Multimodal Pain Intensity Assessment based on Physiological Biosignals

by

PAVLIDOU ELISAVET

B.Sc. Technological Educational Institute of Crete, 2020

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

SCHOOL OF ENGINEERING

HELLENIC MEDITERRANEAN UNIVERSITY

2024

Approved by:

Major Professor

Professor Manolis Tsiknakis

Abstract

Pain is a multifaceted and subjective experience that can affect individuals both physically and emotionally. According to the most widely accepted definition, it is characterized as the "unpleasant sensory and emotional experience associated with, or resembling that associated with actual or potential tissue damage". Visual analog scales, numeric rating scales, and various questionnaires, all relying on patient-reported outcome measurements, are prevalent methods within healthcare and research domains for assessing the presence, incidence, and severity of pain. Nevertheless, self-report evaluation tools require cognitive, linguistic, and social abilities, which may manifest variations in certain populations such as neonates, individuals with intellectual disabilities, and those affected by dementia.

Pain can be classified as either acute or chronic. Acute pain is sudden, sharp, and usually, temporary. Conversely, persistent pain can result in prolonged suffering. Chronic pain is defined as enduring discomfort lasting beyond six months, often representing a disease state. Diagnosing and managing chronic pain can be challenging due to its sustained nature. In the case of individuals hospitalized with chronic conditions, precise pain assessment is imperative for the effective application of pain management strategies.

Currently, pain detection and accurate classification continue to pose a significant challenge in both the medical field and research endeavors. The objective of this master thesis is to automate pain recognition and assessment through physiological-data-driven predictive models. This study emphasizes the substantial importance of integrating multimodal physiological signals and exploring their most efficient combination. Furthermore, we explored the influence of using multiple types of signal (multimodal approach) on any improvements in performance compared to using just one type of signal (unimodal approach). Additionally, we aimed to scrutinize the influence of demographic factors such as gender and age on pain perception and how they may affect the outcomes of our experiments. We used the BioVid Part A dataset as input in our pipeline, incorporating the demographic data with age groups divided into: 20-35 years, 36-50 years, and 51-65 years. The Pan & Tompkins algorithm was applied for ECG feature extraction, while statistical analysis was employed for both GSR and EMG signals. Following data pre-processing, an early fusion approach was implemented in all unimodal and multimodal experiments encompassing binary (no pain vs.

pain) and multiclass tasks (all pain levels, no pain vs. pain levels).

Notably, all models exhibited better performance between baseline versus the highest pain intensity classification, whether applied in unimodal or multimodal tasks. Galvanic skin response emerged as the most effective standalone modality, while the combination of ECG, EMG, and GSR proved to be the most effective pair of fused modalities. Including all fused signals resulted in additional improvements across almost all subjects, especially in the binary classification task of no pain (BL) versus the highest level of pain (PA4). It is noteworthy that in machine learning experiments, the polynomial kernel consistently yielded higher accuracy values in both binary and multiclass classification tasks. To the best of our knowledge, our study reported the highest accuracy in gender-age-based experiments (0.8283) using all physiological data from the BioVid dataset. This performance was observed in a binary task distinguishing between baseline and the highest intensity of pain within the female population aged 20-35.

Keywords: automated pain assessment, pain classification, physiological signals, BioVid, machine learning, LSTM, ECG, GSR, EMG, age, gender

Table of Contents

Abstract **1**

Table of Contents **3**

List of Figures **6**

List of Tables **7**

Acknowledgment **9**

Dedication **10**

1 Introduction **11**

1.1 Autonomic Nervous System 13

1.2 Pain and Emotional Arousal 14

1.3 Automated Pain Recognition 16

1.4 Physiological Signals 17

1.4.1 Electrocardiogram 18

1.4.1.1 Heart Rate 20

1.4.1.2 Heart Rate Variability 20

1.4.2 Galvanic Skin Response 22

1.4.3 Electromyogram 23

1.5 Objectives and Research Questions 25

1.6 Thesis Outline 25

2 Systematic Literature Review **26**

2.1 Search Strategy 27

2.2 Selection Criteria 28

2.3 Related Work 28

2.4 Neurological Signals as Input in Pain Assessment 35

2.5 Physiological-based Pain Datasets 36

2.6 Limitations and Open Issues . . . 37
2.7 Ethical Concerns about Automated Pain Assessment . . . 39

3 Predictive Models for Automated Pain Assessment 40
3.1 Artificial Neural Networks . . . 40
3.2 Recurrent Neural Networks . . . 41
3.3 Long-Short Term Memory Network . . . 42
3.4 Support Vector Machine . . . 44
3.5 Desicion Trees . . . 45
3.6 K-Nearest Neighbor Algorithm . . . 46

4 Implementation 47
4.1 BioVid Dataset Description . . . 49
4.2 Extracted Features . . . 50
4.3 Pan & Tompkins Algorithm . . . 52
4.3.1 Noise Cancelation . . . 53
4.3.2 Derivative Filter . . . 53
4.3.3 Squaring and Integration . . . 54
4.3.4 Fiducial Mark . . . 54
4.3.5 Thresholds . . . 54
4.3.6 Search Back for Missing Peaks . . . 54
4.4 Dataset Separation and Preparation . . . 55
4.5 Electrocardiogram Features . . . 57
4.6 Galvanic Skin Response Features . . . 58
4.7 Electromyogram Features . . . 59
4.8 Machine Learning Pipeline . . . 60
4.9 LSTM Pipeline . . . 61
4.10 LSTM Model Data Preprocessing . . . 63
4.11 Evaluation Methods . . . 63

5 Results 64

5.0.1 Unimodal Experiments 64
5.0.2 Multimodal Experiments 70
5.0.3 Gender Experiments 72
5.0.4 Age Group Experiments 76
5.1 Discussion 78

6 Conclusions **81**
6.1 Study Limitations 82
6.2 Further Research 83

References **85**

Appendices **95**

List of Figures

1 ECG in the baseline and the highest level of pain in Matlab environment 19
2 GSR in the baseline and in the highest level of pain in Matlab environment 22
3 EMG in the baseline and the highest level of pain in Matlab environment 24
4 PRISMA flowchart illustrates the searching and screening process 27
5 Illustration of brain neuron and ANN architecture 41
6 Recurrent Neural Network Architecture 42
7 LSTM Network Architecture 43
8 SVM Classification 44
9 Decision Tree Architecture 45
10 Example of K-NN Algorithm 46
11 Description of the computational pipeline employed, including signal preprocessing, feature extraction, methods & experiments. 48
12 Example of pain stimulus 49
13 Steps of Pan & Tompkins algorithm 52
14 Pan-Tompkins preprocessing 55
15 Normalized data in ECG signal 56
16 Machine Learning Process 60
17 LSTM Model Architecture 62

List of Tables

1 Summary of the reviewed papers 29
2 Pain datasets encompassing physiological signals 37
3 List of the extracted features 51
4 Binary ECG Experiments Between BL vs Pain Levels 65
5 Binary ECG experiments Between Pain Levels 65
6 Accuracy of Multiclass ECG Experiments 65
7 Binary GSR Experiments Between BL vs. Pain Levels 66
8 Binary GSR Experiments Between Pain Levels 66
9 Accuracy of Multiclass GSR Experiments 66
10 Binary EMG Experiments Between BL vs Pain Levels 67
11 Binary EMG Experiments Between Pain Levels 67
12 Accuracy of Multiclass EMG Experiments 67
13 LSTM Binary ECG Experiments 68
14 Multiclass ECG Experiments 68
15 LSTM Multiclass GSR Experiments 68
16 LSTM Binary GSR Experiments 69
17 Multiclass EMG Experiments 69
18 LSTM Binary EMG Experiments 70
19 Binary All Subjects Experiments 70
20 Binary All Subjects Experiments - Hold Out 71
21 Multiclass All Subjects Experiments 71
22 LSTM Binary Experiments 72
23 LSTM Multiclass Experiments 72
24 Binary Males Experiments 73
25 Accuracy of Multiclass Gender Experiments 73
26 Binary Females Experiments 74
27 LSTM - Mean Accuracy of Multiclass Gender Experiments 74
28 LSTM Binary Female Experiments 75

29 LSTM Binary Males Experiments 75
30 Accuracy of Binary Females Experiments 76
31 Multiclass Experiments in Gender-Age Groups 77
32 Accuracy of Binary Males Experiments 77
33 Comparison of multimodal studies used BioVid in accuracy 78
34 Comparison of unimodal studies used BioVid in accuracy 79
35 Multiclass Multimodal Hold-Out Experiments 95
36 Multiclass Unimodal ECG Hold-Out Experiments 95
37 Multiclass Unimodal GSR Hold-Out Experiments 95
38 Multimodal Binary Hold-Out Experiments 96
39 Multiclass Unimodal EMG Hold-Out Experiments 97

Acknowledgment

I would like to express my profound appreciation to my supervisor, Professor Manolis Tsiknakis, who inspired me to explore the biomedical field during my master's studies, and I had the privilege of being a part of the Biomedical Informatics & eHealth Laboratory.

Also, I am grateful to my co-supervisor, Dr. Lefteris Koumakis, for his guidance, and sharing knowledge, throughout the whole time of this thesis and the chance I had working with him in the Computational BioMedicine Laboratory at the Foundation for Research and Technology Hellas (FORTH).

Finally, from the depth of my heart, I want to thank my best friend Stavros Vouros for his unwavering support and companionship all these years, and Eva Petavraki for her precious advice and assistance through the journey of my master's studies.

Dedication

Dedicated to the ones who supported me in my pursuit of education,

my mother & aunt,

~ Sofia & Elli ~

1. Introduction

Pain is a multidimensional and subjective sensation that has an impact on individuals' physical and mental health [1]. It is characterized as "the unpleasant sensory and emotional experience associated with, or resembling that associated with actual or potential tissue damage" [2]. This definition is broadly accepted and has been embraced by healthcare providers, researchers in the field of pain, governmental and non-profit organizations, including the World Health Organization. Pain consists of two aspects: a sensory (physical) and an affective (emotional/psychological) aspect [1]. The first one alerts the body to the source of the pain, whereas the second one leads to psychological distress as a result of a painful engagement [3].

Pain might endure for minutes, months, or years, and its intensity may vary. It can be extremely minor, barely perceptible, or unbearable. Based on its duration it can be defined as acute or chronic [4][5]. The first one is sudden, sharp, and usually, temporary [5]. A specific event, such as wounds, injuries, or a broken bone triggers this kind of pain. Acute pain can be divided into two phases. The initial phase involves the brain alerting the body to potentially dangerous stimuli and lasts for a few seconds. In the second phase, the body employs defensive mechanisms to repair the injured tissue. This phase, also known as the sub-chronic phase, may persist for a few hours or extend over several days [5][6]. On the other hand, when pain persists, it can lead to suffering. Chronic pain is an ongoing situation that lasts longer than six months [6]. This type of pain is considered as a disease state and sometimes can be difficult to be diagnosed and managed appropriately. Chronic pain affects an extensive percentage of the population, with worldwide incidence estimates ranging from 18% to 43% [6]. Additionally, it may have an impact on physical impairment, emotional status, and a decreased quality of life [5][6].

Pain is characterized by three primary aspects: the level of intensity, the duration, and its distribution [6]. The primary factor is the degree of pain, varying from mild to moderate or intense. The second factor concerns the length of time the pain persists, which is categorized as either acute or chronic, as previously explained. The distribution of pain stands as a crucial element in the clinical evaluation of individuals experiencing long-lasting pain. To adequately address an individual's pain situation, it is crucial to identify both the presence of pain and the level of its intensity.

The current benchmark approaches in the healthcare and research domain, for determining the presence, incidence, and severity of pain, are the Visual Analog Scales (VAS), the Verbal Rating Scales (VRS), the Numeric Rating Scales (NRS), or various questionnaires, which rely on patient-reported outcome measurements [7]. The most typical measurements include scales for patients to assess their level of pain on a range of 0–10 or 0–100. As a result, these processes are time-consuming. Self-report, however, demands cognitive, linguistic, and social abilities that may vary in children, and other populations such as newborns, or individuals with dementia [7]. Furthermore, self-reports are an imprecise assessment tool of a patient's discomfort because pain relies also on behavioral observations. Even if the qualified medical staff were able to observe the intensity of pain several times per day, this frequency of assessment is likely to decrease under economic pressure. Therefore, a human observation may miss pertinent pain events or discover changes too late. Consequently, healthcare providers progressively advocate the utilization of reliable pain assessment tools to guarantee the efficacy of pain control and therapy.

Often, unpleasant stimuli like an injury or a wound cause a sense of pain [9][10]. Lesions in the peripheral or central nervous system can also cause pain, as seen in patients with diabetes, or stroke [10]. Specific autonomic and somatic reflexes are originally linked to acute pain when it happens. For example, facial expressions, body motions, and speech that are intimately associated with psychological suffering are examples of nonverbal pain reactions [10].

In recent years, accurate pain detection has remained one of the strongest challenges in the medical field and research studies [7][8]. Today, pain is one of the most significant clinical signs used to assess the severity of an injury in a patient. Pain is regarded as a warning sign of danger and frequently identifies the location of a lesion; the intensity of it also reflects the health status of the patient [7][9]. The lack of adequate pain alleviation not only produces emotional distress and causes changes in the mental health of patients, such as anxiety disorder and depression, but also leads to several pathophysiological consequences, work productivity problems, and has an impact on patients' quality of lives [11]. These consequences impose a significant socioeconomic burden on patients, families, and society.

1.1 Autonomic Nervous System

The autonomic nervous system (ANS) is the branch of the peripheral nervous system that controls the operations of internal organs including the heart and stomach [12]. The main purpose of the ANS is to control actions, regulate free will motion and coordination, transfer data from the brain to the other organs, and activate the "fight-or-flight" response [15].

The ANS can be separated into two subsystems: the Parasympathetic Nervous System (PNS) and the Sympathetic Nervous System (SNS) [13]. The PNS is typically engaged during periods of rest and sleep, whereas the SNS is triggered in reaction to stress or discomfort [14][13]. The SNS controls the activity of involuntary movements when necessary [13], often reinforcing our "fight-or-flight" responses. In contrast, the PNS is engaged to decrease the activity of involuntary movements and supports our rest-and-digest conditions [15][16].

There is growing evidence that the autonomic nervous system and pain perception are related [12][13]. Consequently, several studies have examined and documented changes in the ANS that occur when subjects are exposed to painful stimuli. The main role of the ANS in pain reaction is to highlight physiological markers for pain investigation [12]. The pain process is often initiated by unpleasant mechanical, heat, or cold stimuli applied to the body from an endogenous or exogenous source, activating the sensory neural pathway [**?**]. Researchers have indicated a probable connection between autonomic dysfunction and pain, which is supported by a significant correlation between the brain regions responsible for pain processing and control, and the central autonomic network [11] [14][13].

The central and autonomic nervous systems receive data through neural pathways from the peripheral nervous system [16]. The signal is then transmitted from the spinal cord to the brain's prosencephalon, mesencephalon, and cortex centers at higher levels, where it is processed to produce sympathetic activity [16][17]. As a result, physiological responses typically stem from this process. Thus, skin temperature, heart rate, and muscle activity are some of the physiological signals that are altered when the autonomic nervous system is activated. Due to this, they offer excellent choices for automated pain assessment approaches [16].

It has been established that certain brain areas, such as the forebrain structures, exhibit gender-based changes in activity in response to noxious cutaneous heat stimulation [10]. More specifically, the female population reported higher rates than the male population [10]. According to brain studies, men and women process emotions differently, which may explain why there are gender variations in the actual (or subjectively experienced) emotional processes [11]. Given the correlation between brain structures and autonomic control in pain experience, physiological responses of the human body have been proven to deliver significant information [10][11] [12].

1.2 Pain and Emotional Arousal

Emotions fulfill a vital role in daily life since they affect a person's emotional status and mental health [14]. Positive emotions enhance individuals' effectiveness, while negative emotions can have adverse effects on their well-being. Risk factors for anxiety or depression include the accumulation of negative feelings over time. Emotions typically indicate a mental state that arises spontaneously, without conscious intention, as opposed to mood, which represents a conscious and prevailing emotional state at a particular time [14][15]. They are often followed by biological and behavioral alterations that affect the human organs, including the brain, heart, and stomach [14]. Emotional distress can also exacerbate physical pain in different body parts. Typical discomforts that might be linked to mental distress involve trouble sleeping, headaches, muscle pain especially in the neck, etc.

Sentiments are highly interrelated and tightly linked with pain in the networks of the human brain [15]. Sadness and depression are both types of unpleasant emotional experiences that fall under the umbrella of pain [19][20]. According to neuropsychology, the same brain areas are responsible for both emotion and pain [17][19]. For instance, the hippocampus, and medial frontal cortex are implicated in both pain and unpleasant feelings, indicating that the processing of negative emotions is influenced by the perception of pain [20][21]. Defensive reactions are controlled by a subcortical circuit that includes the unconscious processing of cues that explain the emotions connected to chronic pain [21]. According to Johnson et al.[22] when this subcortical defensive circuit is out of balance, it interacts with the cerebral cortex and results in the conscious experience of dread and anxiety. The dysregulated subcortical defense circuit may therefore contribute to the persistent activation of these cortical regions by causing secondary emotional reactions to pain,

which can lead to more suffering and impairment.

According to research findings, people who experience pain typically experience mental health issues, notably melancholy. In a study where individuals had to rate emotional scenes while receiving painful or non-painful electrical shocks, it was discovered that unpleasant stimuli drastically diminished the emotional ratings of pleasant images and reduced the visual elicitation of emotional reactions in response to pleasant stimuli [23]. Their dysfunctional interpretation of emotional cues implies that pain might influence people's emotional experiences [19][24]. In a different study conducted by Gerdes et al. [25], pressure pain was deliberately induced in participants who were simultaneously exposed to emotional facial expressions. Researchers found that while people's reactions to joyful expressions were minimized by painful stimuli, their emotional judgments of the images did not change.

Emotions and pain are the foundation of subjective experiences that govern a variety of physiological states and influence human behavior. Almost all life forms are motivated by instinctive feelings like fear and desire, which help them successfully adapt to their environment. The human perception of pain is commonly characterized by three dimensions [26]. The position, timing, and physical attributes (such as mechanical, chemical, and thermal) of the unpleasant stimuli are all identified by the sensory-discriminative dimension [27]. The discomfort associated with exposure to noxious stimuli is underpinned by the affective dimension, which is the one most directly connected to emotion. This dimension also triggers defensive responses, which help the person deal with the noxious stimulus. The cognitive-evaluative component also affects how we interpret the significance and outcomes of pain or injury [27].

Pain serves as a defensive mechanism for living organisms, helping them detect and avoid physical threats. Activation of the pain system leads to negative emotions, promoting avoidance behavior and learning to prevent present and future danger. Consequently, there is a natural connection between pain and emotions. Without the painful feeling brought on by nociceptive input, there would be no incentive to avoid dangerous stimuli [27]. The dynamic relationship between pain and emotional arousal creates a feedback loop, where each influences and magnifies the other in a situation that unfolds within the realms of both physiology and psychology.

1.3 Automated Pain Recognition

The Numeric Pain Rating Scale[1] and self-reported questionnaires on physical function, such as Pain Assessment in Advanced Dementia Scale (PAINAD)[2], the Adult Non-Verbal Pain Scale (NVPS)[3], the Defense and Veterans Pain Rating Scale (DVPRS)[4], among others, are instruments that provide a subjective assessment of pain and currently the only methods used as the gold standard for clinical evaluation of pain [28]. The intricate and complex nature of pain poses significant challenges for both patients and healthcare professionals. As a result, accurately determining pain can be a challenging undertaking. Currently, there are no established methods for objective pain assessment in clinical settings. Specifically, there is a lack of reliable tools to aid clinicians in determining if patients are experiencing pain, evaluating its severity, understanding its impact on daily life, and assessing the effectiveness of interventions such as procedures or medications in significantly reducing pain levels.

To date, the detection of pain in individuals has been examined only by observation. Continuous recording is not conceivable, even if skilled healthcare providers use observation tools to measure the level of pain multiple times during the day or night. To create a trustworthy automated pain detection model, it is important to use reliable pain patterns that can be captured in a variety of methods.

The precise pain prevalence and levels of pain for patients remain under the shadow. With the desire to assist and promote better pain management, the research field was focused on using predictive models that have been designed throughout the past few years to further explain pain patterns. Predictive models include a broad spectrum of statistical learning and reasoning techniques that mimic various elements of how the human brain works [8]. The methods and traits that allow machines to function versatile and effectively are divided into many categories. Machine and deep learning approaches use physiological and behavioral data to develop predictive models for automatic recognition of pain [8]. Several research investigations have delved into employing computational models and physiological signals to objectively estimate pain. The rising interest

[1] https://www.painscale.com/article/numeric-rating-scale-nrs

[2] https://www.mdcalc.com/calc/3701/pain-assessment-advanced-dementia-scale-painad

[3] https://www.mdcalc.com/calc/3621/nonverbal-pain-scale-nvps-nonverbal-patients

[4] https://www.va.gov/PAINMANAGEMENT/docs/DVPRS.pdf

in automated pain detection is presently revolving around the evaluation of its severity, and the distinction between the categories of "baseline" vs. "pain threshold" vs. "pain tolerance" [29]. However, the accuracy of these estimations has not met the standards suitable for clinical settings. Consequently, there is a requirement to create an advanced model that can be utilized to develop a dependable pain assessment tool suitable for both clinical and personal applications.

The efforts for developing automatic pain recognition frameworks have made substantial strides within the experimental context, achieving pain intensity recognition rates exceeding 80% for both unimodal and multimodal approaches through machine learning techniques [29]. Unimodal systems, are defined by their reliance on a single data channel, in contrast to multimodal systems, which integrate multiple modalities for analysis. Looking ahead, the next step involves creating a comprehensive framework for assessing pain severity in clinical scenarios. In order to achieve this goal, it is crucial to comprehend the anticipations and requirements of medical professionals and nursing staff regarding automatic pain recognition and assess their readiness to adopt this technology in clinical settings.

1.4 Physiological Signals

Physiological signals, covering a variety of bodily reactions, serve as impartial markers for pain. Key signals utilized in the development of automated pain assessment include heart rate variability, skin conductance, electroencephalogram, and electromyogram. These signals reflect the intricate responses of the autonomic nervous system to pain stimuli, providing valuable insights that extend beyond what can be conveyed through verbal expressions. Both short-term and long-lasting pain are associated with physiological changes, including heightened cardiovascular activity, resulting in fluctuations in heart rate and blood pressure [30] [31] [32].

Physiological signals are the description of a physiological phenomenon, which is generated by the body. Signals that capture the electrical activity generated by organs and tissues are frequently included in these modalities. They can be utilized to assess alterations during episodes of acute or chronic pain, employing minimally invasive or noninvasive methods. The ANS concurrently influences various physiological functions in the body during episodes of acute pain. Consequently, technologies capable of monitoring ANS activities are frequently utilized in order to capture them. Such activities encompass galvanic skin response, heart rate variability, electromyo-

gram, respiration, etc. For instance, GSR can detect sympathetic changes associated with pain by assessing sweat gland activity in response to pain stimuli [33] [34]. The sympathetic nerves affecting the respiratory system stimulate increased oxygen intake during acute pain. Additionally, the cerebral cortex's metabolic regions and muscular activity are also impacted by pain [35]. While these sensing technologies can be applied independently, the use of multimodality systems has also demonstrated promising outcomes.

Biosignals have been examined and analyzed across different scenarios to create pain assessment systems. Various sensors employed during the data collection phase determine the system's quantity and variety. Recent studies have validated the effectiveness of pain assessment methods employing multimodal signals, with some showing significant improvements over single-signal approaches [30]. However, despite these strides, the predominant focus in physiology-driven pain research is on establishing statistically significant correlations between various physiological signals and the presence of pain. Despite numerous commendable studies, a universally accepted pain assessment method based on physiological signals has yet to emerge

Variability in individual responses, the influence of contextual factors, and the need for real-time processing present hurdles in developing universally applicable automated pain assessment systems. The fusion of physiological signals and automated pain assessment represents a paradigm shift in how we understand and address pain. While challenges persist, the potential for improved accuracy, objectivity, and real-time monitoring offers a glimpse into a future where body language becomes a cornerstone in our approach to pain assessment and management.

1.4.1 Electrocardiogram

An electrocardiogram (ECG) offers a visual depiction of the cardiac rhythm, essentially a graphical display illustrating the heart's electrical behavior over a while, usually a few seconds [30]. It is a valuable tool for evaluating the heart's rhythm and detecting any abnormalities in the heart's electrical conduction system. The ECG comprises elements like the P-QRS-T waves, and these components are all affected by the heart's electrical functions. The length of these waves and their intervals supply vital insights into the heart's performance. Considering that the R wave is the most common peak within the QRS complex, its dependable detection is of paramount importance. The two key features that can be identified from an ECG are the heart rate (HR) and the heart rate

variability (HRV). Both of them are linked to ANS activity and can be used as a good proxy for measuring pain intensity [32]. Figure 1 illustrates the representation of the ECG in the Matlab environment for the BioVid subject **071309_w_21-BL1-081** and **071309_w_21-PA4-076** in the baseline and the higher level of pain respectively.

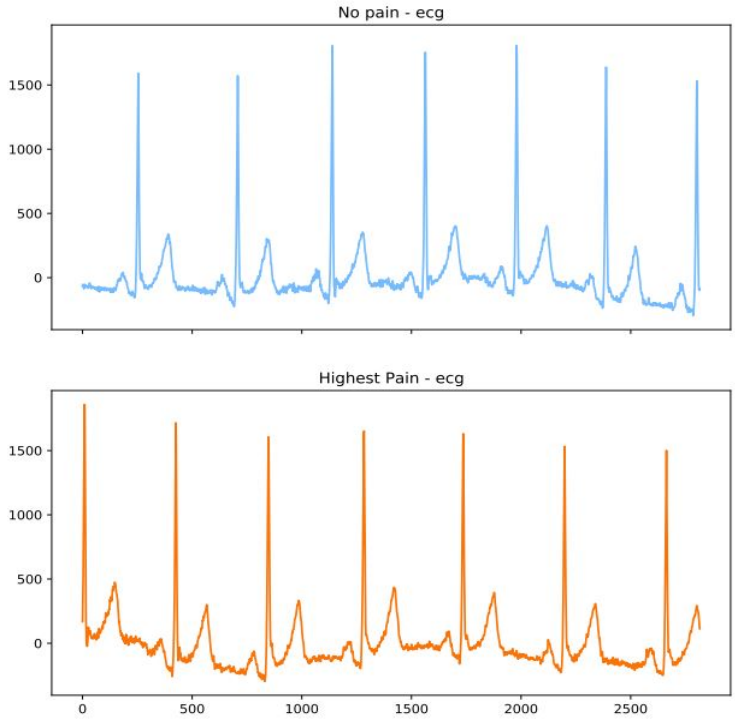


Figure 1: ECG in the baseline and the highest level of pain in Matlab environment

The QRS complex detection as a fundamental morphological element in ECG waves represents the initial stage in the analysis of ECG data for pain research. These peaks are often denoted as RR intervals in the literature. ECGs can show certain patterns or changes that might be associated with conditions that cause chest pain or discomfort. To ascertain whether a patient is encountering pain-related concerns during the examination of an electrocardiogram, healthcare providers typically seek distinct changes in the ECG. For instance, if the ST segment (the flat, isoelectric part of the ECG) is elevated, it can be indicative of myocardial infarction (heart attack) [32]. This is one of the most critical findings associated with severe chest pain. According to Moscato et al. [36], blood pressure and ECG parameters have been extensively examined with a notable emphasis on the LF/HF (low-frequency / high-frequency) ratio, a metric obtained from ECG data, attracting

substantial interest. ECG is the signal that has been studied the most in the literature because of the changes that occur in autonomous activity [20]. Among all the physiological signals recorded by monitors or wearable technology, ECG is useful for detecting the presence of pain.

1.4.1.1 Heart Rate

Heart rate, also referred to as pulse, is the count of heartbeats occurring within a minute and serves as a helpful prognostic marker for the health of the heart [30]. While the characteristics include the pace, volume, and amplitude, the average heart rate can vary from person to person. Adults typically have between 60 to 100 heartbeats per minute; heartbeats above or below this range are referred to as tachycardia and bradycardia, respectively [30]. Over time, changes in pulse and pace might happen, which may be a sign of a cardiac ailment or another physiological or pathological issue that has to be identified.

Several pain symptoms are present in almost all cardiac disorders, including arrhythmia, atrial fibrillation, and heart valve dysfunction [31]. These could include discomfort in the chest, breathing problems, or chest pain. Cases of chronic pain can cause a great deal of worry and anxiety and increase blood pressure and heart rate. A continuous increase in systolic pressure and pulse rate may cause cardiovascular problems, resulting in cardiac arrest, stroke, or death. Chronic tachycardia, or an HR of more than 100 beats per minute, can affect certain people with chronic pain [30]. However, physical activity, rest, anxiety, tension, a chronic disease, and medication are factors that can cause alteration in the heart rate.

1.4.1.2 Heart Rate Variability

Heart rate variability refers to the variation in the time intervals between successive heartbeats (the R-to-R-wave interval), reflecting the adaptability and responsiveness of the cardiovascular system to various stimuli and it can be used to evaluate the ANS's performance [10][12][27]. Heart rate variability is considered an important physiological parameter that reflects the balance between the SNS and PNS branches of the ANS, providing insights into the body's ability to respond to different situations and has found application in numerous clinical and research contexts, one of which involves the evaluation of pain [37].

Heart rate variability (HRV) is not indicative of the speed of the heartbeat; rather, it denotes the fluctuations in the time intervals between consecutive heartbeats [36]. High HRV signifies adaptability and resilience, while low HRV may indicate stress or impaired regulatory mechanisms. Leveraging HRV in automated pain assessment harnesses the body's innate responses to pain stimuli, offering a nuanced and objective perspective [36]. During periods of relaxation, the PNS becomes dominant, resulting in a gentle heartbeat. It is common for heart rate variability to be higher in such situations [37]. Furthermore, HRV has been studied in individuals with chronic pain conditions, such as fibromyalgia and chronic low back pain. Reduced HRV has been associated with chronic pain states, potentially reflecting a persistent dominance of sympathetic activity [32].

Usually, HRV features are calculated over either a short duration of 5 minutes or an extended period of 24 hours. Acute physiological changes may require the use of certain HRV parameters that can be analyzed in a very short time, specifically in less than a minute [27]. RR intervals inside each time frame may be used to extract HRV characteristics in different domains. In the current thesis, we focused on the time domain and the statistical analysis of HRV. Examples of statistical features are the average of RR intervals, the average of HR, the difference between the RR intervals of greatest and smallest duration, the variation in beats per minute throughout daylight and night, etc. [27].

Recent advancements in technology, particularly ML algorithms, have facilitated the assimilation of complex physiological data, including HRV, into automated pain assessment systems. These algorithms, trained on diverse datasets, enable the identification of patterns and correlations, enhancing the precision of pain intensity predictions. The link between technology and physiology opens new horizons in the quest for more objective and efficient pain evaluation.

In the pursuit of establishing a system for discerning between the presence and absence of pain, Lopez-Martinez and Picard [1] employed conventional machine learning methodologies, such as logistic regression, SVMs with various kernels, and RNNs, incorporating HR and HRV. As posited by Koenig et al. [37], HRV holds promise as a potential indicator of the autonomic response in adults to nociceptive stimuli. Consequently, our study centers on the examination of features related to HR and HRV, with the explicit objective of elucidating their correlation with the experience of pain.

1.4.2 Galvanic Skin Response

The Galvanic Skin Response (GSR) is considered an important indicator of the emotional state of an individual [34]. The palmar/plantar surfaces of the palms of the hands and feet have 600 to 700 glands/cm^2 eccrine sweat glands, which is the maximum density [38]. The glands can be stimulated during the "fight-or-flight" response because sympathetic innervation predominates. Their stimulation may be a reliable method for objective pain assessment research. Galvanic skin response can be influenced by pain and stress, and it is used in various contexts to assess emotional or physiological responses, including those related to pain. Figure 2 shows the representation of the GSR in the Matlab environment for the subject **071309_w_21-BL1-081** and **071309_w_21-PA4-076** in the baseline and the higher level of pain respectively.

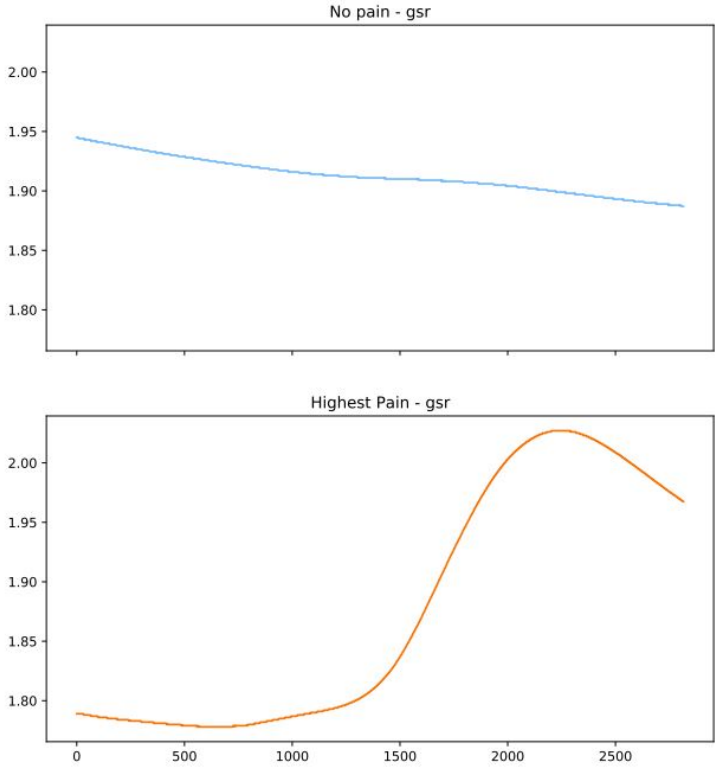


Figure 2: GSR in the baseline and in the highest level of pain in Matlab environment

The galvanic skin response is regarded as a biomarker that relates to pain perception and is later used to quantify pain in children and newborns [33] [34]. Susam et al. [39] discriminated between

different types of pain in children who underwent surgery using GSR data. Although sweat glands are found all over the body, they are most abundant on the palms of the hands, the forehead, and the cheeks. Patients may experience emotional sweating if subjected to certain stimuli that cause discomfort. When sweat glands become active, the presence of moisture on the skin enables a low-voltage current to pass through the measurement electrode. While perspiration evaporates, impedance increases, and conductance decreases once more. Eriksson et al. [33] and Munsters et al. [34] confirmed the link between GSR and newborns, establishing them as an appropriate dataset for automated pain evaluation, given their inability to communicate verbally. During blood sampling or care intervention, they discovered that the data could differentiate between pain and no pain.

1.4.3 Electromyogram

The electromyogram (EMG) records the electrical activity produced by skeletal muscles and represents a broad psycho-physiological state of arousal [40]. The EMG at the trapezius muscle, which is situated posterior to the shoulder, is included in the BioVid dataset. The trapezium's activity indicates a high degree of tension. The activity of the zygomaticus is related to euphoria, whereas the corrugator is related to negative emotions [40]. When someone is experiencing discomfort they are likely to be under a lot of stress, which is indicated by the activation of the trapezius [40]. EMG is commonly employed to diagnose and evaluate conditions that impact both muscles and the corresponding nerve control. Such conditions involve neuromuscular disorders, muscle-related ailments, and nerve compression. During an electromyogram, needles are positioned on the muscle, and the signals generated by muscle contractions are recorded and analyzed. When the pain is triggered, significant levels of stress are expected, and the stimulation of the trapezius is an indication of it [41].

Despite being a noninvasive measuring method, EMG signals are sometimes complicated and time-consuming to employ in practical settings. Figure 3 shows the representation of the EMG in the Matlab environment for the subject **071309_w_21-BL1-081** and **071309_w_21-PA4-076** in the baseline and the higher level of pain respectively.

EMG, skin temperature, HR, SCL, and blood volume pulse were used by the authors of [42] as input signals for defining emotions. Temporal parameters of the EMG signal include mean,

standard deviation, distance, and the first and second differences' absolute values. They also extracted the spectral coherence function's mean and standard deviation as frequency parameters and achieved an 85% recognition rate for various emotions.

Additionally, EMG has been mostly used in studies with low back pain (LBP). The results are consistent with the proposition that individuals undergoing pain frequently manifest initial indications of muscle fatigue in their myoelectrical activity. This implies that EMG may prove to be a valuable diagnostic instrument for identifying and assessing LBP. Recently, Rodrguez-Tapia et al. [35] and Campanini et al. [43] examined muscle activity and its research importance. They offered recommendations for analysis and discussed challenges to its practical implementation. In another study [44], sixty participants were categorized into two groups: patients who suffer from LBP (n=30) and those who do not suffer from LBP (n=30). The results indicate that individuals experiencing pain demonstrated significantly lower intensity levels in comparison to those without pain.

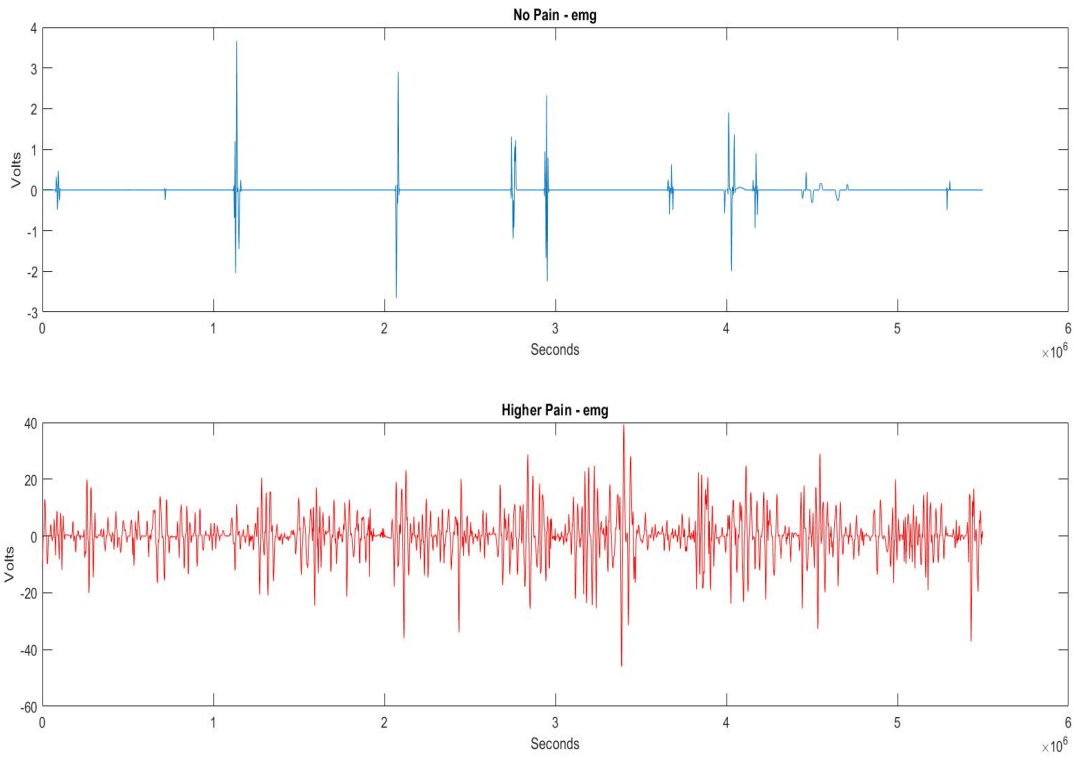


Figure 3: EMG in the baseline and the highest level of pain in Matlab environment

1.5 Objectives and Research Questions

Given that subjective tools such as self-reports are commonly employed for pain measurement, the goal of this thesis is to develop a framework for automated and objective pain assessment using multimodal physiological signals. Based on the elaborate review of relevant literature in Chapter II we have formulated the following research questions to be explored:

RQ1: Which is the most efficient combination of physiological signals for developing an automated pain assessment framework and what enhancements in result performance have been identified in contrast to unimodal strategies?

RQ2: Within the unimodal experiments context, which physiological signal emerges as a reliable indicator for pain assessment?

We conducted binary and multiclass experiments, unimodal and multimodal, to assess pain through physiological data-driven models. Furthermore, we aimed to thoroughly exploit all information within the dataset to investigate potential correlations between demographic data and pain. The BioVid dataset served as an input for the experimental pipeline. Subsequently, support vector machines with diverse kernels (linear/gaussian/polynomial), decision trees, and Long Short-Term Memory models were implemented as part of the experimental process with the Hold-Out and Leave One Subject Out cross-validation methods.

1.6 Thesis Outline

The sections of this thesis are structured as follows: Chapter II reports the related work on automated pain assessment based on physiological signals. Chapter III describes the approaches of the predictive models used in the literature. Chapter IV presents the implementation pipeline and the conducted experiments. Conclusions, study limitations, and considerations for future work are presented in Chapter V.

2. Systematic Literature Review

Pain assessment is an essential part of patient care. The subjective pain reports, which are the gold standard method for pain evaluation, are inadequate to observe the ongoing pain. The unbiased evaluation of pain poses a difficulty for medical professionals to appropriately manage and treat it.

Automated pain assessment, an evolving area of study, has increasingly captured the attention of researchers. The predominant focus in many studies revolves around models trained with images, aiming to capture body movements and facial expressions as indicators of pain. However, these expression-centric techniques come with several limitations, diminishing their practical applicability in clinical settings [30]. One notable concern is the potential discomfort and privacy issues experienced by subjects when being recorded for these assessments. Incorporating images and videos necessitates careful consideration of the surroundings and the challenge of keeping moving subjects in focus. Additionally, a significant drawback is the susceptibility of subjects to overstate the severity of their discomfort through facial expressions, often driven by a desire for medication [30].

Considerable research efforts are being directed toward the development of automated pain recognition systems, aiming to discern the presence and severity of pain by analyzing physiological and behavioral reactions. Over the last decade, researchers have focused on creating frameworks that enable systems to recognize sentiments, including pain. They have explored biological responses and ML approaches for objective pain evaluation, but have not yet achieved performance suitable for clinical contexts. Consequently, there is a need to develop an efficient system and a reliable pain evaluation tool usable in both laboratory and individual settings. The primary motivation for pain estimation using physiological signals stems from the ANS's function in pain perception [31] [32]. EDA, ECG, EMG, and other biosignals are considered strong candidates for creating an ML framework for objective pain evaluation [34] [35].

In this chapter we explored and elucidated methodologies and models designed for pain recognition and assessment, focusing exclusively on physiological signals.

2.1 Search Strategy

An extensive analysis of the available literature was performed according to PRISMA guidelines [45]. For this search, we accessed "PUBMED", "Science Direct", and "IEEE Xplore" online databases to identify publications concerning technology relevant to pain recognition and assessment. Moreover, the search was performed with specific combinations of keywords such as: "automated pain detection", "automated pain assessment", "physiological signals", "ECG", "GSR", "EMG", "machine learning", and "deep learning".

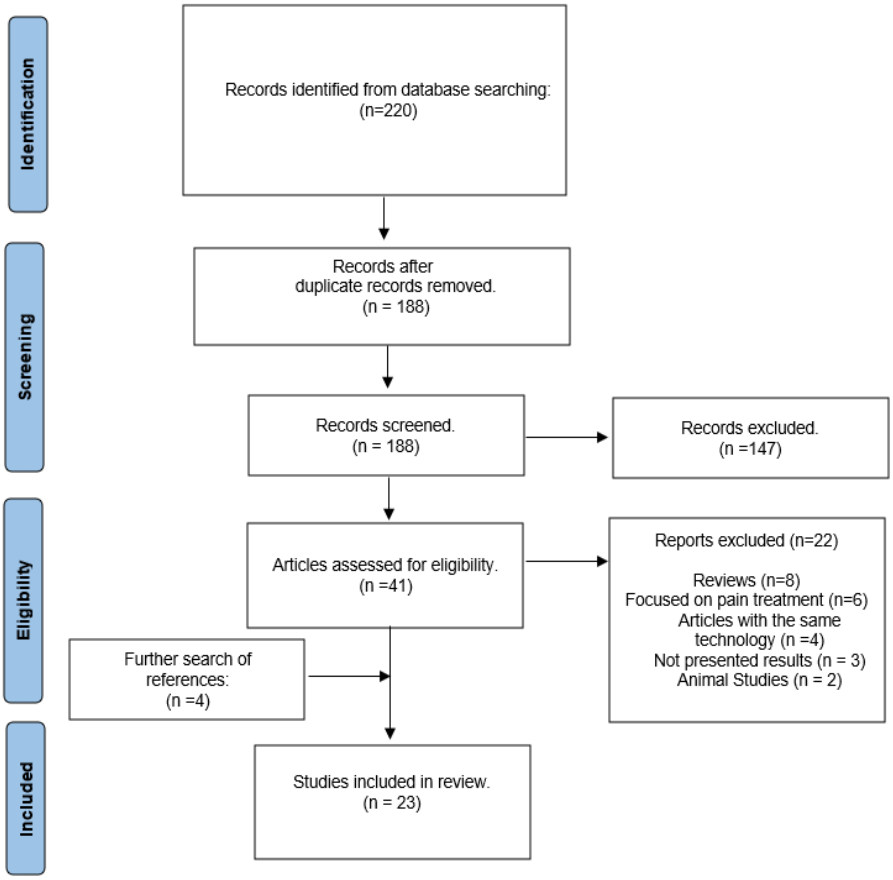


Figure 4: PRISMA flowchart illustrates the searching and screening process

To broaden the scope, we scrutinized the citations of selected literature, and an additional search on Google Scholar was executed to ensure comprehensive coverage. This search was carried out in July 2023. The PRISMA [45] flowchart in Figure 4 illustrates the research and filtering process

employed in the review. The primary focus of this review is on articles and studies that highlight automated pain detection or assessment through physiological signals within the last decade, aligning with current advancements in technology and methodologies.

Articles not directly related to this topic were excluded based on specific criteria. The exclusion criteria applied for this literature review are as follows: 1) Articles incorporating other channels in the experimental process, such as images or videos, 2) Studies of pain estimation lacking results, 3) Studies lacking an implementation approach, 4) Articles not sourced from scientific databases, 5) Animal subject studies, 6) Articles centered on pain treatment, 7) Review papers, and 8) Papers that describe the same technology.

2.2 Selection Criteria

We established inclusion criteria for the articles gathered to ensure the accuracy of the literature review concerning relevant studies. The papers were required to address the research questions and explore frameworks related to automated pain assessment using physiological signals through machine or deep learning approaches. Additionally, articles needed to be sourced from scientific publications or databases and be in either English or Greek. However, the ultimately selected papers for the review were exclusively in English. Another inclusion criterion was the publication date, with a focus on studies from the past ten years to maintain relevance. These four criteria were applied by scrutinizing each paper's title, abstract, and key terms. The selected papers that satisfied these conditions were thoroughly analyzed and are presented in the following subsection.

2.3 Related Work

Due to challenges in interpreting imaging and audio-visual modalities, especially in clinical settings where individuals may feign facial expressions or feel uncomfortable with recording, the majority of published studies rely on biological signals to better understand the pain experience. Consequently, numerous studies have explored the correlation between pain and various physiological responses, including cardiovascular activity, muscle function, electrodermal activity, brain function, and respiratory rate.

The papers that will be discussed employ various validation methods such as different splits

with the Hold-Out method and a special kind of k-fold validation the Leave-One-Subject-Out. Furthermore, the majority of studies place a significant emphasis on accuracy metric, especially in the context of machine learning algorithms. Accuracy serves as a metric for assessing a model's performance by calculating the ratio of correctly classified instances to the overall number of instances within the dataset. Additional information regarding each examined study is available in Table 1.

Table 1: Summary of the reviewed papers

Authors	**Modality**	**Dataset**	**Model**	**Validation**	**Result**	**Year**
Gruss et al. [40]	ECG EMG, GSR	BioVid	SVM	Hold-Out	90.94%	2015
Walter et al. [41]	ECG, EMG, GSR, Video	BioVid	ML	LOSO	70%	2015
Thiam et al. [46]	ECG, EMG, GSR	BioVid	2D-CNN	LOSO	84.40%	2019
Thiam et al. [47]	ECG, EMG, GSR	BioVid	DDCAE	LOSO	83.99%	2020
Lopez-Martinez et al. [48]	ECG, GSR	BioVid	LSTM-NN	LOSO	MAE:1.05, RMSE: 1.29, R^2: 0.24	2021
Aqajari et al. [50]	GSR	Post-operative Patients	KNN, RF	LOSO	84%, 70%	2021
Naeini et al. [51]	ECG	Post-operative Patients	SVM, RF KNN, XGB ADA	LOSO	84.14%, 70.84%, 70.92%, 69.57%, 75.68%	2021

Thiam et al. [52]	ECG, EMG, GSR, Video	BioVid	2D-CNN	LOSO	70%	2021
Wang et al. [53]	ECG, EMG, GSR	BioVid	Hybrid RNN-ANN	LOSO	83.3%	2020
Zhao et al. [54]	PPG, GSR, Skin Temperature	Chronic Pain Subjects	CNN	LOSO	70%	2020
Gkikas et al. [55]	ECG	BioVid	MT-NN	LOSO	69.40%	2023
Chu et al. [12]	BVP, ECG, SCL	Healthy	LDA, k-nn, SVM	LOSO	75%	2017
Kächele et al. [9]	GSR, ECG, EMG	BioVid	Adaptive RF	LOSO	85.7%	2016
Lopez-Martinez et al. [1]	GSR, ECG	BioVid	Neural Networks	10-fold	82.75%	2017
Gkikas et al. [49]	ECG	BioVid	SVM, LDA	LOSO	58.62%	2022
Subramaniam et al. [60]	EDA, ECG	BioVid	CNN-LSTM	ReLU	u: 85.65% m: 94.12%	2020
Nezam et al. [62]	EEG	Healthy Subjects	SVM, k-nn	LOSO	83±5%, 80±5%	2018
Elsayed et al. [63]	EEG	Healthy Subjects	ANN	Adam	94.83%	2020
Yu et al. [64]	EEG	Healthy Subjects	CNN	10-fold	97.37%	2019
Misra et al. [65]	EEG	Healthy Subject	SVM	LOSO	89.58%	2017

Levitt et al. [66]	EEG	Healthy & Lumbar Radiculopathy Patients	SVM	10-fold & LOSO	82.5%	2020
Vuckovic et al. [67]	EEG	Healthy & CNP Patients	ANN, SVM, LDA	LOSO	86%	2018
Susam et al. [39]	EDA	Children following laparoscopic appendectomy	SVM	LOSO	77.66%	2018

Considering that benchmark pain datasets include physiological data sources, Walter et al. [41] were the pioneers in utilizing the BioVid Heat Pain dataset. They employed these physiological sources to create machine learning models that could distinguish four different levels of intensity versus the presence of pain. Their feature extraction involved aspects such as magnitude, rate, stationarity, entropy, linearity, and variance. Gruss et al. [40] later expanded on this work by incorporating similar characteristics. Subsequently, Thiam et al. [46] developed a 2D model using these modalities. They concatenated and inputted it into a 9-layer 2D CNN using an early fusion approach. Their results highlighted a strong association between EDA and pain intensity. Interestingly, the amalgamation of information from all three sources (ECG, EDA, and EMG) did not yield higher scores compared to using EDA alone. In a subsequent study [47], they introduced a multi-modal data fusion strategy for binary tasks using biosignals from the same dataset, relying on deep denoising convolutional autoencoders.

Chu et al. [12] employed linear discriminant analysis, k-nn, and support vector machines in a dataset comprising six subjects with no medical history. The objective was to categorize patients' pain into five different levels using a multimodal approach that included ECG, BVP, and GSR. The results suggested that SVM algorithms exhibited slightly superior performance. As a result, the trained model could serve as a versatile tool for evaluating pain levels in new populations not previously studied, achieving an average accuracy of 75% across four distinct pain intensities. In a related study, Kächele et al. [9] utilized EDA, ECG, and EMG to extract skewness, standard deviation, and QRS complexes. They applied random forest classifiers to categorize pain, focusing on two groups: no pain and a specified discomfort level. Additionally, they developed a less restrictive pain assessment regression model. Their work introduced an initial approach to persistent pain assessment, using an Adaptive Random Forest model with biosignals [9].

Lopez-Martinez et al. [48] proposed an RNN model for assessing pain levels based on EDA. In their earlier investigation [1], they observed that a detailed analysis of HRV measures indicated an overall increase in HRV with pain. However, they noted that the considerable diversity in individual pain responses might constrain its effectiveness. They concluded that including individual differences through Multi-Task Learning led to enhanced accuracy in recognizing pain intensity compared to other approaches, although used only GSR and ECG.

To assist healthcare providers in managing hospitalized individuals, there is a critical need for a precise and reliable method of assessing acute discomfort and pain, particularly in postoperative patients. This entails the continuous monitoring of various biological indicators, such as HR, RESP, electrodermal activity, and others. Aqajari et al. [50] conducted a study involving 25 postoperative patients, collecting GSR signals using the Empatica E4 wristband. Two machine-learning algorithms were employed for model development, with each approach used for four binary classification problems to distinguish between the baseline and the four pain levels. Despite challenges in assessing actual clinical information, their models outperformed the BioVid paper approach [41] for the first three pain models. The maximum prediction accuracy for "baseline" vs "pain level 1", "baseline" vs "pain level 2", and "baseline" vs "pain level 4" was achieved using a random forest classifier (86.0%, 70.0%, and 61.5%, respectively). Additionally, a k-nearest-neighbor classifier was utilized, resulting in an accuracy of 72.1% for baseline versus pain level 3 (PL3). In a different study, Naeini et al. [51] gathered a group of 25 postoperative patients aged between 18 and 65. The

authors of this study extracted 19 time-domain HRV features and devised an automated framework to evaluate the subjects' pain. The highest accuracy was achieved using SVMs, performing 84.14% in distinguishing between the "baseline" and "pain level 2". Similar results were obtained by employing the SVM algorithm and Gini index for the top 8 features, with a performance of 84.79% for the same classification task ("baseline" vs. "pain level 2").

In recent years, there has been a rising interest in DL techniques for pain detection, prompted by their promising performance across various domains. Thiam et al. [52] introduced DL frameworks for pain classification, specifically developing 1D convolutional neural networks (CNN) on GSR data. They improved binary experiments, discerning between no pain and the highest intensity of pain. Additionally, they presented a multi-modal pattern-matching model based on deep denoising convolutional autoencoders for binary experiments utilizing biosignals from Walter et al. [41] dataset. Wang et al. [53] proposed a Bidirectional LSTM (Bi-LSTM) recurrent neural network model for the same dataset. For binary classification tasks, they combined neural networks with traditionally extracted features, unifying EDA and ECG signals. Notably, their model's performance surpassed that of the study by [41]. Moreover, it marked the first instance of an RNN network for binary classification of pain, achieving 82.7% for "no pain" versus "the highest level of pain", compared to 72.4% in the [41] study.

Furthermore, Lopez-Martinez and Picard [1] integrated ECG and EDA, creating a multitask neural network designed for personalized pain identification. In their follow-up research [54], they recommended an RNN based on deconvolved signals. This approach involved extracting tonic-phasic features from GSR and R peaks, along with interbeat intervals (IBIs) from ECG.

Additional integrations of similar signals have been recognized in the scientific community. Zhao et al. [56] combined PPG, GSR, and skin temperature. Following early fusion, they applied 2D convolutions to extract spatial features and leveraged time frames to capitalize on temporal data. In pain severity evaluation, Li et al. [57] utilized Motion Capture (MoCap) and EMG as inputs, conducting experiments with various Long Short-Term Memory (LSTM) structures. They found that employing a 3-layer fully connected network and a 3-layer vanilla LSTM yielded the most significant outcomes.

To the best of our knowledge, there has been limited research on automated pain assessment that considers demographic factors. Recent studies on pain have begun to focus on these factors, investigating their potential impact on discomfort intensity. This shift is driven by the recognition of social factors in nociception, highlighting the importance of a comprehensive examination of pain from both physiological and social viewpoints [58]. There is substantial evidence supporting the notion that gender influences pain modulation, leading to differences in pain perception between males and females. Women, compared to men, tend to experience heightened pain and discomfort in specific body areas [59]. Contemporary research indicates the presence of gender and age biases in pain intensity, becoming an increasingly discussed topic. A comprehensive review by Bartley et al. [58] concluded that, overall, women exhibit greater sensitivity to laboratory-induced pain than men. Additionally, Toomey et al. [59] demonstrated that females have lower pain thresholds, perceiving equivalent stimuli as more acute and severe. Their findings also revealed that older adults tend to have higher pain thresholds and pain ratings compared to younger individuals. Consequently, these findings support the theory that the pain system in the elderly engages slightly later, resulting in the perception of stimuli as more severe when recognized.

In the same experimental framework exploring the relationship between age and gender, Gkikas et al. [49] were involved. They examined ECG signals and employed SVM with various kernels for categorizing pain intensity. Their study involved leveraging the demographic data from Biovid, revealing substantial variations between genders. Males exhibited lower sensitivity, particularly in high pain intensities. The impact of age was also noteworthy, indicating that with increasing age, the sensation of pain diminishes, consequently posing a higher risk for additional injury. In a subsequent study [55], the use of MT-NN was employed utilizing the same signal. Initially, the feature vectors underwent augmentation by incorporating the subject's demographic elements, leading to an enhancement in classification accuracy. Following that, a multi-task learning NN was employed, merging the pain estimation task with gender and age estimation. This approach exhibited superior results in comparison to methods previously discussed in the paper and other related studies. Finally, Subramaniam et al. [60] investigated features of GSR and ECG. They developed a model comprising four convolutional layers followed by a single LSTM as an attribute generator. The combination of these signals for pain discrimination yielded the most favorable outcomes. Furthermore, the study highlighted the challenge of gender-oriented pain detection by

demonstrating differences between male and female populations.

2.4 Neurological Signals as Input in Pain Assessment

Given that the brain governs all responses to stimuli, it is posited that neurological responses are linked to processes such as sensory and motor integration [18] [20]. A growing body of research indicates that both anatomical and functional abnormalities in the brain are correlated with chronic pain [20] [61]. It is noteworthy that electroencephalography (EEG) can be employed to monitor these alterations and is considered another pain indicator. For example, neural activity in the sensorimotor cortex may signify various pain states [60]. EEG is recognized as a particularly useful noninvasive technique for assessing the severity of pain. Furthermore, EEG offers advantages in terms of clinical comfort, cost-effective setup, and maintenance.

Nezam et al. [62] utilized SVM and k-nn decision trees to collect EEG and EMG data, evaluating diverse pain levels. Both models demonstrated classification accuracy exceeding 80%. Due to its high sensitivity to noisy data, k-nn eliminates the requirement for an initial training phase. The integration of k-NN and SVM can create a decision tree, especially in datasets that include both categorical and numerical features. In another study, Elsayed et al. [63] classified neurological responses associated with pain into four severity levels (baseline, mild, moderate, and intense) through a combination of signal processing algorithms and ML methodologies. They identified an explicit association between alpha frequency band power and pain intensity, supported by signal processing evidence. The classifier's accuracy reached 94.83%. The consistent findings across various investigations suggest that normalized alpha power in the central part of the brain could function as a reliable and quantitative indicator of prolonged discomfort. In a related context, Yu et al. [64] introduced an innovative approach to assess subject-dependent tonic cold pain using EEG signals. Their methodology involved extracting diverse feature representations from the frequency bands of a 32-channel EEG signal. Afterward, these representations were combined to create a fully connected network to classify pain into three specific levels of tonic discomfort.

Misra et al. [65] recruited thirty healthy subjects and applied SVM with Gaussian kernel using EEG signals for pain classification into high and low categories. They employed theta and gamma power in the medial prefrontal region, along with lower beta power in the contralateral sensorimotor region, as the three features. The achieved accuracy, assessed through leave-one-subject-out cross-

validation, was 89.58%. In the context of deriving pain phenotypes from EEG data, Levitt et al. [66] employed SVM. This study collected EEG signals from 20 patients with persistent lumbar radiculopathy, 20 non-patient volunteers, and 17 patients with persistent lumbar pain scheduled for treatment and achieved an accuracy of 82.5%.

In the field of pain management, the development of predictive models can be extremely beneficial, particularly for complex chronic pain conditions that pose challenges in treatment. Vuckovic et al. [67] developed ANNs, SVMs, and linear discriminant analysis to identify spinal cord patients prone to developing central neuropathic pain. Each framework demonstrated an accuracy surpassing 85%. Furthermore, it has been established that features derived from electroencephalography (EEG) can be employed to predict the efficacy of pain therapies [68].

In conclusion, automated pain assessment and the analysis of neurological biosignals are essential facets of modern healthcare, presenting opportunities for enhanced diagnostic accuracy, personalized treatment plans, and a deeper understanding of pain mechanisms and neurological disorders. It is imperative to recognize that these technologies are dynamic and continually advancing, emphasizing the significance of ongoing research to refine their clinical applications.

2.5 Physiological-based Pain Datasets

The following table presents data sources used to conduct pain experiments in various studies. The inclusion criteria for the synthesis of Table 2 were datasets that include biosignals. The BioVid dataset is the most popular database in pain assessment through physiological signals. The majority of datasets are publicly accessible or by request, although a few studies experimented with private databases such as the X-ITE [72], EmoPain [71], and Sense-Emotion[73]. The research that focused on infant discomfort is also using private data.

Table 2: Pain datasets encompassing physiological signals

	Subjects	Channels	Annotations	Available
BP4D+ Database [69]	140 healthy	Video, HR, RESP, blood pressure, GSR	stimulus, FACS coding	Yes
BioVid Heat Pain Database [41]	90 healthy	Video, EDA, ECG, EMG	stimulus	By request
NPAD-1 [70]	6 newborns	Video, vital signs	self-report	By request
EmoPain [71]	22 chronic pain patients	Video, audio, EMG, mocap	self-report	No
X-ITE [72]	41 healthy	Video, audio, ECG, EDA, EMG	stimulus	No
Sense-Emotion [73]	45 healthy	Video, SCL, ECG, EMG, RESP	stimulus	No

2.6 Limitations and Open Issues

Automated pain assessment, while promising, is confronted with several limitations and open issues that present challenges to its comprehensive implementation. One significant constraint lies in the inherent subjectivity of pain, as individuals vary in their perception and expression of pain, making it difficult to create a universally applicable model [30]. This subjectivity introduces complexities related to cultural, environmental, and individual factors, impeding the seamless integration of automated systems. Moreover, individual diversity exacerbates these challenges, introducing a broad spectrum of pain responses. This diversity complicates the development of a standardized model that can effectively encompass such varied reactions.

Additionally, the lack of a universally acknowledged benchmark for pain assessment compli-

cates the validation of automated systems. Real-time difficulties emerge from the delay in physiological responses and the imperative for swift and accurate pain detection. External factors, like movement artifacts or unrelated emotional states, may introduce interference, potentially jeopardizing the precision of physiological signal measurements [30].

Another substantial limitation arises from the fact that the majority of proposed methodologies have been tested using induced pain data collected from healthy individuals. Despite the availability of data collections involving patients who have undergone postoperative pain [74] and those dealing with chronic musculoskeletal pain [75][76], there is a pressing need for the development of pain datasets in crucial contexts, such as cancer pain and primary chronic pain conditions. These datasets should include a diverse demographic range such as infants, adolescents, elderlies, and individuals with impairments, ensuring that the outcomes can be tailored to these distinct groups. Multimodal datasets could also include a wide range of data types such as information derived from body temperature, hormonal analyses, genetic profiling, and images obtained from the brain.

The intricacies of pain, being a subjective and complex phenomenon, unfold within a multifaceted terrain influenced by a variety of biological and psychological factors. This complexity is particularly pronounced when considering pain associated with cancer. For instance, cancer-related pain is inherently subjective, varying greatly among patients based on factors such as cancer type, stage, and individual pain thresholds [77]. Different types of cancer induce varying pathophysiological changes, leading to diverse pain manifestations. Cancer-induced neuropathic pain, for example, presents a unique pathophysiology compared to nociceptive pain. Understanding and incorporating these pain mechanisms into automated assessment models are essential but challenging due to the heterogeneity of cancer-related pain. Also, cancer treatments can cause specific issues with the body's autonomic functions, potentially affecting the analysis [77]. The dependability of physiological metrics may be affected by a range of factors, including medications, age, gender, and the particular health condition afflicting a patient [78].

Unresolved issues in the field add an extra layer of complexity. Combining various modalities, such as physiological signals, facial expressions, and contextual information, presents a significant challenge in developing a comprehensive pain assessment model. The lack of longitudinal studies impedes the assessment of the performance and reliability of automated pain assessment systems over extended periods, limiting understanding of their continued effectiveness. Addressing also

chronic pain conditions is a critical gap, as current research primarily focuses on acute pain.

Considerations of user acceptance and potential resistance from both patients and healthcare professionals are vital factors influencing the widespread adoption of automated pain assessment methods. Researchers and practitioners actively engage with these limitations and open issues, striving to advance the field of automated pain assessment and overcome practical challenges associated with its integration into healthcare settings. Continued exploration and innovation are critical for addressing these complexities and unlocking the complete potential of automated pain assessment grounded in physiological signals.

2.7 Ethical Concerns about Automated Pain Assessment

The transition from controlled research settings to real-world clinical applications necessitates addressing not only issues related to usability and integration with existing healthcare systems but also ethical concerns. Developing models that deliver accurate assessments and provide clear explanations for their decisions is imperative for gaining trust and acceptance in clinical settings. Ensuring data security and privacy measures, especially when dealing with sensitive data, demands stringent safeguards against unauthorized access or misuse [79].

The research context includes also considerations of clarity and ethical concerns. Ethical considerations, such as privacy, consent, and the potential misinterpretation of physiological signals, contribute significantly to the complexity of automated pain assessment. These aspects are indispensable components of AI, aiming to establish a transparent, trustworthy, and consistent framework aligned with ethical principles [79]. In AI, "clarity" denotes a model's ability to provide a precise and intelligible justification for each step in its decision-making process. This involves users, researchers, and developers in understanding how systems arrive at their findings and detecting any biases or inaccuracies that may arise. The ethical standards and regulations guiding the creation, implementation, and application of models are collectively known as AI ethics [79]. This encapsulates concerns of confidentiality, equity, openness, responsibility, and human autonomy [79]. To safeguard the rights and welfare of the research community and society, it is important to ensure that AI systems are developed and utilized ethically.

3. Predictive Models for Automated Pain Assessment

In pursuit of advancing pain management, research interest has been directed toward employing predictive models designed to clarify pain patterns. Presently, established methodologies for objectively appraising pain within clinical settings are absent. Specifically, the lack of dependable tools hinders clinicians from discerning the presence of pain in patients, assessing its severity, comprehending its ramifications on daily functioning, and evaluating the efficacy of interventions, whether procedural or pharmaceutical, in substantially mitigating pain levels. Hence, there is a need for predictive models that generate a dependable pain assessment tool applicable for both clinical and personal purposes.

Predictive models center on employing data and algorithms to mimic the learning processes of the human brain, progressively enhancing their outcomes [38]. By applying statistical techniques, algorithms are trained to furnish classifications and predictions that unveil crucial information for data mining projects. The objective is to establish decision-making systems that can impact significant results.

In this chapter, we will present and describe a series of techniques and algorithms employed in Chapter II for the automated classification and assessment of pain.

3.1 Artificial Neural Networks

Artificial Neural Networks (ANNs) are computational models that emulate the behavior of neurons in a biological brain. These networks consist of simulated neurons, each with connections resembling real axon-synapse-dendrite links, enabling communication with other neurons [80]. Figure 5 depicts the pathway of signals starting from inputs received at dendrites and culminating in outputs at the axon terminal. Upon reaching a certain threshold, signals from dendrites are transmitted through a neuron's axon, branching into the dendrites of multiple neurons [79]. Neurons adapt by adjusting the weights assigned to various inputs from other neurons. A basic neural network consists of an input layer, one or more concealed layers, and an output layer [79][80]. The first layer receives information from the external environment for analysis. This information is then passed to the hidden layer, which evaluates the results from the preceding step, refines them, and transmits them to the subsequent layer. The final stage is the output, with the number of nodes depending on

the nature of the problem, such as one node for binary classification or multiple nodes for multiclass scenarios. Each connection between nodes is associated with a weight, and every node processes an activation function on the calculated sum of its weighted inputs [80]. The weights and biases in the network are adjusted during the training process using optimization algorithms, often through a process called backpropagation [80].

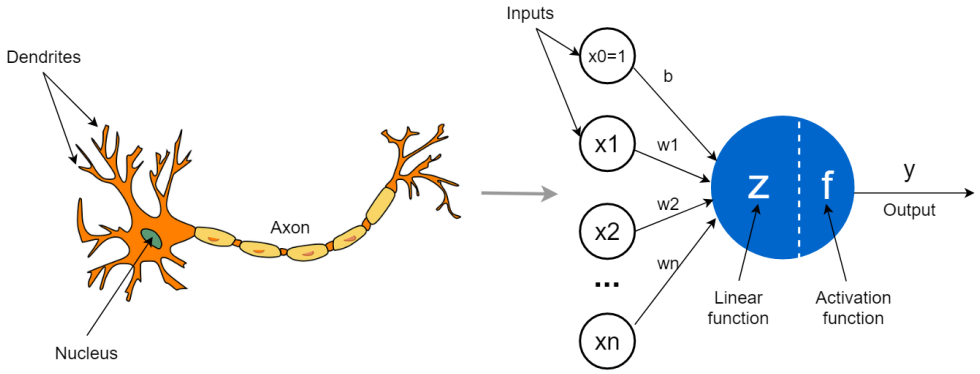


Figure 5: Illustration of brain neuron and ANN architecture

3.2 Recurrent Neural Networks

Recurrent neural networks[5] (RNNs) are a kind of Neural Network that uses the outputs from a preceding phase to serve as parameters for the subsequent phase. In typical neural networks, all inputs and outputs operate independently of each other. RNNs are networks with loops, which allow the persistence of data. They consist of three key elements: the input, the hidden layer, and the output. The input accepts the initial information provided to the network, which can consist of either a single data point or a sequence of data points. The crucial feature of an RNN is the hidden state, which retains specific information about a given sequence. The memory state enables the network to recall prior inputs. The output layer generates the network's outcome by utilizing the insights gleaned from the hidden layer(s). This result can be a singular value, suitable for tasks like regression, or a likelihood distribution across several categories, as needed for classification tasks. It generates results by applying the same procedures to each input or hidden layer while

[5]https://www.ibm.com/topics/recurrent-neural-networks

employing identical parameters for each input. In contrast to other neural networks, this simplifies the parameter set, reducing complexity. This chain-like behavior illustrates the close relationship between sequences, lists, and RNNs. Some of their main applications include image recognition, face detection, text generation, language modeling, time-series forecasting, etc.

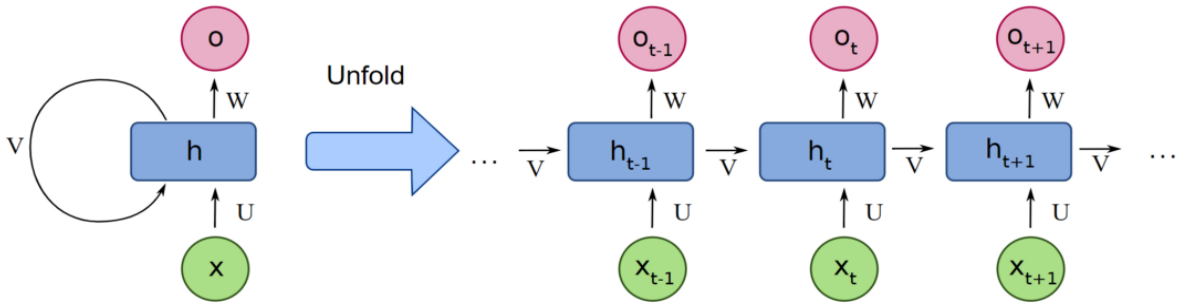


Figure 6: Recurrent Neural Network Architecture

3.3 Long-Short Term Memory Network

Long Short-Term Memory[6] (LSTM) represents a unique form of RNN structure. Hochreiter et al. [81] proposed a solution to tackle the challenge of enduring dependency in Recurrent Neural Networks. This solution aims to enhance the network's capacity to predict upcoming terms with higher precision based on recent data, as opposed to relying solely on terms stored in long-term memory.

The key principle underlying how LSTMs function is their utilization of two separate channels for making predictions, instead of relying on a single feedback loop to capture events from both recent occurrences and distant past [81]. Short-term memory traverses a distinct route, divergent from the path followed by long-term memory. In comparison to a basic vanilla Recurrent Neural Network, which unrolls from a relatively simple unit, the base of an LSTM is a considerably more complex entity. As shown in Figure 7, the fundamental structure of this model expands the boundaries of an RNN model by including three additional characteristics to the hidden layer: the input, the forget, and the output gate. The input determines if the information is important or not to transfer it and the level of its importance. The forget gate determines if the data is essential

[6]https://en.wikipedia.org/wiki/Long_short-term_memory

enough to be sent, and the output gate specifies the subsequent hidden state. Four connected layers communicate and create the structure of the central box, which is known as the LSTM cell. The LSTM cell consists of three states: a) the cell state, which functions as a path that allows data to flow along the sequence of events, b) the hidden state, which is used to encode the most recent time-step data, and c) the gates.

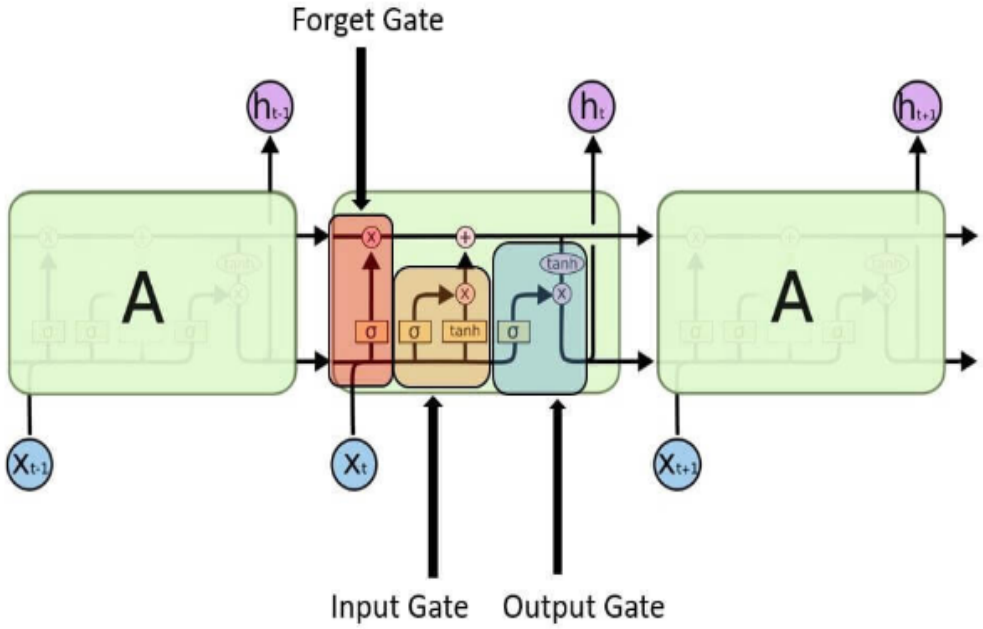


Figure 7: LSTM Network Architecture

Compared to RNNs, the LSTM model has the advantage of retaining memory-stored data for an extended period. LSTM cells are pivotal in the learning process as they possess the capacity to store information from past time steps [81]. Additionally, this type of network is a suitable method for categorizing, analyzing, and making predictions for time series with ambiguous temporal delays. Finally, the learning process of LSTMs enables models to be trained with thousands of time-step sequences, a task that an RNN typically finds challenging.

3.4 Support Vector Machine

Support Vector Machine[7] (SVM) belongs to the category of supervised machine learning techniques used for tasks involving classification and regression analysis. The SVM algorithm aims to identify the most effective hyperplane in an N-dimensional space for segregating samples into distinct classes. This hyperplane is positioned in a manner that maximizes the margin between the classes. By achieving these objectives, SVMs can efficiently classify data into different categories, rendering them a valuable tool in tasks related to machine learning and pattern recognition.

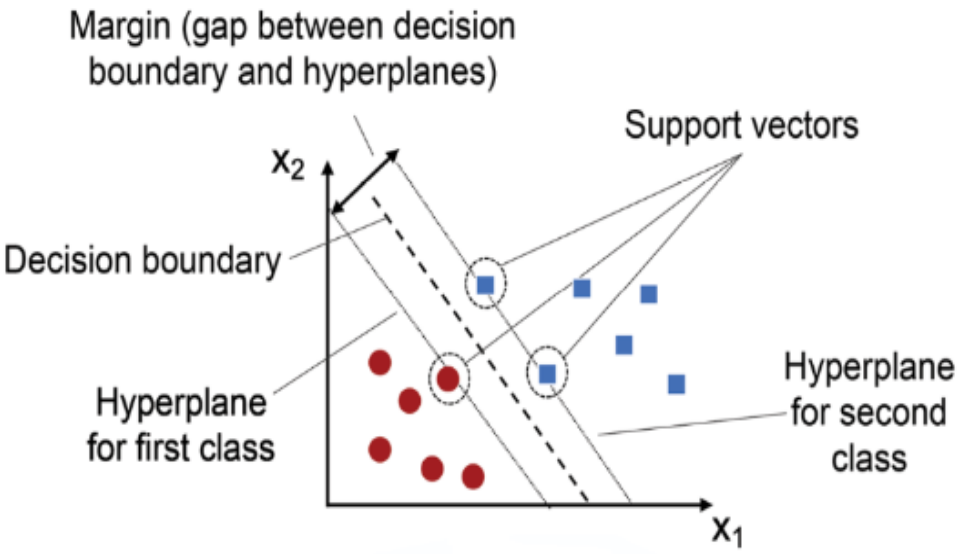


Figure 8: SVM Classification

The maximum width of the slice that extends parallel to the hyperplane with no internal support vectors is referred to as the margin. Support vectors are the closest training data points. The classifier's generalization error decreases as the margin increases. Thus, the decision function is the hyperplane. SVM is capable of efficiently performing both linear and nonlinear categorizations, thanks to techniques like kernel functions. However, the linear SVM formulation is the most straightforward. Support vectors have a range of different kernels[8] to preprocess the input data

[7] https://en.wikipedia.org/wiki/Support_vector_machine

[8] https://www.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html

into the desired form before processing it. For the purpose of this thesis, we utilized all kernels to compare them:

• **Gaussian**: $K(x_1, x_2) = exp(-\frac{||x_1 - x_2||^2}{2\sigma^2})$

It is utilized to execute the transformation, with no prior understanding of the data.

• **Linear**: $K(x_1, x_2) = x_1^T x_2$

It is utilized when data can be separated linearly.

• **Polynomial**: $K(x_1, x_2) = (x_1^T x_2 + 1)^\rho$

It illustrates the relationship between vectors in the training set within a feature space, considering the polynomials of the initial variables employed in the kernel.

3.5 Desicion Trees

The Decision Tree (DT) algorithm[9] stands out as a highly efficient ML technique for both classification and regression tasks. Each internal node represents an attribute, every branch denotes a result, and every terminal node signifies a category, forming a tree structure resembling a graph. The objective is to acquire simple decision rules derived from the attributes to construct a model capable of predicting the outcome of a given parameter.

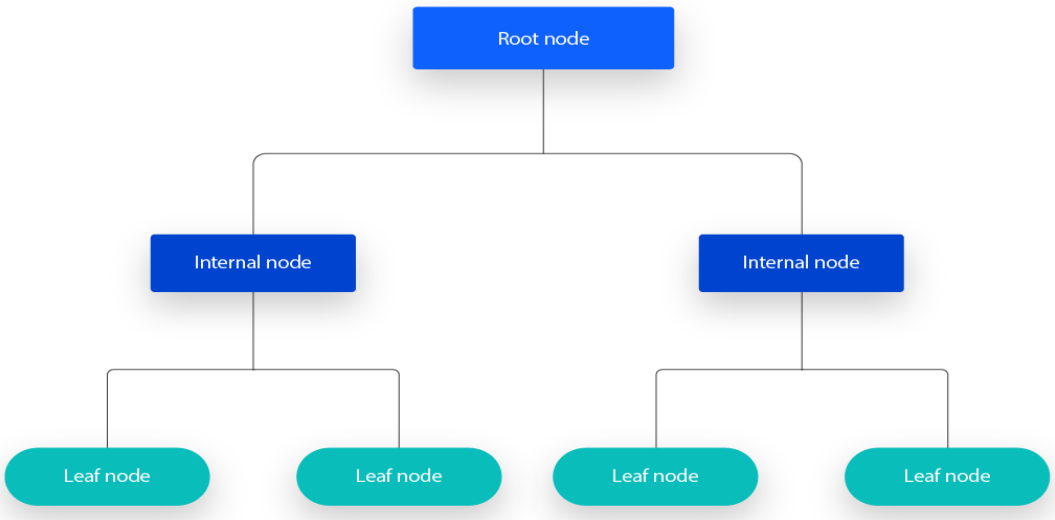


Figure 9: Decision Tree Architecture

[9]https://en.wikipedia.org/wiki/Decision_tree

Moreover, their inherent structure makes them easy to comprehend and interpret. Nevertheless, DT model trainers may produce overly intricate trees that struggle to generalize effectively with new input data. To alleviate overfitting, it is crucial to employ strategies such as specifying the minimum number of samples necessary at a leaf node or setting constraints on the maximum depth of the tree.

3.6 K-Nearest Neighbor Algorithm

KNN[10] is the simplest and most popular classification technique in ML. It belongs to the area of supervised learning and is extensively used in the fields of pattern recognition and data mining. While it is commonly used for classification, it can also be applied to regression problems. Due to its non-parametric architecture, KNN is more universally applicable in real-world situations than other algorithms. The k parameter defines the value of neighbors to classify a given data point. The class of the point is determined by the majority class among its k nearest neighbors. When selecting the value of k, it is crucial to consider a range of numbers to prevent both overfitting and underfitting. Lower numbers for k may result in high variation but low bias, whereas higher values of the parameter may introduce strong bias and low variance.

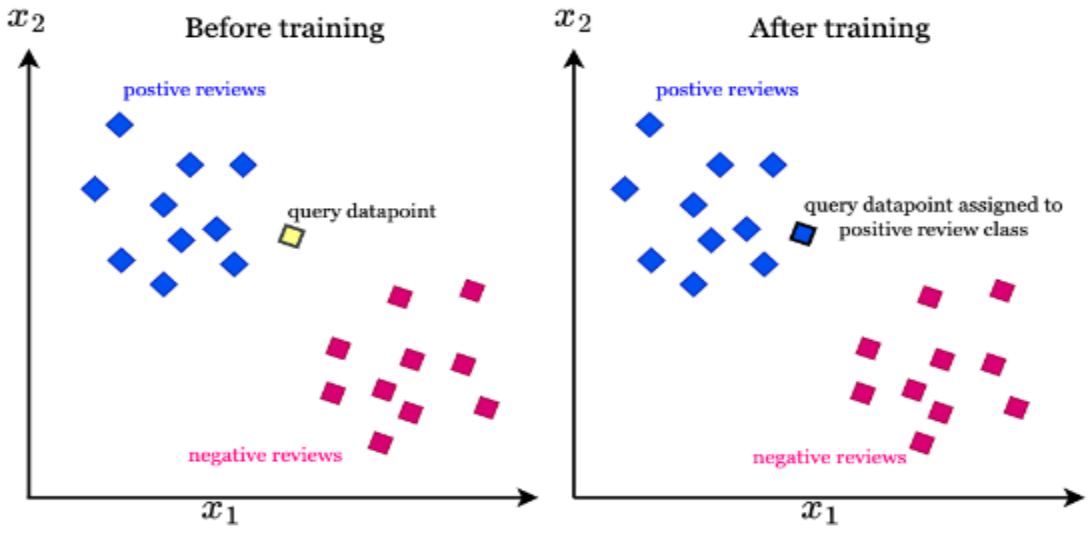


Figure 10: Example of K-NN Algorithm

[10]https://www.ibm.com/topics/knn

4. Implementation

This chapter provides a detailed description of the implementation process for both machine and deep-learning experiments as illustrated in Figure 11. The BioVid Heat Pain Dataset Part A, which includes ECG, GSR, and EMG signals, was employed as input in our binary and multiclass experiments. For the machine-learning approaches, the initial step involved the pre-processing step, aiming to handle missing values, tuples, and noisy data. Subsequently, we proceeded with the data transformation phase. Finally, our dataset was divided into three subsets: a) based on gender, b) based on the age group, and c) based on both gender-age. The next step was the feature extraction phase for 24 features. The Pan & Tompkins algorithm [82] employed ECG time-domain feature extraction, while statistical analysis was applied to GSR and EMG signals. We developed support vector machines with different kernels(Linear/Gaussian/Polynomial) and Decision Trees aimed at categorizing the presence or absence of pain, as well as the intensity of pain. Both approaches used the hold-out method.

Additionally, we explored deep-learning processes by applying Long-Short Term Memory models. Moreover, we transformed the data into a three-dimensional format suitable for LSTMs. The model architecture is based on stacked LSTM layers: the first layer consists of 5 units, the following three of 64 units, and the last layer of 5 units. These layers continue to process and analyze the information, capturing more complex patterns using the leave-one-subject-out cross-validation, ensuring unbiased and robust results.

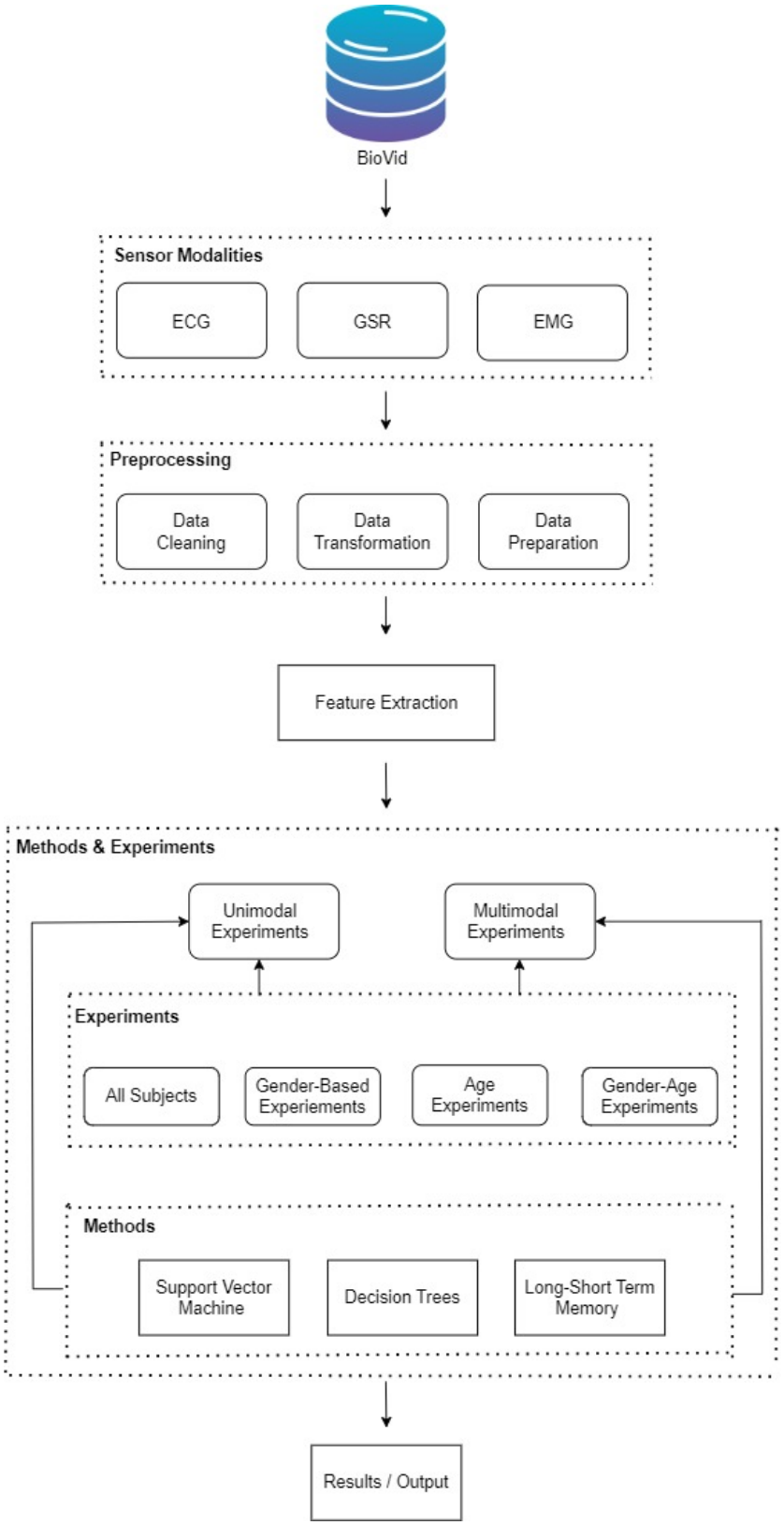


Figure 11: Description of the computational pipeline employed, including signal preprocessing, feature extraction, methods & experiments.

4.1 BioVid Dataset Description

The BioVid Heat Pain Database[11], introduced by Walter et al. [41] in 2013, incorporates biological signals and frontal video material for the detection and classification of heat-induced pain. Given that the BioVid dataset had already been segmented, there was no necessity for additional segmentation procedures. The study involved 90 participants, evenly distributed between genders and across three age categories: 20 to 35 years, 36 to 50 years, and 51 to 65 years. To induce pain in the right arm, a thermode, a small device affixed to the skin, was employed for heat stimulation. The experiment unfolded in six distinct phases, during which four different temperature stimuli were applied for 25 minutes, constituting the pain stimulation paradigm. Each temperature setting underwent 20 repetitions and lasted 4 seconds, interspersed with random intervals ranging from 8 to 12 seconds. The pain experiment occurred in two sessions: one without facial obstruction and another utilizing facial EMG sensors. BioVid includes five sub-datasets, each featuring a diverse set of modalities. This thesis focused on the Part A dataset, which encompasses physiological signals such as ECG, GSR, and EMG (Trapezius muscle). Part A is the most frequently referenced sub-dataset in the context of the research study.

Before the initiation of the experimental process, a thorough examination of every participant's medical history was conducted to ascertain their eligibility based on predefined criteria. The exclusion criteria covered pre-existing brain-related ailments, long-lasting pain, heart-related diseases, the use of pain relievers, and the consumption of painkillers immediately preceding the trial.

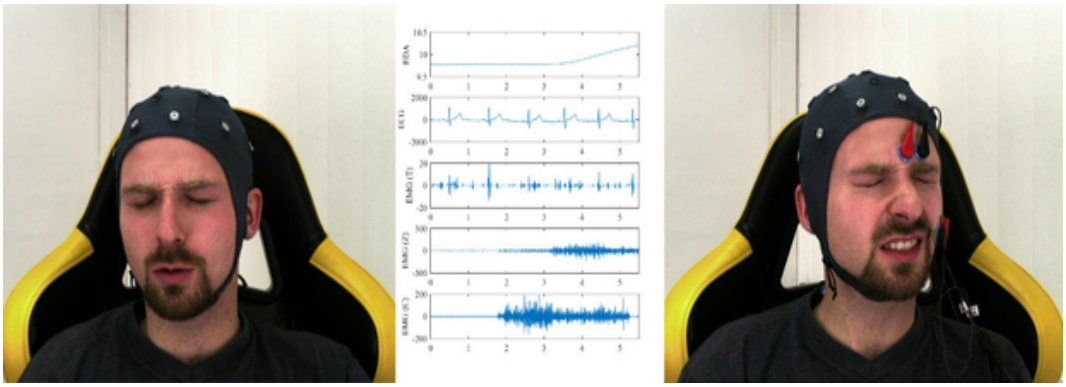


Figure 12: Example of pain stimulus[11]

[11] https://www.nit.ovgu.de/BioVid.html

A Nexus-32 amplifier was employed for the recording of the three physiological responses aimed at identifying pain across different sessions:

• GSR: Two sensors were strategically placed on the index and ring fingers to measure the galvanic skin response.

• ECG: Two electrodes were positioned on the upper right and lower left sides of the body to record participants' cardiac activity.

• EMG: Muscle activity from three specific areas (sEMG) was recorded using a two-channel surface EMG. Sensors were positioned on the Corrugator and Zygomaticus muscles of the face, and one on the shoulder (Trapezius).

The initial sample for the study consisted of 90 subjects; however, three patients' data were excluded due to technical difficulties during recording, resulting in a dataset containing information from 87 subjects. The data were recorded at a sampling rate of 512 Hz. BioVid includes pre-segmented intervals, each lasting 5.5 seconds, with a 3-second delay following the start of the stimulus. Consequently, a training sample from a single sensor channel forms a vector with a length of 2.816 (calculated as the product of a 512 Hz sampling rate and a 5.5-second duration). The discomfort intensities were determined based on data collected at a baseline temperature T_0 of 32 C. The dataset comprises four levels of pain discomfort, each with 100 data samples per individual, and each temperature applied 20 times. For each pain category and the baseline, 20 samples were collected, resulting in 8.700 samples used as input in our experiments.

4.2 Extracted Features

Table 3 presents all features extracted from ECG, GSR, and EMG modalities within the context of developing the pain assessment framework. Various studies have extracted these features by applying the Pan & Tompkins algorithm and statistical analysis. The relevant studies using these features are also shown in Table 3. Our objective was to employ them as inputs in our experimental process, compare our findings with existing research, and draw conclusions based on the results obtained in binary and multiclass tasks.

Table 3: List of the extracted features

#	**Feature**	**Signal**	**Group**	**References**	**Year**
1	Mean of Inter-beat Interval	ECG	Variability	[1],[9],[53]	2017, 2016, 2020
2	Heart Rate	ECG	Variability	[1], [9]	2017, 2016
3	Root Mean square of Successive Difference	ECG	Variability	[1],[9],[53]	2017, 2016, 2020
4	Standard Deviations of the NNs	ECG	Variability	[1],[9],[53]	2017, 2016, 2020
5	Slope of the linear regression	ECG	Variability	[1],[9]	2017, 2016,
6	The ratio of SDNN to RMSSD	ECG	Variability	[1],[9]	2017, 2016
7	Peaks (Maximum)	GSR	Amplitude	[1],[9], [53]	2017, 2016, 2020
8	Standard Deviation	GSR	Variability	[1],[9]	2017, 2016
9	Mean	GSR	Variability	[1],[9],[53]	2017, 2016, 2020
10	Root mean square	GSR	Amplitude	[1],[9], [53]	2017, 2016, 2020
11	Range	GSR	Variability	[1],[9],[53]	2017, 2016, 2020
12	Interquartile range	GSR	Variability	[1],[9],[53]	2017, 2016, 2020
13	MAVFD*	GSR	Amplitude	[1], [9], [53]	2017, 2016, 2020
14	MAVSD**	GSR	Amplitude	[1], [9],[53]	2017, 2016, 2020
15	MAVFD_std***	GSR	Amplitude	[9],[53]	2016, 2020
16	MAVSD_std****	GSR	Amplitude	[9],[53]	2016, 2020
17	Skewness	GSR	Variability	[9],[53]	2016, 2020
18	Kurtosis	GSR	Variability	[9],[53]	2016, 2020
19	Standard Deviation	EMG	Variability	[9],[53]	2016, 2020

20	Peaks (Maximum)	EMG	Amplitude	[9], [53]	2016, 2020
21	MAVFD*	EMG	Amplitude	[9], [53]	2016, 2020
22	MAVSD**	EMG	Amplitude	[9], [53]	2016, 2020
23	MAVFD_std***	EMG	Amplitude	[9], [53]	2016, 2020
24	MAVSD_std****	EMG	Amplitude	[9], [53]	2016, 2020

* Mean absolute value of the first differences;

** Mean absolute value of the second differences;

*** Mean absolute value of the first differences of the standardized signal;

**** Mean absolute value of the second differences of the standardized signal.

4.3 Pan & Tompkins Algorithm

The Pan & Tompkins algorithm [82] was implemented to identify the QRS complex in an electrocardiogram. The desirable features for the current thesis were: heart rate, mean of inter-beat intervals, RMSSD, SDNN, ratioSR, and the slope of linear regression. The primary apparent peak in an ECG signal is the QRS complex, which is indicative of ventricular depolarization. This aspect makes it especially appropriate for heartbeat estimation, which is the main method of evaluating the cardiovascular system.

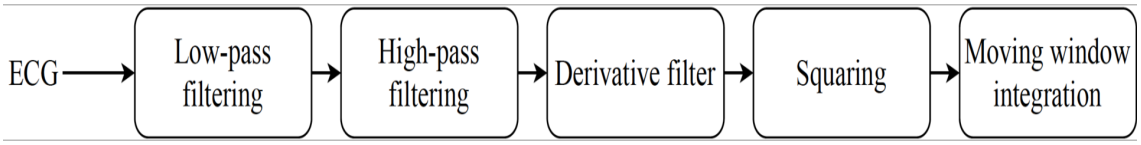


Figure 13: Steps of Pan & Tompkins algorithm

The algorithm consists of two phases. The first one is the preprocessing step and the second is the decision-making part. Preprocessing includes noise removal, signal smoothing, and amplification of the QRS complex's amplitude and slope. Pan & Tompkins's steps employ several filters to emphasize the frequency spectrum of the heart depolarization and eliminate unwanted interference such as baseline-wanders, electrical interference, electrode motion artifacts, or muscular interference [82]. After this step, the noiseless signal is forced to pass through differentiation to obtain the amplitude of the QRS complex. This mostly decreases low-frequency P and T wave parts while providing a substantial amplification to high-frequency components resulting from the QRS complex's steep slopes. Pan & Tompkins recommended applying filters with the transfer parameters that are presented in the following subsections, for data captured at a rate of 200 Hz [82].

4.3.1 Noise Cancelation

The application of a band-pass filter serves to mitigate the impact of "noise". A filter within the range of 5 to 15 Hz was achieved by sequentially cascading the Low Pass Filter (LPF) and High Pass Filter (HPF). The LPF was employed to eliminate high-frequency noise components, such as Power Line Interference, and T-wave interference, thus capturing the low-frequency signals. Simultaneously, the HPF was utilized to diminish low-frequency noise, including baseline wander. The band-pass filter is implemented through the combination of two transfer functions.

a. The second-order LPF transfer function:

$$H(z) = \frac{(1-z^{-6})^2}{(1-z^{-1})^2}$$

b. The HPF transfer function with a unity gain:

$$H(z) = \frac{(-1/32 + z^{-16} - z^{-17} + z^{-32}/32)}{(1-z^{-1})}$$

4.3.2 Derivative Filter

Following the application of the filtering procedure, the ECG signal was isolated, with a specific emphasis on determining the slope characteristics of the optimal QRS complex. During the dif-

ferentiation stage, the low-frequency P and T-waves were eliminated, aiming to isolate the high-frequency signals residing in the steeper slopes of the QRS complex.

$$H(z) = 0.1(-2z^{-2} - z^{-1} + z^{1} + 2z^{2})$$

4.3.3 Squaring and Integration

As a result, all of the sample's points become positive. This demonstrates the signal's higher frequencies, which are mostly attributable to the QRS complex. The final step is to apply the moving window integration in the output signal of squaring, which is used to get waveform feature data regarding the R wave's amplitude [82].

4.3.4 Fiducial Mark

The peaks are located in the integrated signal to identify a QRS complex. When a signal shifts from a higher to a lower position, it is denoted as a peak. No additional peak can be recognized for 200 ms after the last one.

4.3.5 Thresholds

Every peak from the previous step is regarded as a QRS complex. To decrease the chances of choosing the wrong peak as a QRS complex, we compare the peaks with a threshold.

Threshold1 = NoiseLevel1 + 0.25(SignalLevel1 - NoiseLevel1)

The limit value (threshold1) changed automatically after identifying a new peak, according to its categorization as an actual signal peak or noise peak.

4.3.6 Search Back for Missing Peaks

If the aforementioned process fails to detect a QRS complex, an additional search is initiated. If a QRS complex is not identified within a window spanning 166% of the average RR interval, the algorithm considers the maximum peak within that timeframe as a potential QRS and proceeds to identify it using 50% of the threshold values, as per the approach outlined by [82]. This method enhances the reliability of R peak detection by employing adaptive threshold adjustments. The

bandpass filter optimizes the waveform ratio for low thresholds, with the higher of the two thresholds in each set initially applied to the signal. In instances where no QRS is identified within a specified timeframe, the lower threshold is utilized, and a search-back method is employed to retroactively search for any missed peaks.

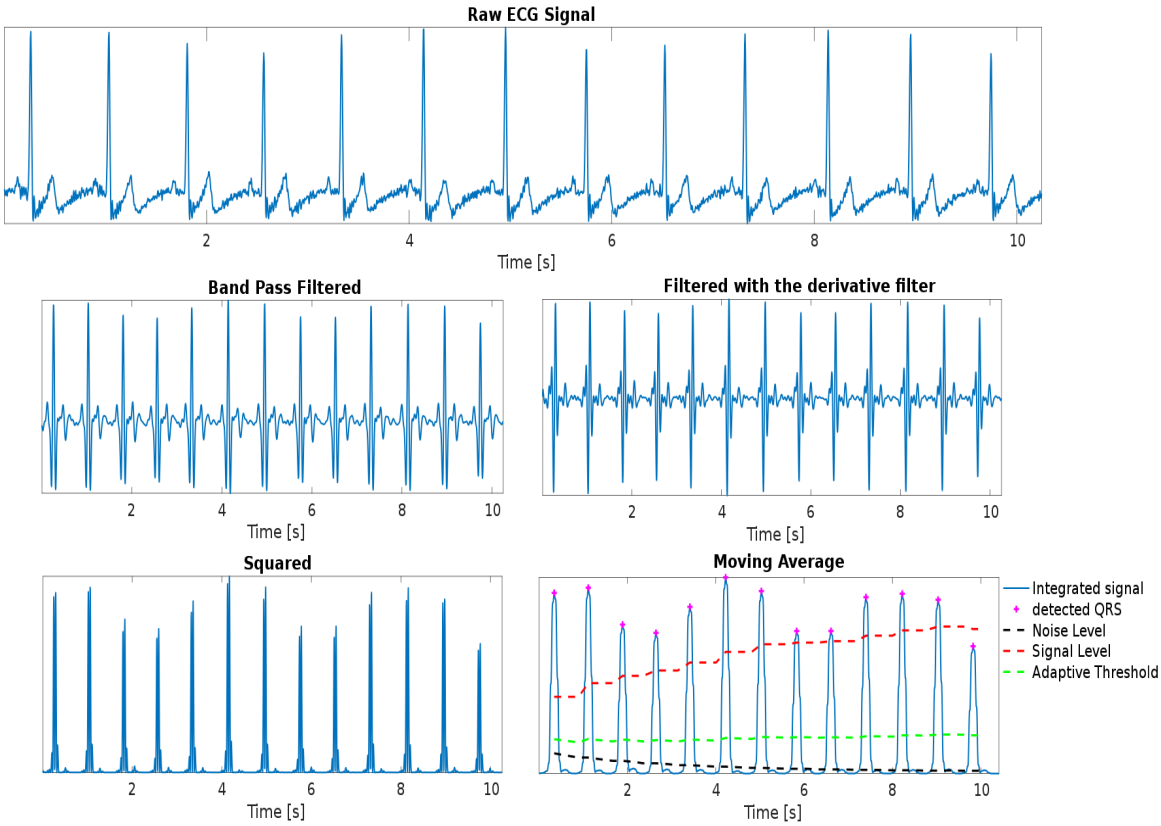


Figure 14: Pan-Tompkins preprocessing

4.4 Dataset Separation and Preparation

The dataset comprises 8.700 samples distributed across five categories, each containing 1.740 samples. We initially segmented the dataset into these five categories and allocated the samples to their respective groups. Additionally, we conducted data annotation to refine the categorization process in our subsequent models, intending to achieve more precise predictions and enhance the usability of the data. Essentially, data annotation contributes to the effectiveness of exploratory data analysis

by providing additional context to datasets.

The subsequent steps involve data importation and preprocessing. Missing values were randomly distributed, and in certain categories, such as the baseline, certain lines were entirely null. To address this, we imputed missing data by automatically replacing it with the mean value after loading the samples. After addressing the missing values, the next step was to obtain the values of our features. We observed variations between the values and features, so we had to normalize them to balance the data values and ensure there was no bias when used by the models.

The next step involved the normalization of features. The prevalent scaling technique employed is the normalization of data, which re-scales values to fall within the range of 0 and 1. This process, commonly known as Min-Max scaling, ensures that each feature contributes uniformly to the model and mitigates the influence of parameters with larger values.

8700x7 table

	1 HR	2 Mean_of_IBIs	3 RMSSD	4 SDNN	5 RatioSR	6 b1	7 class
1	0.7337	0.0738	0.0070	0.0141	0.4927	0.0247	0
2	0.6930	0.0887	0.0112	0.0139	0.3098	0.0380	0
3	0.6283	0.1150	0.0132	0.0180	0.3424	0.0432	0
4	0.7014	0.0855	0.0082	0.0091	0.2744	0.0377	0
5	0.6524	0.1047	0.0114	0.0115	0.2523	0.0555	0
6	0.6954	0.0878	0.0088	0.0132	0.3710	0.0379	0
7	0.6325	0.1132	0.0079	0.0093	0.2900	0.0573	0
8	0.6507	0.1054	0.0074	0.0110	0.3648	0.0566	0
9	0.6558	0.1033	0.0118	0.0104	0.2199	0.0554	0
10	0.7140	0.0808	0.0079	0.0136	0.4275	0.0359	0
11	0.7549	0.0666	0.0066	0.0123	0.4554	0.0346	0
12	0.6608	0.1013	0.0111	0.0162	0.3651	0.0429	0
13	0.6228	0.1174	0.0093	0.0142	0.3783	0.0439	0
14	0.6692	0.0979	0.0042	0.0047	0.2704	0.0543	0
15	0.8061	0.0502	0.0046	0.0077	0.4018	0.0217	0
16	0.6725	0.0966	0.0066	0.0075	0.2774	0.0406	0
17	0.7079	0.0831	0.0094	0.0110	0.2900	0.0383	0
18	0.6636	0.1002	0.0089	0.0102	0.2847	0.0554	0
19	0.6049	0.1254	0.0221	0.0306	0.3498	0.0635	0
20	0.7134	0.0811	0.0075	0.0083	0.2713	0.0271	0
21	0.5280	0.1641	0.0233	0.0310	0.3373	0.0902	0
22	0.5218	0.1675	0.0244	0.0302	0.3124	0.0701	0
23	0.4922	0.1847	0.0090	0.0191	0.5228	0.0962	0
24	0.4183	0.2339	0.0096	0.0134	0.3461	0.1123	0
25	0.4901	0.1859	0.0049	0.0071	0.3465	0.0998	0
26	0.5266	0.1649	0.0104	0.0131	0.3139	0.0704	0
27	0.4444	0.2154	0.0142	0.0120	0.2114	0.1076	0
28	0.4475	0.2133	0.0088	0.0084	0.2349	0.1078	0

Figure 15: Normalized data in ECG signal

4.5 Electrocardiogram Features

Butterworth band-pass filter (0.1-250 Hz) was used to remove unwanted frequencies from the ECG signal. We then extracted the HR and five HRV features through the Pan and Tompkins algorithm according to these studies [1] [52]. More specifically:

• Mean of the Inter-beat intervals (IBIs):

$$\mu = \frac{1}{N}\sum_{i=1}^{N}(RR_{i+1} - RRi)$$

• The heart rate:

$$HeartRate = \frac{60 \cdot Fs}{\mu}$$

• The root mean square of the successive differences (RMSSD):

$$RMSSD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(RR_{i+1} - RR)^2}$$

• The standard deviations of the NNs (SDNN):

$$SDNN = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(RR_i - \overline{RR})^2}$$

• The slope of the linear regression of IBIs:

$$A^T A_x = A^T b$$

• The ratio of SDNN to RMSSD, which is a metric of the heart's rate acceleration:

$$RatioSR = \frac{SDNN}{\text{RMSSD}}$$

4.6 Galvanic Skin Response Features

We extracted amplitude and variability features from the GSR signal, which have been employed in prior pain studies [1] [53]. Through the statistical analysis, we extracted the peak, range, standard deviation, interquartile range, root mean square, and mean value. Additionally, we computed the mean absolute value of the first and second differences for both raw and standardized signals. The standardized signals are then transformed into z-scores. More specifically:

• Maximum

• Standard Deviation

• Mean

• Root mean square

• Range

• Interquartile range

• MAVFD*

$$\frac{1}{N-1}\sum_{i=1}^{N-1}|x_{i+1}-x_i|$$

where N represents the count of data points within a single signal stream

• MAVSD**

$$\frac{1}{N-2}\sum_{i=1}^{N-2}|x_{i+2}-x_i|$$

where N represents the count of data points within a single signal stream

• MAVFD of the standardized signal

• MAVSD of the standardized signal

• Skewness

• Kurtosis

*MAVFD: Mean absolute value of first differences in GSR signal;

** MAVSD: Mean absolute value of second differences in GSR signal

4.7 Electromyogram Features

First, a Butterworth band-pass filter (20–250 Hz) is used for the EMG signal. Through the statistical analysis, we extracted the features below that were utilized in earlier pain studies [53][83]. The extracted features are:

• Maximum

• Standard Deviation

• MAVFD*

$$\frac{1}{N-1}\sum_{i=1}^{N-1}|x_{i+1}-x_i|$$

where N represents the count of data points within a single signal stream

• MAVSD**

$$\frac{1}{N-2}\sum_{i=1}^{N-2}|x_{i+2}-x_i|$$

where N represents the count of data points within a single signal stream

• MAVFD of the standardized signal

• MAVSD of the standardized signal

*MAVFD: Mean absolute value of first differences in EMG signal;

** MAVSD: Mean absolute value of second differences in EMG signal

4.8 Machine Learning Pipeline

The process outlined in Figure 16 delineates our ML approach to collecting, managing, analyzing, and classifying data. As already described, our source was the BioVid dataset, which includes ECG, GSR, and EMG modalities. Following data preparation and pre-processing steps, the feature extraction phase took place. Twenty-four (24) features were extracted and combined with early fusion, as mentioned, for our experimental processes in both unimodal and multimodal tasks, consistent with the approaches employed by most of the studies. This type of fusion involves combining features at the outset, before inputting the dataset into the computational algorithm, in contrast to late fusion which combines the data streams after they have been processed by the computational model separately [87]. Finally, support vector machines and decision tree algorithms were employed to execute the ML experiments.

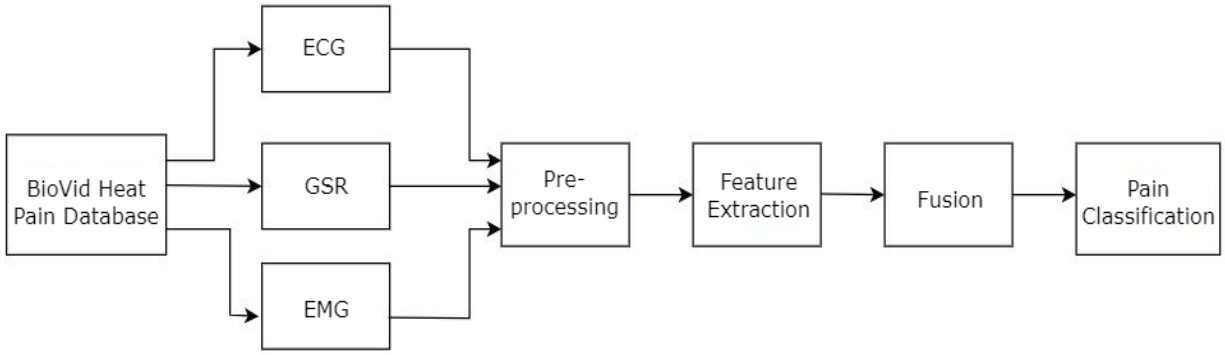


Figure 16: Machine Learning Process

The SVM experiments were carried out in the MATLAB[12] (MATrix LABoratory) environment. MATLAB serves as both a programming language and a widely utilized tool for tasks in machine learning and data science. Specifically designed for numerical computing and data analysis, MATLAB offers comprehensive support for machine learning through its dedicated toolbox. This toolbox encompasses a diverse array of algorithms and tools, addressing various aspects of ML such as supervised and unsupervised learning, feature selection, and model assessment. MATLAB's environment is conducive to the development of machine learning applications, emphasizing

[12]https://www.mathworks.com/products/matlab.html

user-friendliness, interactivity, and seamless integration with diverse disciplines.

4.9 LSTM Pipeline

The process involves several steps in preparing, transforming, and modeling the data for deep learning-based classification. Initially, the data undergoes a transformation process to convert it into a three-dimensional format suitable for Long Short-Term Memory (LSTM) modeling. This transformation facilitates the extraction of temporal dependencies in the data. Subsequently, the target variable is encoded using label encoding to enable binary and multi-class classification. Further preprocessing involves reshaping the input data to incorporate temporal information from preceding time steps, thus enhancing the model's ability to capture sequential patterns. Following these preparatory steps, an LSTM-based neural network model is constructed, comprising multiple LSTM layers to prevent overfitting. These LSTM layers are configured to accommodate the temporal structure of the input data, allowing the model to learn complex patterns. Throughout the training process, model performance metrics such as mean accuracy are monitored to assess their efficacy in classifying the input data.

The model architecture includes stacked LSTM layers for capturing temporal dependencies in the data. The first layer has 5 units and uses the hyperbolic tangent (tanh) activation function. The following three LSTM layers consist of 64 units. These layers continue to process and analyze the information, capturing more complex patterns. The last layer has 5 units and it refines the information further, summarizing key features. Finally, there is a Dense layer with five units for the five different categories (multiclass classification). Each unit in this layer represents a class, and the model endeavors to assign probabilities to different classes based on a given input. For the binary classification tasks, it has one unit, producing a single probability value.

Employing stacked LSTMs proves advantageous for several reasons. Sequential information processing stands out as a key strength of stacked LSTMs. Physiological signals related to pain often exhibit temporal dependencies, requiring a model that can effectively capture nuanced patterns evolving. Stacked LSTMs, designed for sequential data, are well-suited to address this temporal complexity. The hierarchical feature extraction capability of stacked LSTMs enhances their utility in pain assessment. By allowing each layer to learn different levels of abstraction, the model can automatically decipher complex structures within the input data. Moreover, the flexibility of

stacked LSTMs to handle variable-length sequences is essential in the context of pain-related physiological signals. Pain experiences can vary widely among individuals, and the model's capacity to adapt to different durations of signals contributes to its robustness. Stacked LSTMs offer an effective framework for seamlessly combining information from diverse physiological signals. The model's adaptability to multimodal input enhances its ability to capture the complexity inherent in pain perception. The architecture of the LSTM model is presented in Figure 17.

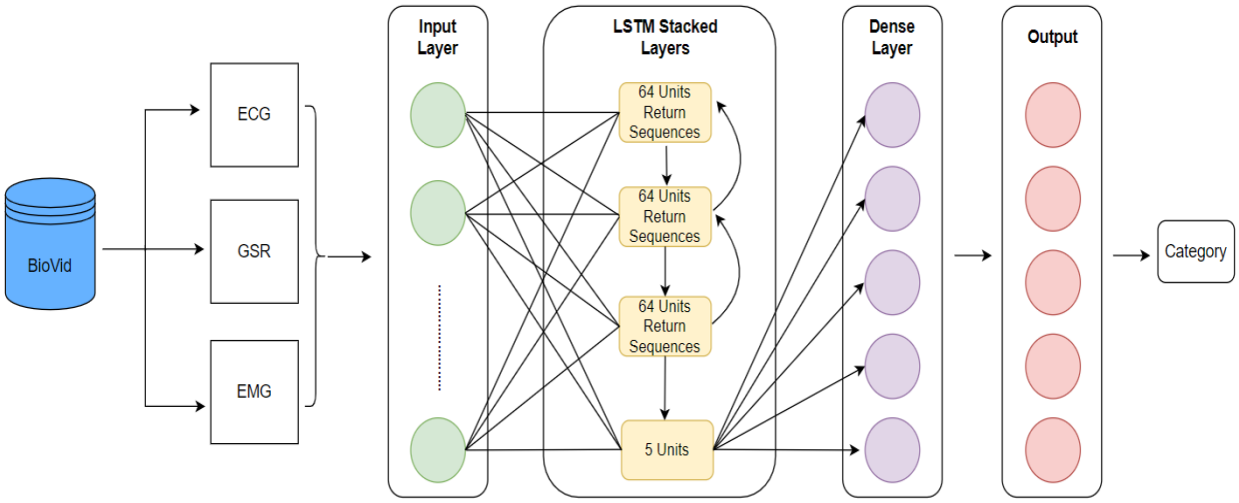


Figure 17: LSTM Model Architecture

The LSTM model was implemented with the Tensorflow[13] package, in Python language, using the Jupyter[14] Notebook environment. TensorFlow is an open-source ML framework developed by Google. It is created to facilitate the development, training, and deployment of machine learning models. Jupyter and TensorFlow are commonly used together for developing LSTM models, which are a type of recurrent neural network commonly used in sequence modeling tasks. By combining Jupyter's interactive and documentation capabilities with TensorFlow's deep learning tools, we can effectively develop, train, and evaluate LSTM models for sequence modeling tasks, such as time series forecasting.

[13]https://www.tensorflow.org/learn

[14]https://jupyter.org/

4.10 LSTM Model Data Preprocessing

The initial step in this process involves employing the missing values and the normalization of data. Furthermore, we used the "look back" technique, which is fundamental for analyzing time-series data. This method aids in capturing the influence of past data points on the current state by retrospectively examining one step at a time. This approach enables our model to discern patterns and trends over time, thereby enhancing its predictive capabilities. Additionally, we transform the data into a three-dimensional format suitable for Long Short-Term Memory modeling. This involves breaking the data into sequences of fixed length, which preserves the relationship between data points over time. We also use label encoding to handle the target variable, making it easier for our model to classify data into binary and multiple categories.

4.11 Evaluation Methods

Hold-Out Method

The dataset is partitioned into two distinct and non-overlapping sets: the training set, and the testing set, denoted. It is imperative to maintain the coherence of the data distribution to the greatest extent possible. In our case scenarios, we used different splits of 20%-80% and 40%-60% for our predictive models. Typically, experiments involve multiple iterations with random partitioning, and the average value is computed as the evaluation outcome. Approximately 2/3 percent of the dataset is commonly allocated for training, with the remaining samples reserved for testing.

Cross Validation Method

One commonly employed cross-validation method includes the leave-one subject out. Leave-one-subject-out (LOSO) involves using one original sample as the testing set, while the remaining samples constitute the training set. Although LOSO yields more precise results, it becomes computationally intensive for large datasets due to prolonged training times. In our experiments, we observed that the specific method required more time compared to the "Hold-Out" approach. Nevertheless, we posit that in experiments involving medical data or scenarios where the outcome pertains to patients, the LOSO method stands out as the most reliable validation option.

5. Results

This section presents the outcomes of the conducted experiments provided by SVM, Decision Trees, and LSTM models with the two validation methods. We employed the classification algorithms to classify pain and pain intensities and discover correlations between pain thresholds and gender-age variables. The classification processes were structured around pain categories and were executed in multiclass and binary approaches. We conducted five categories of experiments, each with a distinct objective: 1) Binary and multiclass classification between BL vs. All Pain Levels for all subjects in both unimodal and multimodal experiments, 2) Binary and multiclass experiments between pain levels in both unimodal and multimodal experiments, 3) Gender-based classification tasks, dividing the dataset into male and female groups, 4) Age-based classification tasks, creating three groups based on subjects' ages: 20-35 years, 36-50 years, and 51-65 years, and 5) Gender-age based classification tasks. The outcomes for each relevant task, along with the model used, are presented in Tables 4-32.

5.0.1 Unimodal Experiments

It has been noted that galvanic skin response (GSR) demonstrates superior performance in unimodal experiments compared to electrocardiography (ECG) and electromyography (EMG). Specifically, the highest accuracy percentages in binary experiments were observed between baseline versus higher-intensity conditions. In machine learning experiments, GSR was achieved in binary and multiclass tasks 75.60% and 35.77%, while in Long Short-Term Memory (LSTM) experiments, achieved mean accuracy of 76.86 and 37.52%.

Table 4: Binary ECG Experiments Between BL vs Pain Levels

		Support Vector Machine Kernels (6)		
All Subjects	Validation	Linear	Gaussian	Polynomial
BL vs P1	LOSO	51.20%	51.63%	**51.86%**
BL vs P2	LOSO	**52.87%**	52.61%	52.84%
BL vs P3	LOSO	55.22%	**55.37%**	55.14%
BL vs P4	LOSO	58.01%	58.30%	**58.39%**

Table 5: Binary ECG experiments Between Pain Levels

		Support Vector Machine Kernels (6)		
All Subjects	Validation	Linear	Gaussian	Polynomial
P1 vs P2	LOSO	**51.26%**	51.12%	50.91%
P1 vs P3	LOSO	53.01%	53.73%	**54.08%**
P1 vs P4	LOSO	56.58%	57.29%	**57.35%**
P2 vs P3	LOSO	51.52%	52.87%	**53.21%**
P2 vs P4	LOSO	55.02%	56.03%	**56.12%**
P3 vs P4	LOSO	53.07%	**53.44%**	53.33%

Table 6: Accuracy of Multiclass ECG Experiments

		Support Vector Machine Kernels (6)		
All Subjects	Validation	Linear	Gaussian	Polynomial
BL vs Pain Levels	LOSO	23.18%	23.79%	**23.87%**
Pain Levels	LOSO	27.93%	29%	**29.06%**

Table 7: Binary GSR Experiments Between BL vs. Pain Levels

		Support Vector Machine Kernels (12)		
All Subjects	Validation	Linear	Gaussian	Polynomial
BL vs P1	LOSO	50.73%	51.03%	**51.46%**
BL vs P2	LOSO	51.60%	53.04%	**57.04%**
BL vs P3	LOSO	59.91%	61.78%	**65.45%**
BL vs P4	LOSO	68.78%	71.37%	**75.60%***

Table 8: Binary GSR Experiments Between Pain Levels

		Support Vector Machine Kernels (12)		
All Subjects	Validation	Linear	Gaussian	Polynomial
P1 vs P2	LOSO	51.12%	51.43%	**53.16%**
P1 vs P3	LOSO	60.08%	58.90%	**62.41%**
P1 vs P4	LOSO	65.77%	68.53%	**72.52%**
P2 vs P3	LOSO	53.85%	55.11%	**55.51%**
P2 vs P4	LOSO	63.10%	63.73%	**67.12%**

Table 9: Accuracy of Multiclass GSR Experiments

		Support Vector Machine Kernels (12)		
All Subjects	Validation	Linear	Gaussian	Polynomial
BL vs Pain Levels	LOSO	25.35%	27.50%	**29.67%**
Pain Levels	LOSO	33.08%	33.72%	**35.77%****

* Highest SVM performance in unimodal binary task classification;

** Highest SVM performance in unimodal multiclass task classification;

BL: baseline, P1: minor, P2: tolerant, P3: intense, P4: very intense;

Pain Levels: P1 vs P2 vs P3 vs P4

Table 10: Binary EMG Experiments Between BL vs Pain Levels

		Support Vector Machine Kernels (6)		
All Subjects	Validation	Linear	Gaussian	Polynomial
BL vs P1	LOSO	**49.97%**	49.88%	49.91%
BL vs P2	LOSO	52.09%	52.41%	**52.67%**
BL vs P3	LOSO	52.09%	52.41%	**53.67%**
BL vs P4	LOSO	53.79%	55.11%	**55.37%**

Table 11: Binary EMG Experiments Between Pain Levels

		Support Vector Machine Kernels (6)		
All Subjects	Validation	Linear	Gaussian	Polynomial
P1 vs P2	LOSO	51.60%	51.95%	**52.24%**
P1 vs P3	LOSO	51.60%	51.95%	**52.24%**
P1 vs P4	LOSO	53.21%	54.79%	**54.82%**
P2 vs P3	LOSO	50%	50%	50%
P2 vs P4	LOSO	50.83%	**52.29%**	52.12%
P3 vs P4	LOSO	50.83%	**52.19%**	52.12%

Table 12: Accuracy of Multiclass EMG Experiments

		Support Vector Machine Kernels (6)		
All Subjects	Validation	Linear	Gaussian	Polynomial
BL vs Pain Levels	LOSO	21.05%	**22.05%**	21.94%
Pain Levels	LOSO	26.10%	27.34%	**27.38%**

Table 13: LSTM Binary ECG Experiments

Group	Classes	Model	Epochs	Batch Size	Accuracy
All Subject	BL vs P1	LSTM	100	64	51.20%
All Subject	BL vs P2	LSTM	100	64	52.47%
All Subject	BL vs P3	LSTM	100	64	54.39%
All Subject	BL vs P4	LSTM	100	64	57.58%
All Subject	P1 vs P2	LSTM	100	64	51.14%
All Subject	P1 vs P3	LSTM	100	64	54.16%
All Subject	P1 vs P4	LSTM	100	64	57.78%
All Subject	P2 vs P3	LSTM	100	64	52.32%
All Subject	P2 vs P4	LSTM	100	64	55.77%
All Subject	P3 vs P4	LSTM	100	64	53.53%

Table 14: Multiclass ECG Experiments

All Subjects	Model	Epochs	Batch Size	Accuracy
BL vs Pain Levels	LSTM	100	64	23.54%
Pain Levels	LSTM	100	64	28.87%

Table 15: LSTM Multiclass GSR Experiments

All Subjects	Model	Epochs	Batch Size	Accuracy
BL vs Pain Levels	LSTM	100	64	30.67%
Pain Levels	LSTM	100	64	**37.52%**××

Table 16: LSTM Binary GSR Experiments

Group	Classes	Model	Epochs	Batch Size	Accuracy
All Subject	BL vs P1	LSTM	100	64	52.01%
All Subject	BL vs P2	LSTM	100	64	55.60%
All Subject	BL vs P3	LSTM	100	64	65.34%
All Subject	BL vs P4	LSTM	100	64	**76.86%**×
All Subject	P1 vs P2	LSTM	100	64	54.91%
All Subject	P1 vs P3	LSTM	100	64	62.70%
All Subject	P1 vs P4	LSTM	100	64	74.02%
All Subject	P2 vs P3	LSTM	100	64	54.56%
All Subject	P2 vs P4	LSTM	100	64	69.33%
All Subject	P3 vs P4	LSTM	100	64	60.54%

Table 17: Multiclass EMG Experiments

All Subjects	Model	Epochs	Batch Size	Accuracy
BL vs Pain Levels	LSTM	100	64	23.03%
Pain Levels	LSTM	100	64	**28.49%**

× Highest LSTM mean accuracy in unimodal binary task classification;

×× Highest LSTM mean accuracy in unimodal multiclass task classification;

BL: baseline, P1: minor, P2: tolerant, P3: intense, P4: very intense;

Pain Levels: P1 vs P2 vs P3 vs P4.

Table 18: LSTM Binary EMG Experiments

Group	Classes	Model	Epochs	Batch Size	Accuracy
All Subject	BL vs P1	LSTM	100	64	50%
All Subject	BL vs P2	LSTM	100	64	53.70%
All Subject	BL vs P3	LSTM	100	64	53.85%
All Subject	BL vs P4	LSTM	100	64	56.83%
All Subject	P1 vs P2	LSTM	100	64	53.10%
All Subject	P1 vs P3	LSTM	100	64	53.13%
All Subject	P1 vs P4	LSTM	100	64	57.87%
All Subject	P2 vs P3	LSTM	100	64	50%
All Subject	P2 vs P4	LSTM	100	64	52.27%
All Subject	P3 vs P4	LSTM	100	64	52.70%

5.0.2 Multimodal Experiments

Table 19: Binary All Subjects Experiments

			Support Vector Machine Kernels (24)		
Group	Classes	Validation	Linear	Gaussian	Polynomial
All Subjects	BL vs P1	LOSO	51.78%	51.52%	**52.38%**
All Subjects	BL vs P2	LOSO	55.31%	55.97%	**58.47%**
All Subjects	BL vs P3	LOSO	59.48%	61.37%	**65.97%**
All Subjects	BL vs P4	LOSO	69.39%	71.46%	**76.69%**
All Subjects	P1 vs P2	LOSO	52.41%	53.01%	**55.57%**
All Subjects	P1 vs P3	LOSO	56.60%	57.90%	**61.40%**
All Subjects	P1 vs P4	LOSO	66.72%	68.33%	**72.87%***
All Subjects	P2 vs P4	LOSO	62.44%	63.79%	**68.18%**
All Subjects	P3 vs P4	LOSO	58.07%	58.64%	**61.12%**

Table 20: Binary All Subjects Experiments - Hold Out

Group	Classes	Model	Validation	Accuracy
All Subjects	BL vs P1	Decision Tree (24)	40%-60%	50.86%
All Subjects	BL vs P2	Decision Tree (24)	40%-60%	54.66%
All Subjects	BL vs P3	Decision Tree (24)	40%-60%	63.43%
All Subjects	BL vs P4	Decision Tree (24)	40%-60%	71.33%
All Subjects	P1 vs P2	Decision Tree (24)	40%-60%	51.07%
All Subjects	P1 vs P3	Decision Tree (24)	40%-60%	59.19%
All Subjects	P1 vs P4	Decision Tree (24)	40%-60%	67.31%
All Subjects	P2 vs P3	Decision Tree (24)	40%-60%	50.71%
All Subjects	P2 vs P4	Decision Tree (24)	40%-60%	64.58%
All Subjects	P3 vs P4	Decision Tree (24)	40%-60%	56.53%

Table 21: Multiclass All Subjects Experiments

			Support Vector Machine Kernels (24)		
Group	Classes	Validation	Linear	Gaussian	Polynomial
All Subjects	BL vs Pain Levels	LOSO	27.51%	27.91%	**30.79%**
All Subjects	All Pain Levels	LOSO	33.21%	33.66%	**37.09%**

Table 22: LSTM Binary Experiments

Group	Classes	Model	Epochs	Batch Size	Accuracy
All Subject	BL vs P1	LSTM	100	64	51.40%
All Subject	BL vs P2	LSTM	100	64	56.81%
All Subject	BL vs P3	LSTM	100	64	63.13%
All Subject	BL vs P4	LSTM	100	64	**77.21%**$^{+}$
All Subject	P1 vs P2	LSTM	100	64	55.94%
All Subject	P1 vs P3	LSTM	100	64	60.45%
All Subject	P1 vs P4	LSTM	100	64	73.33%
All Subject	P2 vs P3	LSTM	100	64	55.11%
All Subject	P2 vs P4	LSTM	100	64	56.60%
All Subject	P3 vs P4	LSTM	100	64	61.29%

Table 23: LSTM Multiclass Experiments

Group	Model	Classes	Accuracy
All Subject	LSTM	BL vs All Pain Levels	30.74%
All Subject	LSTM	P1 vs P2 vs P3 vs P4	**37.74%**$^{++}$

+ Highest mean accuracy in multimodal binary task LSTM;

++ Highest mean accuracy in Multimodal LSTM Experiments;

BL: baseline, P1: minor, P2: tolerant, P3: intense, P4: very intense;

Pain Levels: P1 vs P2 vs P3 vs P4.

5.0.3 Gender Experiments

Furthermore, we investigated the impact of demographic factors on experimental outcomes, emphasizing the pivotal roles of gender in shaping perceptions of pain. Our conducted experiments revealed significant disparities between genders, indicating that males tend to exhibit diminished sensitivity when confronted with high levels of pain. Specifically, a noteworthy distinction emerged in an experiment comparing baseline conditions to the highest level of pain (PA4), wherein female participants achieved an accuracy rate of 77.61% in SVM and a mean accuracy of 79.88% in

LSTM, surpassing the 72.61% and 70.85% observed among males, respectively. Additionally, the divergence in pain perception between the two genders becomes more pronounced in experiments featuring lower pain intensity, with women consistently registering higher percentages compared to men.

Table 24: Binary Males Experiments

			Support Vector Machine Kernels (24)		
Group	Classes	Validation	Linear	Gaussian	Polynomial
Males	BL vs P1	LOSO	52.27%	52.50%	**53.63%**
Males	BL vs P2	LOSO	55.39%	55.45%	**57.67%**
Males	BL vs P3	LOSO	58.35%	54.43%	**65.51%**
Males	BL vs P4	LOSO	67.17%	65.51%	**72.61%**•
Males	P1 vs P2	LOSO	55.28%	55.28%	**60.28%**
Males	P1 vs P3	LOSO	60.68%	62.21%	**67.55%**
Males	P2 vs P3	LOSO	52.27%	52.32%	**54.43%**
Males	P2 vs P4	LOSO	59.37%	59.31%	**64.03%**
Males	P3 vs P4	LOSO	53.75%	54.60%	**56.70%**

Table 25: Accuracy of Multiclass Gender Experiments

			Support Vector Machine Kernels (24)		
Group	Classes	Validation	Linear	Gaussian	Polynomial
Female	BL vs Pain Levels	LOSO	27.25%	28.74%	**31.35%**
Female	Pain Levels	LOSO	33.69%	34.91%	**38.45%**
Male	BL vs Pain Levels	LOSO	25.13%	25.50%	**28.70%**
Male	Pain Levels	LOSO	29.31%	30.96%	**33.86%**

Table 26: Binary Females Experiments

			Support Vector Machine Kernels (24)		
Group	Classes	Validation	Linear	Gaussian	Polynomial
Females	BL vs P1	LOSO	50.69%	50.75%	**51.86%**
Females	BL vs P2	LOSO	54.30%	54.53%	**58.83%**
Females	BL vs P3	LOSO	59.53%	60.29%	**64.47%**
Females	BL vs P4	LOSO	68.89%	71.51%	**77.61%**••
Females	P1 vs P2	LOSO	53.08%	52.73%	**55.93%**
Females	P1 vs P3	LOSO	57.26%	57.67%	**60.87%**
Females	P1 vs P4	LOSO	66.91%	69.01%	**74.88%**
Females	P2 vs P3	LOSO	53.54%	54.94%	**55.93%**
Females	P2 vs P4	LOSO	63.66%	64.94%	**60.87%**
Females	P3 vs P4	LOSO	58.66%	58.89%	**74.88%**

Table 27: LSTM - Mean Accuracy of Multiclass Gender Experiments

Group	Classes	Epochs	Batch	Accuracy
Females	BL vs Pain Levels	100	64	31.44%
Females	Pain Levels	100	64	**38.74%**
Males	BL vs Pain Levels	100	64	28.95%
Males	Pain Levels	100	64	**33.49%**

• Highest performance in male population;

•• Highest performance in female population;

BL: baseline, P1: minor, P2: tolerant, P3: intense, P4: very intense;

Pain Levels: P1 vs P2 vs P3 vs P4.

Table 28: LSTM Binary Female Experiments

Group	Classes	Model	Epochs	Batch Size	Accuracy
Females	BL vs P1	LSTM	100	64	50.41%
Females	BL vs P2	LSTM	100	64	58.37%
Females	BL vs P3	LSTM	100	64	63.95%
Females	BL vs P4	LSTM	100	64	**79.88%**⊕
Females	P1 vs P2	LSTM	100	64	57.32%
Females	P1 vs P3	LSTM	100	64	62.61%
Females	P1 vs P4	LSTM	100	64	77.50%
Females	P2 vs P3	LSTM	100	64	57.32%
Females	P2 vs P4	LSTM	100	64	71.56%
Females	P3 vs P4	LSTM	100	64	64.30%

Table 29: LSTM Binary Males Experiments

Group	Classes	Model	Epochs	Batch Size	Accuracy
Males	BL vs P1	LSTM	100	64	52.10%
Males	BL vs P2	LSTM	100	64	56.70%
Males	BL vs P3	LSTM	100	64	63.69%
Males	BL vs P4	LSTM	100	64	**70.85%**
Males	P1 vs P2	LSTM	100	64	54.14%
Males	P1 vs P3	LSTM	100	64	57.78%
Males	P1 vs P4	LSTM	100	64	67.44%
Males	P2 vs P3	LSTM	100	64	54.94%
Males	P2 vs P4	LSTM	100	64	62.78%
Males	P3 vs P4	LSTM	100	64	57.15%

⊕ LSTM best binary gender performance;

BL: baseline, P1: minor, P2: tolerant, P3: intense, P4: very intense;

Pain Levels: P1 vs P2 vs P3 vs P4.

5.0.4 Age Group Experiments

Regarding the factor of age, notable variations were detected. These differences suggest that as individuals age, their sensitivity to pain decreases. In specific demographic groups, this divergence in pain perception surpasses 10%, underscoring the considerable variability in how individuals undergo the experience of pain. Differences between genders were noticeable in terms of pain sensitivity thresholds, particularly among younger individuals, but these distinctions faded in the elderly population. The study found that young females had significantly lower pain detection and tolerance thresholds compared to young males. However, these gender-related differences became less pronounced and statistically insignificant in older individuals.

Table 30: Accuracy of Binary Females Experiments

				Support Vector Machine Kernels (24)		
Gender	Age Group	Classes	Validation	Linear	Gaussian	Polynomial
Females	20-35	BL vs P1	LOSO	51.66%	54%	**54.50%**
Females	20-35	BL vs P4	LOSO	73.83%	76.83%	**82.83%**$^{\diamond}$
Females	20-35	P1 vs P4	LOSO	56.66%	57.50%	**81.33%**
Females	36-50	BL vs P1	LOSO	48%	48%	**51%**
Females	36-50	BL vs P4	LOSO	57.33%	59%	**75.83%**
Females	36-50	P1 vs P4	LOSO	56.66%	57.50%	**71.33%**
Females	51-65	BL vs P1	LOSO	50.19%	50.96%	**51.15%**
Females	51-65	BL vs P4	LOSO	54.13%	55.38%	**60%**
Females	51-65	P1 vs P4	LOSO	56.15%	56.15%	**62.50%**

Table 31: Multiclass Experiments in Gender-Age Groups

			Support Vector Machine Kernels (24)		
Group	Classes	Validation	Linear	Gaussian	Polynomial
Females 20-35	BL vs All Pain Levels	LOSO	28.53%	29.60%	**33.46%**$^{\diamond\diamond}$
Females 36-50	BL vs All Pain Levels	LOSO	21.86%	22.46%	**27.53%**
Females 51-65	BL vs All Pain Levels	LOSO	28.17%	28.84%	**30.96%**
Males 20-35	BL vs All Pain Levels	LOSO	28.06%	28.46%	**30.86%**
Males 36-50	BL vs All Pain Levels	LOSO	21.21%	20.92%	**23.71%**
Males 51-65	BL vs All Pain Levels	LOSO	21.26%	20.46%	**23.66%**

Table 32: Accuracy of Binary Males Experiments

				Support Vector Machine Kernels (24)		
Gender	Age Group	Classes	Validation	Linear	Gaussian	Polynomial
Males	20-35	BL vs P1	LOSO	**55%**	53.33%	53%
Males	20-35	BL vs P4	LOSO	77.66%	77.83%	**80.16%**$^{\diamond\diamond\diamond}$
Males	20-35	P1 vs P4	LOSO	51.66%	53%	**74.16%**
Males	36-50	BL vs P1	LOSO	48.57%	48.75%	**51.78%**
Males	36-50	BL vs P4	LOSO	56.603%	58.03%	**65.35%**
Males	36-50	P1 vs P4	LOSO	54.46%	55.89%	**60.17%**
Males	51-65	BL vs P1	LOSO	**53%**	51.83%	51.33%
Males	51-65	BL vs P4	LOSO	54.16%	54.50%	**55.83%**
Males	51-65	P1 vs P4	LOSO	49.33%	50.66%	**52%**

$\diamond$Highest multimodal performance in binary task classification between female population;

$\diamond\diamond$ Highest multimodal performance in multiclass classification;

$\diamond\diamond\diamond$Highest multimodal performance in binary task classification between male population;

BL: Baseline, P1: minor, P2: tolerant, P3: intense, P4: very intense;

Pain Levels: P1 vs P2 vs P3 vs P4

5.1 Discussion

The outcomes of the experiments conducted with LSTM, SVM, and DT models, are presented in Tables 4-32, depicting classification accuracy, with the number of features employed in each model indicated in parentheses. The accuracy metric was used in the majority of papers referenced in the literature; therefore, we chose it to compare our results with other studies. This thesis focused on two classification tasks: the identification of pain and the classification of its intensity.

Multimodal approaches outshine unimodal methods, in both binary and multiclass tasks. The integration of all fused modalities resulted in further enhancements in performance compared to a single modality. In machine learning experiments, it is noteworthy that the polynomial kernel consistently yielded superior accuracy across all classification tasks, either binary or multiclass. Our experimental results fell within the range observed by most studies, surpassing some of them, especially in machine learning algorithms. Table 33 summarizes all the multimodal approaches along with the models used with the BioVid dataset as input with machine and deep learning techniques.

Table 33: Comparison of multimodal studies used BioVid in accuracy

Study	Modalities	Machine Learning	BL vs PA4
Werner et al. [2]	ECG, EMG, GSR	Random Forest	74.10%
Lopez-Martinez et al. [54]	ECG, GSR	Support Vector Machine	72.20%
Thiam et al. [46]	ECG, EMG, GSR	2D CNN	76.72%
Thiam et al. [52]	ECG, EMG, GSR, Video	DDCAE	70%
Subramaniam et al. [60]	EDA, ECG	CNN-LSTM	94.12%
Wang et al. [53]	ECG, EDA, EMG	RNN	82.7%
Wang et al. [53]	ECG, EDA, EMG	Hybrid RNN-ANN	83.3%
Walter et al. [41]	ECG, EMG, GSR	Support Vector Machine	70%
Ours	ECG, GSR, EMG	Support Vector Machine	76.69%
Ours	ECG, GSR, EMG	Long short-term memory	77.21%

Unimodal studies along with accuracy and model-used approaches are presented in Table 34. It was established that galvanic skin response emerged as the most effective modality, while the combination of ECG, EMG, and GSR proved to be the most effective of fused modalities. Our results yielded an accuracy of 75.60% in the SVM binary experiment (pain versus no pain) and a mean accuracy of 76.86% in the corresponding LSTM classification task, the best results for both machine and deep learning approaches for unimodal experiments.

Table 34: Comparison of unimodal studies used BioVid in accuracy

Study	Modalities	Machine Learning	BL vs PA4
Werner et al. [2]	ECG	Random Forest	62.0%
Werner et al. [2]	GSR	Random Forest	73.8%
Werner et al. [2]	EMG	Random Forest	57.9%
Lopez-Martinez et al. [54]	ECG	Logistic Regression	57.69%
Thiam et al. [46]	ECG	1D CNN	57.04%
Gkikas et al. [49]	ECG	Support Vector Machine	58.62%
Gkikas et al. [55]	ECG	MT-NN	69.40%
Subramaniam et al. [60]	EDA	CNN-LSTM	80.17%
Subramaniam et al. [60]	ECG	CNN-LSTM	75.21%
Ours	ECG	Support Vector Machine	58.39%
Ours	GSR	Support Vector Machine	75.60%
Ours	EMG	Support Vector Machine	55.37%
Ours	ECG	Long short-term memory	57.58%
Ours	GSR	Long short-term memory	76.86%
Ours	EMG	Long short-term memory	56.83%

As the deviation between categories increases, we observe that higher performances are achieved in binary experiments, as shown in Tables 4-32. For instance, the categories representing the minimum level of pain compared to the baseline (BL vs PA1) do not exhibit significant differences, thereby complicating the model's ability to differentiate between them. Conversely, when comparing pain at the highest level to the minimum level of pain or the baseline (PA1 vs P4 or BL vs PA4), exceptionally high performances are noted due to the distinctiveness of the categories.

The present thesis highlighted substantial differences in pain detection and tolerance thresholds, with young females demonstrating significantly lower thresholds compared to the male population of the same age. Additionally, an observation was made that the pain detection threshold decreases with advancing age, particularly among males. Conversely, the tolerance threshold declined with age in both males and females. The conducted experiments showed significant variability between genders, with males demonstrating reduced sensitivity, particularly in cases of higher pain intensities. Variations in pain perception between men and women involve anatomical, physiological, neural, hormonal, and psychological factors [84]. Analysis reveals that women tend to report pain more often and have a lower pain threshold compared to men. Hormonal fluctuations in women are linked to pain experiences, elucidating the connection between severe pain instances, like childbirth, and hormonal changes [84]. Women may undergo heightened pain from electric shock or thermal pressure, though the influence of social factors on these outcomes is subject to debate. In our study, we observed significantly high performances in the female population in both machine learning and LSTM models.

While BioVid has been extensively employed as an input source in numerous studies over the years, further investigation into the aspects of gender and age is warranted. As we mentioned, there are limited studies that have focused on demographic factors. The study [49] investigated both gender and age using machine learning techniques and achieved 60.69% and 25.11% for the female population experiment in binary and multiclass experiments, respectively, as well as 56.07% and 22.25% for binary and multiclass experiments in males using only the ECG signal. In their subsequent work [55], they proposed neural networks with single-task learning to further explore how demographics influence acute pain. They reported 69.48% and 30.59% in the female population and 66.48% and 28.61% for males for the same classification tasks. Subramaniam et al. [60] explored this hypothesis and achieved 81.71% accuracy for the female population and

76.79% for males using a CNN-LSTM model combining GSR and ECG signals. To the best of our knowledge, our study reported the highest machine learning performance (82.83%) in the binary task distinguishing baseline from the highest pain intensity, particularly among females aged 20-35.

In comparison to experiments conducted on both genders using machine learning approaches, we surpassed the study [49], achieving a mean accuracy of 79.88% accuracy for females and 70.85% for males in the LSTM model and 77.61% and 72.61% in SVM, respectively. The key difference from all other studies that also delved into exploring gender and age impacts was that we fully exploited all physiological signals (ECG, GSR, and EMG) from the dataset.

Regarding age, discernible variations were evident, revealing that as age advances, there is a reduction in pain perception, consequently elevating the susceptibility to further injury. In certain demographic groups, the observed difference exceeded 10%, underscoring the divergence in pain perception among individuals.

Presently, there is a lack of comprehensive understanding regarding gender and age dynamics in the perception of experimental pain. Findings from a study by Pickering et al. [85] indicated that young and elderly females exhibited lower pressure nociception thresholds compared to the male population. However, with the introduction of age as a variable, a discernible reduction in pressure detection and tolerance thresholds emerged, particularly among males. These observations are in concordance with the outcomes of the present study, wherein a diminution in pressure pain detection and tolerance thresholds was discerned in males (N=44). Nevertheless, a study by Lautenbacher et al. [86] failed to ascertain any interaction between sex and age. The authors attributed this lack of disparity to the utilization of a modest experimental sample.

6. Conclusions

To sum up, pain management and assessment have always been vital for patients and public health. While patient-reported outcomes are regarded as the "gold" standard for pain evaluation, they pose challenges due to their subjectivity and the uncertainty surrounding their reliability. In the interdisciplinary domain of pain research, the application of various AI techniques holds the potential for substantial benefits.

The objective of this master's thesis was to develop a framework for automated pain recogni-

tion and assessment through physiological-data-driven predictive models. This study emphasizes the substantial importance of integrating multimodal physiological signals and their most efficient combination. Also, we investigated if the utilization of multimodal signals impacts any observed performance enhancements compared to a single modality. Additionally, we aimed to scrutinize how demographic factors such as gender and age influence pain perception and how they can affect the results of our experiments. The BioVid Part A dataset was used as input in our pipeline, incorporating the demographic data with age groups divided into 20-35 years, 36-50 years, and 51-65 years. The Pan & Tompkins algorithm was applied for ECG feature extraction, while statistical analysis was employed for both GSR and EMG signals. Following data pre-processing, binary and multiclass tasks were conducted in both unimodal and multimodal experiments.

We employed machine learning models such as SVMs, DT, and a deep learning model of stacked Long Short-Term Memory (LSTM). The utilization of stacked LSTMs in automated pain assessment is driven by their proficiency in handling sequential information, hierarchical feature extraction, addressing long-term dependencies, and facilitating multimodal integration. Moreover, models demonstrated superior performance between baseline versus the highest pain intensity classification tasks, in both unimodal and multimodal experiments.

In conclusion, the outcomes are promising, but we recognize that further investigations need to resolve several challenges. However, we aspire that the results stemming from this current thesis will further contribute to research in pain estimation, and generally assist in extracting valuable and efficient information from demographic data.

6.1 Study Limitations

The findings of this study demonstrate significant potential but also reveal certain limitations that must be acknowledged for future research endeavors. The BioVid dataset, commonly used in pain research, provides valuable information such as the gender and age of each participant. However, it lacks data on external factors such as individual stress levels. Psychological states serve as confounding variables that can interfere with the perception of pain, particularly when anticipating a subsequent pain stimulus. Additionally, the relatively small number of subjects (N=87) raises questions about generalizability and objectivity across different populations.

Furthermore, our study solely focuses on physiological signals, excluding image and audio

modalities. Our findings indicate that EMG did not yield high-performance rates. Therefore, substituting or adding other physiological signals, such as EEG, may potentially enhance multimodal fusion and provide further insights into our research.

Finally, it is important to highlight that our study centers on acute thermal pain in a laboratory research setting. The exploration of chronic pain conditions (such as low back pain or in cancer patients) is lacking due to the availability of public datasets in this research area.

6.2 Further Research

In our future research endeavors, we aim to delve into the intricate influences on pain perception arising from diverse stimuli, such as those encountered in cold pain experiments. Furthermore, pain studies confront challenges necessitating in-depth exploration. One notable challenge involves a thorough investigation of factors that potentially impact the intensity of patients' pain. These factors encompass a broad spectrum, from emotional states and mental health conditions such as depression, and anxiety, to sleep patterns. Consequently, the development of a robust framework emerges as a necessity, capable of accurately linking physiological changes to their respective causal factors. This comprehensive approach aims to contribute substantially to a refined understanding of the complex facets of pain, paving the way for the development of more precise and effective tools for pain evaluation.

Challenges also emerge in applying pain assessment models to diverse demographic groups. Future efforts should prioritize enhancing the adaptability of these models to different populations and integrating algorithms to facilitate timely processing for monitoring pain in real time.

As a prospective extension of this work, the utilization of audio and image modalities, integrated with the existing physiological signals could be considered. Additionally, the development of an application for monitoring, tracking, and detecting pain using wearable devices could prove beneficial. Another avenue for expansion involves exploring advanced classification methodologies, particularly focusing on complex DL techniques and different types of fusion such as hybrid or late, to observe potential changes in our results.

Highlighting the importance of utilizing the framework in real-world contexts necessitates additional exploration and meticulous analysis. This includes engaging participants from a broader age spectrum, diverse ethnic backgrounds, and various pain conditions. By integrating such diversity,

the computational models can achieve increased flexibility and relevance, guaranteeing their effectiveness across different groups. This strategy aligns with our dedication to enhancing the strength and applicability of our pain assessment models.

Machine and deep learning techniques establish a strong foundation in automated pain assessment. This ongoing work holds the potential to revolutionize pain assessment, enhancing accuracy, adaptability, and clinical relevance.

References

[1] D. Lopez-Martinez and R. Picard, "Multi-task neural networks for personalized pain recognition from physiological signals," 2017 7th Int. Conf. Affect. Comput. Intell. Interact. Work. Demos, ACIIW 2017, vol. 2018-Janua, pp. 181–184, 2018, doi: 10.1109/ACIIW.2017.8272611.

[2] P. Werner, D. Lopez-Martinez, S. Walter, A. Al-Hamadi, S. Gruss, and R. W. Picard, "Automatic Recognition Methods Supporting Pain Assessment: A Survey," IEEE Trans. Affect. Comput., vol. 13, no. 1, pp. 530–552, 2022, doi: 10.1109/TAFFC.2019.2946774.

[3] W. Li, P. Liu, Y. Hu, and J. Meng, "Pain Modulates Responses to Emotional Stimuli," Front. Psychol., vol. 11, no. November, pp. 1–12, 2020, doi: 10.3389/fpsyg.2020.595987.

[4] H. Merskey et al., "Pain terms: a list with definitions and notes on usage," PAIN, vol. 6, no. 3, 1979.

[5] J. D. Loeser and R. Melzack, "Pain: An overview," Lancet, vol. 353, no. 9164, pp. 1607–1609, 1999, doi: 10.1016/S0140-6736(99)01311-2.

[6] A. Fayaz, P. Croft, R. M. Langford, L. J. Donaldson, and G. T. Jones, "Prevalence of chronic pain in the UK: A systematic review and meta-analysis of population studies," BMJ Open, vol. 6, no. 6, 2016, doi: 10.1136/bmjopen-2015-010364.

[7] Y. Takai, N. Yamamoto-Mitani, Y. Abe, and M. Suzuki, "Literature review of pain management for people with chronic pain," Japan J. Nurs. Sci., vol. 12, no. 3, pp. 167–183, 2015, doi: 10.1111/jjns.12065.

[8] B. Susam et al., "Automated Pain Assessment in Children Using Electrodermal Activity and Video Data Fusion via Machine Learning," IEEE Trans. Biomed. Eng., vol. 69, no. 1, pp. 422–431, 2022, doi: 10.1109/TBME.2021.3096137.

[9] M. Kächele, P. Thiam, M. Amirian, F. Schwenker, and G. Palm, "Methods for Person-Centered Continuous Pain Intensity Assessment from Bio-Physiological Channels," IEEE J. Sel. Top. Signal Process., vol. 10, no. 5, pp. 854–864, 2016, doi: 10.1109/JSTSP.2016.2535962.

[10] S. D. Subramaniam and B. Dass, "Automated Nociceptive Pain Assessment Using Physiological Signals and a Hybrid Deep Learning Network," IEEE Sens. J., vol. 21, no. 3, pp. 3335–3343, 2021, doi: 10.1109/JSEN.2020.3023656

[11] S. Iravanian and L. Tung, "A novel algorithm for cardiac biosignal filtering based on filtered residue method," IEEE Trans. Biomed. Eng., vol. 49, no. 11, pp. 1310–1317, 2002, doi: 10.1109/TBME.2002.804589.

[12] Y. Chu, X. Zhao, J. Han, and Y. Su, "Physiological signal-based method for measurement of pain intensity," Front. Neurosci., vol. 11, no. MAY, pp. 1–13, 2017, doi: 10.3389/fnins.2017.00279.

[13] I. Poláčková Šolcová and A. Lačev, "Differences in male and female subjective experience and physiological reactions to emotional stimuli," Int. J. Psychophysiol., vol. 117, no. March, pp. 75–82, 2017, doi: 10.1016/j.ijpsycho.2017.04.009.

[14] L. Shu et al., "A review of emotion recognition using physiological signals," Sensors (Switzerland), vol. 18, no. 7, 2018, doi: 10.3390/s18072074.

[15] D. J. Hohenschurz-Schmidt et al., "Linking Pain Sensation to the Autonomic Nervous System: The Role of the Anterior Cingulate and Periaqueductal Gray Resting-State Networks," Front. Neurosci., vol. 14, no. February, 2020, doi: 10.3389/fnins.2020.00147.

[16] K. Hamunen, V. Kontinen, E. Hakala, P. Talke, M. Paloheimo, and E. Kalso, "Effect of pain on autonomic nervous system indices derived from photoplethysmography in healthy volunteers," Br. J. Anaesth., vol. 108, no. 5, pp. 838–844, 2012, doi: 10.1093/bja/aes001.

[17] T. D. Yeater et al., "Chronic Pain is Associated With Reduced Sympathetic Nervous System Reactivity During Simple and Complex Walking Tasks: Potential Cerebral Mechanisms," Chronic Stress, vol. 5, no. Zip 32608, 2021, doi: 10.1177/24705470211030273.

[18] R. Fernandez Rojas, N. Brown, G. Waddington, and R. Goecke, "A systematic review of neurophysiological sensing for the assessment of acute pain," npj Digit. Med., vol. 6, no. 1, pp. 1–25, 2023, doi: 10.1038/s41746-023-00810-1.

[19] P. Reicherts, A. B. M. Gerdes, P. Pauli, and M. J. Wieser, "On the mutual effects of pain and emotion: Facial pain expressions enhance pain perception and vice versa are perceived as more arousing when feeling pain," Pain, vol. 154, no. 6, pp. 793–800, 2013, doi: 10.1016/j.pain.2013.02.012.

[20] T. Mokhtari, Y. Tu, and L. Hu, "Involvement of the hippocampus in chronic pain and depression," Brain Sci. Adv., vol. 5, no. 4, pp. 288–298, 2019, doi: 10.26599/bsa.2019.9050025.

[21] E. A. R. Losin, S. R. Anderson, and T. D. Wager, "Feelings of Clinician-Patient Similarity and Trust Influence Pain: Evidence From Simulated Clinical Interactions," J. Pain, vol. 18, no. 7, pp. 787–799, 2017, doi: 10.1016/j.jpain.2017.02.428.

[22] M. K. Johnson, S. Nolen-Hoeksema, K. J. Mitchell, and Y. Levin, "Medial cortex activity, self-reflection and depression," Soc. Cogn. Affect. Neurosci., vol. 4, no. 4, pp. 313–327, 2009, doi: 10.1093/scan/nsp022

[23] Godinho F., Frot M., Perchet C., Magnin M., Garcia-Larrea L. (2008). Pain influences hedonic assessment of visual inputs. Eur. J. Neurosci. 27 2219–2228. 10.1111/j.1460-9568.2008.06196.x

[24] A. Craig et al., "Developing a model of associations between chronic pain, depressive mood, chronic fatigue, and self-efficacy in people with spinal cord injury," J. Pain, vol. 14, no. 9, pp. 911–920, 2013, doi: 10.1016/j.jpain.2013.03.002.

[25] Gerdes A. B. M., Wieser M. J., Alpers G. W., Strack F., Pauli P. (2012). Why do you smile at me while I'm in pain? – Pain selectively modulates voluntary facial muscle responses to happy faces. In. J. Psychophysiol. 85 161–167. 10.1016/j.ijpsycho.2012.06.002

[26] W. Li, P. Liu, Y. Hu, and J. Meng, "Pain Modulates Responses to Emotional Stimuli," Front. Psychol., vol. 11, no. November, pp. 1–12, 2020, doi: 10.3389/fpsyg.2020.595987.

[27] J. L. Rhudy, Emotional Modulation of Pain. Elsevier Inc., 2016.

[28] S. Sheikh et al., "Comparing Pain Intensity Using the Numeric Rating Scale and Defense and Veterans Pain Rating Scale in Patients Revisiting the Emergency Department," Cureus, vol. 13, no. 8, 2021, doi: 10.7759/cureus.17501.

[29] S. Walter et al., "'What About Automated Pain Recognition for Routine Clinical Use?' A Survey of Physicians and Nursing Staff on Expectations, Requirements, and Acceptance," Front. Med., vol. 7, no. December, pp. 1–8, 2020, doi: 10.3389/fmed.2020.566278.

[30] F. Pouromran, S. Radhakrishnan, and S. Kamarthi, "Exploration of physiological sensors, features, and machine learning models for pain intensity estimation," PLoS One, vol. 16, no. 7 July, pp. 1–11, 2021, doi: 10.1371/journal.pone.0254108.

[31] P. Silva and R. Sebastião, "Using the Electrocardiogram for Pain Classification under Emotional Contexts," Sensors, vol. 23, no. 3, 2023, doi: 10.3390/s23031443.

[32] E. K. Naeini et al., "Pain recognition with electrocardiographic features in postoperative patients: Method validation study," J. Med. Internet Res., vol. 23, no. 5, pp. 1–13, 2021, doi: 10.2196/25079.

[33] M. Eriksson, H. Storm, A. Fremming, and J. Schollin, "Skin conductance compared to a combined behavioral and physiological pain measure in newborn infants," Acta Paediatr. Int. J. Paediatr., vol. 97, no. 1, pp. 27–30, 2008, doi: 10.1111/j.1651-2227.2007.00586.x.

[34] J. Munsters, L. Wallström, J. Ågren, T. Norsted, and R. Sindelar, "Skin conductance measurements as pain assessment in newborn infants born at 22-27weeks gestational age at different postnatal age," Early Hum. Dev., vol. 88, no. 1, pp. 21–26, 2012, doi: 10.1016/j.earlhumdev.2011.06.010.

[35] B. Rodríguez-Tapia, I. Soto, D. M. Marínez, and N. C. Arballo, "Myoelectric Interfaces and Related Applications: Current State of EMG Signal Processing-A Systematic Review," IEEE Access, vol. 8, pp. 7792–7805, 2020, doi: 10.1109/ACCESS.2019.2963881.

[36] S. Moscato, P. Cortelli, and L. Chiari, "Physiological responses to pain in cancer patients: a systematic review," Computer Methods and Programs in Biomedicine, vol. 217, Article ID 106682, 2022.

[37] J. Koenig, M. N. Jarczok, R. J. Ellis, T. K. Hillecke, and J. F. Thayer, "Heart rate variability and experimentally induced pain in healthy adults: A systematic review," Eur. J. Pain (United Kingdom), vol. 18, no. 3, pp. 301–314, 2014, doi: 10.1002/j.1532-2149.2013.00379.x.

[38] M. Cascella et al., "Artificial Intelligence for Automatic Pain Assessment: Research Methods and Perspectives," Pain Res. Manag., vol. 2023, pp. 1–13, 2023, doi: 10.1155/2023/6018736.

[39] B. Susam, M. Akcakaya, H. Nezamfar et al., "Automated pain assessment using electrodermal activity data and machine learning," in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 372–375, Honolulu, HI, USA, July 2018.

[40] S. Gruss et al., "Pain intensity recognition rates via biopotential feature patterns with support vector machines," PLoS One, vol. 10, no. 10, pp. 1–14, 2015, doi: 10.1371/journal.pone.0140330.

[41] S. Walter et al., "The biovid heat pain database: Data for the advancement and systematic validation of an automated pain recognition," 2013 IEEE Int. Conf. Cybern. CYBCONF 2013, no. January 2015, pp. 128–131, 2013, doi: 10.1109/CYBConf.2013.6617456.

[42] Gouizi K., Reguig F.B., Maaoui C. Analysis physiological signals for emotion recognition; Proceedings of the 2011 7th International Workshop on Systems, Signal Processing and their Applications (WOSSPA); Tipaza, Algeria. 9–11 May 2011; pp. 147–150.

[43] I. Campanini, C. Disselhorst-Klug, W. Z. Rymer, and R. Merletti, "Surface EMG in Clinical Assessment and Neurorehabilitation: Barriers Limiting Its Use," Front. Neurol., vol. 11, no. September, pp. 1–22, 2020, doi: 10.3389/fneur.2020.00934.

[44] C. T. Candotti et al., "Electromyography for assessment of pain in low back muscles," Phys. Ther., vol. 88, no. 9, pp. 1061–1067, 2008, doi: 10.2522/ptj.20070146.

[45] M. J. Page et al., "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," Int. J. Surg., vol. 88, pp. 1–10, 2021.

[46] P. Thiam, P. Bellmann, H. A. Kestler, and F. Schwenker, "Exploring deep physiological models for nociceptive pain recognition," Sensors (Switzerland), vol. 19, no. 20, pp. 1–21, 2019, doi: 10.3390/s19204503.

[47] Thiam P, Kestler HA, Schwenker F. Multimodal deep denoising convolutional autoencoders for pain intensity classification based on physiological signals. ICPRAM 2020—Proc 9th Int Conf Pattern Recognit Appl Methods. 2020; 289–296.

[48] D. Lopez-Martinez and R. Picard, "Continuous Pain Intensity Estimation from Autonomic Signals with Recurrent Neural Networks," Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf., vol. 2018, pp. 5624–5627, 2018, doi: 10.1109/EMBC.2018.8513575.

[49] S. Gkikas, C. Chatzaki, E. Pavlidou, F. Verigou, K. Kalkanis, and M. Tsiknakis, "Automatic Pain Intensity Estimation based on Electrocardiogram and Demographic Factors," Int. Conf. Inf. Commun. Technol. Ageing Well e-Health, ICT4AWE - Proc., no. January, pp. 155–162, 2022, doi: 10.5220/0010971700003188.

[50] S. A. H. Aqajari et al., "Pain assessment tool with electrodermal activity for postoperative patients: Method validation study," JMIR mHealth uHealth, vol. 9, no. 5, pp. 1–11, 2021, doi: 10.2196/25258.

[51] E. K. Naeini et al., "Pain recognition with electrocardiographic features in postoperative patients: Method validation study," J. Med. Internet Res., vol. 23, no. 5, pp. 1–13, 2021, doi: 10.2196/25079.

[52] Thiam, P.; Hihn, H.; Braun, D.A.; Kestler, H.A.; Schwenker, F. Multi-Modal Pain Intensity Assessment Based on Physiological Signals: A Deep Learning Perspective. Front. Physiol. 2021, 1, 720464.

[53] R. Wang, K. Xu, H. Feng, and W. Chen, "Hybrid RNN-ANN Based Deep Physiological Network for Pain Recognition," Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS, vol. 2020-July, pp. 5584–5587, 2020, doi: 10.1109/EMBC44109.2020.9175247.

[54] Lopez-Martinez, D. and Picard, R. (2018). Continuous Pain Intensity Estimation from Autonomic Signals with Recurrent Neural Networks. In the International Conference of the IEEE Engineering in Medicine and Biology Society., volume 2018, pages 5624–5627.

[55] S. Gkikas, C. Chatzaki, and M. Tsiknakis, "Multi-task Neural Networks for Pain Intensity Estimation using Electrocardiogram and Demographic Factors Multi-task Neural Networks for Pain Intensity Estimation using Electrocardiogram and Demographic Factors," no. July 2023, doi: 10.1007/978-3-031-37496-8.

[56] Y. Zhao et al., "How Much Does It Hurt: A Deep Learning Framework for Chronic Pain Score Assessment," IEEE Int. Conf. Data Min. Work. ICDMW, vol. 2020-November, pp. 651–660, 2020, doi: 10.1109/ICDMW51313.2020.00092.

[57] Y. Li, S. Ghosh, J. Joshi, and S. Oviatt, "LSTM-DNN based Approach for Pain Intensity and Protective Behaviour Prediction," Proc. - 2020 15th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2020, pp. 819–823, 2020, doi: 10.1109/FG47880.2020.00061.

[58] E. J. Bartley and R. B. Fillingim, "Sex differences in pain: A brief review of clinical and experimental findings," Br. J. Anaesth., vol. 111, no. 1, pp. 52–58, 2013, doi: 10.1093/bja/aet127.

[59] M. Toomey, "Gender differences in pain," AANA J., vol. 76, no. 5, pp. 355–359, 2008.

[60] Subramaniam SD, Dass B. Automated Nociceptive Pain Assessment Using Physiological Signals and a Hybrid Deep Learning Network. IEEE Sens J. 2021;21: 3335–3343.

[61] G. Misra, W. E. Wang, D. B. Archer, A. Roy, and S. A. Coombes, "Automated classification of pain perception using high-density electroencephalography data," Journal of Neurophysiology, vol. 117, no. 2, pp. 786–795, 2017.

[62] Nezam T., Boostani R., Abootalebi V., Rastegar K. A novel classification strategy to distinguish five levels of pain using the EEG signal features. IEEE Transactions on Affective Computing. 2021;12(1):131–140. doi: 10.1109/TAFFC.2018.2851236.

[63] M. Elsayed, K. S. Sim, and S. C. Tan, "A novel approach to objectively quantify the subjective perception of pain through electroencephalogram signal analysis," IEEE Access, vol. 8, pp. 199920–199930, 2020.

[64] Yu M, Sun Y, Zhu B, Zhu L, Lin Y, Tang X, et al. Diverse frequency band-based convolutional neural networks for tonic cold pain assessment using EEG. Neurocomputing. 2020;378: 270–282.

[65] G. Misra, W. E. Wang, D. B. Archer, A. Roy, and S. A. Coombes, "Automated classification of pain perception using high-density electroencephalography data," Journal of Neurophysiology, vol. 117, no. 2, pp. 786–795, 2017.

[66] J. Levitt, M. M. Edhi, R. V. Thorpe et al., "Pain phenotypes classified by machine learning using electroencephalography features," NeuroImage, vol. 223, Article ID 117256, 2020.

[67] A. Vuckovic, V. J. F. Gallardo, M. Jarjees, M. Fraser, and M. Purcell, "Prediction of central neuropathic pain in spinal cord injury based on EEG classifier," Clin. Neurophysiol., vol. 129, no. 8, pp. 1605–1617, 2018, doi: 10.1016/j.clinph.2018.04.750.

[68] M. Wei, Y. Liao, J. Liu et al., "EEG beta-band spectral entropy can predict the effect of drug treatment on pain in patients with herpes zoster," Journal of Clinical Neurophysiology, vol. 39, no. 2, pp. 166–173, 2022.

[69] C. Id and W. Published, "BP4D + Multimodal Spontaneous Emotion Database Case ID: Description:," pp. 2–4, 2021.

[70] G. Zamzmi, C. Y. Pai, D. Goldgof, R. Kasturi, T. Ashmeade, and Y. Sun, "A Comprehensive and Context-Sensitive Neonatal Pain Assessment Using Computer Vision," IEEE Trans. Affect. Comput., vol. 13, no. 1, pp. 28–45, 2022, doi: 10.1109/TAFFC.2019.2926710.

[71] J. O. Egede et al., "EMOPAIN Challenge 2020: Multimodal Pain Evaluation from Facial and Bodily Expressions," Proc. - 2020 15th IEEE Int. Conf. Autom. Face Gesture Recognition, FG 2020, pp. 849–856, 2020, doi: 10.1109/FG47880.2020.00078.

[72] E. Othman, P. Werner, F. Saxen, A. Al-Hamadi, S. Gruss, and S. Walter, "Automatic vs. Human recognition of pain intensity from facial expression on the x-ite pain database," Sensors, vol. 21, no. 9, pp. 1–19, 2021, doi: 10.3390/s21093273.

[73] M. Velana et al., "The senseemotion database: A multimodal database for the development and systematic validation of an automatic pain-and emotion-recognition system," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10183 LNAI, pp. 127–139, 2017, doi: 10.1007/978-3-319-59259-$6_1$1.

[74] E. Kasaeyan Naeini, M. Jiang, E. Syrjälä et al., "Prospective study evaluating a pain assessment tool in a postoperative environment: protocol for algorithm testing and enhancement," JMIR Res Protoc, vol. 9, no. 7, Article ID e17783, 2020.

[75] G. H. Chang, D. T. Felson, S. Qiu, A. Guermazi, T. D. Capellini, and V. B. Kolachalama, "Assessment of knee pain from MR imaging using a convolutional Siamese network," European Radiology, vol. 30, no. 6, pp. 3538–3548, 2020.

[76] M. S. H. Aung, S. Kaltwang, B. Romera-Paredes et al., "The automatic detection of chronic pain-related expression: requirements, challenges and the multimodal EmoPain dataset," IEEE Transactions on Affective Computing, vol. 7, no. 4, pp. 435–451, 2016.

[77] S. Moscato, L. Palmerini, P. Palumbo, and L. Chiari, "Quality assessment and morphological analysis of photoplethysmography in daily life," Front Digit Health, vol. 4, Article ID 912353, 2022.

[78] J. Lötsch, A. Ultsch, B. Mayer, and D. Kringel, "Artificial intelligence and machine learning in pain research: a data scientometric analysis," PAIN Reports, vol. 7, no. 6, Article ID e1044, 2022.

[79] M. Goirand, E. Austin, and R. Clay-Williams, "Implementing ethics in healthcare AI-based applications: a scoping review," Science and Engineering Ethics, vol. 27, no. 5, p. 61, 2021.

[80] Z. Zhang, "A gentle introduction to artificial neural networks," Ann. Transl. Med., vol. 4, no. 19, pp. 1–6, 2016, doi: 10.21037/atm.2016.06.20.

[81] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[82] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," IEEE Trans. Biomed. Eng., vol. BME-32, no. 3, pp. 230–236, 1985, doi: 10.1109/TBME.1985.325532.

[83] Phinyomark, A., Quaine, F., Charbonnier, S., Serviere, C., Tarpin-Bernard, F., and Laurillau, Y. (2014). Feature extraction of the first difference of emg time series for EMG pattern recognition. Comput. Methods Programs Biomed. 117, 247–256. doi: 10.1016/j.cmpb.2014.06.013

[84] W. Gutiérrez Lombana and S. E. Gutiérrez Vidal, “Pain and gender differences: A clinical approach,” Colomb. J. Anesthesiol., vol. 40, no. 3, pp. 207–212, 2012, doi: 10.1016/j.rcae.2012.05.006.

[85] Pickering,G., Jourdan,D., Eschalier,A., Dubray,C. (2002). Impact of age, gender and cognitive functioning on pain perception. Gerontology 48, 112-8

[86] Lautenbacher,S., Kunz,M., Strate,P., Nielsen,J., Arendt-Nielsen,L. (2005). Age effects on pain thresholds, temporal summation and spatial summationof heat and pressure pain. Pain 115, 410-418.

[87] G. M. Dimitri, “A Short Survey on Deep Learning for Multimodal Integration: Applications, Future Perspectives, and Challenges,” Computers, vol. 11, no. 11, 2022, doi: 10.3390/computers11110163.

Appendices

Table 35: Multiclass Multimodal Hold-Out Experiments

			Support Vector Machine Kernels		
Group	Classes	Validation	Linear	Gaussian	Polynomial
All Subject	BL vs Pain Levels	20%-80%	**30.05%**	29.25%	26.60%
All Subject	BL vs Pain Levels	40%-60%	30.89%	**31.12%**	26.46%
All Subject	Pain Levels	20%-80%	32.47%	**34.05%**	33.97%
All Subject	Pain Levels	40%-60%	31.93%	33.04%	**34.23%**

Table 36: Multiclass Unimodal ECG Hold-Out Experiments

			Support Vector Machine Kernels		
Group	Classes	Validation	Linear	Gaussian	Polynomial
All Subject	BL vs Pain Levels	20%-80%	21.14%	21.37%	**21.78%**
All Subject	BL vs Pain Levels	40%-60%	**22.67%**	22.44%	21.86%
All Subject	Pain Levels	20%-80%	**28.08%**	26.86%	26.43%
All Subject	Pain Levels	40%-60%	27.55%	27.22%	**27.87%**

Table 37: Multiclass Unimodal GSR Hold-Out Experiments

			Support Vector Machine Kernels		
Group	Classes	Validation	Linear	Gaussian	Polynomial
All Subject	BL vs Pain Levels	20%-80%	26.72%	26.83%	**27.47%**
All Subject	BL vs Pain Levels	40%-60%	25.40%	24.85%	**25.97%**
All Subject	Pain Levels	20%-80%	30.60%	30.81%	**30.96%**
All Subject	Pain Levels	40%-60%	30.20%	30.38%	**31.82%**

Table 38: Multimodal Binary Hold-Out Experiments

		Support Vector Machine Kernels		
Classes	Validation	Linear	Gaussian	Polynomial
BL vs P1	20%-80%	49.56%	**50.28%**	50.14%
BL vs P1	40%-60%	50.93%	50.86%	**51.43%**
BL vs P2	20%-80	55.60%	57.18%	**58.76%**
BL vs P2	40%-60%	56.17%	56.03%	**59.48%**
BL vs P3	20%-80%	61.06%	61.35%	**65.08%**
BL vs P3	40%-60%	60.91%	61.06%	**64.58%**
BL vs P4	20%-80%	**77.01%**	75.71%	74.56%
BL vs P4	40%-60%	**78.66%**	77.72%	76%
P1 vs P2	20%-80%	53.30%	53.87%	**55.60%**
P1 vs P2	40%-60%	54.66%	54.45%	**54.81%**
P1 vs P3	20%-80%	57.61%	60.91%	**63.21%**
P1 vs P3	40%-60%	58.26%	59.41%	**62.57%**
P1 vs P4	20%-80%	68.96%	71.40%	**73.41%**
P1 vs P4	40%-60%	68.10%	71.33%	**73.99%**
P2 vs P3	20%-80%	55.02%	54.45%	**56.46%**
P2 vs P3	40%-60%	52.29%	52.58%	**54.88%**
P2 vs P4	20%-80%	63.21%	64.51%	**66.95%**
P2 vs P4	40%-60%	63.50%	65.94%	**67.31%**
P3 vs P4	20%-80%	57.75%	57.75%	**58.04%**
P3 vs P4	40%-60%	58.76%	58.04%	**59.41%**

Table 39: Multiclass Unimodal EMG Hold-Out Experiments

			Support Vector Machine Kernels		
Group	Classes	Validation	Linear	Gaussian	Polynomial
All Subject	BL vs Pain Levels	20%-80%	21.14%	**21.89%**	20.05%
All Subject	BL vs Pain Levels	40%-60%	19.94%	**20.63%**	20.31%
All Subject	Pain Levels	20%-80%	24.78%	**25%**	23.99%
All Subject	Pain Levels	40%-60%	26.36%	**27.08%**	25.50%