisement in social media channels and military-affiliated organizations. Interested individuals were asked to complete an initial online screening survey. Those determined to be eligible participated in a virtual informational onboarding session to discuss the study protocol and determine their willingness to complete the required components of the study. Bouts of stress in response to triggers or anxiety can happen to everyone regardless of the formal clinical diagnosis of other mental health disorders. For this reason, and because we aimed to learn universal representations for the stress recognition task and evaluate the feasibility of such a system in the absence of any patient-identifying background information, there was no need for a full clinical assessment at the onboarding phase.

7.2.2.1 Data Collection

Participants were instructed to wear the watch continuously for 10 consecutive days. We also needed to gather supervision signals to recognize perceived stress. To this end, we designed a labeling scheme to allow participants to submit self-reported details of their stressful episodes. An online survey particularly designed for this purpose was sent to them via email daily, in which they were asked to provide the following:

1. A rating of their overall perceived stress level for that day (`none`, `low`, `medium`, or `high`).

2. A recounting of any specific stressful incidents, including their description of the context and what they believe to be the cause or trigger, in addition to the approximate time or timespan at which these episodes occurred and perceived magnitude (`low`, `medium`, or `high`).

3. Any times during which the device, based on its solely HRV-based approximation of stress, triggered a high-stress indicator

4. Any times during which the watch was removed and for what reason (e.g., for charging or showering).

Our chosen smartwatch has an internal approximation of physiological stress based solely on heart-rate variability (HRV) analysis. Once this value gets past a certain threshold, the watch will issue notifications to the individual so as to focus their attention on what could be a stressful episode. As an essential intervention measure, the smartwatch then gives the user the option to perform a breathing exercise to help relieve the stress. Additionally, to ensure that sufficient records of stressful episodes were present in our data, with their informed consent, the participants were asked to undergo two remote, stress-inducing laboratory sessions under the direct guidance of study personnel. These sessions were scheduled at least one day apart and around the same time of day. Study personnel would guide the participant over a

video call through a series of three exercises designed to induce a controlled amount of stress while the participant wore the smartwatch. Each exercise has been clinically established as a method for induced stress control, invoking both cognitive, social, and physical stress. The specific exercises were a cold pressor test [BGN16], Stroop Test [MBR05], and Maastricht Acute Stress Test (MAST) [SLC17]. For the Stroop Test, participants were shown the name of a color in a colored "ink." They were asked to say aloud the ink color and press the corresponding key on their computer. For example, the word "purple" appears as a green color. Participants have to say "green" and not purple. The color names were shown one at a time for a total test time of about one minute. The cold pressor test involved the participant first submerging their hand without the watch in room temperature water (around 38°C) for two minutes, followed by submerging their hand in ice cold water (0°C- 4°C) for 2 minutes. Lastly, the MAST exercise consisted of a two-minute timed arithmetic test while they submerged their hand in ice-cold water in the same fashion as the cold pressor test. The participant's self-assessed level of stress (measured as low, medium, or high) was recorded after each test. Participants were also asked about their activities before and after the stress test session to give context for their emotional state before and after the laboratory sessions.

7.2.2.2 Pre-processing and Initial Data Analyses

Data were collected in a continuous manner over regular intervals as predetermined by the smartwatch manufacturer. For the purposes of this study, both raw sensor data, as well as a selection of features produced by the smartwatch were collected. These included the pulse oximeter/oxygen saturation, respiration rate, and heart-rate data. The watch also produced a Stress level indicator derived from the measure of heart rate variability while the individual was sufficiently inactive. Finally, the user's sleep quantity from the previous night was tallied and included.

Significant null values were observed in the data we collected. This is an acknowledged issue with wearable sensors, particularly smartwatches that are not typically affixed tightly

Table 7.2: Number of participants in our cohort per category based on their duty status and service branch

	Duty Status
National Guard	2
Active Duty	3
Veteran	9
	Service Branch
Airforce	2
Marine	3
Navy	4
Army	5

to the skin [CKZ17] when consistent skin contact is required to accurately take a reading. In attempting to address these gaps, Data was down-selected only to include intervals when complete data was available. For episodes with a low degree of missing information, we leveraged a causal interpolation scheme to utilize the last known value and complete the available episode information by that. Stressful events in the data were labeled using the daily journal surveys completed by participants, in which they specified approximate periods of high stress, the stress magnitude, and the type of stress. Owing to the nature of individuals self-reporting stress hours after the fact, the precise time intervals of these stressful events could not be ascertained from the journals. Therefore, the decision was made to time-block intervals to one-hour increments, with the stress label assigned to that hour. Additionally, the dates and times when participants were taking part in stress testing were included and processed similarly.

7.3 Methods

7.3.1 Long-term Anxiety

Leveraging the K-10 entries made by the participants, we applied thresholds and formulated a classification task. First, we select a sub-group of features based on the magnitude of their correlation with the raw K10 score. This helps, given that the limited amount of available data renders it cumbersome to deal with the resulting overfittings expected to happen with many machine learning models. We then would apply classifiers and experiment with classifying the results.

7.3.2 Short-term Stress

Data were collected in a continuous manner over regular intervals as predetermined by the smartwatch manufacturer. For the purposes of this study, both raw sensor data, as well as a selection of features produced by the smartwatch were collected. These included the pulse oximeter/oxygen saturation, respiration rate, and heart-rate data. The watch also produced a Stress level indicator derived from the measure of heart rate variability while the individual was sufficiently inactive. Finally, the user's sleep quantity from the previous night was tallied and included. Tables 7.3 and 7.2 reflect on the data modalities and prominent features as well as the breakdown of subjects by their duty status.

Significant null values were observed in the data we collected. This is an acknowledged issue with wearable sensors, particularly smartwatches that are not typically affixed tightly to the skin [CKZ17] when consistent skin contact is required to accurately take a reading. To address these gaps, data was down-selected to remove time spans exhibiting a high degree of data missingness. For episodes with a low degree of missing information, we leveraged a causal interpolation scheme to utilize the last known value and complete the available episode information by that. Stressful events in the data were labeled using the daily journal surveys

Table 7.3: Data modalities and the features they are focused on

Modality	**Focus**
Daily	heart-rate readings, number of floors climbed, BMR kilocalories, distance traveled, activity levels, aggregated HRV measures
Pulse Ox	SPO2
Respiration	Respiration rate
Stress	HRV-based readings

completed by participants, in which they specified approximate periods of high stress, the stress magnitude, and the type of stress. Owing to the nature of individuals self-reporting stress hours after the fact, the precise time intervals of these stressful events could not be ascertained from the journals. Therefore, the decision was made to time-block intervals to one-hour increments, with the stress label assigned to that hour. Additionally, the dates and times when participants were taking part in stress testing were included and processed similarly.

7.3.2.1 Label Smoothing and Window Augmentation

Given the limited availability of data for enabling efficient training of our pipelines, we needed to generate as many training examples as possible using our corpus. To this end, we considered a label smoothing strategy to allow proposing an estimate of the stress level at each point in time, leveraging a non-causal smoothened version of the events obtained

from the self-reports. For each individual, we aimed to create a continuous and real-valued *teacher function* $f(\cdot)$ enabling us to map a timestamp t (in seconds) to an estimation of the severity of stressful response.

$$f(\cdot) \leftarrow f(\cdot) + \lambda(d) \cdot \exp(-\frac{t - \mu(d)}{2\sigma(d)^2})$$

The iterative process by which we created these functions is as follows: We start from a constant function set to zero, $f(t) = 0$, which follows the intuitive assumption of *default state is non-stressed*. Note that the individuals in our study are not patients, and they were confirmed not to have any diagnosed condition affecting the prevalence of stressful episodes. Then, we sweep the recorded entries and iteratively update the teacher function by adding functional components. Consider each entry similar to the sample depicted in Figure 7.3 as a record d. Each record contains an attribute titled `probe_datetime`, which can take two types of values:

1. *timestamp*: a single specific timestamp assigned to a reported episode of stressful response. Note that in our study, the individual's report is in 15-minute units (e.g., 4:15 pm).

2. *time span*: a pair of timestamps indicating the participant's reported start and end times.

The update step involves the operation below:

$$f(\cdot) \leftarrow f(\cdot) + \lambda(d) \cdot \exp(-\frac{t - \mu(d)}{2\sigma(d)^2}) \tag{7.1}$$

The parameters associated with each update functional, namely, μ and λ, are selected to make the resulting signals qualitatively reasonable to a human reviewer. Figures 7.4 and 7.5 showcase an example data window and its corresponding teacher function.

```
{
    "subject_id": "12345",
    "stress_type": "general",
    "perceived_rate": "low",
    "stress_rate": "low",
    "stress_description": """
    I was stressed because of a deadline
    """,
    "probe_datetime": (
        datetime(
            2021, 5, 10, 9, 0, 0,
            tzinfo=timezone.utc),
        datetime(
            2021, 5, 10, 18, 0, 0,
            tzinfo=timezone.utc))
}
```

Figure 7.3: A sample stress probe datapoint, showing the record structure

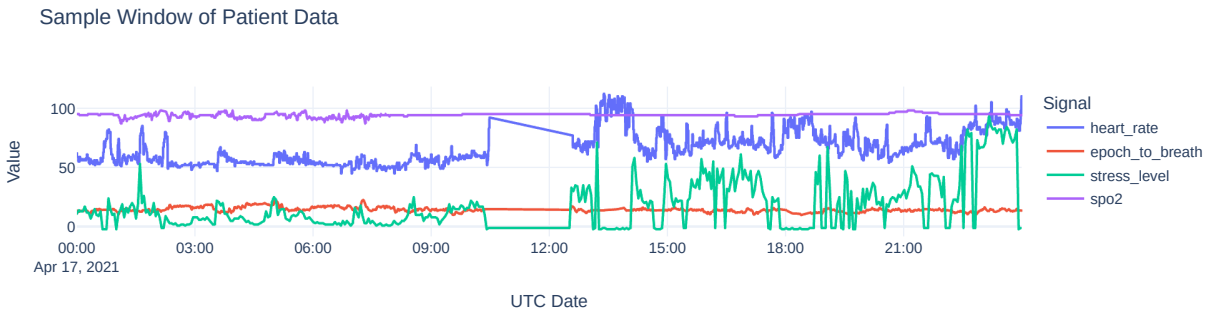


Figure 7.4: Example: A slice of patient physiological signals recorded via smartwatch system

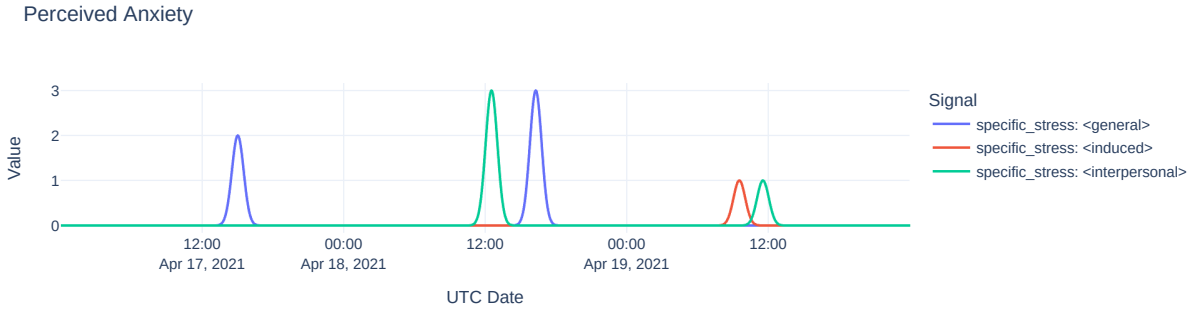


Figure 7.5: Example: Anxiety levels - Subject trajectories through time

7.3.3 Representation Learning

7.3.3.1 Event Observations

The first step towards preparing our inference pipeline is defining the *observations* in this study, or, in other words, the input to our pipeline. We define our entity, the *episode*, as the $1-hour$ episodes, and our observation is the collection of sampled data corresponding to a 1-hour sensory reading episode, with data sampled from the aforementioned modalities. The sampled records contain longitudinal information on the aforementioned data modalities (heart rate and activity aggregates, heart-rate variability, pulse ox, and respiration) throughout the span of the episode. To help with faster convergence and overall stabil-

ity, we leveraged min-max normalization over the magnitudes and learned standardizers for reshaping the data. We define the task of predicting perceived stress below:

Perceived Stress Task: Given the information provided by 1-hour observation of the user's physiological episode, determine whether or not the episode is one that ends in a stressful note, a.k.a, with t_{end} denoting the episode's ending timestamp, determine whether or not the statement $f_{\text{u}}(t_{\text{end}}) > \tau$ is true.

7.3.4 Fusion

When multiple data sources are involved in a single observation, one critical question would be the point at which information merging should occur. In this study, we offer two solutions in which we have early fusion for one representation learning pipeline and late fusion for the other.

7.3.4.1 Early-fused Pipeline

The first approach to fuse the information is the simple choice of creating a *hybrid continuous token* by synchronizing and concatenating the information from all data sources. This approach is more intuitive, however, it is expected to render the system more prone to the impact of inaccurate data.

Tackling Limited Data: In our early-fusion pipeline, given that we are working with a more complex input and a single pipeline responsible to learn the corresponding joint distribution, we designed a targetted information removal objective to reduce the overfitting and help with the more efficient utilization of this data. To this end, we leverage gradient ascent to *unlearn* the information biasing the model towards the peculiarities of data associated with individuals rather than the task. Consider a model acting as a discriminator, denoted as $D(\cdot; \psi_1)$. This model aims to leverage the latent representation resulting from processing the episode by our model and try to distinguish between different participants given their

IDs as the task. We consider our pipeline's stem as the core component of the *generator* model, denoted as $G(\cdot;\psi_2)$, which is to challenge the discriminator in making the aforementioned subject classification task more cumbersome to carry out. We implement $D(\cdot;\psi_2)$ as MLP-based multi-layer head applying on the latent representations:

$$D(\cdot;\psi_1) = f_{\text{adv}}(\cdot;\psi) \tag{7.2}$$

To build the optimization component, we simply add the loss term $\mathcal{L}_D - \mathcal{L}_G$ to the final loss, in which $\mathcal{L}_D$ and $\mathcal{L}_G$ is defined as below:

$$\mathcal{L}_G = \mathbb{E}[l_{\text{ce}}(\text{fr}(f_{\text{adv}})(z(x)), y_{\text{sub}})] \tag{7.3}$$

$$\mathcal{L}_D = \mathbb{E}[l_{\text{ce}}(f_{\text{adv}}(\text{sg}(z(x))), y_{\text{sub}})] \tag{7.4}$$

Note that in the above, $\text{fr}(\cdot)$ *freezes* the parameter set ψ and sg is the stop-gradient operator, cutting off the gradient pathway to what comes before it.

7.3.4.2 Late-fused Pipeline

While intuitive and straightforward to build, the early-fusion pipeline introduces its problems in training and inference. In that setup, we need to make assumptions on the regularity of the data sampling, no missingness[3], and the optimality of following a uniformly distributed importance weights over the input information coming from different sources. To explore an alternative solution, we propose a late-fusion pipeline in which the data from each modality is represented by a dedicated neural network pipeline and mapped to a *shared* semantic space.

The observations representing our episodic one-hour entities are easier to build compared to the early-fusion pipeline. In this context, there is no need for additional transformations

[3]In our case, we do causal interpolation to deal with that.

such as interpolation, upsampling, or synchronization. Each data point contains the information thoroughly representing the episode's time span, denoted as $e = (t_{\text{start}}, t_{\text{end}})$, with the information we have in the corpus:

$$\boldsymbol{x}_{m,e}^{(p)} = \{x_{m,t}^{(p)} \in \mathbf{x}_m^{(p)} | t \in e\} \tag{7.5}$$

This setup has a separate modality-specific dedicated encoder denoted as $f(\cdot; \theta_m)$. Given that our observation data comes from different modalities, for each modality $m \in [M]$, we have:

$$f(\cdot; \theta_m) : \mathcal{X}_m \to \mathcal{S} \quad \forall m \in [M] \tag{7.6}$$

Note that in the above equation, $\mathcal{S}$ is a shared semantic space allowing the encoders from different modalities to collaborate, teach each other, and contribute to the final learned representations. Finally, each encoder creates a representation $z_{m,e}^{(p)}$ for the corresponding data, computed as below:

$$z_{m,e}^{(p)} = f(\boldsymbol{x}_{m,e}^{(p)}; \theta_m) \quad \forall m \in [M] \tag{7.7}$$

There are numerous reasons for which it is plausible to assume that the information from each modality and each episode can contribute differently to the final output if considered by a Bayesian optimal classifier. The quality of the information made available via smartwatch sensors can vary over time (based on how the watch is worn, how subtle the experienced stressful episode was, and so on), which is a crucial matter to consider while deciding on the contribution of these modalities. We designed a pooling pipeline based on attention pooling to enable the model to make such decisions automatically and on the level of *instance* data.

$$a_i^{(m)} \leftarrow g(\boldsymbol{z}_i^{(m)}; \boldsymbol{\psi}) \quad \forall m \in [M] \tag{7.8}$$

Afterward, a softmax operation is applied to make sure that the summation of the pre-

dicted contributions is 1.0; in other words, the contribution matrix is right stochastic:

$$\alpha_i^{(m)} = \frac{\exp(a_i^{(m)})}{\sum_{j\in[M]} \exp(a_i^{(j)})} \tag{7.9}$$

The attention weights will then be used to tune the contribution of the semantics of each modality to the construction of the final aggregated representation embedding the entire episode. $z = \sum_{i=1}^{M} \alpha_i^{(m)} \cdot z_m$.

To observe the content similarity between the latent representation and the aggregated embedding, we use cosine similarity, which is a conventional choice in the contrastive learning domain.

$$\phi(\boldsymbol{u}, \boldsymbol{v}) = \frac{h(\boldsymbol{u})^T \cdot h(\boldsymbol{v})}{\|h(\boldsymbol{u})\|_2 \cdot \|h(\boldsymbol{v})\|} \tag{7.10}$$

The following objective will be involved in the final optimization, then injects the proper incentives and penalties, guiding the encoders to represent the data for each episode closer to each other and the aggregate representations. Note that our approach does not solely rely on bootstrapping but also leverages negative examples. Therefore, training can take place without the use of stop-gradient operation. Given the difficulty of providing comprehensive data transformations that can apply to episodes without guaranteeing that the semantic content will not be altered extensively, we chose this approach.

$$\mathcal{L}_{\text{cl}} = \frac{1}{|\mathcal{B}|} \sum_{i\in\mathcal{B}} \frac{1}{|\mathcal{M}|} \sum_{m\in\mathcal{M}} -\log \frac{\exp(\phi(\boldsymbol{z}_i^{(m)}, \boldsymbol{z}_i)/\tau)}{\sum_{j\in\mathcal{B}, j\neq i} \exp(\phi(\boldsymbol{z}_i^{(m)}, \boldsymbol{z}_j)/\tau)} \tag{7.11}$$

- *Pre-training*: Pre-training the model parameters by optimizing $\mathcal{L}_{\text{cl}}$ through a long training sequence. Afterward, start with the resulting weights as the initial point for the supervised fine-tuning of the model with the cross-entropy objective:

$$\mathcal{L}_{\text{cross-entropy}} = -\sum_{c\in\mathcal{C}} y_c \ln p_c \tag{7.12}$$

  In the equation above, $\mathcal{C}$ is the set of all classes (e.g., in our experiments, the two categories of stressed and non-stressed for each episode), and p_c is the predicted probability of class c for an observation, computed by passing representations through a final projection and Softmax layer.

- *Regularization*: Use $\lambda_{\text{reg}} \cdot \mathcal{L}_{\text{cl}}$ as a regularization term in the overall loss, and train the model by optimizing this loss simultaneously as the supervised learning objective.

There are several points worth remarking upon with regard to the comparison of these two training schemes. To begin with, deciding whether pre-training is going to lead to better generalization performance versus the regularization-based approach depends on model complexity, availability of data, and the challenges of the specific task that one is targeting. That being said, the regularization approach is expected to be considerably faster than the two-stage pre-training and fine-tuning method, and in our experiments on the task of predicting stress labels, it led to better test performance as well.

We have detailed the main intuitive advantages of this approach below:

Modularity: Having a separate dedicated encoder for each modality and fusion via a shared semantic space allows a *modular* inference process. Our optimization scheme's reliance on contrastive regularization focuses on optimizing the individual branch's ability to generate representations optimal for the task.

Knowledge Transfer: The encapsulated view in the early phase of our pipeline allows one to leverage knowledge transfer and employ weights and models that serve as suitable initial points for training the pipelines. In addition, different optimizers and learning rate schedulers can be used for each branch per the requirements of the fine-tuning scheme.

Plug-and-Play: The shared semantic space also has the empirical advantage of allowing researchers and developers to apply the same pipeline with new modalities added or some removed, assuming a corresponding expert model is either trained or fine-tuned to allow the new modalities to work with the shared semantic space[4].

[4]Our analyses show that the empirical contribution of data from different modalities is not uniform. Therefore, further research on larger corpora can shed more light on whether one modality can be used as the *central* data source in our observations, similar to the approach presented in [GEL23] for self-supervised learning from multiple modalities.

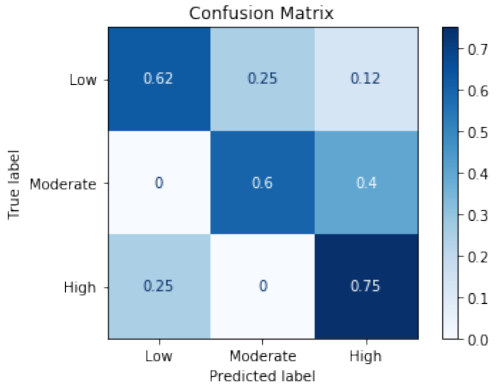


Figure 7.6: 3-class (Low, Moderate, and High distress) Classification Confusion Matrix.

7.4 Experiments

7.4.1 Long-term Anxiety

For the 4-class classification, we used 5-fold Cross-Validation (CV) with four models: K-Nearest Neighbors (KNN), Extra-Trees (ET), Support Vector Machine (SVM), and Multi-layer Perceptron (MLP). The class weight was automatically applied to the models inversely proportional to the class frequencies to train the imbalanced dataset. The highest classification accuracy achieved was around 76% with the extra-trees model. We also applied the under-sampling technique to improve the performance of an imbalanced dataset. Samples from the low distress class were removed randomly to make uniformly distributed class labels. Samples from the very high distress class were also ignored. The confusion matrix in Figure 7.6 demonstrates the performance.

7.4.1.1 Relevant Features

There are some notable and counter-intuitive findings regarding what data elements proved to be most-highly correlated to mental health. It is not surprising to note the presence of features closely related to physical activities (e.g., Duration of time spent biking or walking) as such activities have been definitively linked to mental health [Ste88].

What is somewhat less intuitive is the presence of multiple audio and light-sensing features. Audio sensing was included in the protocol under the hypothesis that a moderate level of sound could be indicative of pro-social activities like being outdoors or in group settings. Conversely, overly loud or quiet noise profiles could be indicative of stressful environments or isolated conditions that could be deleterious to mental health. But while interesting in theory, there are many confounding causes of noise that, by limiting ourselves to solely capturing the frequency and decibel levels of the sound, we would fail to distinguish (intuitively, someone watching TV at home alone could register the same noise profile as someone out to dinner with friends).

Similarly, it was hypothesized that light sensing could be indicative of an individual being outside, which has been shown to positively correlate to mental health[TDC15], however here too, many confounding factors would impact light readings, foremost among them, that the user would actually have to have their phones out and exposed when outside for the light sensor to register it.

We note that sound and light sensing is notable in that both were the most frequently sampled of all features. It is possible that the high degree of granularity of readings afforded to these particular sensing modalities explains their relevance. Regardless, we suggest additional work is needed to understand whether or not these features are indeed more universally indicative of mental health and explore why that is potentially the case.

7.4.1.2 Limitations of the Study

While 10 subjects completing one-months worth of continuous data represents a critical validation of the technology and its potential utility, obtaining a larger dataset is still important to improve the quality and robustness of the findings. Additionally, this pilot was scoped to only include individuals without a clinical diagnosis of Anxiety. Consequently, there were insufficient cases of user-reported mental distress, particularly moderate or severe cases, in order to classify them effectively. Additional studies are planned to enlarge our dataset and

include a cohort of individuals with diagnosed mental health conditions.

7.4.1.3 Accuracy of Labeling

At the time, a primary concern was that participants would fail to submit a sufficient number of survey responses. Therefore the protocol was designed to combat this by prompting users to fill out a daily EMA in the application via push notification, with manual outreach to users who failed to complete an EMA within 48 hours, as well as designing the K10 to be simple to complete multiple-choice assessment. This combination successfully encouraged active participation in the study; however, there was no mechanism designed to confirm or validate that the resulting inputs accurately reflected a user's well-being.

It is highly likely, therefore, that at least some users were motivated to respond quickly, and not necessarily accurately. This would result in users simply selecting the default answer of no reported anxiety to each question.

Furthermore, there may have been a reluctance among users to accurately report mental health issues given the perceived embarrassment or stigma associated with poor mental health. Under-reporting of mental health issues is a persistent issue that plagues the domain more generally and isn't limited to this study [CW02]; however, failing to account for under-reporting is a notable issue.

Finally, even well-meaning participants may have failed to accurately represent their mental health state due to their either overlooking or mischaracterizing the stresses they encountered. This is particularly true when comparing responses across users, where baseline expectations of stress may vary wildly among participants, with prior work, for instance, demonstrating a clear association between gender and reported wellbeing [PS01], the result being that one participant's perception of a 'normal' day, might easily be classified as a low or moderately-stressed day by another.

Solving this challenge is essential for ultimately achieving the intended goal of accurate

classification of mental illness, for unlike alternative labeling exercises, where quantifiable metrics are possible, here the labeling of an objective state, mental illness, particularly when physiological monitoring is not available, is entirely reliant on subjective inputs, ones that are difficult to capture accurately, and even more difficult to standardize across users.

We recommend that future studies work on addressing these concerns by better anticipating and correcting for challenges with accurate labeling of mental health.

There are a number of possible remedies to this. In the questionnaire itself, carefully structuring the questions can engage users to provide more thought-out results. Cross-validating questions designed to ensure internal consistency are also effective for ensuring user accuracy [LHD98].

Consideration should also be given to alternative methods for collecting labels. Interviewing subjects to determine their mental health, for instance, would likely produce more accurate results, although would have attendant trade-offs of its own, such as reducing the number of labels that could effectively be captured.

Educating participants on the presentations of anxiety could also be key to a more accurate and consistent recall of symptoms. Finally, developing participants' trust through engagement and transparency could help solicit more honest engagement.

7.4.1.4 Subject Heterogeneity

The activities tracked by the eWellness app showcase significant heterogeneity across subjects in terms of usage patterns. Variables like distance traveled, number of texts and calls, and physical activity levels are all far more likely to be impacted by the individual's lifestyle than their mental health on any given day.

Figure 7.7 showcases a fairly typical distribution, in this case the duration spent on foot in a given day, bucketed into quartiles, with the 4 labels of interest (with L1 or Level-1 corresponding to Low-distress, L2 to Moderate Distress, L3 to High-Distress, and L4 to

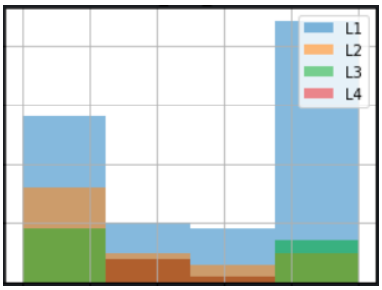


Figure 7.7: Histogram of normalized values of Duration on Foot for the 4 labels of stress

Very high Distress), in this classification superimposed. While intuitively, more time spent on foot may be associated with better mental health, here we observe no clear pattern.

It was therefore assumed that primary success would be achieved by classifying mental health within users across time, once their baselines for normative behavior were established, rather than across users. The limitations of this initial dataset did not allow for adequate classifying by individual; however, the fact that classification success was achieved by bundling samples across all subjects is remarkable in its indication that cross-subject learning in this domain could be possible. We suspect that part of this result likely stems from normalization performed on the data to account for habitual differences in subject usage. By normalizing the data in this manner, the absolute number is rendered largely moot, and instead, variances in user patterns are highlighted, as it is likely the day-to-day variations that are more reflective of shifts in mental health. Additional data collection is necessary to validate this finding.

7.4.1.5 Usability

Attempting to gauge the viability of the concept, participants in the pilot were asked to submit a voluntary anonymized post-study questionnaire regarding their perceptions of the application and its data collection practices. All participants responded. A significant majority described the application as somewhat (40%) or mostly (40%) useful. Likewise, all users endorsed feeling comfortable with the application, and only one user expressed reservations

about the data being collected.

All participants obtained detailed accounting of the data that was collected as part of their onboarding process to the study. No individual declined to participate after learning the precise nature of what was being tracked. This sampling suggests that, particularly among young adults who are more accustomed to digitized lives, there is less concern about data collection through their mobile devices. Limiting the collection of PII could be sufficient to assuage most privacy concerns.

The primary issue users had with the application was its battery consumption resulting from heavy over-sampling of the sensors. Future iterations of the application will seek to optimize battery usage by minimizing the sampling frequency.

7.4.2 Short-term Stress

In this part, we discuss our proposed methods for the recognition of short-term stressful episodes.

7.4.2.1 Labeling

In our experiments, we followed the strategy below in parameterizing the teacher function for every subject:

$$\mu(d) = \begin{cases} d[\text{`probe\_datetime`}] & \text{if single timestamp} \\ \text{avg}(d[\text{`probe\_datetime`}]) & \text{if time range} \end{cases}$$

$$\sigma(d) = \begin{cases} 30 \times 60 \text{ sec} & \text{if single timestamp} \\ \dfrac{\text{len}(d[\text{`probe\_datetime`}])}{(30 \times 60 \text{ sec})} & \text{if time range} \end{cases}$$

In this work and as a proof of concept, we focused on predicting whether a window ends

in a high-stress note. We considered the coefficient $\lambda(d)$ which adjusts the magnitude of each probe to be proportionate to the reported stress levels, and respectively defined the high-stress window as one that the corresponding teacher function returning a value larger than 0.5 for its endpoint, which led to a high consistency between the resulting teacher functions and the reported episodes. The sensory data used in our work was based on heart rate (every 15 seconds), a measure of heart-rate variability (every 3 minutes), pulse oximeter (every 1 minute), and respiration rate (every 2 minutes). On the input side and to help stabilize the training further, we fit min-max normalizers on the features across the time slices in the train set.

7.4.3 Modeling and Optimization

The recurrent neural network module we considered is a 4-layer bi-directional RNN in Long Short Term Memory (LSTM) configuration, leading to a $z \in \mathbb{R}^{256}$ latent representation. To prepare the inputs for processing, we perform early fusion of the sensory readings and create a sequence of vectors representing physiological status at each timestamp, as it is a more intuitive approach for modeling the inputs in this case.

Our optimization protocol employed the Adam algorithm with a learning rate of 1e-3 and a weight decay of 1e-4 to help with overfitting. We also made use of cosine annealing scheduling, reducing the learning rate across our 100 epoch experiments.

With regards to adversarial regularization for enhancing subject invariance, our $f_{\text{adv}}(\cdot; \psi)$ is a two-layer MLP mapping the latent representation to the subject identifier label. The results shown in Table 7.4 indicate an increase in the generalization performance.

Focusing on our real-world perceived stress corpus, we conducted experiments under the main settings of 1) supervised training baseline, 2) pre-training the contrastive objective and fine-tuning via supervised objective, and 3) training the supervised objective and simultaneously optimizing a scaled version of the contrastive term as a regularizing loss.

Table 7.4: Performance overview for the task of recognizing high-stress windows

	Acc
Supervised Setup	62.0
Supervised Setup **+ Adversarial Subject Invariance**	64.5

Table 7.5: Performance comparison for the trained pipeline under different learning setups

Method	**Test Accuracy**
Early-fusion + Supervised Training [FLB22]	64%
Late-fusion + Supervised Training	66%
Late-fusion + Contrastive Pre-training + Fine-tuning	70%
Late-fusion + Supervised Training **+ Contrastive Regularization**	**73%**

We observed that leveraging more features and following a late-fusion protocol for combining modality representations did lead to an improved generalization performance over the supervised setup proposed in [FLB22], which combined the features at the beginning of the pipeline. In the case of our cohort, training with contrastive regularization led to the best generalization on the unseen test data, and the results are shown in Table 7.5. Note that, in general, it is hard to say which self-supervised setup (pre-training versus regularization) is best, as it could depend on other factors, including model complexity, optimization, data availability, and task difficulty. That being said, our approach allows learning high-quality representations by optimizing the modality-contrastive objective via both of these setups.

Additionally, we investigated the interpretability and leveraged the task-specific attention mechanism in our pipeline, which pools the representations from different modalities, to study the utility and contribution of observations from each feature group. This enables the network to dynamically assign weights to each modality's latent representation (in the

shared space) as it processes each instance, allowing us to study their contribution both per instance and in expectation for performing the desired task. In Figure 7.8, we have shown the results on this matter for the contrastive regularization setup[5]. The results indicate that even though the contributions of the different modalities follow a non-uniform distribution as expected, none of them were ignored by the model, and they all play a part in the final predictions.

Figure 7.8: The average contribution of the four modalities to the final episode embeddings.

7.4.3.1 Broader Impact

In the context of remote health monitoring, there are several factors addressing which are of paramount importance. In what follows, we elaborate upon these factors and how the solution proposed in this work attempts to address them:

- *Affordability and Compatibility*: For the scalability of a proposed remote health monitoring framework, focusing on widely available devices sold at affordable prices makes

[5]The label `heart` in Figure 7.8 corresponds to the `daily` modality's information, given that its main focus is heart-rate.

deploying the system easier. We focused on basic physiological signals for which reading sensors are available in most commercially available smartwatches. Nonetheless, the proposed methodology has no intrinsic limitation regarding the modalities used; thus, additional data available in often more expensive devices (e.g., galvanic skin response) can also be utilized in the same methodology, and the main requirement is providing a modality-specific encoder fit for the data domain. Furthermore, this framework offers a more encapsulated view in representing different modalities as the observation from each can be embedded by a dedicated encoder first, and the contrastive objective encourages each local branch to optimize its parameters towards the given task as well. This has clear advantages in terms of transferring knowledge as well, an example of which could be initializing each branch separately via pre-trained weights so as to prepare a better starting point for the model and optimization.

- *Ease-of-use*: Optimizing remote health monitoring with regards to minimizing the amount of required user interaction makes it easier for individuals to use the system. This is why passive monitoring techniques are receiving more attention in the eHealth domain.

- *Interpretation*: In all automated healthcare applications of machine learning, any insight and interpretation into what parts of the observation a model mostly focuses on in determining the final decision is crucial and can help experts better validate the system. We incorporated a task-specific attention mechanism for pooling the representations from different modalities. This helps determine the weights assigned to each modality (per instance and in expectation) to perform the task efficiently.

- *Limited Data*: The data availability for eHealth applications is often limited due to the difficulty and costs of conducting large-scale studies, the exclusivity of data, and privacy reasons. It is, therefore, essential to try to maximize the use of data in model training. This work combines label smoothing with inter-mode self-supervision objectives to go

beyond self-reported supervision objectives.

7.4.3.2 Limitations

It is crucial to discuss the main limitations of this work given the sensitive nature of dealing with health as its objective. In this work, we relied on self-reported entries to decide the supervision signals for individual timelines. This has the issue of being prone to human error, as one might not accurately recall the time and extent to which one has felt stress. Additionally, reports on the intensity of the felt stress are also subject to noise. Another challenge is the small size of our dataset. A primary reason behind our self-supervision component in this work was alleviating the negative impacts associated with the aforementioned limitations.

7.5 Conclusion

In this chapter, we discussed passive-sensing monitoring solutions for short-term episodes of stress as well as longer-term manifestations of anxiety patterns. Our proposed solutions leveraged sensor data obtained in a non-intrusive fashion via smartphone and smartwatch and employed machine learning solutions to make inferences conditioned on their data on individuals' mental health status pertinent to stress and anxiety. Our empirical findings demonstrated the efficient performance of proposed solutions to the defined tasks on stress and anxiety.

CHAPTER 8

Conclusion

In this thesis, we investigated algorithms and methods for effective training and efficient optimization of representation learning pipelines focused on mapping longitudinal health data for various applications. Our works covered a diverse set of domains in health analytics in which the key observations were temporal in nature and discussed our proposed solutions towards learning efficient representations suitable for desired downstream tasks. Specifically, our discussion covered the following domains:

- **Physiological Health**: For outpatient healthcare and remote health monitoring, we remarked upon the importance of effectively modeling time-series data and physiological readings. Specifically, we focused on ECG as a widely used signal in health applications. We demonstrated that focusing on representative supervised tasks can help with learning highly transferable representations, exhibiting high performance when employed towards predictive modeling on tasks defined on the same observation. Additionally, we designed and utilized machine learning pipelines for activity recognition in children, rehabilitation exercise tracking, as well as posture monitoring, and discussed the importance of such problems in the HealthAI domain. In the context of inpatient care, we discussed representation learning for electronic health records. We demonstrated that even without large-scale annotations pertaining to highly representative tasks, one could still leverage the data via unsupervised and self-supervised training from multiple modalities to learn efficient representations. We demonstrated the effectiveness of said representations on commonly defined supervised tasks on EHR

corpora.

- **Mental Health**: We focused on stress and anxiety, which along with depression, are recognized as the most common mental health issues. We conducted studies in recognition and tracking of short-term bouts of stress, as well as longer-term manifestations of anxiety in one's behavior. For short-term stress monitoring, we designed a smartwatch-based monitoring platform. We worked on efficient representation learning from multiple modalities of data that it provides, and with techniques such as adversarial objectives and self-supervised pre-training and regularization, managed to obtain efficient representations for making predictions on whether or not a short-term window is associated with the perception of stress by an individual. For the long-term patterns of anxiety, we proposed a modeling framework centered around phone usage behavior (e.g., number of calls). We gathered a variety of features, from aggregate high-level information on the communication behavior as well as information having to do with one's environment (e.g., ambient sound level and light intensity), and utilized them in making predictions on the levels of distress.

- **Public Health**: Going beyond healthcare solutions focused on individuals, we also discussed the importance of leveraging computerized advancements in the public health domain. Focusing on the recent worldwide pandemic of COVID-19, we proposed an end-to-end platform including dataset creation, monitoring interface, and analytical framework for modeling pandemic outcomes leveraging spatio-temporal representations. The idea was to go beyond the usual observations leveraged in traditional methods such as ARIMA and SIR, and focus on short-term fluctuations in the pandemic outcomes. Our results demonstrated effective performance in county-level predictions of the pandemic outcomes, highlighting the importance of considering spatio-temporal features for predictive modeling and decision-making.

All in all, numerous subdomains of HealthAI benefit from efficient, accurate, and scalable

representation learning methodologies. Compared to the other domains, due to reasons such as patient privacy and the difficulty of obtaining data in many scenarios, representation learning in the HealthAI domain has been challenged to a higher degree by a lack of large-scale data and/or accurate expert annotations. In addition, the temporal and multi-modal nature of the required observations for a wide range of tasks in health analytics makes it particularly important for the researchers to work toward addressing the aforementioned challenges and propose inference pipelines that make the most of the limited available data.

REFERENCES

[A 18] A. Ng, M. Rddy, and S.M. Schueller. "Veterans' perspectives on Fitbit use in treatment for Post-Traumatic Stress Disorder: An interview study." *MIR Mental Health*, 2018.

[AAB16] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. "Tensorflow: Large-scale machine learning on heterogeneous distributed systems." *arXiv preprint arXiv:1603.04467*, 2016.

[Ada] Lady Ada. "Adafruit Feather 32U4 bluefruit le.".

[AFO17] U Rajendra Acharya, Hamido Fujita, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Muhammad Adam. "Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals." *Information Sciences*, **415**:190–198, 2017.

[AJH18] Anand Avati, Kenneth Jung, Stephanie Harman, Lance Downing, Andrew Ng, and Nigam H Shah. "Improving palliative care with deep learning." *BMC medical informatics and decision making*, **18**(4):122, 2018.

[AMB19] Emily Alsentzer, John R Murphy, Willie Boag, Wei-Hung Weng, Di Jin, Tristan Naumann, and Matthew McDermott. "Publicly available clinical BERT embeddings." *arXiv preprint arXiv:1904.03323*, 2019.

[Ami12] Afshin Aminian. "Dynamic orthopaedic chair.", July 31 2012. US Patent 8,231,175.

[anR] "An all-rounded helper to help you improve your sitting postures." `http://ctms0168.epizy.com/2017/12/20/an-all-rounded-helper-to-help-you-improve-your-sitting-postures-2/?i=2`.

[AOH17a] U Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, Arkadiusz Gertych, and Ru San Tan. "A deep convolutional neural network model to classify heartbeats." *Computers in biology and medicine*, **89**:389–396, 2017.

[AOH17b] U. Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, Arkadiusz Gertych, and Ru San Tan. "A deep convolutional neural network model to classify heartbeats." *Computers in Biology and Medicine*, **89**:389–396, 2017.

[APA] APA. "Stress in America 2020." https://www.apa.org/news/press/releases/stress/2020/sia-mental-health-crisis.pdf.

[APA19] 2019.

[AS01] Gavin Andrews and Tim Slade. "Interpreting scores on the Kessler psychological distress scale (K10)." *Australian and New Zealand journal of public health*, **25**(6):494–497, 2001.

[ASN16] Mohammad Abu Alsheikh, Ahmed Selim, Dusit Niyato, Linda Doyle, Shaowei Lin, and Hwee-Pink Tan. "Deep Activity Recognition Models with Triaxial Accelerometers." In *AAAI Workshop: Artificial Intelligence Applied to Assistive Technologies and Smart Environments*, 2016.

[Ass] National Restaurant Association. "State Statistics." http://web.archive.org/web/*/https://www.restaurant.org/research/state. Accessed: 2020-07-15.

[Ass16] American College Health Association et al. "American college health association–national health assessment II: Reference group executive summary spring 2016.", 2016.

[aus] "Chapter - K10 Scoring.".

[Bal12] Pierre Baldi. "Autoencoders, unsupervised learning, and deep architectures." In *Proceedings of ICML workshop on unsupervised and transfer learning*, pp. 37–49, 2012.

[BAS17] Renato Baptista, Michel Antunes, Abd El Rahman Shabayek, Djamila Aouada, and Björn Ottersten. "Flexible feedback system for posture monitoring and correction." In *2017 Fourth International Conference on Image Information Processing (ICIIP)*, pp. 1–6. IEEE, 2017.

[BCB14a] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." *arXiv preprint arXiv:1409.0473*, 2014.

[BCB14b] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. "Neural machine translation by jointly learning to align and translate." *arXiv preprint arXiv:1409.0473*, 2014.

[BCE18] Tian Bai, Ashis Kumar Chanda, Brian L Egleston, and Slobodan Vucetic. "EHR phenotyping via jointly embedding medical concepts and words into a unified vector space." *BMC medical informatics and decision making*, **18**(4):123, 2018.

[BCF18] Mehdi Boukhechba, Philip Chow, Karl Fua, Bethany A Teachman, Laura E Barnes, et al. "Predicting social anxiety from global positioning system traces of college students: feasibility study." *JMIR mental health*, **5**(3):e10101, 2018.

[BCV13] Yoshua Bengio, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives." *IEEE transactions on pattern analysis and machine intelligence*, **35**(8):1798–1828, 2013.

[BGD14] Oresti Banos, Juan-Manuel Galvez, Miguel Damas, Hector Pomares, and Ignacio Rojas. "Window size impact in human activity recognition." *Sensors*, **14**(4):6474–6499, 2014.

[BGN16] Hajra Banoo, Vibha Gangwar, and Nusrat Nabi. "Effect of cold stress and the cold pressor test on blood pressure and heart rate." 2016.

[BHG18] Rebecca H Bitsko, Joseph R Holbrook, Reem M Ghandour, Stephen J Blumberg, Susanna N Visser, Ruth Perou, and John T Walkup. "Epidemiology and impact of health care provider–diagnosed anxiety and depression among US children." *Journal of developmental and behavioral pediatrics: JDBP*, **39**(5):395, 2018.

[BI04] Ling Bao and Stephen Intille. "Activity recognition from user-annotated acceleration data." *Pervasive computing*, pp. 1–17, 2004.

[BKS95] R Bousseljot, D Kreiseler, and A Schnabel. "Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet." *Biomedizinische Technik/Biomedical Engineering*, **40**(s1):317–318, 1995.

[BMC98] Anna M Bardone, Terrie E Moffitt, Avshalom Caspi, Nigel Dickson, Warren R Stanton, and Phil A Silva. "Adult physical health outcomes of adolescent girls with conduct disorder, depression, and anxiety." *Journal of the American Academy of Child & Adolescent Psychiatry*, **37**(6):594–601, 1998.

[BMK18] Ronny Bruffaerts, Philippe Mortier, Glenn Kiekens, Randy P Auerbach, Pim Cuijpers, Koen Demyttenaere, Jennifer G Green, Matthew K Nock, and Ronald C Kessler. "Mental health problems in college freshmen: Prevalence and academic functioning." *Journal of affective disorders*, **225**:97–103, 2018.

[BSG12] Neeltje M Batelaan, Jan Spijker, Ron de Graaf, and Pim Cuijpers. "Mixed anxiety depression should not be included in DSM-5." *The Journal of nervous and mental disease*, **200**(6):495–498, 2012.

[BSW15] Dror Ben-Zeev, Emily A Scherer, Rui Wang, Haiyi Xie, and Andrew T Campbell. "Next-generation psychiatric assessment: Using smartphone sensors to monitor behavior and mental health." *Psychiatric rehabilitation journal*, **38**(3):218, 2015.

[Car06] Sam Cartwright-Hatton. "Anxiety of childhood and adolescence: Challenges and opportunities.", 2006.

[CAT16] Cynthia S Crowson, Elizabeth J Atkinson, and Terry M Therneau. "Assessing calibration of prognostic risk scores." *Statistical methods in medical research*, **25**(4):1692–1706, 2016.

[CBS16] Edward Choi, Mohammad Taha Bahadori, Elizabeth Searles, Catherine Coffey, Michael Thompson, James Bost, Javier Tejedor-Sojo, and Jimeng Sun. "Multi-layer representation learning for medical concepts." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1495–1504. ACM, 2016.

[CBS17] Edward Choi, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and Jimeng Sun. "GRAM: Graph-based Attention Model for Healthcare Representation Learning." In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '17, pp. 787–795, New York, NY, USA, 2017. ACM.

[CCA08] Alison J Campbell, Jonathan A Cook, Gillian Adey, and Brian H Cuthbertson. "Predicting death and readmission after intensive care discharge." *British journal of anaesthesia*, **100**(5):656–662, 2008.

[CDCa] CDC. "Laboraty-confirmed COVID-19 Associated Hospitalizations." `https://gis.cdc.gov/grasp/covidnet/COVID19_3.html`, archived at `https://archive.is/Mw9d1`. Accessed: 2020-06-05.

[CDCb] CDC. "A Weekly Summary of US COVID-19 Hospitalization Data." `https://gis.cdc.gov/grasp/COVIDNet/COVID19_1.html`, archived at `https://archive.is/qs0IJ`. Accessed: 2020-06-05.

[CE12] Jana Campbell and Ulrike Ehlert. "Acute psychosocial stress: Does the emotional stress response correspond with physiological responses?" *Psychoneuroendocrinology*, **37**:1111–34, 01 2012.

[CFH17] Philip I Chow, Karl Fua, Yu Huang, Wesley Bonelli, Haoyi Xiong, Laura E Barnes, and Bethany A Teachman. "Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students." *Journal of medical Internet research*, **19**(3):e6820, 2017.

[CFL17] Lucio Ciabattoni, Francesco Ferracuti, Sauro Longhi, Lucia Pepa, Luca Romeo, and Federica Verdini. "Real-time mental stress detection based on smartwatch." In *2017 IEEE International Conference on Consumer Electronics (ICCE)*, pp. 110–111. IEEE, 2017.

[CGC14] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. "Empirical evaluation of gated recurrent neural networks on sequence modeling." *arXiv preprint arXiv:1412.3555*, 2014.

[CGN18a] Xiangrui Cai, Jinyang Gao, Kee Yuan Ngiam, Beng Chin Ooi, Ying Zhang, and Xiaojie Yuan. "Medical concept embedding with time-aware attention." *arXiv preprint arXiv:1806.02873*, 2018.

[CGN18b] Xiangrui Cai, Jinyang Gao, Kee Yuan Ngiam, Beng Chin Ooi, Ying Zhang, and Xiaojie Yuan. "Medical concept embedding with time-aware attention." *arXiv preprint arXiv:1806.02873*, 2018.

[CH21] Xinlei Chen and Kaiming He. "Exploring simple siamese representation learning." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15750–15758, 2021.

[CKN20] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. "A simple framework for contrastive learning of visual representations." In *International conference on machine learning*, pp. 1597–1607. PMLR, 2020.

[CKZ17] Jenny Chum, Min Suk Kim, Laura Zielinski, Meha Bhatt, Douglas Chung, Sharon Yeung, Kathryn Litke, Kathleen McCabe, Jeff Whattam, Laura Garrick, et al. "Acceptability of the Fitbit in behavioural activation therapy for depression: a qualitative study." *BMJ Ment Health*, **20**(4):128–133, 2017.

[Cli] Mayo Clinic. "Anxiety Disorders Symptoms and Causes." `https://www.mayoclinic.org/diseases-conditions/anxiety/symptoms-causes/syc-20350961`.

[CMM20] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. "Unsupervised learning of visual features by contrasting cluster assignments." *Advances in Neural Information Processing Systems*, **33**:9912–9924, 2020.

[cova] "COVID-19 Simulator." `http://web.archive.org/web/20200730193215/https://www.covid19sim.org/`. Accessed: 2020-07-30.

[covb] "Covid-19/coronavirus real time updates with credible sources in us and canada." `https://coronavirus.1point3acres.com/en`, archived at `https://archive.is/J3Vmg`. Accessed: 2020-06-05.

[CPR11] Pierluigi Casale, Oriol Pujol, and Petia Radeva. "Human activity recognition from accelerometer data using a wearable device." *Pattern Recognition and Image Analysis*, pp. 289–296, 2011.

[CW02] Patrick W. Corrigan and Amy C. Watson. "The Paradox of Self-Stigma and Mental Illness." *Clinical Psychology: Science and Practice*, **9**(1):35–53, 2002.

[CWD04] Kerry A Collins, Henny A Westra, David JA Dozois, and David D Burns. "Gaps in accessing treatment for anxiety and depression: challenges for the delivery of care." *Clinical psychology review*, **24**(5):583–616, 2004.

[CXS18] Edward Choi, Cao Xiao, Walter Stewart, and Jimeng Sun. "MiME: Multilevel Medical Embedding of Electronic Health Records for Predictive Healthcare." In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pp. 4547–4557. Curran Associates, Inc., 2018.

[DBS16] Laura Dwyer-Lindgren, Amelia Bertozzi-Villa, Rebecca W Stubbs, Chloe Morozoff, Michael J Kutz, Chantal Huynh, Ryan M Barber, Katya A Shackelford, Johan P Mackenbach, Frank J van Lenthe, et al. "US county-level trends in mortality rates for major causes of death, 1980-2014." *Jama*, **316**(22):2385–2401, 2016.

[DCL18a] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805*, 2018.

[DCL18b] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *CoRR*, **abs/1810.04805**, 2018.

[Dev] UPRIGHT Posture Training Device. "UPRIGHT Posture Training Device." `https://www.uprightpose.com`.

[DH00] Odo Diekmann and Johan Andre Peter Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, volume 5. John Wiley & Sons, 2000.

[DKF20] Sajad Darabi, Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh. "Taper: Time-aware patient ehr representation." *IEEE journal of biomedical and health informatics*, **24**(11):3268–3275, 2020.

[DKS17] Amirhossein Esmaili Dastjerdi, Mohammad Kachuee, and Mahdi Shabany. "Non-invasive blood pressure estimation using phonocardiogram." In *Circuits and Systems (ISCAS), 2017 IEEE International Symposium on*, pp. 1–4. IEEE, 2017.

[DM18] Dmitriy Dligach and Timothy A. Miller. "Learning Patient Representations from Text." *CoRR*, **abs/1805.02096**, 2018.

[DMH17] Sajad Darabi, Babak Moatamed, Wenhao Huang, Migyeong Gwak, Casey Metoyer, Mike Linn, and Majid Sarrafzadeh. "Heart rate compression & time reduction method for HRV monitoring in athletes." In *2017 IEEE Healthcare Innovations and Point of Care Technologies (HI-POCT)*, pp. 152–155. IEEE, 2017.

[DP] Center of Disease Control and Prevention. "Mental Health and Coping with Stress During COVID-19 Pandemic." https://web.archive.org/web/20200804105944/https://www.cdc.gov/coronavirus/2019-ncov/daily-life-coping/managing-stress-anxiety.html. Accessed: 2020-08-04.

[EKS17] Amirhossein Esmaili, Mohammad Kachuee, and Mahdi Shabany. "Nonlinear Cuffless Blood Pressure Estimation of Healthy Subjects Using Pulse Transit Time and Arrival Time." *IEEE Transactions on Instrumentation and Measurement*, **66**(12):3299–3308, 2017.

[ele] "County Presidential Election Returns2000-2016, 2018." https://doi.org/10.7910/DVN/VOQCHQ, archived at https://archive.is/cLVL5. Accessed: 2020-06-05.

[EV16] Jheanel E Estrada and Larry A Vea. "Real-time human sitting posture detection using mobile devices." In *2016 IEEE Region 10 Symposium (TENSYMP)*, pp. 140–144. IEEE, 2016.

[FAH21] Chandra L Ford, Bita Amani, Nina T Harawa, Randall Akee, Gilbert C Gee, Majid Sarrafzadeh, Consuela Abotsi-Kowu, Shayan Fazeli, Cindy Le, Ezinne Nwankwo, et al. "Adequacy of existing surveillance systems to monitor racism, social stigma and COVID inequities: a detailed assessment and recommendations." *International journal of environmental research and public health*, **18**(24):13099, 2021.

[FDF10] Davide Figo, Pedro C Diniz, Diogo R Ferreira, and João M Cardoso. "Preprocessing techniques for context recognition from accelerometer data." *Personal and Ubiquitous Computing*, **14**(7):645–662, 2010.

[FDN19] Emmanouil Fragkiadakis, Kalliopi V Dalakleidi, and Konstantina S Nikita. "Design and Development of a Sitting Posture Recognition System." In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3364–3367. IEEE, 2019.

[FIO19] Yusuke Fukazawa, Taku Ito, Tsukasa Okimura, Yuichi Yamashita, Takaki Maeda, and Jun Ota. "Predicting anxiety state using smartphone-based passive sensing." *Journal of biomedical informatics*, **93**:103151, 2019.

[FKS] Shayan Fazeli, Mohammad Kachuee, Majid Sarrafzadeh, and Afshin Aminian. "WatChair: AI-Powered Real-time Monitoring of Sitting Posture and Corrective Suggestions using Wearable Motion Sensor System.".

[FL] Shayan Fazeli and Lionel Levine. "tabluence.".

[FLB22] Shayan Fazeli, Lionel Levine, Mehrab Beikzadeh, Baharan Mirzasoleiman, Bita Zadeh, Tara Peris, and Majid Sarrafzadeh. "Passive Monitoring of Physiological Precursors of Stress Leveraging Smartwatch Data." In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2893–2899. IEEE, 2022.

[FLB23] Shayan Fazeli, Lionel Levine, Mehrab Beikzadeh, Baharan Mirzasoleiman, Bita Zadeh, Tara Peris, and Majid Sarrafzadeh. "A Self-supervised Framework for Improved Data-Driven Monitoring of Stress via Multi-modal Passive Sensing." *arXiv preprint arXiv:2303.14267*, 2023.

[FMS21] Shayan Fazeli, Babak Moatamed, and Majid Sarrafzadeh. "Statistical analytics and regional representation learning for covid-19 pandemic understanding." In *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)*, pp. 248–257. IEEE, 2021.

[fre]

[FZO21] Shayan Fazeli, Davina Zamanzadeh, Anaelia Ovalle, Thu Nguyen, Gilbert Gee, and Majid Sarrafzadeh. "COVID-19 and Big Data: Multi-faceted Analysis for Spatio-temporal Understanding of the Pandemic with Social Media Conversations." *arXiv preprint arXiv:2104.10807*, 2021.

[FZS19] Yi Fan, Kai Zhao, Zheng-Li Shi, and Peng Zhou. "Bat coronaviruses in China." *Viruses*, **11**(3):210, 2019.

[GAG00] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. "Physiobank, physiotoolkit, and physionet." *Circulation*, **101**(23):e215–e220, 2000.

[GEL23] Rohit Girdhar, Alaaeldin El-Nouby, Zhuang Liu, Mannat Singh, Kalyan Vasudev Alwala, Armand Joulin, and Ishan Misra. "Imagebind: One embedding space to bind them all." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15180–15190, 2023.

[GFE19] Migyeong Gwak, Shayan Fazeli, Ghazaal Ershadi, Majid Sarrafzadeh, Melina Ghodsi, Afshin Aminian, and John A Schlechter. "EXTRA: exercise tracking and analysis platform for remote-monitoring of knee rehabilitation." In *2019 IEEE 16th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 1–4. IEEE, 2019.

[GHC19] Jiaqi Gong, Yu Huang, Philip I Chow, Karl Fua, Matthew S Gerber, Bethany A Teachman, and Laura E Barnes. "Understanding behavioral dynamics of social anxiety among college students through smartphone sensors." *Information Fusion*, **49**:57–68, 2019.

[GOM18] Diana Galvan, Naoaki Okazaki, Koji Matsuda, and Kentaro Inui. "Investigating the Challenges of Temporal Relation Extraction from Clinical Text." In *Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*, pp. 55–64, Brussels, Belgium, October 2018. Association for Computational Linguistics.

[GS05] Alex Graves and Jürgen Schmidhuber. "Framewise phoneme classification with bidirectional LSTM and other neural network architectures." *Neural Networks*, **18**(5):602–610, 2005.

[GSA20] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al. "Bootstrap your own latent-a new approach to self-supervised learning." *Advances in neural information processing systems*, **33**:21271–21284, 2020.

[HAR19] Kexin Huang, Jaan Altosaar, and Rajesh Ranganath. "ClinicalBERT: Modeling Clinical Notes and Predicting Hospital Readmission." *arXiv:1904.05342*, 2019.

[HBH17] Anahita Hosseini, Chris M Buonocore, Sepideh Hashemzadeh, Hannaneh Hojaiji, Haik Kalantarian, Costas Sideris, Alex AT Bui, Christine E King, and Majid Sarrafzadeh. "Feasibility of a Secure Wireless Sensing Smartwatch Application for the Self-Management of Pediatric Asthma." *Sensors*, **17**(8):1780, 2017.

[He18] Si He. "A Sitting Posture Surveillance System Based on Kinect." In *Journal of Physics: Conference Series*, volume 1026, p. 012022. IOP Publishing, 2018.

[hea] "US Data for Download." `http://web.archive.org/web/20200121125528/http://www.healthdata.org/us-health/data-download`. Accessed: 2020-04-02.

[HFV18] Anahita Hosseini, Shayan Fazeli, Eleanne van Vliet, Lisa Valencia, Rima Habre, Majid Sarrafzadeh, and Alex Bui. "Children activity recognition: Challenges and strategies." In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 4331–4334. IEEE, 2018.

[HHP08] Paul Hammerness, Theresa Harpold, Carter Petty, Chantal Menard, Claire Zar-Kessler, and Joseph Biederman. "Characterizing non-OCD anxiety disorders in psychiatrically referred children and adolescents." *Journal of affective disorders*, **105**(1-3):213–219, 2008.

[HOT06] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. "A fast learning algorithm for deep belief nets." *Neural computation*, **18**(7):1527–1554, 2006.

[HS06] Geoffrey E Hinton and Ruslan R Salakhutdinov. "Reducing the dimensionality of data with neural networks." *science*, **313**(5786):504–507, 2006.

[Hwa19] Andrew Hwang. "Arduino Pressure Sensor Cushion for Tracking and Improving Sitting Posture." *International Journal of Biomedical and Biological Engineering*, **13**(11):454–460, 2019.

[HXL16] Yu Huang, Haoyi Xiong, Kevin Leach, Yuyan Zhang, Philip Chow, Karl Fua, Bethany A Teachman, and Laura E Barnes. "Assessing social anxiety using GPS trajectories and point-of-interest data." In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 898–903, 2016.

[HZR16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[HZY22] Shayan Hassantabar, Joe Zhang, Hongxu Yin, and Niraj K Jha. "Mhdeep: Mental health disorder detection system based on wearable sensors and artificial neural networks." *ACM Transactions on Embedded Computing Systems*, **21**(6):1–22, 2022.

[IGK06] Omer T Inan, Laurent Giovangrandi, and Gregory TA Kovacs. "Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features." *IEEE Transactions on Biomedical Engineering*, **53**(12):2507–2515, 2006.

[IS18] Karlos Ishac and Kenji Suzuki. "Lifechair: A conductive fabric sensor-based smart cushion for actively shaping sitting posture." *Sensors*, **18**(7):2261, 2018.

[J 17] J. Chum, M.S. Zielinski, M. Bhatt, D. Chung, S.Yeung and Z. Samaan. "Acceptabiity of the Fitbit in behavioural activation therapy for depression: a qualitative study." *Evidence-Based Mental health*, 2017.

[JD17] Linpeng Jin and Jun Dong. "Classification of normal and abnormal ECG records using lead convolutional neural network and rule inference." *Science China Information Sciences*, **60**(7), 2017.

[JG93] Kris Jensen and Ben K Graf. "The effects of knee effusion on quadriceps strength and knee intraarticular pressure." *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, **9**(1):52–56, 1993.

[JJD01] Fiona K Judd, Henry Jackson, Julian Davis, Alexandra Cockram, Angela Komiti, Nicholas Allen, Greg Murray, Michael Kyrios, and Gene Hodgins. "Improving access for rural Australians to treatment for anxiety and depression: the University of Melbourne Depression and Anxiety Research and Treatment Group–Bendigo Health Care Group initiative." *Australian Journal of Rural Health*, **9**(2):92–97, 2001.

[JOK11] Xiaoqian Jiang, Melanie Osl, Jihoon Kim, and Lucila Ohno-Machado. "Calibrating predictive model estimates to support personalized medicine." *Journal of the American Medical Informatics Association*, **19**(2):263–274, 2011.

[JPS16] Alistair EW Johnson, Tom J Pollard, Lu Shen, H Lehman Li-wei, Mengling Feng, Mohammad Ghassemi, Benjamin Moody, Peter Szolovits, Leo Anthony Celi, and Roger G Mark. "MIMIC-III, a freely accessible critical care database." *Scientific data*, **3**:160035, 2016.

[KAC12] Andrew MQ King, Michael J Adams, Eric B Carstens, and Elliot J Lefkowitz. "Virus taxonomy." *Ninth report of the International Committee on Taxonomy of Viruses*, pp. 486–487, 2012.

[kaga] "COVID-19 Open Research Dataset Challenge (CORD-19)." `https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge`, archived at `https://archive.is/nokIg`. Accessed: 2020-06-05.

[kagb] "Diversity index of US counties." `https://www.kaggle.com/mikejohnsonjr/us-counties-diversity-index`, archived at `https://archive.is/uX9iX`. Accessed: 2020-06-05.

[kagc] "ICU Beds by county in the US." `https://www.kaggle.com/jaimeblasco/icu-beds-by-county-in-the-uss`, archived at `https://archive.is/QgAoO`. Accessed: 2020-06-05.

[kagd] "Novel coronavirus 2019 dataset." `https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset`, archived at `https://archive.is/zfDjx`. Accessed: 2020-06-05.

[kage] "United States Droughts by County." `https://www.kaggle.com/us-drought-monitor/united-states-droughts-by-county`, archived at `https://archive.is/JgGoj`. Accessed: 2020-06-05.

[kagf] "US Census Demographical Data." `https://www.kaggle.com/muonneutrino/us-census-demographic-data`, archived at `https://archive.is/ZY12v`. Accessed: 2020-06-05.

[kagg] "US county-level mortality." `https://www.kaggle.com/IHME/us-countylevel-mortality`, archived at `https://archive.is/xEVs3`. Accessed: 2020-06-05.

[kagh] "US drought monitor." `https://droughtmonitor.unl.edu/`, archived at `https://archive.is/P76Bb`. Accessed: 2020-06-05.

[kagi] "US Household Income Statistics." https://www.kaggle.com/goldenoakresearch/us-household-income-stats-geo-locations, archived at https://archive.is/iJaLT. Accessed: 2020-06-05.

[kagj] "US mortality rates by county." http://ghdx.healthdata.org/record/ihme-data/united-states-mortality-rates-county-1980-2014, archived at https://archive.is/juSbk. Accessed: 2020-06-05.

[KB14] Diederik Kingma and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*, 2014.

[KBD15] Javad Kojuri, Reza Boostani, Pooyan Dehghani, Farzad Nowroozipour, and Nasrin Saki. "Prediction of acute myocardial infarction with artificial neural networks in patients with nondiagnostic electrocardiogram." *Journal of Cardiovascular Disease Research Vol*, **6**(2):51, 2015.

[KDM19] Mohammad Kachuee, Sajad Darabi, Babak Moatamed, and Majid Sarrafzadeh. "Dynamic Feature Acquisition Using Denoising Autoencoders." *IEEE Trans. Neural Netw. Learning Syst.*, **30**(8):2252–2262, 2019.

[KF19] Dario Krpan and Barbara Fasolo. "Revisiting embodied approach and avoidance effects on behavior: The influence of sitting posture on purchases of rewarding foods." *Journal of Experimental Social Psychology*, **85**:103889, 2019.

[KFS18] Mohammad Kachuee, Shayan Fazeli, and Majid Sarrafzadeh. "Ecg heartbeat classification: A deep transferable representation." In *2018 IEEE international conference on healthcare informatics (ICHI)*, pp. 443–444. IEEE, 2018.

[KIG16] Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks." *IEEE Transactions on Biomedical Engineering*, **63**(3):664–675, 2016.

[accuracy: 99! There were many formulas (just to fill their paper).]

[KKE17] Daphne Koinis Mitchell, Sheryl J Kopel, Cynthia A Esteban, Ronald Seifer, Nico W Vehse, Sath Chau, and Elissa Jelalian. "Asthma Status And Physical Activity In Urban Children." In *B23. NOVEL EPIDEMIOLOGY, MANAGEMENT, AND OUTCOMES IN ASTHMA*, pp. A2993–A2993. Am Thoracic Soc, 2017.

[KKH18] Yuri Kwon, Ji-Won Kim, Jae-Hoon Heo, Hyeong-Min Jeon, Eui-Bum Choi, and Gwang-Moon Eom. "The effect of sitting posture on the loads at cervico-thoracic and lumbosacral joints." *Technology and Health Care*, **26**(S1):409–418, 2018.

[KKM17] Mohammad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade, and Mahdi Shabany. "Cuffless blood pressure estimation algorithms for continuous health-care monitoring." *IEEE Transactions on Biomedical Engineering*, **64**(4):859–869, 2017.

[KM27] William Ogilvy Kermack and Anderson G McKendrick. "A contribution to the mathematical theory of epidemics." *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, **115**(772):700–721, 1927.

[KMC15] Lydia Kogler, Veronika Mueller, Amy Chang, Simon Eickhoff, Peter Fox, Ruben Gur, and Birgit Derntl. "Psychosocial versus physiological stress — Meta-analyses on deactivations and activations of the neural correlates of stress reactions." *NeuroImage*, **119**, 06 2015.

[KMW18] Zachary King, Judith Moskowitz, Laurie Wakschlag, and Nabil Alshurafa. "Predicting Perceived Stress Through Mirco-EMAs and a Flexible Wearable ECG Device." In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pp. 106–109, 2018.

[Kuf20] Tadeusz Kufel. "ARIMA-based forecasting of the dynamics of confirmed Covid-19 cases for selected European countries." *Equilibrium. Quarterly Journal of Economics and Economic Policy*, **15**(2):181–204, 2020.

[KVH19] Patrik Kutilek, Petr Volf, Jan Hejda, Pavel Smrcka, Jindrich Adolf, Vaclav Krivanek, Lenka Lhotska, Karel Hana, Radek Doskocil, Jiri Kacer, et al. "Non-contact Measurement Systems for Physiological Data Monitoring of Military Pilots During Training on Simulators: Review and Application." In *2019 International Conference on Military Technologies (ICMT)*, pp. 1–6. IEEE, 2019.

[KW13] Diederik P Kingma and Max Welling. "Auto-encoding variational bayes." *arXiv preprint arXiv:1312.6114*, 2013.

[KWS20] Benjamin D Killeen, Jie Ying Wu, Kinjal Shah, Anna Zapaishchykova, Philipp Nikutta, Aniruddha Tamhane, Shreya Chakraborty, Jinchi Wei, Tiger Gao, Mareike Thies, et al. "A County-level Dataset for Informing the United States' Response to COVID-19." *arXiv preprint arXiv:2004.00756*, 2020.

[LBA18] CC Lim, SN Basah, MA Ali, and CY Fook. "Wearable Posture Identification System for Good Sitting Position." *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, **10**(1-16):135–140, 2018.

[LCC19] Ping Liu, Wu Chen, and Jin-Ping Chen. "Viral metagenomics revealed Sendai virus and coronavirus infection of Malayan pangolins (Manis javanica)." *Viruses*, **11**(11):979, 2019.

[Leo21] Megan Leonhardt. "What you need to know about the cost and accessibility of mental health care in America.", 2021.

[Lew95a] James R Lewis. "IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use." *International Journal of Human-Computer Interaction*, **7**(1):57–78, 1995.

[Lew95b] James R Lewis. "IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use." *International Journal of Human-Computer Interaction*, **7**(1):57–78, 1995.

[LGZ16] Alex M Lamb, Anirudh Goyal Alias Parth Goyal, Ying Zhang, Saizheng Zhang, Aaron C Courville, and Yoshua Bengio. "Professor forcing: A new algorithm for training recurrent networks." In *Advances In Neural Information Processing Systems*, pp. 4601–4609, 2016.

[LHD98] Fuzhong Li, Peter Harmer, Terry E. Duncan, Sue C. Duncan, Alan Acock, and Takashi Yamamoto. "Confirmatory Factor Analysis of the Task and Ego Orientation in Sport Questionnaire with Cross-Validation." *Research Quarterly for Exercise and Sport*, **69**(3):276–283, 1998. PMID: 9777664.

[LKB98] Guoan Li, Kenji Kawamura, Peter Barrance, Edmund YS Chao, and Ken Kaufman. "Prediction of muscle recruitment and its effect on joint reaction forces during knee exercises." *Annals of biomedical engineering*, **26**:725–733, 1998.

[LLW15] Bin Liu, Jikui Liu, Guoqing Wang, Kun Huang, Fan Li, Yang Zheng, Youxi Luo, and Fengfeng Zhou. "A novel electrocardiogram parameterization algorithm and its application in myocardial infarction detection." *Computers in biology and medicine*, **61**:178–184, 2015.

[LPM15] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. "Effective approaches to attention-based neural machine translation." *arXiv preprint arXiv:1508.04025*, 2015.

[Lun01a] Arnold M Lund. "Measuring usability with the use questionnaire12." *Usability interface*, **8**(2):3–6, 2001.

[Lun01b] Arnold M Lund. "Measuring usability with the use questionnaire12." *Usability interface*, **8**(2):3–6, 2001.

[LYK19] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. "Biobert: pre-trained biomedical language representation model for biomedical text mining." *arXiv preprint arXiv:1901.08746*, 2019.

[LZ16] Taiyong Li and Min Zhou. "ECG classification using wavelet packet entropy and random forests." *Entropy*, **18**(8):285, 2016.

[LZR18] Jingshu Liu, Zachariah Zhang, and Narges Razavian. "Deep ehr: Chronic disease prediction using medical notes." *arXiv preprint arXiv:1808.04928*, 2018.

[MAL13] Roshan Joy Martis, U Rajendra Acharya, Choo Min Lim, KM Mandana, Ajoy K Ray, and Chandan Chakraborty. "Application of higher order cumulant features for cardiac health diagnosis using ECG signals." *International journal of neural systems*, **23**(04):1350014, 2013.

[mbi19] "MetaTracker.", Feb 2019.

[MBR05] Maura Mitrushina, Kyle B Boone, Jill Razani, and Louis F D'Elia. *Handbook of normative data for neuropsychological assessment*. Oxford University Press, 2005.

[Med98] Association for the Advancement of Medical Instrumentation et al. "Testing and reporting performance results of cardiac rhythm and st segment measurement algorithms." *ANSI/AAMI EC38*, **1998**, 1998.

[Men] National Institute of Mental Health (NIMH). "Any Anxiety Disorder." `https://www.nimh.nih.gov/health/statistics/any-anxiety-disorder`.

[MH08] Laurens van der Maaten and Geoffrey Hinton. "Visualizing data using t-SNE." *Journal of machine learning research*, **9**(Nov):2579–2605, 2008.

[MHB10] Kathleen Ries Merikangas, Jian-ping He, Marcy Burstein, Sonja A Swanson, Shelli Avenevoli, Lihong Cui, Corina Benjet, Katholiki Georgiades, and Joel Swendsen. "Lifetime prevalence of mental disorders in US adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A)." *Journal of the American Academy of Child & Adolescent Psychiatry*, **49**(10):980–989, 2010.

[MHS16] Sangyong Ma, Sangpyo Hong, Hyeon-min Shim, Jang-Woo Kwon, and Sangmin Lee. "A Study on Sitting Posture Recognition using Machine Learning." *The Transactions of the Korean Institute of Electrical Engineers*, **65**(9):1557–1563, 2016.

[MIH13] Aizan Masdar, BSKK Ibrahim, Dirman Hanafi, M Mahadi Abdul Jamil, and KAA Rahman. "Knee joint angle measurement system using gyroscope and flex-sensors for rehabilitation." In *The 6th 2013 Biomedical Engineering International Conference*, pp. 1–4. IEEE, 2013.

[MLK16] Riccardo Miotto, Li Li, Brian A Kidd, and Joel T Dudley. "Deep patient: an unsupervised representation to predict the future of patients from the electronic health records." *Scientific reports*, **6**:26094, 2016.

[MM01] George B Moody and Roger G Mark. "The impact of the MIT-BIH arrhythmia database." *IEEE Engineering in Medicine and Biology Magazine*, **20**(3):45–50, 2001.

[MMC11] Jonathan Masci, Ueli Meier, Dan Cireşan, and Jürgen Schmidhuber. "Stacked convolutional auto-encoders for hierarchical feature extraction." In *International Conference on Artificial Neural Networks*, pp. 52–59. Springer, 2011.

[MP18] S Rasouli D Maryam and Shahram Payandeh. "A novel human posture estimation using single depth image from Kinect v2 sensor." In *2018 Annual IEEE International Systems Conference (SysCon)*, pp. 1–7. IEEE, 2018.

[MSC13] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. "Distributed Representations of Words and Phrases and their Compositionality." In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pp. 3111–3119. Curran Associates, Inc., 2013.

[MWD18] James Mullenbach, Sarah Wiegreffe, Jon Duke, Jimeng Sun, and Jacob Eisenstein. "Explainable prediction of medical codes from clinical text." *arXiv preprint arXiv:1802.05695*, 2018.

[MYP20] IL-Young Moon, Chung-Hwi Yi, IL-Woo Park, and Joon-Hyoung Yong. "Effects of Sitting Posture and Bolus Volume on Activation of Swallowing-Related Muscles." *Journal of Oral Rehabilitation*, 2020.

[NGT20] Nematjon Narziev, Hwarang Goh, Kobiljon Toshnazarov, Seung Ah Lee, Kyong-Mee Chung, and Youngtae Noh. "STDD: Short-Term Depression Detection with Passive Sensing." *Sensors*, **20**(5):1396, 2020.

[NH10] Vinod Nair and Geoffrey E Hinton. "Rectified linear units improve restricted boltzmann machines." In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 807–814, 2010.

[NKB18] Carla FJ Nooijen, Lena V Kallings, Victoria Blom, Örjan Ekblom, Yvonne Forsell, and Maria M Ekblom. "Common perceived barriers and facilitators for reducing sedentary behaviour among office workers." *International journal of environmental research and public health*, **15**(4):792, 2018.

[NMW14] Jill M Newby, Louise Mewton, Alishia D Williams, and Gavin Andrews. "Effectiveness of transdiagnostic internet cognitive behavioural treatment for mixed

anxiety and depression in primary care." *Journal of Affective Disorders*, **165**:45–52, 2014.

[NTW16] Phuoc Nguyen, Truyen Tran, Nilmini Wickramasinghe, and Svetha Venkatesh. "`Deepr`: a convolutional net for medical records." *IEEE journal of biomedical and health informatics*, **21**(1):22–30, 2016.

[nyt] "COVID-19 Data in the United States." `https://usafacts.org/visualizations/coronavirus-covid-19-spread-map`, archived at `https://archive.is/WefdJ`. Accessed: 2020-05-10.

[OOB11] Albertine Oldehinkel, Johan Ormel, Nienke Bosch, Esther Bouma, Arie Roon, Judith Rosmalen, and Harriëtte Riese. "Stressed out? Associations between perceived and physiological stress responses in adolescents: The TRAILS study." *Psychophysiology*, **48**:441–52, 04 2011.

[OPC14] Jorge Osma, Inmaculada Plaza, Elena Crespo, Carlos Medrano, and Raquel Serrano. "Proposal of use of smartphones to evaluate and diagnose depression and anxiety symptoms during pregnancy and after birth." In *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI)*, pp. 547–550. IEEE, 2014.

[Org] World Health Organization. "Cardiovascular Diseases." `http://www.who.int/mediacentre/factsheets/fs317/en/`.

[PGC17] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. "Automatic Differentiation in PyTorch." In *NIPS Autodiff Workshop*, 2017.

[PKK01] Tapani Pöyhönen, Heikki Kyröläinen, Kari L Keskinen, Arto Hautala, Jukka Savolainen, and Esko Mälkiä. "Electromyographic and kinematic analysis of therapeutic knee exercises under water." *Clinical Biomechanics*, **16**(6):496–504, 2001.

[PS01] Martin Pinquart and Silvia Sörensen. "Gender Differences in Self-Concept and Psychological Well-Being in Old Age: A Meta-Analysis." *The Journals of Gerontology: Series B*, **56**(4):P195–P213, 07 2001.

[PS20a] Sen Pei and Jeffrey Shaman. "Initial Simulation of SARS-CoV2 Spread and Intervention Effects in the Continental US." *medRxiv*, 2020.

[PS20b] Mohammad Pourhomayoun and Mahdi Shakibi. "Predicting mortality risk in patients with COVID-19 using artificial intelligence to help medical decision-making." *medRxiv*, 2020.

[QMX10] Huimin Qian, Yaobin Mao, Wenbo Xiang, and Zhiquan Wang. "Recognition of human activities using SVM multi-class classifier." *Pattern Recognition Letters*, **31**(2):100–111, 2010.

[RHH17] Pranav Rajpurkar, Awni Y Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y Ng. "Cardiologist-level arrhythmia detection with convolutional neural networks." *arXiv preprint arXiv:1707.01836*, 2017.

[RHW88] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. "Learning representations by back-propagating errors." *Cognitive modeling*, **5**(3):1, 1988.

[ROC18] Alvin Rajkomar, Eyal Oren, Kai Chen, Andrew M Dai, Nissan Hajaj, Michaela Hardt, Peter J Liu, Xiaobing Liu, Jake Marcus, Mimi Sun, et al. "Scalable and accurate deep learning with electronic health records." *NPJ Digital Medicine*, **1**(1):18, 2018.

[RSB20] Ramesh Raskar, Isabel Schunemann, Rachel Barbar, Kristen Vilcans, Jim Gray, Praneeth Vepakomma, Suraj Kapa, Andrea Nuzzo, Rajiv Gupta, Alex Berke, et al. "Apps gone rogue: Maintaining personal privacy in an epidemic." *arXiv preprint arXiv:2003.08567*, 2020.

[saf] "Private kits: Safepaths; Privacy-by-Design Covid19 Solutions using GPS+Bluetooth for Citizens and Public Health Officials." `https://safepaths.mit.edu`, archived at `https://archive.is/zUgQA`. Accessed: 2020-06-05.

[Sch15] Jürgen Schmidhuber. "Deep learning in neural networks: An overview." *Neural networks*, **61**:85–117, 2015.

[SCL20] Cuihua Shen, Anfan Chen, Chen Luo, Jingwen Zhang, Bo Feng, and Wang Liao. "Using Reports of Symptoms and Diagnoses on Social Media to Predict COVID-19 Case Counts in Mainland China: Observational Infoveillance Study." *Journal of Medical Internet Research*, **22**(5):e19421, 2020.

[Sco] Eugene Scott. "4 reasons coronavirus is hitting black communities so hard." `http://web.archive.org/web/20200728222505/https://www.washingtonpost.com/politics/2020/04/10/4-reasons-coronavirus-is-hitting-black-communities-so-hard/`. Accessed: 2020-07-28.

[SDA14] Naser Safdarian, Nader Jafarnia Dabanloo, and Gholamreza Attarodi. "A new pattern recognition method for detection and localization of myocardial infarction using t-wave integral and total integral as extracted features from one cycle of ecg signal." *Journal of Biomedical Science and Engineering*, **7**(10):818, 2014.

[SGG19] Maksim Sandybekov, Clemens Grabow, Maksym Gaiduk, and Ralf Seepold. "Posture Tracking Using a Machine Learning Algorithm for a Home AAL." In *Intelligent Decision Technologies 2019: Proceedings of the 11th KES International Conference on Intelligent Decision Technologies (KES-IDT 2019)*, volume 2, p. 337. Springer, 2019.

[SGH19] Iwan Aang Soenandi, Meriastuti Ginting, and Budi Harsono. "Real Time Floor Sitting Posture Monitoring using K-Means Clustering." In *Proceedings of the 2nd International Conference on Software Engineering and Information Management*, pp. 194–198, 2019.

[SKN16] Shunsuke Suzuki, Mineichi Kudo, and Atsuyoshi Nakamura. "Sitting posture diagnosis using a pressure sensor mat." In *2016 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, pp. 1–6. IEEE, 2016.

[SLC17] Alexandra L Shilton, Robin Laycock, and Sheila G Crewther. "The Maastricht Acute Stress Test (MAST): Physiological and subjective responses in anticipation, and post-stress." *Frontiers in psychology*, **8**:567, 2017.

[SLY12] Li Sun, Yanping Lu, Kaitao Yang, and Shaozi Li. "ECG analysis using multiple instance learning for myocardial infarction detection." *IEEE transactions on biomedical engineering*, **59**(12):3348–3356, 2012.

[SMA87] Gary L Soderberg, Scott Duesterhaus Minor, Kevin Arnold, Thomas Henry, Joyce Kirchner Chatterson, Debra Ridenour Poppe, and Cheryl Wall. "Electromyographic analysis of knee exercises in healthy subjects and in patients with knee pathologies." *Physical therapy*, **67**(11):1691–1696, 1987.

[SMI12] Satoshi Suzuki, Yasue Mitsukura, Hiroshi Igarashi, Harumi Kobayashi, and Fumio Harashima. "Activity recognition for children using self-organizing map." In *RO-MAN, 2012 IEEE*, pp. 653–658. IEEE, 2012.

[Soc18] Heart Society. "Heart Diseases & Disorders.", 2018.

[SP20] Ajitesh Srivastava and Viktor K Prasanna. "Learning to Forecast and Forecasting to Learn from the COVID-19 Pandemic." *arXiv preprint arXiv:2004.11372*, 2020.

[SSC10] Omid Sayadi, Mohammad B Shamsollahi, and Gari D Clifford. "Robust detection of premature ventricular contractions using a wave-based Bayesian framework." *IEEE Transactions on Biomedical Engineering*, **57**(2):353–362, 2010.

[SSG17] Harini Suresh, Peter Szolovits, and Marzyeh Ghassemi. "The use of autoencoders for discovering patient phenotypes." *arXiv preprint arXiv:1703.07004*, 2017.

[SSG18] Lee Smith, Alexia Sawyer, Benjamin Gardner, Katri Seppala, Marcella Ucci, Alexi Marmot, Pippa Lally, and Abi Fisher. "Occupational physical activity habits of UK office workers: cross-sectional data from the active buildings study." *International journal of environmental research and public health*, **15**(6):1214, 2018.

[STD15] LN Sharma, RK Tripathy, and Samarendra Dandapat. "Multiscale energy and eigenspace approach to detection and localization of myocardial infarction." *IEEE Transactions on Biomedical Engineering*, **62**(7):1827–1837, 2015.

[Ste88] Thomas Stevens. "Physical activity and mental health in the United States and Canada: Evidence from four population surveys." *Preventive Medicine*, **17**:35–47, 01 1988.

[SVL14] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. "Sequence to sequence learning with neural networks." In *Advances in neural information processing systems*, pp. 3104–3112, 2014.

[SW09] Murray B Stein and John R Walker. *Triumph over shyness: Conquering social anxiety disorder*. Anxiety Disorders Association of America, 2009.

[Swa12] Melanie Swan. "Health 2050: The realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen." *Journal of personalized medicine*, **2**(3):93–118, 2012.

[SWM06] Trevor J Schraufnagel, Amy W Wagner, Jeanne Miranda, and Peter P Roy-Byrne. "Treating minority patients with depression and anxiety: what does the evidence tell us?" *General hospital psychiatry*, **28**(1):27–36, 2006.

[SXP20] Ajitesh Srivastava, Tianjian Xu, and Viktor K Prasanna. "Fast and Accurate Forecasting of COVID-19 Deaths Using the SIkJa Model." *arXiv preprint arXiv:2007.05180*, 2020.

[TBK19] Sarah Tinitali, Kelly-Ann Bowles, Jennifer L Keating, and Terry Haines. "Sitting posture during occupational driving causes low back pain; evidence-based position or dogma? A systematic review." *Human factors*, p. 0018720819871730, 2019.

[TDC15] Margarita Triguero-Mas, Payam Dadvand, Marta Cirach, David Martínez, Antonia Medina, Anna Mompart, Xavier Basagaña, Regina Gražulevičienė, and Mark J Nieuwenhuijsen. "Natural outdoor environments and mental and physical health: relationships and mechanisms." *Environment international*, **77**:35–41, 2015.

[TEW17] Maurice Topper, Paul MG Emmelkamp, Ed Watkins, and Thomas Ehring. "Prevention of anxiety disorders and depression by targeting excessive worry and

rumination in adolescents and young adults: a randomized controlled trial." *Behaviour research and therapy*, **90**:123–136, 2017.

[Tex17] Texas A&M Engineering. "Veterans take wearable PTSD monitor on test ride." 2017.

[TLA87] Daniel Teres, Stanley Lemeshow, JILL SPITZ Avrunin, and HARRIS Pastides. "Validation of the mortality prediction model for ICU patients." *Critical care medicine*, **15**(3):208–213, 1987.

[TLQ18] Nelson Wai-Hung Tsang, Kam-Yiu Lam, Umair M Qureshi, Joseph Kee-Yin Ng, Ioannis Papavasileiou, and Song Han. "Indoor Activity Tracking for Elderly Using Intelligent Sensors." In *Intelligent Data Sensing and Processing for Health and Well-Being Applications*, pp. 197–222. Elsevier, 2018.

[TMB19] Muhammad Tariq, Hammad Majeed, Mirza Omer Beg, Farrukh Aslam Khan, and Abdelouahid Derhab. "Accurate detection of sitting posture activities in a secure IoT based assisted living environment." *Future Generation Computer Systems*, **92**:745–757, 2019.

[TZ15] Naftali Tishby and Noga Zaslavsky. "Deep learning and the information bottleneck principle." In *2015 IEEE Information Theory Workshop (ITW)*, pp. 1–5. IEEE, 2015.

[usf] "US Facts Dataset." `https://usafacts.org/visualizations/coronavirus-covid-19-spread-map`, archived at `https://archive.is/tt0ih`. Accessed: 2020-07-30.

[VFB15] Thomas Vilarinho, Babak Farshchian, Daniel Gloppestad Bajer, Ole Halvor Dahl, Iver Egge, Sondre Steinsland Hegdal, Andreas Lønes, Johan N Slettevold, and Sam Mathias Weggersen. "A combined smartphone and smartwatch fall detection system." In *2015 IEEE international conference on computer and information technology; ubiquitous computing and communications; dependable, autonomic and secure computing; pervasive intelligence and computing*, pp. 1443–1448. IEEE, 2015.

[VLB08] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. "Extracting and composing robust features with denoising autoencoders." In *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103. ACM, 2008.

[VSP17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." In *Advances in neural information processing systems*, pp. 5998–6008, 2017.

[WCC20] Virginie Woisard, Mireille Costes, Hélène Colineaux, and Benoit Lepage. "How a personalised transportable folding device for seating impacts dysphagia." *European Archives of Oto-Rhino-Laryngology*, **277**(1):179–188, 2020.

[WCY19] Chi-Chih Wu, Chuang-Chien Chiu, and Chun-Yu Yeh. "Development of wearable posture monitoring system for dynamic assessment of sitting posture." *Physical and Engineering Sciences in Medicine*, pp. 1–17, 2019.

[web]

[WEG87] Svante Wold, Kim Esbensen, and Paul Geladi. "Principal component analysis." *Chemometrics and intelligent laboratory systems*, **2**(1-3):37–52, 1987.

[WF01] Lianne J Woodward and David M Fergusson. "Life course outcomes of young people with anxiety disorders in adolescence." *Journal of the American Academy of Child & Adolescent Psychiatry*, **40**(9):1086–1093, 2001.

[WGL20] Guannan Wang, Zhiling Gu, Xinyi Li, Shan Yu, Myungjin Kim, Yueying Wang, Lei Gao, and Li Wang. "Comparing and Integrating US COVID-19 Daily Data from Multiple Sources: A County-Level Dataset with Local Characteristics." *arXiv preprint arXiv:2006.01333*, 2020.

[WKW16] SL Wagner, C Koehn, MI White, HG Harder, IZ Schultz, Kelly Williams-Whitt, O Warje, CE Dionne, M Koehoorn, R Pasca, et al. "Mental health interventions in the workplace and work outcomes: a best-evidence synthesis of systematic reviews." *Int J Occup Environ Med (The IJOEM)*, **7**(1 January):607–1, 2016.

[WRK14] Daryl J Wile, Ranjit Ranawaya, and Zelma HT Kiss. "Smart watch accelerometry for analysis and diagnosis of tremor." *Journal of neuroscience methods*, **230**:1–4, 2014.

[WWG20] Li Wang, Guannan Wang, Lei Gao, Xinyi Li, Shan Yu, Myungjin Kim, Yueying Wang, and Zhiling Gu. "Spatiotemporal dynamics, nowcasting and forecasting of COVID-19 in the United States." *arXiv preprint arXiv:2004.14103*, 2020.

[XCY19] Yifeng Xu, Juan Chen, Qiaoning Yang, and Qing Guo. "Human Posture Recognition and fall detection Using Kinect V2 Camera." In *2019 Chinese Control Conference (CCC)*, pp. 8488–8493. IEEE, 2019.

[YHZ17] Shuochao Yao, Shaohan Hu, Yiran Zhao, Aston Zhang, and Tarek Abdelzaher. "Deepsense: A unified deep learning framework for time-series mobile sensing data processing." In *Proceedings of the 26th International Conference on World Wide Web*, pp. 351–360. International World Wide Web Conferences Steering Committee, 2017.

[YMC17] Leiyue Yao, Weidong Min, and Hao Cui. "A New Kinect Approach to Judge Unhealthy Sitting Posture Based on Neck Angle and Torso Angle." In *International Conference on Image and Graphics*, pp. 340–350. Springer, 2017.

[YSH20] Tong Yang, Kai Shen, Sixuan He, Enyu Li, Peter Sun, Lin Zuo, Jiayue Hu, Yiwen Mo, Weiwei Zhang, Pingying Chen, et al. "CovidNet: To Bring the Data Transparency in Era of COVID-19." *arXiv preprint arXiv:2005.10948*, 2020.

[YWC08] Jhun-Ying Yang, Jeen-Shing Wang, and Yen-Ping Chen. "Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers." *Pattern Recognition Letters*, **29**(16):2213–2220, 2008.

[ZCD17] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. "mixup: Beyond empirical risk minimization." *arXiv preprint arXiv:1710.09412*, 2017.

[ZHL19] Yangyang Zhang, Ying Huang, Baisheng Lu, Yuanming Ma, Jihong Qiu, Yunong Zhao, Xiaohui Guo, Caixia Liu, Ping Liu, and Yugang Zhang. "Real-time sitting behavior tracking and analysis for rectification of sitting habits by strain sensor-based flexible data bands." *Measurement Science and Technology*, 2019.

[ZKH18] Jinghe Zhang, Kamran Kowsari, James H Harrison, Jennifer M Lobo, and Laura E Barnes. "Patient2vec: A personalized interpretable deep representation of the longitudinal electronic health record." *IEEE Access*, **6**:65333–65346, 2018.

[ZMC17] Mark Zimmerman, Jacob Martin, Heather Clark, Patrick McGonigal, Lauren Harris, and Carolina Guzman Holst. "Measuring anxiety in depressed patients: A comparison of the Hamilton anxiety rating scale and the DSM-5 Anxious Distress Specifier Interview." *Journal of psychiatric research*, **93**:59–63, 2017.