UC Santa Barbara
UC Santa Barbara Electronic Theses and Dissertations

Title

Engineering development and exploration of physiological instrumentation for health optimization and diagnostic applications

Permalink

https://escholarship.org/uc/item/8k02v9wz

Author

Ly, Franklin Somchith

Publication Date

2023

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA

Santa Barbara

Engineering development and exploration of physiological instrumentation for health optimization and diagnostic applications

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Mechanical Engineering

by

Franklin Somchith Ly

Committee in charge:

Professor Henry T. Yang, Chair

Professor Paul K. Hansma

Professor Linda Petzold

Professor Michael Miller

December 2023

The dissertation of Franklin Somchith Ly is approved.

_____
Paul K. Hansma

_____
Linda Petzold

_____
Michael Miller

_____
Henry T. Yang, Committee Chair

November 2023

Engineering development and investigation of physiological devices for health improvement and diagnostics

Copyright © 2023

by

Franklin Somchith Ly

ACKNOWLEDGEMENTS

The journey towards my Ph.D. has been an extensive and challenging endeavor that I could not have navigated without the guidance, support, and collaboration of numerous individuals. First and foremost, I express my gratitude to the students and collaborators who journeyed alongside me throughout these research endeavors. My sincerest thanks go to Dahyana Arroyo, Nirmit Ashar, Vincent Billard, John Chen, Kaie Chen, Destinee Cheng, Ericka Dixon, Jordan Gray, Kevin Hoffseth, Anirudh Iyer, Christopher Li, Kaiwen Li, Isaac Kwon, Kian Lonergan, Octavio Lopez, Pedro May, Stephany Pavlov, Aditi Phatak, Tyler Santander, Sergio Sokolovskiy, Elyes Turki, Zachary Vaillancourt, Celia Vann, Kegan Woodhouse, Yanis Yankauskas, Jamie Yoo, Jerry Zhang, Minghao Zhang, Jiayang Zhao, and Yun Zhao.

Heartfelt thanks to my advisors Michael Miller, Linda Petzold, Paul K. Hansma, and Henry T. Yang, who not only shared their invaluable insights and guidance but also enriched my experience through their collaboration. A special mention to Paul K. Hansma whose guidance has been instrumental in shaping my research direction towards physiological technology within the scope of chronic pain applications. I extend another special thanks to Henry T. Yang, who generously provided me with the opportunity to engage in this Ph.D. program and mentored me not only in research but also in life. I additionally acknowledge the Bill and Melinda Gates Foundation for providing me with the opportunity to pursue higher education.

I extend my deepest gratitude to my friends and family who have supported my journey. Whether our relationship comes from childhood, college, or volleyball, I greatly appreciate the relationship we have developed. A special thanks goes to my parents, Nhon and Nancy,

and my step-parents, Ha and Chaiwat, whose sacrifice and decision to immigrate to the United States paved the way for the plethora of opportunities that came my way. Furthermore, I would like to express heartfelt thanks to my sister, Elizabeth. Her guidance and assistance throughout the college admissions process and during my scholarship applications were invaluable. I would also like to thank my extended family, cousins, aunts, and uncles who all supported me through this journey.

I owe a special debt of gratitude to my fiancé, Chanel Singsavaddy. She has been my pillar of support throughout this journey, offering encouragement and understanding during the challenges. Her support, especially over the past year as she carried our baby and taken care of our newborn, Phoenix, has been nothing short of remarkable. Phoenix, both your mom and dad will always love you unconditionally. The completion of this Ph.D. will allow our family to do incredible activities and services together.

Without everyone's invaluable contributions and unwavering support, the journey towards this Ph.D. would not have been possible. Together, we have produced amazing research and developed technology that will undoubtedly benefit the wider community. From the depths of my heart, I express my profound gratitude. This accomplishment is not mine alone but a testament to the collective efforts of us all.

VITA OF FRANKLIN SOMCHITH LY

November 2023

EDUCATION

Bachelor of Science in Mechanical Engineering, University of California, Santa Barbara, June 2017

Doctor of Philosophy in Mechanical Engineering, University of California, Santa Barbara, October 2023 (expected)

PROFESSIONAL EMPLOYMENT

2017-22: Teaching Assistant, Department of Mechanical Engineering, University of California, Santa Barbara

2018-23: Graduate Student Researcher, Department of Mechanical Engineering, University of California, Santa Barbara

PUBLICATIONS

"Home-use and Portable Biofeedback Lowers Anxiety and Pain in Chronic Pain Subjects," Unpublished research article submitted to the American Journal of Lifestyle Medicine, (2023).

"Measurement and analysis of internal deformation under soft material indentation with experiment and finite elements," Unpublished research article submitted to the Journal of Biomimetics (2023).

"Preliminary study: quantification of chronic pain from physiological data," *Journal of Pain Reports*, (2022), 7 (6).

"Dynamic Phase Extraction: Applications in Pulse Rate Variability," *Journal of Applied Psychophysiology and Biofeedback*, (2022), 47 (3), 213-222.

"How much does it hurt: A deep learning framework for chronic pain score assessment," *International Conference on Data Mining Workshops* (ICDMW), (2020), 651-660.

"Significant correlation of bone material strength index as measured by the osteoprobe with vickers and rockwell hardness," *Review of Scientific Instruments*, (2020), *91 (8)*.

"Preliminary design, experiment, and numerical study of a prototype hydraulic bio-inspired damper," *Journal of Sound and Vibration*, (2019), 459, 114845.

"Bio-inspired passive base isolator with tuned mass damper inerter for structural control," *Journal of Smart Materials and Structures*, (2019), 28 (10), 105008.

AWARDS

Bill and Melinda Gates Millennium Fellowship, 2013-2023

FIELDS OF STUDY

Major Field: Mechanical Engineering

Studies in Physiological Sensor Research with Professors Paul K. Hansma and Henry T. Yang

Studies in Machine Learning Methods with Professor Linda Petzold

Studies in Human Subject Research Methods and Statistics with Professor Michael Miller

ABSTRACT

Engineering development and exploration of physiological instrumentation for health optimization and diagnostic applications

by

Franklin Somchith Ly

This dissertation explores the development and effectiveness of various physiological instruments for health optimization and diagnostic applications. It begins by studying the Bone Material Strength index (BMSi) and its significant correlation with conventional hardness measurements. This dissertation then includes the engineering development of prototype Pain Meters for the assessment of chronic pain. The research examines the potential of low-cost physiological sensors to measure chronic pain's physiological changes, establishing a significant correlation between physiological data and chronic pain at both individual and population levels. This discovery suggests the potential for developing accessible and affordable diagnostic tools for chronic pain.

The dissertation also explores pulse rate variability, developing an innovative technology like Dynamic Phase Extraction to correlate breathing rates and inter-beat intervals. This technology proves effective in providing insights into various health conditions. The research then delves into the use of portable biofeedback devices for managing chronic pain and anxiety in remote studies. Pilot studies report substantial reductions in pain and anxiety levels among participants using these devices, affirming their efficacy.

Finally, a study involving a handheld thermal biofeedback device reveals its potential to enhance sleep quality and reduce anxiety levels. The research concludes by underscoring the comprehensive capability of physiological instruments in diverse health applications, including anxiety management and sleep enhancement. The findings lay a strong foundation for future research and innovation in this field, with the potential to substantially enhance healthcare outcomes and patient well-being.

TABLE OF CONTENTS

I. Introduction 4

II. Significant correlation of Bone Material Strength index as measured by the OsteoProbe with Vickers and Rockwell hardness 4

A. Introduction 5

B. Methods 13

C. Results 8

D. Summary 11

III. How Much Does It Hurt: A Deep Learning Framework for Chronic Pain Score Assessment 13

A. Introduction 13

B. Methods 43

C. Results 30

1. Optimizing the Pain Meter Design 36

D. Summary 39

IV. Preliminary Study: Quantification of Chronic Pain from Physiological Data ... 40

A. Introduction 40

B. Methods 43

C. Results 52

1. Individual-level Models 52

2. Population-level Model 54

D. Summary 55

V. Dynamic Phase Extraction: Applications in Pulse Rate Variability 57

A. Introduction 58

B. Methods 130

C. Results 136

D. Summary 72

VI. Home-use and Portable Biofeedback Lowers Anxiety and Pain in Chronic Pain Subjects 713

A. Introduction 713

B. Methods 78

1. Participants 78

2. Procedures and Materials 79

3. Statistical Modeling 84

C. Results 87

D. Summary 94

VII. A Novel Thermal Biofeedback Device to Enhance Sleep Quality and Alleviate Anxiety 97

A. Introduction 97

B. Methods 99

1. Participants 99

2. Procedures and Materials 99

C. Results 101

D. Summary 105

I. Introduction

This dissertation explores various domains within medical science, emphasizing the role of biofeedback technology, deep learning, and physiological instrumentation in understanding and managing chronic pain and other health conditions. This work brings together these cutting-edge areas to offer a cohesive insight into their collective potential in enhancing patient outcomes.

Before delving into specific sections, it is imperative to provide an overview of physiological instrumentation. Physiological instruments, such as the OsteoProbe, prototype Pain Meter, and prototype biofeedback devices, played a crucial role in non-invasively measuring and analyzing various health parameters. These instruments have paved the way for more advanced, precise, and convenient diagnostic and monitoring tools, enhancing the ability to understand, assess, and manage various health conditions.

Section II: The OsteoProbe and Bone Material Strength Index

The OsteoProbe, a microindentation device, quantifies bone's resistance to indentation, providing the Bone Material Strength index (BMSi). This section will critically evaluate the reliability of the OsteoProbe's measurements against established hardness measurement standards such as Vickers and Rockwell tests, offering a comprehensive understanding of its effectiveness and precision.

Section III: Deep Learning in E-Health and Chronic Pain Assessment

Deep learning models have revolutionized sectors like computer vision, speech recognition, and more. With the rise of e-health data analysis, challenges presented by the multifaceted nature of well-being data come to the fore. Chronic pain, affecting a significant portion of the U.S. adult population, has considerable societal and economic repercussions.

The pressing need for an effective, low-cost, and easy-to-use Pain Meter is addressed through the utilization of classical physiological measurements, and the introduction of two new chronic pain datasets collected using prototype Pain Meters. This section presents a deep learning ordinal regression framework for chronic pain score assessment, pitting the innovative model against traditional machine learning methods and showcasing its advantages.

Section IV: Physiological Sensors and Machine Learning in Pain Assessment

Temperature sensors, motion sensors, force resistive sensors, and photoplethysmography (PPG) were engineered into a remote prototype Pain Meter and used to collect data remotely from chronic pain subjects. This section delves into the use of physiological sensors in the prototype Pain Meter, and correlates machine learning model predictions to self-reported pain levels to determine its potential applications in chronic pain assessment.

Section V: Phase Relationships in Heart Rate Variability

The resurgence of interest in phase relationships and the delay between heart rate modulation and breathing leads to the presentation of the Dynamic Phase Extraction (DPE) method. This study delves into its potential benefits compared to traditional methods of measuring heart rate variability (pulse rate variability).

Section VI: Biofeedback Devices and Chronic Pain Management

Chronic pain affects millions globally, with an enormous economic and social toll. This chapter examines the potential of portable biofeedback devices in managing chronic pain, focusing on their effectiveness in physiological self-regulation and relaxation, and their intersection with pain reprocessing therapy.

Section VII: Thermal Biofeedback for Sleep and Anxiety

The rise in global sleep disturbances and anxiety disorders necessitates a deeper examination of thermal biofeedback devices. This exploration evaluates the device's potential to improve sleep quality and reduce anxiety symptoms.

Summary

This dissertation, through a comprehensive integration of physiological technologies like the OsteoProbe, Pain Meters, and biofeedback devices, aims to make a significant contribution to medical science. It offers a holistic and innovative perspective on medical advancements, emphasizing their role and potential in improving chronic pain assessment and management, along with health areas relating to bone, anxiety and sleep. The collective findings and insights from this research will potentially lay a solid foundation for further advancements in e-health, paving the way for more effective, efficient, and patient-centered healthcare solutions in the future.

II. Significant correlation of Bone Material Strength index as measured by the OsteoProbe with Vickers and Rockwell hardness

The BMSi, Bone Material Strength index, as measured by the OsteoProbe, is significantly correlated with Vickers Hardness and Rockwell RW Hardness measurements on conventional materials. The Vickers and Rockwell measurements were done according to ASTM, American Society for Testing and Materials, Standard Test Methods, and OsteoProbe measurements followed published standardized testing methods. The correlations between BMSi and Rockwell, $r = 0.93$, and between BMSi and Vickers, $r = 0.94$, are comparable to the correlation between Rockwell and Vickers, $r = 0.87$. The correlation between BMSi and Rockwell is significant at $p < 0.01$ and the correlation between BMSi and Vickers is significant at $p < 0.01$. These results show that the indentation measurement performed by the OsteoProbe may be considered as a type of hardness measurement comparable to widely used conventional methods, with specific applications targeted by its portable and narrow design.

A. Introduction

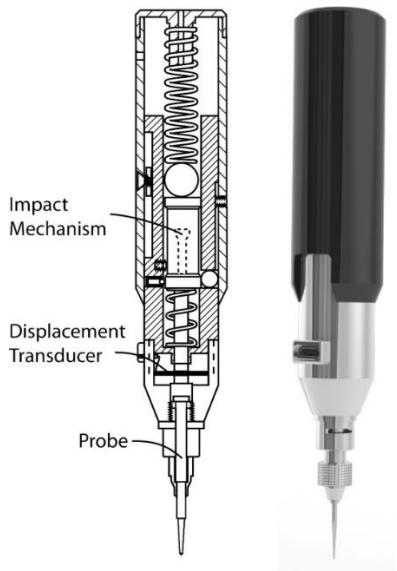


Figure 1. OsteoProbe image with a cross section view [1]. Reproduced from Bridges, D., Randall, C. and Hansma, P.K., 2012. A new device for performing reference point indentation without a reference probe. Review of Scientific Instruments, 83(4), p.044301., with the permission of AIP Publishing

The OsteoProbe [1, 2] is a microindentation device that has been used to quantify bone's resistance to indentation at the micro scale with a singular scalar value, named Bone Material Strength index, or BMSi. This value is calculated by taking the ratio of measured penetration distance into a test material (e.g. bone) and the measured penetration distance into a reference material, then multiplying the ratio by 100.

$$BMSi = 100 \times \left(\frac{penetration\ distance\ into\ reference\ material}{penetration\ distance\ into\ bone} \right)$$

The OsteoProbe has been used in multiple clinical studies of human patients and a standard testing procedure has been published previously [3]. One study found that BMSi is lower in patients with Type II diabetes [4, 5], while another showed that BMSi decreased when patients were treated with glucocorticoids and that this decrease could be eliminated or

reversed with drug therapy [6]. A third work examined a decrease of BMSi in patients with fragility fractures independent of bone mineral density [7].

Some controversy has arisen about what is being measured [8-10]. Though correlations exist for some cases [11, 12], a simple universal relationship between BMSi and bulk material mechanical properties of bone that holds for all bone diseases has not been found. Thus, we cannot appeal to comparison with bulk material mechanical properties to answer the question of what is being measured. Instead, OsteoProbe measurements should be compared with standard mechanical tests for measuring surface mechanical properties. In this study, OsteoProbe measurements are compared to Vickers and Rockwell tests, two standard mechanical tests for measuring hardness – a surface mechanical property. Clinical measurement of BMSi in a patient can range from the low 40s up to the low 100s. The polymers we selected had BMSi values that represent this range, with the softest scoring 44 and the hardest scoring 104. This range of BMSI corresponds to indentation depths of 350μm down to 125μm both in the patients and in the polymers [4-7].

B. Methods

To demonstrate that the OsteoProbe's BMSi is a measure of hardness, measurements performed by the OsteoProbe were compared with conventional indentation-based measurements of hardness, namely measurements collected through Rockwell and Vickers type testing. Using a scanning electron microscope, a micro scale image of an indented polymer by the OsteoProbe can be seen in Figure 2. Conventional engineering materials were used for the comparison testing, namely high-performance polymers. Fourteen different polymers that span the observed range of OsteoProbe measurements in patients were selected and indented five times by each instrument. The fourteen polymers include,

uhmw polyethylene (ultra-high molecular weight polyethylene), hdpe (high density polyethylene), polystyrene, abs (acrylonitrile butadiene styrene), noryl ppo, cast nylon, polyester, extruded nylon, rexolite, delrin, ultem pei (polyetherimide), peek, torlon pai (polyamide-imide), and PMMA (Poly(methyl methacrylate)).

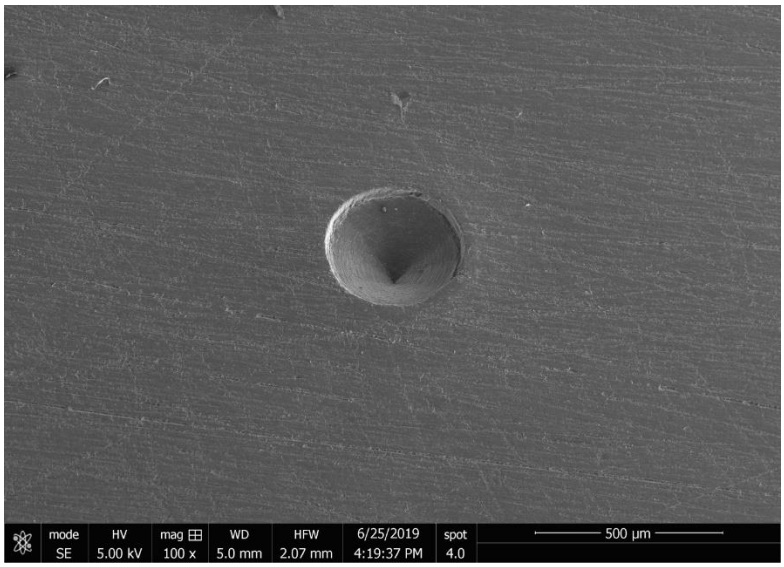


Figure 2. Image of an OsteoProbe indentation for measurements.

Indentation methods chosen for testing were Rockwell Type RW, Vickers, and BMSi using the OsteoProbe. The Rockwell hardness was measured according to ASTM D785-08 (Reapproved 2015) Standard Test Method using a United Tru-Blue II hardness testing instrument. Test parameters used a loading force of 60kg and indenter geometry of 1/2 inch ball diameter.

The Vickers hardness was measured according to ASTM E384 Standard Test Method using a United Tru-Blue II hardness testing instrument. Test parameters used a loading force of 10kg and a 136-degree high-load Vickers diamond indenter. The Vickers indentations were measured with a 200x and 500x microscope. The Vickers hardness number was calculated from these measurements according to ASTM E384. The OsteoProbe BMSi was

measured according to a published, standard testing procedure with 10 indentations for each polymer sample [3]. Every indentation applies a preload force of 10N at which point a trigger system releases a mechanism that applies an additional 30N force at high speed, making a microscopic indentation. The penetration distance into the material is automatically measured by the OsteoProbe.

C. Results

Hardness measurements were performed on fourteen selected polymers using Rockwell, Vickers, and OsteoProbe methods, with results seen in Figures 3-5. OsteoProbe measurements of the selected polymers, as given by values of BMSi, were found to be comparable to the range of BMSi values measured for the bones of patients [4-7].

Results showed correlative trends among the measurement methods. As can be seen in Figures 3 and 4, there are statistically significant correlations between BMSi values and Vickers, with $r = 0.94$ and $p < 0.01$; and between BMSi values and Rockwell values, with $r = 0.94$ and $p < 0.01$. Scatter seen in the results may be expected given the different mechanical properties of the fourteen different polymers. Figure 5 shows comparable scatter even between the two conventional measures of hardness, Rockwell and Vickers.

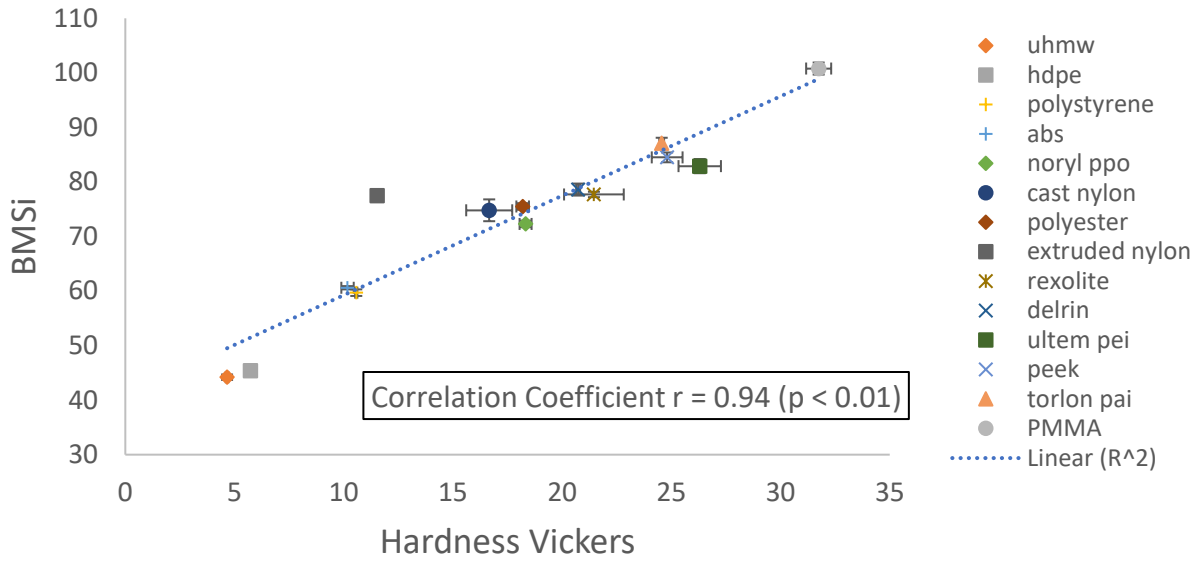


Figure 3. Correlation of BMSi and Vickers Hardness.

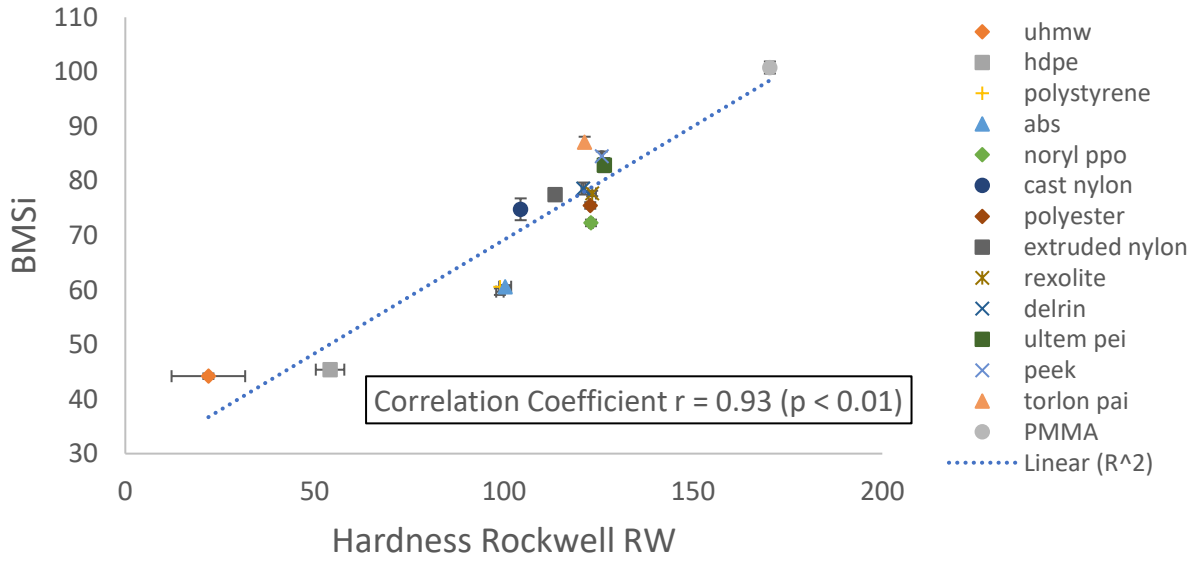


Figure 4. Correlation of BMSi and Rockwell RW Hardness.

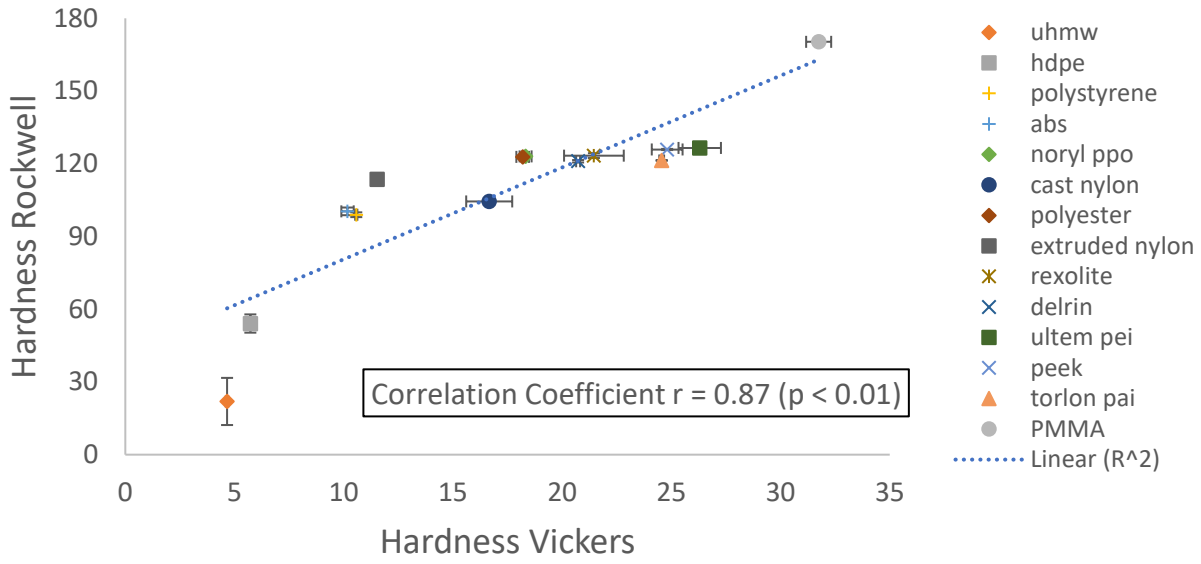


Figure 5. Correlation of Rockwell RW Hardness and Vickers Hardness, with a correlation coefficient that is slightly lower compared to BMSi and Vickers Hardness, as well as between BMSi and Rockwell RW Hardness.

Hardness Data 1	Hardness Data 2	Correlation Coefficient	p-value
BMSi	Vickers	0.94	< 0.01
BMSi	Rockwell	0.93	< 0.01
Rockwell	Vickers	0.87	< 0.01

Table 1. Correlation coefficients and p values for hardness data from BMSi, Vickers and Rockwell measurements.

	Rockwell		**BMSi**		**Vickers**	
N=5	**Mean**	**St. Dev**	**Mean**	**St. Dev**	**Mean**	**St. Dev**
PMMA	74	1.6	100.1	2.2	22.67	0.57
noryl ppo	62	1.1	73.2	0.5	19.91	0.27
delrin	77	1.8	77.0	1.5	21.95	1.23
CPVC	70	0.7	88.5	3.7	19.26	0.18
cast nylon	65	0.7	66.8	0.5	16.66	1.05
POLYPROPYLENE	59	0.7	65.7	1.4	11.61	0.22
abs	41	1.3	61.4	0.2	13.42	0.55

Table 2. Mean and standard deviations for hardness data from BMSi, Vickers and Rockwell measurements.

Using the Table of Critical Values: Pearson Correlation, the calculated correlation between BMSi and Rockwell was found significant at $p < 0.01$ and the correlation between BMSi and Vickers was significant at $p < 0.01$. These tests all evaluated the hardness of the materials within a few hundred microns of the surface [14]. All three test methods were comparably and significantly correlated: between BMSi and Rockwell, $r = 0.93$ at $p < 0.01$, between BMSi and Vickers, $r = 0.94$ at $p < 0.01$, and between Rockwell and Vickers, $r = 0.87$ at $p < 0.01$. The BMSi measurements were slightly more correlated with Rockwell and Vickers hardness measurements than the two standard testing methods were to each other.

D. Summary

The simple answer to the question: "what does the OsteoProbe measure?" is that it measures a normalized indentation distance, BMSi, which is a measure of the hardness on the surface of the bone. It is similar to other instrumented hardness tests in that an indentation is made according to a standard procedure and the resulting indentation depth is measured: the smaller the indentation depth, the larger the hardness.

Indentation has classically been utilized as a method of material characterization with measured results often reported in terms of hardness [15]. The classic Vickers and Rockwell hardness methods use pyramid, cone or spherical indenter shapes to create permanent, localized deformation on the surface of a part under a given load [15]. Measurement of the deformed geometry gives the hardness. Softer materials are more deformed. Hard materials are less deformed. The OsteoProbe uses a conical indenter shape to create permanent, localized deformation that is monitored by the instrument itself. Specifically, strain gauges in the OsteoProbe monitor the distance that the conical indenter goes into the material under test. Thus, the OsteoProbe can be considered an indentation device and its results can be

compared to conventional devices such as Rockwell and Vickers. The comparison reported in this manuscript revealed strong correlations: correlation coefficients above 0.9, $r > 0.9$, between BMSi and Rockwell and Vickers Hardness instruments.

Why is it important to measure the hardness on the surface of the bone? The answer to this question is that the hardness on the surface of bone, as quantified by BMSi, has been shown clinically to be lower in patients with bone degenerating diseases such as Type II diabetes [4, 5] and bone degenerating treatments such as glucocorticoid treatment [6]. Patients with fragility fractures have decreased BMSi [7].

BMSi, as measured with the OsteoProbe, is significantly correlated with both Rockwell and Vickers hardness on fourteen engineering polymers: between BMSi and Rockwell, $r = 0.93$ at $p < 0.01$, between BMSi and Vickers, $r = 0.94$ at $p < 0.01$, and between Rockwell and Vickers, $r = 0.87$ at $p < 0.01$. Results support that the OsteoProbe indentation measurement, BMSi, is a valid measurement of hardness. The value of this measurement of hardness is that it can be done in a clinic on patients without surgically exposing bone. Decreased hardness, as quantified by BMSi, has been clinically demonstrated to be correlated with various causes of bone degeneration [4-7].

III. How Much Does It Hurt: A Deep Learning Framework for Chronic Pain Score Assessment

Chronic pain is defined as pain that lasts or recurs for more than 3 to 6 months, often long after the injury or illness that initially caused the pain has healed. The "gold standard" for chronic pain assessment remains self report and clinical assessment via a biopsychosocial interview, since there has been no device that can measure it. A device to measure pain would be useful not only for clinical assessment, but potentially also as a biofeedback device leading to pain reduction. In this paper we propose an end-to-end deep learning framework for chronic pain score assessment. Our deep learning framework splits the long time-course data samples into shorter sequences, and uses Consensus Prediction to classify the results. We evaluate the performance of our framework on two chronic pain score datasets collected from two iterations of prototype Pain Meters that we have developed to help chronic pain subjects better understand their health condition.

A. Introduction

Deep learning models have achieved remarkable success in computer vision [16], natural language processing [17], speech recognition [18] and the game of Go [19]. Recently there has been increasing interest in applying deep learning for end-to-end health data analysis [20]. However, e-health data analysis is even more challenging since well-being data can be affected by many factors, for example individual differences, measurement errors in data collection and missing data.

Chronic pain is described as persistent or recurrent pain that lasts for at least 3 to 6 months [21]. According to the 2016 National Health Interview Survey (NHIS), roughly

20.4% (50.0 million) of U.S. adults suffer from chronic pain. Chronic pain affects individuals, their families, and society, and results in complications harming both physical and mental health. The economic costs of chronic pain and pain-related disability in the United States cannot be overstated. One influential report [22] conservatively estimated an annual toll of 560−650 billion dollars-far exceeding the costs of cardiovascular disease, cancer, and diabetes. Therefore, identifying the score of chronic pain is of significant value to reduce further complications.

Neurophysiological signals have been used to quantify pain [23–25]. In the last decade, there has been some progress towards discovering the neurobiological substrates of pain. However, none of those methods is low-cost or easy-to-use. Clearly, there is an unmet need for a low-cost, easy-to-use Pain Meter. According to recent research results [26–28], classical physiological measurements can effectively quantify pain. Inexpensive technology is now available to measure relevant physiological features. Taken together, it is clear that objectively quantifying pain is possible. We thus investigated a number of inexpensive commercial sensors capable of measuring physiological variables for pain score assessment. We have thus far built prototype Pain Meters that offer the immense potential to revolutionize pain treatment and the development of therapeutics.

To this end, we collected two new chronic pain datasets, using prototype Pain Meter 1 and 2, respectively. For Dataset 1, our subject has been suffering chronic pain for more than 10 years. Neck and shoulder pain were causing her difficulty to perform daily activities. For Dataset 2, we recruited chronic pain subjects from our local community. All subjects signed an informed consent form according to a protocol approved by a Human Subjects Committee. Our chronic pain datasets are characterized by the following unique properties:

First, we use Photoplethysmography (PPG) [29], which is low cost and comfortable for patients, to collect pulse signals. Secondly, we also include several temperature signals and Galvanic Skin Response(GSR) signals to detect the chronic pain symptoms like nervousness. For Dataset 2, we also use accelerometers and gyros to detect movements.

Given the pain score datasets, we introduce the task of chronic pain score prediction, which is to train an end-to-end oridinal classifier to accurately predict pain score. The illustration of our workflow is shown in Fig. 6. An accurate chronic pain score assessment will

1. Facilitate the development of new therapies both in the laboratory and in Phase II and Phase III clinical trials.
2. Make it possible for physicians to quantify the effects of existing therapies on individual patients.
3. Minimize the harm caused by diagnostic delays and the under/overtreatment of pain due to the influence of gender, race, or age.
4. Allow a patient to decide, with objective personal data, whether current treatments are effective in his/her quest to reduce chronic pain.
5. Serve as a biofeedback device for chronic pain subjects. The unconscious mind can learn to control things it can monitor [30]. If people are enabled to accurately monitor their chronic pain score with a Pain Meter, their unconscious mind can figure out how to decrease the chronic pain score. It is trained and rewarded by the tiny decreases in pain score that are accurately and continuously monitored.

Our main contributions are threefold:

1. We propose a deep learning ordinal regression framework for chronic pain score assessment. To the best of our knowledge, this is the first paper to use deep learning for chronic pain score assessment.
2. We collect two new chronic pain datasets using our prototype Pain Meters to predict the score of chronic pain.
3. We split the long recordings into smaller slices for training, which not only eases the burden on GPU memory but also provides many training samples for deep learning models. We define Consensus Prediction as the majority voting result of the sampled short slices for testing, since not all of the short slices can be expected to contain enough useful information.

Deep Learning for E-Heath

Since the emergence of deep learning in e-health, more and more researchers have been implementing deep learning models for medicine, aiming to improve health care [20], classify diseases [31], and prevent misdiagnosis [32]. More specifically, the prediction of medical events has been popular, including the prediction of death rates [33], prescriptions [34], and successful extubation [35]. Although many researchers have applied deep learning to e-health, to the best of our knowledge no researcher has applied deep learning from pulse signals to the prediction of chronic pain scores [36]. There are image based models for automatic estimation of pain [37], [38]. Deep Pain [37] uses long short term-memory (LSTM) and analyzes images of facial expression. However, images alone are not reliable since images alone may not contain enough relevant information for pain score prediction. Many factors affect pain score, including stress and mood, which might not be fully captured by image based methods. Physio-based models have been developed for

assessments of physiological pain [39] but these have not been applied to chronic pain. Motivated by this need, we propose a Convolutional Neural Network (CNN)-based framework that uses physiological signals to assess chronic pain scores.

Chronic Pain and Traditional Machine Learning

In the context of pain assessment research, physiologically-based pain has been the main focus for many pain researchers [40]. Chronic pain, on the other hand, is prolonged, lasting anywhere from months to years [41]. In the past, traditional machine learning methods have been applied to e-health, but this has required extensive feature extraction [42]. Human feature extraction has many disadvantages. For example, it is costly and can result in the "loss of data interpretability" [43]. Random Forest has been used to monitor the nociception (perception of pain) level, which requires feature extraction [44]. Physiological parameters like heart rate, heart rate variability, plethysmograph wave amplitude, skin conductance level, number of skin conductance fluctuations, and their time derivatives are extracted for prediction. In contrast, our end-to-end deep learning framework requires no feature extraction and little data preprocessing.

Physiological Sensors

Photoplethysmography (PPG) is a widely used non-invasive method for measuring pain intensity, in which changes in blood volume and light absorption are detected [29]. This type of technology is commonly used because it is simple and can easily collect data, while being cost-efficient and accessible [45]. A PPG device has also been developed to monitor respiratory and heart rates of infants, and this has been proven to perform well [46]. PPG is favored not only for clinical use, but also for home use as a biofeedback device. In addition,

it is more comfortable for patients because it does not involve gel and electrodes contacting their skin [47]. Thus, in our prototype Pain Meter, we use PPG [48] to collect pulse recordings. We also included axis accelerometer gyroscope modules (MPU-6050), force sensitive resistors to measure the forces that cause low frequency motion (DF9-40), and a GSR sensor (ZIYUN Grove GSR sensor).

B. Methods

Prototype Pain Meter

Fig. 7 shows the device we used for collecting Dataset 1. Pain Meter 1 contains: 1) two PPG pulse sensors held to the temples via a headband, three at the three arteries supplying blood to the brain mounted in a neck pillow, and two at the fingertip and palm, 2) temperature sensors at each location of the PPG pulse sensors, and 3) GSR sensors embedded in the block on which the hand rests. These provide a total of 15 signals, recorded in Dataset 1.

As is shown in Fig. 8, it turned out that the PPG sensors were sensing more than pulse. They were also sensing subtle motion. Motivated by these phenomena, we added actual motion sensors in our Pain Meter 2. As is shown in Fig. 9, Pain Meter 2 contains: 1) PPG pulse sensors in a headband, in a neck band, and on the fingertip, 2) temperature sensors on the neck and fingertip, 3) 3-axis accelerometers and 3-axis gyros in the head band and wrist band, 4) force sensors on the forehead, back of neck, side of neck, and wrist band, and 5) GSR sensors between the middle and ring fingers. These provide a total of 25 signals, recorded in Dataset 2.

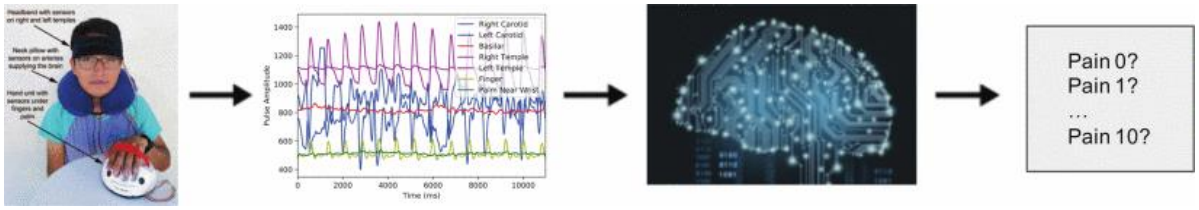


Figure 6. The workflow for chronic pain score assessment. The brain image is from Google.

For both Pain Meters, a Teensy 3.6 microcontroller with a 32-bit 180 MHz ARM Cortex-M4 processor is used to sample all signals every 15ms. A data acquisition program was designed in MegunoLink Pro for Pain Meter 1 data and a customized data acquisition program was written in Python for Pain Meter 2 data.

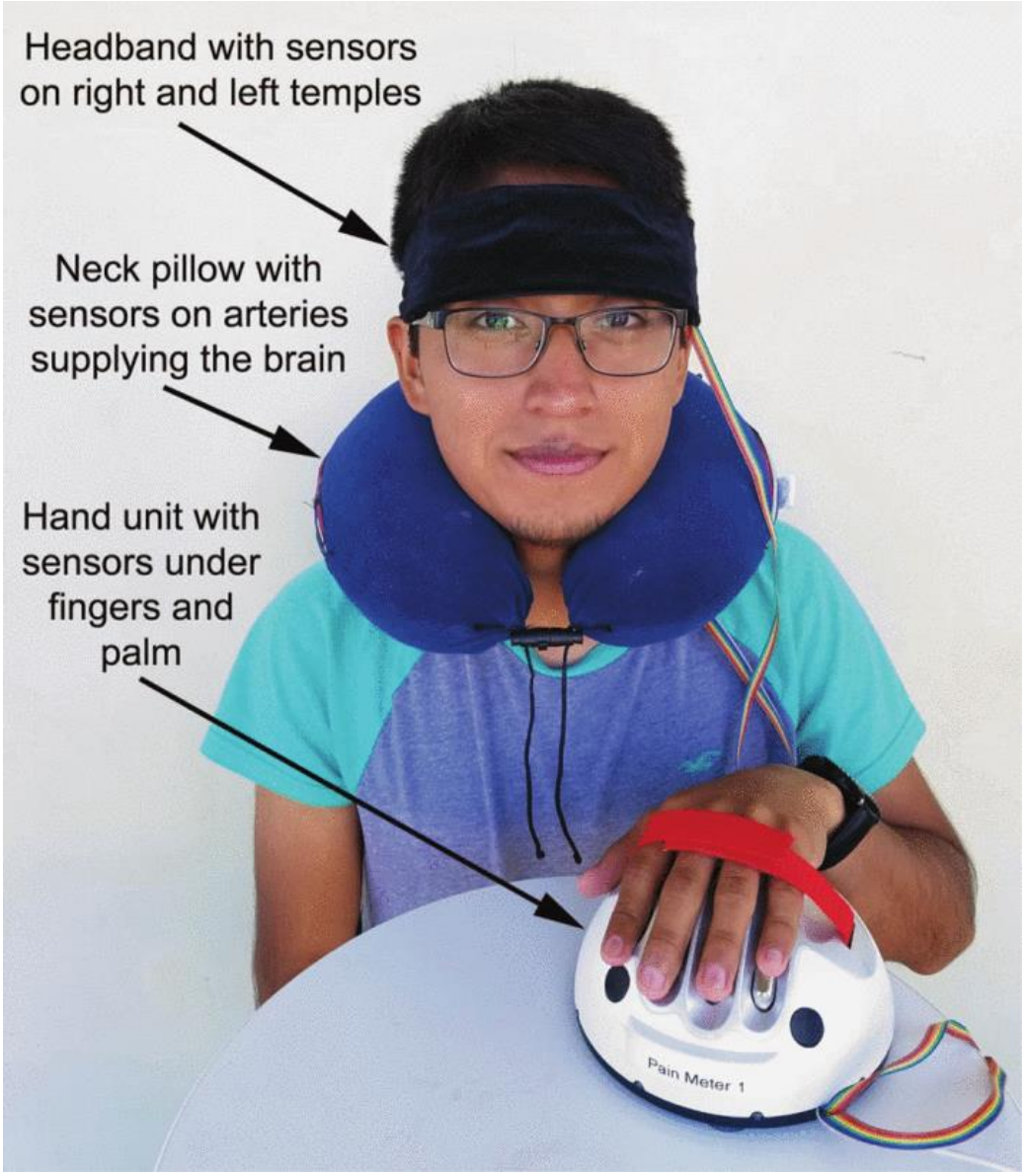


Figure 7. Pain meter 1 sensed temperature, pulse and GSR, but it did not directly sense motion.

Data Collection

Using Pain Meter 1 from Fig. 7, our chronic pain Dataset 1 was collected from one subject who self-reported the pain score. The subject was asked to fix the headband and neck pillow into a comfortable position. Once secured, real-time plots of the pulse and temperature data were viewed to verify that the head and neck sensors were collecting

accurate and reliable signals for the subject. Minor adjustments to the sensor positions can be made if needed. After adjustments were made, the subject placed the left hand on the module to read the hand signals. A final verification step viewing all plots of data was performed. Then the subject was asked to close their eyes and relax before the recording begins. After 10 minutes, the recording was ended. Each recording was taken on a different day at the same time in the afternoon, the same temperature and the same environment brightness. During recordings with Pain Meter 1 we noticed that some of the pulse signals became erratic compared to other pulse signals and that this erratic behavior seemed to correlate with pain. We hypothesized that this was due to subtle movement [49] and constructed Pain Meter 2 (Fig. 9) to have motion sensors. With the addition of motion sensors we could confirm this hypothesis (Fig. 8) and measure the motion directly.

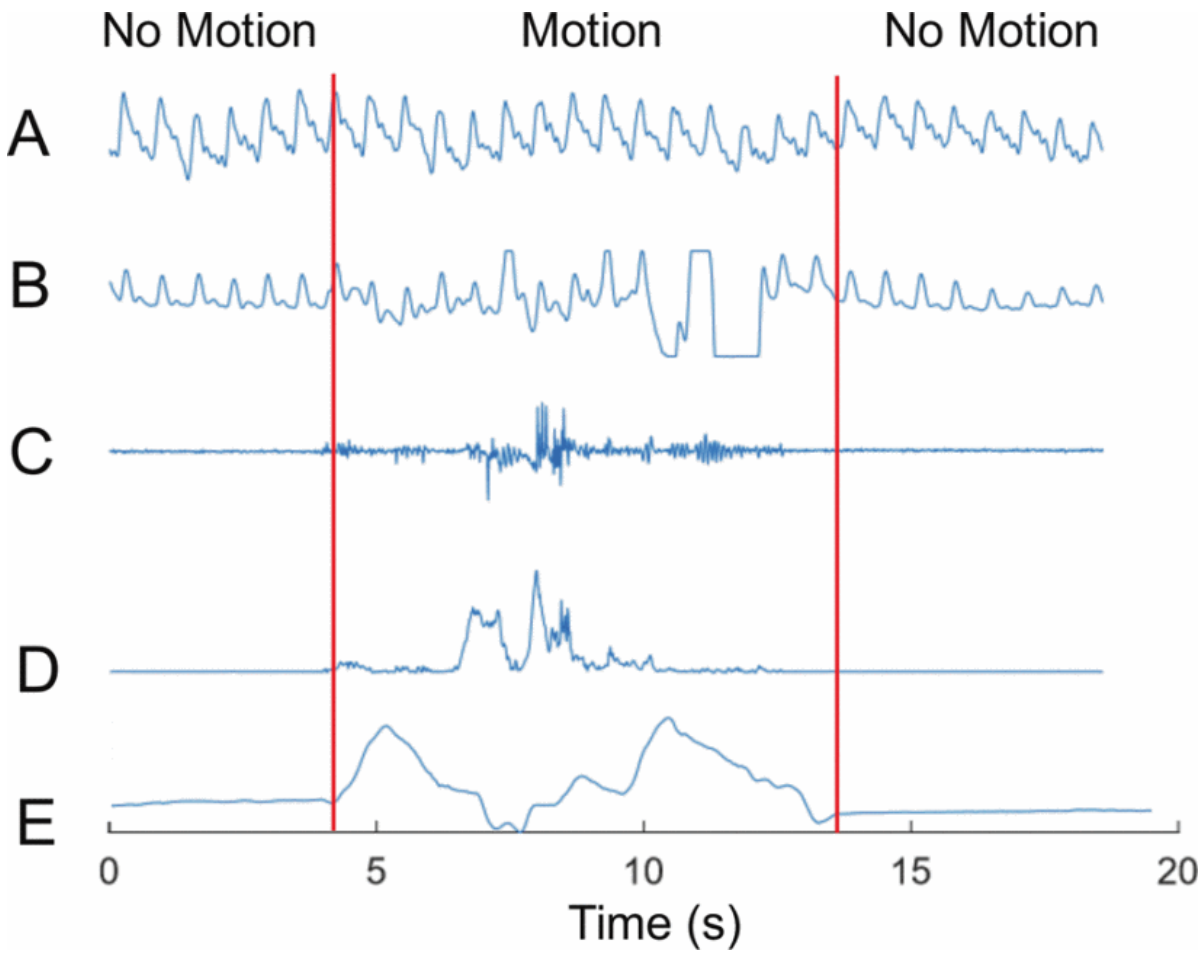


Figure 8. Motion affects the pulse sensor signals, which are based on measuring the intensity of light reflected from the skin. Part A shows a pulse sensor at the carotid artery, far from motion. Part B shows a pulse sensor at the finger. Motion is generated by flexing the wrist and is measured in part C by a wrist acceleration sensor. Part D shows data from wrist gyro sensor and part E shows data from a wrist force sensor. Note that the pulse sensor on the finger (sensor b), which is closer to the motion, is strongly affected by the motion, while the pulse sensor on the carotid artery (sensor a) is not.

Figure 9. Pain meter 2 consists of: 1) PPG pulse sensors in a headband, in a neck band, and on the fingertip, 2) Temperature sensors on the neck and fingertip, 3) 3-axis accelerometers and 3-axis gyros in the head band and wrist band, 4) Force sensors on the forehead, back of neck, side of neck, and wrist band, and 5) GSR sensors between the middle and ring fingers.

Data	Classes	Pain score distribution
Chronic Pain	2	1 : 1

Table 3: Data statistics for dataset 1

Pain score distributions for the 2-class Dataset 1 and 7-class Dataset 2 are shown in Table 3 and Fig 10, respectively. Each recording has a length of 10 minutes, with signals sampled every 15 milliseconds. We have 4 recordings from one subject in Dataset 1 and 62 recordings from 20 subjects in Dataset 2. We divide each 10-minute recording into ten mutually exclusive 1-minute samples.

Classification Problem

The classification problem for chronic pain assessment is to classify a Pain Meter dataset of time sequences into pain scores on a scale of 0 for no pain to 10 for the worst pain possible. The model is first trained on Dataset 1 from a chronic pain subject who self reports her pain score for each of the datasets on a 0 to 10 scale. Note that seldom has our subject reported a pain score higher than 2 in Dataset 1. Thus in Dataset 1, we use pain scores 1 and 2. In Dataset 2, we have pain scores from 0 to 6, which formulates to be a 7 class classification problem. Our goal is to use the trained deep learning model to accurately classify chronic pain datasets into the self reported pain scores.

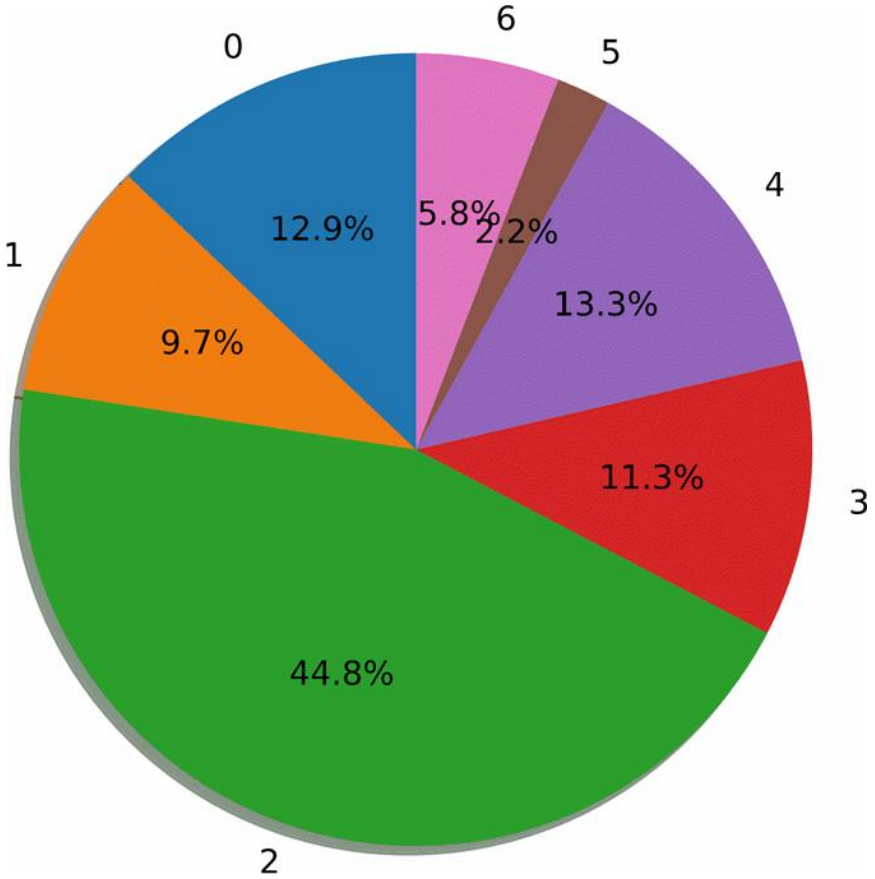


Figure 10. Distribution of pain scores in dataset 2. The numbers outside the pie chart are the corresponding pain scores and the percentages of time for which each pain score is reported are shown in the pie chart.

Methodology

The model architecture, shown in Fig. 10, consists of convolution-pooling layers followed by fully connected layers. To learn temporal and correlation features, the convolution is performed on both the time and sensor dimensions. We split the 1-minute samples into smaller slices with length of seq_length. Sensor recording signals with a dimension of (number of sensors (N) × seq_length) serve as input x for the neural network. A convolution operation involves a filter $w \in R^{st}$, which is applied to a window of s sensors and t samples to produce a new feature. For example, a feature $f_{i,j,}$ ($0 \leq I \leq -N - s + 1, 0 \leq j \leq seq_length - t + 1$) is generated from a window size (s,t) of the sensor signals:

$$f_{i,j} = \mathrm{ReLU}\left(\omega x_{i:i+s-1,j:j+t-1} + b\right), \tag{1}$$

where $b \in R$ is a bias term. This filter is applied to each possible window of the voltage signals to produce a feature map:

$$f = \begin{bmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,seq_length-t+1} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,seq_length-t+1} \\ \cdots & \cdots & \cdots & \cdots \\ f_{N-s+1,1} & f_{N-s+1,2} & \cdots & f_{N-s+1,seq_length-t+1} \end{bmatrix}, \tag{2}$$

with $f \in R^{N-s+1,seq_length-t+1}$. We then apply a max-pooling operation over the feature map and take the maximum value m = max(f) as the feature corresponding to this particular filter. The idea is to capture the most important feature, the one with the highest value, for each feature map. Our model uses multiple filters to obtain multiple features. These features form the penultimate layer and are passed to a fully connected softmax layer whose output is the probability distribution over different pain scores. We adjust the number of convolutional

ReLU layers from 2 to 5, based on the choice of seq_length. The prediction of our model, parameterized by **W**, is given by $p(\hat{y}|x, \boldsymbol{W})$.

Due to the existence of the inherent ordering information in our chronic pain score, we apply ordinal regression, a setting that bridges metric regression and classification, to predict chronic pain scores of ordinal scale. Compared to regular regression problems, these pain scores are discrete. These pain scores are also different from the labels of multiple classes in classification problems due to the existence of the ordering information. The cross entropy loss of our ordinal regression for input vector x_n is as follows,

$$\begin{aligned} L_n(\boldsymbol{W}) =& (1 + |\frac{argmax(\boldsymbol{p}(\hat{\boldsymbol{y}}|\boldsymbol{x}_n, \boldsymbol{W})) - \boldsymbol{y}}{C-1}|) \\ & \times \frac{1}{C}(\sum_{i=1}^{C} -\boldsymbol{y}_i \cdot log(\boldsymbol{p}(\hat{\boldsymbol{y}}|\boldsymbol{x}_n, \boldsymbol{W}))), \end{aligned} \tag{3}$$

where y represents the true pain scores, $p(\hat{y})$ denotes the predicted probability vector with one value for each possible pain score, and C is number of pain scores. Divided by C−1, the absolute error $|\,argmax(p(\hat{y}|x, \boldsymbol{W})) - y|$ is normalized between 0 and 1, with C classes in total. By multiplying $|\frac{argmax(p(\hat{y}|x,\boldsymbol{W}))-y}{C-1}|$, the normalized absolute error between prediction and ground truth, with cross entropy loss, we include the ordinal information in our loss function. We penalize more in our loss function if the absolute error between ground truth and prediction is larger.

For testing, we define Consensus Prediction to measure the performance of predictions for the whole sensor signal sample. Consensus Prediction synthesizes results from multiple short slices by majority voting, which can significantly improve the prediction accuracy for

a long sample. This is because not all of the short time slices can be expected to contain useful information for classification.

We use Batch Normalization [50] to accelerate training. For regularization, dropout [51] and early stopping methods [52] are implemented to avoid overfitting. Dropout prevents coadaptation of hidden units by randomly dropping out a proportion of the hidden units during backpropagation. Model training is ended when no improvement is seen during the last 100 validations. Softmax cross entropy loss is minimized with the Adam optimizer [53] for training. Since both frequency information and correlation between sensors are captured through convolution filters, our CNN-based framework automatically deals with the features needed for classification. We use grid search for hyperparameter tuning. The hyperparameters are described in Table 4.

Experimental Setup

We introduce two baselines: Multilayer Perceptron (MLP) and Logistic Regression, to compare with our proposed CNN framework for the two classification problems.

CNN Based Model and Consensus Prediction

We implement a parallel processing framework that distributes the convolutional neural network into multiple (N) GPUs to ease the burden on GPU memory. Each GPU contains an entire copy of the deep learning model. We first split the training batch evenly into N sub-batches. Each GPU processes only one of the sub-batches. Then we collect gradients from each replicate of the deep learning model, aggregate them together and update all the replicates. We train our CNN based framework with two NVIDIA GeForce GTX 1080s, each of which has a memory of 11178 MB.

Multilayer Perceptron

We use a four layer fully connected network, whose output is the probability distribution over two different classes, as a baseline. We use a parallel processing framework similar to our CNN model implementation. We train our MLP on two NVIDIA FeForce GTX 1080s.

Logistic Regression

Since the signals are periodic, we extract individual voltage signal FFT (x_1) and pairwise sensor signal Pearson Correlation (x_2) as features (X) for a Logistic Regression classifier:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(w_0 + w_1 x_1 + w_2 x_2)}}, \tag{4}$$

where Y is the label for classification and P is the probability of predicting Y as label 1 (pain score 1). w_0, w_1 and w_2 are model parameters to be learned during training.

Split Training and Testing

Since each of our recordings is long enough to split into multiple informative samples and there exists different settings among different recordings, we implement two methods of splitting the training and testing data. We first divide each 10-minute recording into ten mutually exclusive 1-minute samples.

1. Considering the size of our datasets in prediction, we use 5-fold cross validation. For each of the 10-minute recordings, we use between $2*(i-1)$ and $2*i$ minutes as testing, and the rest for training, for the i-th fold. This training/testing split fits well with the scenario of practical use of pain score assessment. It is known that different subjects

have different perceptions of pain. For real use, our pain meters need some self-calibration before getting accurate pain score readouts.

2. Leave-one-recording-out cross validation on all the recordings. Each recording is used once as a test set while the remaining recordings form the training set. We note that, due to differences in subjects' sensitivity to pain, different settings while recordings, and the size of our data, it is much more difficult to predict across subjects than within one recording. We report the result using this splitting method in Section VIII.

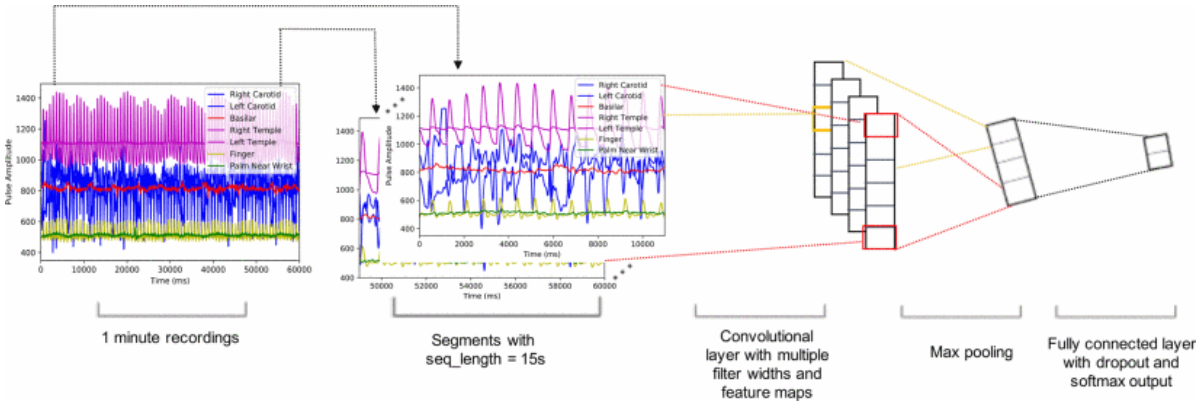


Figure 11. Model architecture: 1-minute samples of the voltages measured on the physiological sensors are collected from our prototype pain meter. Segments with seq_length = 15 s of these samples are individually classified. These individual classifications are conducted for consensus prediction in pain score recordings.

Hyperparameters	Value
Batch size	24
Epoch	2000
Dropout rate	0.5
Seq_length	15 seconds
Learning rate	0.5

Table 4: Hyperparameters

C. Results

Considering the size of our dataset in prediction, we use 5-fold cross validation and report the average results in this section. Note that Dataset 2 is unbalanced. We also report the confusion matrix for evaluation.

Confusion matrices are shown in Fig. 12 and Fig. 13 for chronic pain score prediction. Dominant numbers on the confusion matrix diagonal indicates that our model achieves high accuracy for each class. Fig. 14 shows the individual class prediction distribution for Dataset 2. The infrequent prediction mistakes scatter around the ground truth. For example, although 11% of the predictions from pain score 3 are incorrect, they are still close to 3 (either 2 or 4). Fig. 15 further demonstrates that the probability of making an error decreases as the absolute prediction error increases. This is the benefit from ordinal regression, which penalizes more in the loss if the absolute error between ground truth and prediction is larger. Fig. 16 is a scatter plot of expected predicted pain score vs. self-reported pain score. The relationship between predicted pain score and ground truth is highly linear, with an R-squared (R^2) of 0.9463.

Since we use seq_length to split long recordings into shorter slices, some of the short slices may not contain enough information for pain score prediction. However, these effects can be eliminated using Consensus Prediction. Although we have imblanced data in Dataset 2, we still use accuracy to compare different models and check the benefits obtained from Consensus Prediction, since chronic pain subjects are most interested in prediction accuracy.

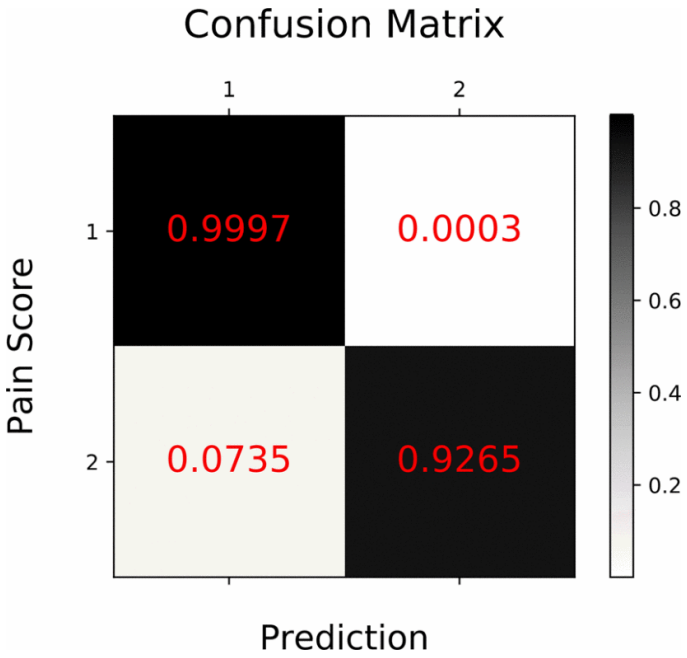


Figure 12. Confusion matrix for chronic pain score prediction in dataset 1.

Results of our framework compared against other machine learning models on chronic pain recordings are shown in Tables 5 and 6 respectively. We compare our CNN based model with two baselines: Multilayer Perceptron (MLP) and feature-based Logistic Regression. For Dataset 1, our CNN-based deep learning approach improves the prediction accuracy by 5.25% compared to feature-based Logistic Regression. Fig. 17 shows the Consensus Prediction accuracy. The accuracy improves by 3.84% using Consensus Prediction. Although not all of the short slices can be expected to contain enough useful recording patterns, we can overcome that when we synthesize multiple individual classification results from these short slices. For Dataset 2, our model achieves accuracy of 95.23% for short recording slices, which is a 2.8% improvement over feature based Logistic Regression. The accuracy further improves to 98.30% with Consensus Prediction as shown in Fig. 18. Our CNN based deep learning model also outperforms MLP on both of the two sets of recordings by 3.32% and 2.91% respectively, which shows CNN's advantage of local

feature extraction using convolutional kernels over MLP. Also from Fig. 17 and Fig. 18, 100 time slices for each recording are sufficient in Consensus Prediction.

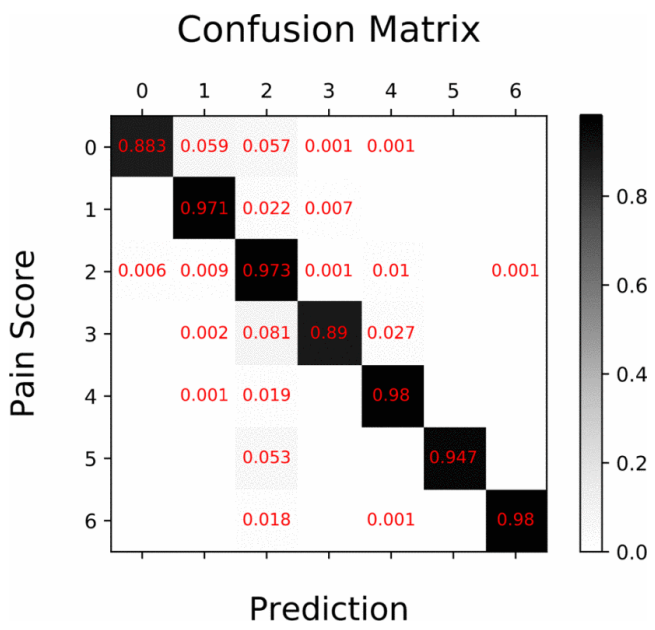


Figure 13. Confusion matrix for chronic pain score prediction in dataset 2. Our model achieves high accuracy for each class.

Model	Accuracy on Testing
Convolutional Neural Network	**0.9630**
Multilayer Perceptron	0.9321
Logistic Regression	0.9150

Table 5: Cross-validation performance comparison of our deep learning model with multilayer perceptrons and logistic regression on dataset 1.

Fig. 19 shows the trend of accuracy versus the choice of seq_length for Dataset 1. For the effect of seq_length on accuracy, there exists a trade-off between number of training samples and representation of a whole recording. The short slices contain less information but can provide more independent training samples. For deep learning models, larger numbers of training slices help more than a larger sample.

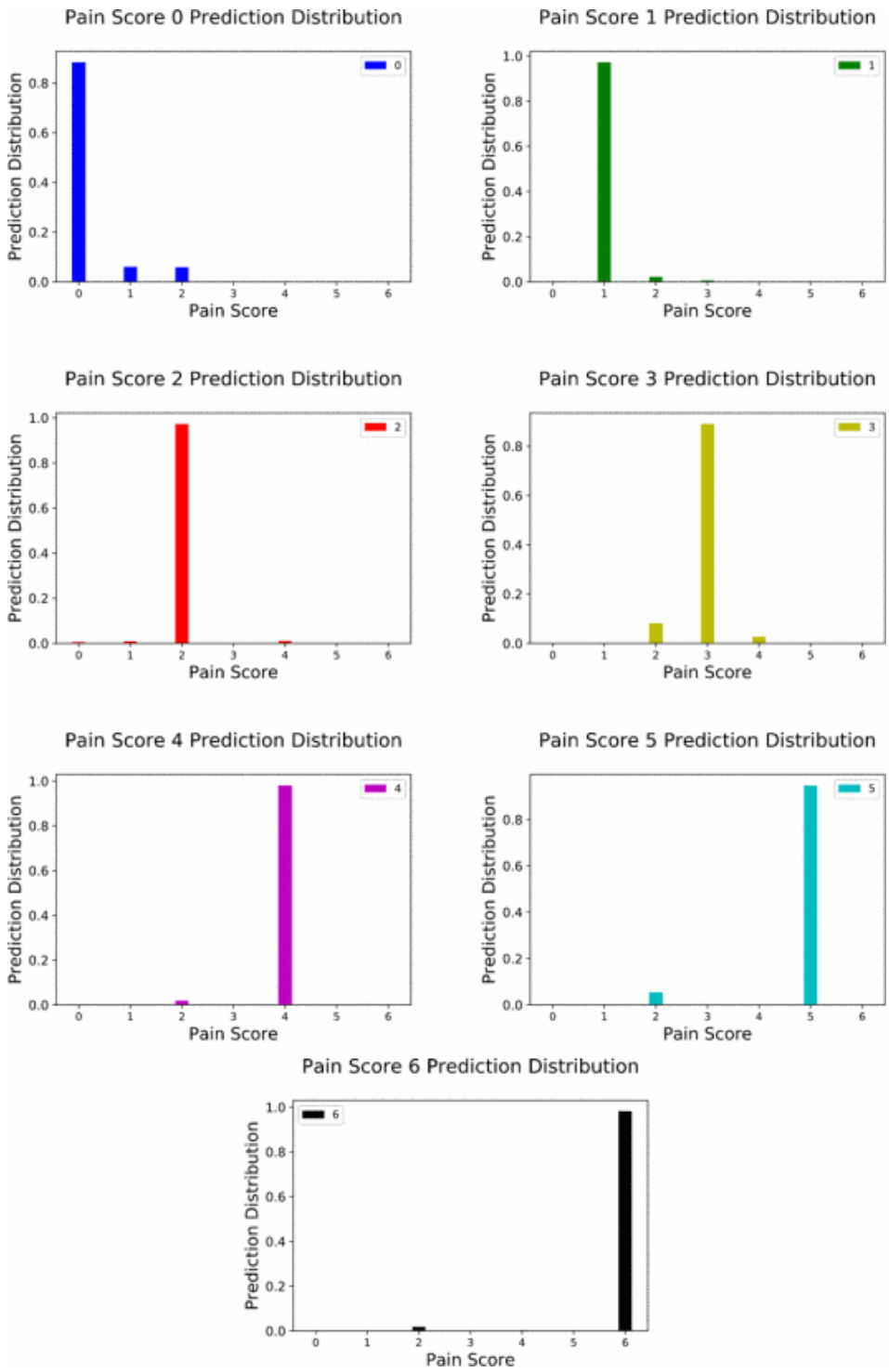


Figure 14. Individual class prediction distribution for dataset 2. The occasional incorrect predictions scatter close to the ground truth.

Model	Accuracy on Testing
Convolutional Neural Network	**0.9523**
Multilayer Perceptron	0.9238
Logistic Regression	0.9063

Table 6: Cross-validation performance comparison of our deep learning model with multilayer perceptrons and logistic regression on dataset 2. Although we have imblanced data for 7 classes, subjects are still most interested in prediction accuracy. We showed confusion matrix for our CNN model in the previous figure.

However, we still cannot choose too small of a seq_length, since a too short slice is not representative for a recording. Given the data we currently have, we use a seq_length of 15 seconds.

We show some failure cases for our CNN classifiers in Fig. 20. The upper figure shows the histogram of short sequence prediction for one recording, where blue indicates predicted successfully and red indicates predicted incorrectly. The x axis is the starting point of the sampled short sequences. The lower figure shows the corresponding voltage signals from our prototype Pain Meter recording. There is a period from 0.55 minute to 0.65 minute, during which the classifier consistently makes incorrect predictions. From the pulse signals, they also perform abnormally compared to other periods. This is also true for the short peak in the data coming from the palm near wrist at 0.28 minute. This illustrates that for those periods, by looking at shorter sequences, it is easy to make a wrong prediction for both human experts and a classifier. This is due to the fact that our sensors are sensitive to movements. Noise can be introduced with a contact position change between sensors and skin. However, by using Consensus Prediction, these errors have no effect on the final prediction.

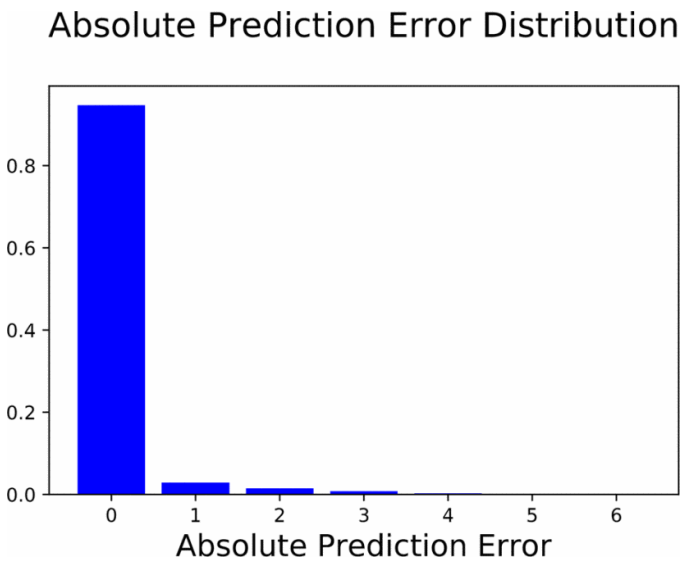


Figure 15. The distribution of absolute prediction error using ordinal regression for dataset 2. The chance of making an error decreases with the increase of absolute prediction error.

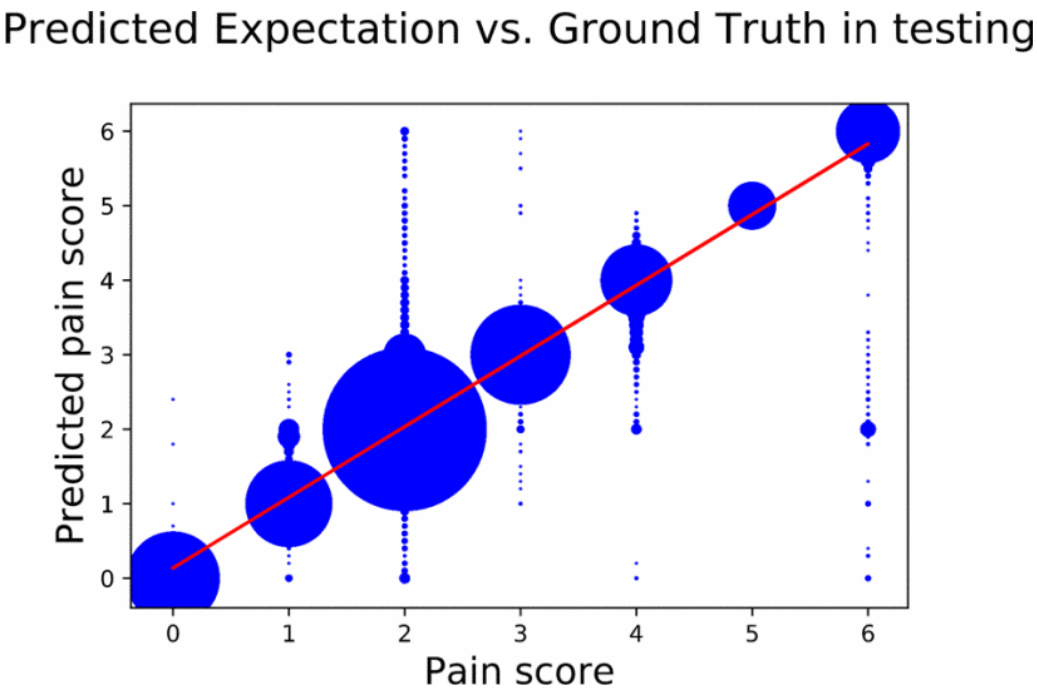


Figure 16. Scatter plot of expected predicted pain score vs. self-reported pain score for dataset 2. The larger size of the point corresponds to the higher appearance frequency of the data point. We keep a decimal in calculating the expected predicted pain score because 0.1 is a reasonable precision to estimate pain in real life. The predicted pain score exerts high linear relationship with ground truth with a high R2 of 0.9463.

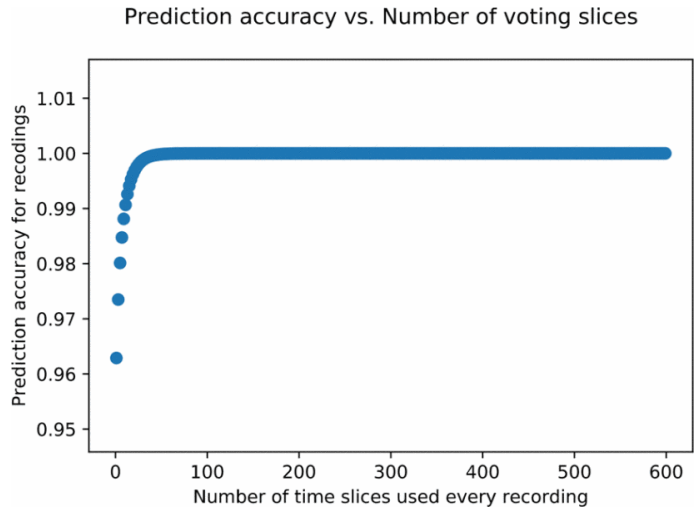


Figure 17. Consensus prediction for chronic pain score prediction in dataset 1.

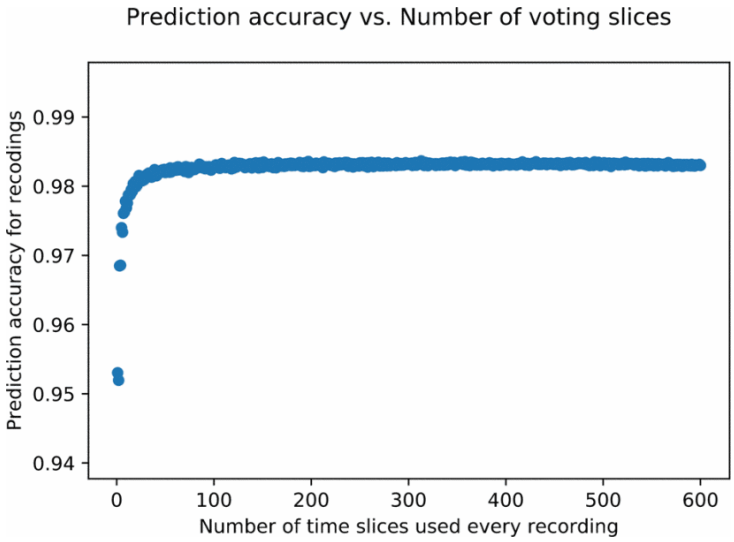


Figure 18. Consensus prediction for chronic pain score prediction in dataset 2.

1. Optimizing the Pain Meter Design

One of the benefits of our deep learning framework is that it can distinguish which sensors are most useful for pain score assessment. We compare the performance of each of the separate signals in Pain Meter 1 as shown in Fig. 21. We report the average accuracy using 5-fold cross validation. This can help us to further improve our chronic Pain Meter. By testing individual signal performance, we find that the accuracy for temperature signals are all around 0.5, which is similar to a random guess in binary classification. Thus the temperature signals are not very informative. We also find that the temple pulse and the palm near wrist pulse contribute substantially to our pain score prediction.

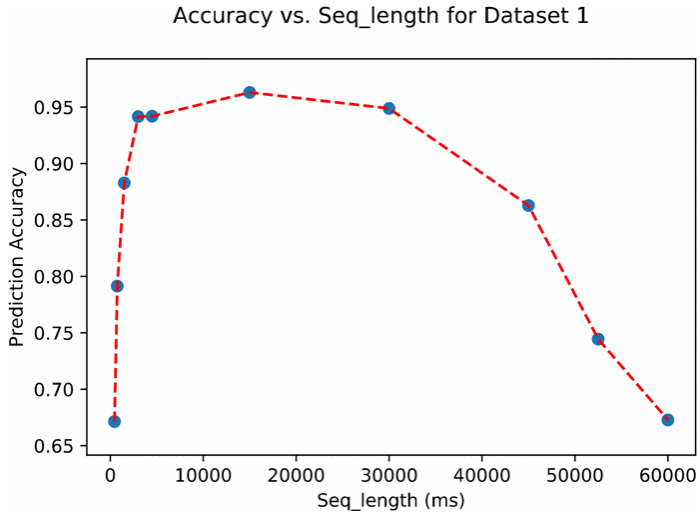


Figure 19. Accuracy vs. seq_length trend for dataset 1.

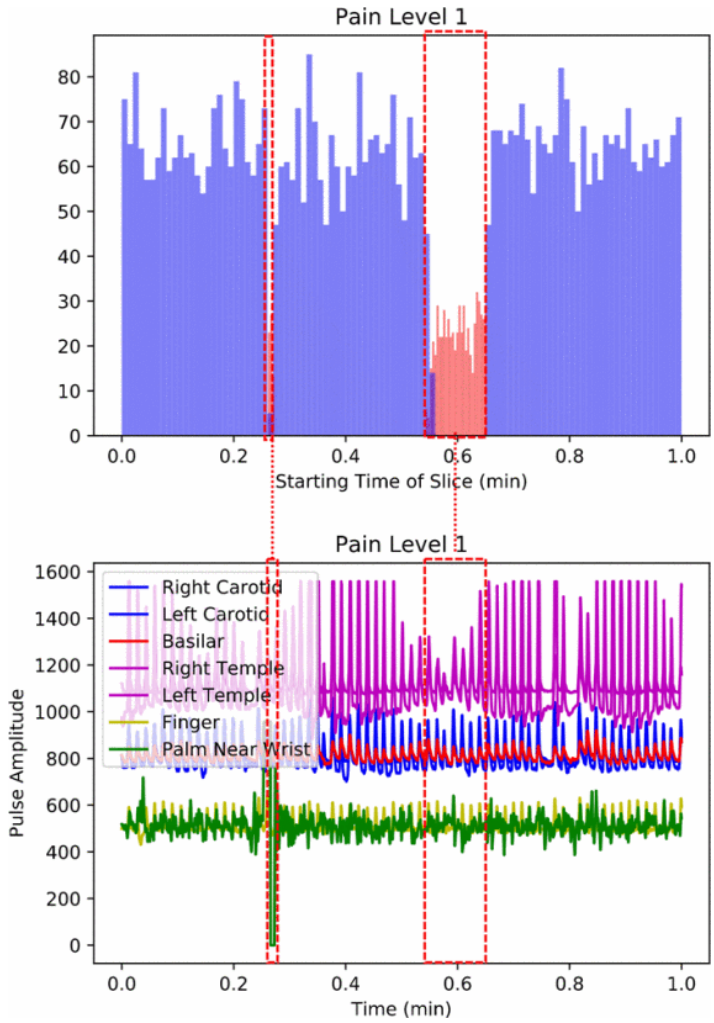


Figure 20. Case study for short slices that the classifier predicted incorrectly. In the upper figure, red color bars represent incorrect predictions while blue bars represent correct predictions.

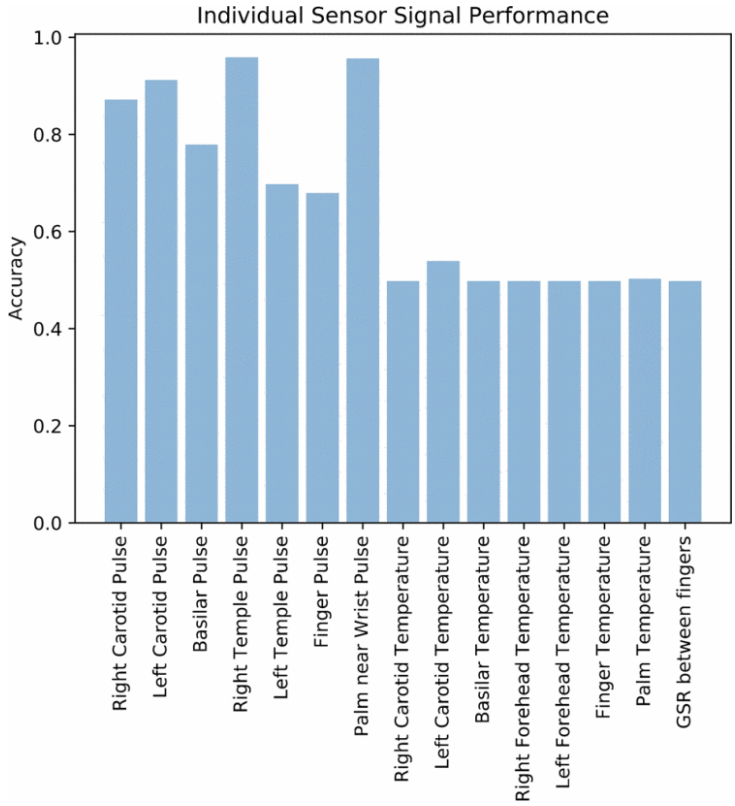


Figure 21. Individual sensor signal classification performance for dataset 1.

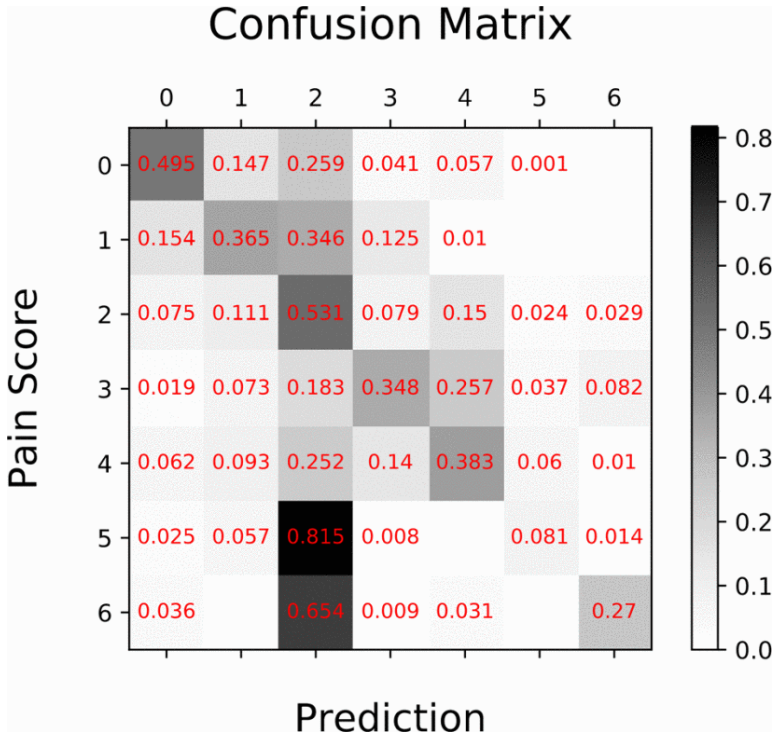


Figure 22. Confusion matrix for chronic pain score prediction in dataset 2 using leave one recording out. The performance is poor due to subjects' differing sensitivities to pain, the use of different settings while recording and the limited unbalanced dataset.

D. Summary

We have addressed the issue of predicting a chronic pain score by proposing a deep learning ordinal regression framework. We split the long recordings into smaller slices, which not only eases the burden on GPU memory but also provides more training samples for the deep learning model. We define and use Consensus Prediction during testing. We present the Confusion Matrix of leave-one-recording-out in Fig. 22. Leave-one-recording-out does not perform well due to subjects' differing perceptions of pain intensities. Also since we have much more pain score 2 samples compared with other pain score samples in Dataset 2, as is shown in Fig. 10, the model is prone to make predictions of pain score 2.

This paper is a proof of principle for chronic pain score assessment via deep learning. It can provide an objective pain assessment for each patient. More data for additional chronic pain subjects is needed before it can be definitively known if deep learning will be a generally useful technique for chronic pain assessment.

IV. Preliminary Study: Quantification of Chronic Pain from Physiological Data

It is unknown if physiological changes associated with chronic pain could be measured with inexpensive physiological sensors. Recently, acute pain and laboratory induced pain have been quantified with physiological sensors. Here we investigate the extent to which chronic pain can be quantified with physiological sensors.

A. Introduction

Chronic pain is a significant public health problem. Recent estimates show that 50 million U.S. adults had chronic pain [75]. Chronic pain and pain-related disability in the United States costs $560-$650 billion dollars, far exceeding the costs of cardiovascular disease, cancer, and diabetes [89]. Yet, despite the high cost and the profound social and personal burdens imposed by chronic pain, the clinical means by which we quantify levels of pain are largely relied on subjective self-reports rather than objective measures of pain intensity [71].

A large body of work demonstrates that chronic pain causes changes in the brain that cause pain to persist beyond tissue healing. In particular, Functional Magnetic Resonance Imaging, fMRI, has been used to demonstrate changes in the brain that are consistent with chronic pain [56-59, 65, 66, 77, 87, 106, 109, 127, 129, 135]. One recent randomized controlled study [56] found significant changes in fMRI imaging of subjects with chronic back pain before and after psychological pain treatment, showing a possibility of using fMRI in pain measurement. However, fMRI is rather expensive and hard to do for routine

use. Less expensive and easy to use devices that can also show changes that are consistent with chronic pain are needed.

This research [56] along with another randomized controlled study [80] demonstrated interventions that not only reduced, but in many cases cured, chronic pain by retraining the brain. Techniques for retraining the brain away from chronic pain include psychological therapies [54,78,81,82,86,91,95,99,100,104,110,115,121–123,140,141,150], education [114,117,120,131,137], biofeedback [84,128,132], activities 55,70,97,107,108,116,139,144,145,149] and meditation [60,85,96,155]. Multimodal combinations of these techniques have been successful [62,72,74,79,88,90,94,103,146].

The changes in the brain that can cause chronic pain may also cause physiological changes that can be measured. Recently there has been significant progress in quantification of pain with physiological sensors [63,67−69,73,83,92,98,111,113,125, 126,130,138,151, 153, 154, 156]. Sensors that measure physiological signals are usually low-cost and easy-to-use. By measuring physiological changes due to chronic pain, such a device could potentially be useful to diagnose chronic pain more accurately, which can in turn assist in pain reduction. Such a device does not directly help reducing the pain, instead, it assists in pain reduction by better assessing the pain. For example, one can use it to monitor pain before and after an intervention to assess the effectiveness of pain reducing techniques described above. Additionally, quantification of pain is important because there is increasing concern that gender, race, age, and intellectual development disabilities may be involved in diagnostic delays as well as over-/under-treatment of pain [61,76,105,119].

A growing body of literature has suggested that self-reported acute pain and experimentally-induced acute pain are associated with differences in physiological

parameters —for example, heart rate and heart rate variability (HRV); blood pressure; peripheral pulsatile components of the cardiac cycle; and electrodermal activity [73,83,98,130,153,154]. Much of this work is consistent with the notion that *multivariate assessments* are superior to single-parameter models when predicting the subjective experience of pain intensity. For example, in an experimental induction of acute thermal pain, *only a linear combination* of physiological data from electrocardiograms (ECG), photoplethysmograms (PPG), and galvanic skin response (GSR) significantly differentiated between no-pain and low-, medium-, and high-pain states—*no individual parameter* was able to distinguish between no-pain and pain states alone [138], suggesting that information from multiple autonomic physiological signals may indeed offer the most promising avenue for objective pain quantification [111]. This modeling approach extended to cases of pure nociception under anesthesia, where the conscious perception of pain was ostensibly impossible: a multivariate approach including HRV, PPG, and pulse transition time accounted for intra-operative clinical measures of noxious intensity of incision [126]. Similarly, in another study, heart rate, HRV, PPG, GSR, and associated temporal/spectral subcomponents of those physiological signals differentiated between noxious and non-noxious surgical events [63].

Of special relevance to the present study, previous research [156] has highlighted the promise of machine learning models for pain assessment. Multilayer artificial neural networks using features derived from skin conductance and ECG distinguished between no pain and moderate-to-high levels of experimentally-induced thermal pain, with a combination of the two features outperforming either measure alone [113]; in the case of chronic pain resulting from sickle cell anemia, multinomial logistic regression using six

physiological features significantly predicted pain scores on an 11-point pain scale—both within- and between-individuals [151]. Thus, there is considerable evidence to suggest that multivariate assessments of classical physiological measures can inform our understanding of pain intensity. Critically, however, *there is no apparent consensus* on the optimal set of features to use: the pioneering work on this question has explored a highly-diverse field of potential physiological parameters [67−69,77,92,93,101,102,112,118,133,134,136, 143,147, 148, 152].

In this study we report preliminary pain prediction results using physiological data collected from chronic pain sufferers on our new Pain Meter. We investigated both individual-level models and an overall, population-level model, spanning various combinations of pulse and temperature features. At the individual-level, our results show that the ability to quantify chronic pain can vary quite considerably, with only half of the participants showing moderately-strong predictive accuracies. Aggregating the data into a population-level model, however, significantly predicted pain scores across each recording session, providing preliminary evidence for a generalizable model of chronic pain quantification using physiological parameters.

B. Methods

Pain Meter

The prototype equipment was built to allow subjects to record at home. For safety, we avoided any sensors that required electrode contact to the subjects' bodies. These in-home subject recordings enable us to capture more natural variation in pain levels for each subject, including high pain levels that would have made travel to our laboratory difficult. The ability to create subject specific chronic pain models is due to the ability to record in homes.

Our Prototype Pain Meter (**Fig. 23**) represents the seventh iteration of our device. The sensors we chose were inexpensive and available commercially: photoplethysmograms pulse sensors (Pulse Sensor Amplified; http://www.pulsesensor.com); temperature sensors (10K Ohm NTC Thermistors MF52-103); IR temperature sensor (MLX 90614); acceleration/gyroscope sensors (HiLetgo GY-521 MPU-6050 MPU6050 3 Axis Accelerometer Gyroscope Modules); and force sensors (Interlink FSR 402 on the forehead and wrist; and Interlink 406 on the basilar and carotid arteries). An example of the signals recorded from Pain Meter in shown in **Figure 24**. The Prototype Pain Meter does not apply stimulation to obtain different pain score recordings.

Figure 23. Pain Meter contains (1) PPG pulse sensors in a headband, in a neck pillow, wristband, and on the fingertip; (2) temperature sensors on the temple, forehead, wrist, and fingertip; (3) 3-axis accelerometers and 3-axis gyros in the headband and wristband; and (4) force sensors on the left carotid and basilar. PPG, photoplethysmograms.

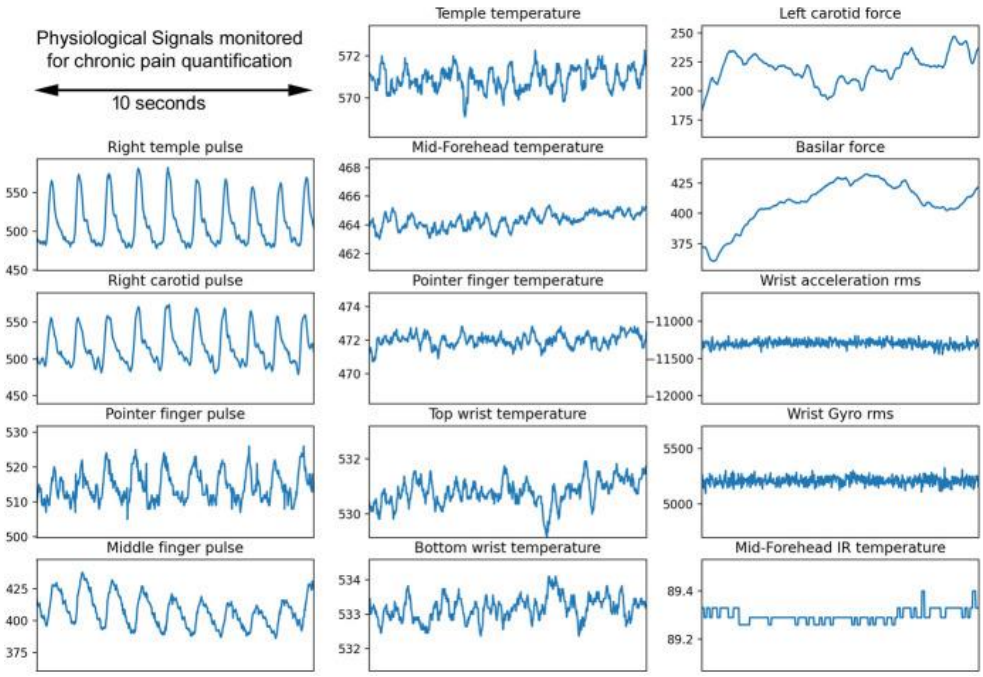


Figure 24. An example of the signals recorded from Pain Meter.

A Teensy 3.6 microcontroller with a 32-bit 180 MHz ARM Cortex-M4 processor was used to sample all signals at a frequency of 66.67 Hz. Data visualization and acquisition programs for our Pain Meter were written in Python. Data was remotely monitored using the Dropbox cloud storage system.

Participants

Data were collected from 12 participants (9 women, 3 men; M_{age} = 38.25, *SD* = 17.10, range = 21 - 65), yielding a total of 183 10-minute recordings. Importantly, we did not specify any explicit exclusion criteria for participation: thus, pain etiology/symptomatology, duration, and intensity varied across individuals.

All subjects were recruited via email and provided written informed consent for a protocol approved by the UCSB Human Subjects Committee.

Data Collection

Subjects first opened a computer program on the Pain Meter computer that was supplied to them. They connected to Wi-Fi on the first use. Then they were asked to rate the best representation of their current pain score using a visual analogue scale, which included pictures and short descriptions of pain states from 0 to 10 (for example, the description of state 0 was 'no pain' and the description of state 10 was 'unimaginable unspeakable').

Subjects next put on the headband and neck pillow sensors, adjusted their position until they felt comfortable, and secured the sensors with Velcro (headband) and a fastener (neck pillow). Next, they put on a wrist band and two finger cuff sensors. Once these were secured, the computer monitored all of the pulse signals and gave green lights for all sensors giving good signals (signals with clearly detected peaks) and yellow for any that were not. If there were any yellow, the subject was asked to readjust the sensor and try again.

Once all of the sensors were giving good signals, a 10-minute recording session was initiated in which pulse, temperature, force, and motion data were recorded and stored in local memory (at the end of the session, the complete record was automatically uploaded to the cloud).

During this recording, the subject was instructed to relax and to look at a fixation cross on the computer screen. The data being recorded was not visible to the subjects during the recording. After the recording was done, the subjects were asked to report their current pain score again with the same visual analogue scale. The protocol instructed subjects to record for two weeks, twice per day.

Data Pre-processing

Movements of the subjects during the recording sessions can disturb the pulse signal and add additional noise into the data. Consequently, it can be difficult to extract robust,

interpretable features from individual pulses. We use several steps to preprocess the data and reduce the influence of noise in our analysis. After unstable segments are detected (as detailed in **Fig. 25**), we remove them and split the stable segments into 10-second samples before proceeding to feature extraction. We used three pulse sensors in our analysis: temple pulse sensor, carotid pulse sensor, and pointer finger pulse sensor. Note that, because we removed all unstable segments, the number of available recordings and samples differed depending on the combination of sensors—**Table 7** shows the number of stable recordings and samples for each combination.

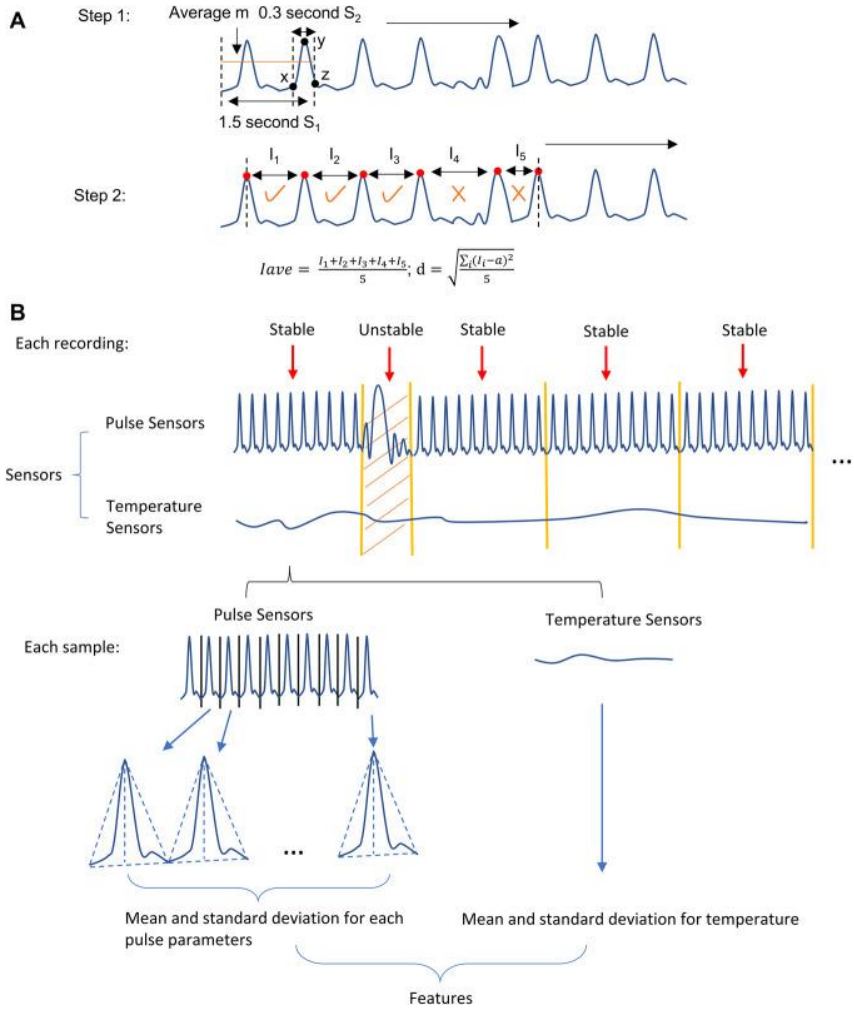


Figure 25. Preprocessing and feature extraction. (A) Step 1: Detect all pulse peaks in the recording by comparing the moving average m of each 100 data-point segment in S1 (S1 is

1.5-second-long) with the last 20 data points in S2 (S2 is 0.3-second-long) of S1. If the first and last data point in S2 is smaller than m and the max value of S2 is larger than m, then a peak is detected. Step 2: Take the first 6 peaks and calculate the mean Iave and SD d of the time between peaks. If Iave is between 0.6 and 1.2 seconds and d is smaller than 0.2 and if the time I between 2 peaks is between 0.9 and 1.1 times Iave, this data segment is classified as stable. Otherwise, this data segment is classified as unstable. Repeat this process for all the data. (B) The data preprocessing and feature extraction steps: (1) Remove unstable segments and divide the rest of the recording into continuous 10-second samples. (2) Inside each sample, extract pulse features such as the mean and SD for each pulse parameters and mean and SD for temperature.

Details of the models trained with different combination of pulse sensors.

Pulse sensors	**# of recordings**	**# of samples**	***r***	**RMSE**
1	173	7755	0.60	1.52
2	168	6671	0.44	1.78
3	179	8900	0.60	1.52
1, 2	156	5446	0.45	1.69
1, 3	168	7033	0.57	1.56
2, 3	163	5978	0.49	1.66
1, 2, 3	152	5044	0.41	1.71

Table 7: Number of recordings, number of samples, Pearson correlation coefficient (r), and root mean square error (RMSE) are shown. For pulse sensors, 1 represents temple pulse sensor; 2 represents carotid pulse sensor; and 3 represents pointer finger pulse sensor.

Participants provided one pain score at the beginning of each recording and then again at the conclusion. We did not give any instructions to the participants such as telling them to relax or in any way attempt to change their pain. Nevertheless participants often reported a different pain level at the end of the session. We therefore split the data and assigned the first pain score to samples from the first half of each recording and the second pain score to samples from the second half. This is, of course, only an approximation because we have no information about when the pain level changed, but it seemed preferable to using only the beginning pain score or only the end pain score for the entire recording.

Feature Extraction

Each 10-second sample included stable signals from the pulse sensors and temperature sensors. We defined a series of pulse features analogously to earlier research [97], which suggested that pulse morphology derived from transmitted-light PPG can be used for postoperative pain assessment. Our device, however, uses green reflective-light pulse sensors with highly-filtered output (because they are both inexpensive and readily available). We therefore investigated whether our pulse sensor parameters can inform the quantification of chronic pain.

We extracted only pulse parameters that are independent of the height of the pulse because the height of the pulse can be influenced by how tightly the device is worn and its position. Specifically, we extracted pulse width parameters: L_R, L_F, PPI_H and PPI_L (**Fig. 26**). PPI_H is defined as the interval between two consecutive high peaks of pulses. PPI_L is defined as the interval between two consecutive low peaks of pulses. The standard deviation of PPI_H is a measure of pulse rate variability, which was used as one of the parameters in our models.

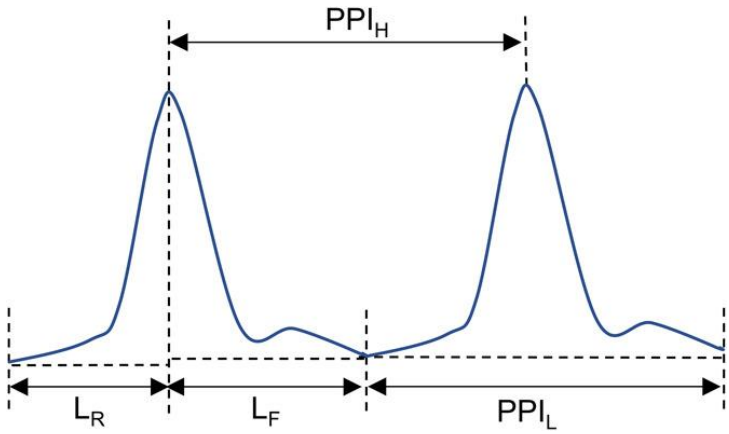


Figure 26. Pulse parameters. LR is the width of the rising part of the pulse; LF is the width of the falling part of the pulse; PPIH is the interval between 2 consecutive highs; and PPIL is the interval between 2 consecutive lows.

From each 10-second sample, after finding the peaks, we segmented the data into individual pulses by finding the local minimum between peaks. **Figure 25B** illustrates this process.

Thus, the following features were extracted from the samples:

1. Mean and standard deviation of all pulse sensor parameters.
2. Mean of temperature at forehead, temple, top of wrist, bottom of wrist and finger.

Additionally, we computed $\overline{T_{Temple}}$ / $\overline{T_{Finger}}$, $\overline{T_{Temple}}$ - $\overline{T_{Finger}}$, $\overline{T_{Forehead}}$ / $\overline{T_{Finger}}$, $\overline{T_{Forehead}}$ - $\overline{T_{Finger}}$, where $\overline{T_{Temple}}$ represents the temple temperature mean for each 10-second sample window, $\overline{T_{Forehead}}$ represents the forehead temperature mean and $\overline{T_{Finger}}$, represents the finger temperature mean. All features were z-scored normalized prior to further analysis.

Recursive Feature Elimination and Random Forest Regression

Feature selection methods are used to identify a subset of 'important' features and simplify a model. These techniques often afford several advantages, such as reducing computational costs, avoiding overfitting, and improving model performance.

One popular method, Recursive Feature Elimination (RFE), selects features by training a model with one set of features, ranking them by an importance metric, and removing the least important. This procedure is recursively repeated until the user-specified number of features is reached. We implemented subject-specific linear models for individualized pain prediction, accounting for the fact that relevant physiological parameters could likely vary on a subject-by-subject basis.

At the overall population level, we used random forest regression: an ensemble method that aggregates the predictions from multiple regression trees to make more accurate

predictions. Regression trees are decision trees when the outcomes are continuous values, recursively partitioning the data space into smaller regions where simple models can be used to describe the relationships between various model inputs and outcomes. In random forest, each regression tree is created using random subsets of features and random subsets of samples. Accordingly, because random forest uses different subsets from the data and features, it is less likely to overfit and performs well with high-dimensional datasets. Additional advantages include fewer constraints and assumptions about the data and its distribution, as well as robustness to outliers and missing values. After the forest is fit, the predicted pain score of the test set is computed as the mean predicted score of the trees in the forest. This part of the analysis was done with the Python package *scikit-learn* [124].

7. Leave-One-Recording-Out Cross Validation

We used leave-one-recording-out cross validation to assess performance in all models (individual- and population-level). For leave-one-recording-out, all samples from one recording were iteratively removed from the dataset, the model was trained on the remaining data, and predicted pain scores were obtained for the left-out recording. Pearson correlation coefficient (r), intraclass correlation coefficient and root mean square error (RMSE) between the predicted pain ratings and the reported values are reported. Intraclass correlation coefficients (ICC(3,1)) are calculated with R package *psych* [142].

We additionally performed 1000 iterations of nonparametric permutation tests to assess model performance against chance. In this procedure, we constructed a robust empirical null distribution of predictive accuracies expected if pain scores were randomly associated with patterns of physiological parameters. Thus, for each permutation, pain scores were shuffled randomly for all recordings, models were fit using cross-validation as above, and

performance was recorded. Significance was determined by the proportion of cases in which null-model accuracy equaled or exceeded the 'true' accuracy.

Since Pearson correlation is not an agreement statistic, a high correlation does not necessarily imply a good agreement between the two scores. Thus, we also ran a Bland-Altman analysis for our population-level model.

Despite these checks, there are still limitations and potential overfit problems with our methods. Future studies are needed to validate stable models against independent subject samples.

C. Results

We were able to capture a fairly-wide range of pain scores (spanning 0-9) through our home recording sessions. However, we note that more extreme ratings (particularly, pain scores 0, 8 and 9) have many fewer samples for analysis relative to the other values. We elected to retain these recordings given the continuous nature of our regression frameworks, but the imbalanced representation of scale endpoints is worth keeping in mind for model assessment.

1. Individual-level Models

We first trained and cross-validated predictive models on an individualized, within-subject basis. Because of the smaller sample size for individual subjects, we used recursive feature elimination to select five features first to avoid overfitting and then fit a linear model to predict the pain scores for each subject. All combinations of pulse sensors were tested for each subject. The models with the combination that gave the best predictive performance (i.e. correlations between reported and predicted pain) are reported in Figure 27. As expected, performance was rather variable across individuals. However, pain scores in 5 out

of 12 subjects were predicted with moderately-strong accuracy ($0.49 < r < 0.78$, $p < 0.05$). The intraclass correlation coefficients for these five subjects are between 0.46 and 0.75. A few features were selected by recursive feature elimination repeatedly in these five models. For example, $\overline{PPI_L}$ was selected in 5/5 models and $\overline{L_F}$, $\overline{T_{Forehead}}$, $\overline{T_{Finger}}$ and $\overline{T_{Forehead}}$ - $\overline{T_{Finger}}$ were selected in 4/5 models. However, the relationships between these features with pain score are not consistent across different individuals. For example, $\overline{T_{Forehead}}$ and $\overline{T_{Finger}}$ increased with the increase of pain score in 2/4 subjects, and decreased in the other two subjects.

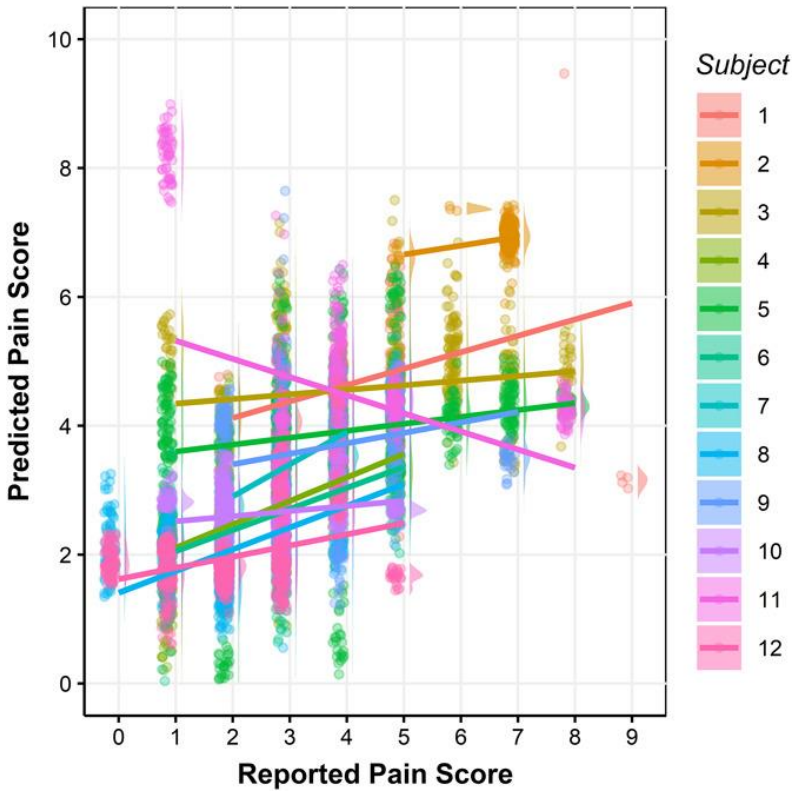


Figure 27. Cross-validated predicted pain scores vs reported pain scores for each subject. Different colors represent different subjects in the raincloud plots. Each dot represents one sample, and each line is the fitted linear regression line for each subject. In total, there are 6982 samples, 99.50% of the data is shown in the figure. The rest were omitted to provide higher resolution for the bulk of the data. The Pearson correlation coefficients between reported and the predicted pain scores are 0.78, 0.29, 0.14, 0.54, 0.19, 0.49, 0.56, 0.50, 0.24, 0.16, −0.31, and 0.35, respectively, for subjects 1 through 12.

2. Population-level Model

We then aggregated the data from all subjects and fit population-level random forest models using different combinations of pulse sensors. Table 7 shows the results from each possible combination of sensors. In general, modeling more pulse sensors simultaneously leads to a smaller number of potential training samples overall (since a given 10-second window was only used if it was stable for all sensors combined). Models including only the pointer finger sensor or only the temple sensor gave the strongest correspondences between actual and predicted pain ($r = 0.60$, RMSE = 1.52). The intraclass correlation coefficient for both models are 0.58, showing a moderate similarity between actual and predicted pain.

However, as aforementioned, predictive accuracy generally suffered the worst at pain scores with fewer available examples (Fig. 28). Because the number of stable recordings and samples varies when using different pulse sensors, there are different number of samples for each pain score. The more extreme ends of pain score (8 and 9) in particular were considerably undershot by the model. Nonetheless, performance on the whole consistently outperformed empirical null models, suggesting that pain prediction was still significantly better than chance ($p < .001$).

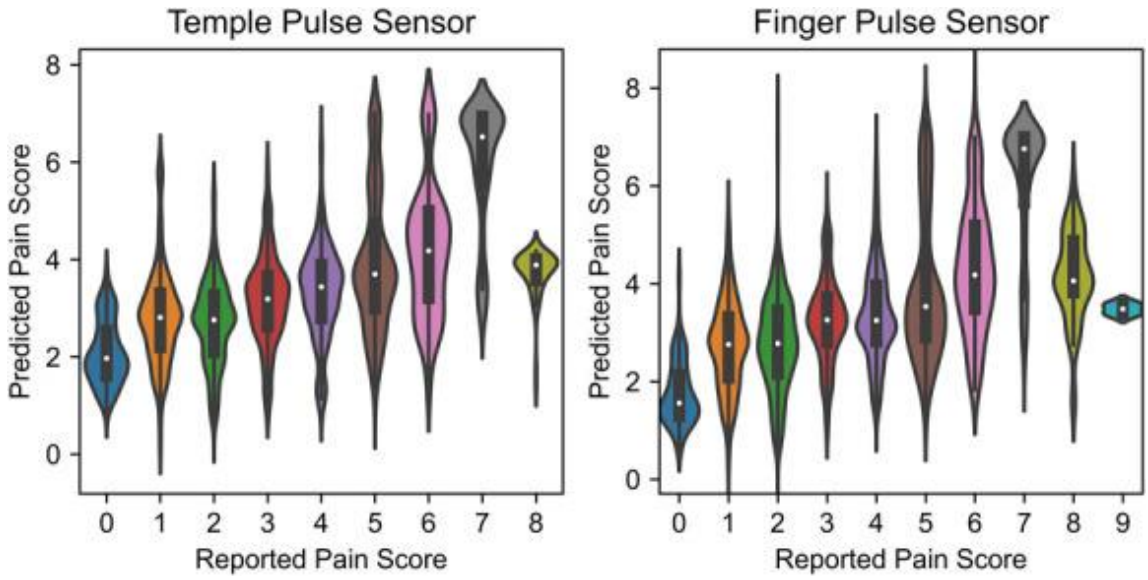


Figure 28. Cross-validated predicted pain scores vs reported pain scores for population-level models. Left panel shows the results from the model using signals from temple pulse sensors, and right panel shows the results from the model using signals from pointer finger pulse sensors. The violin plots for each class shows the distributions of the predicted pain scores.

Because one of the measurements in this paper is the self-reported pain, which we consider as our reference measurement, we used the self-reported pain scores as the x axis in our Bland-Altman plots. Note that the self-reported pain scores are integers from 0 to 10, thus all dots in the plots lie on vertical lines with integer x value. The bias and agreement limits are the same for both models. The bias is 0.02, and the range of agreements is (-2.95, 2.99). Our model tends to overestimate the lower end of the pain scores and underestimate the higher end of the pain scores. Note that plotting against reference measurement instead of average measurements might be responsible for the association between difference and pain scores [64].

D. Summary

This is the first demonstration, to our knowledge, that physiological data can be correlated with chronic pain, both for individuals and populations. In our subject-specific

models for individualized pain prediction, 5 out of 12 subjects yielded Pearson correlation coefficients of 0.49 ($p<0.05$) to 0.78 ($p<0.05$) and intraclass correlation coefficients between 0.46 to 0.75. In our population-level model, we achieved an intraclass correlation coefficient of 0.58 and a significant Pearson correlation between self-reported and predicted pain ($r = 0.60$, $p < .001$) in this preliminary study. It is likely that higher correlations will be achieved with better devices in the future. This first demonstration had significant limitations due to the limited amount of data that we could collect shortly before and during COVID-19. Additional research is urgently needed to explore the effects of the type and duration of chronic pain, the age and gender of the subjects and other factors.

Additional research is also needed to optimize the selection and placement of sensors. As one example of possible future improvements, an optimized pulse sensor would be able to sense the low frequency variations in the baseline due to breathing, which were filtered out in the commercial pulse sensor that we used. It is also possible (perhaps even probable) that the inclusion of electrocardiogram (ECG), electromyography (EMG), or skin impedance sensors would give better correlations. However, in the interest of subject safety, since these were unsupervised measurements in the subjects' own homes, we did not include any devices with electrodes.

Our long-term goal is the development of a relatively low-cost, easy to use system that could be used by patients in their own homes or in physician offices to objectively measure chronic pain. Hopefully this first demonstration of the possibility of achieving this goal will inspire engineers, physicists, computer scientists and psychologists to join in doing the work that will be necessary to move from the demonstration of a possibility to a practical system. The current system of subjective self-reported pain levels results in the over- and under-

prescribing of treatment. Shifting to an objective methodology would provide healthcare professionals the ability to properly prescribe patients with the appropriate dosage of treatment to alleviate pain. This is especially important with the many risks associated with pain medication, including long-term opioid addiction.

V. Dynamic Phase Extraction: Applications in Pulse Rate Variability

Pulse rate variability is a physiological parameter that has been extensively studied and correlated with many physical ailments. However, the phase relationship between inter-beat interval, IBI, and breathing has very rarely been studied. A program based on Lock-in Amplifier technology was written in Python to implement a novel technique, Dynamic Phase Extraction. It was tested using a breath pacer and a PPG sensor on 6 subjects who followed a breath pacer at varied breathing rates. The data were then analyzed using both traditional methods and the novel technique (Dynamic Phase Extraction) utilizing a breath pacer. Pulse data was extracted using a PPG sensor. Dynamic Phase Extraction can extract both the phase between the breath pacer and the changes in IBI due to the respiratory sinus arrhythmia component of pulse rate variability (ΔIBI), but is limited by needing a breath pacer.

A. Introduction

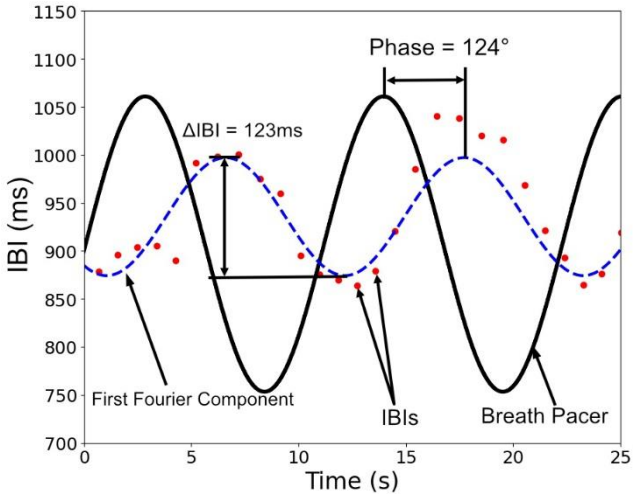


Figure 29. Dynamic Phase Extraction (DPE) can give the phase and magnitude (ΔIBI) of the variation in Inter-beat Interval (IBI) due to Respiratory Sinus Arrhythmia (RSA). We define the phase as a shift to the left being negative, and a shift to the right being positive, in agreement with standard physics and engineering convention. The phase and ΔIBI are determined by DPE, which pulls out the first Fourier component of the IBI Interval curve, as shown with the blue dashed line.

Recent experiments have reawakened interest in the phase relationships caused by the delay between the modulation of the heart rate from the breath and the breath [157, 158]. Lehrer et al. have shown the phase shift at the breathing rate that yields the maximum power in heart rate variability varies systematically with age [157, 158]. This report builds upon those papers and the largely-ignored work on phase by Angelone and Coulter, who demonstrated over 50 years ago that there is an intricate relationship between phase, heart rate variability amplitude, and breathing rate [159]. In this paper we introduce a method to accurately and efficiently extract the phase of the respiratory sinus arrythmia (RSA) induced variations in inter-beat interval (IBI), measured with photoplethysmography, relative to the breath. For the sake of clarity and succinctness, in this paper we will define the change in IBI due to RSA as the *change in IBI* (ΔIBI).

Our Dynamic Phase Extraction (DPE) method is based on lock-in amplifier technology. The lock-in amplifier is a powerful instrument commonly used in astronomy, physics, and engineering that can isolate signals a millionth the amplitude of the noise.

We propose that this method could be used for analyzing the wave characteristics of RSA and may be useful for biofeedback studies of pulse rate variability. Broadly, pulse and heart rate variability (HRV) is negatively associated with a multitude of ailments and physical states, including chronic pain [160, 161], post myocardial infarction death rate [162-164], anxiety [165-167], diabetes [168], and sympathetic nervous system activation [160, 168, 169]; conversely, it is positively associated with athletic performance [170,171]. Some fluctuations in the very low frequency (VLF) range (<0.04 Hz) have been positively associated with sympathetic nervous system activation, with some further complexities in chronic heart failure patients [169]. This has led to many papers seeking optimization parameters, and extensive work has been done in developing biofeedback and breathing protocols to improve and optimize heart rate variability [166, 172, 173], as well as the development of consumer devices that measure and utilize pulse rate variability, e.g. HeartMath, Whoop, Fitbit.

Previous work has modeled the cardiac system as a two-closed-loop system, with the brain, baroreceptors, vascular tone control system, blood pressure control system, and heart rate control system as its constituents. Vaschillo (2002) emphasized the importance of phase delays in his work on biofeedback and approached the system from an engineering perspective, where often the most important components of a feedback system are the phase relationships [174]. This is reason enough to investigate phase relationships, but there has been even more work motivating an investigation into the phase relationships in HRV.

There is evidence to suggest that the delay causing the phase shift between the changes in IBI and breathing rate exists because a delay results in more efficient pulmonary gas exchange [175, 176], and that the changes in IBI are just the apparent manifestations of a system that optimizes this gas exchange [177]. Thus, we are interested in finding two parameters: the phase between IBI and breathing, and the peak-to-peak amplitude of the inter-beat interval, ΔIBI, as defined in **Fig. 29**.

Notably, the preponderance of extant literature measures HRV by electrocardiography (ECG). Here, however, we use photoplethysmography (PPG) to measure pulse rate variability. While we acknowledge limitations to this approach (e.g., difficulties in precisely capturing the timing of peak-to-peak events), PPG is nevertheless the most accessible form of measuring pulse rates. Our future research hopes to use the technique in various applications where ECG is unfeasible.

Here we introduce and discuss the Dynamic Phase Extraction (DPE) method and demonstrate its ability to extract the phase relationships from PPG data. We also compare the measure for the magnitude of pulse rate variability, ΔIBI, that it generates to other methods of extracting pulse rate variability (the LF Power and Standard Deviation of the IBI) and compare the phase that it extracts to the phase extracted using the Fourier transform. ΔIBI is defined as the peak-to-peak amplitude of the waves in the IBI interval vs. time curve. We extracted the phase and amplitude of the pulse rate variability, compared it to other methods of measuring pulse rate variability and phase, and found good agreements between them.

B. Methods

Participants

We collected data from six healthy young adults (5 men and 1 woman, aged between from 19-26). All procedures were approved by the UC Santa Barbara Human Subjects Committee.

Data acquisition procedures

A custom built PPG sensor was used to obtain pulse data. A chest strap with an integrated FUTEK LSB 200 load cell, an instrument that measures force using a Wheatstone bridge and strain gauges, was used to obtain breathing data. These were used because they were conveniently available in the lab; however, any linear chest circumference measurement and pulse rate measurement could be used.

The PPG sensor was attached to the left index finger using a Velcro strap. The voltage powering the LED in the PPG sensor was adjusted until a pulse waveform that occupied about a third of the voltage range on the measurement device with clear peaks was measured. The chest strap was wrapped around the thorax to measure chest movement, and the output was run through a low-pass filter and then to an amplifier. Voltage values from the PPG sensor and chest strap were read using a Wemos LOLIN32 board sampling at 2000Hz and recorded by a computer.

Each subject was tested for at least 8 different breathing rates, ranging from 2.4 to 57 breaths per minute. IBIs were eliminated without replacement if they were more than double or less than half of the previous pulse rate. In order to calculate the phase and perform Fourier transforms, each IBI interval was interpolated between to ensure consistent time interval spacing.

Participants were instructed to follow a sine wave breath pacer at the desired breathing rate, and to breathe in as the sine wave increased and breath out as the sine wave

decreased. Recording sessions were 12 minutes total, with 2 minutes for each breathing rate. Multiple sessions were recorded for each subject, with the goal of finding a breathing pace where a local maximum of ΔIBI could be seen.

The chest strap data was used to verify that the participant followed the breath pacer, this was done visually, just to make sure that the subject took a breath every cycle of the pacer. Inter-beat Interval (IBI) data were extracted by passing the pulse data through a peak detection program and then using those peaks to determine IBI.

Dynamic Phase Extraction and analysis

The lock-in amplifier is an instrument commonly used in engineering and physics. Here, we apply the principles of lock-in amplifier technology, specifically dual-phase lock-in amplifiers, in the DPE method to extract the phase of the IBI variations relative to the breath pacer. The method can be broken down into a few simple parts.

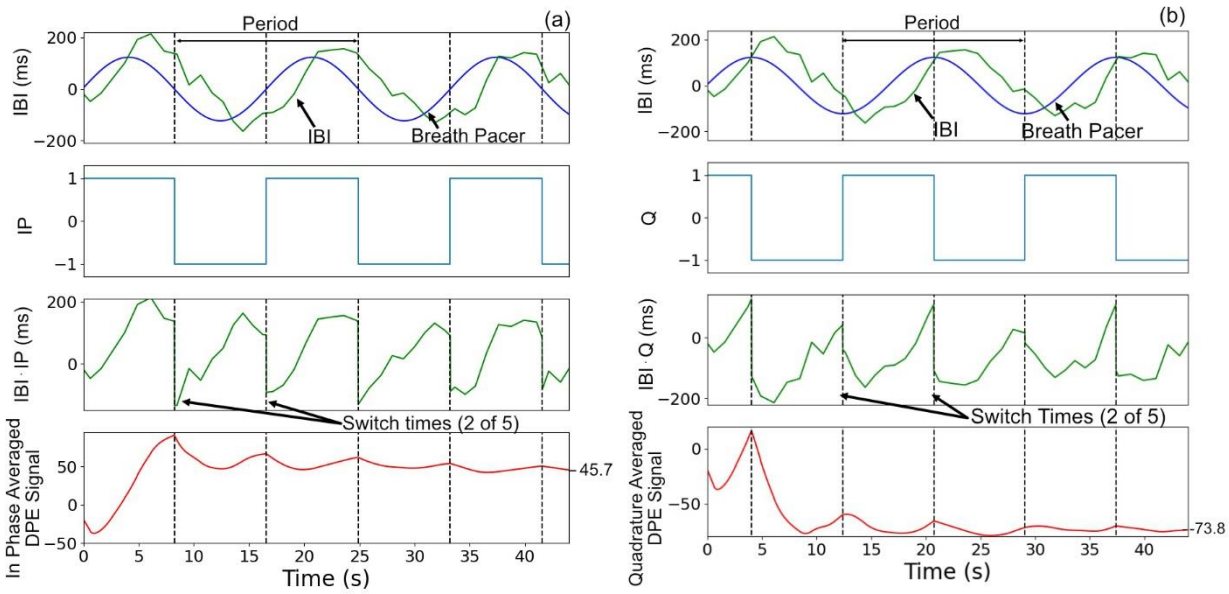


Figure 30. The Dynamic Phase Extraction (DPE) method extracts an In-phase Averaged Signal and a Quadrature (90 degrees out of phase) Averaged Signal. These are used to calculate the amplitude and phase of the Inter-beat Interval (IBI) relative to a Breath Pacer. The average value of the pulse rate was subtracted out to only reflect the changes in IBI.

First, a reference signal is fed into any system which will process and return a signal. In this case, the reference signal is a breath pacer, the system is the human body, and the signal is the Inter-beat Interval. The signal is then processed in two ways: "in-phase" and "quadrature". For the in-phase processing (**Fig. 30a**), the signal is inverted every time the reference signal crosses zero, as seen in the third panel of **Fig. 30a**. Similarly, for the quadrature signal (**Fig. 30b**), the signal is inverted every time the reference signal has a peak or a trough, as seen in the third panel of **Fig. 30b**. The in-phase and quadrature signals are then continuously time-averaged, as seen in the fourth panels of Fig 30a and 30b. In practical terms, we can simply perform a numerical summation.

$$A_0 = \frac{1}{N}\sum_{i=1}^{N} \mathrm{IBI}_i \mathrm{IP}_i \,, \tag{1}$$

$$A_{\pi/2} = \frac{1}{N}\sum_{i=1}^{N} \mathrm{IBI}_i \mathrm{Q}_i \ . \tag{2}$$

In the sum, IP_i is positive for the first half of the reference signal, and negative for the second half (as seen in the second panel of **Fig. 30a**). Similarly, Q_i follows the second panel of **Fig. 30b**, where it starts positive and switches sign every half-wave. i indexes the ith heart beat from where you start counting, and N is the total number of heart beats. After evaluating A_o and $A_{\pi/2,}$ using equations (1) and (2), we can calculate the phase and amplitude (ΔIBI) as follows:

$$\begin{aligned}\text{Phase} := \phi = \text{arctan2}\left(A_0, A_{\pi/2}\right) = -\text{arctan2}(45.7, -73.8) \\ = 58.2^o,\end{aligned} \tag{3}$$

$$\Delta\text{IBI} := 2A = \pi\sqrt{A_0^2 + A_{\pi/2}^2} = \pi\sqrt{73.8^2 + 45.7^2} = 86.8\, ms, \tag{4}$$

where ΔIBI is the peak-to-peak amplitude of the signal, A_0 is the in-phase value, and $A_{\pi/2}$ is the quadrature value. The function arctan2 gives the angle of a ray through the origin to a point (*x*,*y*) relative to the *x*-axis. It is used because the normal arctan only ranges from -90° to $+90^o$, while arctan2 ranges from -180° to $+180^o$. We are justified in using these formulas as DPE extracts the first Fourier component at the reference frequency. This can be seen by converting our sum (eqs. (1) and (2)) to an integral:

$$A_0 = \frac{1}{T}\int_0^T IBI(t)IP(t)\, dt,$$

$$A_{\pi/2} = \frac{1}{T}\int_0^T IBI(t)Q(t)\, dt.$$

If IBI(*t*) is periodic, its Fourier decomposition is given by:

$$IBI(\omega t) \ = \sum_{i=1}^{\infty} a_i \sin(i\omega t) \ + b_i \cos(i\omega t).$$

We can neglect the constant value in the Fourier transform because it is exactly equal to the average value, which we subtracted out of the IBI.

However, when we integrate with the in-phase and quadrature signals, the sines and cosines cause anything other than a wave at the same frequency as the reference to diminish as 1/*T*, thus only the *I* = 1 terms are left. We can choose to start anywhere—in other words, may keep only the sine terms without loss of generality. Thus, we are left integrating:

$$A_0 = \frac{1}{T}\int_0^T a_1 \sin(\omega t - \phi)\,\mathrm{IP}(\omega t)dt + \epsilon = \frac{2a_i}{\pi}\cos(\phi),$$

$$A_{\pi/2} = \frac{1}{T}\int_0^T a_1 \sin(\omega t - \phi)\,\mathrm{Q}(\omega t)dt + \epsilon = -\frac{2a_i}{\pi}\sin(\phi).$$

We can use the result of this integration to easily derive eqs. (3) and (4).

In general, DPE will have a reasonable phase after two breaths, but inevitably, the more breaths that are taken, the better the precision of the method will be. The trade-off between time and precision is important to consider when measuring phase, and depends strongly on the application of DPE.

C. Results

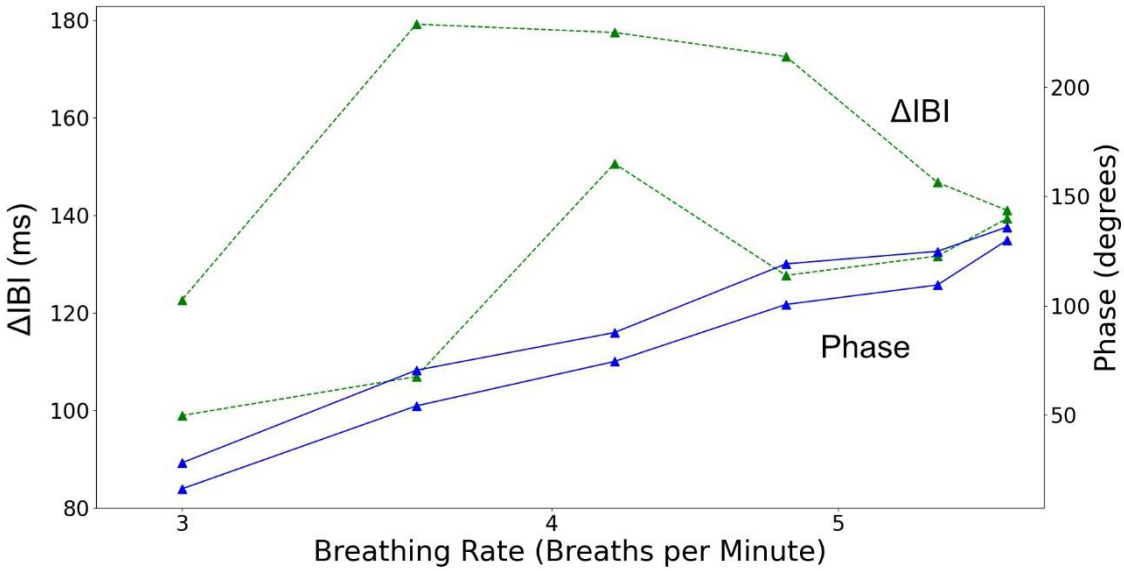


Figure 31. ΔIBI and phase as a function of Breathing Rate for a young, athletic person. The maximum ΔIBI occurs around 100° Phase. The x-axis is plotted on a log scale. Multiple runs are reported.

First, we investigated how changes in the Breathing Rate (Breaths per Minute) of the breath pacer modulated the amplitude, ΔIBI, and Phase, as shown in **Fig. 31**.

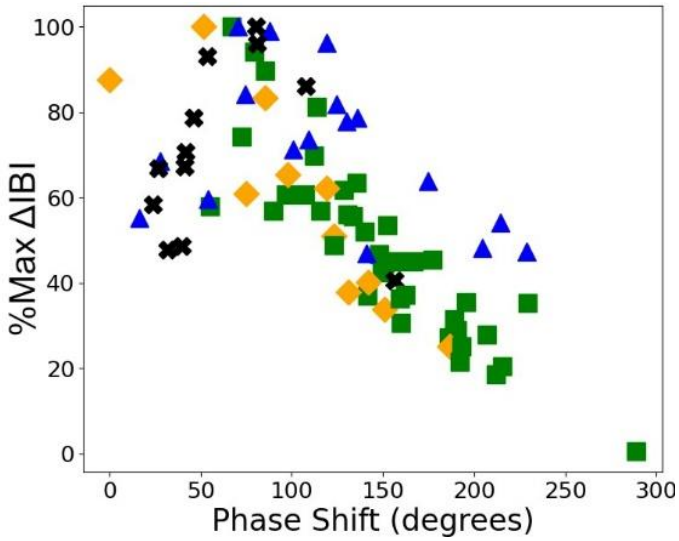


Figure 32. Different people (shown by different symbols and colors) can have slightly different phases at their maximum ΔIBI

This allowed us to identify the breathing rate that maximizes ΔIBI (**Fig. 32**). If we compare multiple trials to each other, we can locate what phase an individual's ΔIBI is maximized at, and compare said phase to other individuals. Aside from the applications in pulse rate variability, the DPE method can be applied to any physiological signal that varies at the same frequency as a reference signal.

As shown in **Fig. 33**, it is possible to extract phases relative to the breath from different physiological systems—even if they are noisy or have other signals mixed into them. It is evident that the pulse and the phase of the R-R interval are different. This is because, although the R-R interval is extracted from the pulse, breathing has a different phase relationship to the *pulse* specifically than it does to the R-R interval. Although the shapes don't resemble sine waves, plugging them into the DPE method as one would with

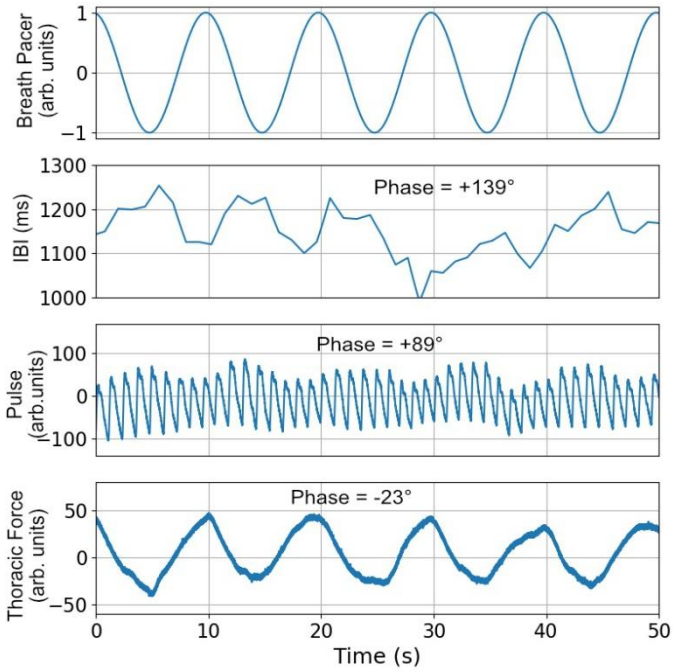


Figure 33. Dynamic Phase Extraction (DPE) can measure the phase and amplitude of the IBI Interval, Pulse Waveform, and Thoracic Force on a chest strap. In general, DPE can measure the phase and amplitude of any signal that is the same frequency as a reference signal. In this case, all phases are calculated with reference to the breath pacer.

the IBI yields the phase of the fundamental Fourier component of the signal relative to the reference signal.

Futhermore, we demonstrate that there is an upward trend in phase as we increase the breathing rate. However, as the breathing rate increases, the phase increases slower, indicating that the change in phase isn't a simple delay in the breathing, but rather, is some more complicated function of breathing. We note that this general trend was not observed in the single female participant, although we cannot determine whether this is due to gender or is simply an outlier. There is also a variation in the phase measured at each breathing rate. Whether this is truly a meaningful variation or an artifact due to measurement noise remains unclear. More research is needed before any conclusions are warranted.

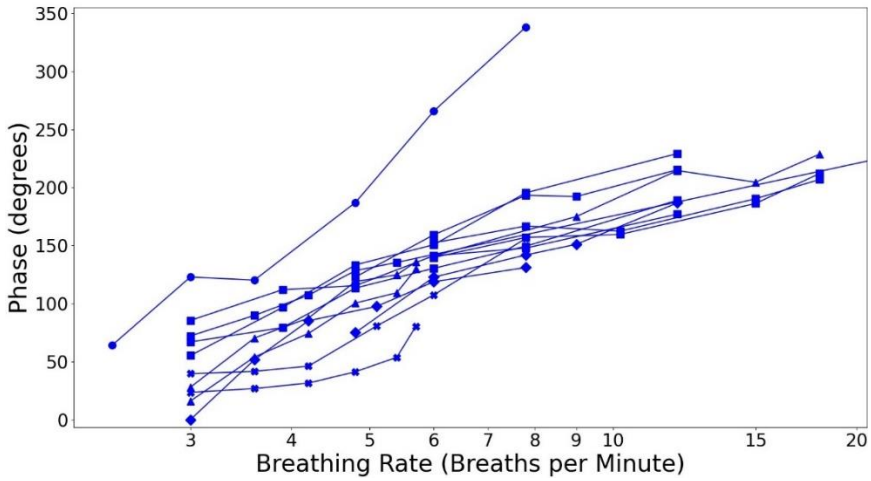


Figure 34. The phase relationships of the population, plotted on a logarithmic x-scale, there is a general upwards trend in phase as breathing gets faster.

Correlations with standard methods

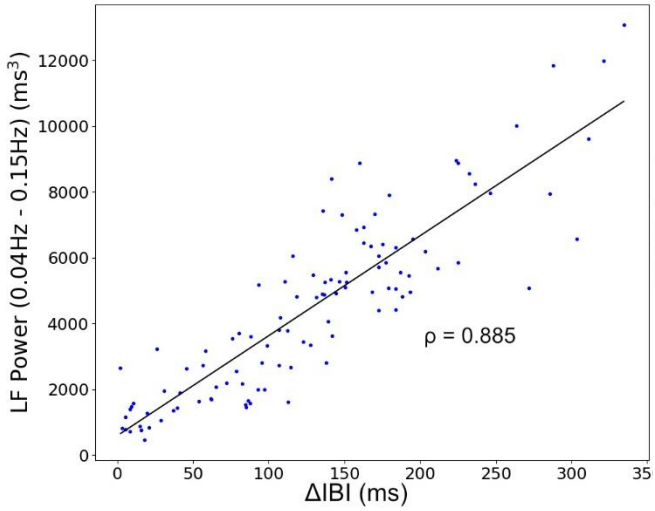


Figure 35. ΔIBI from DPE correlates with the LF Power (0.04Hz-0.15Hz) (ρ = 0.885)

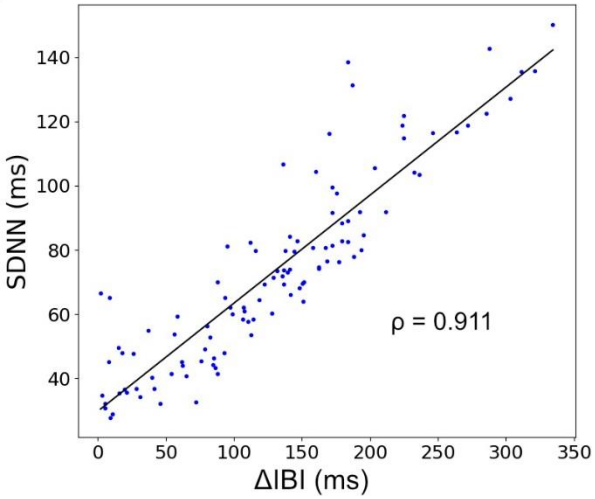


Figure 36. ΔIBI from DPE correlates well with SDNN (ρ = 0.911)

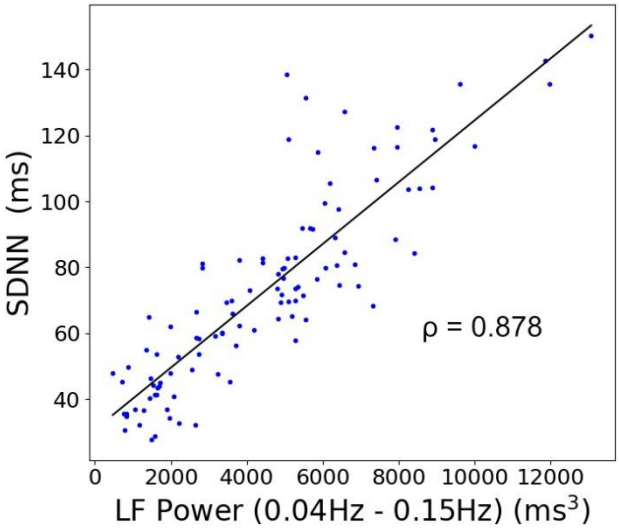


Figure 37. The correlation between two conventional measures of pulse rate variability, LF Power and SDNN, ρ =0.877, is comparable to the correlation of each of them with the novel measure, Dynamic Phase Extraction (DPE) (Figs. 34 and 35).

Our results for ΔIBI correlate well with other standardized measures of pulse rate variability (PRV). Here, we compare our method of analysis to the LF Power, a common frequency domain analysis method for HRV, and the standard deviation of the inter-beat interval (SDNN), a common time-domain analysis method [178, 179]. We performed these comparisons by calculating ΔIBI, SDNN, and LF power for each data set recorded, and

plotting the value obtained for each data record with respect to each other measure obtained from the same data set. The breathing rates were paced but varied, in order to verify that the method generalized well across breathing rates.

Notably, we observed an offset from zero (**Fig. 36)** due to SDNN being based on the standard deviation (which will pick up any variance due to noise in the data), while the DPE naturally filters the noise out. There is also a spread in the data; however, this spread is also present when correlating SDNN to the LF Power (**Fig. 37**). All correlations were assessed via the nonparametric Spearman method and were significant at $p < 0.00001$.

In order to confirm that these correlations were a result of our method's sensitivity to the RSA—and not some other source of variance in the LF band—we took the Fourier

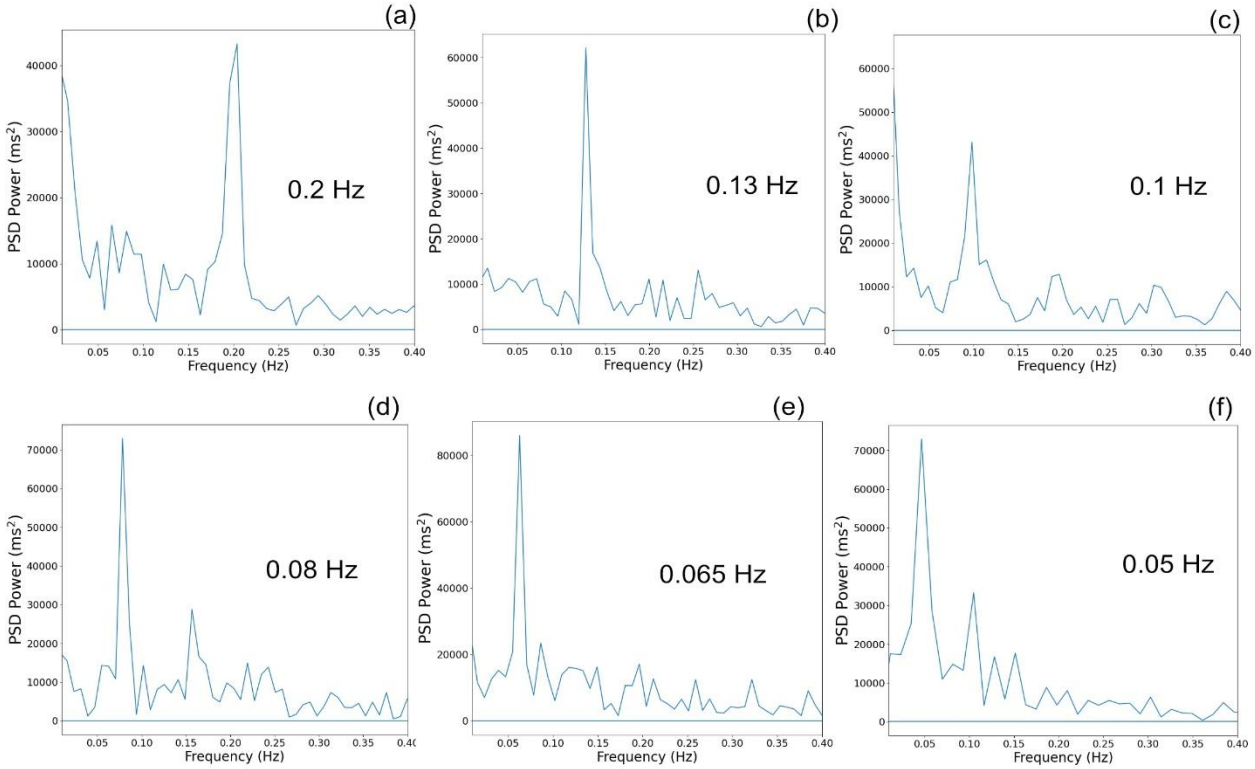


Figure 38. The Fourier transforms of a typical set of data, with the breathing rates indicated on each of the graphs. The dataset shows that the primary source of power comes from frequencies close to the breathing rate. Each subfigure (a-f) represent is a different trail with a breathing rate set to the frequency in the figure.

transform of the pulse rate and verified that the largest peak was at the respiratory rate. **Figure 38** presents a set of power spectra for typical trials.

In most cases, the primary source of power does appear to arise from RSA; however, there is a baseline level of power, and the harmonics are also somewhat visible (**Fig. 38**). Furthermore, at very low breathing rates, power from the VLF band (<0.04 Hz) may contaminate the estimate (although at such low breath frequencies, it becomes very difficult for an individual to keep breathing at the necessary pace).

While there are no standardized methods of phase extraction, one approach for extracting phase relationships utilizes the Fourier transform. In general, the Fourier transform of a time series yields a complex function—the modulus of the values in frequency space gives the power density at that frequency, while the argument of the values in frequency space gives the phase shift of said frequency. In other words, the phase shift, ϕ, can be calculated from the Fourier transform of a function $f(t)$:

$$\phi(\omega) = arctan2\left(\frac{Im\left(\hat{f}(\omega)\right)}{Re\left(\hat{f}(\omega)\right)}\right)$$

Where $\hat{f}(\omega)$ denotes the Fourier transform of a function $f(t)$. Specifically for our case, to find the phase shift of the value in question, we perform a Fourier transform on the pulse rate that we extract from the PPG signal, and then find the peak within $\pm$0.01hz of the breathing rate. We then identify the phase shift of that peak and show that our DPE method is strongly correlated to the traditional Fourier technique (**Fig. 39**).

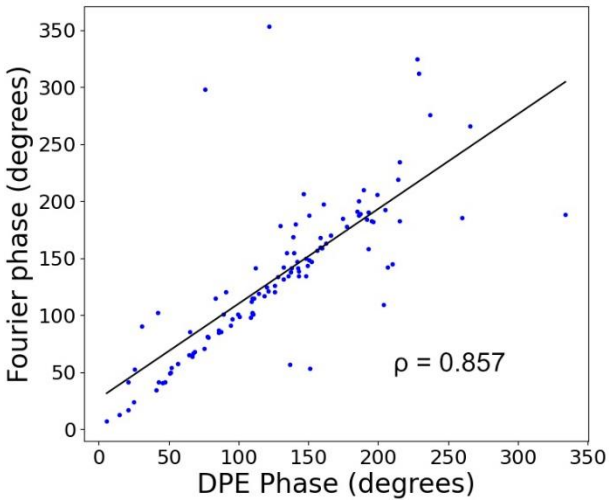


Figure 39. The phase extracted using DPE correlates well with the phase

D. Summary

Dynamic Phase Extraction (DPE) is a powerful methodology that can determine the phase of a signal relative to a reference signal and extract the amplitude of the fundamental Fourier component of the signal at the frequency of the reference signal. There has recently been a resurgence of interest in phase, and several groundbreaking studies have shown that phase can be used to determine the breathing rate that gives the maximum variation in heart rate variability in the LF (0.04-0.15Hz) band [157, 158, 180]. Our results, using DPE analysis of PPG data, agree with previous estimates of where maximum HRV amplitude is in relationship to the phase [157]. The magnitude extracted by DPE also demonstrates a significant correlation to other standardized methods of measuring pulse rate variability, namely SDNN and LF Power (ρ= 0.911 and ρ= 0.885, respectively), and the phase agrees well with that extracted via Fourier transform techniques (ρ=0.857). This suggests that DPE is an effective measurement technique for analyzing changes in inter-beat intervals due to respiratory sinus arrythmia as measured with PPG.

We note that PPG has nontrivial limitations relative to ECG estimates of heart rate variability. Constant et.al have emphasized that *pulse rate variability* (PRV)—an analogue for heart rate variability measured using PPG—is not a surrogate for HRV measured via ECG. In particular, they found that respiratory PRV does not precisely reflect respiratory HRV in standing healthy subjects and in patients with low HRV. Accordingly, Yuda et.al. suggest that it is more appropriate to recognize PRV as a different biomarker than HRV [181]. The literature at large, however, is more conflicted: a number of influential studies have persisted in using *HRV* for PPG-derived signals to emphasize that they are often similar to ECG estimates, especially in the frequency domain [182-191].

There are differences between PPG and ECG sensors, particularly with respect to the RSA spectrum. However, this tends to manifest in an overestimation of the power in the high frequency band [192], while most data taken in this study is in the LF band. Furthermore, other studies have found that PPG and ECG measurements of heart rate/pulse rate variability are well correlated during rest [192, 193]. There is also a delay in where the peaks are detected between PPG and ECG sensors. This delay is on the order of 0.1s [194]. For our application, we are interested in phase delays, and therefore do not investigate non-resting participants. Our ultimate goal is to develop more accessible techniques to assess pulse rate variability. As such, in this context, we believe that PPG is suitable for use.

Additionally, there are some mathematical niceties to consider for the numbers to make sense. The formula $\sqrt{A_{\pi/2}^2 + A_0^2}$ gives the RMS amplitude of the changes in the Inter-beat Interval, with no assumption of the waveform. But if one wants to convert to peak-to-peak amplitude, one needs to know the waveform. For example, the IBI pattern during paced breathing is visually similar to a sine wave, thus we multiply by π, as in **Figures 29** and **32**.

In multiplying by π there is an inherent assumption that the waveform is a sine wave. Strictly speaking, this doesn't change any of the correlation coefficients, but it does link the number generated by DPE to a physical attribute (the peak-to-peak amplitude).

In terms of usage, the primary limitation of this method is that it requires a breath pacer. The most common methods for determining the heart/pulse rate variability are to perform a Fourier transform and integrate the power within the target frequency band (0.04-0.15Hz) [178-180] or to perform time-domain analysis involving some statistical methods, eg SDNN (the standard deviation of the time between heartbeats) [178-180]. These methods do not require a breath pacer, and as such are easier to implement. The primary benefit of using DPE is that it can extract the phase efficiently after a breath is taken, while simultaneously providing a measurement for the amplitude of the signal at the desired frequency, making it ideal for biofeedback.

In this paper we introduce a novel technique, Dynamic Phase Extraction (DPE), that can detect the magnitude (ΔIBI) and phase of a signal relative to a reference signal. ΔIBI correlated well with other standard measures of respiratory sinus arrythmia ($\rho = 0.885$ for LF power and $\rho = 0.911$ for SDNN). We apply this technique to PPG data to determine the magnitude and phase of the IBI, pulse, and thoracic force as measured with a chest strap relative to a breath pacer. In the case of the IBI, the magnitude detected by this novel technique correlates well with pulse rate variability as determined by conventional time and frequency-based techniques. The phases detected by our new technique are consistent with what has been reported with other techniques, and correlates well with the phase extracted using Fourier techniques ($\rho = 0.857$). The phases between various parameters may give

important clues to the detailed operation of the physiological feedback loops that are critical for health.

Work beyond the scope of this initial report on the Dynamic Phase Extraction method would be needed to determine if this method is also useful for ECG data and, if that is successful, to determine how the parameters from Dynamic Phase Extraction in PPG compare to ECG.

VI. Home-use and Portable Biofeedback Lowers Anxiety and Pain in Chronic Pain Subjects

In this study, we investigated the use of novel, home-use and portable biofeedback devices in a remote program for managing chronic pain. In three separate 4-week pilot studies, participants engaged in twice-daily, 10-minute biofeedback sessions, with self-assessed reductions in anxiety and pain levels using the 6-item State-Trait Anxiety Inventory (STAI-6) and Visual Analogue Scale (VAS), respectively in Studies 2 and 3. Among these 113 (Study 2) and 237 (Study 3) biofeedback sessions, 81 (~72%) and 130 (~55%) showed reductions in pain, while 93 (~82%) and 184 (~78%) experienced reductions in anxiety. A positive relationship was found between anxiety and pain reduction, indicating that larger reductions in anxiety correspond to larger reductions in pain. In Study 1, only anxiety reductions were measured: across 143 biofeedback sessions, 127 experienced reductions in anxiety (~89%). Participants in all studies demonstrated reductions in baseline to final results in pain, anxiety, and showed increases in satisfaction and recovery. Our results provide strong evidence that portable biofeedback devices can enhance pain management programs by helping to alleviate anxiety and pain in individuals living with chronic conditions. This study can provide a basis for the integration of biofeedback devices into the expanding research of lifestyle and integrative medicine.

A. Introduction

Chronic pain, impacting millions of individuals globally, stands as a primary contributor to disability and diminished quality of life. However, extensive research efforts are underway to address and alleviate this issue [195-199].

In this paper, we investigate the effectiveness of home-use and portable biofeedback devices in helping to reduce chronic pain. Biofeedback is a mind-body technique that assists users to in gaining conscious control over a physiological process of the body, leading to improved physical and mental health [200-201]. This approach enables the acquisition of self-regulation over physiological responses, thereby fostering relaxation and enhancing well-being.

Many studies have been conducted showing benefits to mindfulness meditation, cognitive therapy, psychophysiologic therapy, and multidisciplinary treatments for chronic pain [202-211]. Biofeedback has been shown to be an effective non-invasive treatment option for chronic pain [200, 212-214], and it has been demonstrated as a highly efficacious intervention for the alleviation of anxiety symptoms [201]. By utilizing biofeedback techniques, patients may experience a reduction in pain intensity and anxiety, and improve well-being.

In a recent controlled study, the “Boulder Study”, 33 out of 50 participants (66%) with long-term lower back pain were randomized to receive 4 weeks of pain reprocessing therapy. At posttreatment, they were pain-free or nearly pain-free [215]. Through pain reprocessing therapy, patients' beliefs about the causes and threat value of pain are shifted. This therapy utilized a combination of cognitive, somatic, and exposure-based techniques to assist patients in reconceptualizing their pain as a result of nondangerous brain activity, rather than peripheral tissue injury The principles of pain reprocessing therapy have been presented in a new book, “The Way Out”, by Alan Gordon [216]. A central component of pain reprocessing therapy involves the reduction of fear and anxiety associated with pain. Our study explores the question: what benefits can be derived from the use of home-based

biofeedback devices combined with a remote group course based on this book for individuals with chronic pain?

Lifestyle medicine, a rapidly developing field of research, has emerged as a systematic approach for managing chronic diseases, with recent investigations focusing specifically on the effects of lifestyle interventions on chronic pain [217-225]. These studies have shown efficacy towards the benefit of lifestyle medicine approach for chronic pain. Biofeedback within lifestyle medicine research has shown benefits for various health conditions [220, 226-227]. A recent paper by Ziya Altug, provides an overview of the benefits of lifestyle medicine strategies to reduce the level of chronic lower back pain including biofeedback [220]. This study serves as a basis for the field of lifestyle medicine and integrative medicine for how home-use and portable biofeedback devices can be integrated into programs.

In this paper, we provide the results of three pilot studies that investigate the effects of novel, home-use and portable biofeedback technology in a 4-week therapy program on anxiety, pain, and satisfaction and recovery in chronic pain subjects.

B. Methods

1. Participants

All participants provided written informed consent as approved by the University of California Institutional Review Board.

Study 1:

A group of 7 participants with chronic lower back pain or fibromyalgia for more than 6 months was recruited from referrals, posters, and emails from the Santa Barbara County

Community. Criteria for inclusion in the study was based on four factors: 1.) Age above 18; 2.) Proximity to UCSB for mailing purposes; 3.) Access to the internet; 4.) Status of experiencing chronic pain for at least 6 months. Demographic information including age, gender, and ethnicity were also collected. Once consented, participants were randomly assigned to complete the biofeedback sessions either the first two weeks of the study, or the latter two weeks of the 4-week study.

Study 2:

A group of 6 primarily chronic lower back pain and fibromyalgia subjects was recruited from referrals, posters, and emails from the Santa Barbara County Community. One subject did not have lower back pain but had lower pelvic pain. Criteria for inclusion in this study was based on 3 factors, all of which were required for a participant to be included: 1.) Age above 18; 2.) Access to the internet and app store; 3.) Status of experiencing chronic pain for at least 6 months. Demographic information including age, gender and ethnicity were also collected.

Study 3:

A group of 9 participants without exclusion to type of chronic pain was recruited from referrals, posters, and emails. Criteria for inclusion in this study was based on 3 factors, all of which were required for a participant to be included: 1.) Age above 18; 2.) Access to the internet and app store; 3.) Status of experiencing chronic pain for at least 6 months. Location was not considered, as the devices were mailed out in this study. Demographic information including age, gender, and ethnicity were also collected.

2. Procedures and Materials

Study 1:

This 4-week study was entirely remote, with a 1 hour Zoom session for the entire group at the beginning and in each following week, for a total of 5 group sessions. A group of 7 participants were provided with a home-use pulse and temperature biofeedback device for 2 weeks of the study and the book "The Way Out" by Alan Gordon [216] for the entire study. The Zoom sessions focused on learning the concepts of the book with PowerPoint presentations, group sharing and discussions.

The two biofeedback methods used in this study were thermal biofeedback and heart rate variability (HRV) biofeedback. Subjects were instructed to perform five minutes of thermal biofeedback and then switch to five minutes of HRV biofeedback in each session. The device displayed the temperature with a six-segment display (Figure 40) using an absolute temperature scale. As the finger temperature increased, segments changed colors in a clockwise direction. Each color set defined a different set of temperature ranges. The heart rate variability biofeedback provided a breath pacer for subjects to follow. It used the interbeat-interval and Dynamic Phase Extraction [228] to measure the heart rate variability, and displayed the integrated magnitude of respiratory sinus arrhythmia. The segments changed in a clockwise direction based on the value of the integrated magnitude.

Anxiety levels were measured on the tablet by a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI-6) [229], which was collected at the beginning and end of each biofeedback session.

In addition to the six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI-6), before and after each biofeedback session, participants completed an online Qualtrics survey before the first group meeting and after the last group

meeting. This survey contained the trait anxiety (STAI) [230], Satisfaction and Recovery Index (SRI) [231], McGill Pain Inventory [232], and questions about maximum, average and minimum pain with a Visual Analogue Scale (VAS) in the preceding week [233].

The custom biofeedback device measured finger temperature with an infrared temperature sensor (MLX 90614) and pulse with photoplethysmography (PPG). The visual biofeedback signal was displayed on Microsoft Surface Pro tablets (**Figure 40**). The microprocessor was a Wemos Lolin ESP32 board that used a serial connection to the tablets to acquire and present the biofeedback signal with a custom python software. The raw temperature and pulse data, together with pulse parameters such as the magnitude and phase of respiratory sinus arrhythmia as computed with Dynamic Phase Extraction [228], were sampled at 100 Hz, stored locally, and transmitted to cloud storage. When the device was connected to the internet, it transmitted the data after each session. If not, the device stored the data and transmitted the data later when the device was connected to the internet.

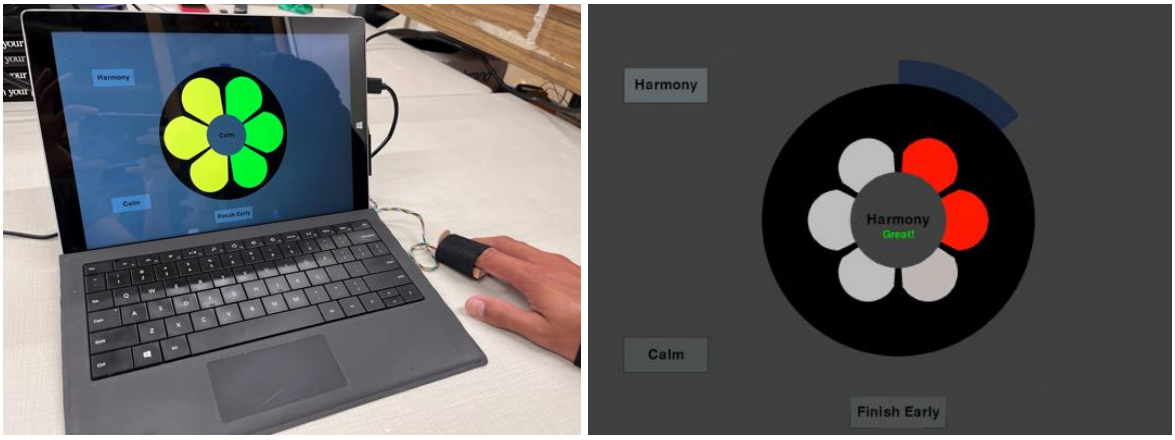


Figure 40. The device used in Study 1 was a home-use biofeedback device that has a temperature and pulse sensor strapped onto the finger. The data acquisition and biofeedback display is performed on a tablet. The left figure shows the display with temperature used to change the color of the segments sequentially. The right figure shows the breath pacer (segments expand and contract) with integrated magnitude of respiratory sinus arrhythmia used to change the color of the segments sequentially.

Study 2:

This 4-week study was entirely remote, with a 1 hour Zoom session for the entire group at the beginning, followed by 1 hour Zoom sessions for the entire group each following week, for a total of 5 group sessions. A group of 6 participants were provided with a portable handheld temperature biofeedback device for the entire study and the book "The Way Out" by Alan Gordon [216]. The Zoom sessions focused on learning the concepts of the book with PowerPoint presentations, group sharing and discussions. These sessions were iterated based on responses from participants in Study 1.

The biofeedback method used in this study was thermal biofeedback. Subjects were instructed to perform ten minutes of thermal biofeedback each session. The device was portable, fit in the subject's hand and displayed the temperature with a six-light display (**Figure 41**) using an absolute temperature scale. As the finger temperature increased, lights changed colors in a clockwise direction. Each color set defined a different set of temperature ranges. Data was collected through a mobile app that included measuring before and after session STAI-6 and VAS, and collected the temperature data.

In addition to the STAI-6 [229] and VAS, before and after each biofeedback session, participants completed an online Qualtrics survey before the first group meeting and after the last group meeting. This survey contained the trait anxiety (STAI), SRI, McGill Pain Inventory [230-233]., and questions about maximum, average and minimum pain in the preceding week that were the same as Study 1.

The custom portable biofeedback device measured hand temperature with an infrared temperature sensor (MLX 90614). The visual biofeedback signal was displayed on the

device with light emitting diodes (LEDs) that represented the temperature measurement. The microprocessor was a Seeed Studio nRF52840 board that used Bluetooth low-energy (BLE) to transmit data to a custom mobile app for data acquisition sampling at 50 Hz.

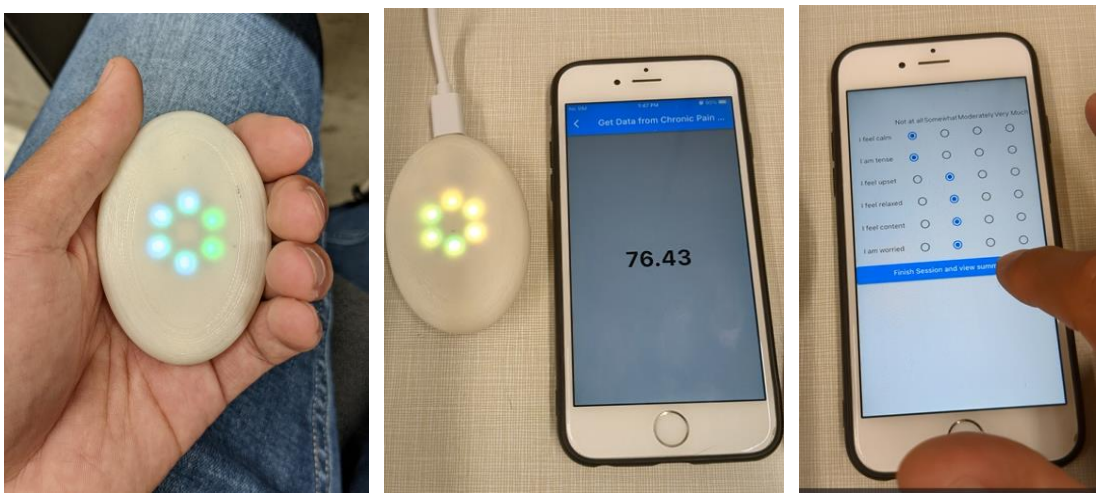


Figure 41. (*Left*) The portable biofeedback devices used in Studies 2 and 3 had a temperature sensor measuring at the palm. (*Middle*) The device paired with a mobile app for temperature data collection. (*Right*) STAI-6 and VAS were recorded before and after each session.

Study 3:

This was a 4-week study with a 1-hour individual meeting in the week leading up to the study which was not required for all participants but was completed by 9 subjects in the group. During this meeting which was held online or in person depending on subject comfortability, the participants were provided with and instructed on how to use the portal handheld temperature biofeedback device and how to download the app for use at home. The rest of the study was remote with a 1 hour Zoom session offered twice a week for the following 5 weeks for a total of 5 group sessions and 1 individual session. Participants were also provided “The Way Out” by Alan Gordon which they were allowed to keep for the entirety of the study [216]. The Zoom sessions focused on learning the concepts of the book

with PowerPoint presentations, group sharing and discussions with minimal iterations made from the previous study, to include terminology encompassing different types of chronic pain, not limited to chronic lower back pain and fibromyalgia.

The biofeedback method and device were the same as in Study 2, and additionally included a breath pacer in the device. The breath pacer was visualized by the increasing and decreasing brightness of the lights and had a 3-second inhalation and 7-second exhalation period [234]. The mobile app was the same, where the data included measuring before and after session STAI-6 and VAS, and collected the temperature data. **Table 8** shows an overview of the biofeedback method and device for each study.

In addition to the STAI-6 [229] and VAS, before and after each biofeedback session, participants completed an online Qualtrics survey before the first group meeting and after the last group meeting. This survey used the trait anxiety (STAI), SRI, McGill Pain Inventory [230-233], and questions about maximum, average and minimum pain in the preceding week that were the same as in Study 1 and Study 2.

	Biofeedback Method	Biofeedback Duration	Biofeedback Device	Biofeedback Survey
Study 1	Temperature and HRV	2 weeks	Finger Insert and Tablet (Home)	STAI-6
Study 2	Temperature	4 weeks	Handheld and Phone (Portable)	STAI-6 and VAS
Study 3	Temperature with breath pacer	4 weeks	Handheld and Phone (Portable)	STAI-6 and VAS

Table 8. Biofeedback overview for use in Studies 1, 2 and 3. The details include the biofeedback method, duration, device, and surveys.

3. Statistical Modeling

Self-reported pain and anxiety measures were analyzed using Bayesian hierarchical models via Stan and brms in R [235-236]. Broadly, these allowed us to assess three hypotheses: 1) that self-reported pain and anxiety decrease when comparing *pre*-biofeedback to *post*-biofeedback ratings, within each biofeedback session; 2) that the *magnitude* of the pre-post difference in anxiety predicts the degree of change in pain, within each session; and 3) that there are longer-term changes in pain and anxiety when comparing survey measures reported at baseline to those at the end of each study period. To assess effects relevant to (1), models were specified according to the following formulae (using Wilkinson notation): y ~ 1 + prePost + (1 + prePost | subjectID), allowing for both random intercepts and random slopes across participants (under the assumption that each participant's pre-biofeedback ratings may be correlated with the degree of pre-post change). Similarly, for (2), models were specified according to: deltaPain ~ 1 + deltaAnxiety + (1 + deltaAnxiety | subjectID). And finally, for (3), we collapsed across all three studies (because each study had a small *N* and, unlike the previous two cases estimated over repeated biofeedback sessions, each participant had only two values per survey measure) using a nested random effects structure over the model intercept to account for inter-study variability: y ~ 1 + prePost + (1 | studyID / subjectID).

For consistency, and to avoid the possibility that any given model could be unduly influenced by arbitrary prior specifications, several key parameters for all models were estimated under the same set of weakly-informative priors:

Intercept ~ Normal(0, 5)

Slopes ~ Normal(0, 2.5)

Random effect *SD* ~ Half-Cauchy(0, 0.5)

When correlated random effects were present, as in (1) and (2) above, we retained the default priors given by brms: Random effect $r \sim$ LKJ(1)

Lastly, we also retained the default, data-dependent priors over the residual variances, which followed: $\sigma \sim$ Half-Student-t(3, 0, 2.5) if the median absolute deviation (*MAD*) of the outcome measure was ≤ 2.5, or $\sigma \sim$ Half-Student-t(3, 0, *MAD*) otherwise.

Robust exploration of the posterior space for all models was performed using Hamiltonian Monte Carlo, using four independent chains with 15000 iterations each (5000 of which were used as warm-up samples). We ensured that all models properly converged to equilibrium for all parameters using classical benchmarks including: the effective sample size (considering the autocorrelation between independent posterior draws); the Monte Carlo standard error (relative to the posterior *SD*); *R*-hat (the variance ratio between each chain relative to *all* chains); and no divergent Monte Carlo transitions after warm-up. For statistical inference, we report 95% credibility intervals (i.e., using the highest posterior density interval) around the posterior median parameter estimates along with the 'probability of direction', capturing the proportion of the posterior density in the hypothesized direction, above or below zero (thus, this can be thought of as a Bayesian analogue to the frequentist *p*-value; [43]. Models (1) and (3) also include measures of effect size (analogous to Cohen's *d*) and Bayes factors (BF10) derived using the Savage-Dickey ratio of the posterior against a point-null prior estimate of zero effect. Effect sizes were computed by taking the posterior draws for each predictor and dividing them by the square root of the summed residual variances and random effects variances (as per convention for hierarchical linear models; [238], yielding a posterior distribution for each effect size. The models designed for (2) above report effect sizes using R^2 for both the full model and the marginal R^2 capturing

variance that can be attributed to the fixed effects alone [239]. We additionally checked for outliers using a robust criterion over the residuals: if the median absolute deviation of each individual residual was greater than three times the total *MAD* over all residuals, these data points were excluded and the model was re-fit. This resulted in 2 data points (1.77%) being removed for model (2) in Study 2 and 8 data points (3.38%) for model (2) in Study 3.

C. Results

We first tested the hypothesis that anxiety and pain would be reduced when comparing *pre*-biofeedback ratings to *post*-biofeedback ratings, using the Bayesian hierarchical models described under **Statistical Modeling**. In Study 1, we observed extremely strong evidence for a large reduction in anxiety (d = -1.01, 95CI = [-1.30, -0.72], BF10 > 1000) under the pre-post contrast (b = -3.25, SD = 0.38, 95CI = [-3.99, -2.48]) with a 100% probability of being negative (i.e., such that anxiety was lower following biofeedback). Similarly, in Study 2, we observed moderate evidence for a fairly-large reduction in anxiety (d = -0.72, 95CI = [-1.27, -0.12], BF10 = 5.96; b = -2.26, SD = 0.88, 95CI = [-3.86, -0.33]) with a 98.67% probability of being negative. Study 3 also showed extremely strong evidence for a moderate reduction in anxiety following biofeedback (d = -0.65, 95CI = [-0.99, -0.31], BF10 > 100; b = -3.05, SD = 0.65, 95CI = [-4.29, -1.69]) with a 99.97% probability of being negative. **Figure 42** highlights the posterior predictive densities for each of these models.

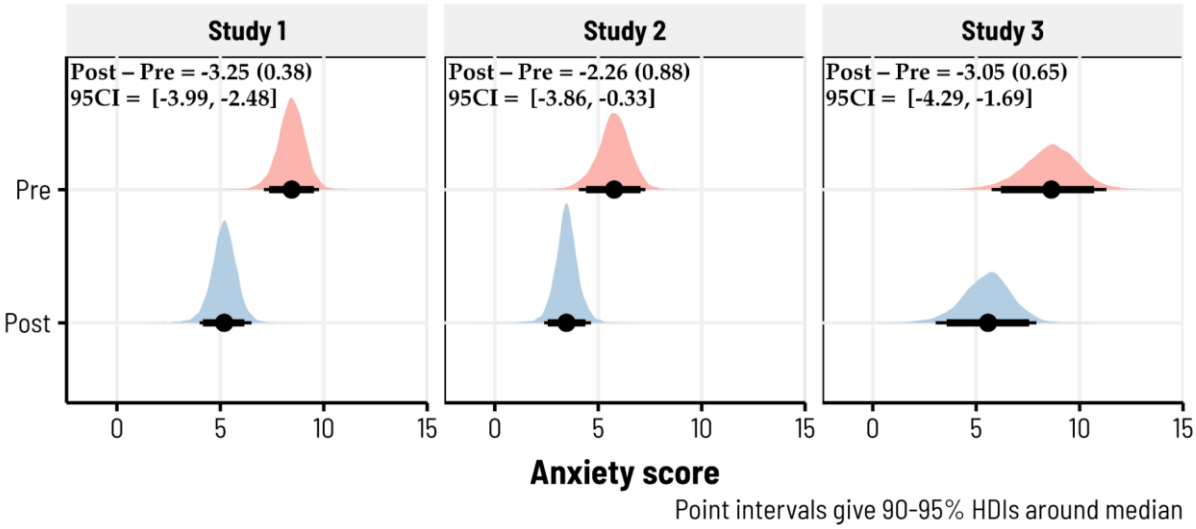


Figure 42. 10-minute biofeedback sessions reduced anxiety across all three studies. Posterior predictive densities, derived via Bayesian hierarchical models, are shown for each study—giving the predicted estimates of pre- and post-biofeedback anxiety ratings. Anxiety was self-reported using the STAI-6, which ranged from 0 to 18. In these 493 biofeedback sessions (totaled across studies), 404 had reductions in anxiety (~82%). The median pre/post differences, along with the posterior standard deviations and 95% credibility intervals, are given in the top left of each subplot.

Changes in pain before and after biofeedback were assessed for Studies 2 and 3 (as we did not collect these ratings in Study 1). We again observed strong evidence for a moderate reduction in pain in Study 2 (*d* = -0.69, 95CI = [-1.06, -0.28], BF10 = 15.17; *b* = -1.03, *SD*.= 0.27, 95CI = [-1.55, -0.46]) with a 99.79% probability of being negative. In Study 3, where we admitted participants across a wider spectrum of chronic pain types, there was a more modest reduction in pain after biofeedback, although we still observed strong evidence against the null hypothesis of no effect (*d* = -0.33, 95CI = [-0.54, -0.14], BF10 = 24.86; *b* = -0.76, *SD* = 0.20, 95CI = [-1.16, -0.37]) and a 99.91% probability of being negative. The posterior predictive densities for each model are shown in **Figure 43**.

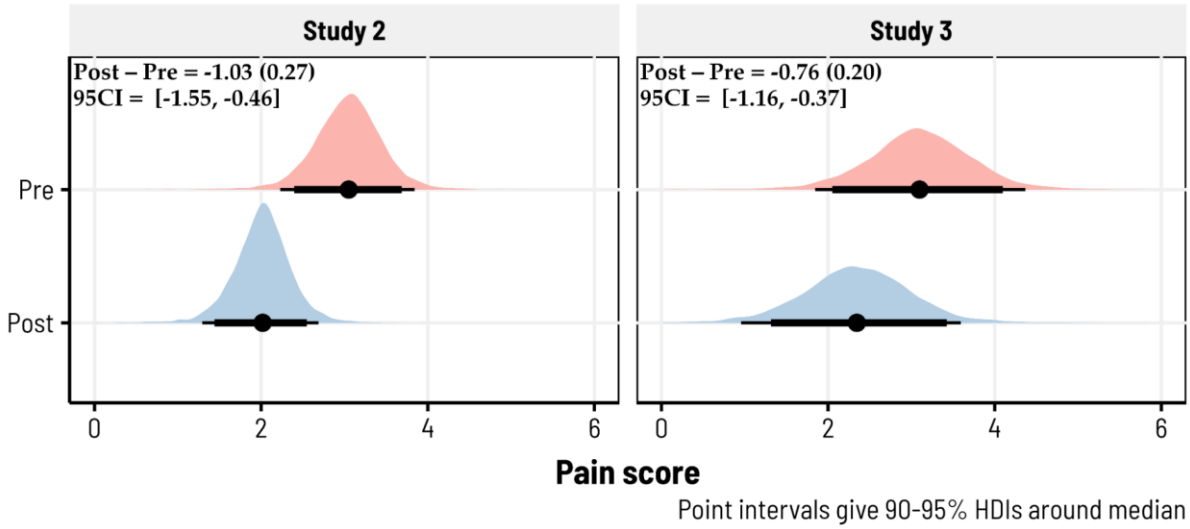


Figure 43. 10-minute biofeedback sessions reduced pain across Studies 2 and 3. Posterior predictive densities, derived via Bayesian hierarchical models, are shown for each study—giving the predicted estimates of pre- and post-biofeedback pain ratings. Pain was self-reported using an 11-point visual analogue scale (VAS), which ranged from 0 to 10. In these 350 biofeedback sessions (totaled across studies), 211 had reductions in pain (~60%). The median pre/post differences, along with the posterior standard deviations and 95% credibility intervals, are given in the top left of each subplot.

We then sought to test the hypothesis that the magnitude of the change in anxiety (pre-/post-biofeedback) would predict the magnitude of the change in pain, such that larger reductions in anxiety would correspond to larger reductions in pain. This was again assessed for Studies 2 and 3. Indeed, this is what we found in Study 2 (b = 0.27, SD = 0.09, 95CI = [0.09, 0.44]), with a 99.49% probability of the slope being positive (i.e., indicating that larger changes in anxiety were accompanied by larger changes in pain). The model had substantial explanatory power considering the fixed and random effects together (R^2 = 0.70, 95CI = [0.64, 0.75], adjusted R^2 = 0.64), with the change in anxiety specifically accounting for over half of the variance (Marginal R^2 = 0.57, 95CI = [0.18, 0.71]). These trends are shown in **Figure 44** for the overall, group-level fit, and **Figure 45** highlights individual

variation in slopes. In Study 3, we observed similar, albeit slightly weaker, trends (b = 0.11, SD = 0.03, 95CI = [0.04, 0.17]), still with a 99.63% probability of being positive. Given the greater range of pain histories represented in Study 3, this model also had weaker total explanatory power relative to Study 2 (R^2 = 0.35, 95CI = [0.26, 0.43]), adjusted R^2 = 0.30) with the change in anxiety captured by the marginal R^2 accounting for 15% of the total variance (95CI = [6.42 x 10^{-7}, 0.28]). The group-level fits and individualized slopes for Study 3 are illustrated in **Figures 44** and **45**, respectively.

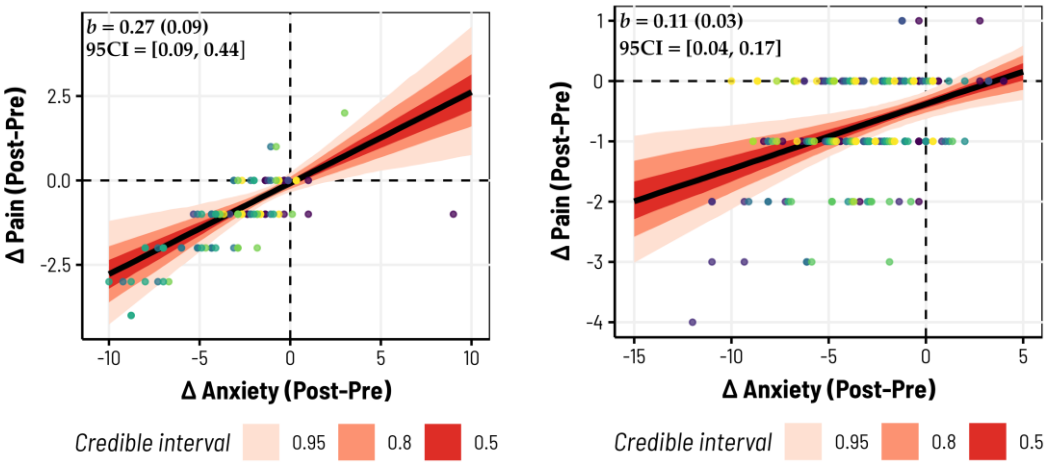


Figure 44. Changes in anxiety before/after biofeedback predict changes in pain. In both Study 2 (*left*) and Study 3 (*right*), the pre/post difference in self-reported anxiety following each biofeedback session corresponded with changes in pain, such that larger reductions in anxiety were accompanied by larger reductions in pain. Here we plot the raw data entered into each Bayesian hierarchical model, where different colored datapoints correspond to different subjects in each study. The population-level slopes are displayed with various uncertainty intervals (shaded in red); in the top left, we provide the posterior median estimates, their standard deviations, and 95% credibility intervals.

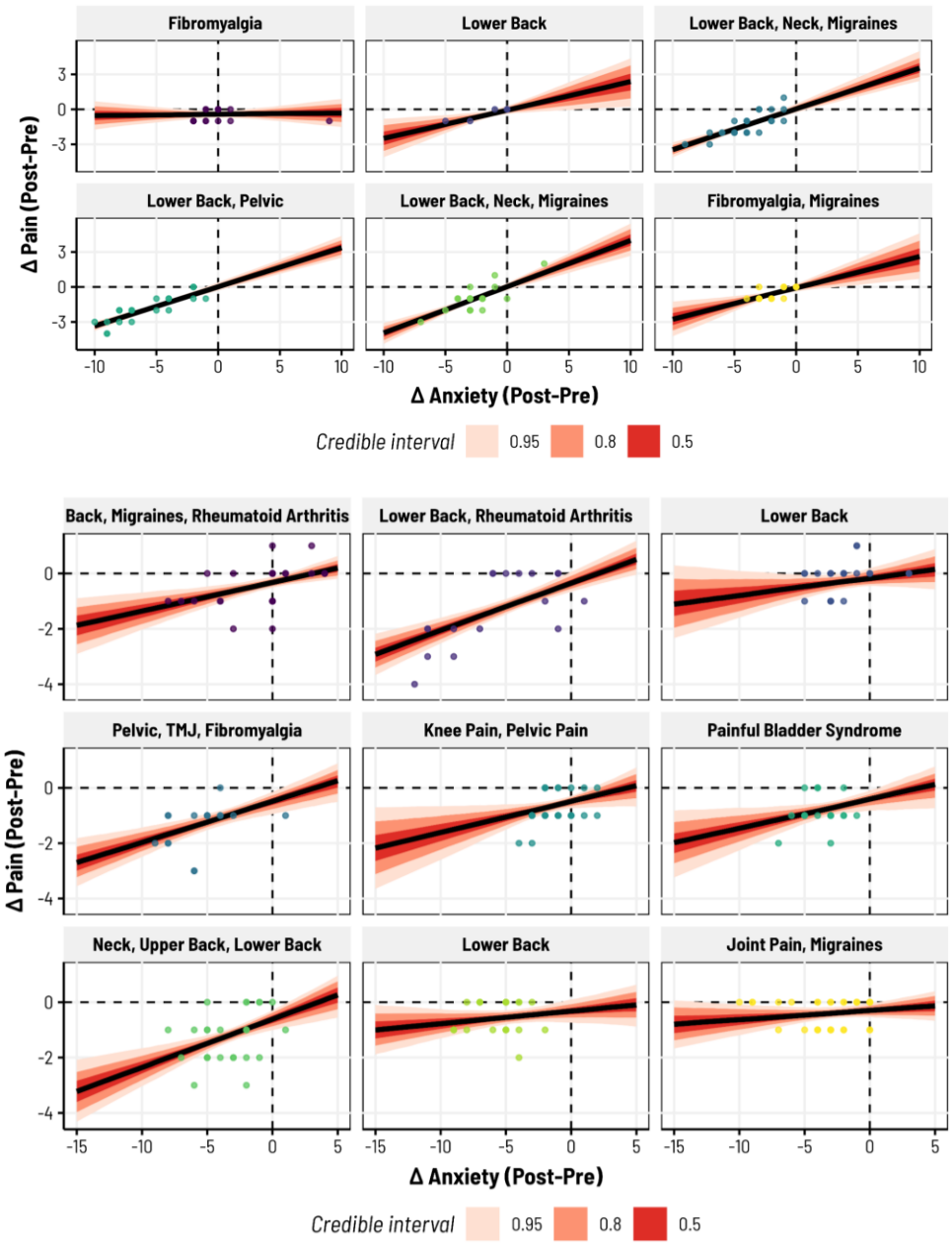


Figure 45. Inter-subject variability in the associations between anxiety and pain, before and after biofeedback. Here we display subject-specific estimates for the relationships between anxiety and pain reductions—that is, the *random effects* from the Bayesian hierarchical models shown in **Figure 44**. Subjects with many types of chronic pain were included in Studies 2 (*top*) and 3 (*bottom*), even subjects with Rheumatoid Arthritis, in which there was structural pain from a physical cause. For all these types of chronic pain,

10-minute biofeedback sessions usually decreased pain and anxiety (data points in the lower left quadrant).

Finally, we examined longer-term changes in pain and anxiety by comparing several measures recorded at baseline and at the end of the study period, including STAI reports of anxiety; the McGill Pain Inventory (MPI); minimum, average, and maximum pain levels; and the Satisfaction & Recovery Index (SRI). As described previously, these measures were collapsed across studies to maximize power, while using nested random effects structures to account for variability between studies. Consistent with our pre-/post-biofeedback models, we observed strong evidence for a longer-term reduction in anxiety (*d* = -0.39, 95CI = [-1.05, -0.03], BF10 = 25.21; *b* = -4.01, *SD* = 1.38, 95CI = [-6.65, -1.23]) with a 99.71% probability of being negative. We also observed moderate evidence for a reduction in pain as per the MPI, although we note potential uncertainty in the estimated effect size and contrast estimate, which have credibility intervals lightly intersecting zero (*d* = -0.30, 95CI = [-0.61, 0.001], BF10 = 4.35; *b* = -2.90, *SD* = 1.42, 95CI = [-5.56, 0]). Still, for MPI, the contrast posterior suggests a 97.49% probability of being negative. With respect to participants' minimum, average, and maximum pain ratings, models consistently revealed moderate-to-strong evidence for a reduction in each. For minimum pain, the change relative to baseline was large (*d* = -0.88, 95CI = [-1.45, -0.31], BF10 = 25.63; *b* = -1.57, *SD* = 0.49, 95CI = [-2.51, -0.61]) with a 99.88% probability of being negative. For average pain, the reduction was slightly lower in magnitude (-*d* = -0.74, 95CI = [-1.28, -0.21], BF10 = 8.83; *b* = -1.19, *SD* = 0.41, 95CI = [-2.00, -0.37]) with a 99.66% probability of being negative. And for maximum pain, we also observed a slightly lower reduction on average (*d* = -0.52, 95CI = [-0.93, -0.13], BF10 = 4.51; *b* = -0.98, *SD* = 0.36, 95CI = [-1.68, -0.27]) with a 99.45% probability of being negative. However, despite these differences, we did not see evidence

for a consistent increase in SRI across the three study periods, which was considerably more variable than the other measures assessed (*d* = 0.04, 95CI = [-0.04, 0.15], BF10 = 1.81; *b* = 2.60, *SD* = 2.39, 95CI = [-2.12, 7.30]) and showed only an 86.16% probability of being positive. **Figure 46** provides a summary of the posterior differences for each of these effects and **Table 9** provides a summary of the average measures for each study.

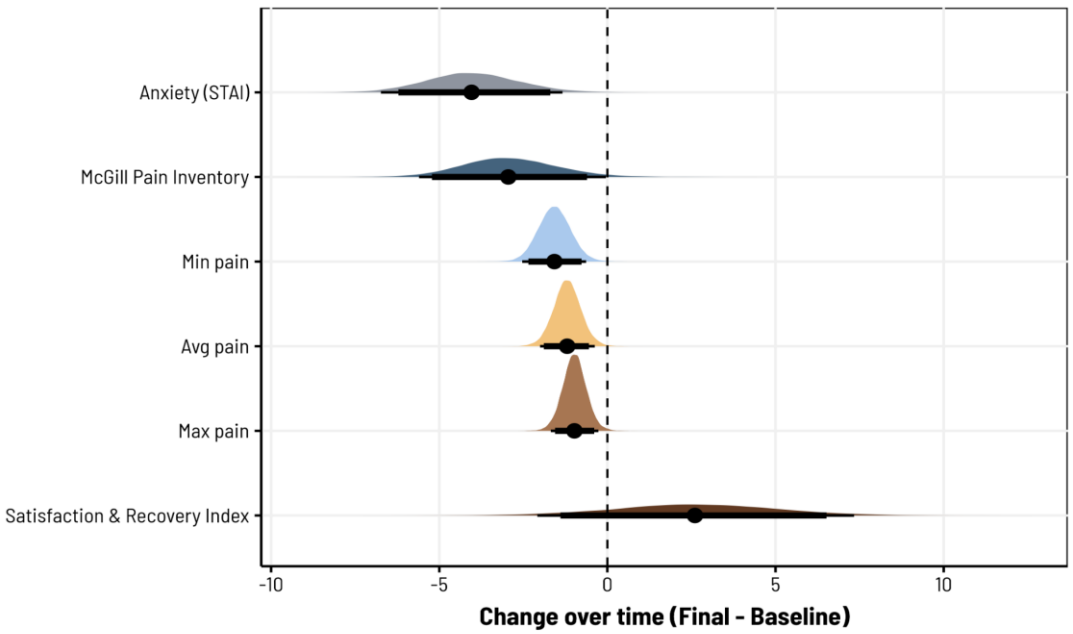


Figure 46. For all three studies, participants showed reductions in pain and anxiety together with increases in satisfaction and recovery from the baseline results before the study to the final results after the study. The measures obtained were maximum, average and minimum pain using VAS, McGill Pain survey, trait anxiety from STAI, and SRI.

	Study 1	Study 2	Study 3	Combined
Max Pain Baseline	8.0 ± 0.6	6.8 ± 2.3	7.4 ± 2.1	7.4 ± 1.7
Max Pain Final	7.4 ± 1.0	5.8 ± 1.0	6.2 ± 1.9	6.5 ± 1.3
Avg Pain Baseline	5.4 ± 2.1	4.2 ± 1.8	4.6 ± 1.3	4.7 ± 1.7
Avg Pain Final	3.4 ± 0.5	3.4 ± 0.8	3.4 ± 1.6	3.4 ± 1.0

Min Pain Baseline	3.6 ± 2.8	2 ± 1.5	3.4 ± 1.4	3.0 ± 1.9
Min Pain Final	0.6 ± 0.8	1.0 ± 0.8	1.4 ± 1.8	1.0 ± 1.1
McGill Pain Baseline	22.2 ± 5.5	16.8 ± 8.7	17.2 ± 9.5	18.7 ±7.9
McGill Pain Final	15.4 ± 4.3	10.6 ± 9.6	14.4 ± 9.8	13.5 ± 7.9
STAI Baseline	24.6 ± 4.3	25.8 ± 1.6	27 ± 5.6	25.8 ± 3.8
STAI Final	15.4 ± 3.8	20.8 ± 6.9	20 ± 4.9	18.7 ± 5.2
SRI Baseline	53.3 ± 15.8	59.3 ± 19.7	41.8 ± 23.6	51.5 ± 19.7
SRI Final	72.5 ± 11.1	64.2 ± 9.5	64.0 ± 18.5	66.9 ± 13.0

Table 9. Average baseline and final survey results for studies 1, 2 and 3. The measures obtained were maximum, average, and minimum pain using VAS, McGill Pain survey, trait anxiety from STAI, and SRI.

D. Summary

Significant decreases in anxiety and pain were observed, with respective posterior medians in Studies 1, 2, and 3 being -3.25, -2.26, -3.05 for anxiety, and -1.03, -0.76 in Studies 2 and 3 for pain. Additionally, a positive correlation was noted between anxiety and pain reductions, with posterior medians of 0.27 and 0.11 in Studies 2 and 3, supporting the hypothesis that the neural circuits that create the experience of pain include not only pain, but also anxiety. The findings highlight the strong efficacy of biofeedback, with reductions in anxiety and pain levels observed with a probability exceeding 98.5%, and the correlation between them exceeding a 99.4% probability. Among these 113 (Study 2) and 237 (Study 3) biofeedback sessions, 81 (~72%) and 130 (~55%) showed reductions in pain, while 93 (~82%) and 184 (~78%) had reductions in anxiety. In Study 1, only anxiety reductions were measured: across 143 biofeedback sessions, 127 had reductions in anxiety (~89%). The results provide strong evidence that portable biofeedback devices could enhance management programs by helping to alleviate anxiety and pain in individuals living with

chronic pain. Limitations of this study include the small participant sample sizes of each study, the lack of control to compare against, and the remote group interactions with participants.

One of the advantages of thermal biofeedback is that subjects gain an increased sense of agency: their ability to act independently and effectively control their own lives. Individuals acquire knowledge regarding their capacity to modulate blood flow, leading to an enhanced sense of self-efficacy that extends to other areas of physiological and psychological control. This increased awareness enables individuals to recognize and harness their ability to exert influence over various aspects of their physiological responses, which may have previously gone unrecognized.

Moreover, thermal biofeedback devices have several advantages. Firstly, thermal feedback eliminates the need for costly equipment, making it a cost-effective solution. Secondly, its portable nature allows for increased convenience, enabling researchers to gather a larger amount of data per subject. Lastly, this versatility enables its utilization in a wide range of settings, further enhancing its practicality and applicability.

A question raised by this work is: How can the temporary reductions in anxiety best be used to help people overcome chronic pain and other problems? Possibilities include: 1. Reducing anxiety before, during or after somatic tracking. 2. Reducing anxiety in situations that might otherwise trigger chronic pain. 3. Reducing anxiety before, during or after physical therapy or cognitive therapy.

As the field of lifestyle medicine and integrative medicine grows, the devices in this study can be used as a tool for research collaborations to investigate the effects of biofeedback devices in combination with other modalities for various health conditions.

Utilizing the capabilities of portable biofeedback devices in remote settings, future research should investigate various methods of integrating biofeedback into programs and engage larger and more diverse samples to further validate these promising results.

VII. A Novel Thermal Biofeedback Device to Enhance Sleep Quality and Alleviate Anxiety

In this study, we investigated the use of a novel, handheld thermal biofeedback device in a study measuring sleep quality and a study measuring anxiety levels. In the 2-week pilot sleep study, participants engaged in once-nightly, 10-minute thermal biofeedback sessions. The 8 participants completed the Pittsburgh Sleep Quality Index (PSQI) before and after the study showing 7 (87.5%) improved PSQI scores with an average of 29.79% ($p < 0.05$) improvement in sleep quality. Daytime dysfunction, the ability to stay awake and maintain enthusiasm, was seen to improve an average of 58.33% ($p < 0.01$). Among the 74 biofeedback sessions there was an average of 87.5% of sessions that had an increase in hand temperature. In the pilot anxiety study, 10 participants engaged in three 10-minute thermal biofeedback sessions on separate days within one week. Among the 30 biofeedback sessions, 25 (83.3%) showed reductions in state anxiety levels with an average of 22.1% ($p < 0.001$) reduction before and after the session using the State-Trait Anxiety Inventory (STAI-10). Our results provide evidence that portable biofeedback devices can help to enhance overall sleep quality and alleviate anxiety. This study provides a foundation for using a thermal biofeedback device in lifestyle medicine.

A. Introduction

Despite major advances in sleep and anxiety research, sleep disturbances and anxiety-related disorders still exist widely today [240-242]. The convergence of psychophysiological research and emerging technologies presents an opportunity to develop non-pharmacological interventions that could help the way we understand, manage, and ultimately improve the

landscape of sleep, anxiety and mental health. The current study explores the potential therapeutic effects of a newly developed thermal biofeedback device, hypothesized to enhance sleep quality and alleviate symptoms of anxiety.

Biofeedback mechanisms, employing real-time displays of physiological processes, have shown promising results in therapeutic contexts, enabling individuals to develop control over these processes and hence enhance their well-being. Among various modalities of biofeedback, thermal biofeedback, which provides immediate feedback on skin temperature changes, has been found particularly useful in relaxation training and stress management. However, the extent to which thermal biofeedback can be employed to improve sleep quality remains largely underexplored with limited recent research [243-245].

Sleep is a complex physiological process, with temperature regulation playing a crucial role in its onset and maintenance [246]. Subtle changes in core and peripheral body temperatures signal the body's transition into different sleep stages. Therefore, the proposed thermal biofeedback device holds potential for non-invasive enhancement of sleep quality.

Similarly, given that anxiety has been linked to aberrations in physiological arousal, including temperature fluctuations, the same device could also be harnessed to manage anxiety symptoms [247]. By teaching individuals to regulate their thermal responses—often associated with anxiety—this device might foster self-regulation skills, thereby reducing anxiety levels.

The current paper presents findings from two pilot studies aimed at evaluating the efficacy of this novel thermal biofeedback device. The outcomes of interest are improvements in sleep quality, as assessed by self-reported sleep indices, and reductions in anxiety symptoms, as measured by validated anxiety scales. The results of this study may

not only shed light on the therapeutic potential of the thermal biofeedback device but also contribute to the broader literature on the application of biofeedback technologies in sleep, anxiety and mental health.

B. Methods

1. Participants

All participants provided written informed consent as approved by the University of California Institutional Review Board.

Sleep Study:

A group of 8 participants was recruited from emails to university students. Criteria for inclusion in the study was based on four factors: 1.) Age above 18; 2) Access to the internet; 3) Owning a mobile phone. Demographic information including age, gender, and ethnicity were also collected.

Anxiety Study:

A group of 10 participants was recruited from emails to university students. Criteria for inclusion in the study was based on four factors: 1.) Age above 18; 2) Access to the internet; 3) Owning a mobile phone. Demographic information including age, gender, and ethnicity were also collected. Once consented, participants were instructed to use the handheld thermal biofeedback device in a 10-minute session.

2. Procedures and Materials

The biofeedback device was portable, fit in the subject's hand and displayed the temperature with a six-light display (Figure 1) using an absolute temperature scale. As the finger temperature increased, lights changed colors in a clockwise direction. Each color set

defined a different set of temperature ranges. Data was collected through a mobile app that included measuring the temperature data.

Figure 1. The thermal biofeedback device and mobile app used for the study.

Sleep Study:

This 2-week study was entirely remote. A group of 8 participants were instructed to use the handheld thermal biofeedback device in a 10-minute session prior to sleeping each night. Subjects recorded biofeedback sessions by connecting the device to a mobile app. Sleep quality was measured with the Pittsburgh Sleep Quality Index (PSQI) before and after the study [248].

Anxiety Study:

A group of 10 participants were instructed to come into an office space on three separate days within one week. In each visit, participants were instructed to use the handheld thermal biofeedback device in a 10-minute session and input their before and after session

anxiety survey using STAI-10 Y1 [249-251]. Subjects recorded biofeedback sessions by connecting the device to a mobile app. Before the start and after the end of the study, participants were also asked to input their trait anxiety using STAI-10 Y2.

C. Results

Sleep Study:

We first tested the hypothesis that sleep quality would be improved when comparing baseline to final PSQI scores [248]. Through calculating the average of the baseline and final scores, we observed an average difference of 1.75 reduction. This reduction corresponds to a 29.79% improvement in sleep quality and can be seen in Table 10. We used a Wilcoxon signed-rank test here since our sample data is paired and not normally distributed. The results of the statistical analysis indicated a significant reduction in PSQI scores between the baseline and final assessments ($p < 0.02846$).

Utilizing the same Wilcoxon test model, we want to examine the differences within each category of PSQI in Table 11. The individual categories of the PSQI are Duration of Sleep (DURAT), Sleep Disturbance (DISTB), Sleep Latency (LATEN), Day Dysfunction Due to Sleepiness (DAYDYS), Sleep Efficiency (HSE), Overall Sleep Quality (SLPQUAL), and Need Medication to Sleep (MEDS). The results of the statistical test demonstrated a significant reduction in Daytime Dysfunction Due to Sleepiness (DAYDYS) between the baseline and final assessments ($p < 0.005367$). No statistically significant reduction in other PSQI categories is observed.

PSQI Scores

Subject	Pre	Post	Difference
1	5	3	-2
2	5	3	-2
3	7	2	-5
4	6	8	2
5	3	1	-2
6	12	11	-1
7	3	1	-2
8	6	4	-2
Average	5.875	4.125	-1.75

Table 10: PSQI scores to measure changes between baseline and final sleep quality scores. There is an average 29.79% improvement in sleep quality after the two week study.

PSQI Category Score Differences

SUBJECT	DURAT	DISTB	LATEN	DAYDYS	HSE	SLPQUAL	MEDS
1	0	1	-1	-1	-1	0	0
2	0	0	1	-1	0	-1	-1
3	-2	0	-1	-1	0	-1	0
4	1	0	1	-1	1	0	0
5	0	0	0	-1	0	-1	0
6	1	0	-2	0	0	0	0
7	-1	0	0	-1	0	0	0
8	0	-1	0	-1	0	0	0
p-value	N/A	N/A	N/A	< 0.01	N/A	N/A	N/A

Table 11: PSQI category differences between baseline and final sleep quality scores. There is moderate evidence to support that biofeedback improves daytime dysfunction, the ability to stay awake and maintain enthusiasm.

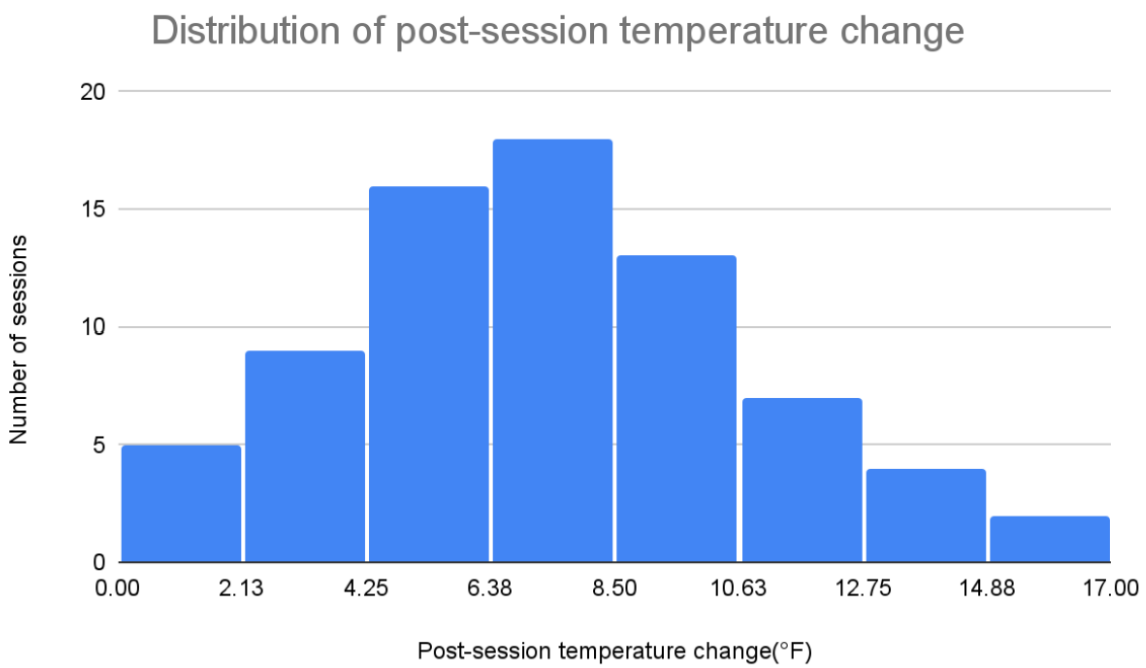


Figure 2. Temperature change distribution before and after each biofeedback session for the sleep study participants.

Anxiety Study:

Among the 30 biofeedback sessions, 25 (83.3%) showed reductions in anxiety levels with an average of 22.1% reduction before and after the session using the State-Trait Anxiety Inventory (STAI-10 Y1) [249-251]. The results of a Wilcoxon signed-rank test indicated a significant reduction in state anxiety scores between the baseline and final assessments ($p < 0.00002888$).

We tested the hypothesis that trait anxiety would be reduced when comparing *pre*-biofeedback ratings to *post*-biofeedback ratings. The average reduction for the trait-anxiety score is 2.3 points, which reflects a 8% reduction on overall anxiety levels. The results of a Wilcoxon signed-rank test indicated that the reduction in trait anxiety scores was not significant.

By using the Wilcoxon signed-rank test, we found the post-biofeedback STAI ratings are significantly smaller than the pre-biofeedback STAI ratings($p < 0.00002888$). We also

analyzed the data in each category of the State-Trait Anxiety Inventory (STAI-10). Our findings revealed significant reductions in various aspects of individuals' emotional states. Specifically, there were significant reductions in the categories of calm ($p < 0.00012$), tense ($p < 0.000066$), ease ($p < 0.0005$), misfortunes ($p = 0.0066$), nervous ($p < 0.02965$), jittery ($p < 0.0061$), relaxed ($p < 0.000035$), worried ($p = 0.011$), and steady ($p < 0.00093$).

STAI Category Significance

Category	p-value
Calm	< 0.001
Tense	< 0.001
Ease	< 0.001
Misfortunes	< 0.01
Frightened	N/A
Nervous	< 0.05
Jittery	< 0.01
Relaxed	< 0.001
Worried	< 0.05
Steady	< 0.001

Table 12: STAI-Y1 category differences between baseline and final anxiety scores. Use of the thermal biofeedback device shows evidence to support that there is improvement in most categories of the anxiety inventory.

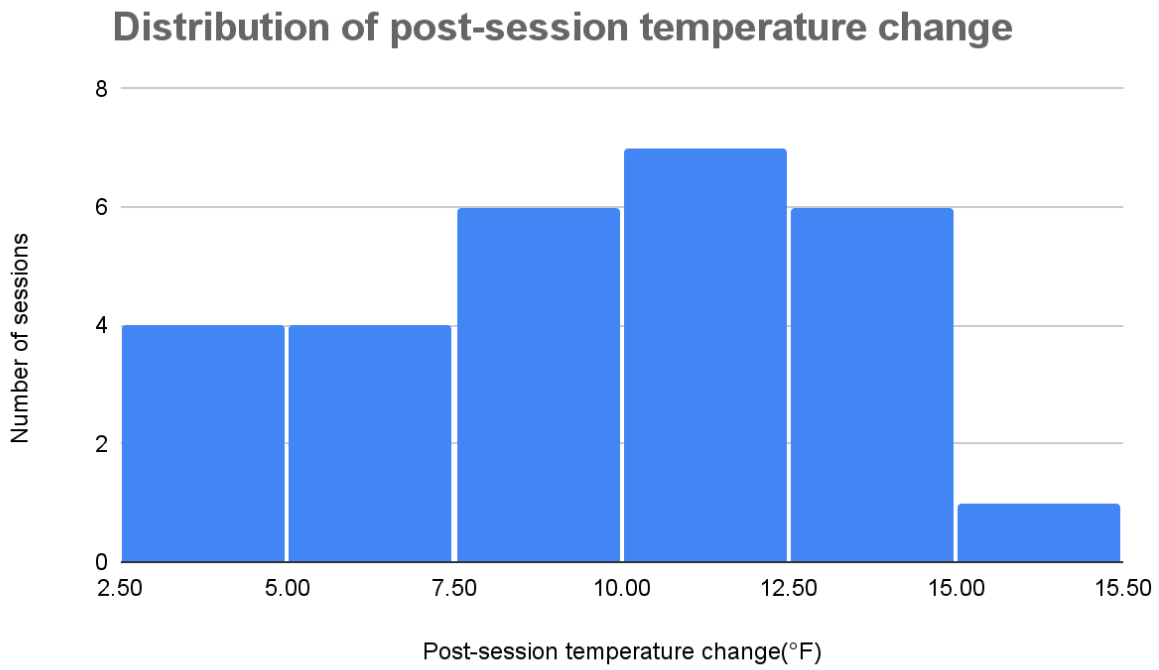


Figure 3. Temperature change distribution before and after each biofeedback session for the anxiety study participants.

D. Summary

Our investigation primarily sought to determine if there was a discernible improvement in sleep quality when comparing initial and concluding PSQI scores. On analyzing the data, we noted a considerable reduction in the average score by 1.75 points. This change translates to an improvement in sleep quality by nearly 29.79%. Given that our dataset comprised paired samples that did not adhere to a normal distribution, we employed the Wilcoxon signed-rank test to validate our findings. The results from this analysis revealed a statistically significant improvement in sleep quality from the baseline to the final assessment, as indicated by the p-value of less than 0.02846.

In a secondary examination, using the same Wilcoxon test model, we delved deeper into individual categories within the PSQI to discern specific areas of change. Here, our

results highlighted a significant amelioration in daytime dysfunction between the initial and concluding evaluations, denoted by a p-value of less than 0.005367.

The marked improvement in sleep quality overall, and specifically in the category of daytime dysfunction, the ability to stay awake and maintain enthusiasm, underscores the effectiveness of whatever intervention or conditions were present during the study period. This change in daytime dysfunction is particularly noteworthy as it directly influences an individual's daily activities, social engagements, and overall well-being. Further studies may be essential to pinpoint the exact mechanisms or interventions driving these improvements and to test the durability of these effects over more extended periods.

Our study focused on the efficacy of biofeedback sessions in mitigating anxiety levels as gauged by the State-Trait Anxiety Inventory (STAI-10 Y1). A significant majority, i.e., 83.3% of the biofeedback sessions, resulted in a reduction of anxiety, with an average decrement of 22.1%. Utilizing the Wilcoxon signed-rank test, our analysis divulged a statistically significant reduction in state anxiety scores between the baseline and subsequent evaluations, denoted by a p-value of less than 0.00002888.

We further explored the potential effect of biofeedback sessions on trait anxiety. However, despite an observed average decline of 2.3 points (equating to an 8% overall reduction), this decrease in trait anxiety scores did not achieve statistical significance. This suggests that while biofeedback might have an immediate effect on situational anxiety (state anxiety), its impact on the more enduring and stable component of anxiety (trait anxiety) requires further elucidation.

A more granular exploration of individual categories within the State-Trait Anxiety Inventory (STAI-10) illuminated significant shifts in various emotional domains post-

biofeedback. Concretely, categories encompassing calmness, tension, ease, perception of misfortunes, nervousness, jitteriness, relaxation, worry, and steadiness all reflected statistically significant reductions.

These findings accentuate the potency of biofeedback as a viable tool for the amelioration of certain aspects of anxiety, especially in the realm of state anxiety and various emotional facets therein. There are significant limitations to these studies, namely, the number of participants and lack of controls. However, these pilot studies provide preliminary evidence for researchers to further investigate the effects of this novel thermal biofeedback device.

One of the key benefits of thermal biofeedback is that it empowers individuals with a heightened sense of autonomy, bolstering their confidence in managing their own lives. Through this method, they gain insights into their potential to regulate blood flow. This newfound understanding not only elevates their self-belief in bodily control but also magnifies their perception of their psychological influence. With this enhanced consciousness, they can tap into their potential to manipulate various physiological reactions, many of which they might have been previously unaware of.

In addition, thermal biofeedback devices offer several notable advantages. To begin with, they remove the requirement for expensive equipment, presenting a more economical option. Their compact design also means greater portability, facilitating more comprehensive data collection from each participant. This adaptability ensures their efficacy in a plethora of environments, amplifying their utility and relevance.

References

1. D. Bridges, C. Randall, and P.K. Hansma, "A new device for performing reference point indentation without a reference probe," *Review of Scientific Instruments*, vol. 83, no. 4, p. 044301, (2012).

2. C. Randall et al., "Applications of a new handheld reference point indentation instrument measuring bone material strength," *Journal of Medical Devices*, vol. 7, no. 4, p. 041005, (2013).

3. A. Diez-Perez et al., "Recommendations for a standard procedure to assess cortical bone at the tissue-level in vivo using impact microindentation," *Bone Reports*, vol. 5, pp. 181-185, (2016).

4. JN Farr et al., "In vivo assessment of bone quality in postmenopausal women with type 2 diabetes," *J Bone Miner Res*, vol. 29, no. 4, pp. 787–95, (2014).

5. J.N. Farr and S. Khosla, "Determinants of bone strength and quality in diabetes mellitus in humans," *Bone*, vol. 82, pp. 28-34, (2016).

6. L. Mellibovsky et al., "Bone tissue properties measurement by reference point indentation in glucocorticoid-induced osteoporosis," *Journal of Bone and Mineral Research*, vol. 30, no. 9, pp. 1651-1656, (2015).

7. F. Malgo et al., "Bone material strength as measured by microindentation in vivo is decreased in patients with fragility fractures independently of bone mineral density," *The Journal of Clinical Endocrinology & Metabolism*, vol. 100, no. 5, pp. 2039-2045, (2015).

8. M.R. Allen et al., "True gold or pyrite: a review of reference point indentation for assessing bone mechanical properties in vivo," *Journal of Bone and Mineral Research*, vol. 30, no. 9, pp. 1539-1550, (2015).

9. J.N. Farr, S. Amin, and S. Khosla, "Regarding 'True Gold or Pyrite: A Review of Reference Point Indentation for Assessing Bone Mechanical Properties In Vivo'," *Journal of Bone and Mineral Research*, vol. 30, no. 12, pp. 2325-2326, (2015).

10. M.R. Allen et al., "Response to Comments on 'True Gold or Pyrite: A Review of Reference Point Indentation for Assessing Bone Mechanical Properties In Vivo'," *Journal of Bone and Mineral Research*, vol. 30, no. 12, pp. 2327-2327, (2015).

11. K. Hoffseth, C. Randall, P. Hansma, and H.T. Yang, "Study of indentation of a sample equine bone using finite element simulation and single cycle reference point indentation," *Journal of the Mechanical Behavior of Biomedical Materials*, vol. 42, pp. 282-291, (2015).

12. A. Idkaidek and I. Jasiuk, "Modeling of Osteoprobe indentation on bone," *Journal of the Mechanical Behavior of Biomedical Materials*, vol. 90, pp. 365-373, (2019).

13. S. Hamada, M. Nakanishi, T. Moriyama, and H. Noguchi, “Re-Examination of correlation between hardness and tensile properties by numerical analysis,” *Experimental Mechanics*, vol. 57, no. 5, pp. 773-781, (2017).

14. E. Broitman, “Indentation hardness measurements at macro-, micro-, and nanoscale: a critical overview,” *Tribology Letters*, vol. 65, no. 1, p. 23, (2017).

15. S. M. Walley, “Historical origins of indentation hardness testing,” *Materials Science and Technology*, vol. 28, nos. 9-10, pp. 1028-1044, (2012).

16. A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. -, no. -, pp. 1097-1105, (2012).

17. R. Collobert and J. Weston, “A unified architecture for natural language processing: Deep neural networks with multitask learning,” *Proceedings of the 25th international conference on Machine learning*, vol. -, no. -, pp. 160-167, (2008).

18. A. Graves, A.-r. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” *2013 IEEE international conference on acoustics, speech and signal processing*, vol. -, no. -, pp. 6645-6649, (2013).

19. D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton et al., “Mastering the game of go without human knowledge,” *Nature*, vol. 550, no. 7676, p. 354, (2017).

20. A. Avati, K. Jung, S. Harman, L. Downing, A. Ng, and N. H. Shah, “Improving palliative care with deep learning,” *BMC medical informatics and decision making*, vol. 18, no. 4, p. 122, (2018).

21. J. Dahlhamer, J. Lucas, C. Zelaya, R. Nahin, S. Mackey, L. DeBar, R. Kerns, M. Von Korff, L. Porter, and C. Helmick, “Prevalence of chronic pain and high-impact chronic pain among adults united states, 2016,” *Morbidity and Mortality Weekly Report*, vol. 67, no. 36, p. 1001, (2018).

22. D. J. Gaskin and P. Richard, “The economic costs of pain in the united states,” *The Journal of Pain*, vol. 13, no. 8, pp. 715-724, (2012).

23. M. M. van der Miesen, M. A. Lindquist, and T. D. Wager, “Neuroimaging-based biomarkers for pain: state of the field and current directions,” *Pain reports*, vol. 4, no. 4, (2019).

24. M. C. Reddan and T. D. Wager, “Brain systems at the intersection of chronic pain and self-regulation,” *Neuroscience letters*, vol. 702, pp. 24-33, (2019).

25. M. C. Reddan and T. D. Wager, “Modeling pain using fmri: from regions to biomarkers,” *Neuroscience bulletin*, vol. 34, no. 1, pp. 208-215, (2018).

26. R. Cowen, M. K. Stasiowska, H. Laycock, and C. Bantel, “Assessing pain objectively: the use of physiological markers,” *Anaesthesia*, vol. 70, no. 7, pp. 828-847, (2015).

27. Y. L. Yang, H. S. Seok, G.-J. Noh, B.-M. Choi, and H. Shin, “Postoperative pain assessment indices based on photoplethysmography waveform analysis,” *Frontiers in physiology*, vol. 9, p. 1199, (2018).

28. L. S. Prichep, J. Shah, H. Merkin, and E. M. Hiesiger, “Exploration of the pathophysiology of chronic pain using quantitative eeg source localization,” *Clinical EEG and neuroscience*, vol. 49, no. 2, pp. 103-113, (2018).

29. H. Liu, Y. Wang, and L. Wang, “A review of non-contact, low-cost physiological information measurement based on photoplethysmographic imaging,” *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. -, no. -, pp. 2088-2091, (2012).

30. J. A. Bargh and E. Morsella, “The unconscious mind,” *Perspectives on psychological science*, vol. 3, no. 1, pp. 73-79, (2008).

31. R. Kanawade, S. Tewary, H. Sardana et al., “Photoplethysmography based arrhythmia detection and classification,” *2019 6th International Conference on Signal Processing and Integrated Networks (SPIN)*, vol. -, no. -, pp. 944-948, (2019).

32. R. Poplin, A. V. Varadarajan, K. Blumer, Y. Liu, M. V. McConnell, G. S. Corrado, L. Peng, and D. R. Webster, “Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning,” *Nature Biomedical Engineering*, vol. 2, no. 3, p. 158, (2018).

33. H. R. Darabi, D. Tsinis, K. Zecchini, W. F. Whitcomb, and A. Liss, “Forecasting mortality risk for patients admitted to intensive care units using machine learning,” *Procedia Computer Science*, vol. 140, pp. 306-313, (2018).

34. B. Jin, H. Yang, L. Sun, C. Liu, Y. Qu, and J. Tong, “A treatment engine by predicting next-period prescriptions,” *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, vol. -, no. -, pp. 1608-1616, (2018).

35. M.-H. Hsieh, M.-J. Hsieh, C.-M. Chen, C.-C. Hsieh, C.-M. Chao, and C.-C. Lai, “An artificial neural network model for predicting successful extubation in intensive care units,” *Journal of clinical medicine*, vol. 7, no. 9, p. 240, (2018).

36. 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, (2017).

37. P. Rodriguez, G. Cucurull, J. Gonzalez, J. M. Gonfaus, K. Nasrollahi, T. B. Moeslund, and F. X. Roca, "Deep pain: Exploiting long short-term memory networks for facial expression classification," *IEEE transactions on cybernetics*, (2017).

38. D. Liu, F. Peng, A. Shea, R. Picard et al., "Deepfacelift: interpretable personalized models for automatic estimation of self-reported pain," arXiv preprint arXiv:1708.04670, (2017).

39. Y. Chu, X. Zhao, J. Han, and Y. Su, "Physiological signal-based method for measurement of pain intensity," *Frontiers in neuroscience*, vol. 11, p. 279, (2017).

40. J. Lotsch and A. Ultsch, "Machine learning in pain research," *Pain*, vol. 159, no. 4, p. 623, (2018).

41. M. H. Pitcher, M. Von Korff, M. C. Bushnell, and L. Porter, "Prevalence and profile of high-impact chronic pain in the united states," *The Journal of Pain*, vol. 20, no. 2, pp. 146–160, (2019).

42. A. E. Johnson, M. M. Ghassemi, S. Nemati, K. E. Niehaus, D. A. Clifton, and G. D. Clifford, "Machine learning and decision support in critical care," *Proceedings of the IEEE. Institute of Electrical and Electronics Engineers*, vol. 104, no. 2, p. 444, (2016).

43. Z. M. Hira and D. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Advances in bioinformatics*, vol. 2015, (2015).

44. N. Ben-Israel, M. Kliger, G. Zuckerman, Y. Katz, and R. Edry, "Monitoring the nociception level: a multi-parameter approach," *Journal of clinical monitoring and computing*, vol. 27, no. 6, pp. 659–668, (2013).

45. U. Rubins, Z. Marcinkevics, I. Logina, A. Grabovskis, and E. Kviesis-Kipge, "Imaging photoplethysmography for assessment of chronic pain patients," in Optical Diagnostics and Sensing XIX: Toward Point-of-Care Diagnostics, vol. 10885. *International Society for Optics and Photonics*, (2019).

46. A. Johansson, P. A. Oberg, and G. Sedin, "Monitoring of heart and respiratory rates in newborn infants using a new photoplethysmographic technique," *Journal of clinical monitoring and computing*, vol. 15, no. 7-8, pp. 461–467, (1999).

47. A. Bonissi, R. D. Labati, L. Perico, R. Sassi, F. Scotti, and L. Sparagino, "A preliminary study on continuous authentication methods for photoplethysmographic biometrics," in 2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications. *IEEE*, (2013), pp. 28–33.

48. pulseSensor.com, Accessed May 26, 2020.

49. I. Zuzarte, P. Indic, D. Sternad, and D. Paydarfar, "Quantifying movement in preterm infants using photoplethysmography," *Annals of biomedical engineering*, vol. 47, no. 2, pp. 646–658, (2019).

50. S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” arXiv preprint arXiv:1502.03167, (2015).

51. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: a simple way to prevent neural networks from overfitting,” *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, (2014).

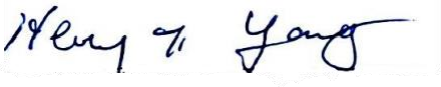
52. L. Prechelt, “Early stopping-but when?” in Neural Networks: Tricks of the trade. *Springer*, (1998), pp. 55–69.

53. D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, (2014).

54. S. Åkerblom, S. Perrin, M.R. Fischer, and L.M. McCracken, “Treatment outcomes in group-based cognitive behavioural therapy for chronic pain: An examination of PTSD symptoms,” *European Journal of Pain*, vol. 24, pp. 807–817, (2020).

55. K.R. Ambrose and Y.M. Golightly, “Physical exercise as non-pharmacological treatment of chronic pain: Why and when,” *Best Practice & Research Clinical Rheumatology*, vol. 29, pp. 120–130, (2015).

56. Y.K. Ashar, A. Gordon, H. Schubiner, et al., “Effect of Pain Reprocessing Therapy vs Placebo and Usual Care for Patients With Chronic Back Pain,” *JAMA Psychiatry*, vol. 79, pp. 13, (2022).

57. S.A. Aytur, K.L. Ray, S.K. Meier, et al., “  ,” *Frontiers in Human Neuroscience*, vol. 15, (2021).

58. M.N. Baliki, D.R. Chialvo, P.Y. Geha, et al., “Chronic Pain and the Emotional Brain: Specific Brain Activity Associated with Spontaneous Fluctuations of Intensity of Chronic Back Pain,” *Journal of Neuroscience*, vol. 26, pp. 12165–12173, (2006).

59. M.N. Baliki, P.Y. Geha, A.V. Apkarian, and D.R. Chialvo, “Beyond Feeling: Chronic Pain Hurts the Brain, Disrupting the Default-Mode Network Dynamics,” *Journal of Neuroscience*, vol. 28, pp. 1398–1403, (2008).

60. E.F. Ball, E. Nur Shafina Muhammad Sharizan, G. Franklin, and E. Rogozińska, “Does mindfulness meditation improve chronic pain? A systematic review,” *Current Opinion in Obstetrics & Gynecology*, vol. 29, pp. 359–366, (2017).

61. C.C. Barney, R.D. Andersen, R. Defrin, L.M. Genik, B.E. McGuire, and F.J. Symons, “Challenges in pain assessment and management among individuals with intellectual and developmental disabilities,” *PAIN Reports*, vol. 5, e821, (2020).

62. N. Becker, P. Sjøgren, P. Bech, A.K. Olsen, and J. Eriksen, “Treatment outcome of chronic non-malignant pain patients managed in a Danish multidisciplinary pain

centre compared to general practice: a randomised controlled trial," *Pain*, vol. 84, pp. 203–211, (2000).

63. N. Ben-Israel, M. Kliger, G. Zuckerman, Y. Katz, and R. Edry, "Monitoring the nociception level: a multi-parameter approach," *Journal of Clinical Monitoring and Computing*, vol. 27, pp. 659–668, (2013).

64. J.M. Bland and D.G. Altman, "Comparing methods of measurement: why plotting difference against standard method is misleading," *The Lancet*, vol. 346, pp. 1085–1087, (1995).

65. D. Borsook and L.R. Becerra, "Breaking down the Barriers: fMRI Applications in Pain, Analgesia and Analgesics," *Molecular Pain*, vol. 2, 1744-8069-2–30, (2006).

66. A.L.H. Buffington, C.A. Hanlon, and M.J. McKeown, "Acute and persistent pain modulation of attention-related anterior cingulate fMRI activations," *Pain*, vol. 113, pp. 172–184, (2005).

67. E. Campbell, A. Phinyomark, and E. Scheme, "Feature Extraction and Selection for Pain Recognition Using Peripheral Physiological Signals," *Frontiers in Neuroscience*, vol. 13, (2019).

68. Y. Chu, X. Zhao, J. Han, and Y. Su, "Physiological Signal-Based Method for Measurement of Pain Intensity," *Frontiers in Neuroscience*, vol. 11, (2017).

69. Y. Chu†, X. Zhao, J. Yao†, Y. Zhao, and Z. Wu, "Physiological Signals Based Quantitative Evaluation Method of the Pain," *IFAC Proceedings Volumes*, vol. 47, pp. 2981–2986, (2014).

70. J.F. Coleman, "Spring Forest Qigong and Chronic Pain," *Journal of Holistic Nursing*, vol. 29, pp. 118–128, (2011).

71. K.F. Cook, W. Dunn, J.W. Griffith, et al., "Pain assessment using the NIH Toolbox," *Neurology*, vol. 80, S49–S53, (2013).

72. A. Corry, G. Linssen, and P. Spinhoven, "Multimodal treatment programmes for chronic pain: A quantitative analysis of existing research data," *Journal of Psychosomatic Research*, vol. 36, pp. 275–286, (1992).

73. R. Cowen, M.K. Stasiowska, H. Laycock, and C. Bantel, "Assessing pain objectively: the use of physiological markers," *Anaesthesia*, vol. 70, pp. 828–847, (2015).

74. R. B. Cutler, D. A. Fishbain, H. L. Rosomoff, E. Abdel-Moty, T. M. Khalil, and R. S. Rosomoff, "Does Nonsurgical Pain Center Treatment of Chronic Pain Return Patients to Work?" Spine (Phila Pa 1976), vol. 19, pp. 643–652, (1994).

75. J. Dahlhamer, J. Lucas, C. Zelaya, R. Nahin, S. Mackey, L. DeBar, R. Kerns, M. Von Korff, L. Porter, and C. Helmick, "Prevalence of Chronic Pain and High-Impact

Chronic Pain Among Adults — United States, 2016," *MMWR Morbidity and Mortality Weekly Repo*rt, vol. 67, pp. 1001–1006, (2018).

76. R. Daoust, J. Paquet, É. Piette, K. Sanogo, B. Bailey, and J-M. Chauny, "Impact of Age on Pain Perception for Typical Painful Diagnoses in the Emergency Department," *The Journal of Emergency Medicine*, vol. 50, pp. 14–20, (2016).

77. K. D. Davis, H. Flor, H. T. Greely, G. D. Iannetti, S. Mackey, M. Ploner, A. Pustilnik, I. Tracey, R-D. Treede, and T. D. Wager, "Brain imaging tests for chronic pain: medical, legal and ethical issues and recommendations," *Nature Reviews Neurology*, vol. 13, pp. 624–638, (2017).

78. M. A. Day, L. C. Ward, D. M. Ehde, B. E. Thorn, J. Burns, A. Barnier, J. B. Mattingley, and M. P. Jensen, "A Pilot Randomized Controlled Trial Comparing Mindfulness Meditation, Cognitive Therapy, and Mindfulness-Based Cognitive Therapy for Chronic Low Back Pain," *Pain Medicine*, vol. 20, pp. 2134–2148, (2019).

79. W. W. Deardorff, H. S. Rubin, and D. W. Scott, "Comprehensive multidisciplinary treatment of chronic pain: a follow-up study of treated and non-treated groups," *Pain*, vol. 45, pp. 35–43, (1991).

80. M. W. Donnino, G. S. Thompson, S. Mehta, M. Paschali, P. Howard, S. B. Antonsen, L. Balaji, S. M. Bertisch, R. Edwards, L. H. Ngo, A v. Grossestreuer, "Psychophysiologic symptom relief therapy for chronic back pain: a pilot randomized controlled trial," *PAIN Report*s, vol. 6, e959, (2021).

81. C. Eccleston, L. Hearn, and A. C. de C. Williams, "Psychological therapies for the management of chronic neuropathic pain in adults," *Cochrane Database of Systematic Reviews*, (2015).

82. C. Eccleston, T. M. Palermo, A. C. de C. Williams, A. Lewandowski Holley, S. Morley, E. Fisher, and E. Law, "Psychological therapies for the management of chronic and recurrent pain in children and adolescents," *Cochrane Database of Systematic Reviews*, (2014).

83. M. Eriksson, H. Storm, A. Fremming, and J. Schollin, "Skin conductance compared to a combined behavioural and physiological pain measure in newborn infants," *Acta Paediatrica,* vol. 97, pp. 27–30, (2007).

84. A. Fahrenkamp, L. Sim, L. Roers, M. Canny, T. Harrison, and C. Harbeck-Weber, "An Innovative and Accessible Biofeedback Intervention for Improving Self-Regulatory Skills in Pediatric Chronic Pain: A Pilot Study," *The Journal of Alternative and Complementary Medicine*, vol. 26, pp. 212–218, (2020).

85. S. B. Finn, B. N. Perry, J. E. Clasing, L. S. Walters, S. L. Jarzombek, S. Curran, M. Rouhanian, M. S. Keszler, L. K. Hussey-Andersen, S. R. Weeks, P. F. Pasquina, and

J. W. Tsao, "A Randomized, Controlled Trial of Mirror Therapy for Upper Extremity Phantom Limb Pain in Male Amputees," *Frontiers in Neurology*, vol. 8, (2017).

86. E. Fisher, E. Law, J. Dudeney, C. Eccleston, and T. M. Palermo, "Psychological therapies (remotely delivered) for the management of chronic and recurrent pain in children and adolescents," *Cochrane Database of Systematic Reviews*, (2019).

87. P. Flodin, S. Martinsen, R. Altawil, E. Waldheim, J. Lampa, E. Kosek, and P. Fransson, "Intrinsic Brain Connectivity in Chronic Pain: A Resting-State fMRI Study in Patients with Rheumatoid Arthritis," *Frontiers in Human Neuroscience*, vol. 10, (2016).

88. H. Flor, T. Fydrich, and D. C. Turk, "Efficacy of multidisciplinary pain treatment centers: a meta-analytic review," *Pain*, vol. 49, pp. 221–230, (1992).

89. D. J. Gaskin and P. Richard, "The Economic Costs of Pain in the United States," The *Journal of Pain*, vol. 13, pp. 715–724, (2012).

90. R. J. Gatchel and A. Okifuji, "Evidence-Based Scientific Data Documenting the Treatment and Cost-Effectiveness of Comprehensive Pain Programs for Chronic Nonmalignant Pain," *The Journal of Pain*, vol. 7, pp. 779–793, (2006).

91. H. R. Gilpin, A. Keyes, D. R. Stahl, R. Greig, and L. M. McCracken, "Predictors of Treatment Outcome in Contextual Cognitive and Behavioral Therapies for Chronic Pain: A Systematic Review," *The Journal of Pain*, vol. 18, pp. 1153–1164, (2017).

92. S. Gruss, M. Geiger, P. Werner, O. Wilhelm, H. C. Traue, A. Al-Hamadi, and S. Walter, "Multi-Modal Signals for Analyzing Pain Responses to Thermal and Electrical Stimuli," *Journal of Visualized Experiments*, (2019).

93. 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, e0140330, (2015). doi:10.1371/journal.pone.0140330.

94. J. Guzman, “Multidisciplinary rehabilitation for chronic low back pain: systematic review,” *BMJ*, vol. 322, pp. 1511–1516, (2001). doi:10.1136/bmj.322.7301.1511.

95. N. Henschke, R.W. Ostelo, M.W. van Tulder, J.W. Vlaeyen, S. Morley, W.J. Assendelft, and C.J. Main, “Behavioural treatment for chronic low-back pain,” *Cochrane Database of Systematic Reviews*, (2010). doi:10.1002/14651858.CD002014.pub3.

96. L. Hilton, S. Hempel, B.A. Ewing, E. Apaydin, L. Xenakis, S. Newberry, B. Colaiaco, A.R. Maher, R.M. Shanman, M.E. Sorbero, and M.A. Maglione, “Mindfulness Meditation for Chronic Pain: Systematic Review and Meta-analysis,” *Annals of Behavioral Medicine*, vol. 51, pp. 199–213, (2017). doi:10.1007/s12160-016-9844-2.

97. C. Holmberg, J. Rappenecker, J. Karner, and C.M. Witt, “The perspectives of older women with chronic neck pain on perceived effects of qigong and exercise therapy on aging: a qualitative interview study,” *Clinical Interventions in Aging*, (2014):403. doi:10.2147/CIA.S54249.

98. L. Holper, A. Gross, F. Scholkmann, B.K. Humphreys, M.L. Meier, U. Wolf, M. Wolf, and S. Hotz-Boendermaker, “Physiological effects of mechanical pain stimulation at the lower back measured by functional near-infrared spectroscopy and capnography,” *Journal of Integrative Neuroscience*, vol. 13, pp. 121–142, (2014). doi:10.1142/S0219635214500071.

99. M.C. Hsu, H. Schubiner, M.A. Lumley, J.S. Stracks, D.J. Clauw, and D.A. Williams, “Sustained Pain Reduction Through Affective Self-awareness in Fibromyalgia: A Randomized Controlled Trial,” *Journal of General Internal Medicine*, vol. 25, pp. 1064–1070, (2010). doi:10.1007/s11606-010-1418-6.

100. J.M. Hughes, E.A. Seemann, J.M. George, and K.D. Willis, “The Effects of Pre-treatment Depressive Symptoms on Quality of Life Across Cognitive Behavioral Therapy for Chronic Pain,” *Journal of Clinical Psychology in Medical Settings*, vol. 26, pp. 97–105, (2019). doi:10.1007/s10880-018-9568-5.

101. M. Jiang, R. Mieronkoski, E. Syrjälä, A. Anzanpour, V. Terävä, A.M. Rahmani, S. Salanterä, R. Aantaa, N. Hagelberg, and P. Liljeberg, “Acute pain intensity monitoring with the classification of multiple physiological parameters,” *Journal of Clinical Monitoring and Computing*, vol. 33, pp. 493–507, (2019). doi:10.1007/s10877-018-0174-8.

102. M. Kachele, P. Thiam, M. Amirian, F. Schwenker, and G. Palm, “Methods for Person-Centered Continuous Pain Intensity Assessment From Bio-Physiological Channels,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, pp. 854–864, (2016). doi:10.1109/JSTSP.2016.2535962.

103. S.J. Kamper, A.T. Apeldoorn, A. Chiarotto, R.J.E.M. Smeets, R.W. Ostelo, J. Guzman, and M.W. van Tulder, “Multidisciplinary biopsychosocial rehabilitation for chronic low back pain,” *Cochrane Database of Systematic Reviews*, (2014). doi:10.1002/14651858.CD000963.pub3.

104. K.A. Karjalainen, A. Malmivaara, M.W. van Tulder, R. Roine, M. Jauhiainen, H. Hurri, and B.W. Koes, “Multidisciplinary rehabilitation for fibromyalgia and musculoskeletal pain in working age adults,” *Cochrane Database of Systematic Reviews*, (1999);2010. doi:10.1002/14651858.CD001984.

105. J. Kennel, E. Withers, N. Parsons, and H. Woo, “Racial/Ethnic Disparities in Pain Treatment,” *Medical Care*, vol. 57, pp. 924–929, (2019). doi:10.1097/MLR.0000000000001208.

106. D. Keszthelyi, Q. Aziz, J.K. Ruffle, O. O'Daly, D. Sanders, K. Krause, S.C.R. Williams, and M.A. Howard, "Delineation between different components of chronic pain using dimension reduction - an ASL fMRI study in hand osteoarthritis," *European Journal of Pain*, vol. 22, pp. 1245–1254, (2018). doi:10.1002/ejp.1212.

107. L.J. Kong, R. Lauche, P. Klose, J.H. Bu, X.C. Yang, C.Q. Guo, G. Dobos, and Y.W. Cheng, "Tai Chi for Chronic Pain Conditions: A Systematic Review and Meta-analysis of Randomized Controlled Trials," *Scientific Reports*, vol. 6, 25325, (2016). doi:10.1038/srep25325.

108. M.S. Lee, M.H. Pittler, and E. Ernst, "External Qigong for Pain Conditions: A Systematic Review of Randomized Clinical Trials," *The Journal of Pain*, vol. 8, pp. 827–831, (2007). doi:10.1016/j.jpain.2007.05.016.

109. T. Li, S. Zhang, E. Ikeda, and H. Kobinata, "Functional connectivity modulations during offset analgesia in chronic pain patients: an fMRI study," *Brain Imaging and Behavior*, (2022). doi:10.1007/s11682-022-00652-7.

110. J-A. Lim, S-H. Choi, WJ. Lee, JH. Jang, JY. Moon, YC. Kim, and D-H. Kang, "Cognitive-behavioral therapy for patients with chronic pain," *Medicine*, vol. 97, e10867, (2018). doi:10.1097/MD.0000000000010867.

111. ML. Loggia and V. Napadow, "Multi-parameter autonomic-based pain assessment: More is more?" *Pain*, vol. 153, pp. 1779-1780, (2012). doi:10.1016/j.pain.2012.05.010.

112. D. Lopez-Martinez and R. Picard, "Continuous Pain Intensity Estimation from Autonomic Signals with Recurrent Neural Networks," 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), *IEEE*, pp. 5624-5627, (2018). doi:10.1109/EMBC.2018.8513575.

113. D. Lopez-Martinez and R. Picard, "Multi-task neural networks for personalized pain recognition from physiological signals," 2017 Seventh International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), *IEEE*, pp. 181-184, (2017). doi:10.1109/ACIIW.2017.8272611.

114. A. Louw, K. Zimney, EJ. Puentedura, and I. Diener, "The efficacy of pain neuroscience education on musculoskeletal pain: A systematic review of the literature," *Physiotherapy Theory and Practice*, vol. 32, pp. 332-355, (2016). doi:10.1080/09593985.2016.1194646.

115. MA. Lumley, H. Schubiner, NA. Lockhart, KM. Kidwell, SE. Harte, DJ. Clauw, and DA. Williams, "Emotional awareness and expression therapy, cognitive behavioral therapy, and education for fibromyalgia: a cluster-randomized controlled trial," *Pain*, vol. 158, pp. 2354-2363, (2017). doi:10.1097/j.pain.0000000000001036.

116. M. Lynch, J. Sawynok, C. Hiew, and D. Marcon, "A randomized controlled trial of qigong for fibromyalgia," *Arthritis Research & Therapy*, vol. 14, R178, (2012). doi:10.1186/ar3931.

117. A. Malfliet, J. Kregel, I. Coppieters, R. de Pauw, M. Meeus, N. Roussel, B. Cagnie, L. Danneels, and J. Nijs, "Effect of Pain Neuroscience Education Combined With Cognition-Targeted Motor Control Training on Chronic Spinal Pain," *JAMA Neurology*, vol. 75, p. 808, (2018). doi:10.1001/jamaneurol.2018.0492.

118. D. Mamontov, I. Polonskaia, A. Skorokhod, E. Semenkin, V. Kessler, and F. Schwenker, "Evolutionary Algorithms for the Design of Neural Network Classifiers for the Classification of Pain Intensity," pp. 84-100, (2019). doi:10.1007/978-3-030-20984-1_8.

119. AJ. Miller, "Gender Disparities in Diagnosis and Pain Management," *Temple University*, (2018).

120. MM. Mittinty, S. Vanlint, N. Stocks, MN. Mittinty, and GL. Moseley, "Exploring effect of pain education on chronic pain patients' expectation of recovery and pain intensity," *Scandinavian Journal of Pain*, vol. 18, pp. 211-219, (2018). doi:10.1515/sjpain-2018-0023.

121. M. Monticone, C. Cedraschi, E. Ambrosini, B. Rocca, R. Fiorentini, M. Restelli, S. Gianola, S. Ferrante, G. Zanoli, and L. Moja, "Cognitive-behavioural treatment for subacute and chronic neck pain," *Cochrane Database of Systematic Reviews*, (2016). doi:10.1002/14651858.CD010664.pub2.

122. S. Morley, C. Eccleston, and A. Williams, "Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache," *Pain*, vol. 80, pp. 1-13, (1999). doi:10.1016/S0304-3959(98)00255-3.

123. TM. Palermo, EF. Law, J. Fales, MH. Bromberg, T. Jessen-Fiddick, and G. Tai, "Internet-delivered cognitive-behavioral treatment for adolescents with chronic pain and their parents," *Pain*, vol. 157, pp. 174-185, (2016). doi:10.1097/j.pain.0000000000000348.

124. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, (2011).

125. L. S. Prichep, J. Shah, H. Merkin, and E. M. Hiesiger, "Exploration of the Pathophysiology of Chronic Pain Using Quantitative EEG Source Localization," *Clinical EEG and Neuroscience*, vol. 49, no. 2, pp. 103–113, (2018).

126. M. Rantanen, A. Yli-Hankala, M. van Gils, H. Yppa¨rila¨-Wolters, P. Takala, M. Huiku, M. Kyma¨la¨inen, E. Seitsonen, and I. Korhonen, “Novel multiparameter approach for measurement of nociception at skin incision during general anaesthesia,” *British Journal of Anaesthesia*, vol. 96, no. 4, pp. 367–376, (2006).

127. M. C. Reddan and T. D. Wager, “Brain systems at the intersection of chronic pain and self-regulation,” *Neuroscience Letter*s, vol. 702, pp. 24–33, (2019).

128. S. Rosenthal, "“Watch the Screen”: Biofeedback Can Improve Mindfulness for Chronic Pain,” *Biofeedback*, vol. 46, no. 1, pp. 15–20, (2018).

129. A. N. Santana, I. Cifre, C. N. de Santana, and P. Montoya, “Using Deep Learning and Resting-State fMRI to Classify Chronic Pain Conditions,” *Frontiers in Neuroscience*, vol. 13, (2019).

130. H. S. Seok, B.-M. Choi, G.-J. Noh, and H. Shin, “Postoperative Pain Assessment Model Based on Pulse Contour Characteristics Analysis,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2317–2324, (2019).

131. R. Shala, N. Roussel, G. L. Lorimer Moseley, T. Osinski, and E. J. Puentedura, “Can we just talk our patients out of pain? Should pain neuroscience education be our only tool?” *Journal of Manual & Manipulative Therapy*, vol. 29, no. 1, pp. 1–3, (2021).

132. R. Sielski, W. Rief, and J. A. Glombiewski, “Efficacy of Biofeedback in Chronic back Pain: a Meta-Analysis,” *International Journal of Behavioral Medicine*, vol. 24, no. 1, pp. 25–41, (2017).

133. S. K. Singh, V. Rastogi, and S. K. Singh, “Pain Assessment Using Intelligent Computing Systems,” Proceedings of the National Academy of Sciences, India Section A: Physical Sciences, vol. 86, no. 3, pp. 285–295, (2016).

134. B. T. Susam, M. Akcakaya, H. Nezamfar, D. Diaz, X. Xu, V. R. de Sa, K. D. Craig, J. S. Huang, and M. S. Goodwin, “Automated Pain Assessment using Electrodermal Activity Data and Machine Learning,” 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 372–375, (2018).

135. R. Tanasescu, W. J. Cottam, L. Condon, C. R. Tench, D. P. Auer, “Functional reorganisation in chronic pain and neural correlates of pain sensitisation: A coordinate based meta-analysis of 266 cutaneous pain fMRI studies,” *Neuroscience & Biobehavioral Reviews*, vol. 68, pp. 120–133, (2016).

136. D. Teichmann, J. Klopp, A. Hallmann, K. Schuett, S. Wolfart, and M. Teichmann, “Detection of acute periodontal pain from physiological signals,” *Physiological Measurement*, vol. 39, no. 9, 095007, (2018).

137. A. C. Traeger, H. Lee, M. Hübscher, I. W. Skinner, G. L. Moseley, M. K. Nicholas, N. Henschke, K. M. Refshauge, F. M. Blyth, C. J. Main, J. M. Hush, S. Lo, and J. H. McAuley, "Effect of Intensive Patient Education vs Placebo Patient Education on Outcomes in Patients With Acute Low Back Pain," *JAMA Neurology*, vol. 76, no. 2, p. 161, (2019).

138. R. Treister, M. Kliger, G. Zuckerman, I. G. Aryeh, and E. Eisenberg, "Differentiating between heat pain intensities: The combined effect of multiple autonomic parameters," *Pain*, vol. 153, no. 8, pp. 1807–1814, (2012).

139. . von Trott, A. M. Wiedemann, R. Lüdtke, A. Reißhauer, S. N. Willich, and C. M. Witt, "Qigong and Exercise Therapy for Elderly Patients With Chronic Neck Pain (QIBANE): A Randomized Controlled Study," *The Journal of Pain*, vol. 10, no. 5, pp. 501–508, (2009).

140. J. A. Turner, M. L. Anderson, B. H. Balderson, A. J. Cook, K. J. Sherman, and D. C. Cherkin, "Mindfulness-based stress reduction and cognitive behavioral therapy for chronic low back pain: similar effects on mindfulness, catastrophizing, self-efficacy, and acceptance in a randomized controlled trial," *Pain*, vol. 157, no. 11, pp. 2434–2444, (2016).

141. J. A. Turner, S. Holtzman, and L. Mancl, "Mediators, moderators, and predictors of therapeutic change in cognitive–behavioral therapy for chronic pain," *Pain*, vol. 127, no. 3, pp. 276–286, (2007).

142. W. R, "psych: Procedures for Psychological, Psychometric, and Personality Research.," (2022). [Online]. Available: https://cran.r-project.org/package=psych.

143. S. Walter, S. Gruss, M. Kächele, F. Schwenker, P. Werner, A. Al-Hamadi, A. Andrade, G. Moreira, and H. Traue, "Data fusion for automated pain recognition," Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare, ICST, (2015).

144. C. Wang, C. H. Schmid, P. L. Hibberd, R. Kalish, R. Roubenoff, R. Rones, and T. McAlindon, "Tai Chi is effective in treating knee osteoarthritis: A randomized controlled trial," *Arthritis & Rheumatism*, vol. 61, no. 11, pp. 1545–1553, (2009).

145. C. Wang, C. H. Schmid, R. Rones, R. Kalish, J. Yinh, D. L. Goldenberg, and Y. Lee, "A Randomized Trial of Tai Chi for Fibromyalgia," New England Journal of *Medicine*, vol. 363, no. 8, pp. 743–754, (2010).

146. F. P. C. Waterschoot, P. U. Dijkstra, N. Hollak, H. J. de Vries, J. H. B. Geertzen, and M. F. Reneman, "Dose or content? Effectiveness of pain rehabilitation programs for patients with chronic low back pain: A systematic review," *Pain*, vol. 155, no. 2, pp. 179–189, (2014).

147. J. A. Waxman, R. R. Pillai Riddell, P. Tablon, L. A. Schmidt, and A. Pinhasov, "Development of Cardiovascular Indices of Acute Pain Responding in

Infants: A Systematic Review," *Pain Research and Management*, vol. 2016, pp. 1–15, (2016).

148. P. Werner, D. Lopez-Martinez, S. Walter, A. Al-Hamadi, S. Gruss, and R. Picard, "Automatic Recognition Methods Supporting Pain Assessment: A Survey," IEEE Transactions on Affective Computing, (2019).

149. L. S. Wieland, N. Skoetz, K. Pilkington, R. Vempati, C. R. D'Adamo, and B. M. Berman, "Yoga treatment for chronic non-specific low back pain," *Cochrane Database of Systematic Reviews*, (2017).

150. A. C. de C. Williams, C. Eccleston, and S. Morley, "Psychological therapies for the management of chronic pain (excluding headache) in adults," *Cochrane Database of Systematic Reviews*, (2012).

151. 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. 7-8, pp. 48–59, (2018).

152. L. Yang, S. Wang, X. Jiang, S. Cheng, H.-E. Kim, "PATTERN: Pain Assessment for paTients who can't TEll using Restricted Boltzmann machiNe," BMC Medical Informatics and Decision Making, vol. 16, no. 1, p. 73, (2016).

153. Y. la Yang, H. S. Seok, G.-J. Noh, B.-M. Choi, and H. Shin, "Postoperative Pain Assessment Indices Based on Photoplethysmography Waveform Analysis," *Frontiers in Physiology*, vol. 9, (2018).

154. N. Zaproudina, Z. Ming, and O. O. P. Hänninen, "Plantar Infrared Thermography Measurements and Low Back Pain Intensity," *Journal of Manipulative and Physiological Therapeutics*, vol. 29, no. 3, pp. 219–223, (2006).

155. F. Zeidan and D. R. Vago, "Mindfulness meditation-based pain relief: a mechanistic account," Ann N Y Acad Sci, vol. 1373, pp. 114–127, (2016).

156. Y. Zhao, F. Ly, Q. Hong, Z. Cheng, T. Santander, H. T. Yang, P. K. Hansma, and L. Petzold, "How Much Does It Hurt: A Deep Learning Framework for Chronic Pain Score Assessment," 2020 International Conference on Data Mining Workshops (ICDMW), *IEEE*, pp. 651–660, (2020).

157. P. M. Lehrer, E. G. Vaschillo, and V. Vidali, "Heart rate and breathing are not always in phase during resonance frequency breathing," *Applied Psychophysiology and Biofeedback*, vol. 45, no. 3, pp. 145–152, (2020).

158. L. R. Fisher and P. M. Lehrer, "A method for more accurate determination of resonance frequency of the cardiovascular system, and evaluation of a program to perform it," *Applied Psychophysiology and Biofeedback*, (2021).

159. A. Angelone and N. A. Coulter, “Respiratory sinus arrhythmia: A frequency dependent phenomenon,” *Journal of Applied Physiology*, vol. 19, no. 3, pp. 479–482, (1964).

160. S. Evans, L. Seidman, J. Tsao, Z. Lung, L. Zeltzer, and B. Naliboff, “Heart rate variability as a biomarker for autonomic nervous system response differences between children with chronic pain and Healthy Control Children,” *Journal of Pain Research*, (2013).

161. L. M. Tracy, L. Ioannou, K. S. Baker, S. J. Gibson, N. Georgiou-Karistianis, and M. J. Giummarra, “Meta-analytic evidence for decreased heart rate variability in chronic pain implicating parasympathetic nervous system dysregulation,” *Pain*, vol. 157, no. 1, pp. 7–29, (2016).

162. R. M. Carney, J. A. Blumenthal, P. K. Stein, L. Watkins, D. Catellier, L. F. Berkman, S. M. Czajkowski, C. O’Connor, P. H. Stone, and K. E. Freedland, “Depression, heart rate variability, and acute myocardial infarction,” *Circulation*, vol. 104, no. 17, pp. 2024–2028, (2001).

163. M. Sosnowski, P. W. MacFarlane, Z. Czyż, J. Skrzypek-Wańha, E. Boczkowska-Gaik, and M. Tendera, “Age-adjustment of HRV measures and its prognostic value for risk assessment in patients late after myocardial infarction,” *International Journal of Cardiology*, vol. 86, no. 2-3, pp. 249–258, (2002).

164. S. Hohnloser, T. Klingenheben, M. Zabel, and Y. LI, “Heart rate variability used as an arrhythmia risk stratifier after myocardial infarction,” *Pacing and Clinical Electrophysiology*, vol. 20, no. 10, pp. 2594–2601, (1997).

165. J. A. Chalmers, D. S. Quintana, M. J.-A. Abbott, and A. H. Kemp, “Anxiety disorders are associated with reduced heart rate variability: A meta-analysis,” *Frontiers in Psychiatry*, (2014).

166. V. C. Goessl, J. E. Curtiss, and S. G. Hofmann, “The effect of heart rate variability biofeedback training on stress and anxiety: A meta-analysis,” *Psychological Medicine*, vol. 47, no. 15, pp. 2578–2586, (2017).

167. N. M. De Souza, L. C. Vanderlei, and D. M. Garner, “Risk evaluation of diabetes mellitus by relation of chaotic globals to HRV,” *Complexity*, vol. 20, no. 3, pp. 84–92, (2014).

168. C. M. van Ravenswaaij-Arts, “Heart rate variability,” A*nnals of Internal Medicine*, vol. 118, no. 6, (1993).

169. J. Sztajzel, “Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system,” *Swiss Medicine Weekly*, vol. 134, no. 35-36, pp. 514–522, (2004).

170. D. J. Plews, P. B. Laursen, J. Stanley, A. E. Kilding, and M. Buchheit, “Training adaptation and heart rate variability in elite endurance athletes: Opening the door to effective monitoring,” *Sports Medicine*, vol. 43, no. 9, pp. 773–781, (2013).

171. A. M. Kiviniemi, A. J. Hautala, H. Kinnunen, and M. P. Tulppo, “Endurance training guided individually by daily heart rate variability measurements,” *European Journal of Applied Physiology*, vol. 101, no. 6, pp. 743–751, (2007).

172. P. M. Lehrer and W. E. Slime, “Biofeedback Training to Increase Heart Rate Variability,” in Principles and Practice of Stress Management, Third ed., R. L. Woolfolk, Ed. essay, Guilford Press, pp. 227–248, (2007).

173. P. Lehrer, B. Vaschillo, T. Zucker, J. Graves, M. Katsamanis, M. Aviles, and F. Wamboldt, “Protocol for heart rate variability biofeedback training,” *Biofeedback*, vol. 41, no. 3, pp. 98–109, (2013). https://doi.org/10.5298/1081-5937-41.3.08

174. E. Vaschillo, P. Lehrer, N. Rishe, and M. Konstantinov, *Applied Psychophysiology and Biofeedback*, vol. 27, no. 1, pp. 1–27, (2002). https://doi.org/10.1023/a:1014587304314

175. J. Hayano, F. Yasuma, A. Okada, S. Mukai, and T. Fujinami, “Respiratory sinus arrhythmia,” *Circulation*, vol. 94, no. 4, pp. 842–847, (1996). https://doi.org/10.1161/01.cir.94.4.842

176. N. D. Giardino, R. W. Glenny, S. Borson, and L. Chan, “Respiratory sinus arrhythmia is associated with efficiency of pulmonary gas exchange in healthy humans,” *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 284, no. 5, (2003). https://doi.org/10.1152/ajpheart.00893.2002

177. P. Y. Sin, M. R. Webber, D. C. Galletly, P. N. Ainslie, S. J. Brown, C. K. Willie, A. Sasse, P. D. Larsen, and Y.-C. Tzeng, “Interactions between heart rate variability and pulmonary gas exchange efficiency in humans,” *Experimental Physiology*, vol. 95, no. 7, pp. 788–797, (2010). https://doi.org/10.1113/expphysiol.2010.052910

178. Task Force, “Heart rate variability,” *Circulation*, vol. 93, no. 5, pp. 1043–1065, (1996). https://doi.org/10.1161/01.cir.93.5.1043

179. F. Shaffer and J. P. Ginsberg, “An overview of heart rate variability metrics and norms,” *Frontiers in Public Health*, vol. 5, (2017). https://doi.org/10.3389/fpubh.2017.00258

180. P. M. Lehrer and R. Gevirtz, “Heart rate variability biofeedback: How and why does it work?,” *Frontiers in Psychology*, vol. 5, (2014). https://doi.org/10.3389/fpsyg.2014.00756

181. E. Yuda, M. Shibata, Y. Ogata, N. Ueda, T. Yambe, M. Yoshizawa, and J. Hayano, “Pulse rate variability: a new biomarker, not a surrogate for heart rate variability,” *Journal of physiological anthropology*, vol. 39, (2020). https://doi.org/10.1186/s40101-020-00233-x

182. A. Natarajan, A. Pantelopoulos, H. Emir-Farinas, and P. Natarajan, “Heart rate variability with photoplethysmography in 8 million individuals: a cross-sectional study,” The Lancet Digital Health, vol. 2, no. 12, pp. e650-e657, (2020). https://doi.org/10.1016/S2589-7500(20)30246-6

183. B. Vescio, M. Salsone, A. Gambardella, and A. Quattrone, “Comparison between electrocardiographic and earlobe pulse photoplethysmographic detection for evaluating heart rate variability in healthy subjects in short-and long-term recordings,” *Sensors*, vol. 18, no. 3, 844, (2018). https://doi.org/10.3390/s18030844

184. X. Yu, T. Laurentius, C. Bollheimer, S. Leonhardt, and C.H. Antink, “Noncontact monitoring of heart rate and heart rate variability in geriatric patients using photoplethysmography imaging,” *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1781-1792, (2020). https://doi.org/10.1109/JBHI.2020.3018394

185. M.R. Bozkurt, M.K. Uçar, F. Bozkurt, and C. Bilgin, “In obstructive sleep apnea patients, automatic determination of respiratory arrests by photoplethysmography signal and heart rate variability,” A*ustralasian Physical & Engineering Sciences in Medicine*, vol. 42, no. 4, pp. 959-979, (2019). https://doi.org/10.1007/s13246-019-00796-9

186. D.F. Jhang, Y.S. Chu, J.H. Cai, Y.Y. Tai, and C.C. Chuang, “Pain Monitoring Using Heart Rate Variability and Photoplethysmograph-Derived Parameters by Binary Logistic Regression,” *Journal of Medical and Biological Engineering*, vol. 41, no. 5, pp. 669-677, (2021). https://doi.org/10.1007/s40846-021-00651-x

187. E. Lam, S. Aratia, J. Wang, and J. Tung, "Measuring heart rate variability in free-living conditions using consumer-grade photoplethysmography: Validation study," *JMIR Biomedical Engineering*, vol. 5, no. 1, e17355, (2020). https://doi.org/10.2196/17355

188. G.P. Chrousos, N. Papadopoulou-Marketou, F. Bacopoulou, M. Lucafò, A. Gallotta, and D. Boschiero, "Photoplethysmography (PPG)-determined heart rate variability (HRV) and extracellular water (ECW) in the evaluation of chronic stress and inflammation," *Hormones*, pp. 1-8, (2022). https://doi.org/10.1007/s42000-021-00341-y

189. U. Rubins, Z. Marcinkevics, I. Logina, A. Grabovskis, and E. Kviesis-Kipge, "Imaging photoplethysmography for assessment of chronic pain patients," Optical Diagnostics and Sensing XIX: Toward Point-of-Care Diagnostics, vol. 10885, pp. 16-23, (2019). https://doi.org/10.1117/12.2508393

190. L.H.R. Batista, W.J.R. Domingues, A.D.A.C. e Silva, K.A.T. Lopes, M.L. de Castro Amorim, and M. Rossato, "Heart rate variability responses determined by photoplethysmography in people with spinal cord injury," *Biomedical Signal Processing and Control*, vol. 69, 102845, (2021). https://doi.org/10.1016/j.bspc.2021.102845

191. I. Constant, D. Laude, I. Murat, and J.L. Elghozi, "Pulse rate variability is not a surrogate for heart rate variability," *Clinical Science*, vol. 97, no. 4, pp. 391-397, (1999). https://doi.org/10.1042/cs0970391

192. K. Charlot, J. Cornolo, J.V. Brugniaux, J.P. Richalet, and A. Pichon, "Interchangeability between heart rate and photoplethysmography variabilities during sympathetic stimulations," *Physiological Measurement*, vol. 30, 1357, (2009). http://doi.org/10.1088/0967-3334/30/12/005

193. W.H. Lin, D. Wu, C. Li, H. Zhang, and Y.T. Zhang, "Comparison of Heart Rate Variability from PPG with That from ECG," *The International Conference on Health Informatics*, vol. 42, pp. 213-215, (2014). https://doi.org/10.1007/978-3-319-03005-0_54

194. C.E. Dunn, D.C. Monroe, C. Crouzet, J.W. Hicks, and B. Choi, "Speckleplethysmographic (SPG) Estimation of Heart Rate Variability During an Orthostatic Challenge," *Scientific Reports*, vol. 9, 14079, (2019). https://doi.org/10.1038/s41598-019-50526-0

195. K.N. Sá, L. Moreira, A.F. Baptista, et al., "Prevalence of chronic pain in developing countries: systematic review and meta-analysis," *Pain Rep*., vol. 4, no. 6, e779, (2019).

196. J. Dahlhamer, J. Lucas, C. Zelaya, et al., "Prevalence of chronic pain and high-impact chronic pain among adults—United States, 2016," *MMWR Morb Mortal Wkly Rep*., vol. 67, no. 36, p. 1001, (2018).

197. R.J. Gatchel, D.D. McGeary, C.A. McGeary, and B. Lippe, "Interdisciplinary chronic pain management: past, present, and future," *Am Psychol*., vol. 69, no. 2, p. 119, (2014).

198. Z. Cheng, F. Ly, T. Santander, et al., "Preliminary study: quantification of chronic pain from physiological data," *Pain Rep*., vol. 7, no. 6, e1039, (2022).

199. Y. Zhao, F. Ly, Q. Hong, et al., "How much does it hurt: A deep learning framework for chronic pain score assessment," in 2020 International Conference on Data Mining Workshops (ICDMW), *IEEE*, pp. 651-660, (2020).

200. R. Sielski, W. Rief, and J.A. Glombiewski, "Efficacy of biofeedback in chronic back pain: a meta-analysis," *Int J Behav Med*., vol. 24, no. 1, pp. 25-41, (2017). DOI: 10.1007/s12529-016-9572-9. PMID: 27307013.

201. V.C. Goessl, J.E. Curtiss, and S.G. Hofmann, "The effect of heart rate variability biofeedback training on stress and anxiety: a meta-analysis," *Psychol Med*., vol. 47, no. 15, pp. 2578-2586, (2017).

202. M.A. Day, L.C. Ward, D.M. Ehde, et al., "A pilot randomized controlled trial comparing mindfulness meditation, cognitive therapy, and mindfulness-based cognitive therapy for chronic low back pain," *Pain Med*., vol. 20, no. 11, pp. 2134-2148, (2019).

203. W.W. Deardorff, H.S. Rubin, and D.W. Scott, "Comprehensive multidisciplinary treatment of chronic pain: a follow-up study of treated and non-treated groups," *Pain*, vol. 45, no. 1, pp. 35-43, (1991).

204. M.W. Donnino, G.S. Thompson, S. Mehta, et al., "Psychophysiologic symptom relief therapy for chronic back pain: a pilot randomized controlled trial," *Pain Re*p., vol. 6, no. 5, e959, (2021).

205. H. Flor, T. Fydrich, and D.C. Turk, "Efficacy of multidisciplinary pain treatment centers: a meta-analytic review," *Pain*, vol. 49, no. 2, pp. 221-230, (1992).

206. R.J. Gatchel and A. Okifuji, "Evidence-based scientific data documenting the treatment and cost-effectiveness of comprehensive pain programs for chronic nonmalignant pain," *J Pain*, vol. 7, no. 11, pp. 779-793, (2006).

207. S. Rosenthal, ""Watch the screen": biofeedback can improve mindfulness for chronic pain," *Biofeedback*, vol. 46, no. 1, pp. 15-20, (2018).

208. A. Fahrenkamp, L. Sim, L. Roers, et al., "An innovative and accessible biofeedback intervention for improving self-regulatory skills in pediatric chronic pain: a pilot study," *J Altern Complement Med*., vol. 26, no. 3, pp. 212-218, (2020).

209. E. Fisher, E. Law, J. Dudeney, C. Eccleston, and T.M. Palermo, "Psychological therapies (remotely delivered) for the management of chronic and recurrent pain in children and adolescents," *Cochrane Database Syst Rev*., vol. 2019, no. 4, CD011118, (2019).

210. J. Guzman, "Multidisciplinary rehabilitation for chronic low back pain: systematic review," *BMJ*, vol. 322, no. 7301, pp. 1511-1516, (2001).

211. J.A. Turner, M.L. Anderson, B.H. Balderson, et al., "Mindfulness-based stress reduction and cognitive behavioral therapy for chronic low back pain: similar effects on mindfulness, catastrophizing, self-efficacy, and acceptance in a randomized controlled trial," *Pain*, vol. 157, no. 11, pp. 2434-2444, (2016).

212. K.P. Kapitza, T. Passie, M. Bernateck, et al., "First non-contingent respiratory biofeedback placebo versus contingent biofeedback in patients with chronic low back pain: a randomized, controlled, double-blind trial," *Appl Psychophysiol Biofeedback*, vol. 35, no. 3, pp. 207-217, (2010).

213. M.P. Jensen, K.J. Gertz, A.E. Kupper, et al., "Steps toward developing an EEG biofeedback treatment for chronic pain," *Appl Psychophysiol Biofeedback*, vol. 38, no. 2, pp. 101-108, (2013).

214. H. Flor and N. Birbaumer, "Psychophysiological methods in the assessment and treatment of chronic musculoskeletal pain," *Clin Appl Psychophysiol*, pp. 171-184, (1994).

215. Y.K. Ashar, A. Gordon, H. Schubiner, et al., "Effect of pain reprocessing therapy vs placebo and usual care for patients with chronic back pain: a randomized clinical trial," *JAMA Psychiatry*, vol. 79, no. 1, pp. 13-23, (2022).

216. A. Gordon, "The Way Out: A Revolutionary, Scientifically Proven Approach to Healing Chronic Pain," *Avery*, (2021).

217. J.M. Rippe, ed., "Lifestyle Medicine," *CRC Press*, (2019).

218. R.F. Kushner, K.W. Sorensen, "Lifestyle medicine: the future of chronic disease management," *Curr Opin Endocrinol Diabetes Obes*, vol. 20, no. 5, pp. 389-395, (2013).

219. C. Hayes, R. Naylor, G. Egger, "Understanding chronic pain in a lifestyle context: the emergence of a whole-person approach," *Am J Lifestyle Med*, vol. 6, no. 5, pp. 421-428, (2012).

220. Z. Altug, "Lifestyle medicine for chronic lower back pain: an evidence-based approach," *Am J Lifestyle Med*, vol. 15, no. 4, pp. 425-433, (2021).

221. J. Znidarsic, DO, et al., "“Living Well with Chronic Pain”: Integrative Pain Management via Shared Medical Appointments," *Pain Med*, vol. 22, no. 1, pp. 181-190, (2021).

222. A. Lazaridou, M. Paschali, E.S. Vilsmark, et al., "Biofeedback EMG alternative therapy for chronic low back pain (the BEAT-pain study)," *Dig Health*, vol. 9, (2023).

223. R. Naylor, C. Hayes, G. Egger, "The relationship between lifestyle, metaflammation, and chronic pain: a systematic review," *Am J Lifestyle Med*, vol. 7, no. 2, pp. 130-137, (2013).

224. P. Rajguru, M.J. Kolber, A.N. Garcia, et al., "Use of mindfulness meditation in the management of chronic pain: a systematic review of randomized controlled trials," *Am J Lifestyle Med*, vol. 9, no. 3, pp. 176-184, (2015).

225. C. Geyer, "The complex interplay between weight, chronic pain, and mood: how team-based care and personalized approaches can improve function and quality of life," *Am J Lifestyle Med*, vol. 13, no. 4, pp. 362-366, (2019).

226. A.M. Lynch, C.I. Jarvis, R.J. DeBellis, A.K. Morin, "State of the art reviews: nonpharmacologic approaches for the treatment of insomnia," *Am J Lifestyle Med*, vol. 1, no. 4, pp. 274-282, (2007).

227. T. Hubkova, "No more pain in the gut: lifestyle medicine approach to irritable bowel syndrome," *Am J Lifestyle Med*, vol. 11, no. 3, pp. 223-226, (2017).

228. C.H. Li, F.S. Ly, K. Woodhouse, et al., "Dynamic Phase Extraction: Applications in Pulse Rate Variability," *Appl Psychophysiol Biofeedback*, vol. 47, no. 2, pp. 213-222, (2022).

229. T.M. Marteau, H. Bekker, "The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI)," B*r J Clin Psychol*, vol. 31, no. 3, pp. 301-306, (1992).

230. C.D. Spielberger, F. Gonzalez-Reigosa, A. Martinez-Urrutia, L.F. Natalicio, D.S. Natalicio, "The state-trait anxiety inventory," *Rev Interam Psicol*, vol. 5, no. 3-4, (1971).

231. D.M. Walton, R.J. Werstine, J.C. MacDermid, W. Nielson, R.W. Teasell, J. Pretty, P.W. Stratford, "Development and initial validation of the Satisfaction and Recovery Index (SRI) for measurement of recovery from musculoskeletal trauma," *Open Orthop J*, vol. 8, pp. 316, (2014).

232. R. Melzack, "The McGill Pain Questionnaire: major properties and scoring methods," *Pain*, vol. 1, no. 3, pp. 277-299, (1975).

233. M.E. Wewers, N.K. Lowe, "A critical review of visual analogue scales in the measurement of clinical phenomena," *Res Nurs Health*, vol. 13, no. 4, pp. 227-236, (1990).

234. I. Van Diest, K. Verstappen, A.E. Aubert, et al., "Inhalation/Exhalation Ratio Modulates the Effect of Slow Breathing on Heart Rate Variability and Relaxation," *Appl Psychophysiol Biofeedback*, vol. 39, no. 3-4, pp. 171-180, (2014).

235. Stan Development Team, "Stan Modeling Language Users Guide and Reference Manual, v2.32," (2023). Available at: https://mc-stan.org.

236. P.C. Bürkner, "brms: An R Package for Bayesian Multilevel Models Using Stan," *J Stat Softw*, vol. 80, no. 1, pp. 1-28, (2017).

237. D. Makowski, M.S. Ben-Shachar, S.H.A. Chen, D. Lüdecke, "Indices of effect existence and significance in the Bayesian framework," *Front Psychol*, vol. 10, 2767, (2019).

238. J. Westfall, D.A. Kenny, C.M. Judd, "Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli," J *Exp Psychol Gen*, vol. 143, no. 5, pp. 2020-2045, (2014).

239. A. Gelman, B. Goodrich, J. Gabry, A. Vehtari, "R-squared for Bayesian regression models," *Am Stat*, vol. 73, no. 3, pp. 307-309, (2019).

240. S. Bhaskar, D. Hemavathy, S. Prasad, "Prevalence of chronic insomnia in adult patients and its correlation with medical comorbidities," J Family Med Prim Care, vol. 5, no. 4, pp. 780-784, (2016). DOI: 10.4103/2249-4863.201153. PMID: 28348990; PMCID: PMC5353813.

241. M. R. Irwin, R. Olmstead, J. E. Carroll, "Sleep disturbance, sleep duration, and inflammation: a systematic review and meta-analysis of cohort studies and experimental sleep deprivation," Biological Psychiatry, vol. 80, no. 1, pp. 40-52, (2016).

242. U.S. Department of Health and Human Services, "Any anxiety disorder," National Institute of Mental Health. [Online]. Available: https://www.nimh.nih.gov/health/statistics/any-anxiety-disorder. [Accessed (n.d.)].

243. G. Forest et al., "Temperature biofeedback and sleep: limited findings and methodological challenges," ChronoPhysiology and Therapy, (2012): pp. 59-66.

244. S. Ancoli-Israel, A. R. Seifert, M. Lemon, "Thermal biofeedback and periodic movements in sleep: patients' subjective reports and a case study," Biofeedback and Self-Regulation, vol. 11, (1986): pp. 177-188.

245. K. Kräuchi, "The thermophysiological cascade leading to sleep initiation in relation to phase of entrainment," Sleep Medicine Reviews, vol. 11, no. 6, pp. 439-451, (2007).

246. E. JW VanSomeren, "More than a marker: interaction between the circadian regulation of temperature and sleep, age-related changes, and treatment possibilities," Chronobiology International, vol. 17, no. 3, pp. 313-354, (2000).

247. B. A. Thyer et al., "Autonomic correlates of the subjective anxiety scale," Journal of Behavior Therapy and Experimental Psychiatry, vol. 15, no. 1, pp. 3-7, (1984).

248. D. J. Buysse, C. F. Reynolds III, T. H. Monk, S. R. Berman, D. J. Kupfer, "The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research," © University of Pittsburgh 1989 and 2010. All Rights Reserved.

249. C. D. Spielberger, "State-trait anxiety inventory for adults," (1983).

250. C. D. Spielberger, R. L. Gorsuch, R. Lushene, P. R. Vagg, G. A. Jacobs, "Manual for the State-Trait Anxiety Inventory," Consulting Psychologists Press, (1983).

251. T. M. Marteau, H. Bekker, "The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI)," British Journal of Clinical Psychology, vol. 31, no. 3, pp. 301-306, (1992).