University of Rhode Island
DigitalCommons@URI

Open Access Master's Theses

1-1-2022

PREDICTING PAIN PRESENCE IN ICU PATIENTS USING PHYSIOLOGICAL SIGNALS

Derek Jacobs
University of Rhode Island, djacobs@uri.edu

Follow this and additional works at: https://digitalcommons.uri.edu/theses

Recommended Citation

Jacobs, Derek, "PREDICTING PAIN PRESENCE IN ICU PATIENTS USING PHYSIOLOGICAL SIGNALS" (2022). *Open Access Master's Theses.* Paper 2152.
https://digitalcommons.uri.edu/theses/2152

This Thesis is brought to you for free and open access by DigitalCommons@URI. It has been accepted for inclusion in Open Access Master's Theses by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons@etal.uri.edu.

PREDICTING PAIN PRESENCE IN ICU PATIENTS USING PHYSIOLOGICAL SIGNALS

BY

DEREK JACOBS

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

IN

COMPUTER SCIENCE

UNIVERSITY OF RHODE ISLAND

2022

MASTER OF SCIENCE THESIS

OF

DEREK JACOBS

APPROVED:

Thesis Committee:

Major Professor Krishna Kumar Venkatasubramanian

Marco Alvarez

Ryan Chapman

Brenton DeBouf
DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2022

ABSTRACT

Physiological data can be used to detect the presence of pain, a problem that up to this point has entirely subjective solutions. While there are general indicators of pain, physiological signals have been shown to alter as a response to painful stimuli. Prior work has primarily focused on predicting a level of pain reported by a patient based on the assumption that pain is present. In this work, we present a means of using machine learning to identify the presence of pain using data collected from a freely available database MIMIC-III. Our methodology involves constructing an image reconstruction based classifier and evaluating our optimal classifiers on totally unseen testing data. Using both a 2 physiological stream and a 3 physiological stream approach, our models produced an accuracy of 80.56% and 87.18%, respectively. Each model is able to detect pain given less than a minute of data, although the 2 stream approach requires less data to work with. The proposed method for identifying pain presence has not been attempted before to our knowledge.

ACKNOWLEDGMENTS

I would like to begin by thanking Dr. Ryan Chapman, Dr. Marco Alvarez, and Dr. Krishna Kumar Venkatasubramanian for being a part of my thesis committee. In addition, I would like to thank Dr. Edmund Lamagna for serving as defense chair as well. I would like to thank those others who supported me through the years, namely my close friends and Karen especially. I also want to thank Dr. Marc Rigatti, Dr. Stephanie Carreiro, and Brittany Chapman from the University of Massachusetts Medical school for their advising throughout the process. Lastly I would like to thank all of my track and field coaches throughout the years: Leanne, Travis, Pat, Nick, Rock, Fed, Ben, Copeland, Whitten, Trent, and Doyle. You've all encouraged me to always strive to be the best man that I can and I never could have gotten this far without your guidance along the way.

TABLE OF CONTENTS

Page

LIST OF FIGURES

CHAPTER 1

Introduction

1.1 Introduction and Motivation

In a hospital setting, reporting and treating pain are two of the most common occurrences. Proper management and treatment of pain is imperative, as both chronic and acute pain have been shown to negatively impact people's lives [1]. Unfortunately for caregivers and medical workers, because the experience of pain is unique to every individual [2] the ability to accurately identify the existence and level of pain a patient is experiencing remains a challenge. Pain assessment relies upon the subjective reporting from the patients themselves [3], a methodology that remains in practice today. When a patient is experiencing pain, they typically communicate their pain verbally to a medical worker and rate their pain on a scale of 1 to 10. Most commonly doctor's offices display a poster of the Wong-Baker FACES Pain Rating Scale which provides simple, visual references of happy and sad faces to help patients rate their pain levels [4]. While for the most part doctors can trust patients to accurately report their own pain, there are multiple cases where a patient may be unable or reluctant to communicate their pain accurately, such as unconsciousness, a lack of trust in health care providers, fear of increased time in care, financial concerns, and many others [5]. In other words, pain assessment and treatment can be a challenge for both healthcare providers and patients alike.

The problem of objective pain assessment in the realm of modern medicine has existed for several decades now; the possibility alone of being able to look into the physiology of a patient and determine if that patient is experiencing pain or not could yield many benefits. First and foremost, it removes the requirement that patients effectively communicate their pain to their doctors. Regardless of ability or lack thereof, pain should nearly always be mediated when appropriately warranted

as even the unconscious brain registers pain in the body [6]. By eliminating the communication barriers, we can help health workers do their jobs more effectively, and assist patients by treating their pain properly and improve their quality of life [7]. Without proper treatment or some mediation, chronic and acute pain have been shown to negatively impact people's lives. From deterioration of mental health to insomnia to employment issues and troubles in personal relationships [8], pain can greatly affect people when left untreated.

It should be noted that there exists some general indicators that could act as non verbal communication that a patient is in pain. These include but are not limited to facial expressions such as clenched teeth or rapid blinking, body movements like rubbing a body area or guarding, changes in mood or mental state such as depressive symptoms or agitation, and multiple other indicators [9]. Though these indicators have been used in the past to assess and consequently treat pain in nonverbal adults [10], these indicators could be fabricated or hidden in some manner and make their identification harder, and thus warrants the need for fully objective pain assessment methodologies.

A promising objective pain assessment method is looking towards biological processes. An aggregate of physiological signals provides a snapshot of what is happening in the human body at a given time. In general, when a person experiences pain, their central nervous system reacts changing various parts of their physiology including their heart rate, skin conductivity, blood pressure and others [11], with some exceptions due to disorders like congenital insensitivity to pain. Knowing this, it is possible that machine learning methods could be used to read in these physiological signals, detect differences between the values of pain-free physiological signals and painful ones.

In this work, *we present a method that uses machine learning to identify the*

presence or absence of pain using physiological data collected from hospital records. To develop and evaluate our approach for detecting pain presence using physiological data, we rely on physiological time-series data collected from patients in a hospital intensive care unit (ICU). Data was collected from 77 patients who were admitted to the ICU for various purposes. Our subjects were real patients and data had been collected with approval from both the University of Rhode Island Institutional Review Board (IRB), as well as the Massachusettes Institute of Technology IRB.

In order to detect pain presence or absence, we use simple machine learning techniques to develop a classifier that extracts information about the shape and relativity of physiological signals stored in the hospital record system during times of pain, and times where no pain is observed. The physiological measurements collected are electrocardiogram lead II, arterial blood pressure, and respiration rate. While patients were held in the ICU, they were assessed hourly for either presence or absence of pain. All instances of this assessment involved verbal communication between a health provider and the patient. Because patients were asked hourly as to whether they were experiencing pain or not, we used up to an hour long interval surrounding the time the patient was assessed for training and testing our models. Testing various windows allows us to evaluate how short of a time frame we must record physiological data in order to predict whether they are experiencing pain or not. The classifier developed were successful at distinguishing pain presence and absence.

The dataset as a whole had data from 77 patients, 49 of which were male, 28 were female. From these patients, we initially had 947 *pain events* to work with. A pain event is defined as an instance of the patient reporting experiencing or not experiencing pain. Of these initial pain events, 229 were *positive pain events*, or pain

events where the patient reported experiencing pain, and 718 were *negative pain events*, or pain events where the patient reported not experiencing pain. In other words, there were many more negative pain events than positive ones; this class imbalance had to be taken into account during the model development process.

While the results of this work improve upon previous related works, this still demonstrates solely the **viability** of our method. We were able to achieve a maximum classification accuracy of 87.18%. The testing data used to find this accuracy was completely unseen by our model during the training process. We test various time frames for detecting pain in hopes of identifying a model that can predict pain presence objectively with the highest accuracy and requiring minimal information to do so.

List of References

[1] N. Katz. Journal of Pain and Symptom Management. "The impact of pain management on quality of life." July 2002. [Online]. Available: https://www.jpsmjournal.com/article/S0885-3924(02)00411-6/fulltext

[2] R. C. Coghill. U.S. Natnioal Library of Medicine. "Individual differences in the subjective experience of pain: New insights into mechanisms and models." Oct. 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2959190/

[3] J. Katz. Surgical Clinics of North America. "Measurement of pain." Apr. 1999. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0039610905703819?via%3Dihub

[4] Wong-Baker FACES Foundation. [Online]. Available: https://wongbakerfaces.org/

[5] B. Boring, *How and Why Patient Concerns Influence Pain Reporting: A Qualitative Analysis of Personal Accounts and Perceptions of Others' Use of Numerical Pain Scales.*, Frontiers Std., Jan. 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.663890/full

[6] D. D. Price, *Unconscious and Conscious Mediation of Analgesia and Hyperalgesia.*, National Academy of Sciences of the United States of America Std., June 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4485153/

[7] N. Wells, *Improving the Quality of Care through Pain Assessment and Management.*, U.S. National Library of Medicine Std., Apr. 2008. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK2658/

[8] B. McCarberg, *The Impact of Pain on Quality of Life and the Unmet Needs of Pain Management: Results From Pain Sufferers and Physicians Participating in an Internet Survey.*, American Journal of Therapeutics Std., July 2008. [Online]. Available: https://journals.lww.com/americantherapeutics/Abstract/2008/07000/The_Impact_of_Pain_on_Quality_of_Life_and_the.4.aspx

[9] S. Booker, *Assessing Pain in Nonverbal Older Adults.*, U.S. National Library of Medicine Std., May 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4991889/

[10] K. Herr, *Tools for Assessment of Pain in Nonverbal Older Adults with Dementia: A State-of-the-Science Review.*, Journal of Pain and Symptom Management Std., Feb. 2006. [Online]. Available: https://www.jpsmjournal.com/article/S0885-3924(05)00611-1/fulltext

[11] Institute of Medicine (US) Committee on Pain, Disability, and Chronic Illness Behavior. "The anatomy and physiology of pain." Jan. 1987. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK219252/

CHAPTER 2

Literature Review

2.1 Related Work

While the area of pain in general has been studied well, not much work has been completed relating to objective identification of pain. Most research relating to pain involves the actual treatment of pain and how pain processes work in the body, including much work on the central nervous system. Numerous studies have been conducted evaluating current methodologies for assessing pain, and the effectiveness of various scales used to measure and quantify pain. In fact, as of 2017 well over 10,000 paper abstracts related directly to pain assessment or more specifically pain assessment in the eldery [1]. In the last few years there has been a slight shift from evaluating existing methodologies and scales towards attempting to objectively evaluate pain altogether; several attempts have been made with respect to incorporating machine learning into the area of pain research, and although multiple studies utilize machine learning for classifying levels of pain, few attempts have been made at diagnosing the existence or absence of pain [2].

Among these works relating to machine learning and pain, a few noteworthy studies stand out among the rest. The first work "Normalized skin conductance level could differentiate physical pain stimuli from other sympathetic stimuli" by Sugimine et al [3] exhibits the promise of looking at physiological data and its relation to pain. In this work, the authors begin by introducing their experimental setup, which involved inducing pain in participants in the form of thermal stimuli. While pain was induced, the skin conductance of each participant was required. Below in Figure 1 is a plot presented with sample readings from this study. In this plot, we see that when pain is induced, the skin conductance levels in the participant increase, and gradually fall off once the pain stimulus has been halted. With

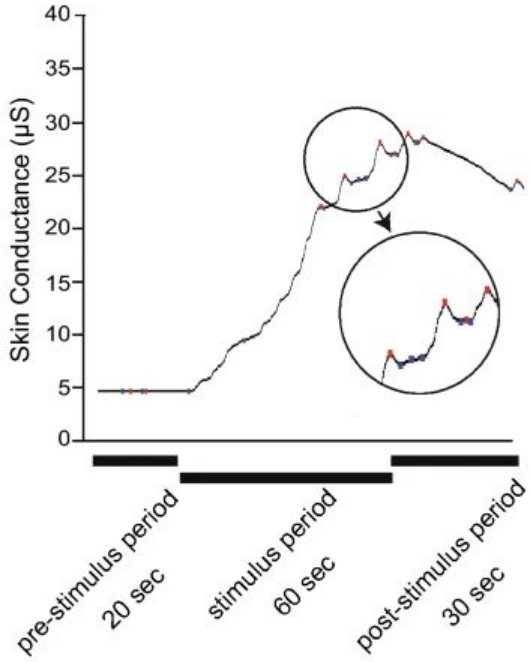

Figure 1. Skin conductance sample reading during pain induction

data like this, the authors reach the conclusion that signals like skin conductance could be used to evaluate pain, but do not attempt to do so themselves. Other works regarding skin conductance have shown promise in the area in relation to pain as well [4, 5]

Another work relating specifically to machine learning and pain prediction is entitled "Machine learning-based prediction of clinical pain using multimodal neuroimaging and autonomic metrics" by Lee et al [6]. This study involved the induction of pain in study participants, except instead of measuring skin conductance, patients underwent magnetic resonance imaging (MRI), and the results of these scans were used as input in the context of predicting pain intensity. Once all data was collected, the authors used a support vector regression algorithm. Support vector algorithms effectively try to find a hyperplane in an n-dimensional space that classifies data points distinctly; this is common knowledge in the machine learning community. Overall, the authors found that the support vector

machine approach made predictions that correlated well with actual pain levels reported by the participants.

The last, most notable work, "Objective Pain Assessment Using Vital Signs" by Erdoğan and Oğul [7] is one of the very few works attempting to predict the existence of pain. These authors begin by acquiring data from Medical Information Mart in Intensive Care (MIMIC)-III, which will be described in detail later on in this paper and has been used in other machine learneing related works before[8]. Erdoğan and Oğul collected measurements of the following physiological parameters for the whole of 8 hours prior to a pain observation: Glasgow Coma Scale, Heart Rate, Oxygen Saturation, Pupil Size, Respiration Rate, Skin Temperature, and Urine Color

Like all instances where physiological data is involved, their dataset had relatively large quantities of missing data. In order to handle this issue, when missing data is encountered, the latest valid measurement is carried until the next valid measurement is reached. Once the dataset was properly setup, the authors tested four machine learning models to make predictions on pain existence: AdaBoost, Multilayer Perceptron (MLP), LogitBoost, and Random Forest. These models are all very popular and widely used in machine learning literature. Within the context of this study, Erdoğan and Oğul tested their models using the aforementioned physiological signals recorded for 3,6, and 8 hours prior to the pain observation. A summary of their results can be seen in Figures 2-5. Basically, Random Forest consistently outperformed the other models, but all performed similarly for each test case. In all of these figures we observe that with respect to accuracy, each model performs drastically better when given data from 3 hours until the pain observation as opposed to when the models are given more data. AUROC scores are generally about the same across all time settings with the exception of the random

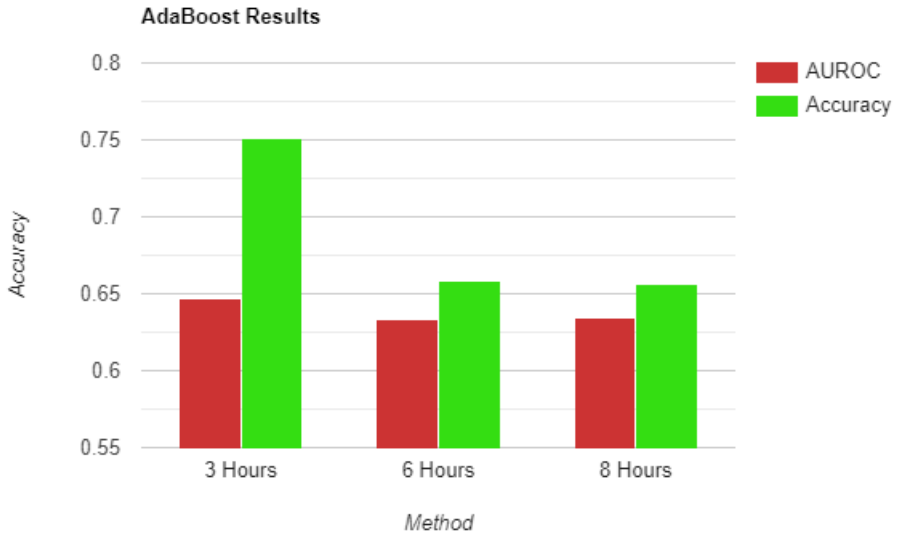

Figure 2. AdaBoost model scores from Erdoğan and Oğul [7]. 3 Hours provides the highest .647 AUROC and a .751 accuracy.

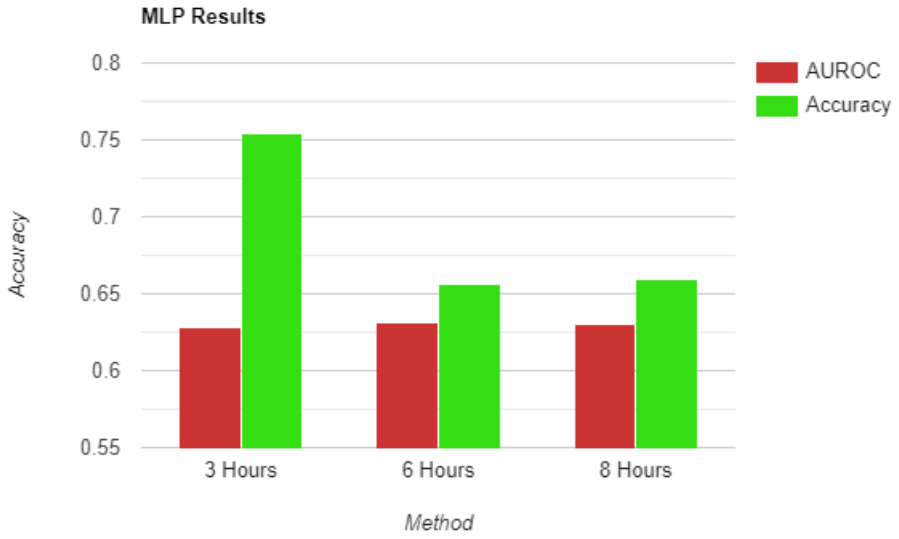

Figure 3. MLP model scores from Erdoğan and Oğul [7]. All provide similar AUROC values but 3 hours produces a .754 accuracy.

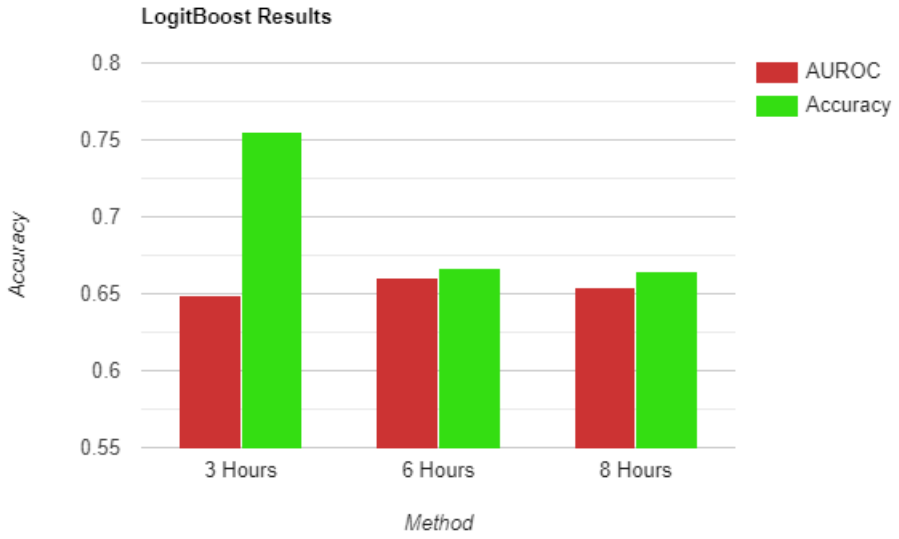

Figure 4. LogitBoost model scores from Erdoğan and Oğul [7]. 6 hours provides a highest .660 AUROC and 3 hours yields a .755 accuracy.

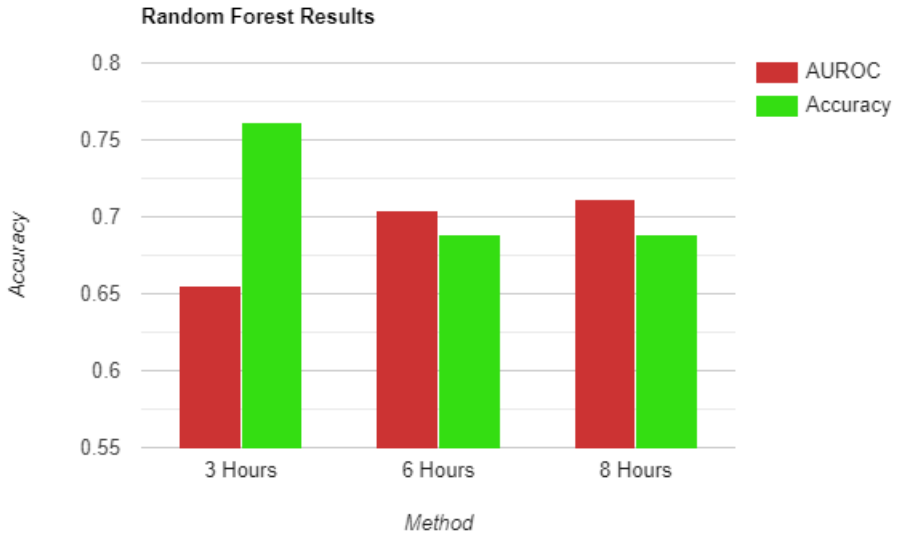

Figure 5. Random Forest model scores from Erdoğan and Oğul [7]. 8 hours produces the highest AUROC at .711 and 3 hours produces an accuracy of .761

forest, in which 3 hours prior to pain observation yields the lowest AUROC and 8 hours yields the highest AUROC of .711. As a whole the results of this study show promise in predicting pain presence or absence but with accuracies no better than 75.4%, these models are unlikely to be used in a real medical setting.

Though the problem of objective pain assessment has been lightly studied, a majority of works place emphasis on predicting the level of pain as opposed to predicting the existence of pain. Those works that do experiment with the pain existence problem exhibit moderate accuracy, with results that show that predicting the existence of pain is possible. Therefore, there is a need to deeply explore objective identification of pain in order to improve model performance in a way that could yield usage and viability in a medical setting.

2.2 Problem Statement

The purpose of this paper is to explore the use of machine learning in the context of identifying pain presence. The basic idea behind this is to construct a model that learns variations between the physiological signals of patients experiencing pain, and patients not experiencing pain. Once created, this model will be able to identify whether a snippet of physiological data represents a patient in pain or a patient not in pain.

List of References

[1] Y.-S. Kim, *Assessment of pain in the elderly: A literature review*, The National Medical Journal of India Std., Apr. 2017. [Online]. Available: https://nmji.in/assessment-of-pain-in-the-elderly-a-literature-review/

[2] M. Matsangidou, *Machine Learning in Pain Medicine: An Up-To-Date Systematic Review*, Springer Std., 2021. [Online]. Available: https://link.springer.com/article/10.1007/s40122-021-00324-2

[3] S. Sugimine, *Normalized Skin Conductance Level Could Differentiate Physical Pain Stimuli from Other Sympathetic Stimuli.*, Nature News Std., July 2020. [Online]. Available: https://www.nature.com/articles/s41598-020-67936-0#::

text=Skin%20conductance%20C%20especially%20nSCL%2C%20was,
including%20tactile%20and%20mental%20stimuli

[4] B. Susam, *Automated Pain Assessment Using Electrodermal Activity Data and Machine Learning.*, U.S. National Library of Medicine Std., July 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30440413/

[5] H. Storm, *Changes in Skin Conductance as a Tool to Monitor Nociceptive Stimulation and Pain.*, U.S. National Library of Medicine Std., Dec. 2008. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18997532/

[6] J. Lee, *Machine Learning-Based Prediction of Clinical Pain Using Multimodal Neuroimaging and Autonomic Metrics.*, U.S. National Library of Medicine Std., Mar. 2019. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30540621/

[7] B. Erdoğan, *Objective Pain Assessment Using Vital Signs.*, Procedia Computer Science Std., Apr. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S187705092030541X

[8] D. Lopez-Martinez, *Deep Reinforcement Learning for Optimal Critical Care Pain Management with Morphine Using Dueling Double-Deep Q Networks.*, Cornell University Std., Apr. 2019. [Online]. Available: https://arxiv.org/abs/1904.11115

CHAPTER 3

Methodology

In this chapter we explore the process of data collection, cleaning our dataset and methods of handling missing data. From there we describe the process behind image reconstruction, and delve into more specifics related to the training and testing procedures, including how the models will be evaluated later on.

3.1 Collecting Data

With inspiration from Erdoğan and Oğul, we also look to the most recent version of the MIMIC-III dataset, version 1.4. The MIMIC database is a freely available database containing health-related data of more than 40,000 patients who were admitted to the intensive care unit (ICU) of the Beth Israel Deaconess Medical Center in Boston, Massachusettes between 2001 and 2012. The database contains information including vital sign measurements, demographics, procedures, medications, any caregiver notes, and many other elements useful in a medical context [1]. All physiological data were recorded at 125 Hz and was collected through either the Philips CareVue Clinical Information System and iMDsoft MetaVision ICU. It should be noted that data collection for MIMIC-III was approved by the institutional review boards of the Massachusetts Institute of Technology for primary collection, and the University of Rhode Island as part of a secondary data analysis as mentioned previously. All data within the database is deidentified using date shifting and other anonymizing techniques.

3.2 Stream Selection

We collected the measurements of 3 different physiological parameters for up to an hour surrounding pain events, giving us half an hour of data before and

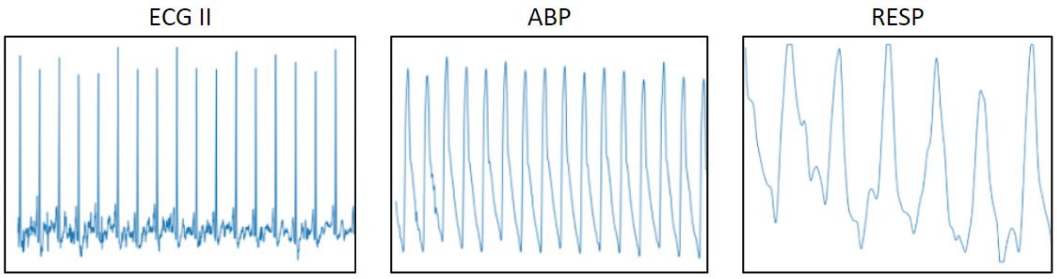

Figure 6. Sample II, ABP, RESP streams

half an hour after a pain event for the model to extract information from. These parameters are as follows: Electrocardiogram Lead II (II), Arterial Blood Pressure (ABP), and Respiration Rate (RESP). Several factors led to the decision to include these three signals specifically. Based on works mentioned previously among others, these three signals have been correlated well with the existence of pain [2, 3, 4]. Another major factor that led to our selection of these signals is data availability. Even though we handle missing data in our working dataset and there are multiple methods to do so depending on how the missing data distribution is defined, be it missing completely at random or missing not at random [5]. Options for handling missing data include but are not limited to mean substitution [6], last observation carried forward [7], and multiple imputation [8]. These options are taken into account later on when handling our own missing data problem. A sample of the 3 data streams collected can be seen in Figure 6.

3.3 Dataset Pruning - Handling Missing Data and Class Imbalance via SMOTE and Downsampling

We begin with 947 completely independent pain events, 229 of which are positive pain events and 718 are negative pain events. All of these pain events come from 49 male patients and 28 female patients for a total of 77 patients. Inherently this indicates that patients each have multiple pain events. Nevertheless, pain events are all separated by ample time to allow our hour long time window to

Figure 7. Comparing imputation techniques

remain completely independent of one another. There are a couple important problems that must be dealt with prior to using this dataset to predict pain existence: missing data and class imbalance.

With respect to missing data, we have seen multiple approaches to handling the missing data problem [2, 6, 8]. Some methods, however, are more applicable to time-series or health data than others. Erdoğan and Oğul demonstrated last observation carry over could be a viable candidate for this dataset specifically. Even still, with consultation of Marc Rigatti of the University of Massachusetts Medical School, we have come to the conclusion that simple carry over will not suffice for the purposes of this experiment. As a result we decide to use linear interpolation as it has been used extensively in the context of physiological stream correction in the past [9, 10]. Linear interpolation is an improvement upon last observation carry over. Instead of repeating the same value at 125 Hz until jumping to the next value, we interpolate data forming a straight line between the latest valid observation and the next one following a segment of missing data. A visual representation of linear interpolation compared with latest observation carry over can be seen in Figure 7. Looking back towards Figure 6, we can observe that sharp vertical jumps in any of the streams of data is rare whereas relatively linear changes are more representative of the signals. With few exceptions, a majority

of our data was missing less than 10% of data. Some pain events were missing a majority of their respective data, something that we fixed through our class imbalance solution.

Given our initial dataset, there is a significant class imbalance, with more than 75% of our data representing negative pain events, and the remaining few samples representing positive pain events. In general, passing imbalanced classes into a machine learning model proves problematic as most models typically assume relatively equal class distributions [11]. As a result, we aim to bring our dataset at least into more balance. However, because positive pain events are generally less common occurring than negative ones, we should maintain a slight imbalance to reflect the reality of our data. Therefore, we will be aiming for a distribution of approximately 55% negative pain events, and 45% positive pain events. There are multiple methods of handling imbalanced classes, as this problem occurs frequently in the world of machine learning. Among these methods are downsampling the majority class, oversampling the minority class, and using different metrics to account for the imbalance [12, 13]. Depending on the context of the problem, it is also common to combine a couple or more of these methods into an ensemble. In this work, we utilize a combination of undersampling the majority whilst also oversampling the minority in the hopes of reaching a class distribution of 300 positive events and 400 negative events.

We begin with downsampling our majority class as this process is much simpler than the oversampling procedure. In order to remove samples from our majority class, we start by ordering the negative pain events in nonincreasing order of missing data percentages. For instance, if there are 5 negative pain events, 5 of which (V, W, X, Y, Z) missing 5% of data, another 2 (A, B) missing 2% of data, the dataset would initially be ordered [V,W,X,Y,Z,A,B]. From there, we randomly

scramble the ordering of pain events with the same amount of missing data. For instance, the previous ordering could be reordered to [X,Z,W,V,Y,B,A]. The non-increasing attribute of missing data quantity is still maintained. Now we begin eliminating events from the beginning of our list until we have the target number of negative pain events remaining. In our case, we begin with 718 negative events, and remove a total of 318, leaving 400 to work with. Because we have altered the missing data distribution of our dataset, we must make sure that both classes are treated equally. After removing extra events from our negative class, the highest amount of missing data in the negative class is approximately 8%, and so in order to maintain fair comparisons we also remove positive pain events that are missing more than 8% of their data. Doing so leaves us with just 192 positive pain events to work with, which brings us to oversampling our minority to fix the overall class distribution.

A highly used method for oversampling minority classes is synthetic minority oversampling technique (SMOTE) [14]. This methodology generates new instances of the minority class by performing certain operations on real data, operating in a feature space. Although the original SMOTE technique was developed for singular point data as opposed to time series data like we are working with, there exists adaptations of SMOTE specifically developed for time series data [15]. When looking to generate new data we pass in 3 arguments, D, the data set, ng, the number of synthetic cases to generate, and k, the number of neighbors used in case generation. We start by pulling a random sample of ng instances of our positive events. For each of these instances, we find the k nearest neighbors of that event. Let's look more closely at one stream of a random sample, like ABP. In order to find the k nearest neighbors of this ABP stream called *tempStream* for this example, we find the k ABP streams from all positive pain events whose euclidean distance

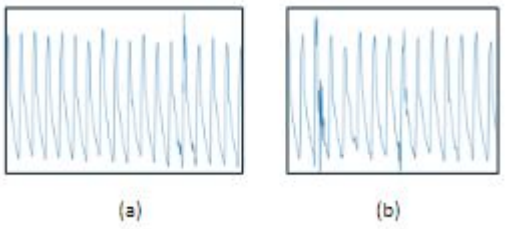

Figure 8. Real ABP data stream (a) shown with a SMOTE generated ABP stream (b)

to *tempStream* are minimal. With a pool of k nearest neighbors to this stream, we pick a single random neighbor denoted *rNeighbor* and for all data points in our initial stream *tempStream*, we conduct the following procedure. We calculate the difference between a point at index i in *tempStream* and *rNeighbor* $diff = tempStream[i] - rNeighbor[i]$. Then we create a new data point at index i in our syntheticStream $syntehticStream[i] = tempStream[i] + Random(0,1) * diff$. By applying a random scale to our diff, we ensure variation exists between real data and synthetic data. Once this process is completed for a single stream, we repeat the process for the other two streams until we have a full synthetic positive pain event. In total, we generate 108 new positive pain events, giving us a total of 300 positive pain events to work with. A comparison of a real data stream and a SMOTE generated stream can be seen in Figure 8. We can observe that there are very few noticeable differences between the two streams, giving us confidence that SMOTE successfully recreates physiological streams. Ultimately, this brings our class distribution to 300 positive pain events and 400 negative pain events, or approximately 42.8% positive cases and 57.2% negative cases, with 700 pain events in total.

3.4 Model Selection and Cross Validation

With a working dataset in our grasp, we can now look to begin develop a model for our data. While other works mentioned previously have used common approaches like random forest and adaboost, we propose an application of other methodology to the problem of pain classification: image reconstruction. This process has been utilized in several applications such as object detection [16] and more importantly for our context, detecting data manipulation attacks on physiological sensor measurements [17]. Image reconstruction is a viable candidate for our pain classification problem for several reasons. Firstly, other classic machine learning methods have been tested against this dataset. Though the dataset of [2] is not exactly the same as ours, the area of objective pain classification must be expanded in order to find an optimal model for the problem, and this expansion includes testing other models and methods like image reconstruction. Another reason to use image reconstruction is its simplicity and efficiency. All of the training and classification aspects of this model are simple mathematical equations that are easy to follow and easy to implement in practice. Additionally, this model is fast relative to other options. One popular method related to image reconstruction is the use of a convolutional neural network (CNN). In our case we opt not to use a CNN because they tend to train slowly compared to image reconstruction, and CNNs typically perform best with much larger datasets. 700 samples are workable but CNNs are very well suited to datasets with thousands of data samples or more.

Having selected the image reconstruction route, we can now develop, train, and test our model. For this process we will maintain a 90%/10% stratified split of the data for training and testing. In other words, each of our training and testing split will maintain the overall class distribution established through our class imbalance solution above, with the training set containing 90% of our data

and the testing set containing 10% of the data. This will give 270 positive pain events and 360 negative pain events for training, and leave 30 positive pain events and 40 negative pain events for testing. During the training process we make use of leave one out cross validation (LOOCV). Cross validation in general is a method used to evaluate the performance of a machine learning model. K-fold cross validation involves splitting the training dataset into k groups. Once the k groups are established, one group will be held for a validation set, and the remaining groups used as a training set. The model is fit on the training set and evaluated on the validation set. The evaluation results are maintained and the model discarded, and the process repeated until each group has been used as the validation set. Once all of this training and validation has been completed, the evaluation scores can be combined to estimate the performance of the model [18]. LOOCV is a variation of k-fold cross validation in which the validation set is a single sample of data, as opposed to a group of data samples. Similarly to k-fold validation, the process of using different validation sets until all data points have been tested at the validation set still occurs. A diagram showing the data splits and process of LOOCV can be seen in Figure 9. Each iteration of the model uses a different singular point as its testing case, a point which is highlighted in each different model of our figure.

3.5 Image Reconstruction Start to Finish

The idea of image reconstruction is relatively straightforward. We start with a set of images, and conduct principal component analysis (PCA). Then, when we are given new images, we can use the principal components to determine whether the given image represents a positive or negative pain event. An overview of how this classification method will work can be seen in Figure 10. Image reconstruction relies upon the idea that different classes of images *look* different, or have distinct

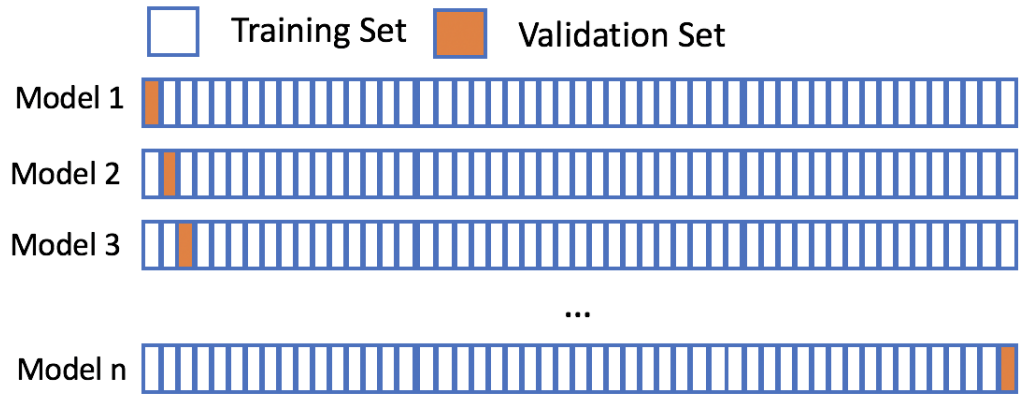

Figure 9. Model splits using LOOCV

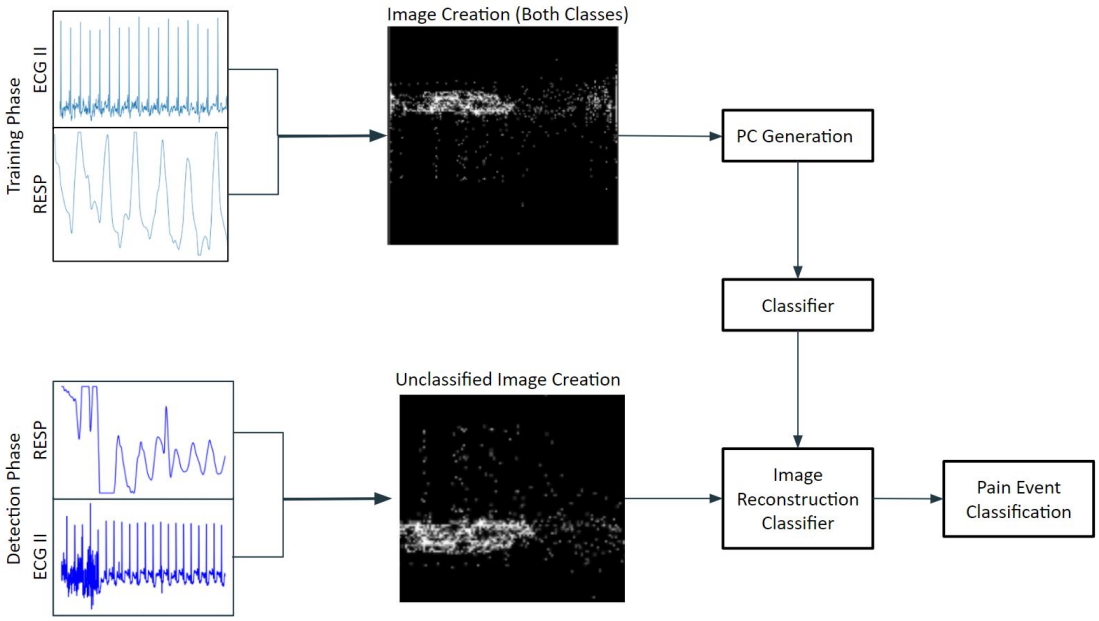

Figure 10. Overarching Diagram of Image Reconstruction Process

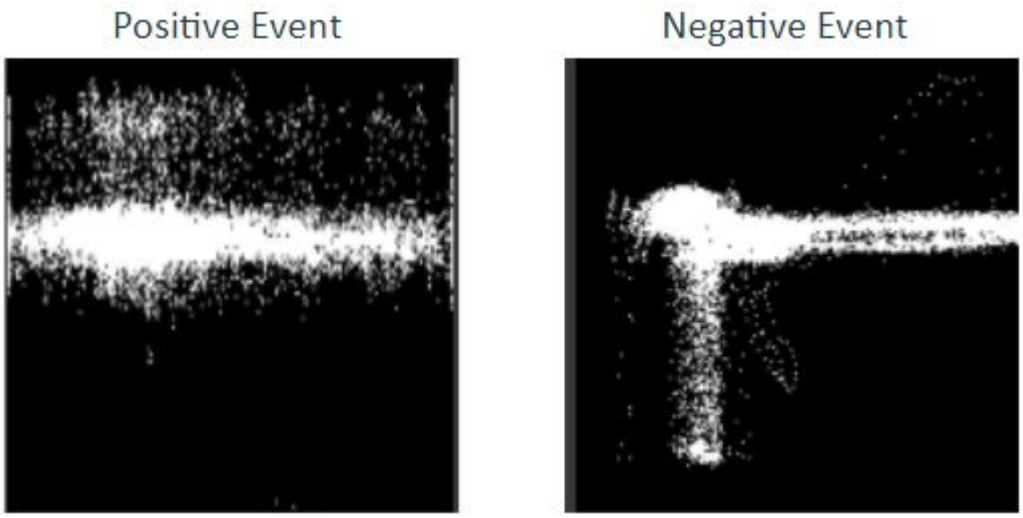

Figure 11. Comparison of positive and negative pain event

features that can be used to uniquely identify them. We expect that image reconstruction will work for pain classification because images of the given physiological streams exhibit different features. For an example, we have extracted an image of a positive and negative pain event with II and RESP from our dataset in Figure 11. By these two images it is clear that each class of images will have distinct features from one another. If each class of images looked exactly the same, then the classifier would have no way to discern which is which as all calculations would effectively be the same and image reconstruction would be useless for our purposes.

Before introducing the processes behind image reconstruction, we must first describe relevant hyperparameters that will be used in the process, as well as their individual purposes. The main hyperparameters we will be referencing are a window size w, a block size Δ, the number of principal components k, and the image resolution r. W is the length of time in seconds that we pull data in order build our images. Δ is the larger time frame where we will pull the length w time segments from. The k principal components will be used to extract the features from our

images and ultimately to classify our images. And lastly, image resolution r will not be tested much, but is important to include in our set of parameters as the resolution of images must be large enough to capture the inter-signal relationships of our data but small enough to reduce the complexity of our resulting images.

To get started with the process of image reconstruciton, we begin with the **training phase**. In this phase, we generate images of our data and conduct PCA, which gives us vectors that can be used for actual reconstruction in our detection phase. First, we take the time series data of our physiological streams and create a *portrait*. A portrait is an n-dimensional representation of a relationship of time series streams. For our purposes, the streams in our portraits will be either ABP, II, and RESP, or II and RESP. When given w seconds of our time series data, we normalize the data streams so that all values fall into the range $[0, 1]$. We do this because ABP, II, and RESP all are recorded with values of incomparable scales, and normalizing them allows us to make reasonable comparisons [19]. Then, we plot each of the streams against each other so that for each of the n dimensions, we have one stream of data. This process effectively produces a scatter plot of our physiological streams for the given time window. For simplification purposes, all visual representations of our processes will be pulled from the 2-dimensional approach of using just II and RESP as opposed to the full 3-dimensional approach. One sample of the resulting portrait from converting given II and RESP streams can be seen in Figure 12.

Once we have portraits generated, we then construct *images* from these portraits. To do so we attempt to view the given portrait as an image of resolution r. We pass the portrait in as an $n*n$ (or $n*n*n$) grid, each element of which describes whether any points are stored within them. To convert this grid to an image, we translate the information to an $n * n$ (or $n * n * n$) matrix that is binarized. This

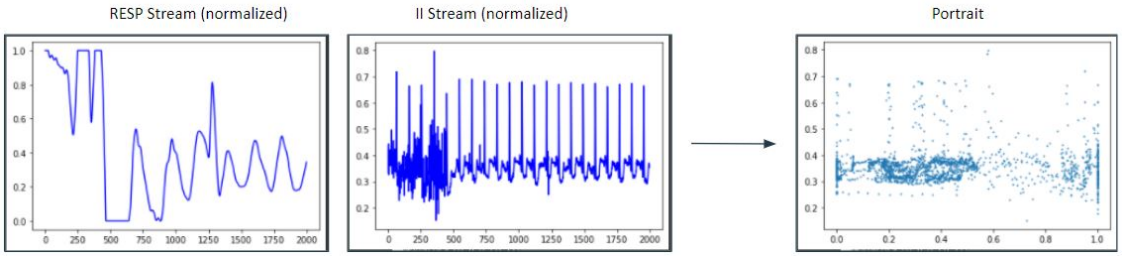

Figure 12. Sample portrait generation. Two streams of data are converted to a single portrait

means that where there is no data present at row i, column j, a 0 will be placed at this point in the matrix. But, if there is data present at row i, column j, a 1 will be placed at this point in the matrix. A sample of the resulting image after converting a portrait to a binarized state can be seen in Figure 13. The process of creating portraits and converting them to images is repeated until a total of Δ time units have been transformed into images. For instance, if w is 5 seconds and Δ is 10 seconds, the process will end with 2 images.

Now with all images generated for the positive pain event class and negative pain event class from our base working dataset, we can create a classifier used to identify which class a given image belongs to and conduct PCA to do so. The aim of PCA is to construct a set of principal components that will be used to explain variation in images and bring out patterns from the given data [20]. The following algorithms are adapted from Cai and Venkatasubramanian[17]. If we define m as the total number of images for a given class (positive or negative), we can represent each image as a column vector v_i, then we can generate a set of principal components by first computing the mean of the column vectors of our given class. This mean can be found using the following equation:

$$\mu_x = \frac{1}{m} \sum_{i=1}^{m} v_i$$

Once we have our mean, we can then create a covariance matrix C_x where x is the

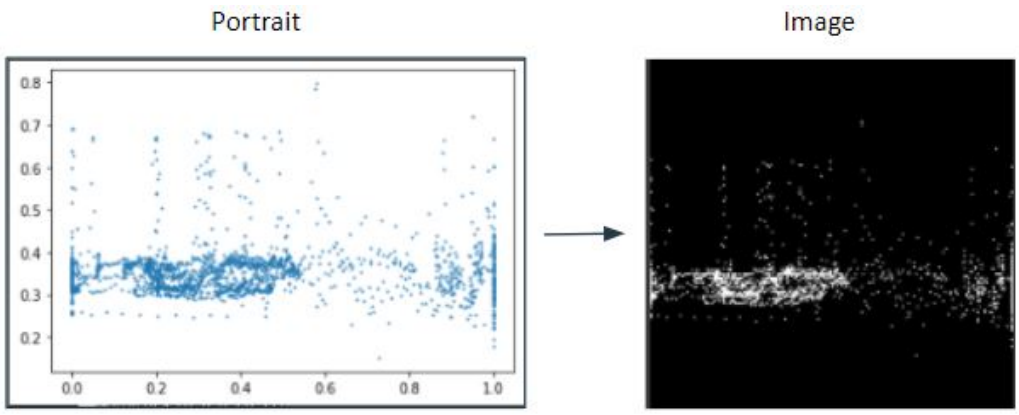

Figure 13. Converting a portrait to an image

class of either positive or negative pain events depending on which class we are constructing the covariance matrix for. This matrix can be calculated using the equation:

$$C_x = \sum_{i=1}^{m} (v_i - \mu_x)(v_i - \mu_x)^T$$

These equations were introduced in [16] and used in [17] for the same purposes. After we have conducted this process for each class, we will have two sets of principal components: one for positive pain events, and one for negative pain events. This is the main element of our classifier because each set will extract the major variations of our classes. In order to use our covariance matrix for classification, we first conduct an eigenanalysis on C_x. We start by finding the eigenvectors of C_x and sort them in non increasing order based on their corresponding eigenvalues. Then, we take the first k eigenvectors, and use them as rows of data to build a matrix denoted P_x.

The training phase is complete once we have created P_x for each class, and we move on to the **detection phase** from there. In order to make our detections, we take a brand new input image u, and begin by projecting the image onto an eigenspace using P_x using the following formula:

$$p = P_x(u - \mu_x)$$

Then, with this projection we can reconstruct the image with the formula:

$$u'_x = P_x^T p + \mu_x$$

This process is done twice for each image u: once attempting to reconstruct it as a positive pain event, and the other trying to reconstruct as a negative pain event. Then, for each reconstructed image we compute a *reconstruction error* via euclidean distance between each reconstructed image and the original. The reconstruction error can be calculated as follows:

$$error = |u'_x - u|$$

With both errors calculated we can classify our image. If the reconstruction error for the positive class is less than that of the negative class, we classify the image as positive. If the reconstruction error for the negative class is greater than or equal to the positive class, we classify the image as negative.

3.6 Hyperparameter Tuning Process Overview

In this section we describe the process of tuning our hyperparameters before presenting our target metrics and ultimately our results. We start with the least important hyperparameter mentioned above, the image resolution r. The optimal resolution, $r = 40$ was found empirically by testing observing images at resolutions 20, 30, 40, 50, 60, 100, 150, 200, and 300. We then moved onto selecting the window size w. To select the window size, we set Δ to a fixed value of 15 minutes and k to a fixed value of 5. Then we evaluated our model with various window sizes, including 5, 10, 15, 20, 25, 50, 100, and 150 seconds. After testing the various window sizes, we attempted to find an optimal Δ value. To do so, we evaluated our model with a fixed w of 25 seconds and fixed k of 5, and modified the Δ value to be 5, 10, 15, 20, or 30 minutes. These time windows were centered around the pain observation, meaning a 30 minutes Δ contains 15 minutes of data before and

15 minutes of data after the pain event. Lastly, we evaluated various k values with a fixed Δ of 15 minutes and a fixed w of 25 seconds. These k values includ 5, 10, 15, 20, and 25.

3.7 Metrics to be Used in Evaluation

While evaluating our models, there are multiple metrics that relate to our project. Firstly, we look to the false negative rate (FN) and the false positive rate (FP). The fasle negative rate is defined as the fraction of cases where positive pain events are classified as negative events. The false positive rate is defined as the fraction of cases where negative pain events are classified as positive ones. Next, we will use the true positive rate (TP) and true negative rate (TN). True positive rate refers to the fraction of positive pain events that are classified as positive ones, and true negative rate refers to the fraction of negative pain events classified as negative. These two rates are used in our calculation of a balanced accuracy rate (BAC), which can be found with the equation:

$$BAC = 0.5 * TP + 0.5 * TN$$

Beyond the BAC, we also observe both precision and recall values. Precision is the fraction of true positive classifications over the number of total positive classifications. This metric tells us of all positive classifications made by the model how many are actually positive pain events. Recall is the fraction of true positives classifications over the total number of actual positive values in our dataset, which provides us with the number of positive predictions made out of all positive predictions that could have been made.

List of References

[1] A. Johnson, *Mimic-III Clinical Database.*, PhysioNet Std., 2016. [Online]. Available: https://physionet.org/content/mimiciii/1.4/

[2] B. Erdoğan, *Objective Pain Assessment Using Vital Signs.*, Procedia Computer Science Std., Apr. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S187705092030541X

[3] Y. Chu, *Physiological Signal-Based Method for Measurement of Pain Intensity*, Fronteirs Media Std., May 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5445136/

[4] M. Jiang, *Acute pain intensity monitoring with the classification of multiple physiological parameters*, U.S. National Library of Medicine Std., June 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6499869/

[5] H. Kang, *The prevention and handling of the missing data*, U.S. National Library of Medicine Std., May 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3668100/

[6] N. Malhotra, *Analyzing Marketing Research Data with Incomplete Information on the Dependent Variable*, Journal of Marketing Research Std., Feb. 1987. [Online]. Available: https://journals.sagepub.com/doi/pdf/10.1177/002224378702400107

[7] R. Hamer, *Last observation carried forward versus mixed models in the analysis of psychiatric clinical trials*, U.S. National Library of Medicine Std., June 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19487398/

[8] S. Sinharay, *The use of multiple imputation for the analysis of missing data*, U.S. National Library of Medicine Std., Dec. 2001. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/11778675/

[9] S.-H. Kim, *Physiocover: Recovering the Missing Values in Physiological Data of Intensive Care Units.*, The Korea Contents Association Std., June 2014. [Online]. Available: https://www.koreascience.or.kr/article/JAKO201420249945774.view?orgId=kocon

[10] X. Chen, *Forecasting acute hypotensive episodes in intensive care patients based on a peripheral arterial blood pressure waveform*, ResearchGate Std., Oct. 2009. [Online]. Available: https://www.researchgate.net/publication/224130340_Forecasting_acute_hypotensive_episodes_in_intensive_care_patients_based_on_a_peripheral_arterial_blood_pressure_waveform

[11] G. Volpi, *Class Imbalance: a classification headache*, Towards Data Science Std., June 2019. [Online]. Available: https://towardsdatascience.com/class-imbalance-a-classification-headache-1939297ff4a4

[12] R. Barandela, *Strategies for Learning in Class Imbalance Problems.*, ResearchGate Std., Mar. 2003. [Online]. Available: https://www.researchgate.net/publication/220604068_Strategies_for_Learning_in_Class_Imbalance_Problems

[13] *10 Techniques to deal with Imbalanced Classes in Machine Learning*, Analytics Vidhya Std., July 2020. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/07/10-techniques-to-deal-with-class-imbalance-in-machine-learning/

[14] N. Chawla, *SMOTE: Synthetic Minority Over-sampling Technique*, Cornell University Std., June 2002. [Online]. Available: https://arxiv.org/pdf/1106.1813.pdf

[15] N. Moniz, *Resampling Strategies for Imbalanced Time Series Forecasting*, Springer Std., Feb. 2017. [Online]. Available: https://link.springer.com/content/pdf/10.1007%2Fs41060-017-0044-3.pdf

[16] L. Malagón-Borja, *Object Detection Using Image Reconstruction with PCA.*, Association for Computing Machinery Std., Jan. 1970. [Online]. Available: https://dlnext.acm.org/doi/abs/10.1016/j.imavis.2007.03.004

[17] H. Cai, *Detecting Data Manipulation Attacks on Physiological Sensor Measurements in Wearable Medical Systems*, Springer International Publishing Std., 2018. [Online]. Available: https://jis-eurasipjournals.springeropen.com/articles/10.1186/s13635-018-0082-y

[18] J. Brownlee, *A Gentle Introduction to k-fold Cross-Validation*, Machine Learning Mastery Std., May 2018. [Online]. Available: https://machinelearningmastery.com/k-fold-cross-validation/

[19] S. Lakshmanan, *How, When, and Why Should You Normalize / Standardize / Rescale Your Data?*, Towards AI Std., May 2019. [Online]. Available: https://towardsai.net/p/data-science/how-when-and-why-should-you-normalize-standardize-rescale-your-data-3f083def38ff

[20] I. Jolliffe, *Principal component analysis: a review and recent developments*, The Royal Society Std., Apr. 2016. [Online]. Available: https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202#:~:text=Principal%20component%20analysis%20(PCA)%20is,variables%20that%20successively%20maximize%20variance.

CHAPTER 4

Results

In this section we discuss the results of our modeling procedures. We first explore the results of our hyperparameter tuning and cross validation. From there we explore the success of our model on unseen data. For each section, we begin by discussing the results of our 2 dimensional approach involving just II and RESP, followed by the results of our 3 dimensional approach involving II, RESP, and ABP.

4.1 Hyperparameter Tuning Results with Metrics from LOOCV Outcomes

As mentioned previously, we test various values for each hyperparameter by locking the keeping the other testable hyperparameters constant and modifying the value of the hyperparameter at hand. In order to better predict the performance of our models, for each testing run, we ran LOOCV as described previously. This was conducted using 90% of the dataset to split into training and validation. Beginning with the 2 dimensional testing, Figure 14 exhibits the results of testing various values of the window size w. A window size of 20 seconds produced the best overall results with a BAC of 79.26%. Additionally, its precision and recall values were the highest our of all tested values meaning that a window size of 20 seconds makes the most accurate positive classifications out of all the values. Following window size testing, we tested out values for Δ. Figure 15 displays these results, and we can observe that a Δ of 10 minutes produced the best results. This pairing of parameters resulted in a BAC of 82.5%. Like in testing the w, the optimal Δ also produced the highest precision and recall values. And lastly we tested out various values for k, Figure 16 shows the results that $k = 5$ yields optimal BAC

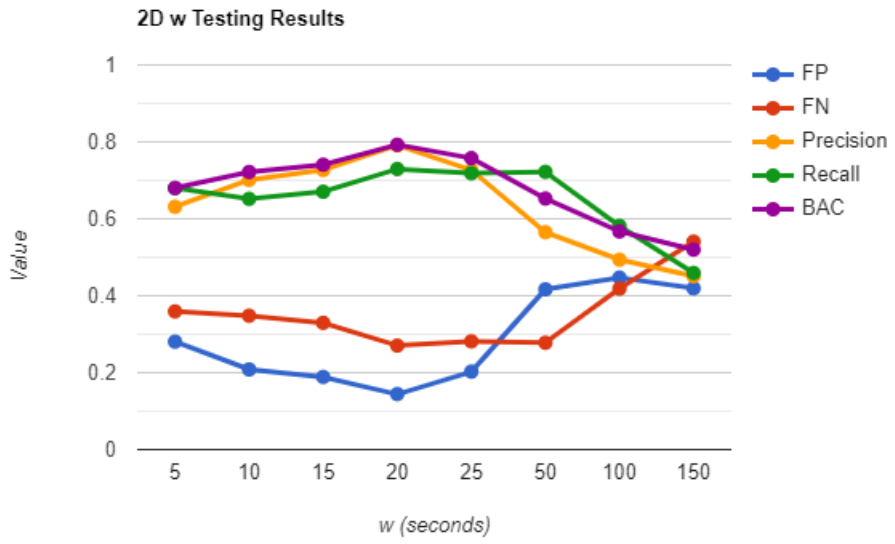

Figure 14. 2D w testing. 20 seconds results in the highest BAC of 79.26%

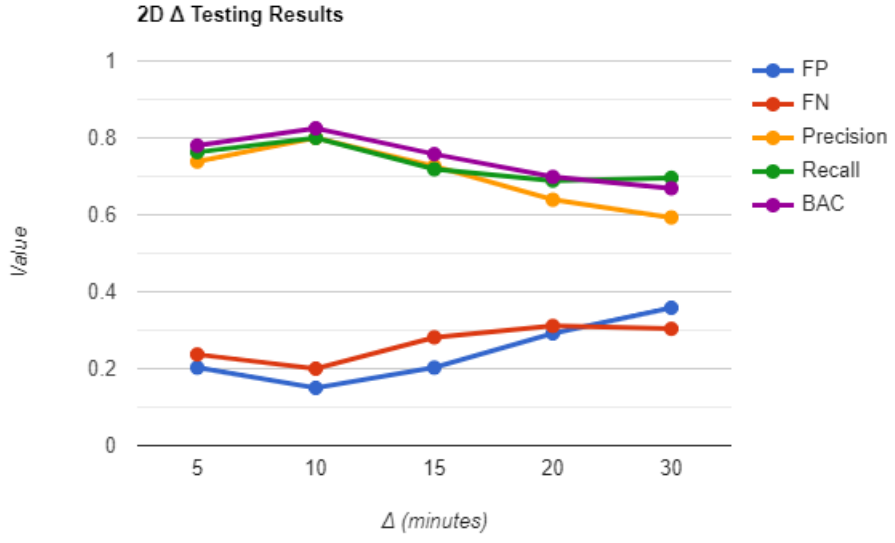

Figure 15. 2D Δ testing. 10 minutes results in the highest BAC of 82.5%

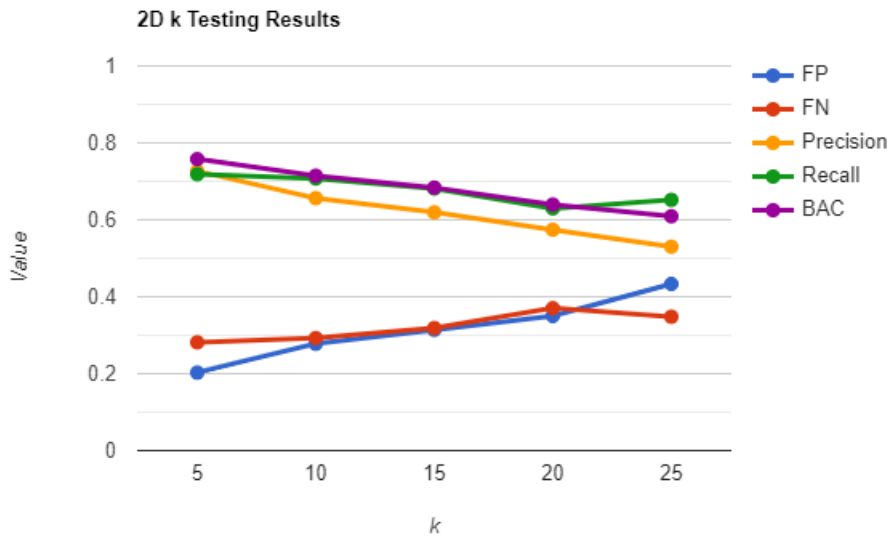

Figure 16. 2D k testing. 5 principal components results in the highest BAC of 75.79%

at 75.79%. Ultimately, through our testing we have concluded that the optimal hyperparameters for our 2 dimensional model are $w = 20$ seconds, $\Delta = 10$ minutes, and $k = 5$ principal components as each performs best through each respective testing series.

After this round of testing was completed, we then found optimal hyperparameters for the 3 dimensional model. Beginning with w, we tested 5, 10, 15, 20, 25, 30, 50, 100, and 150 seconds. In this case we test 30 seconds uniquely for the 3 dimensional case because in testing, 25 seconds was optimal, but there was a large jump from 25 to 50 seconds and so we tested new values at increments of 5 seconds until the accuracy went down. Figure 17 shows the results of our testing. A w of 25 seconds yielded optimal results for our 3 dimensional model, with a BAC of 85.42%, and the highest precision and recall values of the tested options. Like before, we then tested values for Δ in the range 5, 10, 15, 20, 25, and 30 minutes. Again due to the larger jump from 20 to 30, we additionally tested 25 minutes to

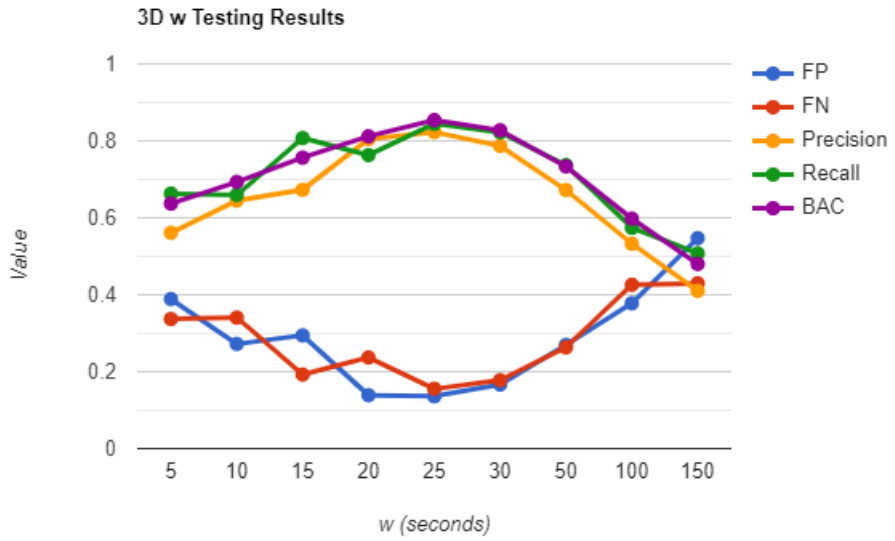

Figure 17. 3D w testing. 25 seconds results in the highest BAC of 85.42%

be sure that our results stood. A 20 minutes Δ produced the best results with a BAC of 86.71%. The results of this testing series can be viewed in Figure 18. Finally we evaluated the same k values for our 3 dimensional model. As can be seen in Figure 19 and similarly to the 2 dimensional approach, $k = 5$ principal components produced the best results at a balanced accuracy of 85.42% with the best precision and recall values of the testing series.

4.2 Final Testing Results

This section summarizes the final results found in the testing phase of both our 2 dimensional and 3 dimensional approach. The results were obtained using the test set described previously in section 3.4 (30 positive pain events and 40 negative pain events). During model training, none of the data in our test set was seen by the models, and so this will demonstrate how our models will generalize to unseen data. In order to test our final models, we used each optimal hyperparameter in a combination together. In other words, for the 2 dimensional approach we

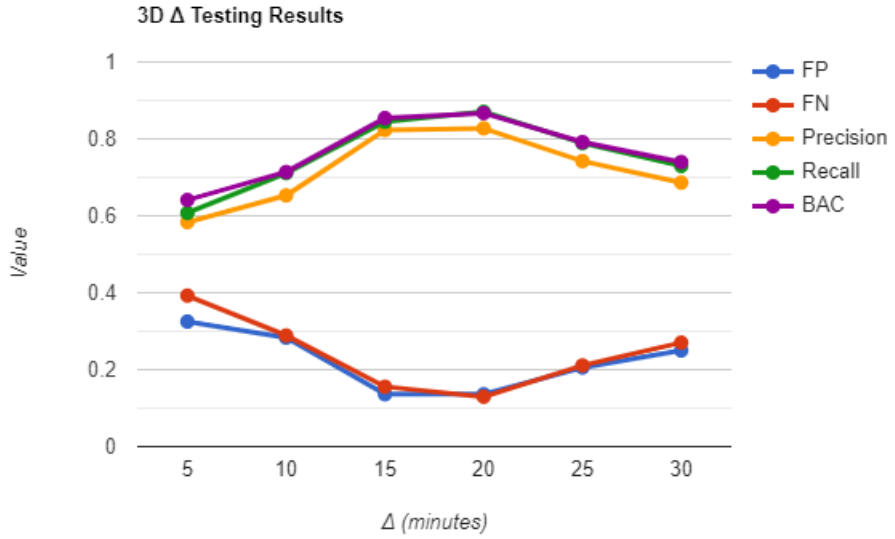

Figure 18. 3D Δ testing. 20 minutes results in the highest BAC of 86.71%

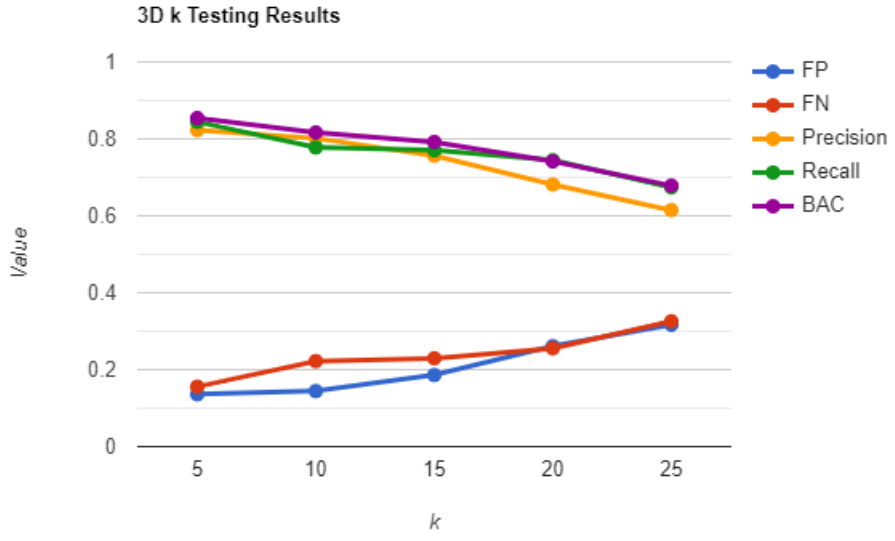

Figure 19. 3D k testing. 5 principal components results in the highest BAC of 85.42%

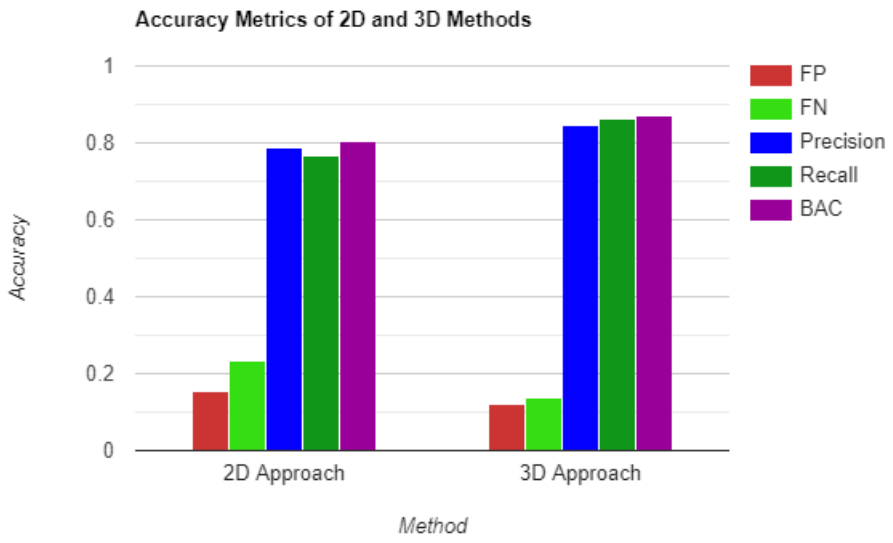

Figure 20. Final testing results presented. The 3 dimensional approach achieves a BAC of 87.18% while the 2 dimensional approach achieves a BAC of just 80.56%

use $w = 20$ seconds, $\Delta = 10$ minutes, and $k = 5$ principal components. For the 3 dimensional approach, we utilize $w = 25$ seconds, $\Delta = 20$ minutes and $k = 5$ principal components. Results of both approaches are presented in Figure 20. Following a general presentation of these results, we move onto their meaning and conclusions we can draw from them in the next chapter. Looking at the 2 dimensional approach first, we see a false positive rate of 15.56% and a false negative rate of 23.33%. The 2 dimensional approach yielded a precision score of 78.71%, recall of 76.67%, and a balanced accuracy of 80.56%. The 3 dimensional approach performed better across the board, with a false positive rate and false negative rate of 11.94% and 13.70%, respectively. This approach also resulted in a precision of 84.42%, a recall of 86.30%, and a balanced accuracy of 87.18%.

4.3 Discussion of Results

As mentioned previously, FN tells us how often positive pain events are classified as negative ones and FP tells us how often negative events are classified

as positive ones. Universally, we observe that the FN of our image reconstruction model is larger than the FP. This indicates that our model generally misses positive pain events more often than it does negative pain events. In a medical context this is the preferred scenario, as even though pain can affect quality of life as mentioned previously, pain in itself is not fatal [1], and so not observing pain immediately will not directly lead to patient mortality. On the other hand, falsely identifying pain may lead to health care workers prescribing pain medications that are unneeded, which can lead to adverse side effects [2]. Looking at the recall and precision values, both classifiers accurately predict positive pain events a vast majority of the time, leading us to believe that our model is successful in its task of predicting pain in general. Compared with previous work, our model performs better. We utilize only 3 physiological streams to produce an accuracy of 87.18% while other works have used more than double this to reach an accuracy of just 76.1%. Ultimately our study displays the viability of objective pain assessment when a model is presented with physiological data. As an extension of this idea, image reconstruction based classifiers appear useful in the context of performing binary classification when given physiological data streams presented as image data. There are some higher upfront computational costs in this methodology for the training phase, but once a base classifier is created, detection is extremely fast and simple, and our overall model is highly interpretable. Moreover, adding additional data to the dataset does not prompt us to fully reconstruct a classifier, as we would still have access to the original averages, covariance matrices and other values required to classification. The only modification would be the data set size, and adding in the new values or the additional data into the calculations. Although our model performs well, a maximum accuracy of 87.18% is far from the perfect classifier. Similarly to the conclusions of other works, models used for objective pain predic-

tion should be utilized in conjunction with current subjective methodology and the professional opinion of health care workers as any errors could prove problematic for health care workers and patients.

List of References

[1] C. Freudenrich, *How Pain Works*, HowStuffWorks Std., Feb. 2019. [Online]. Available: https://heal-responsibly.com/wp-content/uploads/2020/08/ebfed-introduction-to-how-pain-works-_-howstuffworks.pdf

[2] *Overprescribing of medicines must stop, says government*, BBC Std., 2021. [Online]. Available: https://www.bbc.com/news/health-58639253

CHAPTER 5

Conclusions and Future Work

5.1 Conclusions

In this study, we proposed a new method for classifying the presence or absence of pain using physiological signals. We began by describing the general field of machine learning and its relation to pain classification. We found that while some work has been conducted in the area of machine learning in pain research, a majority of it relates to predicting levels of reported pain as opposed to presence. These works operate on the assumption that a patient is reporting pain and show some success in their mission. We then transition towards building a model for binary pain classification. We walk through the process of trimming, imputing within, and splitting our dataset, and describe our model; image reconstruction based methodology has never been attempted within this context to our knowledge. The main goal of this study is to present an alternative to classical machine learning methods, with the hope of improving upon the results of previous studies.

We attempted two approaches for our image reconstruction based classifier: a 3 dimensional and a 2 dimensional approach. The 2 dimensional method included ABP and II for each pain event. With this classifier, we ended with a precision of 78.71% and a recall of 76.67%. This generally means that the majority of the time, positive predictions are accurate. This model also makes false negative classifications more often than false positive classifications, each at 23.33% and 15.56%, respectively. For the purposes of this study, we declare that false negatives are more acceptable than false positives, depending on context. It is known that pain itself is not fatal, as pain is simply a signal sent from the brain. While some physiological processes such as altered vital signs can occur as a result of pain, the pain itself is not deadly [1]. With this in mind, if a patient is determined to

not be in pain by our model, the worst immediate case is that they do not receive treatment. Assuming the event is acute pain as opposed to chronic, there will be no long term effects as mentioned previously in section 1.1 [2]. However, with false positives, there may be the risk of prescribing unneeded pain medications could lead to adverse events. For instance, a patient who is prescribed morphine could ultimately develop an addiction [3] and thus, false negatives are much more acceptable than false positives. In the end, the 2 dimensional approach produced a balanced accuracy of 80.56% when presented with 20 seconds of physiological data.

Moving onto our 3 dimensional approach, which included ABP, II, and RESP. This model far outperforms the 2 dimensional approach, producing a false positive rate of 11.94% and false negative rate of 13.70%. Its precision and recall values were 84.42% and 86.30%, respectively, indicating that for a vast majority of cases, positive predictions made by this model are accurate. The 3 dimensional approach produced an ultimate balanced accuracy of 87.18%, given just 25 seconds of data to work with. Overall this indicates that when more streams of data are included in the classifier, more information must be extracted from the reconstruction based classifier in order to make a decision on the class of an input image, which intuitively makes sense. As an extension, more streams also appears to produce better results in our modeling as well, a phenomenon which can be explored more in the future.

5.2 Limitations

Though the results of our study are promising there are some limitations that should be addressed. Firstly, the problem being solved in this work is a binary classification of pain presence or absence. This lends itself more towards exploring the **viability** of objective pain assessment. Within specific medical contexts, de-

Figure 21. Subjective pain rating scale

tecting the presence of pain may not always be fully useful. For instance, it may be more useful to classify the level of pain a nonverbal patient is experiencing so that proper treatment can be administered. However, it is important to still establish that pain exists first before classifying a level.

Another key limitation of our study is that our dataset contains fully subjective assessments of pain. Patients often referenced the Wong-Baker pain chart pictured in Figure 21 while rating their pain; this model has been used for many years in pain assessment. Though the goal of our study is to objectively identify the presence of pain, we may still have skewed instances of data where a patient not experiencing pain reports a positive pain event or vice versa.

A third limitation of our work is that we rely upon the relationship between several physiological signals in order to make our predictions. While our model is not user-specific, having just 700 pain events to work with may not greatly represent the human population as a whole, and so if used in practice, our model may perform better with some populations than others. The obvious solution to this is training a model on much larger sample sizes, but this in itself is a problem in the area of pain research.

5.3 Future Work

In the future we plan to improve upon this work in a few different ways. Firstly, we plan to collect more data from patients experiencing pain. This increase in data could help to address the class imbalance problem before the dataset is altered, and will assist in better identifying our model generalizability. Next, we will expand the study to incorporate more physiological streams. As we found in our study that more streams appears to produce better results, we plan to incorporate other streams including skin conductance or heart rate variability as they have been associated with pain as well. [4, 5, 6]. Incorporating more streams into our datasets could allow us to further investigate the idea that more streams produces better models at the cost of potentially more prediction runtime and other costs associated with more data usage. Lastly, in the future we plan to attempt to incorporate image reconstruction into the world of predicting pain levels. There are several successful models that predict levels of pain as mentioned previously, but predicting beyond a binary classification problem expands beyond this work and [7], and could help understand the application of image reconstruction to physiological signal data in general.

List of References

[1] C. Freudenrich, *How Pain Works*, HowStuffWorks Std., Feb. 2019. [Online]. Available: https://heal-responsibly.com/wp-content/uploads/2020/08/ebfed-introduction-to-how-pain-works-_-howstuffworks.pdf

[2] N. Wells, *Improving the Quality of Care through Pain Assessment and Management.*, U.S. National Library of Medicine Std., Apr. 2008. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK2658/

[3] *Opioid Addiction*, Johns Hopkins Medicine Std. [Online]. Available: https://www.hopkinsmedicine.org/opioids/science-of-addiction.html

[4] B. Susam, *Automated Pain Assessment Using Electrodermal Activity Data and Machine Learning.*, U.S. National Library of Medicine Std., July 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30440413/

[5] H. Storm, *Changes in Skin Conductance as a Tool to Monitor Nociceptive Stimulation and Pain.*, U.S. National Library of Medicine Std., Dec. 2008. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18997532/

[6] S. Telles, *Heart rate variability in chronic low back pain patients randomized to yoga or standard care*, BMC Software Std., Aug. 2016. [Online]. Available: https://bmccomplementmedtherapies.biomedcentral.com/articles/10.1186/s12906-016-1271-1#:~:text=Chronic%20pain%20is%20an%20emotionally\protect\protect\leavevmode@ifvmode\kern+.1667em\relaxin%20pain%20regulation%20%5B3%5D.

[7] H. Cai, *Detecting Data Manipulation Attacks on Physiological Sensor Measurements in Wearable Medical Systems*, Springer International Publishing Std., 2018. [Online]. Available: https://jis-eurasipjournals.springeropen.com/articles/10.1186/s13635-018-0082-y

BIBLIOGRAPHY

10 Techniques to deal with Imbalanced Classes in Machine Learning, Analytics Vidhya Std., July 2020. [Online]. Available: https://www.analyticsvidhya.com/blog/2020/07/10-techniques-to-deal-with-class-imbalance-in-machine-learning/

Barandela, R., *Strategies for Learning in Class Imbalance Problems.*, ResearchGate Std., Mar. 2003. [Online]. Available: https://www.researchgate.net/publication/220604068_Strategies_for_Learning_in_Class_Imbalance_Problems

Overprescribing of medicines must stop, says government, BBC Std., 2021. [Online]. Available: https://www.bbc.com/news/health-58639253

Booker, S., *Assessing Pain in Nonverbal Older Adults.*, U.S. National Library of Medicine Std., May 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4991889/

Boring, B., *How and Why Patient Concerns Influence Pain Reporting: A Qualitative Analysis of Personal Accounts and Perceptions of Others' Use of Numerical Pain Scales.*, Frontiers Std., Jan. 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.663890/full

Brownlee, J., *A Gentle Introduction to k-fold Cross-Validation*, Machine Learning Mastery Std., May 2018. [Online]. Available: https://machinelearningmastery.com/k-fold-cross-validation/

Cai, H., *Detecting Data Manipulation Attacks on Physiological Sensor Measurements in Wearable Medical Systems*, Springer International Publishing Std., 2018. [Online]. Available: https://jis-eurasipjournals.springeropen.com/articles/10.1186/s13635-018-0082-y

Chawla, N., *SMOTE: Synthetic Minority Over-sampling Technique*, Cornell University Std., June 2002. [Online]. Available: https://arxiv.org/pdf/1106.1813.pdf

Chen, X., *Forecasting acute hypotensive episodes in intensive care patients based on a peripheral arterial blood pressure waveform*, ResearchGate Std., Oct. 2009. [Online]. Available: https://www.researchgate.net/publication/224130340_Forecasting_acute_hypotensive_episodes_in_intensive_care_patients_based_on_a_peripheral_arterial_blood_pressure_waveform

Chu, Y., *Physiological Signal-Based Method for Measurement of Pain Intensity*, Fronteirs Media Std., May 2017. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5445136/

Coghill, R. C. U.S. Natnioal Library of Medicine. "Individual differences in the subjective experience of pain: New insights into mechanisms and models." Oct. 2010. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2959190/

Erdoğan, B., *Objective Pain Assessment Using Vital Signs.*, Procedia Computer Science Std., Apr. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S187705092030541X

Freudenrich, C., *How Pain Works*, HowStuffWorks Std., Feb. 2019. [Online]. Available: https://heal-responsibly.com/wp-content/uploads/2020/08/ebfed-introduction-to-how-pain-works-_-howstuffworks.pdf

Hamer, R., *Last observation carried forward versus mixed models in the analysis of psychiatric clinical trials*, U.S. National Library of Medicine Std., June 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19487398/

Herr, K., *Tools for Assessment of Pain in Nonverbal Older Adults with Dementia: A State-of-the-Science Review.*, Journal of Pain and Symptom Management Std., Feb. 2006. [Online]. Available: https://www.jpsmjournal.com/article/S0885-3924(05)00611-1/fulltext

Institute of Medicine (US) Committee on Pain, Disability, and Chronic Illness Behavior. "The anatomy and physiology of pain." Jan. 1987. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK219252/

Jiang, M., *Acute pain intensity monitoring with the classification of multiple physiological parameters*, U.S. National Library of Medicine Std., June 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6499869/

Opioid Addiction, Johns Hopkins Medicine Std. [Online]. Available: https://www.hopkinsmedicine.org/opioids/science-of-addiction.html

Johnson, A., *Mimic-III Clinical Database.*, PhysioNet Std., 2016. [Online]. Available: https://physionet.org/content/mimiciii/1.4/

Jolliffe, I., *Principal component analysis: a review and recent developments*, The Royal Society Std., Apr. 2016. [Online]. Available: https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202#:~:text=Principal%20component%20analysis%20(PCA)%20is,variables%20that%20successively%20maximize%20variance.

Kang, H., *The prevention and handling of the missing data*, U.S. National Library of Medicine Std., May 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3668100/

Katz, J. Surgical Clinics of North America. "Measurement of pain." Apr. 1999. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0039610905703819?via%3Dihub

Katz, N. Journal of Pain and Symptom Management. "The impact of pain management on quality of life." July 2002. [Online]. Available: https://www.jpsmjournal.com/article/S0885-3924(02)00411-6/fulltext

Kim, S.-H., *Physiocover: Recovering the Missing Values in Physiological Data of Intensive Care Units.*, The Korea Contents Association Std., June 2014. [Online]. Available: https://www.koreascience.or.kr/article/JAKO201420249945774.view?orgId=kocon

Kim, Y.-S., *Assessment of pain in the elderly: A literature review*, The National Medical Journal of India Std., Apr. 2017. [Online]. Available: https://nmji.in/assessment-of-pain-in-the-elderly-a-literature-review/

Lakshmanan, S., *How, When, and Why Should You Normalize / Standardize / Rescale Your Data?*, Towards AI Std., May 2019. [Online]. Available: https://towardsai.net/p/data-science/how-when-and-why-should-you-normalize-standardize-rescale-your-data-3f083def38ff

Lee, J., *Machine Learning-Based Prediction of Clinical Pain Using Multimodal Neuroimaging and Autonomic Metrics.*, U.S. National Library of Medicine Std., Mar. 2019. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30540621/

Lopez-Martinez, D., *Deep Reinforcement Learning for Optimal Critical Care Pain Management with Morphine Using Dueling Double-Deep Q Networks.*, Cornell University Std., Apr. 2019. [Online]. Available: https://arxiv.org/abs/1904.11115

Malagón-Borja, L., *Object Detection Using Image Reconstruction with PCA.*, Association for Computing Machinery Std., Jan. 1970. [Online]. Available: https://dlnext.acm.org/doi/abs/10.1016/j.imavis.2007.03.004

Malhotra, N., *Analyzing Marketing Research Data with Incomplete Information on the Dependent Variable*, Journal of Marketing Research Std., Feb. 1987. [Online]. Available: https://journals.sagepub.com/doi/pdf/10.1177/002224378702400107

Matsangidou, M., *Machine Learning in Pain Medicine: An Up-To-Date Systematic Review*, Springer Std., 2021. [Online]. Available: https://link.springer.com/article/10.1007/s40122-021-00324-2

McCarberg, B., *The Impact of Pain on Quality of Life and the Unmet Needs of Pain Management: Results From Pain Sufferers and Physicians Participating*

in an Internet Survey., American Journal of Therapeutics Std., July 2008. [Online]. Available: https://journals.lww.com/americantherapeutics/Abstract/2008/07000/The_Impact_of_Pain_on_Quality_of_Life_and_the.4.aspx

Moniz, N., *Resampling Strategies for Imbalanced Time Series Forecasting*, Springer Std., Feb. 2017. [Online]. Available: https://link.springer.com/content/pdf/10.1007%2Fs41060-017-0044-3.pdf

Price, D. D., *Unconscious and Conscious Mediation of Analgesia and Hyperalgesia.*, National Academy of Sciences of the United States of America Std., June 2015. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4485153/

Sinharay, S., *The use of multiple imputation for the analysis of missing data*, U.S. National Library of Medicine Std., Dec. 2001. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/11778675/

Storm, H., *Changes in Skin Conductance as a Tool to Monitor Nociceptive Stimulation and Pain.*, U.S. National Library of Medicine Std., Dec. 2008. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18997532/

Sugimine, S., *Normalized Skin Conductance Level Could Differentiate Physical Pain Stimuli from Other Sympathetic Stimuli.*, Nature News Std., July 2020. [Online]. Available: https://www.nature.com/articles/s41598-020-67936-0#::text=Skin%20conductance%20C%20especially%20nSCL%2C%20was,including%20tactile%20and%20mental%20stimuli

Susam, B., *Automated Pain Assessment Using Electrodermal Activity Data and Machine Learning.*, U.S. National Library of Medicine Std., July 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30440413/

Telles, S., *Heart rate variability in chronic low back pain patients randomized to yoga or standard care*, BMC Software Std., Aug. 2016. [Online]. Available: https://bmccomplementmedtherapies.biomedcentral.com/articles/10.1186/s12906-016-1271-1#:~:text=Chronic%20pain%20is%20an%20emotionally\protect\protect\leavevmode@ifvmode\kern+.1667em\relaxin%20pain%20regulation%20%5B3%5D.

Volpi, G., *Class Imbalance: a classification headache*, Towards Data Science Std., June 2019. [Online]. Available: https://towardsdatascience.com/class-imbalance-a-classification-headache-1939297ff4a4

Wells, N., *Improving the Quality of Care through Pain Assessment and Management.*, U.S. National Library of Medicine Std., Apr. 2008. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK2658/

Wong-Baker FACES Foundation. [Online]. Available: https://wongbakerfaces.org/