2023

Data-driven Strategies for Pain Management in Patients with Sickle Cell Disease

Swati Padhee
Wright State University

DATA-DRIVEN STRATEGIES FOR PAIN MANAGEMENT IN PATIENTS WITH SICKLE CELL DISEASE

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

By

SWATI PADHEE
M.S., Wright State University, USA, 2019.
M.Tech, Veer Surendra Sai University of Technology, India, 2013.
B.Tech., Biju Patnaik University of Technology, India, 2010

2023
Wright State University

WRIGHT STATE UNIVERSITY

COLLEGE OF GRADUATE PROGRAMS AND HONORS STUDIES

February 7, 2023

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Swati Padhee ENTITLED Data-driven Strategies for Pain Management in Patients with Sickle Cell Disease BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

_____

Tanvi Banerjee, Ph.D.
Dissertation Director

_____

Thomas Wischgoll, Ph.D.
Director, Computer Science and
Engineering Ph.D. Program.

_____

Shu Schiller, Ph.D.
Interim Dean, College of
Graduate Programs & Honors Studies

Committee on Final Examination:

_____

Tanvi Banerjee, Ph.D.

_____

Krishnaprasad Thirunarayan, Ph.D.

_____

Michael Raymer, Ph.D.

_____

Nirmish Shah, M.D.

Abstract

Padhee, Swati. Ph.D.,Department of Computer Science and Engineering, Wright State University, 2023. Data-driven Strategies For Pain Management in Patients with Sickle Cell Disease.

This research explores data-driven AI techniques to extract insights from relevant medical data for pain management in patients with Sickle Cell Disease (SCD). SCD is an inherited red blood cell disorder that can cause a multitude of complications throughout an individual's life. Most patients with SCD experience repeated, unpredictable episodes of severe pain. Arguably, the most challenging aspect of treating pain episodes in SCD is assessing and interpreting the patient's pain intensity level due to the subjective nature of pain. In this study, we leverage multiple data-driven AI techniques to improve pain management in patients with SCD. The proposed approaches have been evaluated on physiological, medicinal and pain measurements collected from Electronic Health Records (EHRs), demonstrating their ability to digitize the medical essence of patients, thereby assisting in multiple aspects of clinical decision making in pain management. First, we propose to explore the feasibility of estimating subjective pain from objective physiological signals collected from EHRs irrespective of the nature of hospital visits in large patient cohorts. Second, we propose to learn deep feature representations of the subjective pain trajectories from objective physiological signals collected from EHRs. Third, we propose to learn future pain from historical patient EHR data using time-series forecasting methods. Our initial results indicate promise in pursuing each of these three efforts, and our study can be a valuable addition to ongoing studies that utilize EHR data to help providers better understand and design real-time pain management strategies.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Electronic Health Records (EHR) 3

2 Background and Related Work **6**

2.1 Machine Learning in Healthcare 6

2.2 Machine Learning for Sickle Cell Disease 9

2.3 Machine Learning in Pain Medicine 14

2.4 Deep Learning in Healthcare 18

2.5 Forecasting in Healthcare : Sickle Cell Disease 20

2.5.1 Pain Forecasting 22

3 Pain Assessment from Physiological Signals in Electronic Health Records **25**

3.1 Overview 25

3.2 Material and Methods 25

3.2.1 Data 25

3.2.2 Type of Hospital Visit: 27

3.2.3 Pain Prediction 28

3.3 Results And Discussion 30

3.3.1 Intra-individual Pain Prediction 30

3.3.2 Inter-individual Pain Prediction 31

3.3.3 Pain Change Prediction 32

3.4 Summary 34

4 Pain Assessment from Physiological Signals and Medication Data in Electronic Health Records 35

4.1 Overview 35

4.2 Materials and Methods 36

4.2.1 Data 36

4.2.2 Variational Autoencoders 36

4.2.3 Proposed Framework 39

4.2.4 Experimental Study 41

4.3 Results and Discussion 42

4.3.1 Visualization 47

4.4 Summary 50

5 Pain Forecasting from Physiological Signals in Electronic Health Records 52

5.1 Overview . . . 52

5.2 Materials and Methods . . . 54

5.2.1 Data . . . 55

5.2.2 Models . . . 57

5.3 Experiments . . . 64

5.3.1 Autoregressive Integrated Moving Average model . . . 64

5.3.2 Multilayer Perceptron and Long Short-Term Memory models . . . 64

5.3.3 Contrastive Predictive Coding (CPC) and Variational Autoencoder (VAE) . . . 65

5.3.4 Predictive Clustering . . . 66

5.4 Results . . . 66

5.5 Discussion . . . 68

5.5.1 Changing Patient Phenotype over time : Long-term forecasting . . . 69

5.6 Summary . . . 71

6 Conclusion and Future Work 73

6.0.1 Conclusion . . . 73

6.0.2 Future Work . . . 74

REFERENCES **75**

List of Figures

1.1 Figure (A) shows normal red blood cells flowing freely through a blood vessel. The inset shows a cross-section of a normal red blood cell with normal haemoglobin. Figure (B) shows abnormal, sickled red blood cells sticking at the branching point in a blood vessel. The inset image shows a cross-section of a sickle cell with long polymerized sickle haemoglobin (HbS) strands stretching and distorting the cell shape to look like a crescent moon. 2

1.2 Overview of the proposed research 4

3.1 Distribution of visits for 50 patients. (Study Patient identifiers 17, 37 and 50 are absent.) 26

3.2 Pearson correlation of six vital signs and visit types in the original dataset. 27

3.3 Intra-individual pain prediction accuracy results on vital signs data. 29

3.4 Intra-individual pain prediction accuracy results on vital signs and visit information. 29

4.1 An illustration of VAE architecture for one input data modality. 38

4.2 Deep representation learning for Pain Prediction 39

4.3 Confusion Matrix for best performing model with original data representations for 2 pain score levels (Pain Scores:No/mild: 0–5, Severe: 6–10) 43

4.4 Confusion Matrix for best performing model with VAE data representations for 2 pain score levels (Pain Scores:No/mild: 0–5, Severe: 6–10) 43

4.5 Confusion Matrix for best performing model with original data representations for 4 pain score levels (Pain Scores:None: 0,Mild: 1–3,Moderate: 4–6 ,Severe: 7–10) 44

4.6 Confusion Matrix for best performing model with VAE data representations for 4 pain score levels (Pain Scores:None: 0,Mild: 1–3,Moderate: 4–6 ,Severe: 7–10) 44

4.7 Qualitative comparison of the learned representations using t-SNE projections. 48

4.8 Distribution of medication dosage with pain score for sample patients with high and moderate mean pain intensity. 49

5.1 An illustration of our (real-time) clustering procedure. Here, a new patient is assigned over time to one of the three phenotypes (purple, blue, pink) based on the expected future pain score – as new observations are collected. 54

5.2 Initial Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots showing significant lag at 5. 58

5.3 Final model Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots showing no significant lag. 58

5.4 An illustration of an LSTM memory cell. 60

5.5 The architecture of Contrastive Predictive Coding (CPC) network illustrating using four time steps in the past sequence $(x_{t-3}, x_{t-2}, x_{t-1}, x_t)$ to predict a sequence of four time steps $(x_{t+1}, x_{t+2}, x_{t+3}, x_{t+4})$ in the future. 61

5.6 An illustration of run-time examples of our clustering based long term pain forecasting on six representative patients. 70

5.7 Change in systolic blood pressure (mean) across years for the six representative patients. 71

List of Tables

3.1 Intra-individual pain prediction results (accuracy) 28

3.2 Inter-individual pain prediction results (accuracy) 33

3.3 Inter-individual pain prediction results with varying pain scales on vitals data (accuracy) arranged from higher resolution to lower resolution. (6 Pain Scores: None:0, Very mild:1-2,Mild: 3–4, Moderate: 5–6, Severe: 7–8, Very severe:9-10; 4 Pain Scores: None: 0,Mild: 1–3,Moderate: 4–6 ,Severe: 7–10; 2 Pain Scores:No/mild Pain: 0–5, Severe Pain: 6–10) 33

3.4 Pain change prediction results (accuracy) 33

4.1 Data modalities and variables considered in this study. 37

4.2 Pain prediction results with varying pain scales on vitals data (accuracy) arranged from higher resolution to lower resolution. 42

4.3 Pain prediction results with varying pain scales on vitals data (accuracy) as compared to additional medication data arranged from higher resolution to lower resolution. 43

4.4 Sample of patient profiles from the learned VAE representations clustered together using t-SNE projections as shown in Figure 7 (Region 1- High Pain, Region 2,3 - Moderate Pain, Region 4- Low Pain) 50

5.1 Percentage of missing data before interpolation (raw data) and after interpolation. 56

5.2 Pain forecasting results for individualised patient model 67

5.3 Pain forecasting results for mixed patient model 67

5.4 Predictive Clustering for Long-term pain forecasting (AUROC) 68

5.5 Average pain scores of six representative patients for case study 70

Acknowledgement

As I wrote this acknowledgment, my 20-month-old daughter let out a long wail from her bed. I winced and waited a moment to pray that she would gently fall back asleep. Quiet soon settled in, and I breathed a sigh of relief. It takes an entire village to raise a Ph.D. researcher. These words define my Ph.D. journey. I am grateful to have several people to guide me, mentor me, laugh with me, listen to my woes, and share a smile during challenging periods. Thank you for being my support village during this beautiful Ph.D. journey! It's going to be a long night writing this acknowledgment. This has been my life during the past few years; it has felt like a juggling act where someone throws in some extra items to add to the mix of things I am trying to graciously balance.

First, I want to acknowledge my advisor, Dr.Tanvi Banerjee. Thank you for your dedication, commitment, guidance, and mentorship during my scientific journey. You have always believed in me and my research (even when I often did not). Thank you for the several email exchanges, Webex and Zoom conversations, and hours-long in-person conversations to appreciate my research strengths and patiently encourage me in my weaker areas. Graduate school can be a challenging, draining experience. I am proud to say my experience with you has been intellectually exciting and fun. Thank you for making me learn not just technical skills but an effective way of presenting them with good communication skills. While technical skills are part of the learning, emotional support is essential to a Ph.D. journey. I am short of words while thanking Dr.Banerjee for being a fellow doctoral mom and my biggest support during my pregnancy to post-partum times. I found a lifelong mentor and friend in her and will always seek her advice.

Thank Dr.Krinaprasad Thirunarayan for always opening his door and encouraging me to improvise my experiments and writing. My role as a wife and a mother has always taken precedence over my role as a Ph.D. researcher. This has been challenging as a woman in

her thirties when most students I encounter in my field are much younger and have limited family responsibilities. I have been at crossroads, and thank you for listening to me for hours and hours, guiding me, believing in me, and showing me the light at the end of the tunnels. I want to thank Dr.Michael Raymer, the first faculty I met at Wright State University before I started my Masters program and showed interest in pursuing research. Thank you for encouraging me to apply for the Ph.D. program and sharing your experience with paper rejection and ultimately getting the best paper awards. Thank you for your guidance at the end of every semester in this long journey following my trajectory to keep me on track. I am grateful to Dr.Nirmish Shah for patiently answering my long list of silly questions regarding pain and sickle cell disease. It has given me the ultimate satisfaction hearing that one word "absolutely" from you whenever I have asked you if my problem statement, approach, or results made sense clinically.

At different stages of my Ph.D. journey, I have received mentorship from several other faculty members across multiple collaborations. I want to acknowledge Dr.Amit Sheth for officially admitting me into the Ph.D. program at Wright State University. Thank you for showing confidence in my transition from Electrical Engineering to Computer Science. Thank you for encouraging me to apply for all those travel awards, resulting in traveling across continents and getting international recognition for my research. I thank Dr. Daniel M. Abrams for helping me improve all my papers for this dissertation project. I would also like to thank Dr.Valerie Shalin for introducing me to interdisciplinary research in cognitive psychology and showing confidence in me during challenging times. I am grateful to Dr. William Romine for questioning my empirical analysis and improving my work. I thank Dr.Zifeng Yang for pushing me to the edge and encouraging me to publish our interdisciplinary work on fluid dynamics.

I have also been blessed to have excellent mentors. I want to thank my first mentor, Dr. Sarasi Lalithsena, whom I met when I started working at Kno.e.sis. Research Center at

Wright State University. I remember I asked you on your defense day how you did it, and you told me I could do it too. I kept reminding myself of those words all throughout. Dr.Pavan Kapanipathi, thank you for your support and encouragement when I needed it the most, even though I could not complete the work I started with you. I met Dr. Maulik Kamdar and Dr. Prashant Khare in Austria at ISWC 2018. I can not thank them enough for continuously guiding and checking on me. Dr. Manas Gaur, thank you for being the much-needed fun mentor cum friend. I thank Dr. Alejandro (Alex) James, Dr. Tanay K Saha, Ian Micir, and Dr. Kimberly Swygert for mentoring me during my two successful internships and publications. Ph.D. research is not possible without the support of the administrative staff. I acknowledge Tonya Davis, Jennifer Limoli, Wendy Chetcuti, and Gwana C. Snell for always being on top of things and ensuring that my stipends are processed regularly.

I made amazing friends during my Ph.D. I want to thank all of you - Ankita Saxena, Revathy, Lakshika, Ankita Agarwal, RoopTeja, Venkatesh, Pavan Kalyan, Ruwan, Thilini, Dipesh, Amanuel, and Goonmeet. Revathy and Ankita Agarwal - I see part of my struggle in your journeys and have benefitted a lot in discussing how to handle certain situations. I want to shout out to three special friends cum mentors without whom my Ph.D. journey would have been very dull - Dr. Sanjaya Wijeratne, Dr. Shreyansh Bhatt, and Dr. Soonjye Kho. Thank you for being my unofficial therapists, counselors, and cheerleaders during this journey. Particularly, Shreyansh, thank you for teaching me how to talk to certain people in specific scenarios.

Finally, a special thanks to my friends and family. I want to start with my sister-in-law Shruti. Thank you for being the best gift I got from my marriage. You have supported me consistently, encouraging me at every step and believing in me. There has not been an exam all these years that went by without me panic calling you and you patiently telling me the words I wanted to hear. You, girl, know how to handle the emotional and crazy sensitive

part of me. I want to acknowledge the support of my friend cum brother Nippon who takes pride in everything I do since childhood. I appreciate you always speaking your truth and telling me things I don't want to hear but need to hear. It may not be a pleasure to hear, but it's always helpful. This journey would not have been as beautiful as it turned out to be without the hours long gossips with my undergrad friends Manisha, Swagatika, Fariha, and Madhu. Thanks for all of the laughs, advice, and sweet (and naughty) memories throughout time.

My husband Som - I am truly grateful for your love, support, and patience. Without you, I would not have been able to thrive in my Ph.D. program or balance my research with everything else. It was not easy to be a doctoral mom. Furthermore, being an international student from a traditional Indian family is even more complicated. Thank you for shielding me from all those who believe that women should do a stable job and care for their children, husband, and parents once they get married. Thank you for commuting more than 70 miles one way for four long years in snow and rain on the same boring route between Columbus and Dayton. But I am glad I get to enjoy driving now that you are so done! My daughter Sara- thank you for waiting for all these years to join us in this world while I got some publications out. Thank you for supporting me with a no-nausea, no-sickness pregnancy and following an extremely timely schedule afterward. My parents Souri Narayan Padhee and Anandini Behera, thank you for being the driving force behind everything I do. Thank you for putting up with my mood swings always. For all those who think I am very calm and quiet, please never say this to my parents! From tricking me into playing word puzzles and oral Mathematics games during power cuts in India during my childhood years to reviewing each of my papers and critiquing them, thank you for planting the seeds of creative thinking in me. You are where I found peace of mind, zest, and inspiration for life. Both of you have waited even longer for this day than I have. To take a quote from Albert Schweitzer, "At times, our own light goes out and is rekindled by a spark from another person. Each of us has cause to think with deep gratitude of those who have lighted the flame within us." You,

Mom (Maa), and Dad (Bapa) have been that spark for me when my light blew out. Thank you for your unwavering love and support along my journey. I love you both always and forever.

Dedicated to my parents

Anandini Behera & Souri Narayan Padhee

1 Introduction

Thesis Statement:

This research explores data-driven AI techniques to extract insights from relevant medical data for pain management in patients with Sickle Cell Disease (SCD). This can be done in a stepwise manner: (i) Predict subjective pain from objective physiological signals collected from Electronic Health Records(EHRs) irrespective of the nature of hospital visits, (ii) Predict subjective pain from objective physiological signals and medication data collected from Electronic Health Records(EHRs) using self-supervised learning, (iii) Forecasting future pain based on past data from Electronic Health Records(EHRs)?

1.1 Motivation

Sickle cell disease (SCD) is the most typical inherited blood disorder, affecting millions of people worldwide. The production of an altered type of hemoglobin characterizes it. The altered hemoglobin deoxygenates while passing through blood vessels, polymerizes, and becomes fibrous, causing the red blood cells to become rigid and change their shape to sickle-shaped. The altered red blood cells can occlude blood vessels, a phenomenon known as vaso-occlusion, resulting in a lack of oxygen to tissues and thereby causing pain [1]. Most patients with SCD experience repeated, unpredictable episodes of severe pain. These pain episodes are the leading cause of emergency department visits and may last for as long as several weeks. Arguably, the most challenging aspect of treating pain episodes in SCD is assessing and interpreting the patient's pain intensity level.

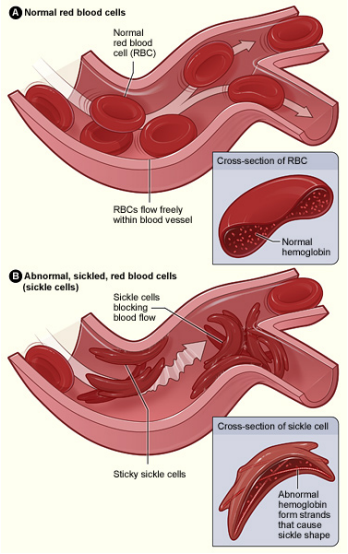

Figure 1.1: Figure (A) shows normal red blood cells flowing freely through a blood vessel. The inset shows a cross-section of a normal red blood cell with normal haemoglobin. Figure (B) shows abnormal, sickled red blood cells sticking at the branching point in a blood vessel. The inset image shows a cross-section of a sickle cell with long polymerized sickle haemoglobin (HbS) strands stretching and distorting the cell shape to look like a crescent moon.

However, in current clinical practice, a patient self-report is the gold standard approach for determining the absence, presence, and intensity of pain. Due to the subjective nature of pain, it becomes challenging for clinicians to ascertain the severity of the patient's pain precisely. Besides, effective treatment is palliative, including intravenous opioid therapy. While these self-described pain intensity levels provide important clinical reference indicators and have been proven to help treat patients suffering from pain in most situations [2], it might have challenges when applied to certain vulnerable populations.

Current clinical guidelines recommend frequent observations of vital signs during the assessment and treatment of painful episodes as they are an objective measurement of the essential physiological functions and are potential indicators for patients' subjective pain levels. It has been previously reported that Machine Learning (ML) techniques can be used to design objective pain assessment models using vital signs from inpatient EHR data.

1.2 Electronic Health Records (EHR)

Electronic health records (EHR) were initially designed to record for billing purposes rather than for research and quality improvement efforts. The utility of EHRs on quality healthcare delivery research has focused on physician performance, and billing precision [3] for quite some time. Studies using EHRs have focussed on process quality metrics, analyzing physician-level variability, and guideline compliance, and recently moving on to overall quality improvement or patient outcomes [4, 5, 6, 7, 8]. Analyzing EHR data has the potential to significantly decrease medical errors by providing better access to relevant information, improved communication and coordination of care among multiple providers and hospitalizations, and more efficient documentation and monitoring [9]. It has been proven that EHRs can be used to decrease prescribing errors by providing real-time clinical decision support [10, 11, 12, 13]. Further recent studies have shown that EHRs can be analyzed to improve the tracking and monitoring of adverse patient outcomes. For example, [14, 15] showed that patient safety outcomes could be improved by providing critical care in cases of urinary tract infections involving catheter usage, deep vein thrombosis, or pulmonary embolism. Overall, improvements in patient condition management and outcomes associated with EHR analysis still need to be well documented, although several studies have looked at changes in quality attributed to electronic healthcare systems. In particular, the effect of data extracted from a well-implemented EHR system on inpatient adverse events, such as pain management, the role of medication dosage for specific chronic medical conditions, and future patient condition prediction, has yet to be explored. We thus explore the association of hospital-level EHR systems with patient management in SCD to better understand the relationship between physiological signals, medicine dosages, and pain.

Specific research questions addressed in this dissertation as shown in Figure 1.2 are detailed as below:

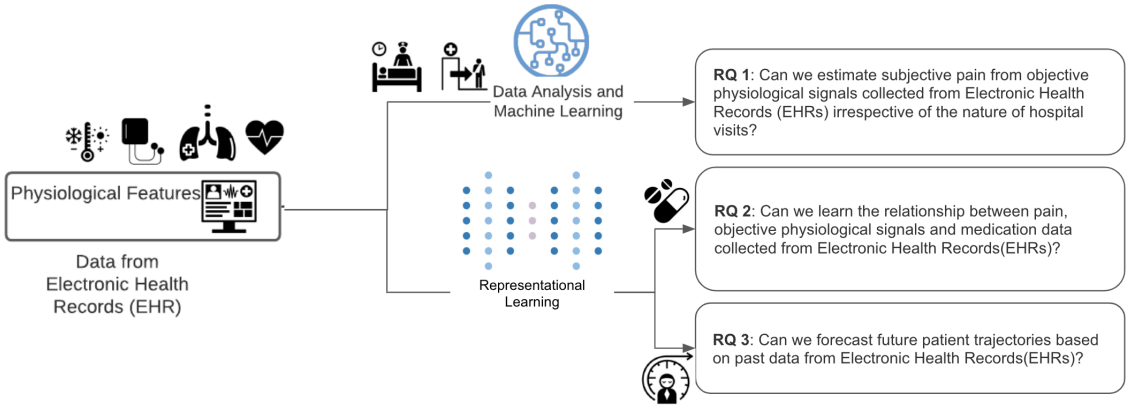

Figure 1.2: Overview of the proposed research

1. **Can we estimate subjective pain from objective physiological signals collected from Electronic Health Records (EHRs) irrespective of the nature of hospital visits?**

   We leveraged multiple machine learning algorithms on six physiological measures of patients with SCD to predict pain scores. We propose dealing with missing data and conducting a series of experiments at intra-individual and inter-individual levels. In this, our purpose is to examine whether, using objective physiological measurements, technology can predict subjective pain in SCD patients and may be generalizable for a larger cohort of patients.

2. **Can we learn the relationship between pain, objective physiological signals, and medication data collected from Electronic Health Records(EHRs)?**

   Deep learning has been successfully employed in an unsupervised manner on EHRs to achieve both specific and general goals [16]: for instance, “Deep Patient” [17] and “Doctor AI” [18] both used unsupervised deep learning before supervised learning. We propose to investigate the feasibility of deep feature representation for predictive modeling. In particular, we propose representing multiple data modalities in EHRs in high-level abstraction by utilizing deep auto-encoder networks such as variational

autoencoders (VAEs).

3. **Can we forecast future patient trajectories based on past data from Electronic Health Records(EHRs)?**

   Probabilistic models, which aim to provide the best assessment of future events, help map a patient's trajectory and forecast the likely course of disease [19]. A large body of evidence suggests that interventions are more precise and practical when individualized models are used to capture the status of a particular patient over time [20, 21]. These models can be used to forecast the progression of a disease and improve the effectiveness of interventions when applied in real-world practice. Forecasting is a sub-discipline of prediction, and generating predictions based on historical time-series data is forecasting. We propose to take our analysis one step deeper by considering the temporal dimension and building a model for simulating the clinical trajectories of patients experiencing episodes of pain over time. However, patient trajectories for pain might be random due to unobservable factors, such as the patient's environment and medical history. As such, accurate forecasting of the progression of clinical measures should be able to account for this random nature. Hence, to address this randomness, we propose to utilize a variational autoencoder (VAE) architecture for our disease progression model. VAE's disentangled latent space could allow for the development of models which incorporate not only prediction but also forecasting using the latent space's distilled information.

2 Background and Related Work

The International Association for the Study of Pain Committee (IASP) pointed out that "Pain is always subjective" [22]. The assessment and management of pain are always a challenge in both clinical practice and academic research. The current definition of pain implies that self-report is the gold standard for pain. However, in practice, clinicians usually integrate patients' self-report with their assessment of non-verbal behaviors, such as voice, body activity, and facial activity, to enhance understanding. Current clinical guidelines recommend considering vital signs during assessing and treating adverse events such as painful episodes. These physiological measurements include: blood pressure, respiratory rate, oxygen saturation, temperature, and pulse [23]. Due to the growing volume of clinical data and the requirement for highly accurate predictive models, machine-learning techniques have been increasingly utilized in pain studies in recent decades. Machine learning techniques allow efficient knowledge mining from high-volume data.

2.1 Machine Learning in Healthcare

The origin of artificial intelligence (AI) dates back to when Alan Turing proposed an experiment involving two players (either human or artificial) trying to convince a third human player that they are also humans [24]. AI is successful when the third player fails to identify the machine. Significant milestones in the development of machine learning (ML) include the first creation of the computer learning program, which was a checker game [25], and the first neural network called the perceptron [26]. Machine learning learns from data aiming at optimization and improved performance of an algorithm instead of analyzing the

probabilities of observations in the underlying data distribution.

While the applications of machine learning and artificial intelligence in medicine have their roots in the earliest days of the field [27], it is only recently that there has been a push towards recognizing the need to have healthcare solutions powered by these technologies. This has led researchers to suggest that it is only a matter of time before machine learning will be ubiquitous in healthcare [28]. Over the recent years, AI techniques have been sending significant waves across healthcare, even fuelling an extremely active discussion on whether AI doctors may eventually replace human physicians. While in the foreseeable future, human physicians can not be replaced by machines; AI can assist physicians in clinical decision-making and human judgment in specific functional areas of healthcare. The rising accessibility of healthcare data with rapid developments in big data analytic methods is empowering recent successful applications of AI in healthcare. These techniques can unlock clinically relevant information hidden within the massive amount of data guided by relevant questions, which can assist clinical decision-making [29, 30, 28]. Multiple applications of AI have been extensively discussed in the medical literature [31, 29, 32]. AI can use sophisticated algorithms to “learn" multiple features in a large volume of healthcare data and then utilize the derived insights to assist with clinical practice.

Furthermore, it can also be equipped with learning and self-correcting capabilities to improve its accuracy based on feedback. In the context of information extraction, an AI system can also provide up-to-date medical information from journals, textbooks, and clinical practices to physicians and assist them with informed proper patient care [33]. In addition, an AI system can contribute to reducing diagnostic and therapeutic errors that are inevitable in human clinical practice [29, 31, 34, 35, 36, 37]. Moreover, an AI system can extract useful information from large patient populations to assist in making real-time inferences for health risk alerts, and health outcome predictions [38, 39].

However, before AI systems can be deployed in healthcare applications, they need to be

“trained" through the data which is generated from multiple clinical activities, including screening, diagnosis, treatment, and so on, to learn similar groups of subjects, associations between subject features, and the outcomes of interest. These multimodal clinical data often exist in but are not limited to demographics, medical notes, electronic recordings from medical devices, physical examinations, clinical laboratory examinations, and images. For example, in the diagnosis stage, a substantial proportion of the AI literature analyses data from diagnosis imaging, genetic testing, and electrodiagnosis. Furthermore, physical examination notes and clinical laboratory results are the other two significant data sources comprising many unstructured narrative texts, such as clinical notes, which can not be directly analyzed in their original format. As a result, the corresponding AI applications focus on first converting the unstructured text to machine-understandable electronic health records (EHRs). Depending on the data, AI techniques mainly fall into two major categories. One category includes natural language processing (NLP) methods to retrieve information from unstructured data, such as clinical notes or external medical knowledge sources, to enrich further and supplement the structured medical data. These NLP procedures are used to convert texts to machine-readable structured data, which ML techniques can then analyze for problems of interest [40]. Another category includes machine learning (ML) techniques that analyze structured data such as imaging, genetic and electrophysical data. Both types of techniques can be used to cluster patients’ traits or infer the probability of the disease outcomes [41, 42].

Inputs to these ML algorithms include patient “traits" and sometimes the interested medical outcomes. A patient’s traits primarily include baseline data, such as age, gender, disease history, and so on, as well as disease-specific data, such as diagnostic imaging, gene expressions, physical examination results, clinical symptoms, medications, etc. Besides the traits, patients’ medical outcomes, such as disease indicators, patient’s response, and quantitative disease levels, for example, tumor sizes, are also collected as data. Depending on whether to include the desired outcomes or not, ML algorithms can be classified into

unsupervised learning and supervised learning. Unsupervised learning is commonly known for feature extraction, while supervised learning is known for predictive modeling by finding some relationships between the patient traits (as input) and the outcome of interest (as output). Unsupervised learning can also be used as part of the pre-processing steps, which makes the follow-up supervised learning step efficient. Relevant techniques include Linear Regression, Logistic Regression, Naïve Bayes, Decision Tree, Nearest Neighbour, Random Forest, Support Vector Machine (SVM), and Neural Networks [43].

Despite the increasingly rich AI literature in healthcare, the research mainly concentrates on a few disease types such as cancer, nervous system disease, and cardiovascular disease [44, 45, 46, 47, 48, 49]. Besides these major diseases, AI has been applied in other diseases as well such as asthma, congenital cataract disease and diabetes [50, 51, 52, 53, 54].

2.2 Machine Learning for Sickle Cell Disease

Current literature shows increased attention on machine learning techniques to understand various complexities associated with patient health in SCD. Lazakidou et. al. [55] developed a personal electronic health record (pEHR) to evaluate the deployment of an advanced web-based application platform that assessed healthcare professionals and patients to provide a more efficient and effective solution compared to daily clinical routine. In their research, the purpose of the web-based solutions is to enable patients to update and access their medical information. The system was examined with three various patient groups consisting of 150 patients who had Parkinson's disease, diabetes, and congenital heart disease that were engaged within three European clinics. The outcomes indicated that the pEHR could provide better services in terms of user-friendly, data management, comprehensiveness, and valuable content.

Du E et. al. [56] developed a microfluidic device that can examine the blood behavior

of SCD patients. This device also has the ability to measure how long blood cells take to become stiff and stuck in blood vessels. The authors claim that the future innovation of this device can easily prevent and predict vaso-occlusive crises. It could assist many researchers in testing the device's efficacy, which happens in about three hundred thousand new-borns per year, mainly in Africa. Twenty-five SCD patients were involved in their study; the researchers, by using this device to evaluate blood samples, were able to decide how deoxygenation affects red blood cells' (RBCs) sickling rates, capillaries stuck rates, how quickly the RBCs re-shape, especially, when oxygen levels are restored.

Knowlton et al.[57], present a sensitive, label-free, and specific testing platform to diagnose SCD using blood samples based on the density of sickle RBCs under deoxygenated circumstances. The Sickle Mobile Tester device was designed in an online application for computer-aided design (Tinker CAD). The platform was implemented with a compact 3D-printer and lightweight add-on installed on a commercial mobile phone. This attachment comprises an optical lens to illuminate the sample of RBCs. The sample collected from patients is suspended in a paramagnetic medium loaded in a microcapillary tube with sodium metabisulfite, which is inserted around the magnets. Eventually, using this model, they could differentiate between the levitation patterns of sickle versus control RBCs in association with their degree of confinement.

Shah et. al. [58] from our research group explored the receptiveness of SCD patients to use mobile applications (apps) that can mitigate the disease. There were two phases in their experiment. Phase one involved 100 patients who finished the task inquiring about their interest in communicating with healthcare providers and self-care management system using the mobile app. Phase 2 surveyed another 17 patients who were asked to test a newly developed SCD app, to report its utility and usability. In the outcomes of this survey, participants stated that the mobile app tested was effective and useful with 94%, 88% to track pain, and valuable for self-care management. In addition, all patients involved

in this experiment reported that the app was an effective tool for communicating with healthcare providers. Overall, this study recommended that patients with SCD, regardless of education or age, are agreeable to using technology to cope with their related pain and disease symptoms. Mobile apps could provide a suitable environment for SCD medical management.

Milton et al. [59] developed an ensemble model exploring 14 algorithms to predict Hemoglobin F (HbF) in patients associated with different configurations of Single Nucleotide Polymorphisms (SNPs). The goal of the study was the prediction of Haemoglobin F in patients suffering from SCD. A sample of 814 patients were involved in their experiments, for which a variety of blood features were measured, such as platelets and hemoglobin. The ensemble outcomes of classifiers labeled 23.4% of the variability in the discovery cohort, while the association between predicted and observed HbF in the three independent cohorts ranged between 0.28 and 0.44.

Allayous et al. [60] demonstrated the high risk of an acute splenic sequestration crisis, a severe symptom of SCD. Solanki [61] implemented two models, including Decision Trees (DT), to classify specific blood groups. In their research, the main aim was to learn how to predict the level of severity depending on the training dataset. The dataset including 15 features was collected from "Centre Caribéen de ladrépanocytose" over a period of 10 years for 42 children. The authors used multiple machine learning algorithms to evaluate the risk of acute splenic sequestration crisis in classifying patients between severe and mild symptoms with the highest accuracy of 92% achieved using the Adaboost algorithm.

Solanki et. al. [61] proposed data mining methods utilizing WEKA tools for patients with SCD. The research has used two classification methods comprising DT and RF, to compare them for classifying specific blood groups. The study's outcome declared that the RF algorithm is better to use than DT, in terms of classifying specific blood groups for individuals affected by SCD.

Artificial Neural Networks (ANNs) have been proposed as a connectionist approach to the classification and determination of medical results, including blood inflammations [62]. The model has been employed widely to automate the assessment of blood disorders such as SCD using morphological attributes of erythrocytes in the cell. Dalvi and Vernekar [63] developed an anemia detection model using statistical and ensemble learning methods to yield high accuracy in Red Blood Cell (RBC) classification. Their outcomes showed that stacking ensemble techniques achieved the highest accuracy. In their experiments, ANN provided the best outcomes in comparison to the K-Nearest Neighbour (KNN), which obtained poor results. They combined the KNN classifier and the Decision Tree classifier using stacked ensembles to obtain satisfactory results. The combination of various models is indicated as providing superior performance than that of individual models. The evaluation measures used in the study included Accuracy, Specificity, Sensitivity, and Precision, with 10-fold cross-validation used in their experiment. The training set comprised 441 instances, while the testing set comprised 49 cases. Sharma and Khullar [64] represented a comparative analysis between the fuzzy expert system and ANN for better efficiency in diagnosing sickle cell patients. The authors have summarised that the best model for diagnosing sickle cell Anaemia is ANN.

Escandell-Montero et al. [65] proposed an approach based on Reinforcement Learning (RL) for sickle cell anemia patients. Using a Markov Decision Processes (MDP) framework, RL was able to learn to discover optimal solutions using clinical datasets automatically. The RL technique applied in the proposed methodology is fitted Q iteration (FQI), which stands out for its capability to effectively and efficiently use data. In order to achieve high accuracy and performance in the medical data, FQI was combined with a function approximator constructed using regression trees to handle a continuous state space and to produce the learned policy applied to the cases not covered by the dataset. Thus, although prospective validation is needed, empirical studies have demonstrated the potential benefits of RL in SCD.

Khalaf et. al. [66] presented a system to examine patient data and provide a suitable amount of Hydroxycarbamide drugs/liquid for each patient. The datasets utilized in their experiments for SCD patients were collected within four years from the Alder Hay Hospital in the UK. Each sample involved 13 attributes deemed vital factors for predicting the SCD trait, i.e., Haemoglobin(Hb), Aspartate Aminotransferase, Mean Corpuscular Volume, Alanine aminotransferase, Neutrophils, Reticulocyte Count (A), Hb F, Platelets, Lactate dehydrogenase, Weight, Bilirubin, Body Bio Blood, and Reticulocyte Count. They also considered two additional attributes age and gender. They proposed a neural network (NN) consisting of a stacked model, where NNs are assembled into levels, and the outputs from one network are fed into the next along with desired outputs. Each level is comprised of two NNs, each with a hidden layer of 20 neurons. The model of NN was formulated with 10 input units. Their MLP method had the lowest error rates in terms of MSE, RMSE, MAE, and MAPE at 17887.55, 133.74, 90.20, and 0.1345, respectively. The authors extended the work to further classify the dosage of medication required for treating patients with SCD [67].

Prior studies have reported that fluctuations in vital signs can be used for assessing pain [68] as acute pain leads to changes in vital signs [69]. These physiological measures include blood pressure, respiratory rate, oxygen saturation, temperature, and pulse. From our research group, Yang et.al. [70] showed the feasibility of ML techniques on a limited dataset of 5363 records from 40 patients during inpatient hospital visits to predict subjective pain scores from six objective vital signs. Alambo et. al. [71] employed 424 clinical notes of the same cohort of 40 patients to assess the prevalence of pain in patients and whether pain increases, decreases, or stays constant. In this study, we investigate the generalizability of ML techniques for a broader group of people during inpatient, outpatient, and outpatient evaluation visits. As validated by our clinical collaborators in Section 3.2.2, we provide definitions of these visits. Specifically, we utilize five years of EHR data from 50 patients suffering from SCD to build pain prediction models using objective physiological measures

as features at both the intra-individual and inter-individual levels based on an 11-point numeric rating scale (NRS) [72]. We further investigate whether the variation in the hospital visit type affects our model performance.

2.3 Machine Learning in Pain Medicine

In the context of pain-related data, machine learning models are desired to learn a mapping of complex features to a particular pain phenotype class. After the machine has learned to predict a pain-related phenotype, the algorithm can then be used to predict the class membership of a new subject. Pain and pain chronification is unresolved medical problems that are not entirely understood and continue to have a high prevalence [73]. Pain has been accepted as a complex phenomenon [74, 75, 76]. Contemporary computational science methods utilize complex clinical and experimental data to understand pain's complexity better. Machine learning-based efforts include techniques to automatically detect hidden patterns in data and then use those uncovered patterns to predict or classify new/future data, to observe structures or subgroups within the data, or to extract relevant information from the data suitable to derive new insights [77, 78].

The primary classifiers provided by supervised machine learning are symbolic [79], or sub-symbolic [80]. For symbolic classifiers, the classification can be interpreted by a domain expert as a combination of conditions on the features. For example, a decision tree-based symbolic classifier was designed from approximately 30 acquired features, including demographic (age, sex, and weight), biomedical (e.g., blood pressure, diabetes, and arterial hypertension), surgery-related therapy and analgesic-related therapy parameters [81] to predict patient-controlled analgesia consumption. Sipila et. al. designed another symbolic classifier - a Bayesian diagnostic tool to predict the persistence of pain in breast cancer surgery from demographic, pain, and surgery-related parameters [82]. They showed a sensitivity and specificity of 33% and 95%, respectively. Their classification procedure can

be directly interpreted via the Bayesian decision limits calculated for the single parameters. However, in subsymbolic classifiers, the possibility of understanding the details is skipped, and better performance of a machine-learned algorithm is sought. Hence, it is challenging to provide biomedical explanations for the functioning of the algorithm. For instance, random forests utilize hundreds or thousands of simple decision trees that escape this interpretation, and the classification is obtained through the complete set of trees, the "forest" [83]. Braundmeier et. al. designed a classifier based on various stool-based markers to diagnose a bladder pain syndrome [84]. Similarly, clusters of patients with neuropathic pain from controls and then various types of neuropathy, such as peripheral neuropathy with or without pain and neuropathy associated with amyotrophic lateral sclerosis, were obtained with a projection method for high-dimensional data; specifically, minimum curvilinear embedding on complex proteomics data [85]. ML algorithms were further applied to predict thermal pain sensitivity from bioresponses acquired through electromyography, skin conductance level, and electrocardiography [86]. Subjects were instructed to determine the pain threshold: "Please press the stop button immediately when you experience a burning, stinging, piercing, or pulling sensation in addition to the feeling of heat." In order to determine the tolerance threshold, the following instruction was given: "Please press the stop button immediately when you can no longer tolerate the heat, taking into account the burning, stinging, piercing, or pulling sensation." Specifically, using support vector machines (SVMs), [87], individual pain threshold and tolerance to thermal stimulation were predicted from the noninvasive measurements at accuracies of $>91\%$ and 79%, respectively [88]. This aimed to draw insights about pain in subjects with verbal and cognitive impairments for whom pain-related queries, such as the standard visual rating scales, cannot be applied. This study is limited to finding classification results with high accuracy and an automated selected feature pattern of biopotentials representing "pain" and "no-pain", respectively. However, for better pain management strategies, it is essential to identify the severity of pain in addition to just having pain or no pain, which we address in our studies.

Kringel et. al. designed a subsymbolic classifier based on k-nearest neighbors calculations to predict which patients needed high opioid doses for analgesia, based on a next-generation sequencing–derived opioid receptor genotype[89]. Patients have different diseases underlying the pain, and the high opioid doses may be accidental as most patients were sent from University tertiary care centers where the physicians were more inclined to raise the opioid doses, whereas, in the periphery, the same patients might have been labeled as opioid resistant already at doses below 400 mg OME per day and therefore not included into this analysis. This analysis showed that opioid receptor genotyping, consistent with biological plausibility, has the potential to provide the desired predictively of particular (clinical) phenotypes as demonstrated with high opioid dose demands in pain patients.

Furthermore, Nickerson et. al. predicted pain scores between 40 to 120 minutes after administering 10 mg oxycodone from pain score values before drug administration using elastic net regression models and SVMs [90]. They extracted 200 features for each patient in the electronic medical records (EMR) data. Essential features included patient age, gender, Chalson comorbidity index, body mass index (BMI), ethnicity, and the International Classification of Diseases 9th edition (ICD9) code class. They applied LSTM to predict the next measured pain score after administering an analgesic drug and compared the results with more straightforward techniques. The Elastic Net model was found to be the top performer with an MSE score of 4.96; however, their data set was limited to postoperative pain scores, and thus, the model only had information on scores before and after drug administration. The authors concluded that with a more descriptive record collection and the inclusion of more temporal data (e.g., other vital signs), these results would likely improve, which we explore in our study by utilizing both vital signs and medication data to predict the severity of pain at varying levels.

Machine learning methods are designed on data-driven research approaches in contrast to classical statistical methods, where knowledge and presumptions about the distributions

and functional dependencies of the data are required. Also, feature selection in machine learning enables one to identify relevant modulators of pain-related outcomes in data-driven and hypothesis-free experimental research methods. For example, Sipila et. al. designed a Bayesian optimal prediction model (sensitivity 33%, specificity 95%), as discussed earlier, to show that among many biomedical parameters, demographic, psychological, and pain-related parameters are the most relevant to explain the persistence of pain among women who underwent breast cancer surgery [82]. They considered moderate to severe pain (NRS>=4 out of 10) as a clinically significant cutoff, as women who had undergone breast cancer surgery reported only some interference by pain in daily activities when pain intensity was estimated to be 1–3 out of 10. However, for better pain management in chronic pain episodes, we must include all the pain levels - which we explore in our work.

Besides, unsupervised machine learning methods can be used to assess whether the acquired biomedical parameters demonstrate the efficacy of a treatment provided, that is, to detect data structures congruent with a known pre-classification, such as the presence of a modulator of the pain phenotype. For example, after treatment of 82 subjects with local UV-B irradiation or capsaicin application and assessing the pain phenotype using 10 different QST parameters, a 246 * 10-sized data matrix was obtained in a human experimental pain study [91]. Using unsupervised machine learning implemented as emergent self-organizing maps [92], data structures were detected that coincided with applied known treatments indicating that modulation of the complex pain phenotype had been obtained [91]. A machine learning algorithm consisting of a classification and regression tree analysis was applied to 8034 independent observations of baseline thermal nociceptive sensitivity in mice [93]. The analysis identified the mouse genotype as predictive of the pain phenotype; however, it also revealed that the experimenter performing the test and additional laboratory factors, including season/humidity, cage density, time of day, sex, and within-cage or order of testing modulated the pain phenotypes [93].

Finally, natural language progressing methods [94], which combine linguistics with computer science to analyze human language in speech or written text, were used to extract signs from clinical notes using features such as the occurrence of terms, for example, keywords that hint at a clinically incident, in a document [95]. The authors created a tool that extracts blood pressure, heart rate, temperature, respiratory rate, blood oxygen saturation, and pain level from nursing and other clinical notes recorded during inpatient care to supplement structured vital sign data. Prediction precision of this method validated using 1,000 manually annotated documents for extracting the patient's pain level was reported to be better than 99%. Our work differs from this study as we propose to predict self-reported pain severity based on vital signs and medication data from EHRs for patients with SCD. Due to data availability constraints, we do not include the clinical notes for the current scope of work. When available, our work can be extended to utilize similar natural language processing techniques to augment the information from EHRs.

2.4 Deep Learning in Healthcare

Deep learning is a subset of machine learning extending the classical neural networks with three or more layers. The rapid development of modern computing enables deep learning to build neural networks with many layers. This is infeasible for classical neural networks. Hence, deep learning can analyze highly complex non-linear patterns in large datasets. As the accessibility to large and complex data is increasing, deep learning techniques are gaining popularity [96]. However, most deep learning techniques are used for image processing tasks, which makes sense, provided that visual data can be naturally complex and of high volume. In such cases, unlike deep learning models using more hidden layers than the classical neural networks can handle the complex structures [43].

In medical applications, the commonly used deep learning algorithms include convolution neural network (CNN), recurrent neural network, deep belief network, and deep neural

network. Recently, CNN has been successfully implemented in the medical field to diagnose disease. Long et. al. used it to diagnose congenital cataract disease through learning the visual images [50] and showed over 90% accuracy on diagnosis and treatment suggestion. Esteva et. al. utilized CNN to identify skin cancer from clinical images [46]. The proportions of correctly predicted malignant lesions (i.e., sensitivity) and benign lesions (i.e., specificity) are both over 90%, indicating CNN's superior performance. Gulshan et. al. applied CNN to detect referable diabetic retinopathy through retinal fundus photographs [53]. They reported over 90% sensitivity and specificity of the algorithm, demonstrating the effectiveness of using the technique in diagnosing diabetes. Chen et al.[97] designed an approach to detect hemolytic anemia. They employed concavity information to isolate the overlapping cells and used different classifiers to classify different types of hemolytic anemia. Acharya et al. [98] split up the erythrocytes from other blood components using an image processing method and then classified them into 11 types using the K-Medoids algorithm.

Elsalamony [99] detected two types of anemia using the shape signature technique and then classified them using a neural network. Albayrak et al. [100] implemented a Circular Hough Transform (CHT) to distinguish between healthy and sickle cells, where a distance tool is utilized to determine the radius range of the cells. Chy et al. [101] extracted features such as ratio, entropy, mean, standard deviation, and variance, which were used to train SVM to classify normal and sickle cells. Chy et al. [102] have employed Extreme Learning Machine (ELM), Support Vector Machine (SVM), and K Nearest Neighbor (KNN) classifiers to classify normal verse sickle cells. Recently, Convolutional Neural Networks (CNNs) have been used to classify RBC and diagnose sickle cell disease [103] with accuracies of 86.34% and 87.50%, respectively. These accuracies are still not the goal for the sickle cell disease classification task since precise classification is the first step toward accurate diagnosis. Alzubaidi, Laith, et al. [104] proposed three deep learning models to classify erythrocytes in three classes, namely: circular (normal), elongated (sickle cells), and other

blood content. Their model obtained state-of-the-art performance by achieving an accuracy of 99.98% with a multiclass SVM classifier on the erythrocytesIDB dataset, which has images of peripheral blood smears samples taken from patients with Sickle Cell Disease in the Special Hematology Department of the General Hospital from Santiago de Cuba [105].

Although the emerging discipline of computational pain research provides contemporary tools to understand pain, it uses computer-based processing of complex pain-related data and relies on "intelligent" learning algorithms. The application of deep learning for pain research–related non-imaging problems is limited by the availability and quality of data. In this proposed study, we explore deep-learning methods to predict the severity of self-reported subjective pain scores from objective vital signs and medication data for patients with SCD. Our work can be utilized to acquaint pain domain experts with the methods and current applications of machine learning in pain research, possibly facilitating the awareness of the methods in current and future projects.

2.5 Forecasting in Healthcare : Sickle Cell Disease

Time series forecasting can be defined as predicting future events based on prior knowledge acquired through a systematic process or intuition [106, 107]. The early works on utilizing forecasting for healthcare problems date back to Hippocrates of Cos (460 BC–370 BC), who studied the natural history of diseases and their primary environmental sources (including food and water) [108], and believed that prognosis was an essential part of medical treatment because by forecasting disease outcome, the physician established his expertise for treating the patient [109]. He developed and forecasted the occurrence of many diseases and conditions. One of the classical terms in medicine, 'Hippocratic facies,' describes the procedure for impending forecasting death based on the observation of distinctive signs and symptoms that he identified [110]. Forecasting has advanced over time and has increased in sophistication in many specialized areas, including healthcare [111, 112, 113, 114, 115,

116]. Data-driven approaches to personalized medicine have the potential to improve patient outcomes while minimizing costs and reducing levels of risk to the patient. Probabilistic models, which aim to provide the best assessment of future events, help map a patient's trajectory and forecast their likely course of disease [19]. A large body of evidence suggests that interventions are more precise and practical when individualized mathematical models are used to capture the status of a particular subject over time [117, 20]. These mathematical models can forecast the progression of the disease and may improve the effectiveness of interventions when applied in real-world practice.

Time series forecasting is a vast field of science, and it can be further categorized as short-term and long-term forecasting. Short-term forecasting can be utilized for intensive analysis and calculations of the underlying characteristics and variations of the time series to provide a robust and precise prediction of the future up to hours ahead of time [118]. In contrast, long-term prediction generally analyses the trend of the available data and the effect of the associated parameters to provide estimates for years in the future [119, 120]. As the technique requires tremendous analysis and calculations, short-term forecasting techniques are not used for long-term prediction. Because of their differentiated abilities, their potential can be applied in different clinical situations. Short-term forecasting, for example, is extremely useful in assessing patients' mortality in emergency care units where immediate action is crucial, allowing doctors to respond immediately before a vital situation can take place [121, 122]. Meanwhile, long-term forecasting thrives at assessing health conditions for many years after hospital discharge taking into account the effects of different types of treatment and the associated risks, thus allowing doctors to provide timely healthcare services for these patients [119, 120].

A growing body of work is utilizing physiological data to design forecasting models for healthcare problems. Ghassemi et al. [123] proposed a multi-task Gaussian process (GP) model to forecast patients' severity of illness using noisy, incomplete, sparse, heterogeneous,

and unevenly sampled clinical data, including physiological signals as well as clinical notes with a root mean squared error score of 0.69. ElMoaqet et al. [124] presented a framework to forecast the near future patterns of continuously monitored physiological signals based on auto-regressive (AR) models with a focus on predictions of critical levels of abnormality. Lee and Hauskrecht [125] proposed a time-series model based on the long short-term memory architecture (LSTM) to predict multiple future clinical events on electronic health record (EHRs) data derived from the MIMIC-III dataset. Their LSTM model relied on two sources of information to predict future events. One source was derived from the set of recently observed clinical events (medication, lab, procedure, and physiological events). The other one was based on the hidden state space defined by the LSTM to abstract past, more distant patient information that is predictive of future events. Fox et al. [126] explored various methods to predict blood glucose trajectories achieving a 5.31 absolute percentage error (APE) in predicting future blood glucose levels using an autoregressive multi-output forecaster model. Lim et al. [127] introduced a model based on recurrent neural networks to forecast the expected response of a patient to a series of planned treatments using simulations of a state-of-the-art pharmacokinetic-pharmacodynamic (PK-PD) model of tumor growth.

2.5.1 Pain Forecasting

Recently, studies on pain forecasting are gaining attention. There have been attempts to forecast pain, specifically postoperative pain, using data other than physiology and activity measurements. Tighe et al. [128] explored various classification algorithms to forecast whether a patient was at risk for moderate to severe postoperative pain for postoperative day 1 and day 3 using 796 clinical variables from Electronic Medical Records (EMRs) in a retrospective cohort of 8,071 surgical patients. In forecasting moderate to severe postoperative pain for the postoperative day (POD) 1, the LASSO algorithm, using all 796 variables, had the highest accuracy with an area under the receiver-operating curve (ROC)

of 0.704. Next, the gradient-boosted decision tree had a ROC of 0.665, and the k-nearest neighbors (k-NN) algorithm had a ROC of 0.643. For POD 3, the LASSO algorithm, using all variables, again had the highest accuracy, with a ROC of 0.727. Logistic regression had a lower ROC of 0.5 for predicting pain outcomes on POD 1 and 3. The same group also developed a model based on RNNs to forecast pain levels after administering specific pain medication and trained it on pain score patterns [90].

Pain forecasting is a critical issue in pain management in SCD. It can provide an objective criterion for the timing and dosage of the administration of opioids based on current physiological data. However, there needs to be more research in pain forecasting for patients with SCD, primarily due to a lack of large, structured datasets. Based on the recent advancements in the field and previous works from our group [129], we believe designing data-driven machine learning models from physiological data is a promising approach for pain forecasting. Data-driven approaches can potentially improve patient outcomes in precision medicine while minimizing costs and reducing patient risk levels. Probabilistic models help map a patient's trajectory and forecast their likely course of disease [19] as they aim to provide the best assessment of future events. Prior works have shown that interventions are more accurate and effective when individualized mathematical models are built to capture the status of a particular subject over time [20, 21, 117]. These models can forecast disease progression and improve the effectiveness of interventions if applied in real-world practice. A simulation of the patient's disease progression that behaves identically to the patient in terms of disease status is a digital twin of the patient. These patient trajectories can be used to model a patient's future conditions if no external intervention changes their clinical course. However, it is very challenging to precisely anticipate the efficacy of a drug in an individual patient [117]. Therapies for an average patient profile may not be well adapted to an individual. It is often difficult to confidently make individual-level forecasts using predictive modeling due to the inherent heterogeneity across patients. Hence, predictive modeling may benefit from approaches that allow for accurate characterization and forecasting

of disease progression at an individual level. Particularly, data-driven representational learning approaches in precision medicine may be useful in managing health conditions that elicit complex patterns in disease progression and treatment response [130]. In this research, we evaluate multiple models to forecast the trajectories of several clinical measurements in patients suffering from SCD experiencing a pain episode.

3 Pain Assessment from Physiological Signals in Electronic Health Records

3.1 Overview

Can we estimate subjective pain from objective physiological signals collected from Electronic Health Records (EHRs) irrespective of the nature of hospital visits?

In this chapter, we describe the detailed approach for our data driven strategies for pain management in patients with sickle cell disease. We describe a pain assessment model that predicts subjective self-reported pain scores from objective physiological measures from Electronic Health Records (EHRs). We provided a structured definition to extract nature of hospital visits from EHR data. Next, investigated the relationship between each physiological signal available and explored the feasibility of using these objective measures to predict subjective pain scores at both intra-individual level and inter-individual level. We presented an evaluation of the pain prediction models under varying Likert scales of pain ratings.

3.2 Material and Methods

3.2.1 Data

In this study, we utilized 67927 records from EHR data collected from 50 participants at Duke University Hospital over five consecutive years. Each record contained measures for six vital signs as follows: (i) peripheral capillary oxygen saturation (SpO2), (ii) systolic

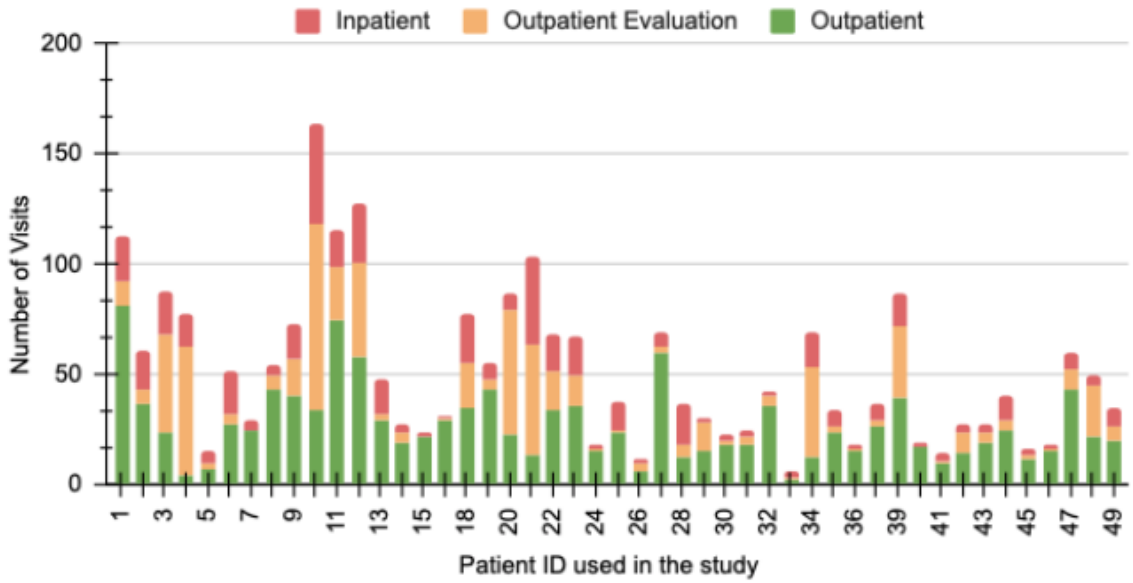

Figure 3.1: Distribution of visits for 50 patients. (Study Patient identifiers 17, 37 and 50 are absent.)

blood pressure (SystolicBP), (iii) diastolic blood pressure (DiastolicBP), (iv) heart rate (Pulse), (v) respiratory rate (Resp), and (vi) temperature (Temp). Along with the vital signs, each record also included the patient's self-reported pain score with an ordinal range from 0 (no pain) to 10 (severe and unbearable pain).

The data were de-identified using study labels to label the patient without identification. The timestamp for each data entry was also de-identified, preserving temporality. The dataset had missing values for one or more of the vital signs and the pain score. Our analysis is done on 59728 records containing at least one of the six vital signs or pain score values from 47 patients as we observed that no data was extracted for three patients. As the percentage of complete records in our dataset was only 7.6%, we employed an imputation method to impute the missing data values. We utilized Multiple Imputations by Chained Equations [131], sometimes called "fully conditional specification" or "sequential regression multiple imputations," as it is widely used in clinical practice and recent healthcare studies [132, 70].

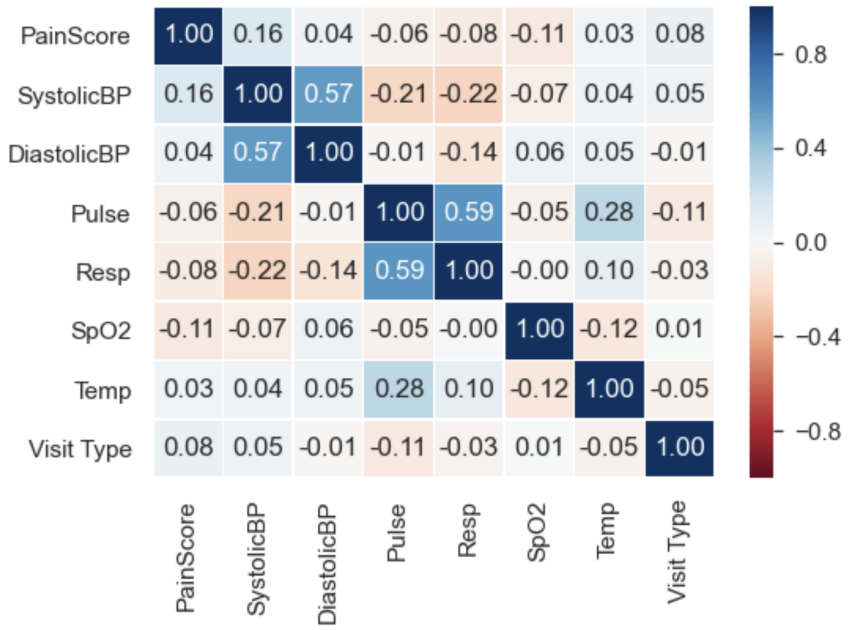

Figure 3.2: Pearson correlation of six vital signs and visit types in the original dataset.

3.2.2 Type of Hospital Visit:

Some patients with SCD have higher inpatient requirements than others due to the subjectivity and frequency of pain crisis. Furthermore, because of SCD-related complications, many people with SCD may visit hospitals more frequently. However, limited information is available related to various hospital visits' characteristics, including emergency department visits among SCD patients. Information related to the type of hospital visit by patients with SCD can help develop services and strategies for best meeting patients' healthcare needs with SCD. To understand the variations in the nature of visits, we followed the definitions below recommended by our co-author clinician to extract information about the nature of visits for every record in our dataset.

- **Visit**: For each patient, we consider a record to be of a different visit if there is a gap of at least two days between the records.
- **Outpatient visit**: We define a visit to be an outpatient visit if the patient has not

Table 3.1: Intra-individual pain prediction results (accuracy)

	SVM	DT	kNN	MLR	RF
Vitals	0.522	**0.653**	0.625	0.535	0.485
Vitals + Visit	0.506	**0.653**	0.625	0.535	0.486
Yang e.t al. [70]	0.582	—	0.522	0.578	0.523

stayed in the hospital for a day or longer and has two or less recordings taken.

- **Inpatient visit**: We define a visit to be an inpatient visit if the patient has stayed for two or more consecutive days in the hospital.

- **Outpatient evaluation visit**: We define a visit to be an outpatient evaluation visit if the patient has stayed in the hospital for one day or has more than two recordings taken in a single day.

Figure 3.1 shows the distribution of the three types of visits in our data.

3.2.3 Pain Prediction

We examined the Pearson correlation between the six vital signs and the type of visit in our dataset as we plan to use them as features influencing pain scores. As shown in Figure 3.2, in addition to a moderate correlation of 0.57 between systolic and diastolic blood pressure, we observe a correlation of 0.59 between pulse and respiratory rate in the original dataset. The other variables are poorly correlated or uncorrelated with one another. Hence, we utilize all six vital signs and visit information as predictors of our pain prediction models.

We implemented five supervised ML classification algorithms to predict patients' pain scores based on their vital signs: k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Multinomial Logistic Regression (MLR), Decision Tree (DT), and Random Forest (RF). We investigated both the intra-individual level and inter-individual level (i.e., treating all the patients as a single entity) analysis. For intra-level analysis, we used the six vital signs, visit information, and pain scores as patient labels were unused since samples from

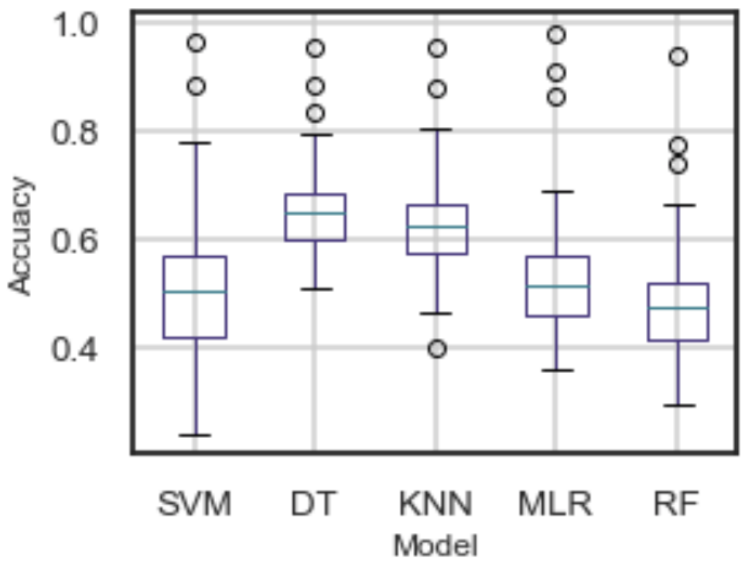

Figure 3.3: Intra-individual pain prediction accuracy results on vital signs data.

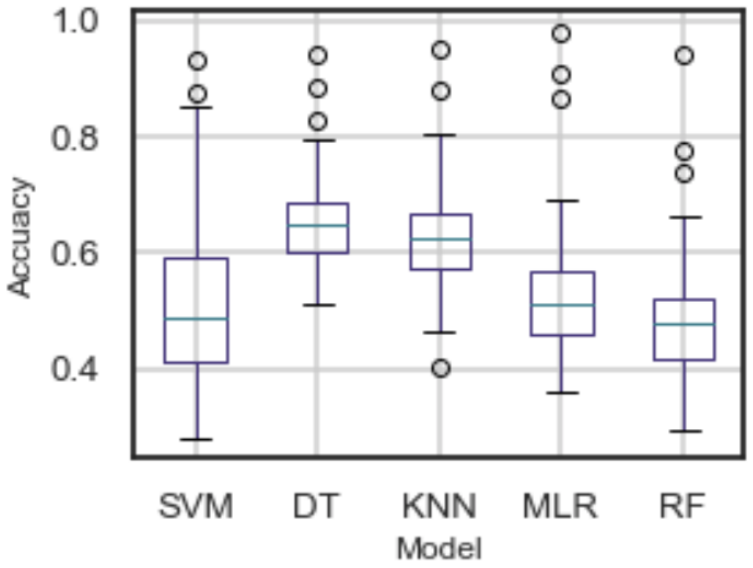

Figure 3.4: Intra-individual pain prediction accuracy results on vital signs and visit information.

the same patient were employed to build the personal model in this analysis. However, for the inter-individual level analysis, we employed the four different scenarios as reported by Yang et al. [70] i.e., Case 1: imputation with patient labels and prediction with patient labels; Case 2: imputation with patient labels and prediction without patient labels; Case 3: imputation without patient labels and prediction with patient labels; Case 4: imputation without patient labels and prediction without patient labels. For each experiment, we report the results with and without visit information. We used 10-fold cross-validation to evaluate our prediction models. We reported the model prediction accuracy as it is the ratio of correctly predicted pain scores over the total number of pain scores.

3.3 Results And Discussion

3.3.1 Intra-individual Pain Prediction

We present the intra-individual pain prediction results for 47 patients in terms of accuracy in Table 3.1. Figure 3.3 shows the accuracy distribution of predictions for all five classifiers. DT achieved the highest accuracy ranging from 0.503 to 0.953, and an average accuracy of 0.653 when trained on the six vital signs described in Section 4.1. Thus, our models trained on the vital signs of a patient could correctly predict the self-reported pain scores of the same patient on an average of 65.3% of the records. Similar performance with additional visit information (Table 3.1, row 2) indicates that for the same patient, our models can learn the differences between vital signs and pain intensity experienced by a patient during different types of visits. Our results show that a model trained on the same patient's historical data can predict the pain intensity levels for the same patient in the future based on their vital signs during outpatient, inpatient, and outpatient evaluation visits. Such a model can provide medical teams with additional information about the severity of a patient's pain, which does not rely on the patient's subjective response.

3.3.2 Inter-individual Pain Prediction

In real-time scenarios, when a new patient visits a hospital, intra-individual level models can not be applied until sufficient data is collected. We report the inter-individual pain prediction results for 47 patients in Table 3.2. The best performance was achieved in Case 4 by DT (accuracy 0.728) compared to 0.429 by MLR in Case 1 by Yang et. al. [70]. This indicates that our DT model trained on more vitals signs data collected over a more extended period (five years) from a larger cohort of people could predict the severity of pain for a new patient more accurately. Also, with more data, the models can generalize better, as we did not consider the patient-level differences during both data imputation and pain prediction (Case 4). Additional visit information seemed important in predicting pain scores when considering patient information at data imputation and prediction (Case 1). This indicates that we need to consider the type of visit to predict pain scores from vital signs for a personalized prediction from a generalizable model.

In our original dataset, we have 11 unique self-described pain scores ranging from 0 to 10. It is challenging for one person to distinguish between such broad and granular pain intensity levels and be consistent with every pain episode. Hence, we reported our model performance at an inter-individual level by transforming our dataset on a 6-point rating scale, a 4-point rating scale, and a binary rating scale [70] in Table 3.3. The higher accuracy associated with the narrow scales is attributed to the narrow space to misclassify many records by our models, thereby improving the chances of correctly predicting a pain score.

We believe the lower performance of RF compared with DT is attributed to the replacement with duplicates based bootstrap approach to sub-sampling used in training a Random Forest model that could lead to training records not representative of the test sample the model is tested on. Furthermore, as a bagging approach to ensemble models, the misclassification error from the first bootstrap sample in random forests is not used to improve a model

trained on a different bootstrap sample. Finally, test accuracy is computed by taking the average of the different bootstrap samples' accuracy where bad bootstrap samples might harm the aggregate test accuracy. With DT, however, it is possible that our model captured the ideal training records for a given test set, thereby yielding better accuracy.

3.3.3 Pain Change Prediction

Additional information about the change in pain (increase/decrease/no change) would help determine the effectiveness of therapy and the consideration for management, such as either giving more pain medication, keeping a medication dosage stable or decreasing a pain dosage. Predicting a change in pain may be more critical than having an estimate of the pain score since the medical team can make treatment decisions based on this information and ultimately improve a patient's pain more quickly. Hence, we formulate a three-class pain change classification problem, i.e., increase, decrease, and no change. In this case, a baseline chance accuracy can be 0.33 (1/3). We report the results of our two best performing DT and kNN based inter-individual level classifiers in Table 3.4. It is not surprising that our DT model was able to predict a pain change correctly 52.2% times (0.7% more) when provided with additional visit information. It indicates our model learned that the change in pain severity might be different for each type of visit for different patients (i.e., Case 1). It is essential to consider whether a visit is an inpatient or outpatient visit to estimate a change in pain intensity for each patient.

Table 3.2: Inter-individual pain prediction results (accuracy)

	Vitals					Vitals + Visit					Yang et. al.[70]	
	SVM	DT	kNN	MLR	RF	SVM	DT	kNN	MLR	RF	MLR	SVM
Case 1	0.585	0.676	0.662	0.448	0.336	0.595	**0.697**	0.668	0.525	0.357	0.429	0.421
Case 2	0.422	**0.647**	0.644	0.345	0.335	0.593	0.643	0.530	0.427	0.335	0.215	0.236
Case 3	0.561	**0.701**	0.658	0.405	0.406	0.591	0.701	0.595	0.460	0.410	0.313	0.305
Case 4	0.590	**0.728**	0.708	0.401	0.404	0.659	0.704	0.595	0.472	0.401	0.257	0.246

Table 3.3: Inter-individual pain prediction results with varying pain scales on vitals data (accuracy) arranged from higher resolution to lower resolution. (6 Pain Scores: None:0, Very mild:1-2,Mild: 3–4, Moderate: 5–6, Severe: 7–8, Very severe:9-10; 4 Pain Scores: None: 0,Mild: 1–3,Moderate: 4–6 ,Severe: 7–10; 2 Pain Scores:No/mild Pain: 0–5, Severe Pain: 6–10)

	11 Pain Score						6 Pain Score					
	SVM	DT	kNN	MLR	RF	Yang et.al.[70]	SVM	DT	kNN	MLR	RF	Yang et.al.[70]
Case 1	0.585	0.676	0.662	0.448	0.336	0.429	0.767	0.779	0.761	0.606	0.481	0.546
Case 2	0.422	0.647	0.644	0.345	0.335	0.215	0.672	0.762	0.732	0.55	0.485	0.347
Case 3	0.561	0.701	0.658	0.405	0.406	0.313	0.766	0.771	0.773	0.599	0.586	0.449
Case 4	0.590	**0.728**	0.708	0.401	0.404	0.257	0.772	**0.814**	0.777	0.605	0.589	0.397
	4 Pain Score						2 Pain Score					
	SVM	DT	kNN	MLR	RF	Yang et.al.[70]	SVM	DT	kNN	MLR	RF	Yang et.al.[70]
Case 1	0.849	0.832	0.809	0.683	0.583	0.681	0.923	0.937	0.904	0.926	0.84	0.821
Case 2	0.788	0.821	0.788	0.659	0.589	0.521	0.915	0.919	0.893	0.903	0.835	0.680
Case 3	0.837	0.824	0.815	0.685	0.66	0.607	0.923	0.939	0.907	0.9267	0.874	0.730
Case 4	0.85	**0.853**	0.818	0.687	0.671	0.563	0.935	**0.941**	0.907	0.927	0.871	0.678

Table 3.4: Pain change prediction results (accuracy)

	Vitals			Vitals + Visit			Yang. et. al [70]
	DT	kNN	MLR	DT	kNN	MLR	MLR
Case1	0.515	0.490	0.514	**0.522**	0.504	0.517	0.403
Case2	**0.508**	0.494	0.508	0.518	0.494	0.503	0.363
Case3	**0.518**	0.466	0.517	0.520	0.492	0.518	0.390
Case4	**0.520**	0.492	0.520	0.517	0.466	0.516	0.404

3.4 Summary

In this study, we leveraged multiple machine learning algorithms on six physiological measures of patients with SCD to predict pain scores. We were able to deal with missing data and conduct a series of experiments at both intra-individual and inter-individual levels. In each of the experiments, we observed higher accuracy with an increase in data. All the models were able to capture the variation in the type of visits at an inter-individual level when considering the diversity between patients in data imputation and prediction. Our results show Decision Tree as the most promising model, followed by k-Nearest Neighbours and Support Vector Machines. The evaluation demonstrates that using objective physiological measurements to predict subjective pain in SCD patients may be generalizable for a larger cohort of patients. In the future, we look forward to extending our work to visit level analysis and exploring the patients' medication information.

4 Pain Assessment from Physiological Signals and Medication Data in Electronic Health Records

4.1 Overview

Can we learn the relationship between pain, objective physiological signals and medication data collected from Electronic Health Records(EHRs)?

In this chapter, we present a self-supervised learning model utilizing physiological data as well as medication data and self-reported pain scores collected by nurses to predict a pain score. The ability to objectively and accurately predict pain severity and onset could result in more prompt and effective treatment of pain crises, leading to improved outcomes, as well as encouraging more diligent use of medications [133]. We evaluated the role of medication data in pain estimation model by leveraging self-supervised learning. Additionally, we compared the performance of regression methods and classification methods in pain prediction problem.

This study aims to learn deep feature representations of subjective pain trajectories from objective physiological signals collected from electronic health records (EHRs). This study aims to use the vital signs and medication information collected from EHR data of patients with SCD to predict patient-reported pain scores using machine learning techniques. In this paper, we propose to represent multiple data modalities in EHRs in highlevel abstraction, vital signs, and medication information by utilizing deep auto-encoder networks such as variational autoencoders (VAEs) to predict pain intensity at varying Likert scales. Our

specific contributions are as below: (i) To the best of our knowledge, we analyzed the most extensive electronic health records data of 126,519 records from 496 patients suffering from Sickle Cell Disease collected over five consecutive years and demonstrated that a larger patient cohort data improves the model performance in pain prediction. (ii) We show that pain medication information with vital signs data can improve pain prediction at varying pain rating scales (i.e., different granularities). (iii) We demonstrate that deep representational learning cannot just improve pain prediction results but also provide a better understanding of the role of medication and physiology on the patient's pain response with a patient profiles study.

4.2 Materials and Methods

4.2.1 Data

In this study, we analyzed 1,26,519 records from EHR data collected for 496 participants at Duke University Hospital over five consecutive years. Each record contained measures for six vital signs as shown in Table 4.1. Along with the vital signs, each record also included the patient's self-reported pain score with an ordinal range from 0 (no pain) to 10 (severe and unbearable pain). Furthermore, we explore the medication information in the records. We extracted three medicinal features upon consultation with our co-author physician as shown in Table 1. The total medication dosage is the sum of all medication dosages given to a patient at a given time. For the purpose of this study, we removed the data points with status as Hold MAR; i.e., during those time stamps, the Medication Administration Record (MAR) was on hold.

4.2.2 Variational Autoencoders

In this aim, we use variational autoencoders (VAEs) to impute missing values within the data based on other samples. Autoencoders are a class of unsupervised deep learning

Table 4.1: Data modalities and variables considered in this study.

Data Modality	Variables
Vital Signs	(i) peripheral capillary oxygen saturation (SpO2) (ii) systolic blood pressure (SystolicBP) (iii) diastolic blood pressure (DiastolicBP) (iv) heart rate (Pulse) (v) respiratory rate (Resp) (vi) temperature (Temp)
Medication	(i) Medication Type (5 classes): HYDROMORPHONE ACETAMINOPHEN KETOROLAC OXYCODONE FENTANYL (ii) Medication Status (2 classes) Given/Applied Missed/Removed/Due (iii) Total Medication Dosage (in mg/ml)
Pain	Self-reported pain score on a scale of 0-10 (0-no pain to 10-severe and unbearable pain)

techniques in which we leverage neural networks for the task of representation learning. We design a neural network architecture to impose a bottleneck in the network, forcing a compressed knowledge representation of the original input data modalities. If the input features were such that each was independent of one another, this compression and subsequent reconstruction would be an arduous task. However, if some association exists in the data (for example, correlations between input data modalities), this structure can be learned and consequently leveraged when forcing the input through the network's bottleneck. VAEs are probabilistic generative models and have the same architecture as vanilla autoencoders but consider specific assumptions about the distribution of middle/latent layer variables. They learn the true distribution of input features from latent variable distributions using a Bayesian approach and present a theoretical framework for reconstruction and regularization [134].

A VAE learns the distribution of data with an encoder network by fitting it to a gaussian distribution and generates data with a decoder by sampling from the learned distribution. We utilized autoencoders to reconstruct input data (x) in the output ($\hat{x}$) layer by an encoding and decoding process. As shown in Figure 4.1, the encoder network converts the input data

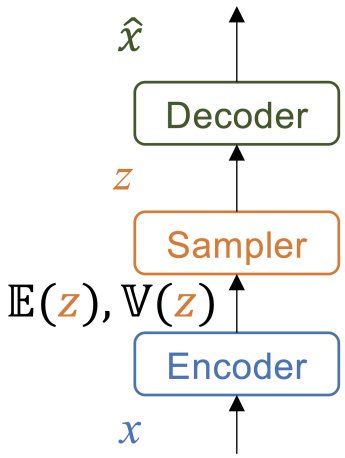

Figure 4.1: An illustration of VAE architecture for one input data modality.

(x) into a latent representation (z). The hidden state comprised two additional layers: $\mathbb{E}(z)$ and $\mathbb{V}(z)$, where the latent variable z follows a Gaussian distribution with mean $\mathbb{E}(z)$ and variance $\mathbb{V}(z)$. We sampled z from the distribution parameterized by the encoder; the decoder network then remodeled the input from the latent representations by using z to generate $\hat{x}$. The fundamental property of autoencoders is to minimise this reconstruction error using a loss function that is composed of a reconstruction term as well as a regularization term, as shown in Eq. (4.1). The loss function we minimize to train the VAE contains a reconstruction term and a regularization term as shown below:

$$l(x, \hat{x}) = l_{reconstruction} + \beta l_{KL}(z, \mathbb{N}(0, I_d)) \tag{4.1}$$

The term $l(x, \hat{x})$ is on the final layer and the regularization term enforces a specific Gaussian structure on the latent layer through a penalty term $l_{KL}(z, \mathbb{N}(0, I_d))$.

Variational means encoder network estimates μ and σ parameters (that we would call them latent variable) of the Gaussian distribution. However, real-world applications, including healthcare, almost always have missing values. In correspondence with missing values in the raw temporal data, we substitute the corresponding categories with a unique integer to

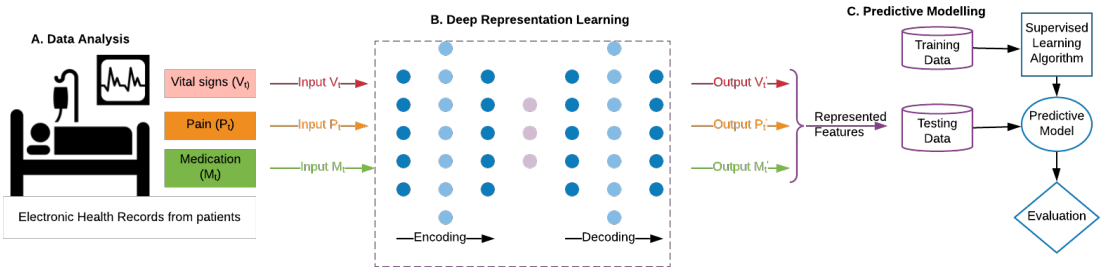

Figure 4.2: Deep representation learning for Pain Prediction

properly encode the status of the missing information. The encoder consists of a Long Short Term Memory (LSTM) cell. It receives input sequences resulting from the concatenation of the raw physiological data and the extracted categorical medicinal features. As in every encoder in a VAE architecture, it produces an output used to approximate the latent distribution's mean and variance. The decoder samples from the latent distribution form output sequences. This approach helps us to develop an unsupervised framework that can fill the missing pieces appearing in the real-world EHR data volume streams in not only patients with SCD but for other healthcare applications as well.

4.2.3 Proposed Framework

Fig. 4.2 provides an overview of our approach in three consecutive steps. In the first step (A), we pre-processed the raw data to overcome the data challenges such as missing values. Next, in the second step (B), we applied unsupervised deep representation learning to generate higher-level abstraction of the input data modalities. Finally, in the third step (C), we investigated supervised algorithms for predictive modeling and performed the evaluation.

Step A: Data Analysis In this study, we utilized records from EHR data collected at Duke University Hospital and de-identified them using study labels to label the patient without identification. The timestamp for each data entry was de-identified, preserving the

temporality. The dataset had missing values for one or more vital signs, medication, and pain scores. The dataset contained 126,519 records from 496 patients collected over five consecutive years. However, we included 33,000 records in this study due to more than four features missing in the remaining records. Of the 33,000 records, 18,291 had at least one of the five medication types (as in Table 4.1) administered. We did not have the demographic information of the patients. Data for each patient varied as while 70 patients had a one-time visit to the hospital, 240 patients visited over at least 100 days. The most patient records were for a patient staying for 1705 days with a high mean pain score of 8 and received pain medication 219 times (an average of 338 mg of total pain medication dosage). We did not consider the effect of any other medical condition on the patients in this study.

Step B: Deep Representation Learning In the second step, we represent all the input data modalities in high level abstraction by utilizing multiple deep auto-encoder networks including variational autoencoders (VAEs). We evaluate the performance of each network while considering the tuning of hyper-parameters such as learning rate, batch size, number of epochs, and number of hidden layers and hidden units for precise training to avoid overfitting.

Step C (Predictive Modeling): In this step, we apply supervised learning techniques to the represented dataset using linear and nonlinear approaches such as Random Forests, SVM and LASSO. Our experiments consists of three main phases: (1) training VAE, where the training samples are used to train the VAE, and the reconstruction loss for each training data sample is stored according to the target pain score; (2) generating new pain scores, where the VAE decoder generates new pain score samples based on specified classes, and each newly generated data sample is merged into the original training data set under the condition that the class reconstruction loss is satisfied; (3) predicting pain scores, where the VAE decoder is used to initialize the weight of the hidden layers, the merged training

data set is used to train the classifier, and the trained classifier is used to predict pain scores on the testing data set.

4.2.4 Experimental Study

In our experimental study, we implemented our methodology on the de-identified EHR dataset. This study design helped us discover our method's performance in predictive modeling for patients with SCD. Across several attributes consisting of patient clinical records and individual health status, nine attributes, including vital signs and medication information, were considered for data analysis related to 496 patients. As mentioned before, the goal is to predict pain scores based on high-dimensional features.

We implemented Variational autoencoder (VAE) by using PyTorch and Keras libraries with tensorflow backend in Python. The VAE architecture has 5 hidden layers (two hidden layers of encoders and decoders and one middle layer). We applied parameter tuning for major parameters such as learning rate, activation functions and batch size to select the best parameters. We employed a hidden dropout component with a rate of 0.2 and a sigmoid activation function to the final layers. The models were trained for 100 epochs using an Adam optimizer with a learning rate of 0.001 (with exponential decay rates of first- and second-moment estimates $\beta_1 = 0.9$ and $\beta_2 = 0.999$) and a batch size of 64. Once the latent features were extracted, they were fed into supervised learning step for pain score prediction. For the supervised learning step we consider three well known supervised classifiers: Random Forests (RF) [135] (with 50 trees and 1/2 of the features considered at every split), Lasso Regression [136] and Support Vector Machine (SVM) [137] (with RBF kernel C = 1.5 and gamma set to 1/N_f, where N_f denotes the number of features). We used average accuracy as our evaluation measure for performance evaluation in testing process. Finally, we visually inspected the learned representations of the whole data set and compared them to the represented data. For this task we employed the t-distributed

stochastic neighboring embedding (tSNE).

4.3 Results and Discussion

We performed our approach for VAE (represented data) and supervised classifiers as well as original data (unrepresented data), and compared their performance based on the results obtained from testing process with 5-folds cross validation (for each fold we considered 80% of the data for training, 10% for validation set and 10% for test set). This comparison has been shown in Table 4.2. We would like to make a note here that missing values of pain scores are not imputed.

According to these results, our approach with representation learning reduces the prediction error and achieves a better accuracy rather than using the original features. Furthermore, we also analyze the performance in pain score prediction with only vital signs as compared to including the medication information. We show in Table 4.3 that our approach with RF classifier achieves better accuracy with medication and vital signs information rather than only vital signs information in predicting the respective pain scores. This indicates that when provided with additional medication information, our approach can learn better representations of patient profiles from vital signs to predict their pain levels.

Table 4.2: Pain prediction results with varying pain scales on vitals data (accuracy) arranged from higher resolution to lower resolution.

Approach	11 pain scores			6 pain scores			4 pain scores			2 pain scores		
	RF	SVM	Lasso	RF	SVM	Lasso	RF	SVM	Lasso	RF	SVM	Lasso
Original Data	**0.301**	0.242	0.216	**0.363**	0.352	0.337	**0.432**	0.426	0.392	**0.535**	0.513	0.486
VAE Data	**0.343**	0.321	0.307	**0.391**	0.371	0.348	**0.472**	0.452	0.439	**0.603**	0.561	0.549

Table 4.3: Pain prediction results with varying pain scales on vitals data (accuracy) as compared to additional medication data arranged from higher resolution to lower resolution.

Approach	11 pain scores		6 pain scores		4 pain scores		2 pain scores	
	Vitals	Vitals + Med	Vitals	Vitals + Med	Vitals	Vitals + Med	Vitals	Vitals + Med
Original Data	0.301	**0.442**	0.363	**0.463**	0.432	**0.689**	0.535	**0.787**
VAE Data	0.343	**0.476**	0.391	**0.493**	0.472	**0.706**	0.603	**0.828**

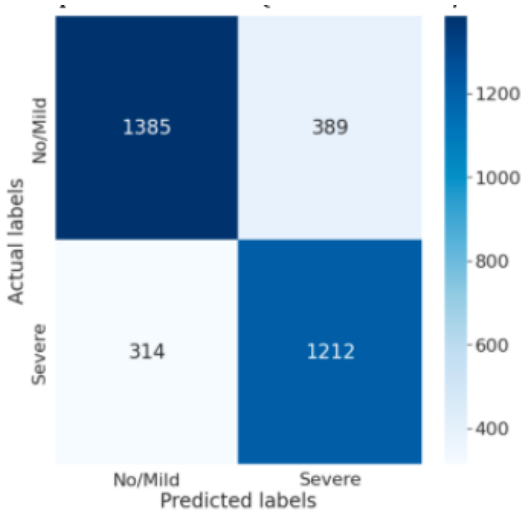

Figure 4.3: Confusion Matrix for best performing model with original data representations for 2 pain score levels (Pain Scores:No/mild: 0–5, Severe: 6–10)

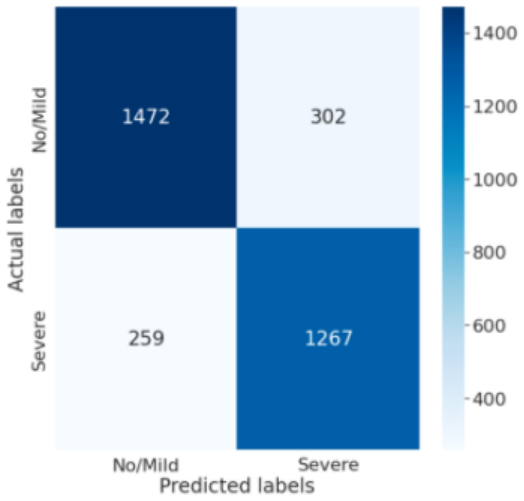

Figure 4.4: Confusion Matrix for best performing model with VAE data representations for 2 pain score levels (Pain Scores:No/mild: 0–5, Severe: 6–10)

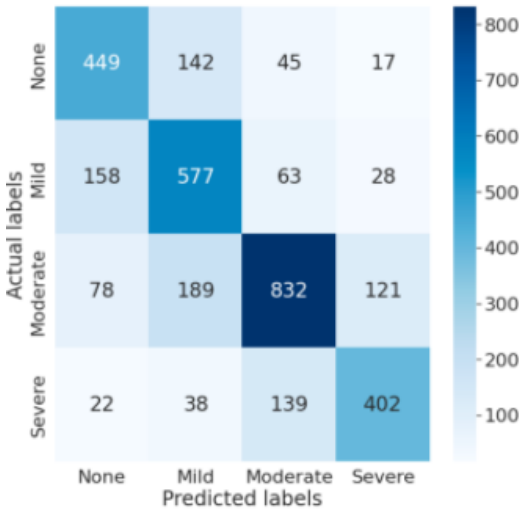

Figure 4.5: Confusion Matrix for best performing model with original data representations for 4 pain score levels (Pain Scores:None: 0,Mild: 1–3,Moderate: 4–6 ,Severe: 7–10)

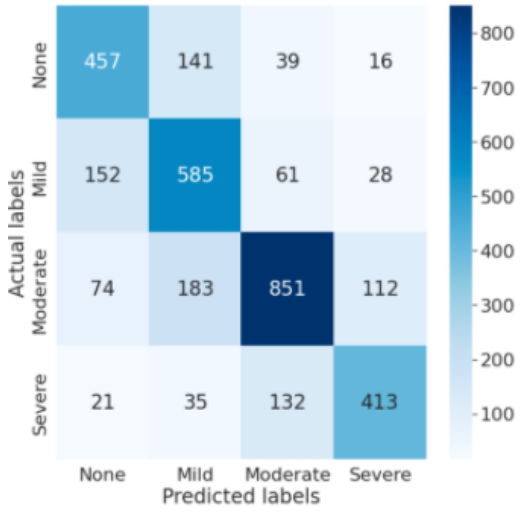

Figure 4.6: Confusion Matrix for best performing model with VAE data representations for 4 pain score levels (Pain Scores:None: 0,Mild: 1–3,Moderate: 4–6 ,Severe: 7–10)

Deep Representation Learning

In this research, we propose a study for the evaluation of deep feature representation in predicting pain scores of patients with SCD from their vitals signs and medication information. The results emphasize that representation learning plays an effective role in the performance of clinical prediction. As shown in Table 4.2, our models trained on deep represented features are able to identify pain scores for not just 6.8 % more patients at an abstract pain intensity level of no/mild pain or severe pain but display significant improvement by detecting pain intensity for 4.2% more patients at a highly granular pain score intensity (i.e., on 11 pain ratings) when compared with models trained on unrepresented raw vitals data. We observe the similar performance of deep represented features as compared with raw data features when medicinal data is included in the modeling. Our models trained on VAE represented features generated on both vitals and medicinal data are able to identify pain scores for 3.4%, 3%, 1.7%, and 4.1% more patients from higher to lower resolution of pain intensity than models trained on raw data. In order to investigate further, we show in the confusion matrices of the best performing models trained on vitals and medicinal data in Figures 4.3 - 4.6. As shown in Figures 4.3 and 4.4, it is interesting to note that with deep feature representations, our model is able to identify 4.9% more cases of no/mild pain and 3.6% more cases of severe pain accurately while reducing the misclassification. Similarly, Figures 4.5 and 4.6 show that with more granular 4 point pain intensity levels (Pain Scores: None: 0, Mild: 1–3, Moderate: 4–6, Severe: 7–10), our model trained on deep represented features is able to identify more than 1% more cases accurately as compared to original data representations while reducing the misclassification to other pain groups. Also, the models are able to identify more instances with moderate pain as compared to none, mild and severe pain. It is noteworthy that the misclassification for each category of pain reduces with the stretch between the pain severity levels. For example, as shown in Figure 4.6, our best model for 4 pain scores (Table 4.3 row 2, column 6) wrongly predicts 21 instances

of severe pain data as no pain, 38 instances as mild pain, and 139 instances as moderate pain. Similarly, it predicts 16 instances of low pain samples as severe pain, 39 instances as moderate pain, and 141 instances as mild pain. The misclassification reduces with the granularity of pain intensity reflecting the subjective nature of pain.

Role of Medication Data

While prior works have shown the efficacy of data mining techniques to implement medical decision making with treatment outcome prediction [138, 139], to the best of our knowledge, this is the first work in analyzing the role of medication in pain level prediction for patients with SCD. Our results show that for abstract pain levels, a representational learning-based approach can predict whether or not a patient is having pain for 22.% more patients when provided with their medication information (Table 4.3). This means that when our model is provided with not just vitals information but also the medication type, total medication dosage, and status, it is able to predict whether or not the patients are having pain for more patients as compared to when provided with only vital signs. This finding is substantiated by current medical literature on pain management [140] where clinical research focuses on finding the optimal medication dosage for individual patients. By building a model that incorporates medication information along with physiological data, we are one step closer to future pain forecasting that can utilize current physiological information and pain medication to predict the pain at a future time point as a means of assessing the next medication dosage and time. Furthermore, for higher resolution of pain levels i.e., 11 levels, our deep representational learning-based approaches could predict subjective pain scores for 13.3% more patients when provided with medication information. Also, our model is able to predict such highly subjective pain scores for 38.5% more patients than the random assignment of 9.09% (i.e. 1/11 pain scores) when provided with both vitals and medication information of the patients.

Empirically, our results demonstrate that: 1) Medical feature representation can improve the performance of prediction, and 2) Medication information can lead to significant improvement on pain level prediction.

4.3.1 Visualization

We visually inspected the learned representations of the whole data set obtained from the VAE representations. Using t-SNE plots, shown in Figure 4.7, we compared the level of disentanglement of the represented data and raw data. The t-SNE projections clearly show that VAE is able to produce more sparse and disentangled representations in comparison to raw data. While the t-SNE projections of the raw data also indicate data separability, the deep representations are able to identify the variations in mean pain scores (low, moderate, high). This may explain the competitive performance produced by the benchmark classifiers in the previous section, as well as the advantage of integrating vital signs and medicinal data. While some embeddings are clearly clustered closer to the same pain range, we observe some overlap as well. Specifically, we observe better alignment among the low pain and high pain profiles as compared to the moderate pain profiles. This may be due to the variation and frequency of data recordings taken for the patients. These preliminary visualization results indicate that our VAE method may require additional data in order to generate representations to obtain a more granular separation between pain scores of patients.

Patient Profiles:

In order to understand the alignment of the representations learnt by our best performing VAE model, we illustrate six sample patient profiles clustered into the three pain range categories (no/low: 0-3, moderate: 4-6, high: 7-10) by the t-SNE projections of the embeddings (as shown in Figure 4.7). As shown in Table 4.4, we present two patient profiles from each

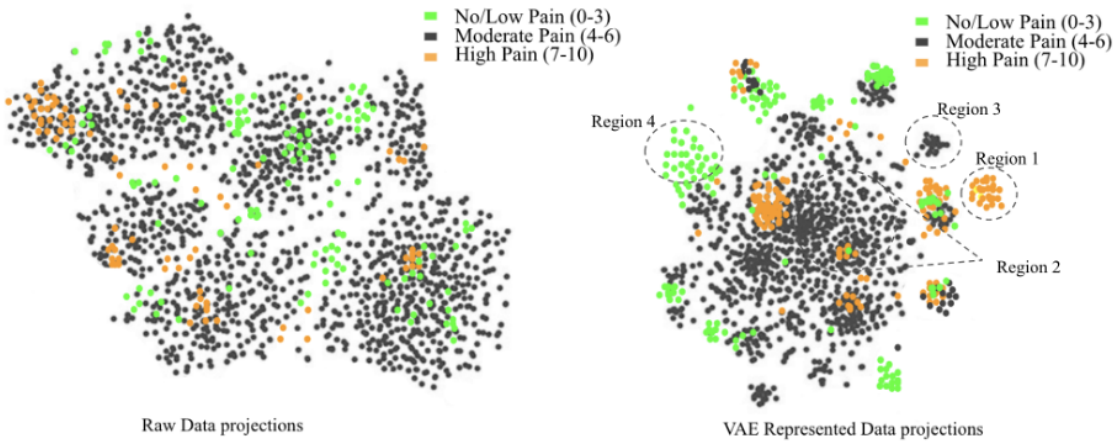

Figure 4.7: Qualitative comparison of the learned representations using t-SNE projections.

of the three categories of pain score with regards to the medication administered and vital signs. It is to be noted that we have specifically chosen regions where the pain profiles belong to one of the three pain levels. While we chose two patient profiles each from better aligned Region 1 and Region 4, we compared two patient profiles from a more spread out moderate pain intensity (Regions 2 and 3). The patient numbers are patient identifiers used in this work.

We observe a positive correlation between the medications administered and pain scores in all four patients with high and moderate pain levels. This reflects that the patients reporting higher pain scores were administered increased medication dosage (as shown in Figure 4.8 for Patients 1 and 3), and our model has learned that relationship. For both patients (Patient 1 and 2) illustrated with high pain, we observed a positive correlation between Hydromorphone dosage and pain score as well as a high correlation between total medication dosage and vital signs which may be reflective of more pain medications being given when a patient has high pain. This indicates that our model has learned the interplay between the medications, vital signs, and pain intensity. It will be interesting to analyze these correlations pre- and post-medication in the future.

For both the patients (Patient 3 and 4) with moderate pain, in addition to a positive correlation

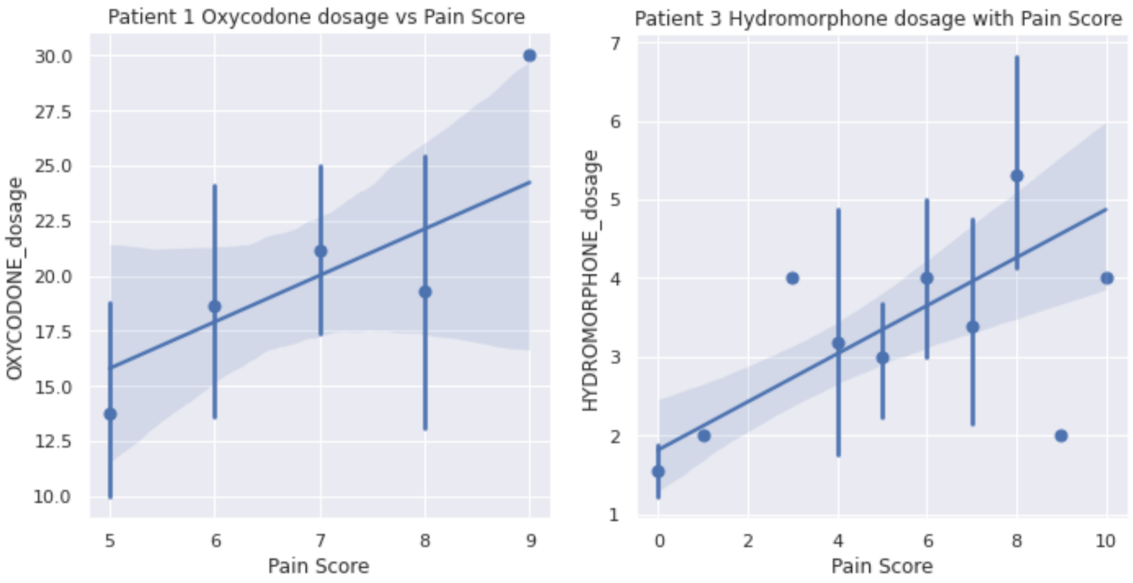

Figure 4.8: Distribution of medication dosage with pain score for sample patients with high and moderate mean pain intensity.

of Hydromorphone with pain score, we observed a positive correlation of medication with blood pressure. This indicates that our model has learned a possible association of administering Hydromorphone for moderate pain intensity levels during which time they have elevated blood pressure. However, Patient 4 has a higher positive correlation between medication dosage and vital signs as compared to Patient 3. This may be a possible reason that they were not closer in the embedding space and belong to distant regions, as shown in Figure 4.7.

Although we did not observe any significant correlation between medication and pain scores for both patients (Patient 5 and 6) with no/low pain, we observed a positive correlation between medication and vital signs. This may again be suggestive of the elevated vital signs that occur with pain and lead to medication administration. While both patients might have reported varying pain scores between 0-3, it is highly challenging to differentiate between pain scores 1 and 2 or 2 and 3. Hence, there might be a case that with medication, their vitals have improved (as indicated by the positive correlation), making them feel better. This sample patient profile study indicates that deep feature representations can be

Table 4.4: Sample of patient profiles from the learned VAE representations clustered together using t-SNE projections as shown in Figure 7 (Region 1- High Pain, Region 2,3 - Moderate Pain, Region 4- Low Pain)

Region	Patient Number	Medication Administered	Correlation Between Medication Dosage and Pain Score	Vital Signs	Correlation Between Medication Dosage and Vital Signs
1	1	Oxycodone	0.23	Temperature	0.65
		Hydromorphone	0.11		
		Acetaminophen	0.17		
1	2	Ketorolac	0.55	Systolic Blood Pressure	0.47
		Hydromorphone	0.24		
2	3	Hydromorphone	0.35	Systolic Blood Pressure	0.099
		Acetaminophen	-0.20		
		Ketorolac	0.41	Pulse	0.04
3	4	Oxycodone	0.59	Systolic Blood Pressure	0.27
		Hydromorphone	0.08		
4	5	Fentanyl	-	Peripheral capillary oxygen saturation SpO2 level	0.13
4	6	Acetaminophen	-	Temperature	0.47
				Pulse	0.38
				Peripheral capillary oxygen saturation SpO2 level	0.28

used to learn complex relationships between various factors influencing pain management. With more data for each patient, this study can be extended to design personalized pain management tools to assist clinicians.

4.4 Summary

In this study [141], we proposed a simple and effective pain prediction model based on objective vital signs and pain medication usage. Our experiments demonstrate that the information about pain medication (type, dosage, and status) can improve pain intensity prediction at both abstract levels as well as granular levels. Our findings indicate the importance of medication information (achieving an accuracy of 82.3%) as well as demonstrate that a larger patient cohort data with deep representational learning improve the model performance (by 17.5% as compared to Padhee et. al.[142] and by 24.6% as compared to Yang et. al.[70]. Furthermore, from our unsupervised analysis, we were able to distinguish unique patient profiles (see Table 4.4) that can help isolate different patient profiles for further understanding of the role of physiology and medication on their pain response. In the future, this study can be extended to further dig deeper into the effect of variation in medication protocols, such as the changes in vital signs before and after medication, and time elapsing between medication doses. This would be an essential part of a real-time pain

forecasting system and can be extended as a trial that evaluates the timing of administration of additional doses of opioids based on physiologic and objective data alone. Our initial results indicate promise in pursuing each of these efforts, and our study is a valuable addition to ongoing studies investigating how objective vital signs and medication data can be used to help providers better understand and design pain management strategies.

5 Pain Forecasting from Physiological Signals in Electronic Health Records

5.1 Overview

In Chapter 3 and Chapter 4, we have shown promising results in pain assessment based on physiological signals and medication data. The promising results in pain prediction motivated us to move to the next step in our research:

Can we use objective physiological signals to forecast future pain?

In other words, can we utilize temporal, historical knowledge of patients for a better understanding of the future pain trajectory of patients? Such forecasts would provide an additional dimension of information in pain treatment plans. We treat pain forecasting as a supervised learning problem. The input is a past sequence of physiological signals, and the output label is the self-reported pain score at a future time. In clinical practice, it is challenging to obtain a large amount of pain records since it is mostly recorded by patients' self-report. Therefore, it is expensive and painful (due to the need for patient compliance) to solve pain forecasting problems in a purely supervised manner. In light of this challenge, we proposed to solve the pain forecasting problem based on the temporal clustering of patient profiles learned using self-supervised methods. Chronic diseases are heterogeneous, with widely differing outcomes even in narrow patient subgroups. Disease progression manifests through a broad spectrum of clinical factors, collected as a sequence of measurements in electronic health records, which gives rise to complex progression patterns among patients [143, 144]). For example, cystic fibrosis evolves slowly, allowing

for the development of comorbidities and bacterial infections, and creating distinct responses to therapeutic interventions, which in turn makes the survival and quality of life substantially different [145, 146]. Identifying patient phenotypes with similar progression patterns can be advantageous for understanding such heterogeneous diseases. Temporal predictive clustering has been recently used as a data-driven framework to partition patients with time-series observations into clusters of patients. Recent research has typically focused on either finding fixed-length and low dimensional representations [147, 148] or modifying the similarity measure [149, 150] both in an attempt to apply the existing clustering algorithms to time series observations. However, clusters identified from these approaches are purely unsupervised – they do not account for patients' observed outcomes (e.g., adverse events, the onset of comorbidities, etc.) – which leads to heterogeneous clusters if the clinical presentation of the disease differs even for patients with the same outcomes. Thus, a typical prognosis in each cluster remains unknown, which can mystify the understanding of the underlying disease progression [151, 152]. To overcome this limitation, we focus on predictive clustering [153] to combine predictions on future outcomes with clustering. More specifically, we aim to find cluster assignments by learning discrete representations of time series that best describe the future outcome distribution. By doing so, patients in the same cluster share similar future outcomes to provide a prognostic value.

Figure 5.1 illustrates a pictorial depiction of our clustering procedure. X-axis represents pain level over time t, and the y-axis represents the state of x_n vital signs. A patient at time t_1, with the state of vital signs x_1, x_2, ..., x_n, can be clustered as belonging to a patient phenotype (purple color in Figure 5.1) with future pain score to be P_1 at time t_2. The vital signs for the same patient changing from the state at t_1 at time t_2 might belong to a phenotype of patients with similar state of vital signs (blue color in Figure 5.1).

In this chapter, we propose to leverage the self-supervised representation learning methods that learn from extensive unlabeled physiological data to solve pain forecasting with limited

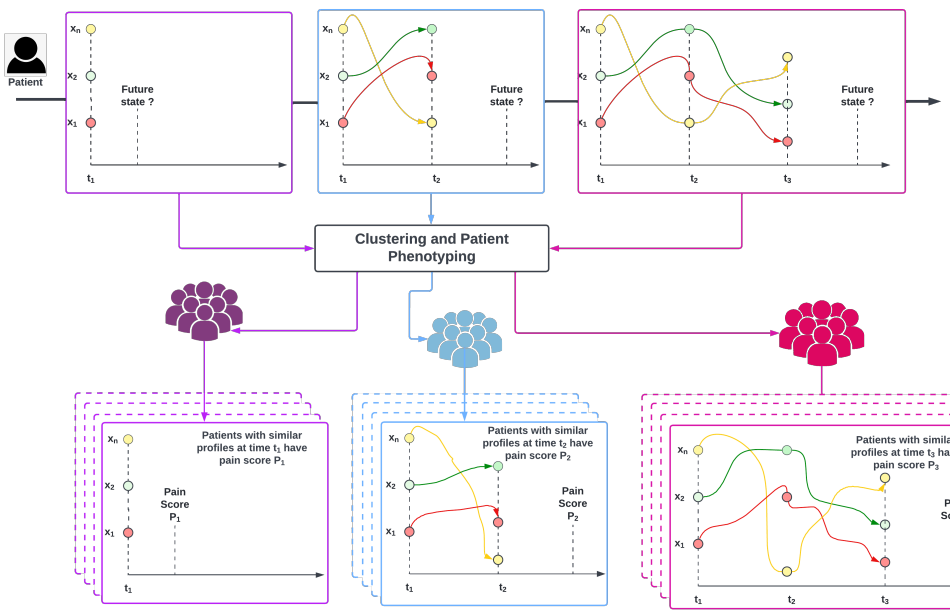

Figure 5.1: An illustration of our (real-time) clustering procedure. Here, a new patient is assigned over time to one of the three phenotypes (purple, blue, pink) based on the expected future pain score – as new observations are collected.

pain labels. We analyzed the performance of self-supervised learning tasks under various training settings and their impact on the pain forecasting downstream task. Then, we demonstrated that the self-supervised-based model performs significantly better than the pain forecasting models trained in a purely supervised fashion. Finally, we showed that models trained with self-supervised approaches could be used to learn the evolution of a patient's pain levels over time.

5.2 Materials and Methods

Accessing healthcare data is a significant challenge due to privacy concerns of patients, hospitals, insurance companies, and pharmaceutical companies. One way of overcoming this challenge is to anonymize healthcare records and medication information so that relationships to specific individuals or entities can not be identified. Using anonymization, prior research has attempted time-series forecasting different healthcare costs [154, 155, 156, 157]. We employ similar anonymization techniques in our dataset.

Although prior research has performed time-series forecasting in healthcare data, another challenge is selecting the appropriate predictive model to use for performing analyses (there are very few suggestive forecasting algorithms for healthcare data due to the newness of this domain and its datasets). A way of overcoming the model selection challenge is by evaluating the predictions from existing forecasting methods with other recent methods in literature [158]. For example, recurrent machine-learning methods (e.g., long short-term models) have recently been proposed [159]. These recurrent methods have been used for supervised learning of features for time-series forecasting [159, 17]. These recurrent methods could provide improvements over existing statistical time-series techniques [e.g., autoregressive integrated moving average (ARIMA) [160]], which are often dependent on hand-crafted features requiring expert knowledge in the field.

In this work, we address the above challenges by evaluating the performance of memory-less neural network models [e.g., multilayer perceptron (MLP)] with memory-based neural network models [e.g., long short-term memory (LSTM)] for performing time-series predictions of longitudinal healthcare data. Due to the popularity of the ARIMA model [161, 162], we evaluate the performance of this model against both memory-less and memory-based approaches in our anonymized data from patients with SCD. Furthermore, motivated by patient phenotyping, we evaluate the potential of a clustering model.

5.2.1 Data

In this study, we utilized 51718 records from 498 participants at Duke University Hospital over a maximum of five consecutive years. Each record contained measures for six vital signs as follows: (i) peripheral capillary oxygen saturation (SpO2), (ii) systolic blood pressure (SystolicBP), (iii) diastolic blood pressure (DiastolicBP), (iv) heart rate (Pulse), (v) respiratory rate (Resp), and (vi) temperature (Temp). Along with the vital signs, each record also included the patient's self-reported pain score with an ordinal range from 0 (no

Table 5.1: Percentage of missing data before interpolation (raw data) and after interpolation.

Variable	Raw Data	After Interpolation
BP	74.481765	37.137940
SpO2	65.178842	3.758846
Pulse	67.738294	0.425384
Resp	67.274355	1.674465
Temp	78.210408	2.225531
Pain Score	68.536177	12.732511

pain) to 10 (severe and unbearable pain). The data were de-identified using study labels to label the patient without identification. The timestamp for each data entry was also de-identified, preserving temporality. The dataset had missing values for one or more of the vital signs and the pain score. Furthermore, we generated the visit information for our dataset following the definitions by Padhee et. al. [142].

Our data had missing values in one or more variables. The percentage of missing data is shown in Table 5.1. There are two main processing methods for missing data: deleting data containing missing data and interpolating missing data [163]. The deletion method is to delete the instance sample data that contains missing data in the data set to obtain the remaining complete data set for subsequent analysis. This method is simple and feasible, but its advantages and disadvantages are quite obvious. When the proportion of missing data is small, especially when a data sample contains multiple missing data, the overall impact of deleting data containing missing data is small. However, it may also lead to sample imbalance and loss of important data information. With the increase of the proportion of missing data, after the deletion of missing data, the remaining data will be difficult to reflect the true information, especially in the case of nonrandom missing data [164]. Therefore, upon consultation with our clinical advisor, due to the challenges with missing data points within small windows in time-series forecasting, we employed linear interpolation within 2-hour time window for the records within each visit. As shown in Table 5.1, the percentage of missing data significantly reduced after interpolation.

5.2.2 Models

In this section, we explain the working of different models like ARIMA, MLP, LSTM, and predictive clustering models.

Autoregressive Integrated Moving Average

In an ARIMA model (Newbold, 1983), the future value of a variable is assumed to be a linear function of several previous observations and random errors. An ARIMA model is defined as ARIMA (p, d, q):

p: order of the autoregressive part (AR);

d: degree of first differencing involved;

q: order of the moving average part (MA).

The Autoregressive (AR) process is a stochastic process where the output is linearly dependent on the weighted sum of the previous values, and a white noise error [165]. One of the critical tasks while designing an ARIMA model is finding the best value of p, q, and d [165]. The value of p represents the previous time steps of the time series to be used in predicting the future value. The value of q represents the previous error terms used to predict the future value [162]. The value of d indicates the number of times we need to differentiate the time series to make it stationary. Autocorrelation and partial autocorrelation function plots are used in the literature to find an approximate range of p and q parameters [165]. Next, the p and q parameters range is used in a grid search approach [166] to find the optimal values.

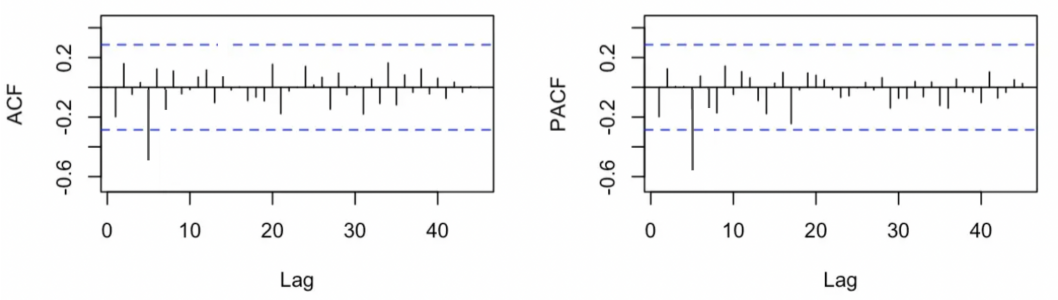

Figure 5.2: Initial Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots showing significant lag at 5.

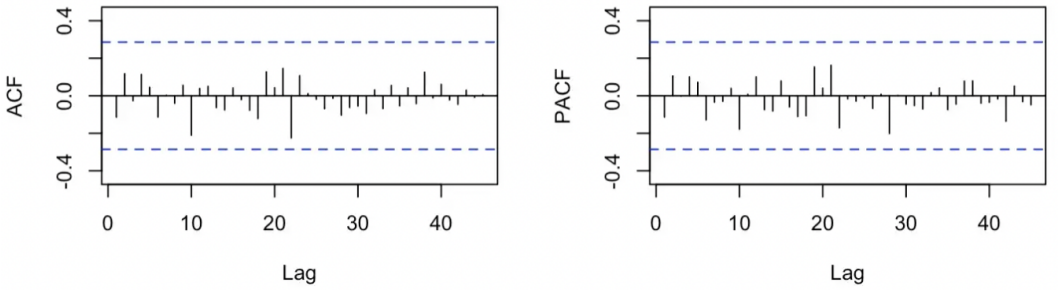

Figure 5.3: Final model Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots showing no significant lag.

Multilayer Perceptron

A multilayer perceptron (MLP) is a variant of the original perceptron model proposed by Rosenblatt [167]. A neuron computes a weighted sum of the inputs, followed by a non-linear activation φ of the calculated sum, as shown in Equation 5.1. The activation function in a neural network helps to generate mappings from inputs to outputs, and the neural networks learn complex data representations [168]. Classically, several activation functions exist, such as sigmoid, tanh, and ReLU [168, 169]. According to Krizhevsky et al. [169], the sigmoid and tanh activation functions suffer from the vanishing gradient problem, and the ReLU activation function overcomes the vanishing gradient problem, providing faster convergence and is computationally efficient. Hence, we used the ReLU activation function in our MLP model. We defined output o_i of a neuron in our MLP model as per the following equation:

$$o_i = \varphi(\sum_{j=1}^{d}(x_j w_{ij} + b_j)) \tag{5.1}$$

Long Short-Term Memory

Long short-term memory (LSTM) is a Recurrent Neural Network (RNN) type, i.e., a multi-layer NN. Hochreiter and Schmidhuber [170] originally introduced the LSTM architecture to overcome the vanishing or exploding gradient problem. n derivatives are multiplied in a network of n hidden layers. The gradient will increase exponentially if the derivative is big, and as we propagate down the model, it eventually explodes, known as the problem of exploding gradient. Alternatively, the gradient will decrease exponentially if the derivatives are small, and as we propagate through the model, it eventually vanishes, known as the vanishing gradient problem. LSTM allows flow gates, i.e., the input gate, the forget gate, the control gate, and the output gate, as shown in Figure 5.4. The input gate, the forget gate, the control gate, and the output gate are denoted by i_t, f_t, c_t, and o_t, respectively. The input gate decides which information can be transferred from the earlier cell to the current cell. The forget gate is used to store the information from the input of previous memory or otherwise. The control gate controls the update of the cell. Finally, the output gate is used to update the hidden layer h_{t-1} and update the output.

Contrastive Predictive Coding (CPC)

Traditional methods for handling missing data often involve filling in the missing values and then applying predictive models on the imputed data [171]. Choosing a suitable imputation scheme is complex, dataset-specific, and relies on domain expertise. Furthermore, this results in a two-step process that prevents the prediction model from adequately exploring the missingness patterns [171]. As shown in [171], such informative missingness may

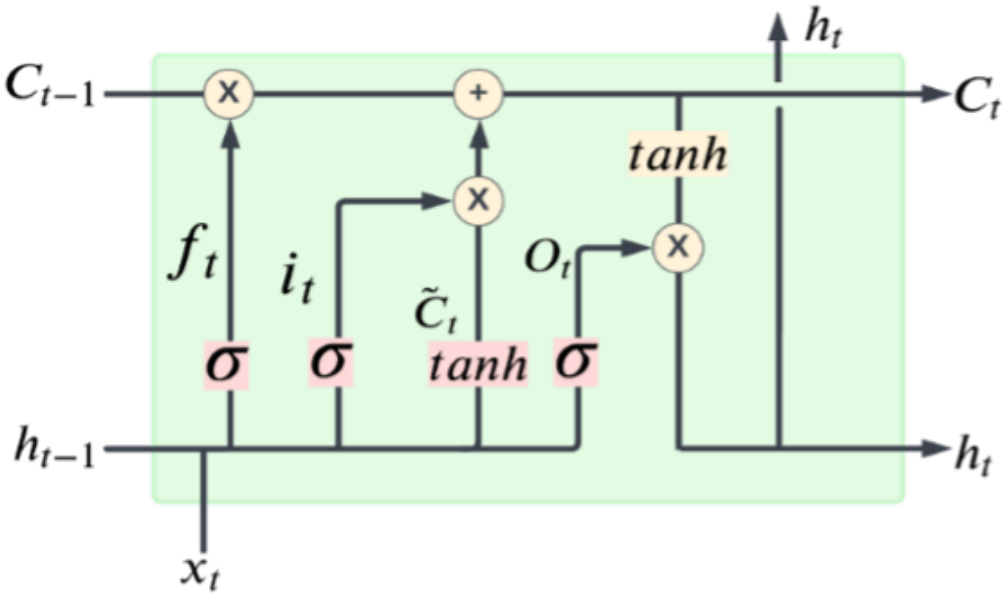

Figure 5.4: An illustration of an LSTM memory cell.

encode helpful information about the target labels. Regarding state-of-the-art time series models, only a few incorporate these missingness patterns when learning from the data. One example is the Bidirectional Recurrent Imputation for Time Series (BRITS) [172], which simultaneously imputes the missing values and performs classification/regression within a joint neural graph. Another similar method is the GRU with trainable Decays (GRU-D) [171]. Both these methods take advantage of two representations of informative missingness: masking and time interval [171]. Two models based on ordinary differential equations, the ODE-RNN and the Latent-ODE, have also shown promising results on irregularly-sampled data [173]. However, the computational complexity of these models is high, which might lead to not finding the optimal hyperparameters [174]. Self-supervised pre-training aims to learn a good initialization point for the supervised setting instead of changing the supervised learning objective. Most models used for this purpose are based on autoencoders [175], while some of the most recent and promising methods are based on the idea of predictive coding, such as Contrastive Predictive Coding (CPC) [176]. It learns from sequential data by trying to predict $n \geq 1$ steps ahead of the current step. We explore

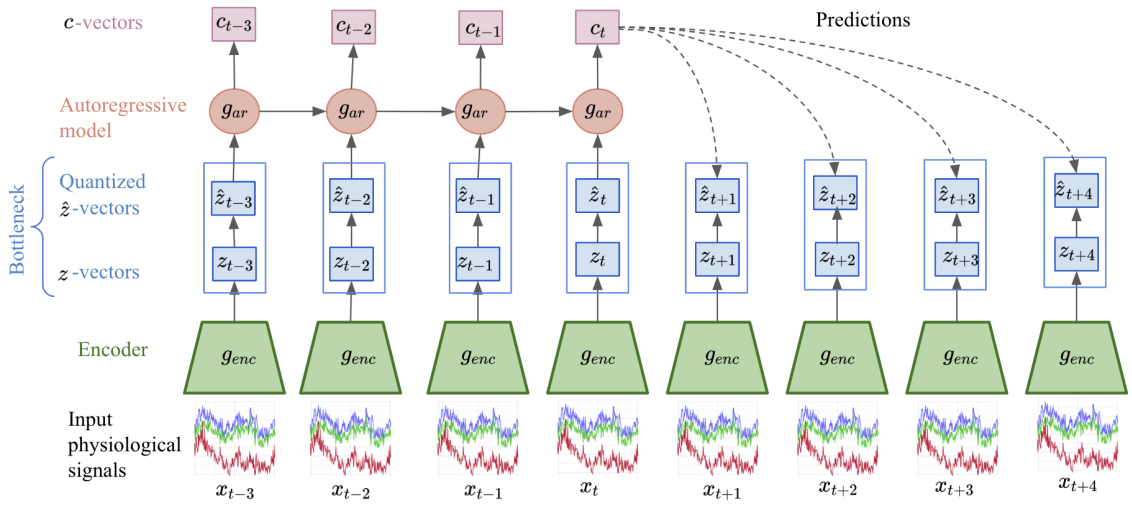

Figure 5.5: The architecture of Contrastive Predictive Coding (CPC) network illustrating using four time steps in the past sequence $(x_{t-3}, x_{t-2}, x_{t-1}, x_t)$ to predict a sequence of four time steps $(x_{t+1}, x_{t+2}, x_{t+3}, x_{t+4})$ in the future.

two self-supervised representation learning algorithms for pain forecasting: Contrastive Predictive Coding (CPC) [176] and Variational Auto-encoders (VAE) [134].

We show the architecture of the CPC network n Figure 5.5. A past vital sign sequence is partitioned into multiple non-overlapping sub-sequences $x_1, x_2, ..., x_t$ of the same length. An encoder network g_{enc} maps each of the input sub-sequence of observations x_t to a latent representation $z_t = g_{enc}(x_t)$. In other words, the encoder network is used to learn a low-dimensional representation z_t of high-dimensional sensor input x_t. Similarly, a future physiological signal sequence is divided into k sub-sequences $x_{t+1}, x_{t+2}, ..., x_{t+k}$, and latent representation $z_{t+1}, z_{t+2}, ..., z_{t+k}$ are extracted by the same encoder network g_{enc}. Then, an autoregressive model g_{ar} summarizes all z_1 to z_t learned from past sequence and produces a context latent representation $c_t = g_{ar}(z_1, z_2, ..., z_t)$. The latent context representation c_t contains all information in the long past vital signs sequence. Therefore, we can use the context latent representation c_t to make the prediction $\widehat{z}_{t+k}$ for the latent representation of every step k in the future sequence. Then, the ground truth representation z_{t+k} and the predicted representation $\widehat{z}_{t+k}$ are compared. The dot product $z_{t+k}^T \cdot \widehat{z}_{t+k}$ is

used to denote the similarity between the truth and the prediction. Then a density ratio $f_k(x_{t+k}, c_t) = exp(z_{t+k}^T \cdot \widehat{z}_{t+k})$ is introduced to measure the quality of the prediction $\widehat{z}_{t+k}$ generated by x_{t+k} and c_t. Based on the idea of Noise-Contrastive Estimation (NCE) [177], given a set X of N random samples containing one positive sample and $N-1$ negative samples, the cross-entropy loss for classifying the positive samples correctly can be defined as:

$$\mathcal{L}_N = -\mathbb{E}_x \left[log \frac{f_k(x_{t+k}, c_t)}{\sum_{x_j \epsilon X} f_k(x_j, c_t)} \right] \tag{5.2}$$

Predictive Clustering

Clustering is an unsupervised learning process where an algorithm brings similar data points closer without any "ground truth" labels. The similarity between data points is measured with a distance metric, commonly Euclidean distance. In general, the Euclidean distance metric (or other types of Minkowski metric) is used to find an average of all the data within the clusters. However, its one-to-one mapping nature cannot capture the average shape of the two time series, in which case Dynamic Time Warping (DTW) [178] is more favorable.

Clustering different time series data is challenging as each data point acts as an ordered sequence. Classically, the most common approach involves flattening the time series (sequence) into a table, with a column for each time point (or an aggregation of the entire sequence), and applying standard clustering algorithms like k-means. However, these clustering algorithms use standard measures such as Euclidean distance, which is often not the best for time series (ordered sequences). Hence, we replace the default distance measure with DTW to compare the temporal sequences that can measure the similarity between two sequences that do not align with each other rigidly in time, speed, or length.

Unlike the Minkowski distance function, dynamic time warping breaks the one-to-one alignment limitation and supports non-equal-length time series. It uses a dynamic programming technique to find all possible paths and selects the one that yields a minimum distance between the two sequences of time series using a distance matrix, where each element in the matrix is a cumulative distance of the minimum of the three surrounding neighbors. Suppose we have two time series, a sequence $Q = q_1, q_2, \ldots, q_i, \ldots, q_n$ and a sequence $C = c_1, c_2, \ldots, c_j, \ldots, c_m$. First, we create an n x m matrix, where every (i, j) element of the matrix is the cumulative distance of the distance at (i, j) and the minimum of the three elements neighboring the (i, j) element, where $0 < i \leq n$ and $0 < j \leq m$. We can define the (i, j) element as:

$$e_{ij} = d_{ij} + min\left\{e_{(i-1)(j-1)}, e_{i(j-1)}, e_{(i-1)j}\right\} \tag{5.3}$$

where $d_ij = (c_i + q_j)^2$ and e_ij is (i, j) element of the matrix which is the summation between the squared distance of q_i and c_j, and the minimum cumulative distance of the three elements surrounding the (i, j) element. Then, to find an optimal path, we have to choose the path that gives minimum cumulative distance at (n, m). The distance is defined as:

$$D_{DTW}\left(Q, C\right) = \underset{\forall w \in P}{min}\left\{\sqrt{\sum_{k=1}^{K} d_{wk}}\right\} \tag{5.4}$$

where P is a set of all possible warping paths, and wk is (i, j) at k^th element of a warping path and K is the length of the warping path.

5.3 Experiments

5.3.1 Autoregressive Integrated Moving Average model

As explained above, the ARIMA model possessed three parameters p, q, and d. Significant lags at 5 in the autocorrelation and partial autocorrelation function plots (Figure 5.2) extend beyond the dashed blue lines and indicate poor model fit. We decided the p and q parameters range for the grid search as 0 and 5. Thus, we applied a grid to search by passing the integer values in the range [0, 5] for both p and q and decided on the value of the p and q parameters to be 1 with AIC value of the model as 584.9 and both ACF and PACF plots showed no significant lags (Figure 5.3). Based on the ADF test [179], we found the time series of pain score to be non-stationary and required one-time differencing (d = 1).

5.3.2 Multilayer Perceptron and Long Short-Term Memory models

MLP and LSTM models were trained on physiology and pain scores on the interpolated time series dataset. We adjusted the hyper-parameters (layers, neurons, batch size, and epochs) in MLP and LSTM models to have an idea of the range (minimum and maximum) value of the parameters. The hyperparameters higher and lower than the values for which we did not get any improvements in the model's fit provided us with the minimum and the maximum range. Next, we varied the hyper-parameters in the obtained range and evaluated the model performance. We validated the model training on test dataset. Since we did not want to provide the memory-based LSTM model an advantage over the memory-less MLP model; we tested the same range of hyper-parameters for both the MLP and LSTM models. After getting the best set of hyper-parameters, we evaluated the model performance on these best hyper-parameters 40 times as there is a run-to-run variability in the model's output on training data. We finalized the model with the least objective function value among the 40 model runs as the final prediction from the model. For training both models

(MLP and LSTM), we used the ReLU activation function. These models were created in Keras using the Tensorflow backend.

5.3.3 Contrastive Predictive Coding (CPC) and Variational Autoencoder (VAE)

In self-supervised learning, the network is trained to predict future physiological data from extensive unlabeled past physiological data. During the training process, our self-supervised learning network learned latent representations that were used to infer future pain states using a regression model. We used a similar architecture by Yang et.al. [129] to learn representations from physiological signals using a CPC network and a VAE architecture by our previous work discussed in Chapter 4 [141]. Specifically, we used a three-layer Convolutional Neural Network (CNN) [180] as the encoder in CPC model. We then used a gated recurrent unit (GRU) based Recurrent Neural Network (RNN) [180] for the autoregressive part of the model with 64 dimensional of hidden states. The output of the GRU based RNN model c_t is then used as the feature for pain forecasting task. The pretext task network was trained using the Adam optimizer with a batch size of 128 and a learning rate of 10^{-3}. The network structure and hyperparameters were tuned based on experiments to maximize the accuracy of the pretext task.

Next, we trained a regression model to predict future pain values using the representations learned (the output of the CPC and VAE network) as input features. To summarize, we used a past physiological signal sequence in the trained self-supervised network to generate the latent representation, which is used as the input feature of a regression model to predict the pain score reported at a future time step. In the downstream task, we trained a regression model to predict future pain values using the latent context representation c_t (the output of the autoregressive network) as input features. Specifically, a past vital signs sequence was fed into the trained CPC network to generate the context latent representation c_t. Then c_t

was used as the input feature of a regression model to predict raw (not interpolated) pain scores reported at a future time step. Due to the lack of pain score labels (as we used raw pain scores instead of the interpolated pain scores for prediction), we utilized random forest [83] as the supervised regression model for pain forecasting. We chose this model because ensemble methods are more robust and have advantages in dealing with small sample sizes [181].

5.3.4 Predictive Clustering

We compute the cluster centroids with respect to DTW by minimizing the sum of squared DTW distance between the cluster centroid and the series in the cluster. We employed k-means clustering for each year of patient data and generated cluster labels for all patients for each year. Next, we used the cluster labels as an additional feature to our pain forecasting models to predict the future cluster label (ground truth) for next year.

5.4 Results

Tables 5.2 and 5.3 show the MAE and R^2 for forecasting pain scores for mixed patients and individual patient models respectively.

We present the results using the best predictions from 40 runs from each model. For mixed patient forecasting, we combined all the patient data. The training data consisted of a past sequence from a patient, and we generated the forecast on a future sequence from a random patient, including the patient in the training set. In the individualized patient experiment, we included past and future sequences from the same patient.

Next, for each year, we used K means clustering using DTW distance on all patients' physiology and pain scores data to obtain an optimal number of seven clusters [with a Normalized Mutual Information (NMI) score of 0.35, purity score of 0.67, Silhouette Index of 0.12]. We treat the cluster labels generated as ground truth labels for each patient for

Table 5.2: Pain forecasting results for individualised patient model

	1 hour		2 hour		4 hour	
	MAE	R^2	MAE	R^2	MAE	R^2
VAE + RF Regression Model	**1.16**	**0.91**	**1.295**	**0.91**	**1.484**	**0.90**
CPC + RF Regression Model	1.184	0.89	1.369	0.89	1.57	0.88
LSTM	1.261	0.84	1.482	0.84	1.663	0.84
Dense Regression (MLP)	1.294	0.79	1.342	0.79	1.47	0.78
ARIMA	1.63	0.74	1.64	0.73	1.70	0.73
RF Regression Model	1.85	0.62	1.86	0.63	1.90	0.61

Table 5.3: Pain forecasting results for mixed patient model

	1 hour		2 hour		4 hour	
	MAE	R^2	MAE	R^2	MAE	R^2
VAE + RF Regression Model [141]	**0.58 (+/- 0.39)**	**0.91**	**0.63 (+/- 0.42)**	**0.89**	**0.78(+/- 0.42)**	**0.82**
CPC + RF Regression Model [129]	0.76 (+/-0.54)	0.90	0.79 (+/- 0.54)	0.88	0.96 (+/- 0.58)	0.88
LSTM	0.78 (+/-0.66)	0.88	0.73 (+/- 0.62)	0.87	0.98 (+/- 0.65)	0.87
Dense Regression (MLP)	1.01 (+/- 0.75)	0.76	1.142 (+/- 0.78)	0.76	1.27 (+/- 0.84)	0.76
ARIMA	1.24 (+/-0.87)	0.65	1.28 (+/-0.91)	0.61	1.45 (+/-0.93)	0.63
RF Regression Model	1.37 (+/-0.89)	0.62	1.42 (+/-0.91)	0.61	1.53 (+/-0.94)	0.62

each year. Our goal is to understand if we can predict future cluster alignment of patients. Hence, we train the same models discussed above on a past sequence of physiology data, pain scores, and cluster labels to forecast the cluster label for the following year. We report the AUROC of our models with raw pain scores as test data for years 2,3,4 and 5 in Table 5.4. The training for each is the data for all previous years, i.e., for the forecast of year 5, we trained models on interpolated data from years 1,2,3, and 4 and tested on original data available for year 5.

The best MAE (= 0.58) on test data was obtained for our LSTM based VAE model which contained 2 hidden layers, 4 neurons in each hidden layer, 20 batch size, and for 30 epochs. In general, all the models resulted in lower MAE and higher R^2 scores in individualised models. As seen in the tables, both the MLP and LSTM models outperformed the RF Regression model (baseline) as well as the ARIMA model. Also, the self-supervised LSTM based VAE model performed the best among all the models.

We report the area under the receiver operating characteristic (AUROC) score for evaluating our long-term cluster forecasting models. Also, we compare the performance of our best

Table 5.4: Predictive Clustering for Long-term pain forecasting (AUROC)

	Year 2	Year 3	Year 4	Year 5
MLP	0.633	0.691	0.729	0.775
CPC + RF	0.721	0.793	0.852	0.886
VAE + RF	**0.743**	**0.832**	**0.893**	**0.921**

performing models from short-term forecasting as shown in Table 5.4. First, as expected, our LSTM based self supervised VAE netowrk performed best in long-term pain forecasting. Second, each model performed best in forecasting the cluster assignment for year 5 as it had more data to learn from (year 1-4).

5.5 Discussion

The primary objective of this research was short-term pain forecasting and evaluating the performance of existing statistical (ARIMA), supervised neural (MLP and LSTM), and self-supervised (CPC, VAE) models for individual and mixed patient scenarios. Another objective of this paper was to systematically evaluate a predictive clustering based approach for long-term pain forecasting. Overall, we expected the self-supervised models to perform better than the statistical and supervised neural models. First, as per our expectation, the best performance in terms of error was found from the VAE trained network, followed by CPC trained network, LSTM, MLP, ARIMA and RF regression models. A likely reason for these results is that ARIMA models are perhaps not able to capture the non-linearities present in the time-series data. Thus, overall, these models tended to perform not as well compared to other models. Also, overall, the neural network models (MLP and LSTM) performed similarly and better than the persistence and ARIMA models. That is likely because our dataset is were non-linear and neural network models, by their design, could account for the non-linear trends in datasets. However, another reason for this result could be simply because the self-supervised network models possess several weights (parameters), whereas the ARIMA model possesses only three parameters.

5.5.1 Changing Patient Phenotype over time : Long-term forecasting

In this subsection, we demonstrate run-time examples of how our predictive clustering approach was able to flexibly update the cluster assignments over time with respect to the future pain in the next year. We present a case study of six representative patients as discussed below and shown in Figure 5.6. We show the average pain range for each of these patients in Table 5.5.

• Patient A had mild pain score (pain score 1-3) in the first year at the beginning of the study. In the second year, he/she had moderate average pain (pain score 4-7). Our clustering model was able to predict the temporal phenotype assigned to this patient as similar to that of patient F who had moderate average pain (pain score 4-7). As shown in Figure 5.7, we see that the systolic blood pressure for both these patients follow a similar trend (decreasing). Furthermore, our clustering algorithm phenotyped patient A and patient E together in the third year to a cluster predominantly having low/mild pain scores. Both of them had mild pain in the first year. So, our approach could change the phenotype of patient A from low/mid pain to moderate pain and again back to low/mild pain. We can see from Figure 5.7 that patients A and E followed a decreasing trend in systolic blood pressure from first year till third year. Furthermore our model predicted accurately patient D to be in the same cluster with patient A and F having moderate pain and decreasing blood pressure in the second year. Patient D had mild pain in first year, and moderate pain in the second year.

• Patient B had an average moderate pain score in the first year and maintained a moderate pain score in second year, no pain in third and fourth year, and an average moderate pain in the fifth year. Our clustering model predicted that patient B and patient C belonged to the same cluster in second year (both had moderate pain). In the third year also, they were clustered together although patient B had no pain and patient C had moderate pain. We

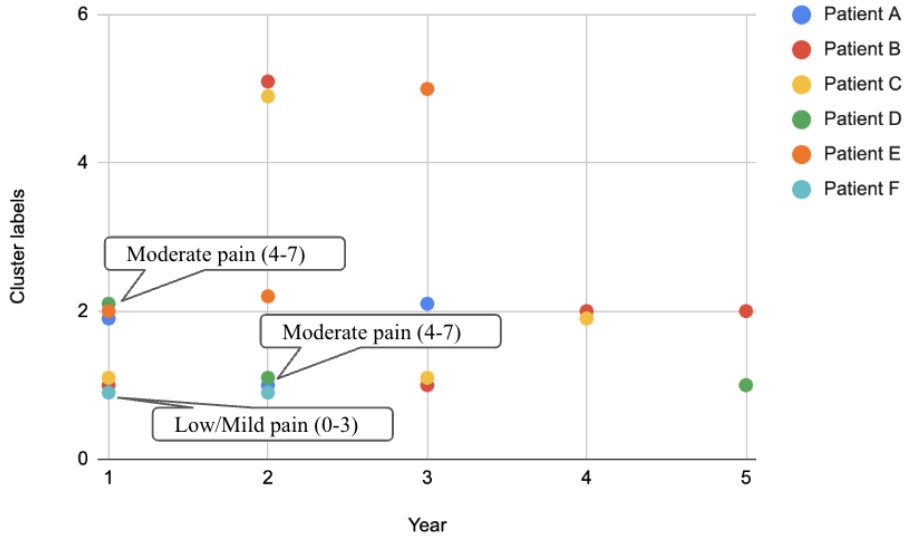

Figure 5.6: An illustration of run-time examples of our clustering based long term pain forecasting on six representative patients.

Table 5.5: Average pain scores of six representative patients for case study

Year	Patient A	Patient B	Patient C	Patient D	Patient E	Patient F
1	3.3	6.5	6.6	1.6	2.2	7.5
2	5.7	4.5	7	6.1	0	7.1
3	0	0	7.2		0	7.3
4		0	6.4	5.5		8
5		5.9				

analyzed from the Figure 5.7 that although belonging to different pain range, they followed a similar trend in systolic blood pressure. Interestingly, our algorithm then allocated them to separate clusters in the fourth year. While patient B was allocated to a cluster with mixed pain scores, patient C was clustered with patients with moderate pain score.

Based on our observations, we would like to point out a limitation in predictive clustering, the trade-off between the clustering performance (for better interpretability) – which quantifies how the data samples are homogeneous within each cluster and heterogeneous across clusters with respect to the future outcomes of interest – and the prediction performance is a common issue. The most critical parameter that governs this trade-off is the number of clusters. More specifically, increasing the number of clusters will give the predictive

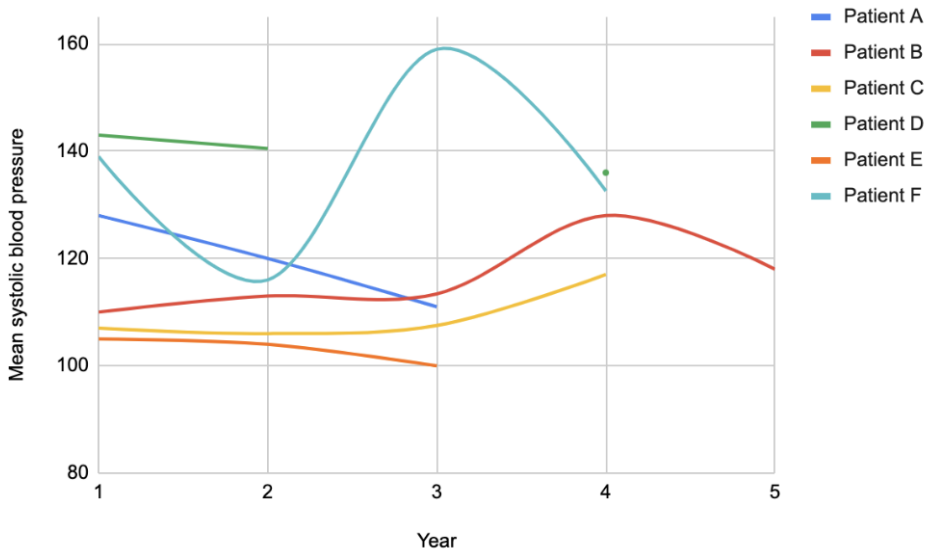

Figure 5.7: Change in systolic blood pressure (mean) across years for the six representative patients.

clusters higher diversity to represent the output distribution and, thus, increase the prediction performance while decreasing the clustering performance. One extreme example is that there are as many clusters as data samples which will make the identified clusters fully individualized; consequently, each cluster will lose interpretability as it no longer groups similar data samples.

5.6 Summary

In this chapter, we evaluated the performance of memory-less neural network models [e.g., multilayer perceptron (MLP)] with memory-based neural network models [e.g., long short-term memory (LSTM) based self-supervised models] and statistical time-series techniques for performing short-term (hourly) and long-term (yearly) time-series predictions of longitudinal healthcare data. Our data-driven self-supervised approaches outperformed the purely supervised learning methods, as well as statistical methods obtaining a MAE is of 0.58 with a standard deviation of 0.39 and a R^2 of 0.91 when the forecast horizon is 1 hour. We observed that the forecast results for a mixed patient scenario is significantly better than that for an individualized model. We observed similar trend in performance with our predictive

clustering approach with a random forest model trained on our LSTM based VAE network achieving an AUROC of 0.921 in long-term patient phenotype forecasting. Finally, we demonstrated that self-supervised predictive clustering approach can be utilized to interpret the changes in patient phenotyping and better understand future pain trajectories.

6 Conclusion and Future Work

In this chapter, we provide a brief overview of this dissertation. Then, we discuss potential future research related to improving pain management in SCD.

6.0.1 Conclusion

The objective of the study in this dissertation is to leverage data-driven strategies to improve pain management in patients with in SCD. We tackled this problem in an incremental step-by-step manner.

Firstly, we presented a pain assessment model based on physiological signals data from EHRs. We evaluated our pain prediction model at intra-individual level and inter-individual level at varying pain rating scales. We showed that intra-indivial pain prediction models had better performances than inter-individual models when trained with sufficient data irrespective of the nature of hospital visit (outpatient, evaluation, or inpatient). We also found that pain prediction based on the 4-point rating scale was appropriate for clinical practices with a high prediction accuracy and pain assessment sensitivity.

Secondly, we showed that medication information in addition to the physiological signals data improves the performance of our pain estimation models. Furthermore, we observed that self-supervised models perform better than traditional machine learning models when provided with sufficient data.

Finally, we solved the pain forecasting problem with restricted pain labels using a self-supervised learning method for both short-term and long-term forecasting. Our self-supervised learning

model was consistently better than the model trained in a purely supervised manner at varying forecast horizons. We also demonstrated that our self-supervised patient phenotyping approach can be used to understand the evolution of future patient pain profiles.

6.0.2 Future Work

The work presented in this dissertation can be further extended by implementing a real-time pain management system such as a mobile application to be used by the clinicians as another dimension to their treatment strategy. We showed that for pain estimation as well for future pain forecasting (to understand future patient response), our models perform better at an individualised level. It highlighted the importance of implementing individualized pain management models. Another dimension to consider in the modeling can be the effect of weather in the modeling process. Also, the role of demographics can be explored in pain assessment and patient response management. This research has a number of implications for healthcare data analytics. First, an implication from our results is that data driven self-supervised network models could be used in forecasting pain from EHR data. Although we considered pain forecasting in short-term and long-term in this work, our results are likely to hold for other medical diagnosis scenarios. Second, another implication of our results is that it may be expensive to train neural network models; however, once these models are trained, they are easy and computationally and temporarily inexpensive to apply on new patient data. Therefore, we believe that the self-supervised approaches would be useful to clinicians, caregivers and patients. For example, the proposed models could be bundled into a mobile or desktop application that helps patients better manage their health by forecasting their future pain, or it can help the clinicians to forecast a patient's future pain and design appropriate treatment strategy.

REFERENCES

[1] D. Manwani and P. S. Frenette, "Vaso-occlusion in sickle cell disease: pathophysiology and novel targeted therapies," *Blood, The Journal of the American Society of Hematology*, vol. 122, no. 24, pp. 3892–3898, 2013.

[2] J. E. Brown, N. Chatterjee, J. Younger, and S. Mackey, "Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation," *PloS one*, vol. 6, no. 9, p. e24124, 2011.

[3] W. R. Hersh, M. G. Weiner, P. J. Embi, J. R. Logan, P. R. Payne, E. V. Bernstam, H. P. Lehmann, G. Hripcsak, T. H. Hartzog, J. J. Cimino *et al.*, "Caveats for the use of operational electronic health record data in comparative effectiveness research," *Medical care*, vol. 51, no. 8 0 3, p. S30, 2013.

[4] H. Harkema, W. W. Chapman, M. Saul, E. S. Dellon, R. E. Schoen, and A. Mehrotra, "Developing a natural language processing application for measuring the quality of colonoscopy procedures," *Journal of the American Medical Informatics Association*, vol. 18, no. Supplement_1, pp. i150–i156, 2011.

[5] A. Mehrotra, E. S. Dellon, R. E. Schoen, M. Saul, F. Bishehsari, C. Farmer, and H. Harkema, "Applying a natural language processing tool to electronic health records to assess performance on colonoscopy quality measures," *Gastrointestinal endoscopy*, vol. 75, no. 6, pp. 1233–1239, 2012.

[6] J. S. Ancker, L. M. Kern, A. Edwards, S. Nosal, D. M. Stein, D. Hauser, R. Kaushal, and with the HITEC Investigators, "Associations between healthcare quality and use

of electronic health record functions in ambulatory care," *Journal of the American Medical Informatics Association*, vol. 22, no. 4, pp. 864–871, 2015.

[7] ——, "How is the electronic health record being used? use of ehr data to assess physician-level variability in technology use," *Journal of the American Medical Informatics Association*, vol. 21, no. 6, pp. 1001–1008, 2014.

[8] J. I. Allen, "Quality measures for colonoscopy: where should we be in 2015?" *Current gastroenterology reports*, vol. 17, no. 3, p. 10, 2015.

[9] G. D. Schiff, D. W. Bates, P. Hartzband, J. Groopman, and G. Schiff, "Can electronic clinical documentation help prevent diagnostic errors?" *New England Journal of Medicine*, vol. 362, no. 12, p. 1066, 2010.

[10] D. W. Bates and A. A. Gawande, "Improving safety with information technology," *New England journal of medicine*, vol. 348, no. 25, pp. 2526–2534, 2003.

[11] R. Kaushal, L. M. Kern, Y. Barrón, J. Quaresimo, and E. L. Abramson, "Electronic prescribing improves medication safety in community-based office practices," *Journal of general internal medicine*, vol. 25, no. 6, pp. 530–536, 2010.

[12] E. B. Devine, W. Hollingworth, R. N. Hansen, N. M. Lawless, J. L. Wilson-Norton, D. P. Martin, D. K. Blough, and S. D. Sullivan, "Electronic prescribing at the point of care: a time–motion study in the primary care setting," *Health services research*, vol. 45, no. 1, pp. 152–171, 2010.

[13] E. Abramson, R. Kaushal, and J. Vest, "Improving immunization data management: an editorial on the potential of electronic health records," *Expert review of vaccines*, vol. 13, no. 2, pp. 189–191, 2014.

[14] J. Shepard, E. Hadhazy, J. Frederick, S. Nicol, P. Gade, A. Cardon, J. Wilson, Y. Vetteth, and S. Madison, "Using electronic medical records to increase the

efficiency of catheter-associated urinary tract infection surveillance for national health and safety network reporting," *American journal of infection control*, vol. 42, no. 3, pp. e33–e36, 2014.

[15] C. M. Rochefort, A. D. Verma, T. Eguale, T. C. Lee, and D. L. Buckeridge, "A novel method of adverse event detection can accurately identify venous thromboembolisms (vtes) from narrative electronic health record data," *Journal of the American Medical Informatics Association*, vol. 22, no. 1, pp. 155–165, 2015.

[16] B. Shickel, P. J. Tighe, A. Bihorac, and P. Rashidi, "Deep ehr: a survey of recent advances in deep learning techniques for electronic health record (ehr) analysis," *IEEE journal of biomedical and health informatics*, vol. 22, no. 5, pp. 1589–1604, 2017.

[17] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2018.

[18] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor AI: Predicting clinical events via recurrent neural networks," in *Machine learning for healthcare conference*. PMLR, 2016, pp. 301–318.

[19] J. R. Walsh, A. M. Smith, Y. Pouliot, D. Li-Bland, A. Loukianov, and C. K. Fisher, "Generating digital twins with multiple sclerosis using probabilistic neural networks," *arXiv preprint arXiv:2002.02779*, 2020.

[20] C. K. Fisher, A. M. Smith, and J. R. Walsh, "Machine learning for comprehensive forecasting of alzheimer's disease progression," *Scientific reports*, vol. 9, no. 1, pp. 1–14, 2019.

[21] A. Moser, C. Appl, S. Brüning, and V. C. Hass, "Mechanistic mathematical models as a basis for digital twins," in *Digital Twins*. Springer, 2020, pp. 133–180.

[22] H. Merskey and N. Bogduk, "Iasp press; seattle: 1994," *Classification of chronic pain*, p. 1.

[23] D. C. Rees, A. D. Olujohungbe, N. E. Parker, A. D. Stephens, P. Telfer, J. Wright, B. C. for Standards in Haematology General Haematology Task Force by the Sickle Cell Working Party *et al.*, "Guidelines for the management of the acute painful crisis in sickle cell disease," *British journal of haematology*, vol. 120, no. 5, pp. 744–752, 2003.

[24] C. Machinery, "Computing machinery and intelligence-am turing," *Mind*, vol. 59, no. 236, p. 433, 1950.

[25] A. L. Samuel, "Some studies in machine learning using the game of checkers," *IBM Journal of research and development*, vol. 3, no. 3, pp. 210–229, 1959.

[26] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain." *Psychological review*, vol. 65, no. 6, p. 386, 1958.

[27] T. Addis, "Towards an 'expert'diagnostic system," *ICL Technical Journal*, vol. 1, pp. 79–105, 1956.

[28] T. B. Murdoch and A. S. Detsky, "The inevitable application of big data to health care," *Jama*, vol. 309, no. 13, pp. 1351–1352, 2013.

[29] S. E. Dilsizian and E. L. Siegel, "Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment," *Current cardiology reports*, vol. 16, no. 1, p. 441, 2014.

[30] E. Kolker, V. Özdemir, and E. Kolker, "How healthcare can refocus on its super-customers (patients, n= 1) and customers (doctors and nurses) by leveraging

lessons from amazon, uber, and watson," *Omics: a journal of integrative biology*, vol. 20, no. 6, pp. 329–333, 2016.

[31] V. L. Patel, E. H. Shortliffe, M. Stefanelli, P. Szolovits, M. R. Berthold, R. Bellazzi, and A. Abu-Hanna, "The coming of age of artificial intelligence in medicine," *Artificial intelligence in medicine*, vol. 46, no. 1, pp. 5–17, 2009.

[32] B. X. Tran, G. T. Vu, G. H. Ha, Q.-H. Vuong, M.-T. Ho, T.-T. Vuong, V.-P. La, M.-T. Ho, K.-C. P. Nghiem, H. L. T. Nguyen *et al.*, "Global evolution of research in artificial intelligence in health and medicine: a bibliometric study," *Journal of clinical medicine*, vol. 8, no. 3, p. 360, 2019.

[33] D. Juric, G. Stoilos, A. Melo, J. Moore, and M. Khodadadi, "A system for medical information extraction and verification from unstructured text," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 08, 2020, pp. 13 314–13 319.

[34] M. L. Graber, N. Franklin, and R. Gordon, "Diagnostic error in internal medicine," *Archives of internal medicine*, vol. 165, no. 13, pp. 1493–1499, 2005.

[35] N. S. Weingart, R. M. Wilson, R. W. Gibberd, and B. Harrison, "Epidemiology of medical error," *Bmj*, vol. 320, no. 7237, pp. 774–777, 2000.

[36] B. Winters, J. Custer, S. M. Galvagno, E. Colantuoni, S. G. Kapoor, H. Lee, V. Goode, K. Robinson, A. Nakhasi, P. Pronovost *et al.*, "Diagnostic errors in the intensive care unit: a systematic review of autopsy studies," *BMJ quality & safety*, vol. 21, no. 11, pp. 894–902, 2012.

[37] C. S. Lee, P. G. Nagy, S. J. Weaver, and D. E. Newman-Toker, "Cognitive and system factors contributing to diagnostic errors in radiology," *American Journal of Roentgenology*, vol. 201, no. 3, pp. 611–617, 2013.

[38] D. B. Neill, "Using artificial intelligence to improve hospital inpatient care," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 92–95, 2013.

[39] M. Anderson and S. L. Anderson, "How should ai be developed, validated, and implemented in patient care?" *AMA journal of ethics*, vol. 21, no. 2, pp. 125–130, 2019.

[40] H. J. Murff, F. FitzHenry, M. E. Matheny, N. Gentry, K. L. Kotter, K. Crimin, R. S. Dittus, A. K. Rosen, P. L. Elkin, S. H. Brown *et al.*, "Automated identification of postoperative complications within an electronic medical record using natural language processing," *Jama*, vol. 306, no. 8, pp. 848–855, 2011.

[41] A. M. Darcy, A. K. Louie, and L. W. Roberts, "Machine learning and the profession of medicine," *Jama*, vol. 315, no. 6, pp. 551–552, 2016.

[42] G. K. Nave Jr, S. Padhee, A. Alambo, T. Banerjee, N. Shah, and D. M. Abrams, "Clustering of pain dynamics in sickle cell disease from sparse, uneven samples," *arXiv preprint arXiv:2108.13963*, 2021.

[43] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[44] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.

[45] S. Somashekhar, R. Kumarc, A. Rauthan, K. Arun, P. Patil, and Y. Ramya, "Abstract s6-07: Double blinded validation study to assess performance of ibm artificial intelligence platform, watson for oncology in comparison with manipal multidisciplinary tumour board–first study of 638 breast cancer cases," 2017.

[46] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[47] C. E. Bouton, A. Shaikhouni, N. V. Annetta, M. A. Bockbrader, D. A. Friedenberg, D. M. Nielson, G. Sharma, P. B. Sederberg, B. C. Glenn, W. J. Mysiw *et al.*, "Restoring cortical control of functional movement in a human with quadriplegia," *Nature*, vol. 533, no. 7602, pp. 247–250, 2016.

[48] D. Farina, I. Vujaklija, M. Sartori, T. Kapelner, F. Negro, N. Jiang, K. Bergmeister, A. Andalib, J. Principe, and O. C. Aszmann, "Man/machine interface based on the discharge timings of spinal motor neurons after targeted muscle reinnervation," *Nature biomedical engineering*, vol. 1, no. 2, pp. 1–12, 2017.

[49] B. Marr, "First fda approval for clinical cloud-based deep learning in healthcare," *Forbes. https://www. forbes. com/sites/bernardmarr/2017/01/20/first-fda-approval-for-clinical-cloud-based-deep-learning-in-he* 2017.

[50] E. Long, H. Lin, Z. Liu, X. Wu, L. Wang, J. Jiang, Y. An, Z. Lin, X. Li, J. Chen *et al.*, "An artificial intelligence platform for the multihospital collaborative management of congenital cataracts," *Nature biomedical engineering*, vol. 1, no. 2, pp. 1–8, 2017.

[51] J. Finkelstein and I. cheol Jeong, "Machine learning approaches to personalize early prediction of asthma exacerbations," *Annals of the New York Academy of Sciences*, vol. 1387, no. 1, p. 153, 2017.

[52] M. A. Sarwar, N. Kamal, W. Hamid, and M. A. Shah, "Prediction of diabetes using machine learning algorithms in healthcare," in *2018 24th International Conference on Automation and Computing (ICAC)*. IEEE, 2018, pp. 1–6.

[53] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[54] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine learning and data mining methods in diabetes research," *Computational and structural biotechnology journal*, vol. 15, pp. 104–116, 2017.

[55] A. A. Lazakidou, *Web-based applications in healthcare and biomedicine*. Springer Science & Business Media, 2009, vol. 7.

[56] E. Du, M. Diez-Silva, G. J. Kato, M. Dao, and S. Suresh, "Kinetics of sickle cell biorheology and implications for painful vasoocclusive crisis," *Proceedings of the National Academy of Sciences*, vol. 112, no. 5, pp. 1422–1427, 2015.

[57] S. Knowlton, I. Sencan, Y. Aytar, J. Khoory, M. Heeney, I. Ghiran, and S. Tasoglu, "Sickle cell detection using a smartphone," *Scientific reports*, vol. 5, no. 1, pp. 1–11, 2015.

[58] N. Shah, J. Jonassaint, and L. De Castro, "Patients welcome the sickle cell disease mobile application to record symptoms via technology (smart)," *Hemoglobin*, vol. 38, no. 2, pp. 99–103, 2014.

[59] J. N. Milton, V. R. Gordeuk, J. G. Taylor, M. T. Gladwin, M. H. Steinberg, and P. Sebastiani, "Prediction of fetal hemoglobin in sickle cell anemia using an ensemble of genetic risk prediction models," *Circulation: Cardiovascular Genetics*, vol. 7, no. 2, pp. 110–115, 2014.

[60] C. Allayous, S. Clémençon, B. Diagne, R. Emilion, and T. Marianne, "Machine learning algorithms for predicting severe crises of sickle cell disease," 2008.

[61] A. V. Solanki *et al.*, "Data mining techniques using weka classification for sickle cell disease," *International Journal of Computer Science and Information Technologies*, vol. 5, no. 4, pp. 5857–5860, 2014.

[62] B. Eftekhar, K. Mohammad, H. E. Ardebili, M. Ghodsi, and E. Ketabchi, "Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data," *BMC medical informatics and decision making*, vol. 5, no. 1, pp. 1–8, 2005.

[63] P. T. Dalvi and N. Vernekar, "Anemia detection using ensemble learning techniques and statistical models," in *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*. IEEE, 2016, pp. 1747–1751.

[64] N. Sharma and V. Khullar, "Comparative review of artificial neural network machine learning for diagnosing aneamia in pregnant ladies," *i-Manager's Journal on Information Technology*, vol. 5, no. 4, p. 33, 2016.

[65] P. Escandell-Montero, M. Chermisi, J. M. Martinez-Martinez, J. Gomez-Sanchis, C. Barbieri, E. Soria-Olivas, F. Mari, J. Vila-Francés, A. Stopper, E. Gatti *et al.*, "Optimization of anemia treatment in hemodialysis patients via reinforcement learning," *Artificial intelligence in medicine*, vol. 62, no. 1, pp. 47–60, 2014.

[66] M. Khalaf, A. J. Hussain, D. Al-Jumeily, R. Keenan, R. Keight, P. Fergus, and I. O. Idowu, "Applied difference techniques of machine learning algorithm and web-based management system for sickle cell disease," in *2015 International Conference on Developments of E-Systems Engineering (DeSE)*. IEEE, 2015, pp. 231–235.

[67] M. Khalaf, A. J. Hussain, R. Keight, D. Al-Jumeily, P. Fergus, R. Keenan, and P. Tso, "Machine learning approaches to the application of disease modifying therapy for

sickle cell using classification models," *Neurocomputing*, vol. 228, pp. 154–164, 2017.

[68] C. Arbour, M. Choinière, J. Topolovec-Vranic, C. G. Loiselle, and C. Gélinas, "Can fluctuations in vital signs be used for pain assessment in critically ill patients with a traumatic brain injury?" *Pain research and treatment*, vol. 2014, 2014.

[69] P. E. Macintyre, S. Schug, D. Scott, E. J. Visser, S. M. Walker *et al.*, *Acute pain management: scientific evidence*. Australian and New Zealand College of Anaesthetists, 2010.

[70] F. Yang, T. Banerjee, K. Narine, and N. Shah, "Improving pain management in patients with sickle cell disease from physiological measures using machine learning techniques," *Smart Health*, vol. 7, pp. 48–59, 2018.

[71] A. Alambo, R. Andrew, S. Gollarahalli, J. Vaughn, T. Banerjee, K. Thirunarayan, D. Abrams, and N. Shah, "Measuring pain in sickle cell disease using clinical text," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 5838–5841.

[72] W. Downie, P. Leatham, V. Rhind, V. Wright, J. Branco, and J. Anderson, "Studies with pain rating scales." *Annals of the rheumatic diseases*, vol. 37, no. 4, pp. 378–381, 1978.

[73] R. A. Elzahaf, O. A. Tashani, B. A. Unsworth, and M. I. Johnson, "The prevalence of chronic pain with an analysis of countries with a human development index less than 0.9: a systematic review without meta-analysis," *Current medical research and opinion*, vol. 28, no. 7, pp. 1221–1229, 2012.

[74] A. I. Basbaum, D. M. Bautista, G. Scherrer, and D. Julius, "Cellular and molecular mechanisms of pain," *Cell*, vol. 139, no. 2, pp. 267–284, 2009.

[75] D. Julius and A. I. Basbaum, "Molecular mechanisms of nociception," *Nature*, vol. 413, no. 6852, pp. 203–210, 2001.

[76] C. A. von Hehn, R. Baron, and C. J. Woolf, "Deconstructing the neuropathic pain phenotype to reveal neural mechanisms," *Neuron*, vol. 73, no. 4, pp. 638–652, 2012.

[77] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[78] D. M. Blei and P. Smyth, "Science and data science," *Proceedings of the National Academy of Sciences*, vol. 114, no. 33, pp. 8689–8692, 2017.

[79] A. Newell and H. A. Simon, "Computer science as empirical inquiry: Symbols and search," in *ACM Turing award lectures*, 2007, p. 1975.

[80] P. Smolensky, "On the proper treatment of connectionism," *Behavioral and brain sciences*, vol. 11, no. 1, pp. 1–23, 1988.

[81] Y.-J. Hu, T.-H. Ku, R.-H. Jan, K. Wang, Y.-C. Tseng, and S.-F. Yang, "Decision tree-based learning to predict patient controlled analgesia consumption and readjustment," *BMC medical informatics and decision making*, vol. 12, no. 1, pp. 1–15, 2012.

[82] R. Sipilä, A. Estlander, T. Tasmuth, M. Kataja, and E. Kalso, "Development of a screening instrument for risk factors of persistent pain after breast cancer surgery," *British journal of cancer*, vol. 107, no. 9, pp. 1459–1466, 2012.

[83] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[84] A. Braundmeier-Fleming, N. T. Russell, W. Yang, M. Y. Nas, R. E. Yaggie, M. Berry, L. Bachrach, S. C. Flury, D. S. Marko, C. B. Bushell *et al.*, "Stool-based biomarkers of interstitial cystitis/bladder pain syndrome," *Scientific reports*, vol. 6, no. 1, pp. 1–10, 2016.

[85] C. V. Cannistraci, T. Ravasi, F. M. Montevecchi, T. Ideker, and M. Alessio, "Nonlinear dimension reduction and clustering by minimum curvilinearity unfold neuropathic pain and tissue embryological classes," *Bioinformatics*, vol. 26, no. 18, pp. i531–i539, 2010.

[86] M. Gram, C. Graversen, A. E. Olesen, and A. Drewes, "Machine learning on encephalographic activity may predict opioid analgesia," *European Journal of Pain*, vol. 19, no. 10, pp. 1552–1561, 2015.

[87] B. Schölkopf, A. J. Smola, F. Bach *et al.*, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[88] S. Gruss, R. Treister, P. Werner, H. C. Traue, S. Crawcour, A. Andrade, and S. Walter, "Pain intensity recognition rates via biopotential feature patterns with support vector machines," *PloS one*, vol. 10, no. 10, p. e0140330, 2015.

[89] D. Kringel, A. Ultsch, M. Zimmermann, J. Jansen, W. Ilias, R. Freynhagen, N. Griessinger, A. Kopf, C. Stein, A. Doehring *et al.*, "Emergent biomarker derived from next-generation sequencing to identify pain patients requiring uncommonly high opioid doses," *The pharmacogenomics journal*, vol. 17, no. 5, pp. 419–426, 2017.

[90] P. Nickerson, P. Tighe, B. Shickel, and P. Rashidi, "Deep neural network architectures for forecasting analgesic response," in *2016 38th annual international conference of the ieee engineering in medicine and biology society (EMBC)*. IEEE, 2016, pp. 2966–2969.

[91] J. Lötsch, G. Geisslinger, S. Heinemann, F. Lerch, B. Oertel, and A. Ultsch, "Qst response patterns to capsaicin-and uv-b-induced local skin hypersensitization in healthy subjects: a machine-learned analysis," *Pain*, vol. 159, pp. 11–24, 2018.

[92] A. Ultsch and J. Lötsch, "Machine-learned cluster identification in high-dimensional data," *Journal of biomedical informatics*, vol. 66, pp. 95–104, 2017.

[93] E. J. Chesler, S. G. Wilson, W. R. Lariviere, S. L. Rodriguez-Zas, and J. S. Mogil, "Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive," *Neuroscience & Biobehavioral Reviews*, vol. 26, no. 8, pp. 907–923, 2002.

[94] T. Winograd, "Understanding natural language," *Cognitive psychology*, vol. 3, no. 1, pp. 1–191, 1972.

[95] O. V. Patterson, M. Jones, Y. Yao, B. Viernes, P. R. Alba, T. J. Iwashyna, and S. L. DuVall, "Extraction of vital signs from clinical notes." *Studies in health technology and informatics*, vol. 216, pp. 1035–1035, 2015.

[96] D. Ravì, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Z. Yang, "Deep learning for health informatics," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 4–21, 2016.

[97] H.-M. Chen, Y.-T. Tsao, and S.-N. Tsai, "Automatic image segmentation and classification based on direction texton technique for hemolytic anemia in thin blood smears," *Machine vision and applications*, vol. 25, no. 2, pp. 501–510, 2014.

[98] V. Acharya and P. Kumar, "Identification and red blood cell classification using computer aided system to diagnose blood disorders," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2017, pp. 2098–2104.

[99] H. A. Elsalamony, "Anaemia cells detection based on shape signature using neural networks," *Measurement*, vol. 104, pp. 50–59, 2017.

[100] B. Albayrak, M. B. Darici, F. Kiraci, A. S. Öğrenci, A. Özmen, and K. Ertez, "Orak hücreli anemi tespiti sickle cell anemia detection," in *2018 Medical Technologies National Congress (TIPTEKNO)*. IEEE, 2018, pp. 1–4.

[101] T. S. Chy and M. A. Rahaman, "Automatic sickle cell anemia detection using image processing technique," in *2018 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*. IEEE, 2018, pp. 1–4.

[102] ——, "A comparative analysis by knn, svm & elm classification to detect sickle cell anemia," in *2019 International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST)*. IEEE, 2019, pp. 455–459.

[103] L. Alzubaidi, O. Al-Shamma, M. A. Fadhel, L. Farhan, and J. Zhang, "Classification of red blood cells in sickle cell anemia using deep convolutional neural network," in *International Conference on Intelligent Systems Design and Applications*. Springer, 2018, pp. 550–559.

[104] L. Alzubaidi, M. A. Fadhel, O. Al-Shamma, J. Zhang, and Y. Duan, "Deep learning models for classification of red blood cells in microscopy images to aid in sickle cell anemia diagnosis," *Electronics*, vol. 9, no. 3, p. 427, 2020.

[105] M. Gonzalez-Hidalgo, F. Guerrero-Pena, S. Herold-García, A. Jaume-i Capó, and P. D. Marrero-Fernández, "Red blood cell cluster separation from digital images for use in sickle cell disease," *IEEE journal of biomedical and health informatics*, vol. 19, no. 4, pp. 1514–1525, 2014.

[106] C. Chatfield, *Time-series forecasting*. Chapman and Hall/CRC, 2000.

[107] J. S. Armstrong, *Principles of forecasting: a handbook for researchers and practitioners*. Springer, 2001, vol. 30.

[108] S. A. Antoniou, G. A. Antoniou, F. A. Granderath, A. Mavroforou, A. D. Giannoukas, and A. I. Antoniou, "Reflections of the hippocratic oath in modern medicine," *World journal of surgery*, vol. 34, no. 12, pp. 3075–3079, 2010.

[109] A. Chang, E. M. Lad, and S. P. Lad, "Hippocrates' influence on the origins of neurosurgery," *Neurosurgical focus*, vol. 23, no. 1, pp. 1–3, 2007.

[110] G. Rosen, *A history of public health*. Jhu Press, 2015.

[111] J. Soni, U. Ansari, D. Sharma, and S. Soni, "Predictive data mining for medical diagnosis: An overview of heart disease prediction," *International Journal of Computer Applications*, vol. 17, no. 8, pp. 43–48, 2011.

[112] T. E. Getzen, "Forecasting health expenditures: short, medium and long (long) term," *Journal of Health Care Finance*, vol. 26, no. 3, pp. 56–72, 2000.

[113] B. R. Kirkwood and J. A. Sterne, *Essential medical statistics*. John Wiley & Sons, 2010.

[114] W. A. Knaus, D. P. Wagner, E. A. Draper, J. E. Zimmerman, M. Bergner, P. G. Bastos, C. A. Sirio, D. J. Murphy, T. Lotring, A. Damiano *et al.*, "The apache iii prognostic system: risk prediction of hospital mortality for critically iii hospitalized adults," *Chest*, vol. 100, no. 6, pp. 1619–1636, 1991.

[115] P. Marno, M. Chalder, T. Laing-Morton, M. Levy, P. Sachon, and D. Halpin, "Can a health forecasting service offer copd patients a novel way to manage their condition?" *Journal of health services research & policy*, vol. 15, no. 3, pp. 150–155, 2010.

[116] D. P. Rogers, M. A. Shapiro, G. Brunet, J.-C. Cohen, S. J. Connor, A. A. Diallo, W. Elliott, K. Haidong, S. Hales, D. Hemming *et al.*, "Health and climate–opportunities," *Procedia Environmental Sciences*, vol. 1, pp. 37–54, 2010.

[117] K. Bruynseels, F. Santoni de Sio, and J. Van den Hoven, "Digital twins in health care: ethical implications of an emerging engineering paradigm," *Frontiers in genetics*, p. 31, 2018.

[118] G. Rünstler, K. Barhoumi, S. Benk, R. Cristadoro, A. Den Reijer, A. Jakaitiene, P. Jelonek, A. Rua, K. Ruth, and C. Van Nieuwenhuyze, "Short-term forecasting of gdp using large datasets: a pseudo real-time forecast evaluation exercise," *Journal of forecasting*, vol. 28, no. 7, pp. 595–611, 2009.

[119] H. Sangrody, N. Zhou, S. Tutun, B. Khorramdel, M. Motalleb, and M. Sarailoo, "Long term forecasting using machine learning methods," in *2018 IEEE Power and Energy Conference at Illinois (PECI)*. IEEE, 2018, pp. 1–5.

[120] J. S. Armstrong, "Long-range forecasting, 2nd," *Available at SSRN 666990*, 2010.

[121] M. A. Cohen and J. A. Taylor, "Short-term cardiovascular oscillations in man: measuring and modelling the physiologies," *The Journal of physiology*, vol. 542, no. 3, pp. 669–683, 2002.

[122] G. Abraham, G. B. Byrnes, and C. A. Bain, "Short-term forecasting of emergency inpatient flow," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 3, pp. 380–388, 2009.

[123] M. Ghassemi, M. Pimentel, T. Naumann, T. Brennan, D. Clifton, P. Szolovits, and M. Feng, "A multivariate timeseries modeling approach to severity of illness assessment and forecasting in icu with sparse, heterogeneous clinical data," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 29, no. 1, 2015.

[124] H. ElMoaqet, D. M. Tilbury, and S. K. Ramachandran, "Multi-step ahead predictions for critical levels in physiological time series," *IEEE transactions on cybernetics*, vol. 46, no. 7, pp. 1704–1714, 2016.

[125] J. M. Lee and M. Hauskrecht, "Recent context-aware lstm for clinical event time-series prediction," in *Conference on Artificial Intelligence in Medicine in Europe*. Springer, 2019, pp. 13–23.

[126] I. Fox, L. Ang, M. Jaiswal, R. Pop-Busui, and J. Wiens, "Deep multi-output forecasting: Learning to accurately predict blood glucose trajectories," in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 1387–1395.

[127] B. Lim, "Forecasting treatment responses over time using recurrent marginal structural networks," *advances in neural information processing systems*, vol. 31, 2018.

[128] P. J. Tighe, C. A. Harle, R. W. Hurley, H. Aytug, A. P. Boezaart, and R. B. Fillingim, "Teaching a machine to feel postoperative pain: combining high-dimensional clinical data with machine learning algorithms to forecast acute postoperative pain," *Pain Medicine*, vol. 16, no. 7, pp. 1386–1401, 2015.

[129] F. Yang, "Improving pain management in patients with sickle cell disease using machine learning techniques," Ph.D. dissertation, Wright State University, 2020.

[130] R. Laubenbacher, J. P. Sluka, and J. A. Glazier, "Using digital twins in viral infection," *Science*, vol. 371, no. 6534, pp. 1105–1106, 2021.

[131] M. J. Azur, E. A. Stuart, C. Frangakis, and P. J. Leaf, "Multiple imputation by chained equations: what is it and how does it work?" *International journal of methods in psychiatric research*, vol. 20, no. 1, pp. 40–49, 2011.

[132] A. D. Shah, C. Langenberg, E. Rapsomaniki, S. Denaxas, M. Pujades-Rodriguez, C. P. Gale, J. Deanfield, L. Smeeth, A. Timmis, and H. Hemingway, "Type 2 diabetes and incidence of cardiovascular diseases: a cohort study in 1· 9 million people," *The lancet Diabetes & endocrinology*, vol. 3, no. 2, pp. 105–113, 2015.

[133] J. Payne, I. Aban, L. M. Hilliard, J. Madison, C. Bemrich-Stolz, T. H. Howard, A. Brandow, E. Waite, and J. D. Lebensburger, "Impact of early analgesia on hospitalization outcomes for sickle cell pain crisis," *Pediatric blood & cancer*, vol. 65, no. 12, p. e27420, 2018.

[134] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[135] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1. IEEE, 1995, pp. 278–282.

[136] R. Tibshirani, "Regression shrinkage and selection via the lasso," *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, vol. 58, pp. 267–288, 1994.

[137] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[138] F. P. Bravo, A. A. D. B. García, M. M. Gallego, A. B. G. Veiga, M. Ruiz, A. G. Peral, and J. L. Ayala, "Prediction of patient's response to onabotulinumtoxina treatment for migraine," *Heliyon*, vol. 5, no. 2, p. e01043, 2019.

[139] A. Athreya, R. Iyer, D. Neavin, L. Wang, R. Weinshilboum, R. Kaddurah-Daouk, J. Rush, M. Frye, and W. Bobo, "Augmentation of physician assessments with multi-omics enhances predictability of drug response: a case study of major depressive disorder," *IEEE computational intelligence magazine*, vol. 13, no. 3, pp. 20–31, 2018.

[140] B. V. Lakkakula, R. Sahoo, H. Verma, and S. Lakkakula, "Pain management issues as part of the comprehensive care of patients with sickle cell disease," *Pain Management Nursing*, vol. 19, no. 6, pp. 558–572, 2018.

[141] S. Padhee, G. K. Nave Jr, T. Banerjee, D. M. Abrams, N. Shah *et al.*, "Improving pain assessment using vital signs and pain medication for patients with sickle cell disease: Retrospective study," *JMIR Formative Research*, vol. 6, no. 6, p. e36998, 2022.

[142] S. Padhee, A. Alambo, T. Banerjee, A. Subramaniam, D. M. Abrams, G. K. Nave Jr, and N. Shah, "Pain intensity assessment in sickle cell disease patients using vital signs during hospital visits," *arXiv preprint arXiv:2012.01126*, 2020.

[143] J. Yoon, C. Davtyan, and M. van der Schaar, "Discovery and clinical decision support for personalized healthcare," *IEEE journal of biomedical and health informatics*, vol. 21, no. 4, pp. 1133–1145, 2016.

[144] L. Samal, A. Wright, B. T. Wong, J. A. Linder, and D. W. Bates, "Leveraging electronic health records to support chronic disease management: the need for temporal data views," *Informatics in primary care*, vol. 19, no. 2, p. 65, 2011.

[145] K. J. Ramos, B. S. Quon, S. L. Heltshe, N. Mayer-Hamblett, E. D. Lease, M. L. Aitken, N. S. Weiss, and C. H. Goss, "Heterogeneity in survival in adult patients with cystic fibrosis with fev1< 30% of predicted in the united states," *Chest*, vol. 151, no. 6, pp. 1320–1328, 2017.

[146] C. Lee, J. Yoon, and M. Van Der Schaar, "Dynamic-deephit: A deep learning approach for dynamic survival analysis with competing risks based on longitudinal data," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 1, pp. 122–133, 2019.

[147] X. Zhang, J. Chou, J. Liang, C. Xiao, Y. Zhao, H. Sarva, C. Henchcliffe, and F. Wang, "Data-driven subtyping of parkinson's disease using longitudinal clinical records: a cohort study," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[148] A. Rusanov, P. V. Prado, and C. Weng, "Unsupervised time-series clustering over lab data for automatic identification of uncontrolled diabetes," in *2016 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 2016, pp. 72–80.

[149] A. Giannoula, A. Gutierrez-Sacristán, Á. Bravo, F. Sanz, and L. I. Furlong, "Identifying temporal patterns in patient disease trajectories using dynamic time warping: a population-based study," *Scientific reports*, vol. 8, no. 1, pp. 1–14, 2018.

[150] D. T. A. Luong and V. Chandola, "A k-means approach to clustering disease progressions," in *2017 IEEE International conference on healthcare informatics (ICHI)*. IEEE, 2017, pp. 268–274.

[151] A. Boudier, S. Chanoine, S. Accordini, J. M. Anto, X. Basagaña, J. Bousquet, P. Demoly, J. Garcia-Aymerich, F. Gormand, J. Heinrich *et al.*, "Data-driven adult asthma phenotypes based on clinical characteristics are associated with asthma outcomes twenty years later," *Allergy*, vol. 74, no. 5, pp. 953–963, 2019.

[152] W. M. Wami, F. Buntinx, S. Bartholomeeusen, G. Goderis, C. Mathieu, and M. Aerts, "Influence of chronic comorbidity and medication on the efficacy of treatment in patients with diabetes in general practice," *British Journal of General Practice*, vol. 63, no. 609, pp. e267–e273, 2013.

[153] H. Blockeel, S. Dzeroski, J. Struyf, and B. Zenko, "Predictive clustering," 2019.

[154] A. S. Ash, R. P. Ellis, G. C. Pope, J. Z. Ayanian, D. W. Bates, H. Burstin, L. I. Iezzoni, E. MacKay, and W. Yu, "Using diagnoses to describe populations and predict costs," *Health care financing review*, vol. 21, no. 3, p. 7, 2000.

[155] Y. Zhao, A. S. Ash, R. P. Ellis, J. Z. Ayanian, G. C. Pope, B. Bowen, and L. Weyuker, "Predicting pharmacy costs and other medical costs using diagnoses and drug claims," *Medical care*, pp. 34–43, 2005.

[156] Y. Zhao, R. P. Ellis, A. S. Ash, D. Calabrese, J. Z. Ayanian, J. P. Slaughter, L. Weyuker, and B. Bowen, "Measuring population health risks using inpatient diagnoses and outpatient pharmacy data." *Health services research*, vol. 36, no. 6 Pt 2, p. 180, 2001.

[157] J. F. Farley, C. R. Harley, and J. W. Devine, "A comparison of comorbidity measurements to predict healthcare expenditures," *American Journal of Managed Care*, vol. 12, no. 2, pp. 110–118, 2006.

[158] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PloS one*, vol. 13, no. 3, p. e0194889, 2018.

[159] J. C. B. Gamboa, "Deep learning for time-series analysis," *arXiv preprint arXiv:1701.01887*, 2017.

[160] B. K. Nelson, "Time series analysis using autoregressive integrated moving average (arima) models," *Academic emergency medicine*, vol. 5, no. 7, pp. 739–744, 1998.

[161] G. P. Zhang, "Time series forecasting using a hybrid arima and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[162] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[163] T. Iliou, C.-N. Anagnostopoulos, M. Nerantzaki, and G. Anastassopoulos, "A novel machine learning data preprocessing method for enhancing classification algorithms performance," in *Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS)*, 2015, pp. 1–5.

[164] S. Sarwar, Z. Ul-Qayyum, and A. Kaleem, "Machine learning based intelligent framework for data preprocessing." *Int. Arab J. Inf. Technol.*, vol. 15, no. 6, pp. 1010–1015, 2018.

[165] P. Newbold, "Arima model building and the time series analysis approach to forecasting," *Journal of forecasting*, vol. 2, no. 1, pp. 23–35, 1983.

[166] D. Whitley, "A genetic algorithm tutorial," *Statistics and computing*, vol. 4, no. 2, pp. 65–85, 1994.

[167] F. Rosenblatt, "Principles of neurodynamics. perceptrons and the theory of brain mechanisms," Cornell Aeronautical Lab Inc Buffalo NY, Tech. Rep., 1961.

[168] H. Chung, S. J. Lee, and J. G. Park, "Deep neural network using trainable activation functions," in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 348–352.

[169] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks. 2012 advances in neural information processing systems (nips)," *Neural Information Processing Systems Foundation, La Jolla, CA*, 2012.

[170] S. Hochreiter, "Ja1 4 rgen schmidhuber (1997)."long short-term memory"," *Neural Computation*, vol. 9, no. 8.

[171] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *Scientific reports*, vol. 8, no. 1, pp. 1–12, 2018.

[172] W. Cao, D. Wang, J. Li, H. Zhou, L. Li, and Y. Li, "Brits: Bidirectional recurrent imputation for time series," *Advances in neural information processing systems*, vol. 31, 2018.

[173] Y. Rubanova, R. T. Chen, and D. K. Duvenaud, "Latent ordinary differential equations for irregularly-sampled time series," *Advances in neural information processing systems*, vol. 32, 2019.

[174] M. Horn, M. Moor, C. Bock, B. Rieck, and K. Borgwardt, "Set functions for time series," in *International Conference on Machine Learning*. PMLR, 2020, pp. 4353–4363.

[175] A. Sagheer and M. Kotb, "Unsupervised pre-training of a deep lstm-based stacked autoencoder for multivariate time series forecasting problems," *Scientific reports*, vol. 9, no. 1, pp. 1–16, 2019.

[176] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[177] M. Gutmann and A. Hyvärinen, "Noise-contrastive estimation: A new estimation principle for unnormalized statistical models," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2010, pp. 297–304.

[178] K. Zhang, S. Lin, H. Sun, L. Ma, and J. Xu, "Dynamic time warping based clustering for time series analysis," in *IoT and Big Data Technologies for Health Care*. Springer, 2022, pp. 376–385.

[179] D. A. Dickey and W. A. Fuller, "Likelihood ratio statistics for autoregressive time series with a unit root," *Econometrica: journal of the Econometric Society*, pp. 1057–1072, 1981.

[180] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[181] P. Yang, Y. Hwa Yang, B. B Zhou, and A. Y Zomaya, "A review of ensemble methods in bioinformatics," *Current Bioinformatics*, vol. 5, no. 4, pp. 296–308, 2010.