Universidade de Aveiro
Ano 2021

Daniela Filipa da Silva Pais

Alterações Fisiológicas em Eletrocardiogramas durante o Período Pós-operatório

Physiological Alterations in Electrocardiogram during the Postoperative Period

Universidade de Aveiro
Ano 2021

Daniela Filipa da Silva Pais

Alterações Fisiológicas em Eletrocardiogramas durante o Período Pós-operatório

Physiological Alterations in Electrocardiogram during the Postoperative Period

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Biomédica, realizada sob a orientação científica da Doutora Ana Raquel Ferreira de Almeida Sebastião, Investigadora do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro e da Doutora Susana Manuela Martinho dos Santos Baía Brás, Investigadora do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro.

Dedico este trabalho à minha família e aos meus amigos.

o júri

presidente

Prof. Doutor Luiz Fernando Ribeiro Pereira
Professor Auxiliar da Universidade de Aveiro

Prof. Doutora Rita Paula Almeida Ribeiro
Professora Auxiliar da Faculdade de Ciências - Universidade do Porto

Doutora Ana Raquel Ferreira de Almeida Sebastião
Investigadora Auxiliar da Universidade de Aveiro

agradecimentos

Expresso os meus sinceros agradecimentos a todos os que tornaram possível o desenvolvimento desta dissertação.

Às minhas orientadoras Doutora Raquel Sebastião e Doutora Susana Brás, pela confiança que depositaram em mim, pelo tempo despendido e pela orientação e acompanhamento incansável ao longo deste percurso.

À Ana, pelo companheirismo e amizade ao longo deste ano.

À Joana e Sara, pela cumplicidade e todo o apoio.

A todos os amigos que fiz durante o meu percurso académico, por me terem acompanhado nos últimos cinco anos. Obrigada por me fazerem sentir em casa.

Um agradecimento especial ao Miguel, Lília, Regina e Flávia, com quem partilhei experiências inesquecíveis.

Por último, aos meus pais e irmã, pelo amor e apoio incondicional.

palavras-chave

Avaliação Fisiológica da Dor, Dor Pós-operatória, Eletrocardiograma, Extração de Características, Métodos de Agrupamento, Seleção de Características

resumo

Neste trabalho, é estudada a informação de eletrocardiogramas (ECG) adquiridos de 19 doentes durante o período pós-operatório. Após procedimentos cirúrgicos, os doentes experienciam frequentemente dor pós-operatória, que deve ser controlada de forma efetiva.

A avaliação da dor é um passo essencial para garantir a escolha adequada do procedimento analgésico, e, atualmente, é baseada numa avaliação feita pelo próprio doente. No entanto, estes métodos de avaliação são subjetivos, descontínuos e inadequados para avaliar a dor de doentes que têm dificuldade ou não conseguem comunicar verbalmente. Neste contexto, o desenvolvimento de um método objetivo e contínuo para avaliar e monitorizar a dor pós-operatória, e que não necessita que o doente auto-relate a sua dor, pode ajudar a ultrapassar estas limitações. Um método mais objetivo é clinicamente relevante, uma vez que pode melhorar o controlo da dor do doente no serviço de recobro e promover uma melhor recuperação.

Nos últimos anos, a avaliação da dor através de indicadores fisiológicos tem sido estudada, uma vez que consistem numa medida não invasiva da atividade do sistema nervoso. Este estudo teve como objetivo encontrar relações entre alterações fisiológicas e mecanismos da dor e descobrir informação relevante extraída dos sinais ECG para estudar os processos da dor, tendo sido realizada uma análise univariada e uma análise multivariada. Os métodos implementados neste trabalho podem ser divididos em pré-processamento, extração e seleção de caraterísticas, e métodos de agrupamento. Também foram implementados testes estatísticos para estudar a normalidade e correlação entre variáveis e as diferenças significativas entre grupos. Assim, foi investigada que caraterística ou combinação de caraterísticas descreve melhor a resposta da dor pós-operatória e qual a técnica de agrupamento que apresenta melhores resultados. A abordagem multivariada provou ser mais promissora que a abordagem univariada, permitindo a identificação e distinção de diferentes tipos de caraterísticas da dor baseadas na informação extraída dos sinais ECG.

keywords

Clustering, Electrocardiogram, Feature Extraction, Feature Selection, Physiological Pain Assessment, Postoperative Pain

abstract

In this work, the information of electrocardiogram (ECG) signals collected from 19 patients during the postoperative period is studied. After surgical procedures, patients often experience postoperative pain, which should be managed effectively.

The assessment of pain is a crucial step to guarantee a suitable analgesic control of postoperative pain, and currently, it is based on the self-reports of the patients. However, these assessment methods are subjective, discontinuous, and inadequate for evaluating the pain of subjects unable or with limited capacity to communicate verbally. In this context, developing an objective and continuous tool for assessing and monitoring postoperative pain that does not require patient reports can overcome these limitations. Such a tool is of great clinical relevance because it can improve pain control during the patient stay in the post-surgery care unit and ultimately promote better recovery.

In the last years, the evaluation of pain through physiological indicators has been studied since they represent non-invasive measures of the nervous system activity. This study aimed to find relationships between physiological alterations and pain mechanisms and discover relevant information extracted from ECG signals to study pain processes. Both univariate and multivariate analyses were carried out. The methods implemented in this work can be divided into pre-processing, feature extraction and selection, and clustering analysis. Statistical tests were also employed to assess the normality and correlation of the features and significative differences between groups. It was investigated which feature or combination of features better describe postoperative pain response and which clustering technique has better performance.

The multivariate approach proved to be more promising than the univariate approach, allowing different types of pain characteristics to be identified and distinguished based on the extracted features from the ECG signals.

Contents

List of Figures

List of Tables

List of Acronyms

ANI	Analgesia/Nociception Index
BMI	Body Mass Index
CHTV	Centro Hospitalar Tondela Viseu
DGS	Direção Geral da Saúde
IASP	International Association for the Study of Pain
PD	Pupil Diameter
PLRA	Pupillary Light Reflex Amplitude
SNS	Sympathetic Nervous System

Chapter 1
Introduction

In this first chapter, an introduction to the subject of the dissertation is made. Firstly, it is introduced the context in which this dissertation is inserted. Then, the importance and motivation that led to the realization of this work are presented, and the corresponding objectives are established. Lastly, a brief description of the document outline is provided.

1.1 Background

After surgery, patients commonly suffer from postoperative pain, which should be controlled to prevent complications [1]. The acute pain that arises after surgery can resort to chronic postoperative pain. Chronic pain can persist months after the procedure, which is reflected in a longer medical care period and higher medical expenses. The transition to chronic pain can be due to preoperative risk factors, such as age and gender, and intraoperative risk factors, including surgery duration and the possible damaging of nerves during surgical manipulation. Lastly, it can appear because of postoperative factors, for example, undertreatment or overtreatment of pain [1], [2]. In addition to this, pain and its inadequate management can lead to psychological and physiological adverse effects [3]. Undertreatment may contribute to tachycardia and increased cardiac metabolic demand, whereas overtreatment may lead to respiratory complications, delirium, oversedation, urinary retention, ileus, and postoperative nausea and vomiting [4]. Inadequate treatment is also related to deep vein thrombosis, pulmonary embolism, coronary ischemia, myocardial infarction, pneumonia, poor wound healing, insomnia, demoralization, postoperative stress, and opioid addiction [5], [6]. It is also believed that children and neonates exposed to untreated pain can experience long-term consequences, such as altered pain sensitivity (e.g., hyperalgesia) and maladaptive pain (e.g., chronic pain) [3], [7]. Therefore, postoperative pain should be alleviated rapidly and managed effectively to decrease patient discomfort, promote better recovery, and, ultimately, reduce postoperative morbidity and mortality [1], [4], [5].

Pain is a personal and subjective experience that depends on biological, psychological, and social factors. Even when two individuals receive similar sensory stimuli (e.g., same surgical procedure), they can have different sensory experiences [8], [9]. Therefore, a suitable choice of analgesic techniques for pain control is dependent on an accurate assessment [10], [11]. As patient needs and characteristics, such as age or cognitive ability, can differ, various pain assessment tools suitable for the different patient groups are required (an extensive description of such tools is provided in subchapter *3.1 Current Pain Assessment Tools*). The tools can be primarily divided into the ability of the patient to self-report pain. When he is capable, the standard postoperative assessment methods are based on self-reports through scoring systems. One example is the Numeric Rating Scale (NRS) that requires the patients to rate their pain on a scale from 0 to 10, where 0 is no pain, and 10 is the worst pain imaginable [12]. Faces Pain Scale-Revised (FPS-R) is another widely chosen self-report measure that uses facial expressions instead of numbers to assess pain intensity, mostly used with children [7]. However, when patients have limited ability or are unable to communicate verbally or understand the scales, the assessment cannot rely on self-reports [12], [13]. In these cases, some behaviors exhibited when in pain can be observed to assess pain in the patient, for example, indicators in facial expression and body movements [7], [12], [14]. In addition, physiological parameters such as heart rate (HR), blood pressure (BP), respiratory rate, and

oxygen saturation can also be measured along with patient observation [7], [12]. A wide variety of scales that combine different behavior and physiological parameters exist [7].

1.2 Motivation

The current pain assessment tools have limitations that complicate the assessment task and may lead to a poor assessment of pain. The following limitations were found:

- **The methods are not continuous and provide short-term measures.** The available tools are intermittent, resulting, most of the time, in large time intervals between assessments [15];
- **Pain evaluation tasks can be too complex and time-consuming.** For example, tools based on an observer are multidimensional. These tools describe several physiological and behavioral parameters increasing the complexity of the scale and the time required to assess pain [7], [12];
- **The subjective experience of pain can be challenging to evaluate from a third-person perspective.** Due to individual differences in pain expression, the intensity assessed may not reflect the actual pain being experienced by the patient [9], [16];
- **Behavior and physiological changes are not always specific to pain.** Physiological changes may be due to other causes, such as illness, whereas behavioral responses, such as crying in neonates, might occur due to stressors other than pain [3], [17].

The problems identified on the available methods promoted the description of a set of characteristics that a pain assessment tool should comprise. Firstly, in the postoperative period, pain intensity should be monitored and recorded continuously to optimize analgesic control and pain relief in clinical settings [15], [18]. In addition, measurements should be easy and practical to acquire without requiring a heavy workload for health care professionals. Moreover, ideally, the assessment should be objective and not involve a subjective perspective from an observer evaluating the pain [19]. Ultimately, it is expected that such a tool could result in better care and improve the patient's quality of life. Therefore, besides the physiological description of pain, the work developed in this dissertation had as main motivation the search for physiological characteristics that could be potentially used as a reliable and feasible instrument to assess postoperative pain in clinical settings.

1.3 Objectives

In the context of the motivation described above, this dissertation aims to study the physiological description of postoperative pain based on electrocardiogram (ECG) signals to better comprehend pain in humans through physiological changes. The main goals of this work are:

- Analyze each extracted feature individually as a postoperative pain indicator;
- Observe the influence of pain on the features of the ECG signal accordingly with self-reported pain;
- Investigate if pain-relief medication posed a signature on the ECG features;
- Perform a multivariate study to explore if more than one feature improves the description of postoperative pain;
- Evaluate which combination of features better distinguishes pain parameters, i.e., intensity and persistence.

1.4 Document Outline

This dissertation consists of 5 chapters, organized as follows:

Chapter 2 – "Postoperative Pain" firstly explains the sensory mechanisms of the nervous system that compose postoperative pain. Then, the components that contribute to the subjectivity of pain are detailed and, after, pharmacological interventions associated with surgery are mentioned.

Chapter 3 - "Pain Assessment" provides a summary of current methods for measuring pain intensity. After, it reviews related work in the area of objective assessment of postoperative pain using different physiological signals, while introducing the signals.

Chapter 4 - "Physiological Assessment of Postoperative Pain" describes the data collection protocol as well as all the work done regarding the implementation of the methods to study postoperative pain with ECG signals. It also details and discusses the results that were obtained.

Chapter 5 - "Conclusions and Further Research" comprises a summary of the results. In addition, conclusions regarding the study of postoperative pain are established, and the associated limitations and proposals for future work are discussed as well.

Chapter 2
Postoperative Pain

In order to have a better understanding of the pain phenomenon, several topics that compose pain experience will be explored in this chapter. Firstly, the concept and descriptive variables of pain are introduced. After, the physiological mechanisms of nociception and inflammation that compose the sensory dimension of postoperative pain are described. Then, the components that modulate the inter-individual differences in the experience of pain are explored. Finally, the influence of the anesthesia given before the surgery, a pharmacological intervention specific to postoperative pain, is considered.

2.1 Pain

The International Association for the Study of Pain (IASP) defines pain as "An unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage." Pain may vary in intensity (mild, moderate, and severe) and duration (transient, intermittent, or persistent), and it can be divided into four types: nociceptive, inflammatory, neuropathic, and functional [22]. In addition, two concepts of pain perception can be differentiated, pain threshold and pain tolerance. Pain threshold is linked to pain sensitivity and corresponds to the intensity at which a person perceives a stimulus as a painful one. On the other hand, pain tolerance is the maximum level of pain a person can endure. When evaluating acute postoperative pain in the clinical setting, the analysis of pain tolerance may be more relevant [4], [30].

The sensation of pain results from mechanisms and pathways that involve the central nervous system, which consists of the brain and the spinal cord, and the peripheral nervous system, which is composed of nerves [20], [23], [24]. However, pain is not only related to these pathways, as it is strongly affected by inter-individual variability, which is modulated by multiple contributions [8], [9].

2.2 Postoperative Pain Pathways and Mechanisms

Postoperative pain corresponds to a pain of a long duration that arises from tissue injury and trauma from recent surgical intervention [6], [15]. Pain that surges from surgery combines various processes of the nervous system. Not only is it due to nociceptive mechanisms, but it is also related to inflammatory processes [1].

2.2.1 Nociceptive pain

IASP describes nociceptive pain as "pain that arises from actual or threatened damage to non-neural tissue and is due to the activation of nociceptors" [8]. Nociceptive pain occurs when the tissue suffers trauma from noxious stimuli, e.g., mechanical force (surgical incision), thermal or chemical stimulation [21]. The sensory experience initiates when nociceptors, peripheral neurons that act as pain receptors, detect the external stimulus. When the noxious stimuli exceed a threshold, they activate the receptors and transduction occurs, and the painful external stimulus is converted into electrical signals. The electrical information is then conducted from the peripheral nervous system to the central nervous system. The electrical activity travels, through unmyelinated (C-fiber) or thinly myelinated (Aδ- fiber)

axons of the nociceptors, from the peripheral terminal to the central terminal of the neurons, reaching the spinal cord. The nociceptors, first-order neurons, end in the dorsal horn of the spinal cord and synapse with second-order neurons that transmit the signals via ascending pathways to the brain, more specifically to the thalamus. From there, third-order neurons, also known as relay neurons, transfer the information to the somatosensory cerebral cortex, where the stimulus is perceived as an intense and unpleasant sensation known as pain [20], [22]–[25].

2.2.2 Inflammatory pain

Inflammation occurs after the tissue is injured to promote healing and repair the tissue trauma and eliminate the necrotic cells. During the inflammation process, inflammatory chemical factors are released at the site of the injury or inflammation. The chemical factors are detected by the peripheral terminal of the nociceptors of the necrotic tissue [26]. These chemical agents can activate the nociceptors originating pain or can alter the sensitivity of the terminals without producing direct nociceptor activation. An increase in sensitivity is due to a reduction of the activation threshold of the nociceptors' terminals, making tissues more sensitive and responsive to contact or movement. This hypersensitivity, known as hyperalgesia, is an increased and exaggerated response to a stimulus that normally causes pain. In addition, the inflammatory process may lead to sensitization of tissues, known as allodynia, in which a stimulus that usually is not painful provokes pain [22], [26].

2.3 Pain Subjectivity

As described previously, the sensory component of postoperative pain relates to inflammation and noxious stimulation of the nociceptors. However, pain perception is affected by several factors, such as genetics, age, emotional state, personality, and cognitive components, that contribute to inter-individual differences. The experience of pain is, therefore, a complex and subjective event. Due to these differences, pain can be challenging to generalize and diagnose, and, consequently, it can be difficult to achieve optimal pain relief [9], [22]. The study of the contribution of each factor could provide a better understanding of the differences between individuals. Pain sensitivity can be studied by applying different painful stimuli, such as pressure, cold, heat, or chemicals [9], [27]. Even though the experimental models developed to study sensitivity do not represent clinical pain, some studies prove the clinical relevance of evaluating sensitivity via cold or heat stimulus and reveal a prospective linkage to postoperative pain [28].

Variations related to genetic factors can be explored through twin studies, i.e., studies conducted on monozygotic (identical) or dizygotic (fraternal) twins since they allow to distinguish the differences in pain perception that result from genetics and the variability that result from the environment [4], [9], [27]. Nielsen et al. performed a twin study concluding that the contribution of genetic and environmental factors to variance in pain sensitivity depends on the pain modalities. Results revealed 60% and 26% of variation due to genetics for pain induced by cold-pressor and heat, respectively, showing that heritability is not equal across modalities [27]. Angst et al. also concluded that 49% of the variance of pain tolerance under a cold pressor stimulus could be assigned to heritability [28]. In addition, Norbury et al. analyzed heritability in women in two separated populations and discovered 45% to 53% of variation due to genetics after heat-induced pain [29].

Furthermore, gender and ethnicity, which are also connected to genetics, can contribute to pain differences. However, studies regarding these two parameters should not rely only on the genetics

context since they are strongly influenced by social components [4], [27]. One study reported that pain sensitivity to heat and cold-pressor is superior in women when compared to men [28]. Moreover, studies that specifically investigate postoperative pain show that there is still conflict in the findings relative to gender differences. Some studies argue that women report higher pain, while others say that men do. In contrast, others show that there is no difference in the reports between genders. However, the type of surgery and the study design varied between studies, which could cause inconsistent conclusions [4].

Age can also influence pain sensation. In the study conducted by Angst et al., older individuals had decreased sensitivity to heat and cold pressor pain [28]. Another study indicates that older age is related to less sensitivity to heat-induced pain than the group of individuals with middle age, which is concordant with the first study's results [30]. Psychological and physiological alterations related to age, such as alteration of the peripheral nerves' density and the increase of degenerated sensory fibers, are associated with advanced age, which helps support the results of the studies [31]. However, the revision of several studies suggests a consensus still does not exist on the variation of pain sensitivity during the life of an individual [30].

Moreover, redirection of attention from pain to a cognitive process during a task can reduce pain sensation. In addition, individual experience related to previous knowledge of a specific stimulus and expectation of pain intensity can influence nociceptive processing [9]. Chang et al. collected information concerning patients' emotions about surgery, revealing that most patients considered postoperative pain fearful and would cause tiring-exhaustion [32]. Besides, a personality profile that exhibits specific psychological characteristics, such as anxiety and catastrophizing, can also contribute to variability in the pain experience. In a clinical scenario, anxiety can aggravate pain, and depression is connected to more frequent and severe pain [9], [28].

2.4 Anesthesia and Analgesia

Beyond the contributors to differences in pain experience specified above, postoperative pain also integrates pharmacological interventions. Before the surgical procedure, anesthesia is given to the patient, blocking pain signals during surgery, while after, analgesia control of pain can be performed.

Balanced anesthesia combines three types of drugs, a hypnotic, an analgesic, and a muscle relaxant (as shown in Figure 2.1), and induces a reversible state of unconsciousness, antinociception, and paralysis during the surgery [33], [34].

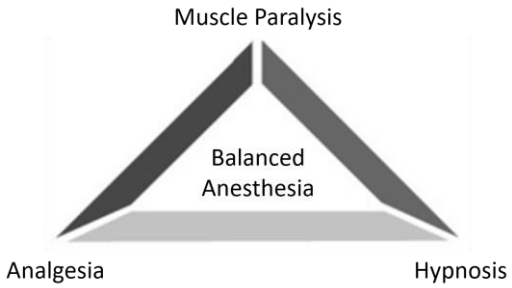

Figure 2.1: Anesthetic state results from the balance of three drugs: analgesics, hypnotics, and muscle relaxants. Adapted from [35].

In the postoperative period, the effects of general anesthesia may extend longer in some patients as a result of changes in pharmacokinetics, for example, the removal of anesthesia from the body can be different. The pharmacokinetics of drugs can be approximated by mathematical models [36], and the three-compartment model commonly describes the pharmacokinetics of intravenous anesthetic [34]. The three-compartment model approximates the body as a central compartment (plasma) and two peripheral compartments, one which equilibrates rapidly (muscle) and another which equilibrates slowly (fat) [34]. The pharmacokinetic properties of drugs may change due to extremely high or low body weight, which means that their metabolism and elimination may change [37]. In practice, clinicians usually classify obesity and malnutrition using BMI (Body Mass Index), defined as the ratio between weight (kg) and the height squared (m^2) [37], and obesity is defined as BMI > 30 $kg \cdot m^{-2}$ [38]. Obese individuals have a higher mass of poorly vascularised adipose tissue which is associated with prolonging effect and delaying terminal elimination of the drug [37]. Therefore, the removal of anesthesia from the body can take longer in patients with more fat tissue.

As postoperative management, adequate analgesia must be given to the patient when needed. Analgesics act as an antinociceptive to reduce the transmission of stimuli that cause pain through the sensory system. Postoperative analgesia may be obtained with a variety of techniques and therapeutics. It can be provided with opioids (e.g., morphine) and antipyretic analgesics (e.g., paracetamol), for example [33].

Chapter 3
Pain Assessment

This chapter begins to address the most commonly used pain assessment tools in clinical settings. The exposed methods are primarily divided into the patient's ability to self-report their pain and secondly into the patient's age. Afterward, deriving from the identified challenges of the currently used assessment methods, an overview of the literature on the objective assessment of postoperative pain is given. A set of studies exploring physiological signals for pain evaluation after a surgical procedure is inspected. Firstly, the physiological signals recorded for this purpose are defined, and after, the methods implemented in each study are detailed. Lastly, the implementation of ECG signals for pain evaluation is discussed.

3.1 Current Pain Assessment Tools

Assessing pain intensity consists of quantifying a painful sensation through clinically validated, reliable and safe tools that are appropriate to the type of pain and patient characteristics (e.g., age, cognitive and clinical condition, verbal communication capabilities) in order to optimize pain management [3], [39].

3.1.1 Self-report pain assessment tools

Currently, standard postoperative pain assessment methods are based on patients' self-reports [7]. *Direção Geral da Saúde* (DGS) establishes the prioritization of self-assessment from the age of 3, whenever possible, as a guideline [39]. When it comes to children with age superior to 3 and inferior to 8, it is usually chosen the Faces Pain Scale (FPS), and the patient is asked to classify the intensity of pain based on facial expressions (see FPS-R in Figure 3.1) [3], [7]. Wong-Baker Faces Pain Scale is another facial expression scale available [3]. For ages superior to 8 years old, numerical scales can be chosen since it is expected the understanding of numerical order and value. A common method is the NRS, which is considered a good choice for assessing acute pain after a surgery, where the patient rates their pain on a scale from 0-10 (Figure 3.1) [3], [40]. The Visual Analogue Scale (VAS) is also commonly applied. The VAS requires the patient to mark on a 10 cm horizontal line the point that he feels that represents his pain (Figure 3.1) [18]. Assessment questionnaires are also recurrent systems [7], [14]. For example, in the case of the Short-form McGill Pain Questionnaire (SF-MPQ), the patient has to score, on a scale from 0 to 3, eleven sensory (e.g., throbbing, shooting, and stabbing) and four affective (e.g., sickening and fearful) descriptors [40]. The Verbal Rating Scale (VRS) is another method that is based on the use of adjectives and descriptive phrases. For simplicity of language, each description is converted to a number, for example, 0 - "no pain," 1 - "mild pain," 2 - "moderate pain," and 3 - "severe pain" [41].

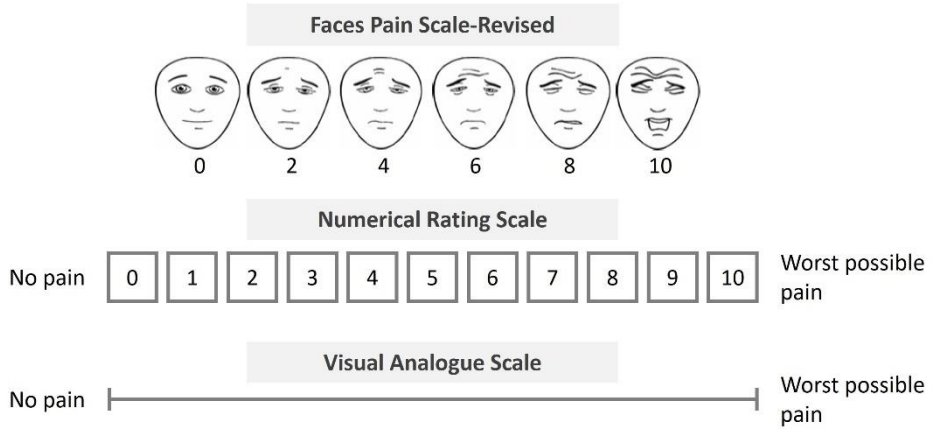

Figure 3.1: Self-report tools for pain assessment: FPS-R, NRS, and VAS. Adapted from [18], [19].

3.1.2 Pain assessment tools in patients with communication problems

Patients in the extreme of age (children and elderly), cognitive-impaired, unconscious, in the intensive care setting, or with a language barrier, have limited ability or are unable to self-report their pain [7], [14], [40]. Children may struggle with understanding and communicating their pain verbally due to their emotional and cognitive development [3]. In addition, neonates are nonverbal, so they depend on others to recognize and assess their pain [7]. Furthermore, cognitive impairment can also hazard the capacity of patients to self-report. For instance, delirium, a neurocognitive disorder common in aged and palliative populations, affects attention, memory, perception, and speech [14]. Additionally, critically ill patients in the intensive care unit can be sedated and, consequently, unable to self-report. Due to anesthesia, patients may also be unconscious during the earlier postoperative period [42], [43]. In these situations, a substitute for the verbal report must be used. Patients experiencing pain often display a variety of behaviors that indicate a painful experience. In addition, physiological markers can be measured [14], [40]. Pain may be associated with higher HR and BP, lower oxygen saturation, and more rapid, shallow, or irregular breathing patterns [17].

Neonatal pain assessment is currently based on behavioral and physiological indicators, since facial expressions consist of an indicator of acute and short-term pain, and arm and leg activity are considered typical body movements during the experience of pain. For instance, increased activity or higher muscle tension (clenched fists and toes and tense arms or legs) are characteristic in the presence of pain. In addition to these, crying sounds can also be informative to evaluate pain in neonates. Consolability is also a parameter often analyzed by nurses when assessing pain in neonates [7], [17]. Many neonatal pain assessment tools are available, including the N-PASS (Neonatal Pain, Agitation and Sedation Scale) [39]. The N-PASS is indicated for gestational age between 0-100 days, and it considers behavioral components (crying, irritability, behavior state, facial expression, extremities of tone) and vital signals (HR, BP, respiratory rate, and oxygen saturation) [7], [11]. In older children, other scales can be selected. The COMFORT Pain Scale is used for infants and small children under the age of 3, being an appropriate scale to assess postoperative pain or pain in unconscious and ventilated infants. This instrument tests nine parameters divided into physiologic components (BP, HR, and respiratory distress) and behavioral components (alertness, calmness, crying, physical movement, muscle tone, and facial tension) [7], [40]. Moreover, for children under the age of 4 or unable to verbally communicate, the FLACC (Face, Legs, Activity, Cry, and Consolability), which considers five behavior categories, can be implemented. This tool

is validated for scoring postoperative pain in infants and children [39]. On the other hand, in cognitively impaired elderly individuals, such as patients with dementia, facial expressions, verbalizations, vocalizations, body movements, changes in interpersonal interaction, and physical changes can be used as behavioral indicators of pain [14]. The Abbey Pain Scale is a common scale, composed of both behavioral and physiological (e.g., temperature) components, for this group of patients [44].

To conclude, a large variety of pain assessment tools has been validated. Nevertheless, pain assessment from a third person can be a very challenging task in an already subjective process [9], [15], [16]. These tools vary widely from each other, not only in the combination of physiological and behavioral parameters that they evaluate but also in the metric range. For example, COMFORT Pain Scale varies in the range of 9-45, and FLACC varies in the range of 0-10, which can create confusion to understand whether a score of the scale corresponds to low, moderate, or severe pain. Thus, developing one standard measure to assess pain in all patients could facilitate communication and understanding of the pain scores [7], [19], [39]. Furthermore, behavior changes and physiological measures are not always specific to pain and may be due to other causes. Pain evaluation with physiological signals identifies changes from a normal baseline state, assuming that all individuals have the same baseline. However, ill individuals can have baselines values distinct from healthy ones, resulting in physiological indicators less specific for pain. Also, pain behaviors are not noticeable if neuromuscular blocking agents or sedatives are administrated to the patient [3], [17]. In addition to the problems detailed here, others were identified in the first chapter of this document. The assessment provides non-continuous and often short-term measures and can be complex and time-consuming when the tools are multidimensional.

3.2 State of the Art of Objective Postoperative Pain Assessment

In recent years, objective and non-invasive measurement of pain has been a topic under investigation due to its clinical relevance. The use of physiological signals as objective markers of nociception and pain is one of the strategies proposed in the literature as they could allow monitoring changes in the nervous system induced by pain [45]. Several physiological indicators have been investigated, yielding good results in the postoperative period. While given a special focus on ECG, as it is the signal addressed in this dissertation, the studies discussed in this section include ECG, electrodermal activity (EDA), surface electromyography (sEMG), pupil reflexes, and accelerometry signals.

When the nervous system is stimulated by pain, stress, fear, or trauma, the body adapts to it, and physiological alterations occur. Unconscious and involuntary processes may happen, which are related to the Autonomic Nervous System (ANS). The ANS regulates physiologic processes including HR, BP, pupil diameter (PD), sweat glands, respiration, and body temperature and is composed of the sympathetic and parasympathetic divisions, which can have an excitatory or inhibitory effect. A stress-related stimulus (pain) will activate the sympathetic division, whereas the parasympathetic tone will be more prominent if the body is at rest. The balance between sympathetic and parasympathetic tones varies according to the body's needs and creates the autonomic tone. Only the sympathetic division innervates the sweat glands leading to sweat secretion when activated. The alterations in sweat glands' secretion can be measured with EDA. In addition, the sympathetic division increases HR and dilates the pupil, whereas the parasympathetic division decreases HR and constricts the pupil. These alterations can be studied through the ECG and pupil reflexes, respectively. By measuring these physiological alterations, the autonomic response to noxious stimuli can be detected non-invasively. Also, muscle

contractions and movements, generated by signals carried to the skeletal muscles, can be provoked by pain experience and measured with sEMG and accelerometry signals [23].

This section is divided into the different signals used to assess pain, and for each signal, the studies that were found are discussed. The studies focus mainly on physiological indicators. However, video and audio signals are also presented as tools to assess pain. Table 3.1 summarizes the features extracted from the collected signals, the number and characteristics of the subjects enrolled in the studies, and the main conclusions of the authors.

3.2.1 ECG

An ECG detects the electrical activity of the heart recorded with electrodes placed on the body surface. The ECG signal represents a sequence of cardiac cycles, and it allows to see the P wave, the QRS complex, and the T wave, which compose one cardiac cycle (Figure 3.2). Each wave is correlated with electrical events (myocardial action potentials). The P wave indicates the atrium depolarization (contraction), the QRS complex represents ventricular depolarization (contraction), and the T wave is associated with ventricular repolarization (relaxation) [23], [46]. An import measure extracted from the ECG signal is the time elapsed between two successive R peaks, representing the length of a ventricular cardiac cycle, and it is referred to as the R-R interval (RR) [46]. These intervals allow determining the HR and the variability in RR intervals, known as heart rate variability (HRV). In the articles found in the literature, postoperative pain was studied with HR and HRV measures in patients after a surgical procedure.

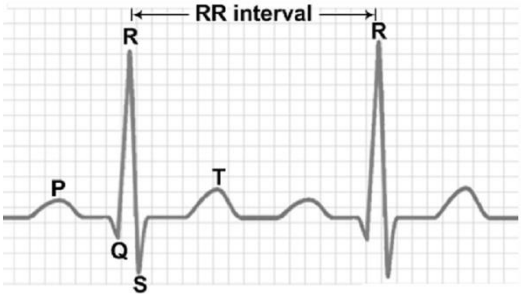

Figure 3.2: ECG signal [47].

Firstly, two studies investigating HR in the postoperative period were found. In one study, HR showed no correlation with the NRS [13]. Contrary, another study found a statistically significant association between HR and the NRS ($r=0.19$; $p=0.02$). However, HR did not alter after pain control and had low performance in discriminating patients with NRS ≤ 4 from patients with NRS > 4 [48].

Secondly, a study was conducted by Chang et al., who investigated the correlation between postoperative pain intensity, assessed using VAS and SF-MPQ, and HRV indices, measured for 5 minutes one day after abdominal surgery. The HRV indices under investigation included the standard deviation of RR intervals (SDNN), in the time domain, and very low frequency (VLF), low frequency (LF), high frequency (HF), and LF/HF ratio, in the frequency domain. Firstly, it was possible to see that the HRV parameters measured the next day after surgery were higher than the mean values measured and established for healthy individuals. Furthermore, LF/HF ratio correlated positively with the assessment tools, whereas the other HRV indices correlated negatively. An important remark withdrawn from this study is that different variables can influence HRV measures and, therefore, need to be pondered in future postoperative pain studies. For instance, the authors investigated the influence of the patient's

position in which the HRV parameters were measured, concluding that patients in the semi-fowler position had higher values than those in the supine position. In addition, it was discovered that gender and age could affect HRV [32].

Additionally, the use of the Analgesia Nociception Index (ANI) was reported as a measure of immediate postoperative pain in awake patients in different age groups. Although, this index was also implemented to monitor and measure pain in unconscious patients. ANI corresponds to an analysis of the RR series variability that considers the influence of respiratory patterns on the HF component of the RR intervals [49], [50]. It consists of a non-invasive and continuous measure of the sympathetic/parasympathetic balance that ranges from 0 (maximal nociception) to 100 (maximal analgesia). Higher ANI values indicate an increase of parasympathetic tone, which occurs in the case of adequate analgesia. However, when nociception increases because of the surgical procedure, or when analgesia fades, sympathetic tone increases, parasympathetic tone decreases, and ANI values decrease [50]–[52]. Firstly, Boselli and colleagues studied ANI as an immediate postoperative pain intensity index in adults who self-reported pain using NRS [52], [53]. The authors found a statistically significant negative relationship between ANI and NRS: $ANI = 68.1 - 4.2 \times NRS$, $R^2 = 0.33$; $p < 0.01$. In addition, a threshold of 50 allowed ANI to discriminate NRS > 3 with a sensitivity of 86%, a specificity of 86%, and an Area Under the Receiver Operating Characteristic Curve (AUC) of 0.89 [52]. Later, Gall et al. investigated the use of ANI in children whose postoperative pain was assessed with FLACC. The study included a control group and a surgical group to compare ANI values and tested the behavior of instantaneous ANI and ANI averaged over a 256 second period. Both measures were significantly lower in the surgical group than in the control group ($p < 0.001$). It was also observed how ANI measurements behaved to predict moderate-to-severe pain (FLACC ≥ 4), with a cut-off value of 66 to the instantaneous ANI and of 56 to the 256 second period ANI. Instantaneous ANI had more measurements of the control group below the cut-off value than ANI calculated over 256 seconds, so instantaneous ANI performs worse in predicting moderate-to-severe intensities. The threshold of 56 offered a sensitivity of 83%, a specificity of 93%, and an AUC of 0.94 to predict moderate-to-severe pain. Finally, ANI measured over the 256 seconds resulted in the following negative linear relationship with FLACC: $ANI = 76.1 - 4.3 \times FLACC$, $R^2 = 0.54$; $p < 0.01$ [50]. Some limitations on the use of ANI were found. Firstly, studies in adults revealed that the anesthetic agent and technique could alter ANI response [52], [53]. Secondly, some drugs can induce changes in HRV that are not specific to nociception, and, therefore, patients who had received them were excluded from the studies [50], [52], [53]. Therefore, although HRV may help healthcare providers to improve pain management, several aspects that need further study may compromise its use.

3.2.2 EDA

EDA, commonly referred to as skin conductance (SC), corresponds to the variation of the electrical conductance of the skin in response to sweat secretion. When sympathetic activation occurs by emotional alterations and stress associated with pain, changes in the SC signal that arise from stimuli may occur [54]–[56]. If sweat glands are activated, they produce sweat that increases SC, and when the sweat is reabsorbed, SC decreases, producing a peak in the signal. The SC signal was explored in the postoperative period as the mean SC and the number of fluctuations within the mean SC per second (NFSC). Studies in different age groups were collected.

Ledowski and colleagues performed an observational pilot study that aimed to investigate the influence of immediate postoperative pain on SC in adults and its correlation with pain rates obtained with NRS after arrival in the recovery room [13]. The promising results of this pilot study led to

conducting a prospective study to test the initial conclusions with a bigger sample of participants [10]. In relation to the first study, two conclusions can be highlighted. Firstly, the mean SC was considered inappropriate as a possible tool for pain evaluation. Secondly, NFSC showed a significant correlation with the NRS rating of patients ($r=0.625$; $p<0.01$), allowing to differentiate moderate and severe pain (NRS > 3) with sensitivity of 89% and specificity of 74%, using a cut-off value for NFSC of 0.1 [13]. In addition, NFSC was significantly different among the no, mild, moderate, and severe pain categories in both populations ($p<0.0001$) [10], [13]. The prospective study also demonstrated that NFSC is significantly lower after administration of analgesic for pain management than before ($p<0.013$). Finally, the study with the bigger sample resulted in a similar performance of NRS in distinguishing no and mild pain from moderate to severe pain, with the cut-off value for NRS of 0.1, obtaining a sensitivity of 88.5%, specificity of 67.7%, and AUC of 75.5% [10]. After, Hullet et al. studied NFSC as a postoperative pain assessment tool in children. Different scales were implemented accordingly with age groups: FLACC, FPS-R, and VAS, for 1-3 years, 4-7 years, and 8-16 years, respectively. It was found that a cut-off value for NFSC of 0.13 can distinguish moderate and severe pain, with sensitivity of 90%, specificity of 64%, and AUC of 0.82, for all age groups. It was also verified that NFSC could be important when used as a negative predictor (predict no or mild pain) for the assessment in pediatric patients because only 3.2% (negative predictive value=96.8%) of the patients with moderate and severe pain were defined incorrectly as no/mild pain. Oppositely, 64.5% (positive predictive value=35.5%) of the patients were incorrectly defined as moderate/severe pain [57]. It can be concluded that NFSC is a tool with potential for pain assessment in the postoperative period. However, since SC is an index of the sympathetic nervous system, it can be influenced by other common sympathetic activators in the postoperative period, such as stress, anxiety, nausea, vomiting and drugs [7], [13], [57], [58]. For instance, it was demonstrated that neostigmine and glycopyrrolate, reversal drugs used for reverse muscle relaxant effects, significantly influence SC measures [59]. For these reasons, these authors argue on the feasibility of relying solely on the SC to assess pain.

3.2.3 sEMG and accelerometry signals

sEMG is a non-invasive technique that measures the electrical activity generated by skeletal muscles through electrodes attached to the skin over the muscle of interest [60], [61], while accelerometry signals consist of the measurement of body movement [15]. As noted above, an increase in body movements and muscle contraction during periods of pain may occur compared to periods of no pain. A study was discovered in which the authors observed which alterations in EMG and accelerometry signals of infants occurred in the postoperative period.

Schasfoort et al. proposed the study of body part activity and muscle activity as behavioral indicators of immediate postoperative pain, assessed using the COMFORT-behavior scale, in preverbal infants. Body part activity was obtained by three piezo-resistive acceleration sensors, two attached to both legs (uni-axial sensors) and one attached to one arm (bi-axial sensor). The extracted motility values resulted in the overall extremity acceleration. On the other hand, the EMG sensor was attached to the wrist flexor muscle. EMG envelopes were extracted from the signals and normalized with the baseline of each infant. Firstly, the authors found a statistically significant correlation between accelerometry-based overall extremity activity and COMFORT-behaviour scale ($r=0.76$; $p<0.001$) and between EMG-based wrist flexor activity and COMFORT-behaviour scale ($r=0.55$; $p<0.001$). Secondly, overall extremity activity showed better results in discriminating pain (COMFORT behavior score ≥ 17), with sensitivity of 45% and specificity of 97%, than wrist flexor muscle activity, with sensitivity of 36% and specificity of 96%. However, the combination of accelerometry and EMG measures produced the best results

corresponding to a sensitivity of 64% and a specificity of 96%. Therefore, EMG signals could be valuable to support the accelerometry measures in discriminating between pain and no pain. However, although specificity values are very high, sensitivity is low, and, therefore, these indicators need to be further studied. In addition, infants medicated with sedative drugs or muscle relaxants needed to be excluded. Nevertheless, the results were satisfactory and sufficient to prove that measuring pain using accelerometry and muscle activity in preverbal infants is feasible and has the potential to discriminate pain [15].

3.2.4 Other studies

Other studies used pupil reflexes to study postoperative pain since pupil dilatation might be a sympathetic response to noxious stimuli. Pupil reflexes can be evaluated as a measure of PD, pupillary light reflex amplitude (PLRA), and pupillary dilatation reflex (PDR). PDR is the difference between pupillary diameter before and after painful stimulation, whereas PLRA is the diameter variance after light stimulation [48], [62]. Aissou et al. found a significant correlation between VRS and PDR. In addition, PDR proved to discriminate VRS ≥ 1 with sensitivity of 91% and specificity of 94%, using a threshold of 23% of variation [63]. Contrarily, Kantor et al. did not find an association between acute postoperative pain and PD or PLRA. Authors argue that the lack of association with postoperative pain may be due to the nature of this type of pain because it is continuous. Moreover, the mean intensity of postoperative pain was small (NRS=4.7±3.1) and may not be enough to evoke an alteration in pupil diameter. Pupil reflexes may also be affected by the remaining effects of anesthesia, which may weaken pain experience, or by other drugs administered to the patient, for instance, anticholinergic agents or vasoactive drugs. In conclusion, PD and PLRA did not revealed to be adequate indicators of acute postoperative pain. Furthermore, PDR requires a painful stimulation which is not an optimal strategy [48].

Not only can physiological signals be used for this purpose, as video and audio signals were also investigated in a study conducted by Salekin et al. to assess postoperative pain in neonates, whose pain was evaluated with the N-PASS scale. The video signal allows an analysis of behavioral indicators of pain, including facial expression and body movement, and the audio signal aims to record crying sounds. The methodology comprehended a unimodal approach in which the performance for facial expression, body movement, and crying sounds was examined individually, and a multimodal approach using all three measures. The crying sound classifier produced the highest performance metrics within unimodal approaches compared to the classifiers using body or face. In comparison to the crying sound, the multimodal approach has a similar performance. Nonetheless, the authors concluded that a multimodal approach is more reliable for assessing pain in the clinical context because multiple conditions can affect the measurements of the signals. For example, the face and body may not be visible due to hospital equipment, oxygen masks, or tapes, baby movements may be constricted, or babies may be under sedation. Furthermore, as pain experience varies between individuals, some babies may move around while others may cry when feeling pain [11].

Table 3.1: Objective markers for the assessment of postoperative pain. (legend: y.o., - years old).

Features	Subjects	Results/ Key findings	Reference
HR	25 adults (21-67 y.o.; 14 female)	HR showed no correlation with NRS	[13] (2006)
HR	145 adults (50 ± 17 y.o.; 75 female)	HR showed a significant correlation with NRS (r=0.19; p=0.02) HR did not change after administration of analgesia HR had low performance in discriminating NRS > 4 (AUC=0.62)	[48] (2014)
VLF, LF, HF, LF/HF, and SDNN	34 adults (18-79 y.o.; 22 female)	The patient's age, gender, and position during signal acquisition could influence HRV values LF/HF ratio correlated positively with pain, whereas SDNN VLF, LF, and HF correlated negatively	[32] (2012)
ANI	200 adults (44 ± 18 y,o., 51 ± 17 y.o.; 88 female)	ANI has an inverse linear relationship with NRS scores An ANI threshold of 50 allowed to discriminate moderate to severe pain (sensitivity=86%, specificity=86%, AUC=0.89)	[52] (2013)
ANI	Surgical group: 32 children (0.4-16 y.o.) Control group: 31 children (0.8-15 y.o.)	ANI has an inverse linear relationship with FLACC scores An ANI threshold of 56 allowed to discriminate moderate to severe pain (sensitivity=83%, specificity=93%, AUC=0.94) using ANI values average over 256 seconds	[50] (2015)
NFSC and Mean SC	25 adults (21-67 y.o.; 14 female)	NFSC showed a significant correlation with NRS (r=0.625; p<0.01) NFSC was significantly different between no, mild, moderate and severe pain reports (p<0.0001) A NFSC threshold of 0.1 differentiated moderate to severe pain (sensitivity=89%, specificity=74%)	[13] (2006)
NFSC	73 adults (18-81 y.o.; 30 female)	NFSC was significantly different between no, mild, moderate and severe pain reports (p<0.001) NFSC was significantly lower after administration of analgesia than before (p<0.013) A NFSC threshold of 0.1 differentiated moderate to severe pain (sensitivity=88.5%, specificity=67.7%, AUC=75.5%)	[10] (2007)
NFSC	165 children (1-3 y.o., 4-7 y.o., 8-16 y.o.)	A NFSC threshold of 0.13 distinguished moderate to severe pain (sensitivity=90%, specificity=64%, AUC=0.82)	[57] (2009)

Mobility features (overall extremity acceleration) and EMG signal envelope of the wrist flexor muscles	Accelerometer: 14 infants (45-400 days) EMG: 7 infants (45–379 days)	Accelerometry-based overall extremity activity and EMG-based wrist flexor activity showed a significant correlation with COMFORT-behavior scale (r=0.76; p<0.001 and r=0.55; p<0.001, respectively) The accelerometery-EMG based indicator showed better results in discriminating pain than accelerometery or EMG individually (sensitivity=64%, specificity=96%)	[15] (2007)
PDR	100 adults (52-62 y.o.; 58 female)	PDR showed a statistically significant correlation with VRS (r=0.88; p<0.0001) A threshold of 23% of variation allowed PDR to discriminate VRS ≥ 1 (sensitivity=91%, specificity=94%)	[63] (2012)
PD and PLRA	145 adults (50 ± 17 y.o.; 75 female)	No statistically significant association was found between NRS and PD (r=0.10; p=0.54) or PLRA (r=0.03; p=0.72)	[48] (2014)
Visual features from face and body images and spectrogram of crying sounds	45 neonates (30-41 weeks)	Multimodal approach is better and more reliable than a unimodal approach in neonates Multimodal approach had good performance to predict postoperative pain (accuracy=0.7936, precision=0.8028, recall=0.7936, f1-score=0.7920, AUC=0.9010)	[11] (2020)

Concluding, several strategies produced important developments and findings in the field of objective indicators of postoperative pain. Some measures exhibited significant correlation with the stimulus and proved their potential to detect pain presence, not being able, however, to discern its severity. Furthermore, some authors debate the use of more than one signal as a more reliable method. Essentially, although the results of the studies are encouraging, the clinical viability of these tools still needs to be further investigated and validated for clinical use.

Since this dissertation will explore ECG signals as an indicator for postoperative pain, a brief discussion will be made on the feasibility of its use. Firstly, the use of some drugs can alter the function of the ANS by enhancing or suppressing sympathetic and parasympathetic activity, and consequently, influence the physiological measurements, limiting their use in some patient groups [23]. The authors of the studies discussed above report this limitation in using HRV and other measures regulated by ANS. Drugs may also induce changes in the morphology, amplitude, and duration of the ECG curve. Moreover, HR can vary in response to hemodynamic changes influenced by circulatory changes, illness, and cardioactive and vasoactive drugs. Hence, in hemodynamically unstable patients, their use may not be reliable to evaluate pain [64]. The studies also showed that the age and gender of the patient and the anesthetic agent and technique might affect HRV measurement [48], [50], [51]. In addition, stress and anxiety, commonly experienced in the postoperative period, are activators of sympathetic activity and, therefore, may influence the physiological response [52], [57]. Concluding, several modulators of the nervous system may induce physiological changes that are not specific to pain. Nonetheless, the studies evaluating HRV proved that HRV metrics could accomplish binary discrimination between no/mild pain and moderate/severe pain with good performance and, therefore, demonstrated the usefulness of using ECG measures as a surrogate for postoperative pain.

Chapter 4
Physiological Assessment of Postoperative Pain

In this chapter, a study is conducted on physiological changes of ECG signals associated with patients' pain experience during the postoperative period. Firstly, the protocol and setup implemented for data collection are presented. After, the dataset used for the work developed in this dissertation is described. Following, the methods of the study performed are detailed, accompanied by a discussion of the results.

4.1 Study Contextualization

As discussed earlier, the body's reaction to pain could be detected through physiological and behavioral measurements. The study of physiological alterations during the postoperative period after a surgical procedure can help understand the relation between postoperative pain and physiological response. Therefore, this chapter aims to study the individual behavior of several extracted features from the ECG signal, the relationship between features, and to evaluate which combination of features better describes the pain experience. Figure 4.1 represents the methods that were implemented in this study. The data was collected postoperatively in the recovery room. The raw collected signals were then pre-processed, and 5-minutes and 10-minutes windows of information were extracted regarding pain assessment and management annotations. The features calculated in each time interval were firstly explored individually. Afterward, a multivariate analysis exploring feature combinations was conducted, after performing feature selection to discard the non-relevant features.

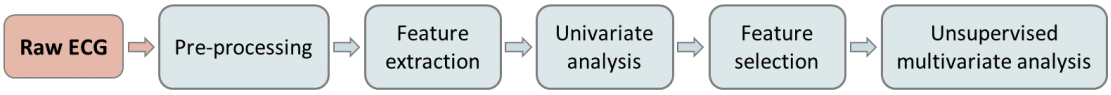

Figure 4.1: Methodology for the study of physiological alterations of ECG signal during postoperative pain experience.

4.2 Protocol and Setup for Data Collection

Due to the ongoing COVID-19 pandemic, the acquisition of physiological signals from patients in clinical settings was not viable. Thus, in this work, a set of previously collected data containing ECG signals was used, which was collected during July and August of 2019. This study was a collaboration between the Institute of Electronics and Informatics Engineering of Aveiro (IEETA) from the University of Aveiro (UA) and the Centro Hospitalar Tondela-Viseu, E.P.E. (CHTV). The study was approved, and it is registered by the Ethics Committee for Health of the CHTV (document number 894) and by the Ethics and Deontology Council of the University of Aveiro (document number 36/2018).

The participants were recruited from volunteer patients undergoing elective abdominal surgery at Hospital de São Teotónio - CHTV. The informed consent provided to the volunteers is available online[1].

The exclusion criteria included the following:

- Age inferior to 18 years and superior to 85 years;
- Cardiac disorders, major systemic diseases, congenital syndromes, pacemakers, and psychological/psychiatric/mental disorders;
- Comorbid diagnosis of chronic pain (e.g., Rheumatoid arthritis);
- Inability to comprehend or answer the self-report measures;
- Refusal to provide informed consent.

The protocol implemented in this study followed several steps:

1. After the patients' selection, an interview was held to explain to the patient the procedures during the data collection and the different sensors he would wear on his body. They were also informed that they could quit the data collection at any time;
2. At the entrance in the recovery room, after being stabilized by the clinical team, electrodes were placed on the body of the patient, and the positioning was maintained through the recording. The ECG signal was collected during the standard clinical practices of analgesia (which were performed according to the clinicians' decisions/demands) and without compromising the patient's wellbeing;
3. A waiting time of 5 minutes was considered to stabilize the signal, while signal recordings were visually inspected;
4. While in the recovery room, patients were asked to describe their pain level at different time points during the recording. The patients' pain evaluations were based on self-report (several assessments, as necessary, until discharge);
5. The clinical interventions were registered, and every time an analgesic therapeutic was applied/changed, a trigger was manually deployed or annotated;
6. If necessary, according to the analgesic therapeutics, and if the patients are in conditions approved by the clinical team, the patients' pain was evaluated based on self-report;
7. With respect to the clinical conditions after surgery, a data collection instrument[1] was applied and fulfilled by the clinician team, addressing multiple variables, such as age, gender, diagnosis, health issues, surgical intervention and related procedures, type of pain, location of pain, level of pain, and pain relief intervention/therapeutics (both in surgery and recovery rooms);
8. As the clinical team considered that the patient was stable for discharge, the data collection ended, and the electrodes were removed.

4.3 Dataset

The dataset used in this work is composed of ECG signals, patient's age and gender, and type of surgical intervention. In addition, the procedures performed during the patient stay in the recovery room were also collected, including self-reports of pain, pain relief therapeutics, and other medical

[1]https://github.com/danielapais/Mestrado_Dissertacao

interventions, such as patient repositioning. These procedures are associated with time triggers that mark the event occurrence in the ECG signal.

ECG recordings with a sampling frequency of 500 Hz from twenty patients were recorded, one recording for each patient. However, one patient was withdrawn from the study because of the lack of pain assessment annotations during the ECG recording. Therefore, the patients consisted of ten females and nine males, with ages ranging from 23 to 83 years. The pain scale consisted of a verbal description of pain through adjectives and phrases by the patients. The self-reports from patients were converted and grouped into four categories: 0 – "no pain", 1 – "mild pain", 2 – "moderate pain", and 3 – "severe pain". A variable number of pain assessment points was obtained for each patient during ECG recording, and a total of 32 pain readings were acquired, with 12 assessments in patients with no pain, 4 in those with mild pain, 11 with moderate pain, and 5 in those with severe pain, as shown in Table 4.1. Category "no pain" included 4 readings of three female individuals and 8 of four male individuals. The mild intensity contains 3 readings from three different females and 1 from a male patient. "Moderate" readings correspond to 8 reports from six females and 3 reports from three males. Lastly, the severe category comprehends 2 readings from two male patients and 3 readings from a female patient. Among all patients, four (two females and two males) did not receive any pain analgesia during their stay in the recovery room since three of them did not experience pain at any time and one only reported mild pain at entering the recovery room. In addition, any analgesic was given after a time point of pain assessment (of mild pain) to another different patient, as this patient was already being medicated for pain management (Table 4.1).

Table 4.1: Number of postoperative pain assessments for each pain category.

Pain	Total pain assessments	Number pain assessments followed by pain control
0 – no pain	12	0
1 – mild	4	2
2 – moderate	11	11
3 – severe	5	5
Total assessments for all categories	32	18

An analysis of the number of mild, moderate, and severe reports of each patient was also done. Table 4.2 represents how many patients reported a non-zero level of pain several times during their stay in the recovery room, both before and during ECG recording, which varied from 0 to 6 reports. In general, patients who did not experience pain or that only report pain one time (number of reports ≤ 1) are younger males (3 males, 23 ≤ age ≤ 58 y.o.) and older females (5 females, 64 ≤ age ≤ 83 y.o.). On the other hand, male patients who reported a non-zero level of pain more than one time and fewer than four times (1 < number of reports ≤ 3) are older (5 males, 59 ≤ age ≤ 83 y.o.) than the ones with a number of reports inferior or equal to one, and the two females had ages 34 and 51 y.o. The three patients who reported pain more than three times (number of reports > 3) were women aged 37, 69, and 82. In general, the seven patients who described more than three times a non-zero level of pain showed persistent pain during the ECG recording, despite being medicated with analgesic therapeutics.

Table 4.2: Number of patients and the respective number of non-zero levels of pain.

Number of reports	0	1	2	3	4	5	6
Number of patients	3	6	3	4	1	1	1

Furthermore, different surgical procedures were performed, which may lead to superior or inferior pain experience according to the type of surgical manipulation. For instance, the two patients who underwent gastric bypass complained of pain several times, whereas patients subjected to bilateral sympathectomy reported a less painful experience in the recovery room. The fact that patients underwent different procedures creates a wider variability in pain experiences. However, the reduced number of patients makes it difficult to draw conclusions on the pain intensity variation within types of surgery. In addition, since the available dataset comprehends a reduced representation of genders and age groups, it is not possible to observe relevant patterns of pain intensity regarding these two variables.

4.4 Pre-processing

ECG signals were pre-processed by filtering and normalization processes. However, firstly, the last samples of the ECG signals, which do not have useful information due to periods with excessive noise, were eliminated.

Filtering the noise of the signals is an essential step of pre-processing, as it allows to increase the quality of the signal. Noise can have a variety of origins, including physiological, equipment, or environmental interference. Common noises observed in ECG signals are powerline interference, which corresponds to meddling at a frequency of 50 Hz or 60 Hz [65], [66], and baseline wander (or baseline drift), whose content usually has frequencies below 1 Hz [46], [67]. To identify the noise in the signals and select the filtering technique, the frequency components of the ECG signals were analyzed using Welch's power spectral density estimate. This analysis allowed to identify content in the low frequencies and, in some spectrums, content at 50 Hz and respective harmonics and, therefore, bandpass filtering was considered suitable for removing the noise content found in the signals. This filter removes low-frequency noise and high-frequency noise content, according to the cut-off frequencies used. This filtering technique is widely used, but different cut-off frequencies are suggested [68]. In order to choose the filter parameters, the performance of different combinations of filter orders and cut-off frequencies were tested. The noise was then removed by filtering the ECG signals with a bandpass Butterworth filter, of order 4, with cut-off frequencies of 0.5 Hz and 40 Hz. An example of the filtering process is shown in Figure 4.2.

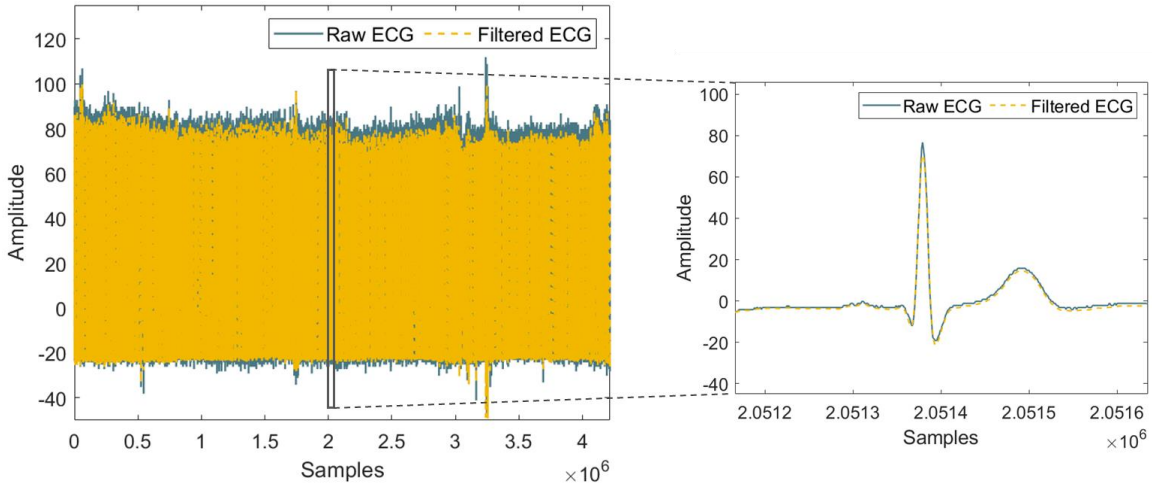

Figure 4.2: Raw and filtered ECG signal.

After filtering the signals, their amplitude was normalized in relation to the individual patient baseline to reduce inter-individual differences. The normalized signal is given by Equation 4.1, where mean_baseline and stdev_baseline are the mean and the standard deviation of the baseline signal, respectively.

$$Normalized\ Signal = \frac{Signal - mean_{baseline}}{stdev_{baseline}} \tag{4.1}$$

Inter-subject variability corresponds to variations in the heartbeats among different subjects that can be a result of heart geometry (heart size and shape), individual attributes (age and weight), physical exercise, cardiac condition, emotions and fatigue, and electrodes characteristics and placement [69]. An ECG signal has the same waves for all individuals, but its geometrical characteristics, such as the amplitude of the waves, can vary between subjects [70]. Since the baseline was not collected previously to the surgery, the baseline state of each patient was assumed to be the last 10 minutes of the ECG recording (after the removal of the noisy samples mentioned above). It is expected that the patient is no longer under the influence of any pharmacologic intervention and no longer in pain during this period. The quality of the baseline segment selected from the filtered ECG was visually checked for artifacts to confirm that the information was viable.

4.5 Feature Extraction

Postoperative pain is a long-lasting event that results from the recent damage of tissue. Therefore, selecting shorter time intervals from the physiological signal is necessary to study this phenomenon and its relationship with the ECG signal. These intervals of interest correspond to pain assessment and management moments, as they are associated with the intensity perceived and reported by the patient at that moment and with the relief of pain. Hence, postoperative pain intensity and experience can be studied in terms of physiological changes in those intervals. Two intervals were considered to investigate the influence of pain intensity and if pain-relief medication altered the features of the ECG signal. Therefore, after pre-processing the data, 5-minutes (150 000 samples) and 10-minutes (300 000 samples) of ECG data were selected before assessment for all self-reports (no, mild, moderate, and

severe pain). Regarding the second proposed goal, 5-minutes and 10-minutes of ECG data, both before and after pain control, were selected for self-reports of mild, moderate, and severe pain. Before proceeding to feature extraction on these time segments, they were examined visually to inspect the presence of artifacts to prevent inaccurate measurement induced by those. Some artifacts were found, yet overall segments presented good quality. The feature extraction performed on those intervals allows to derive important variables that could be more informative and descriptive than the ECG signal [71]. Two Python packages were used to extract the features, Neurokit[2] and Neurokit2[3], along with statistical functions. The statistical operations allowed to obtain more attributes from the ones firstly extracted. Table 4.3 lists the 45 features that were calculated.

Primarily, HRV indices were obtained from R peaks samples, and it was only analyzed the ones considered appropriate for short-term measurements in the literature. Short-term indices are based on approximately 5-minutes intervals of HRV data [72]. For time-domain measurements, RMSSD, SDNN, pNN50, pNN20, HTI and TINN were referenced as suitable for short-term measures [72], [73], [74]. Additionally, MeanNN [73], SDSD, and CVNN were also considered appropriate for 5-minutes intervals [75], and MedianNN and CVSD were also calculated. For frequency-domain indices, spectral power density of the LF and the HF bands, and LF/HF ratio were introduced as short-term measures [72], [74]. In addition, the normalized LF (LFn), the normalized HF (HFn), and the log-transformed HF (LnHF) were also determined. For non-linear measures, the sample entropy of HRV (HRV_SampEn), the approximate entropy of HRV (HRV_ApEn), SD1, SD2, and SD1/SD2 were selected [72], [73].

Furthermore, R peaks also allowed to calculate HR. After HR calculation, the signal amplitude was normalized regarding the patient's baseline with Equation 4.1, and the minimum, maximum, mean, median, standard deviation, and the difference between the maximum and minimum of the normalized HR signal were computed.

Additionally, features related to the ECG waves peaks and location of the waves were calculated. The P and Q waves were not used because only a small number of waves was detected in several segments, compromising the results. Firstly, it was calculated the mean and median amplitude of each wave peak. Secondly, the mean and median differences between consecutive waves were computed, with the wave differences normalized in relation to patients' baselines.

Lastly, it was also computed the mean and median area of the cardiac cycles of each segment and the sample and approximate entropy of the ECG signal.

Besides the HR signal, which has the same length as the ECG segments, the other measurements are represented by a single value for each time interval.

After feature extraction, an initial exploratory analysis was conducted, observing feature behavior to remove the ones without useful information, such as features with constant or missing values.

[2] https://neurokit.readthedocs.io/en/latest/

[3] https://neurokit2.readthedocs.io/en/latest/

Table 4.3: Description of the ECG features extracted in the processing stage.

No	Symbol	Unit	Feature description
Time-domain HRV features			
1	MeanNN	ms	Mean of RR intervals
2	MedianNN	ms	Median of the absolute values of the successive differences between RR intervals
3	SDNN	ms	Standard deviation of RR intervals
4	SDSD	ms	Standard deviation of successive differences of RR intervals
5	RMSSD	ms	Root mean square of successive differences of RR intervals
6	CVNN	%	Standard deviation of the RR intervals (SDNN) divided by the mean of the RR intervals (MeanNN)
7	CVSD		Root mean square of the sum of successive differences (RMSSD) divided by the mean of the RR intervals (MeanNN)
8	pNN50	%	Proportion of RR intervals greater than 50ms, out of the total number of RR intervals
9	pNN20	%	Proportion of RR intervals greater than 20ms, out of the total number of RR intervals
10	TINN	ms	Baseline width of the RR intervals distribution obtained by triangular interpolation
11	HTI		Total number of RR intervals divided by the height of the RR intervals histogram
Frequency-domain HRV features			
12	LF	ms^2	Spectral power density of the low-frequency band (0.04 - 0.15 Hz)
13	HF	ms^2	Spectral power density of the high-frequency band (0.15 - 0.4 Hz)
14	LFHF		Ratio of LF-to-HF power
15	LnHF		Natural logarithm of HF
16	LFn		Normalized low frequency
17	HFn		Normalized high frequency
Nonlinear HRV features			
18	SampEn		Sample entropy
19	ApEn		Approximate entropy
20	SD1	ms	Poincaré plot standard deviation perpendicular the line of identity
21	SD2	ms	Poincaré plot standard deviation along the line of identity
22	SD1/SD2		Ratio of SD1-to-SD2
Other features			
23	HR	bpm	Heart rate signal
24	MeanHR	bpm	Mean heart rate
25	MedianHR	bpm	Median heart rate
26	MinHR	bpm	Minimum heart rate
27	MaxHR	bpm	Maximum heart rate
28	StdevHR	bpm	Standard deviation of heart rate
29	MaxMinHR	bpm	Difference between the highest and lowest heart rates
30 31 32	Mean_Diff_R_peaks Mean_Diff_S_waves, Mean_Diff_T_waves,	ms	Mean of successive differences of R-R, S-S, and T-T intervals
33 34 35	Median_Diff_R_peaks Median_Diff_S_waves, Median_Diff_T_waves	ms	Median of successive differences of R-R, S-S, and T-T intervals
36 37 38	Mean_Amplitude_R_peaks, Mean_Amplitude_S_waves, Mean_Amplitude_T_waves		Mean of R, S, and T waves peaks amplitude
39 40 41	Median_Amplitude_R_peaks, Median_Amplitude_S_waves, Median_Amplitude_T_waves		Median of R, S and T waves peaks amplitude
42	Mean_Heartbeats_Area		Mean area of all cardiac cycles
43	Median_Heartbeats_Area		Median area of all cardiac cycles
44	ECG_SampEn		Sample entropy of ECG
45	ECG_ApproxEn		Approximate entropy of ECG

4.6 Univariate Analysis

The univariate analysis was performed in all 45 features, and it was divided into two studies, one based on the pain reports and the other on the analgesia effects, which will be explained in the following sections.

4.6.1 Pain-reported analysis

Initially, the computed features were studied using boxplots to visualize the distribution of the feature according to pain reported. The boxplots were drawn to assess differences between pain absence (no pain reports) and pain presence (mild, moderate, and severe pain reports), using notched boxes to facilitate visual comparison. If the notch-intervals do not overlap, there is evidence that the medians of the two groups being compared are different at the 5% level [76]. Figure 4.3 represents the behavior of some of the extracted features. These features were selected among the ones with low correlation and were chosen because they seemed to expose differences between the two groups.

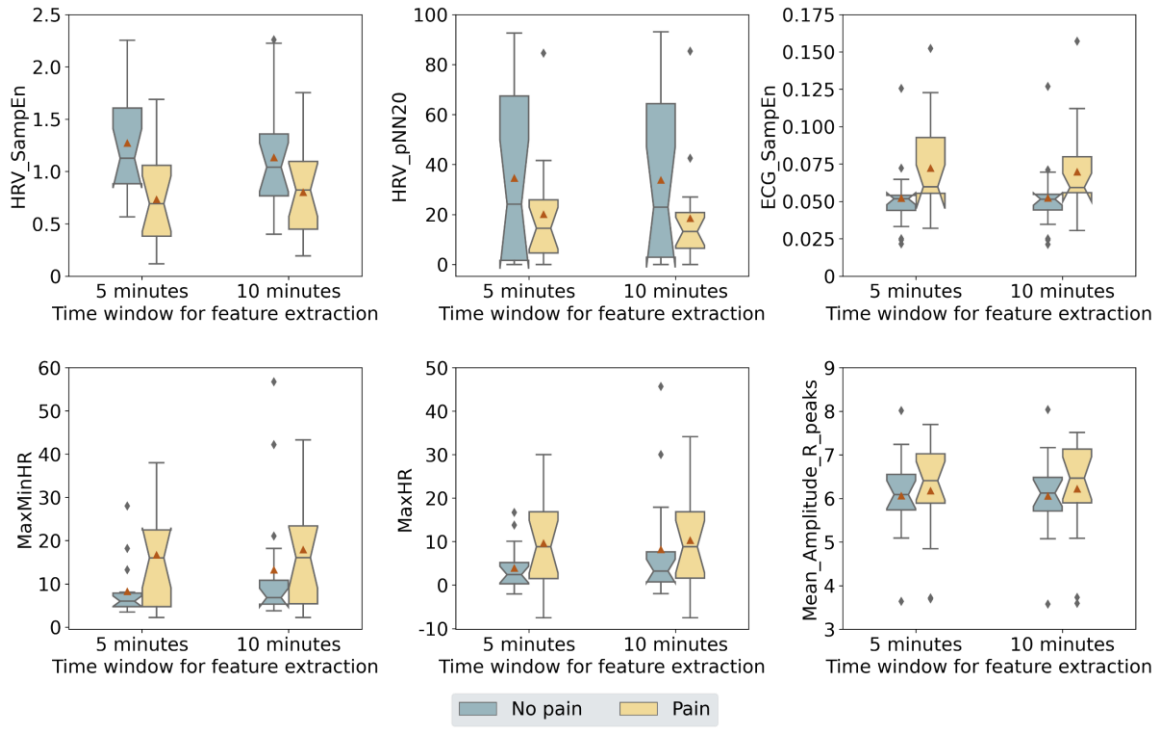

Figure 4.3: Boxplots of ECG features representing the distribution among two groups: pain absence and pain presence. The features include HRV_SampEn, HRV_pNN20, ECG_SampEn, MaxMinHR, MaxHR, and Mean_Amplitude_R_peaks, extracted from 5 minutes and 10 minutes intervals. The red triangle represents the mean.

Firstly, we can see that all boxes overlap, although some differences between the two groups can be discerned. For the range of HR (MaxMinHR), the notch intervals do not overlap, which could indicate that likely there is a difference between the two groups. In addition, the “pain” category presents higher variability, and the “no pain” category has lower values of MaxMinHR. This behavior is also observed for maximum HR, sample entropy of ECG, and mean amplitude of R peaks. Contrarily, pain presence presents the lowest variability and lower values of pNN20 than pain absence. In relation to HRV_SampEn, the reports respective to pain presence have lower values than no pain reports.

4.6.2 Analgesia effect analysis

Afterward, with the goal of investigating if pain-relief medication posed a signature on the features, the significant difference between features corresponding to before and after pain control was tested with Wilcoxon signed-rank test (as features were not normally distributed). The statistical significance level was set at 5%. In addition, boxplots were created to visualize how the data of each feature behaves before and after analgesia. Figure 4.4 shows the boxplots, and Table 4.4 contains the p-values of the Wilcoxon test obtained for three of the features, as an example of this univariate analysis results. The same analysis was done for the remaining features.

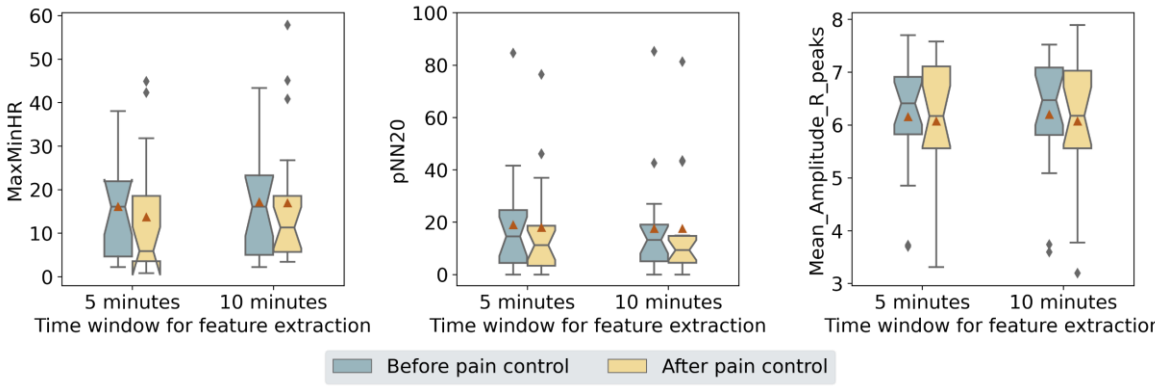

Figure 4.4: Boxplots representing ECG features differences before and after analgesia. The features include MaxMinHR, pNN20, and Mean_Amplitude_R_peaks, for 5 and 10 minutes intervals. The red triangle represents the mean.

Table 4.4: Wilcoxon signed-rank test's p-value obtained to test differences of ECG features before and after analgesia. The features include MaxMinHR, pNN20, and Mean_Amplitude_R_peaks, extracted from 5 and 10 minutes intervals.

Time interval	MaxMinHR	pNN20	Mean_Amplitude_R_peaks
5 minutes	0.799	0.522	0.196
10 minutes	0.966	0.609	0.054

In general, it was not observed significant differences between the two groups. The resulting boxplots show that the boxes of both groups (before and after pain analgesia) overlap for both time intervals. In addition, compared to the boxplots results in Figure 4.3, the features do not behave as expected. On the one hand, in Figure 4.3, no pain ratings have lower and less variable mean amplitude of R peaks and more variable pNN20. However, after analgesia, this behavior is not observed. On the other hand, MaxMinHR behaves similarly to the pain intensity analysis, but the differences are less noticeable and prominent. In addition, the Wilcoxon results support what has been observed in the boxplots since the test exposed no statistically significant differences between the two groups for any of the features in neither 5 minutes intervals nor 10 minutes intervals.

4.6.3 Discussion

It is expected that analgesia had reduced pain and, consequently, physiological changes would occur and reveal pain relief. However, the features extracted from both time segments could not discriminate

nor indicate pain relief after pharmacological intervention. On the one hand, the analgesic effect may not have been enough to provoke an alteration in the ECG signal. Moreover, the individual features may not provide a complete description of the pain response and may not be sensitive enough. In other words, they may not be appropriate or may be insufficient to detect and discriminate differences.

It is important to refer that the results were drawn on a limited dataset and, therefore, the information may have been insufficient. The study of the individual behavior of each feature should be conducted on a sample with more patients and with more pain assessment points. In addition, to complement this analysis in the future, pain should be assessed after analgesia to perceive if patients report pain reduction.

4.7 Unsupervised Multivariate Analysis

While the behavior of the individual features proved to be unsatisfactory, when combined, they may complement their information, and the pain description could improve. Therefore, an unsupervised multivariate analysis was explored to determine whether combinations of features are associated with and expose pain characteristics. Unsupervised learning is carried out by clustering algorithms which are performed considering that the data is unlabelled. The main goals of this task are to examine the data, extract useful information, and find groups that are similar. In short, it aims to discover patterns, associations, and relationships within the features from ECG signals concerning the postoperative pain experience.

The studies proposed for clustering analysis will investigate the pain scores and the number of non-zero levels of pain reported by each patient. Using the features computed on both time segments, two hypotheses will be explored:

i. If the features are grouped accordingly to the reported pain intensity, presenting similar behavior within pain intensities;

ii. If the features expose similarities respective to the number of times a patient reported pain during the postoperative period.

In addition, this study will explore the influence of the number of features used as input to the clustering algorithm, which varied from 2 to 10 features, and evaluate the performance of three clustering algorithms: k-means, agglomerative and spectral clustering. The library scikit-learn[4] of Python will be used to perform the multivariate analysis.

Firstly, this subchapter will describe feature transformation and feature selection, which precedes the clustering tasks. After, the clustering models will be described, and the performance metrics used to evaluate and compare the results are listed. Finally, the case studies will be introduced, and the results of the unsupervised approach will be displayed and discussed.

4.7.1 Feature transformation

Since the extracted features have different ranges and order of magnitudes, they should be transformed into the same scale to ensure a fair comparison, avoid misleading results, and improve the performance of machine learning models. One approach is to scale features between 0 and 1 with the min-max scaling [71].

[4] https://scikit-learn.org/stable/

4.7.2 Feature selection

Feature selection is crucial before clustering tasks because the presence of irrelevant and redundant features can degrade the model's performance [77]. Moreover, having a smaller set of features reduces the complexity of the model and reduces run time.

There are different types of feature selection methods, and, in this dissertation, feature selection was performed through feature correlation and variance analysis followed by backward sequential selection.

Feature selection using correlation and variance

One way of feature selection is to remove redundant variables using the correlation between features [71]. As features were not normally distributed, the Spearman correlation coefficient was computed. The resulting matrices obtained for both time intervals were represented as correlation heatmaps to observe the correlation's strength and direction (positive or negative) through a color code of the 44 features (HR signals were not used). The heatmaps are available in a Github repository[1] for the features extracted from both 5 minutes and 10 minutes intervals.

Features were considered strongly correlated if the correlation coefficient was superior to 0.9 and the statistical significance level was set at 5%. The one with higher variance was selected from the pair of highly correlated features because data with higher variance provide more information. For the features extracted from 5-minutes intervals, 24 of them have a correlation greater than 0.9, whereas, for the 10-minutes, 23 features are highly correlated. For example, literature sustains that RMSSD is equivalent to SD1, both representing short-term HRV [72], and it was obtained a perfect positive correlation between these two features, for both time intervals (r=1; p=0.0).

Table 4.5 presents the selected features, and Figure 4.5 shows the correlation heatmap for the selected features in the 5-minutes interval, according to this procedure.

Table 4.5: Selected features after studying features' correlation and variance, for both 5 and 10 minutes intervals.

Time interval	Features	Total
Subset 5 min	MedianNN, pNN20, TINN, HTI, LF, LFHF, LnHF, SampEn, SD2, ECG_SampEn, ECG_ApproxEn, MinHR, MaxHR, StdevHR, MaxMinHR, Mean_Amplitude_R_waves, Mean_Amplitude_S_waves, Median_Amplitude_T_waves, Mean_Diff_R_peaks, Mean_Heartbeats_Area	20
Subset 10 min	Subset 5 min + pNN50	21

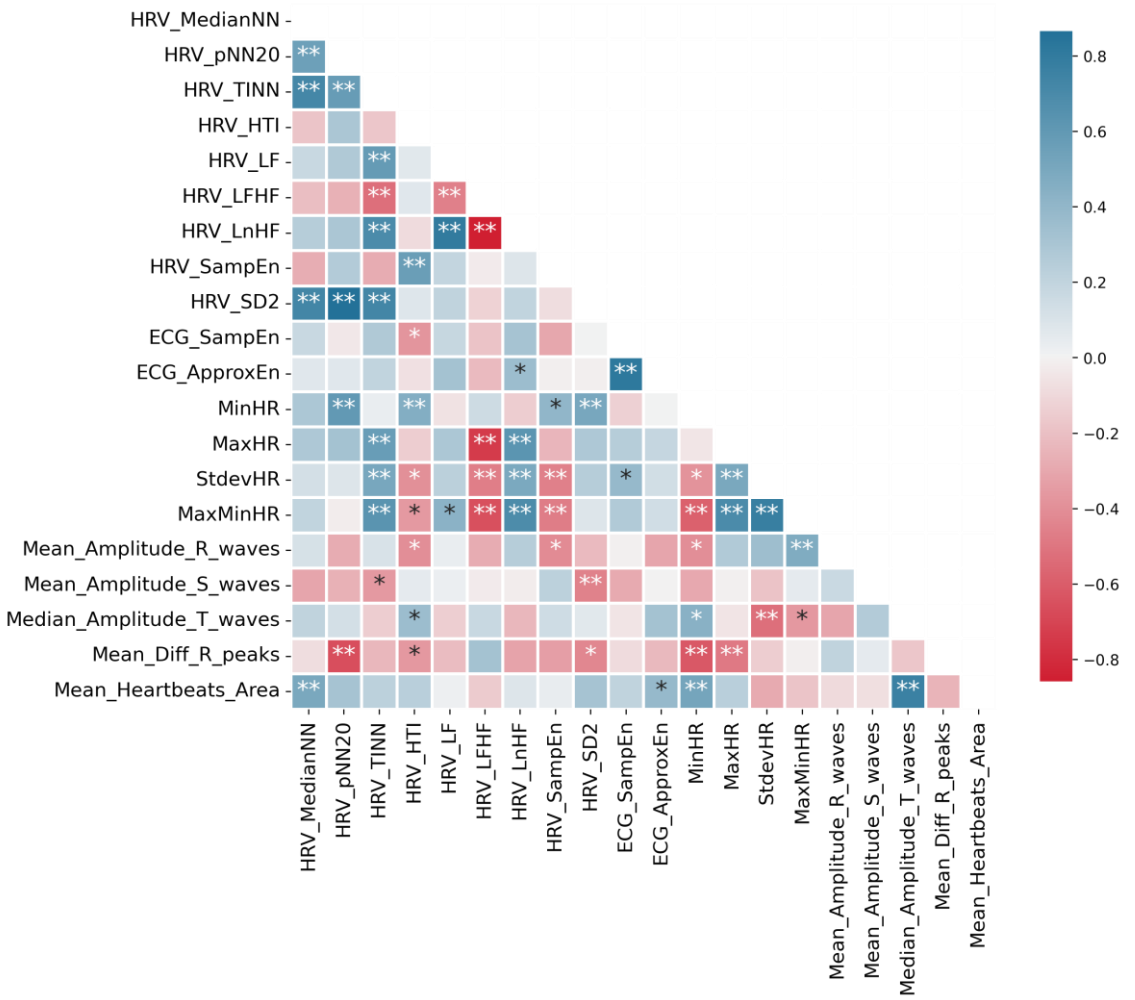

Figure 4.5: Heatmap of the correlation matrix of the features selected after correlation and variance analysis, for 5-minutes interval. Symbols * and ** represent statistical significance for $p<0.05$ and for $p<0.01$, respectively.

Feature Selection using backward elimination

A second selection technique was applied to reduce the input variables to the clustering model by choosing a subset of important features from the previously selected features. In this work, a varying number of features from 2 to 10 was tested. This selection was made with the backward elimination (using scikit-learn from Python[5]), selecting the important variables based on the relation with the desired output. Backward elimination starts with the set of all available features, and one feature at a time is removed. This algorithm performs k-fold cross-validation when removing the feature at each iteration to find the best one to remove from the set based on a greedy procedure that improves the estimator score. The algorithms K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) were used as estimators. The scores of the estimators are calculated for the subset of features and compared to the other subsets, and the feature selected to be removed is the one belonging to the subset that shows the least performance loss [77]. It is important to note that the target variable varied according

[5] https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.SequentialFeatureSelector.html

to the case study, so different subsets of features were selected for the different case studies (for pain-reported and the number of reported non-zero levels of pain analyses).

4.7.3 Clustering models

Three clustering algorithms were tested, including k-means, agglomerative and spectral techniques.

K-means clustering[6]

The k-means clustering algorithm considers the distance of the samples to a centroid as a measure of similarity. It chooses centroids following these steps:

1. Select initial centroids;
2. Measure the distance between each sample and the centroids and assign the sample to the nearest centroid;
3. Calculate the mean of all samples assigned to the centroid, which value corresponds to the new centroid;
4. The last two steps are repeated until the difference between the new centroids, and the previous one is lower than a selected threshold.

Agglomerative hierarchical clustering[7]

Hierarchical clustering works by merging or splitting the clusters successively. In agglomerative clustering, the algorithms work as a "bottom-up" approach as it starts with each sample in its cluster and then merges pairs of clusters as the hierarchy moves up. This clustering technique considers the distance between samples as a measure of similarity.

1. Assign each sample to one cluster;
2. Determine the distance measurement between samples;
3. Determine the linkage criteria to merge the clusters;
4. In the end, all samples are grouped into one cluster.

Spectral clustering[8]

Spectral clustering examines graph distance as a metric of similarity.

1. An affinity matrix, also called a similarity matrix, is created;
2. Compute the components of the eigenvectors to define a feature vector for each sample in a low dimensional space. For k clusters, compute the first k eigenvectors;
3. Perform k-means on the feature vectors to separate the samples into clusters.

4.7.4 Evaluation metrics

Three setups were tested with the unsupervised approach: the size of the time window used for feature extraction, the number of features as input for the clustering tasks, and the clustering algorithm. For each combination of these variables, a confusion matrix and several performance metrics were computed and examined. Briefly, a shorter interval would be optimal since less information would be needed. Additionally, a smaller input to the algorithm, that is, a smaller number of features, may improve the algorithm performance and reduce the complexity and computing time.

[6] https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html
[7] https://scikit-learn.org/stable/modules/generated/sklearn.cluster.AgglomerativeClustering.html
[8] https://scikit-learn.org/stable/modules/generated/sklearn.cluster.SpectralClustering.html

The performance of the clustering algorithms can be evaluated by confusion matrices, which allow easy and quick visualization of the performance of an algorithm by considering actual and predicted values. In binary classification, it can be represented by the matrix in Figure 4.6.

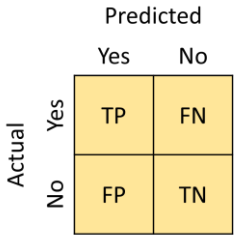

Figure 4.6: Confusion matrix for binary problems. TP: True Positive, FP: False Positive, FN: False Negative, TN: True Negative.

The performance metrics based on the confusion matrix obtained to evaluate the results from clustering methods were accuracy, precision, recall, f1-score, and Matthews correlation coefficient.

- **Accuracy** is the number of correct predictions to the total number of input samples, and it is measured using the formula in Equation 4.2 [78]. If the data is imbalanced, the number of elements in one class is larger than in the other, and accuracy is not an appropriate measure.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2}$$

- **Precision** identifies how many of the predicted as positive are actually positive, and is expressed in Equation 4.3 [78].

$$Precision = \frac{TP}{TP + FP} \tag{4.3}$$

- **Recall/Sensitivity** corresponds to the number of correct predicted positives out of all positives, and it is calculated with Equation 4.4 [78].

$$Recall = \frac{TP}{TP + FN} \tag{4.4}$$

- **F1-score** combines precision and recall into one measure (the harmonic mean), as seen in Equation 4.5 [78].

$$F_1 = 2\frac{precision \times recall}{precision + recall} \tag{4.5}$$

- **Matthews correlation coefficient (MCC)** can be calculated by Equation 4.6, and it varies between -1, if TP=TN=0, and 1, if FP=FN=0. MCC is a good metric for imbalanced datasets, and it indicates that both classes are predictable well if the value is close to 1. Therefore, MCC can be used if both classes are of interest [79].

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{4.6}$$

4.7.5 Clustering results

Three case studies will be presented, which will be treated as binary clustering problems. Case studies A and B consider the pain scores reported by the patients during the postoperative period. The data points were divided accordingly to their intensity to investigate similarities within pain intensities. Cases study C analyze the number of mild, moderate, or severe pain reports each patient gives (see Table 4.2). The last case study aims to investigate the patient profiles during the postoperative period and explore if patients who reported pain more times have physiological alterations different from those who reported fewer times and that seem to have pain managed more efficiently. This analysis is done independently of the intensity of the pain, meaning that scores corresponding to different intensities will belong to the same category. Although a patient may describe different pain intensities, the clustering algorithms may find similarities within the different scores. It is also important to refer that the number of reports considered in this analysis includes the assessments made before ECG collection.

Case-study A: "no pain" versus "pain"

Concerning the pain levels, it was tested if scores of 0 (12 samples) were grouped in a different cluster from scores of 1, 2, and 3 (20 samples). The negative class corresponds to the "no pain" condition, whereas the positive corresponds to the "pain" state. The goal is to assess if it is possible to discriminate between the absence and presence of pain.

Figures 4.7 and 4.8 display the performance results of the shorter (5-minutes) and longer (10-minutes) intervals, respectively, obtained by varying the number of features and the clustering model. It is clinically important to accurately group and predict no pain and pain samples. Therefore, accuracy is not represented in this case study since this measure can be biased and misdealing when classes are imbalanced. That is, accuracy can be high even though the minority class examples are grouped incorrectly.

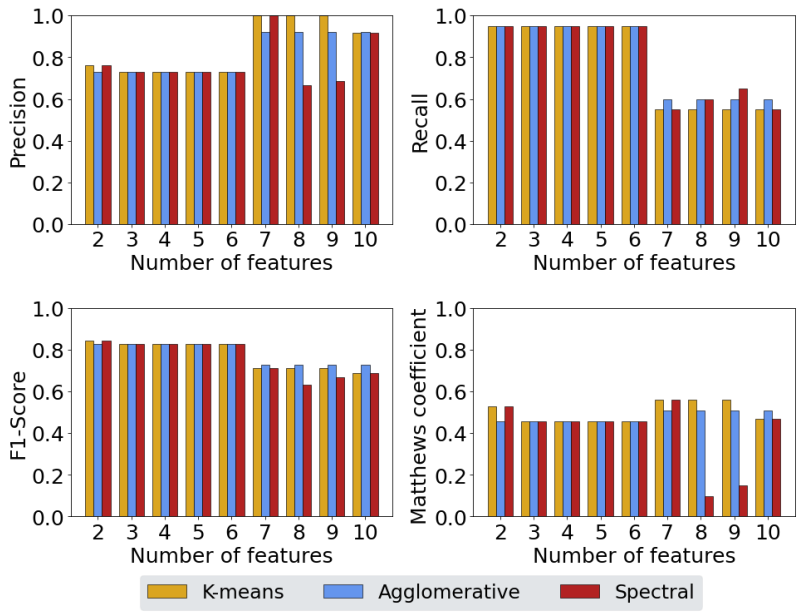

Figure 4.7: Precision, recall, f1-score, and MCC obtained for the study of absence and presence of pain using features extracted from 5-minutes intervals.

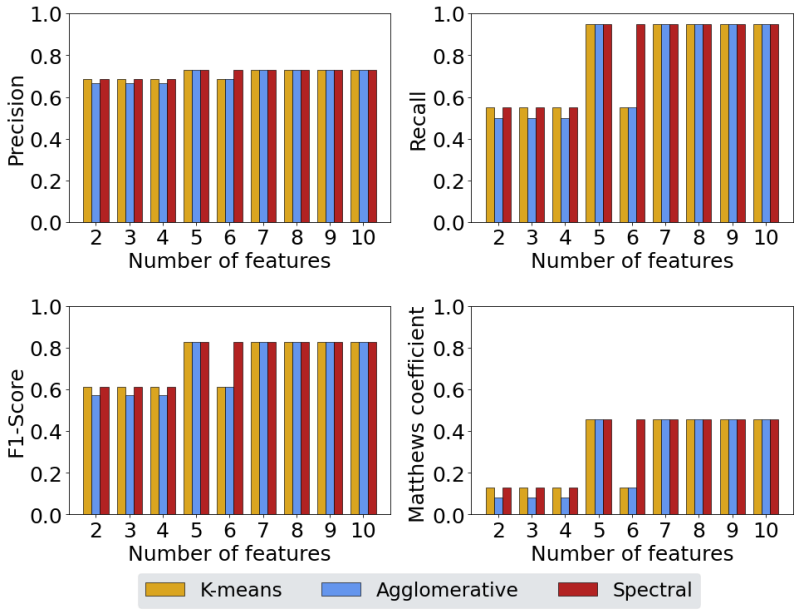

Figure 4.8: Precision, recall, f1-score, and MCC obtained for the study of absence and presence of pain using features extracted from 10-minutes intervals.

The bar plots in Figure 4.7 allow concluding that k-means and spectral clusters outputted the best overall performance for the smallest subset of features: pNN20 and mean area of heartbeats. Another observation is that the precision increases for bigger subsets of features. In some cases, the precision reaches the maximum value (precision=1), which means there is no FP. Even though precision improves significantly for a higher number of features (≥ 7 features), recall significantly decreases as the FN increases. Therefore, even with a higher number of features, the data is still not perfectly separable by the two clusters.

In Figure 4.8, for the 10-minutes interval, the performance metrics only vary within the subsets with two to six features. For the remaining subgroups of features, the metrics remain constant for the three algorithms. The selected subset of two features is different from the previous one and comprehends two HRV metrics, MedianNN and pNN20, which did not lead to good results. Lastly, the results show that the data is not perfectly separable into two clusters, however the three clustering techniques demonstrated that the subset of 5 features achieved the highest and the same overall performance than subsets of more than 6 features.

Figure 4.9 shows the confusion matrices corresponding to the best overall results of both time segments.

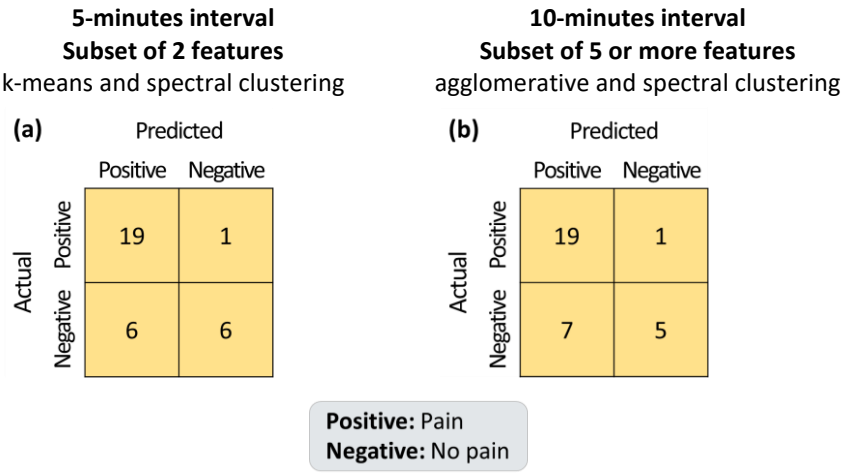

Figure 4.9: Confusion matrices of the best results for clustering analysis between "no pain" and "pain" conditions.

For the 5-minutes study, the best overall performance was achieved, with k-means and spectral clustering, for the subset of 2 features (pNN20 and the mean area of the heartbeats), obtaining accuracy=78.1%, precision=76.0%, recall=95.0%, f1-score=84.4%, and MCC=0.53, which performance is superior to the one obtained with the features extracted in the 10 minutes intervals. Accuracy is high, but the results of accuracy can be misleading. Although 95% of "pain" are accurately grouped in the same cluster, only 50% of "no pain" are placed in the correct group, as we can see in the confusion matrix in Figure 4.9 (a). Consequently, half of the negative class would be incorrectly diagnosed, which would have resulted in the unnecessary treatment of patients. MCC result shows that the predicted class and the true class are weakly correlated, which is attributed to the fact that clustering with the two features is not so good at grouping "no pain" reports together.

In contrast, combinations of a higher number of features could accomplish good discrimination of negative class (FP=0) but failed to set apart the positive one (FN=9). Hence, using more features, more patients would have remained undiagnosed and potentially undertreated.

Therefore, a trade-off needs to be accomplished: with a small subset of features, there is an increase in the number of false alarms (FP), whereas a greater number of features increase the missed detections (FN).

For the 10-minute analysis in Figure 4.9 (b), the subsets of 5 or more features achieved the same results for the three clustering methods. Therefore, with accuracy=75.0%, precision=73.1%, recall=95.0%, f1-score=82.6%, and MCC=0.45, the subset of 5 features achieved the highest performance, allowing to describe the problem with fewer features and proving that the remaining are not relevant to separate both classes. Thus, regardless of the clustering method, the obtained results for the subset of five features (pNN20, Mean_Heartbeats_Area, MedianNN, HTI, and Mean_Diff_R_peaks) indicate a suitable separation into two clusters. However, less than half of the negative class is well predicted.

Scatter plots representing the samples in two-dimensional feature space were also determined to visualize the relationship between variables. Although the algorithms were tested using multiple variables to cluster the data, only scatter plots displaying two variables will be drawn. This tool allows us to understand better how the algorithm performs the task of grouping the data points and, since the division performed by k-means is also represented in the plots, it also allows observing the scores provided in the confusion matrix for this clustering technique.

The scatter plots of HRV_pNN20 and Mean_Heartbeats_Area for both time intervals are represented in Figure 4.10. Although this feature set was not selected by backward elimination in the 10 minutes study, the plot was also drawn to compare it with the data points distribution obtained with 5 minutes intervals. Each input feature defines an axis, and each point in the plot represents a report from a patient. A third variable was added that indicates the categorical values of the patients' reports to distinct each point, encoding through color, between “no pain” and “pain” conditions. Figure 4.10 also contains the cluster centers obtained with k-means for both time segments. For both cases, the cluster assigned to the positive class (“pain”) typically had lower values of pNN20 and lower values of the heartbeats' mean area, compared with the one assigned to the negative class (“no pain”). A decision boundary between the two clusters was created. In k-means, the line represents the equidistant points from the two centroids and the two regions separated by the decision surface correspond to the two classes generated by the algorithm. Each observation in the space is assigned to one of the classes, and this way, it is easier to see how the k-means model divides the data points between the two clusters for a two-dimensional feature space.

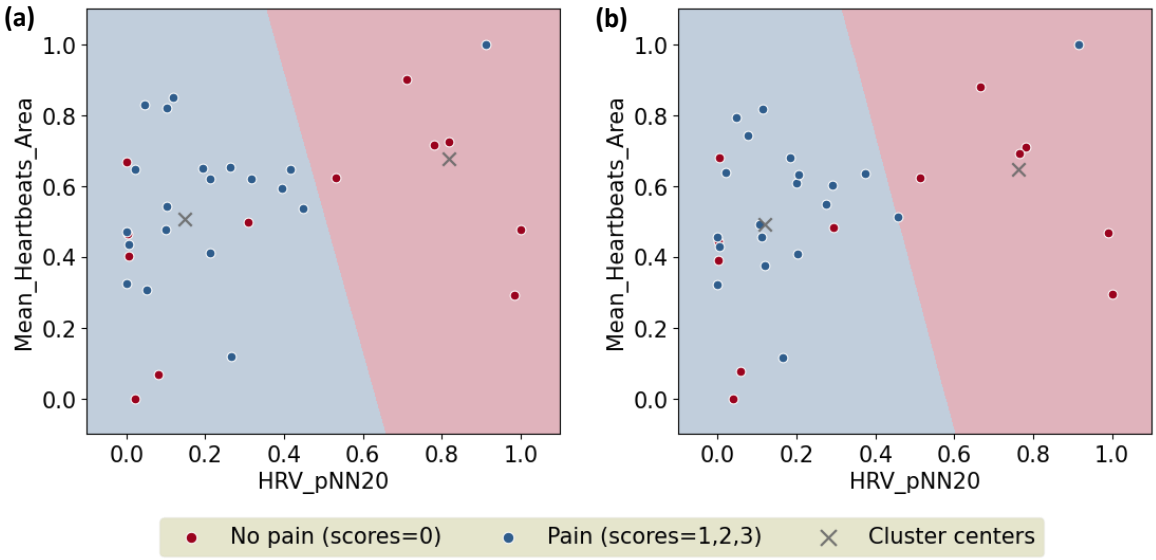

Figure 4.10: Scatter plot representing pNN20 and mean heartbeats area, under the presence and absence of pain. **(a)** Data points for the features extracted from 5-minutes intervals. **(b)** Data points for the features extracted from 10-minutes intervals.

Firstly, by checking the results in Figure 4.10, it is observed that the behavior of the pair of features is identical in both time segments. However, the performance of the k-means algorithm decreases for the 10-minutes interval because the number of FN is superior due to a data point close to the boundary in the 5-minutes graph that was assigned to the opposite class.

Secondly, for both time segments, it is possible to notice that the data points are dispersed and do not seem to resemble any pattern to group into the two different conditions. Additionally, the majority of the data points are placed in the same group, as the clustering algorithms acted poorly at grouping the negative class. The plots display how scores corresponding to the “no pain” category were evenly divided between the two clusters. These results lead to believe the physiological characteristics exposed by pNN20 and mean area of heartbeats of some patients who referred not feeling pain may be similar to physiological alterations of patients that reported the presence of pain. It may be related to the

subject's pain perception, which depends on several variables. For instance, a person may be under pain, and however, due to past painful events, they may perceive and underestimate the postoperative pain as discomfort.

Additionally, we can also see an extreme value, with a high value of pNN20 and mean heartbeats area, that corresponds to the one FN identified in the confusion matrix in Figure 4.9 (a), which could be an outlier as it falls outside the other examples of the positive class. This observation, marked as "pain" condition, should be closer to the positive class centroid in the plot's blue region. This data point belongs to a patient who informed that he was not feeling pain in the two previous assessments, and at the time of this report, he may have been under low pain sensation.

Furthermore, it was possible to see that the performance metrics results dramatically changed from the 5 minutes to the 10 minutes analysis for the smaller subsets of features, which shows that the correct selection of features dramatically influences the results in terms of performance. For example, in the case of 10 minutes, the set of two features that the model selected conduced to lower performance than the features represented in the scatter plot in Figure 4.10 (b), which revealed that pNN20 and mean heartbeats behaved better in the 10 minutes and outputted greater results than the two features selected by backward elimination (pNN20 and MedianNN). Hence, the performance can increase by performing manual feature selection and testing other combinations of features.

In summary, the best overall performance for positive prediction in the study of "no pain" *versus* "pain" was obtained using the smallest time window and only two features, implementing k-means and spectral clustering algorithms. The clustering techniques are successful in group non-zero pain scores in the same cluster (high recall values), and, even though these results suggest that the two features could be used as a positive predictor for pain, no pattern of grouping apart the two classes are recognized when looking at the scatter plots. In addition, this set of two features proved not to be enough to discriminate zero pain scores as 50% of the examples were grouped along with the examples of non-zero pain scores. Finally, the overall performance of the clustering models indicates that as the number of features increases, the algorithms still fail to form two distinct groups accordingly to the pain scores reported by the patients.

Finally, it is also critical to acknowledge that the data consists of only 32 samples. In a small dataset, as is the case, the impact of an outlier can be much greater. Moreover, as there are very few examples in the dataset, finding feature characteristics for different pain conditions is even more challenging. In addition, the distribution of examples across the classes is not equal. The samples of the negative class (n=12 samples) correspond to 37.5% of the data points. Because there are so few total examples, the imbalance becomes more significant even if the distribution of examples is uneven by a small amount. For this reason, the results need to be carefully interpreted, and further study is necessary using a substantially larger dataset.

Case study B: "*tolerable pain*" *versus* "*intense pain*"

Following the previous case study, an exploration of how the clustering performs to separate "tolerable pain", reported as no or mild pain, (16 samples) from "intense pain", reported as moderate or severe pain, (16 samples) will be performed. In this case, the observations are evenly divided across the two classes. The negative corresponds to the "tolerable pain" condition, whereas the positive class corresponds to "intense pain".

Figure 4.11 and Figure 4.12 display the performance metrics obtained by varying the features and the clustering algorithm for 5 and 10 minutes segments, respectively.

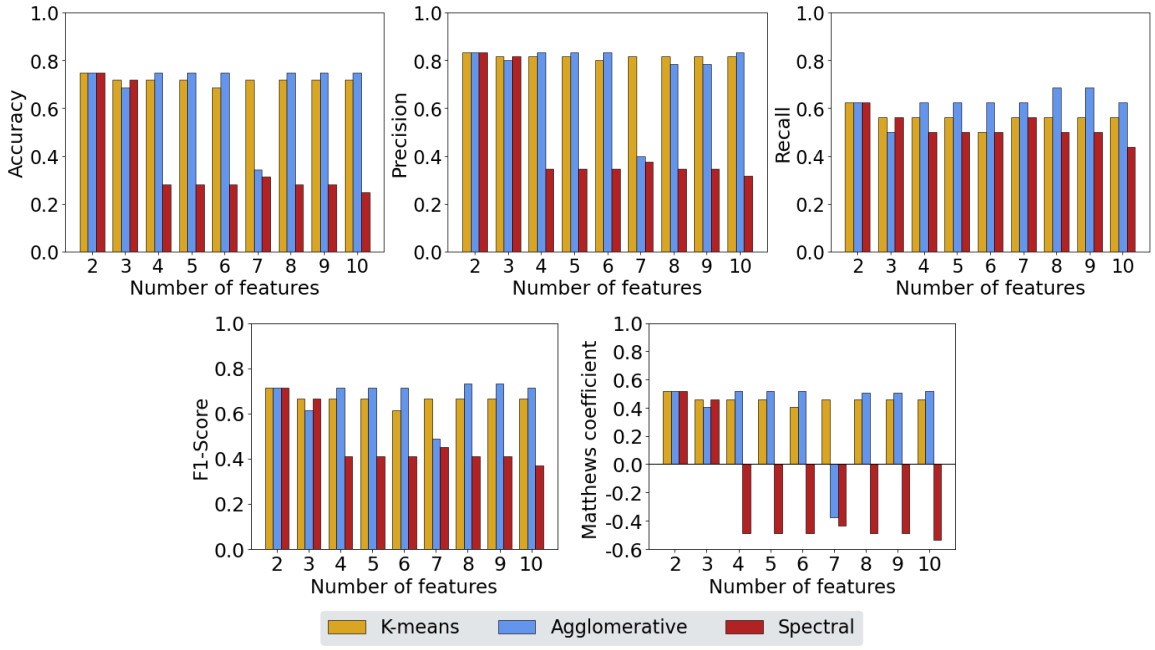

Figure 4.11: Accuracy, precision, recall, f1-score, and MCC obtained for the study of "tolerable pain" *versus* "intense pain" using features extracted from 5-minutes intervals.

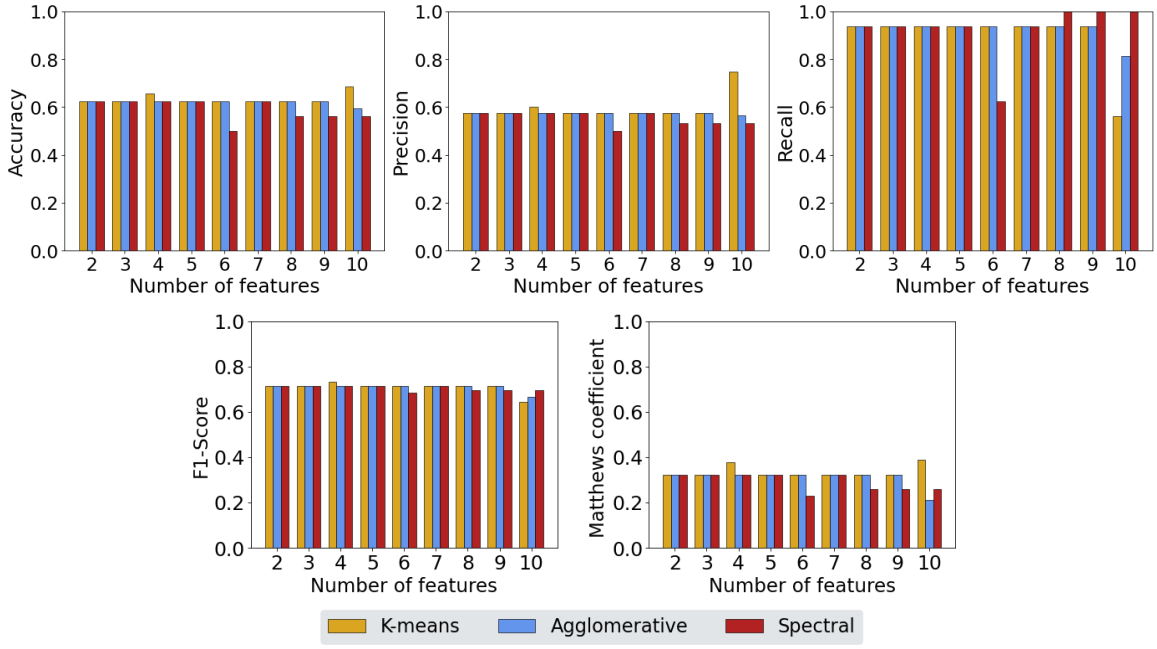

Figure 4.12: Accuracy, precision, recall, f1-score, and MCC obtained for the study of "tolerable pain" *versus* "intense pain" using features extracted from 10-minutes intervals.

Comparing the results of the two figures, it is observed that the features selected in the 10-minutes analysis provided inferior results and that the best result for this case study was obtained in the 5-minutes analysis, using two features and the three clustering algorithms. The two selected features by the backward elimination technique were a non-linear HRV feature, the sample entropy (HRV_SampEn), and the difference between the maximum and minimum heart rate (MaxMinHR). Even though this subset obtains the best result among the various subsets of selected features, f1-score and MCC are

low. Regarding the 10-minutes analysis, the best performance is respective to the k-means clustering with four features, including pNN20, LnHF, Mean_Heartbeats_Area, and Median_Amplitude_T_waves.

Figure 4.13 presents the two confusion matrices corresponding to the best results for each time analysis.

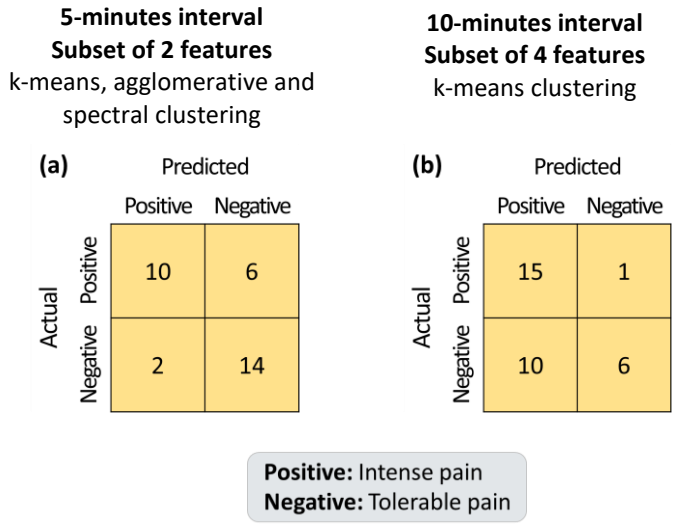

Figure 4.13: Confusion matrices of the best results for clustering analysis between "tolerable pain" and "intense pain".

On the left, for the 5-minutes analysis, the matrix resulted in accuracy=75.0%, precision=83.3%, recall=62.5%, f1-score=71.4%, and MCC=0.52. The recall is low, as several "intense pain" samples fall incorrectly in the "tolerable pain" group (FN=6), which would correspond to undertreatment. On the right, the best result for the 10-minutes analysis has accuracy=65.6%, precision=60.0%, recall=93.7%, f1-score=73.2%, and MCC=0.38. Here, oppositely, the precision is low since most negative class samples are placed in the group of "intense pain" (FP=10). Those examples would be incorrectly diagnosed with "intense pain", and correspond to overtreatment. To sum up, although the 5-minutes intervals achieved feasible results, in neither of the cases, it was possible to separate the two groups perfectly.

The data points corresponding to the two features extracted from 5 minutes intervals are distributed in the scatterplot in Figure 4.14. As in the previous analysis, the plot is divided into two regions of each cluster. The centroid of "intense pain" has lower HRV_SampEn and higher MaxMinHR than the centroid correspondent to "tolerable pain". As we can see, there are several observations from both classes, six "intense pain" reports and fourteen "tolerable pain" reports, concentrated in lower MaxMinHR values, that were placed in the "tolerable pain" group. The other data points in the opposite cluster have higher MaxMinHR values. The majority of them (10) correspond to "intense pain" reports, which indicates that a wider range between extremes in the HR could be characteristic of more acute pain. Even though these differences are detected among the data points, there are very few examples. In addition, the way the "intense pain" cluster observations are dispersed through the cluster region makes it more difficult to identify a distinct group from the other cluster when observing the plot, using the two features information. Therefore, these results need to be tested with more examples to corroborate what has been observed.

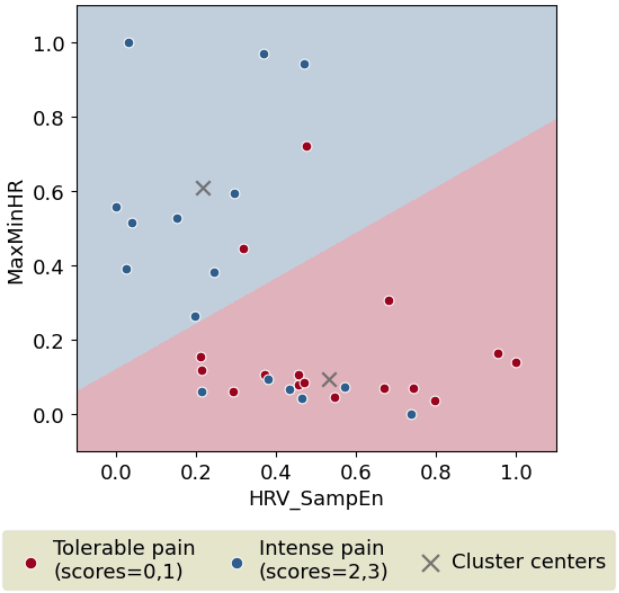

Figure 4.14: Scatter plot representing sample entropy of HRV and the difference between the maximum and minimum HR to study the differences between "tolerable pain" and "intense pain".

As final remarks on this case study, the reports are not perfectly divided between the two groups. Especially, several reports of "intense pain" expose physiological behaviors similar to the ones found on patients who reported "tolerable pain". The pain induced by the surgical procedure may not have an effect on some patients, and, therefore, the features' behavior will not be similar to patients in which pain had a bigger impact. The type of surgery could create more variability within the patients, and as each person has very distinct characteristics, patients reporting the same pain intensity may present similar or very different reactions from each other.

Case study C: Number of pain levels reported

Next, the clustering analysis was based on the number of reports of mild, moderate, or severe pain given by each patient. It was tested if the unsupervised learning algorithms can distinguish between assessments of patients who reported less than four times a non-zero level of pain (23 samples) and the reports of patients who stated that they felt pain four or more times (9 samples). The positive class corresponds to the inferior number of reports (no. assessments < 4), whereas the negative corresponds to the higher number of reports (no. assessments ≥ 4).

Figure 4.15 and Figure 4.16 represent the performance metrics for the 5-minutes interval and the 10-minutes interval analysis, respectively, obtained by varying the number of features and the clustering algorithm. As in case study A, accuracy will not be represented since the classes are imbalanced.

Figure 4.17 exhibits the confusion matrices corresponding to the highest f1-score and MCC values.

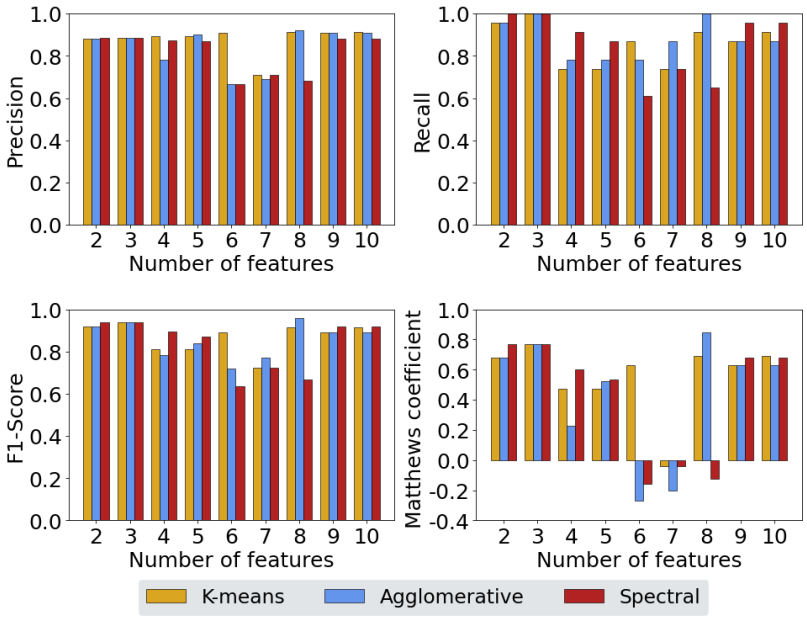

Figure 4.15: Precision, recall, f1-score, and MCC obtained for the study between "no. assessments < 4" and "no. assessments ≥ 4" using features extracted from 5-minutes intervals.

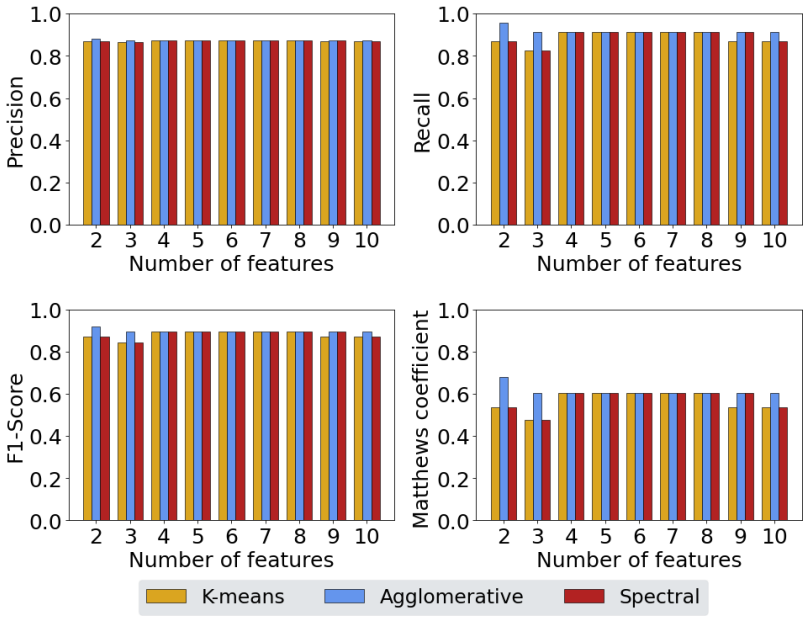

Figure 4.16: Precision, recall, f1-score, and MCC obtained for the study between "no. assessments < 4" and "no. assessments ≥ 4" using features extracted from 10-minutes intervals.

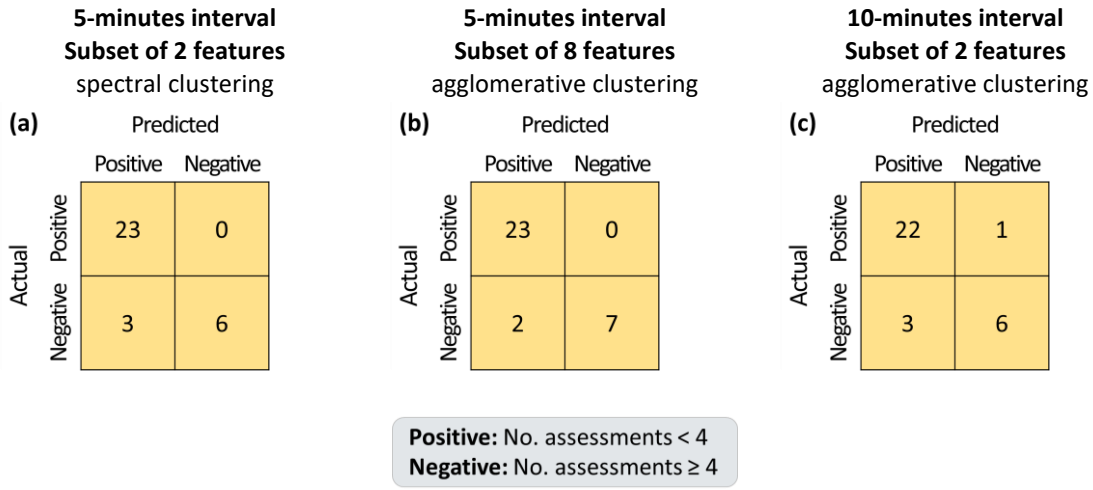

Figure 4.17: Confusion matrices of the best results for clustering analysis between "no. assessments < 4" and "no. assessments ≥ 4".

In Figure 4.15, three performance results can be remarked. Firstly, the performance of the spectral technique had good results using only two features of the 5-minutes HR signal, the standard deviation (StdevHR) and the maximum value (MaxHR). The corresponding confusion matrix is in Figure 4.17 (a). The performance metrics are accuracy=90.6%, precision=88.5%, recall=100.0%, f1-score=93.9%, and MCC=0.77. Secondly, the subset of three features, with StdevHR, MaxHR, and Mean_Heartbeats_Area, worked well in the three clustering models, achieving the same confusion matrix. Finally, the performance of the agglomerative algorithm is increased with eight features as input. In addition to the last three listed features, the set of eight also includes Mean_Amplitude_S_waves, Mean_Diff_R_peaks, Mean_Amplitude_R_waves, ECG_ApproxEn, and HRV_HTI. The matrix in Figure 4.17 (b) corresponds to the results of agglomerative clustering with eight features, having accuracy=93.8%, precision=92.0%, recall=100.0%, f1-score=95.8%, and MCC=0.85.

Figure 4.16 shows that, for the 10-minutes analysis, the agglomerative model behaved better. The best result belongs to agglomerative clustering using two features (StdevHR and Mean_Amplitude_R_waves), with the respective confusion matrix in Figure 4.17 (c), with accuracy=87.5%, precision=88.0%, recall=95.7%, f1-score=91.7%, and MCC=0.68. These results are equal to the performance using two features in the 5-minute analysis, for the same clustering algorithm and k-means.

Figure 4.18 represents the distribution of data points for the sets of two features extracted for each time window. In this study, the categorical variable in the scatterplot corresponds to the number of non-zero pain score reports of the patients, in accordance with being less than 4 reports and equal or greater than 4 pain assessments. For the 5-minute intervals, in Figure 4.18 (a), the cluster center assigned to the positive class has lower StdevHR and lower MaxHR values than the cluster centroid assigned to the negative class. Same as in case study A, clustering successfully groups the class with more examples in the same cluster. However, one-third of the negative class is also in that cluster. In this case study, the data is even more imbalanced than in case study A, which can compromise the results. Furthermore, one possible outlier is visible, which corresponds to a FN as represented in the confusion matrix in Figure 4.17 (c). In Figure 4.18 (b), the data points seem more dispersed. Six examples of patients who reported pain four or more times are located in higher values of the standard deviation of HR and mean amplitude of R waves, although the other three examples (one-third of the total sample of the negative class) have

lower StdevHR values. Comparing the results of both time intervals, two different sets of features were selected, but both provided satisfactory results.

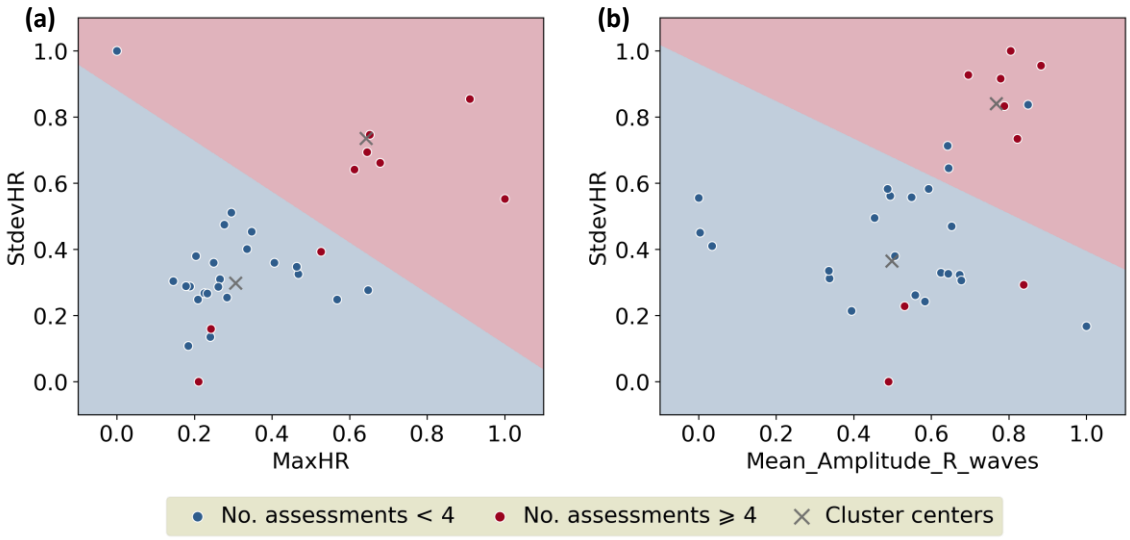

Figure 4.18: Scatter plot representing the two selected features to study the differences between "no. assessments < 4" and "no. assessments ≥ 4". **(a)** For the 5-minutes analysis. **(b)** For the 10-minutes analysis.

4.7.6 Discussion

As final remarks, after the exposition and discussion of the results of the three presented case studies analyzed through unsupervised learning, regarding the three conditions evaluated, the length of the time window of physiological data (5 and 10 minutes), the number of selected features of interest and the clustering algorithm, the following observations can be perceived:

- The dataset comprehend a significantly small amount of examples;
- Two case studies have imbalanced classes, which poses additional challenges to the cluster task;
- For each case study, different features were selected when comparing both time analyses;
- The features selected in the study of pain intensity in case studies A and B were different between them, and also different from the case study C while evaluating the number of pain assessments;
- Case study A: the best overall performance for positive prediction in the study of "no pain" *versus* "pain" was obtained using a 5-minutes time window and only two features (HRV_pNN20 and Mean_Heartbeats_Area), implementing k-means and spectral clustering algorithms (precision=76.0%, recall=95.0%, f1-score=84.4%, and MCC=0.53);
- Case study B: the best result for the study between "tolerable pain" and "intense pain" was obtained in the 5-minutes interval using two features (HRV_SampEn and MaxMinHR) and implementing any of the three clustering algorithms (accuracy=75.0%, precision=83.3%, and recall=62.5%);
- Case study C: the best result was obtained in the 5-minutes analysis with eight features as input to the agglomerative algorithm (precision=92.0%, recall=100.0%, f1-score=95.8%, and MCC=0.85). The eight features are StdevHR, MaxHR, Mean_Heartbeats_Area, Mean_Amplitude_S_waves,

Mean_Diff_R_peaks, Mean_Amplitude_R_waves, ECG_ApproxEn, and HRV_HTI. However, the problem can be described using only two features (StdevHR and MaxHR), with high performance values as well (precision=88.5%, recall=100%, f1-score=93.9%, and MCC=0.77).

Chapter 5
Conclusions and Further Research

This chapter presents some conclusions drawn during this dissertation and the limitations encountered related to the complexity of postoperative pain experience and on the available dataset. Lastly, to finish this chapter, some topics are addressed related to future work that can be performed for further investigation on this topic.

5.1 Discussion

Patients commonly manifest postoperative pain after surgery. It consists of an unpleasant experience that needs to be treated effectively. Considering that the choice of analgesic technique for pain treatment depends on assessing the pain intensity, adequate assessment is mandatory. Besides, the need to develop an objective measure also emerged from the inability of some patients to self-report pain. In these situations, current scales require a third-person perspective from those observing the person experiencing pain and, therefore, produce an even more subjective report of pain. Consequently, in recent years, many tools have been proposed to investigate objective assessment aiming to improve clinical care.

The study conducted in this dissertation investigated the potentiality of using physiological ECG signals as an assessment tool by performing an exploratory work in which physiological alterations evoked by postoperative pain were addressed. The work consisted of univariate and multivariate analyses on how the features extracted from the ECG signals, collected during patients' stay in the recovery room, are related to postoperative pain processes.

The goal of the univariate procedure was to examine the individual behavior of each feature. It was explored the differences of the features within pain intensity to perceive if it is possible to detect pain presence and how ECG features react to pain management by studying the behavior of each feature in detecting alterations after administration of analgesia. Although some features revealed differences in pain presence compared to no pain reports, none of the features exposed pain reduction after treatment. Nonetheless, the individual features may not provide the complete description of the pain response, but a combination of features may improve pain description as the physiological information could be complemented.

The second approach investigated behaviors and patterns of sets of features through clustering and visual analysis. It first aimed to discover feature combinations that could distinguish patients' reports within pain intensities through binary discrimination. After, clustering intended to explore pain prevalence during the immediate postoperative period through a study on how the reports were grouped according to the number of non-zero levels of pain reported by the patients. Although the univariate approach did not achieve satisfactory results, a multivariate approach proved to be promising in studying pain behaviors described by physiological changes of ECG signals. The results suggest that combinations of different features may be helpful and produce improved results, but more research is needed to address this issue.

The goal is to find a reliable and valid measure that correlates and is sensitive and specific to pain. Ideally, this tool would be more accurate and less time-consuming to assess by the healthcare professionals and should allow continuous monitoring and, thus, assist pain management, whether by

providing a measure for those who cannot communicate or alerting when the patient is experiencing pain. One of the identified strategies for this purpose corresponds to physiological parameters that can be collected non-invasively, which was investigated in this work. The concept is that a painful stimulus will induce autonomic or behavioral responses that can be measured and employed as physiological markers of nociception. By continuously monitoring physiological alterations during patients' stay in the recovery room, it would be possible to evaluate the intensity, manage analgesic delivery, and verify the efficacy of the pain treatment to adjust accordingly. Nonetheless, several challenges were encountered.

Firstly, in this work, and in accordance with some of the studies reviewed in this document, patients self-reported the pain they were feeling, and these assessments were used to validate the tools. However, pain corresponds to a conscious response involving psychological and social components to a nociceptive stimulus. Hence, the patient may be neglecting pain when he compares it to more painful experiences in the past, or due to fears about adverse effects or dependency, or even exacerbating it due to expectation. In addition, gender expectations of pain may also play a role in the subjective response. Therefore, profound differences within similar or identical procedures create doubt if the patient is reporting extremely high or low pain improperly. These concerns may compromise the validation through self-reports of the developed tools.

In general, the physiological variables studied as indirect measures of pain are affected by multiple factors. As most correspond to autonomic responses to nociception, sympathetic activators could alter the physiological outcome. These include drugs, nausea, and vomiting. In addition, common psychological activators correspond to stress and anxiety, and the pathological changes resulting from diseases can also affect the physiological response. It leads to uncertainty about whether the changes are related to pain or pharmacological, pathological, and psychological factors, and, ultimately, it could compromise the specificity of these indicators to postoperative pain.

Pain itself is complex due to inter-individual variability. However, during the course of this dissertation, it became evident that postoperative pain is an elaborated process to study. When pain is induced by a stimulus, which intensity and duration can be controlled, in an awake subject, that experiences and reacts to it at the moment, the results are easier to study since the experimental conditions are the same for all participants. However, postoperative pain, being derived from a surgical procedure, is more complex. Firstly, the pain reaction is derived from a past stimulation that caused tissue trauma during which patients are under a state of unconsciousness, and no reflex movement and conscious recall or awareness of the noxious stimulation occur. Secondly, the anesthesia protocol and residual effects may influence the experience at arrival in the recovery room. In addition, the stimulus, that is, the type of surgery, can vary between individuals, which can lead to an increase in variability in the pain experience. Different surgical procedures could cause different pain states and, therefore, produce a procedure-specific pain state. As some surgical procedures are more invasive than others, the painful intensity felt by the patient may be superior.

Moreover, it is essential to acknowledge the limitations of this study and interpret the findings carefully. Several limitations regarding the available dataset to accomplish the work discussed in this document and the nature of physiological measures were encountered. The study comprehends a limited number of participants, with only 19 subjects. Additionally, the number of samples (n=32) is also significantly small, and some of them belong to the same patient. Moreover, the patient's baseline corresponds to a signal segment collected after the postoperative period. Although being stated in this work that the signal segment used as baseline corresponded to a period of minimal pain influence, acquisition of the baseline before the surgical procedure would be a better strategy. Lastly, some measures show high inter-individual differences, not allowing immediate inter-individual comparison of

values as they need to be normalized regarding patient baseline state, which is slightly compromised due to the period of acquisition.

5.2 Future Work

The exploration of physiological alterations during the postoperative period proved to be a great challenge, since postoperative pain is a complex and multidimensional problem in which physiological reaction manifests from autonomic, psychological, and behavioral responses. Therefore, some topics are suggested for future research.

Firstly, combinations of more than one physiological signal may be valuable. The human body and inherent mechanisms make up a complex system, so a multisignal pain description could be more informative about pain experience during the postoperative period. In addition to the ECG signal, EDA and sEMG signals, which have shown promising results on the topic, can also be collected. Hence, it is proposed the study of different signals and a multisignal approach built with parameters from different physiological signals.

Furthermore, the surgical procedure and the anesthesia protocol are variables that change among patients and their impact should be analyzed in future work. In addition, the elimination of the anesthetic drugs by the human body may be altered in patients with extremely high body weight and the remaining effects could mask pain sensation at the beginning of data collection. To study this, since extremes of body weight are commonly evaluated by BMI in clinical settings, the weight and height of patients should be added to the data collection instrument in the following study. Therefore, the effects caused by the variation in the relationship between weight and height (given by BMI) can be explored in the early postoperative period. The proposal is to check for inter-individual differences, by examining if patients with excess adipose tissue have different pain ratings from patients with lower fat tissue at the arrival in the recovery room.

Individual differences in sensory experiences are important in the treatment of pain, and, therefore, the contributors and the magnitude of their influence on inter-individual differences of pain experience could also be investigated. In specific, it is suggested the investigation of age and gender differences in postoperative pain phenomenon.

In addition, an evaluation within the participant itself, and not grouped, can also be performed. It would be interesting to evaluate the differences between a state of no pain and a state of pain within the same patient and discover if differences are detected, and if those differences are more prominent or pronounced in patients who report superior pain levels. Secondly, reports of a non-zero level of pain in which the patient was given a pharmacological pain treatment can be followed up to determine if the pharmacological intervention was effective. That is, assess pain relief by asking the patients to compare pain after medication.

Finally, it is crucial to collect a bigger sample of patients, more representative of age groups and gender, to explore pain variations and pain treatment over the time of the stay of the patients in the recovery room until discharge.

References

[1] E. M. Pogatzki-zahn, D. Segelcke, and S. A. Schug, 'Postoperative pain — from mechanisms to treatment', *PAIN Reports*, vol. 2, 2017.

[2] E. A. Shipton, 'The transition of acute postoperative pain to chronic pain: Part 1 - risk factors for the development of postoperative acute persistent pain', *Trends Anaesth. Crit. Care*, vol. 4, no. 2–3, pp. 67–70, 2014, doi: 10.1016/j.tacc.2014.04.001.

[3] K. Brand and A. Al-rais, 'Pain assessment in children', *Anaesth. Intensive Care Med.*, vol. 20, no. 6, pp. 314–317, 2019, doi: 10.1016/j.mpaic.2019.03.003.

[4] A. H. Kwon and P. Flood, 'Genetics and Gender in Acute Pain and Perioperative Opioid Analgesia', *Anesthesiol. Clin.*, vol. 38, no. 2, pp. 341–355, 2020, doi: 10.1016/j.anclin.2020.01.003.

[5] J. L. Apfelbaum, C. Chen, S. S. Mehta, and T. J. Gan, 'Postoperative pain experience: Results from a national survey suggest postoperative pain continues to be undermanaged', *Anesth. Analg.*, vol. 97, no. 2, pp. 534–540, 2003, doi: 10.1213/01.ANE.0000068822.10113.9E.

[6] D. Segelcke, B. Pradier, and E. Pogatzki-Zahn, 'Advances in assessment of pain behaviors and mechanisms of post-operative pain models', *Curr. Opin. Physiol.*, vol. 11, pp. 85–92, 2019, doi: 10.1016/j.cophys.2019.07.002.

[7] L. G. Maxwell, M. V. Fraga, and C. P. Malavolta, 'Assessment of Pain in the Newborn: An Update', *Clin. Perinatol.*, vol. 46, no. 4, pp. 693–707, 2019, doi: 10.1016/j.clp.2019.08.005.

[8] 'IASP Terminology', *International Association for the Study of Pain*. https://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1698#Pain (accessed Apr. 29, 2021).

[9] R. C. Coghill, 'Individual differences in the subjective experience of pain: New insights into mechanisms and models', *Headache*, vol. 50, no. 9, pp. 1531–1535, 2010, doi: 10.1111/j.1526-4610.2010.01763.x.

[10] T. Ledowski, J. Bromilow, J. Wu, M. J. Paech, H. Storm, and S. A. Schug, 'The assessment of postoperative pain by monitoring skin conductance: Results of a prospective study', *Anaesthesia*, vol. 62, no. 10, pp. 989–993, 2007, doi: 10.1111/j.1365-2044.2007.05191.x.

[11] M. S. Salekin, G. Zamzmi, D. Goldgof, R. Kasturi, T. Ho, and Y. Sun, 'Multimodal spatio-temporal deep learning approach for neonatal postoperative pain assessment', *Comput. Biol. Med.*, vol. 129, 2021, doi: 10.1016/j.compbiomed.2020.104150.

[12] H. Storm, 'Changes in skin conductance as a tool to monitor nociceptive stimulation and pain', *Curr. Opin. Anaesthesiol.*, vol. 21, no. 6, pp. 796–804, 2008, doi: 10.1097/ACO.0b013e3283183fe4.

[13] T. Ledowski, J. Bromilow, M. J. Paech, H. Storm, R. Hacking, and S. A. Schug, 'Monitoring of skin conductance to assess postoperative pain intensity', *Br. J. Anaesth.*, vol. 97, no. 6, pp. 862–865, 2006, doi: 10.1093/bja/ael280.

[14] T. Fischer, A. Hosie, T. Luckett, M. Agar, and J. Phillips, 'Strategies for Pain Assessment in Adult Patients With Delirium: A Scoping Review', *J. Pain Symptom Manage.*, vol. 58, no. 3, pp. 487-502.e11, 2019, doi: 10.1016/j.jpainsymman.2019.05.020.

[15] F. C. Schasfoort, M. A. G. Formanoy, J. B. J. Bussmann, J. W. B. Peters, D. Tibboel, and H. J. Stam, 'Objective and continuous measurement of peripheral motor indicators of pain in hospitalized infants: A feasibility study', *Pain*, vol. 137, no. 2, pp. 323–331, 2008, doi: 10.1016/j.pain.2007.09.011.

[16] J. H. Jang, W. H. Park, H. Kim, and S. O. Chang, 'Ways of Reasoning Used by Nurses in Postoperative Pain Assessment', *Pain Manag. Nurs.*, vol. 21, no. 4, pp. 379–385, 2020.

[17] P. Hummel and M. van Dijk, 'Pain assessment: Current status and challenges', *Semin. Fetal Neonatal Med.*, vol. 11, no. 4, pp. 237–245, 2006, doi: 10.1016/j.siny.2006.02.004.

[18] D. G. da Saúde, 'Circular Normativa nº 9/DGCG', *A Dor como 5º sinal vital. Registo sistemático da intensidade da Dor*. https://www.aped-dor.org/documentos/DGS-dor_como_5_sinal_vital_-_2003.pdf (accessed May 01, 2021).

[19] C. L. Hicks, C. L. Von Baeyer, P. A. Spafford, I. Van Korlaar, and B. Goodenough, 'The Faces Pain Scale - Revised: Toward a common metric in pediatric pain measurement', *Pain*, vol. 93, no. 2, pp. 173–183,

2001, doi: 10.1016/S0304-3959(01)00314-1.

[20] C. E. Steeds, 'The anatomy and physiology of pain', *Surg.*, vol. 27, no. 12, pp. 507–511, 2009, doi: 10.1016/j.mpsur.2009.10.013.

[21] M. Eriksson and M. Campbell-Yeo, 'Assessment of pain in newborn infants', *Semin. Fetal Neonatal Med.*, vol. 24, no. 4, p. 101003, 2019, doi: 10.1016/j.siny.2019.04.003.

[22] C. J. Woolf, 'Pain: Moving from Symptom Control toward Mechanism-Specific Pharmacologic Management', *Ann. Intern. Med.*, vol. 140, no. 6, pp. 441–451, 2004, doi: 10.7326/0003-4819-140-8-200404200-00010.

[23] S. K. Saladin, *Human Anatomy*, 5th ed. McGraw-Hill Education, 2017.

[24] E. L. Garland, 'Pain Processing in the Human Nervous System: A Selective Review of Nociceptive and Biobehavioral Pathways', *Prim. Care Clin. Off. Pract.*, vol. 39, no. 3, pp. 561–571, 2012, doi: 10.1016/j.pop.2012.06.013.

[25] S. A. Armstrong and M. J. Herr, 'Physiology, Nociception'.
https://www.ncbi.nlm.nih.gov/books/NBK551562/ (accessed May 20, 2021).

[26] M. F. Yam, Y. C. Loh, C. S. Tan, S. K. Adam, N. A. Manan, and R. Basir, 'General pathways of pain sensation and the major neurotransmitters involved in pain regulation', *Int. J. Mol. Sci.*, vol. 19, no. 8, 2018, doi: 10.3390/ijms19082164.

[27] C. S. Nielsen, A. Stubhaug, D. D. Price, O. Vassend, N. Czajkowski, and J. R. Harris, 'Individual differences in pain sensitivity: Genetic and environmental contributions', *Pain*, vol. 136, no. 1–2, pp. 21–29, 2008, doi: 10.1016/j.pain.2007.06.008.

[28] M. S. Angst *et al.*, 'Pain Sensitivity and Opioid Analgesia: A Pharmacogenomic Twin Study', *Pain*, vol. 153, no. 7, pp. 1397–1409, 2012, doi: 10.1016/j.pain.2012.02.022.

[29] T. A. Norbury, A. J. MacGregor, J. Urwin, T. D. Spector, and S. B. McMahon, 'Heritability of responses to painful stimuli in women: A classical twin study', *Brain*, vol. 130, no. 11, pp. 3041–3049, 2007, doi: 10.1093/brain/awm233.

[30] J. L. Riley *et al.*, 'Age and race effects on pain sensitivity and modulation among middle-aged and older adults', *J. Pain*, vol. 15, no. 3, pp. 272–282, 2014, doi: 10.1016/j.jpain.2013.10.015.

[31] S. J. Gibson and M. Farrell, 'A review of age differences in the neurophysiology of nociception and the perceptual experience of pain', *Clin. J. Pain*, vol. 20, no. 4, pp. 227–239, 2004, doi: 10.1097/00002508-200407000-00004.

[32] L. Chang, T. Ma, S. Tsay, and G. Jong, 'Relationships between pain intensity and heart rate variability in patients after abdominal surgery: a pilot study', *Chin. Med. J. (Engl).*, vol. 125, pp. 1964–1969, 2012, doi: 10.3760/cma.j.issn.0366-6999.2012.11.023.

[33] E. N. Brown, K. J. Pavone, and M. Naranjo, 'Multimodal General Anesthesia : Theory and Practice', vol. 127, no. 5, pp. 1246–1258, 2018, doi: 10.1213/ANE.0000000000003668.

[34] A. Savoca and D. Manca, *Control strategies in general anesthesia administration*. Elsevier Inc., 2020.

[35] J. W. Bard, 'The BIS monitor: a review and technology assessment', *AANA J.*, vol. 69, no. 6, pp. 477–483, 2001.

[36] H. Jones and K. Rowland-Yeo, 'Basic Concepts in Physiologically Based Pharmacokinetic Modeling in Drug Discovery and Development', *CPT: Pharmacomet. Syst. Pharmacol.*, vol. 2, no. 8, p. e63, 2013.

[37] C. P. Hebbes and J. P. Thompson, 'Pharmacokinetics of anaesthetic drugs at extremes of body weight', *BJA Educ.*, vol. 18, no. 12, pp. 364–370, 2018, doi: 10.1016/j.bjae.2018.09.001.

[38] 'Obesidade'. https://www.sns24.gov.pt/tema/doencas-cronicas/obesidade/ (accessed Apr. 29, 2021).

[39] D. G. da Saúde, 'Orientações técnicas sobre a avaliação da dor nas crianças', 2010. [Online]. Available: https://www.aped-dor.org/images/documentos/avaliacao_dor_criancas/Avaliacao_da_Dor_nas_Criancas.pdf.

[40] H. Breivik *et al.*, 'Assessment of pain', *Br. J. Anaesth.*, vol. 101, no. 1, pp. 17–24, 2008, doi: 10.1093/bja/aen103.

[41] A. Williamson and B. Hoggart, 'Pain: A review of three commonly used pain rating scales', *J. Clin. Nurs.*, vol. 14, no. 7, pp. 798–804, 2005, doi: 10.1111/j.1365-2702.2005.01121.x.

[42] S. Pereira-Morales, C. M. Arroyo-Novoa, A. Wysocki, and L. S. Eller, 'Acute Pain Assessment in Sedated Patients in the Postanesthesia Care Unit', *Clin. J. Pain*, vol. 34, no. 8, pp. 700–706, 2018, doi: 10.1097/AJP.0000000000000593.

[43] P. Severgnini, P. Pelosi, E. Contino, E. Serafinelli, R. Novario, and M. Chiaranda, 'Accuracy of Critical Care Pain Observation Tool and Behavioral Pain Scale to assess pain in critically ill conscious and unconscious patients: Prospective, observational study', *J. Intensive Care*, vol. 4, no. 1, pp. 1–8, 2016, doi: 10.1186/s40560-016-0192-x.

[44] D. Brown, 'Pain assessment with cognitively impaired older people in the acute hospital setting', *Rev. Pain*, vol. 5, no. 3, pp. 18–22, 2011.

[45] R. Cowen, M. K. Stasiowska, H. Laycock, and C. Bantel, 'Assessing pain objectively: The use of physiological markers', *Anaesthesia*, vol. 70, no. 7, pp. 828–847, 2015, doi: 10.1111/anae.13018.

[46] L. Sörnmo and P. Laguna, *Bioeletrical Signal Processing in Cardiac and Neurological Applications*. 2005.

[47] 'Intelligent health monitoring system based on smart clothing - Scientific Figure on ResearchGate'. https://www.researchgate.net/figure/Illustration-of-QRS-complexes-and-RR-interval-of-ECG-signals_fig4_326588784 (accessed Jun. 04, 2021).

[48] E. Kantor, P. Montravers, D. Longrois, and J. Guglielminotti, 'Pain assessment in the postanaesthesia care unit using pupillometry: A cross-sectional study after standard anaesthetic care', *Eur. J. Anaesthesiol.*, vol. 31, no. 2, pp. 91–97, 2014, doi: 10.1097/01.EJA.0000434966.96165.c9.

[49] R. Logier, M. Jeanne, J. De Jonckheere, A. Dassonneville, M. Delecroix, and B. Tavernier, 'PhysioDoloris: A monitoring device for Analgesia / Nociception balance evaluation using Heart Rate Variability analysis', *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Int. Conf.*, pp. 1194–1197, 2010, doi: 10.1109/IEMBS.2010.5625971.

[50] O. Gall *et al.*, 'Postoperative pain assessment in children: a pilot study of the usefulness of the analgesia nociception index', *Br. J. Anaesth.*, vol. 115, no. 6, pp. 890–895, 2015, doi: 10.1093/bja/aev361.

[51] MDoloris Medical Systems, 'Analgesia Nociception Index'. https://mdoloris.com/technologies/adults-and-pediatrics/ (accessed Jun. 09, 2021).

[52] E. Boselli *et al.*, 'Prediction of immediate postoperative pain using the analgesia/nociception index: A prospective observational study', *Br. J. Anaesth.*, vol. 112, no. 4, pp. 715–721, 2014, doi: 10.1093/bja/aet407.

[53] E. Boselli *et al.*, 'Prospective observational study of the non-invasive assessment of immediate postoperative pain using the analgesia/nociception index (ANI)', *Br. J. Anaesth.*, vol. 111, no. 3, pp. 453–459, 2013, doi: 10.1093/bja/aet110.

[54] Z. Visnovcova, M. Mestanik, M. Gala, A. Mestanikova, and I. Tonhajzerova, 'The complexity of electrodermal activity is altered in mental cognitive stressors', *Comput. Biol. Med.*, vol. 79, pp. 123–129, 2016, doi: 10.1016/j.compbiomed.2016.10.014.

[55] C. Kappeler-Setz, F. Gravenhorst, J. Schumm, B. Arnrich, and G. Tröster, 'Towards long term monitoring of electrodermal activity in daily life', *Pers. Ubiquitous Comput.*, vol. 17, no. 2, pp. 261–271, 2013, doi: 10.1007/s00779-011-0463-4.

[56] Y. Hu, C. Converse, M. C. Lyons, and W. H. Hsu, 'Neural control of sweat secretion: a review', *Br. J. Dermatol.*, vol. 178, no. 6, pp. 1246–1256, 2018, doi: 10.1111/bjd.15808.

[57] B. Hullett *et al.*, 'Monitoring electrical skin conductance: a tool for the assessment of postoperative pain in children?', *Pediatr. Anesth.*, vol. 19, no. 5, pp. 556–556, 2009, doi: 10.1111/j.1460-9592.2009.02993_4.x.

[58] S. Chatterjee, A. Rudra, and S. Sengupta, 'Current concepts in the management of postoperative nausea and vomiting', *Anesthesiol. Res. Pract.*, vol. 2011, 2011, doi: 10.1155/2011/748031.

[59] T. Ledowski, J. Preuss, and S. A. Schug, 'The effects of neostigmine and glycopyrrolate on skin conductance as a measure of pain', *Eur. J. Anaesthesiol.*, vol. 26, no. 9, pp. 777–781, 2009, doi: 10.1097/EJA.0b013e32832bb678.

[60] R. Merletti, A. Botter, A. Troiano, E. Merlo, and M. A. Minetto, 'Technology and instrumentation for detection and conditioning of the surface electromyographic signal: State of the art', *Clin. Biomech.*,

vol. 24, no. 2, pp. 122–134, 2009, doi: 10.1016/j.clinbiomech.2008.08.006.

[61] M. Hakonen, H. Piitulainen, and A. Visala, 'Current state of digital signal processing in myoelectric interfaces and related applications', *Biomed. Signal Process. Control*, vol. 18, pp. 334–359, 2015, doi: 10.1016/j.bspc.2015.02.009.

[62] D. Wildemeersch, M. Baeten, N. Peeters, V. Saldien, M. Vercauteren, and G. Hans, 'Pupillary dilation reflex and pupillary pain index evaluation during general anaesthesia: A pilot study', *Rom. J. Anaesth. Intensive Care*, vol. 25, no. 1, pp. 19–23, 2018, doi: 10.21454/rjaic.7518.251.wil.

[63] M. Aissou, A. Snauwaert, C. Dupuis, A. Atchabahian, F. Aubrun, and M. Beaussier, 'Objective Assessment of the Immediate Postoperative Analgesia Using Pupillary Reflex Measurement: A Prospective and Observational Study', *Anesthesiology*, vol. 116, no. 5, pp. 1006–1012, 2012.

[64] H. Storm, 'The Capability of Skin Conductance to Monitor Pain Compared to Other Physiological Pain Assessment Tools in Children and Neonates', *Pediatr. Ther.*, vol. 3, no. 4, 2013, doi: 10.4172/2161-0665.1000168.

[65] R. M. Rangayyan, *Biomedical Signal Analysis*, 2nd ed. Wiley-IEEE Press, 2015.

[66] S. Luo and P. Johnston, 'A review of electrocardiogram filtering', *J. Electrocardiol.*, vol. 43, no. 6, pp. 486–496, 2010, doi: 10.1016/j.jelectrocard.2010.07.007.

[67] A. Appathurai *et al.*, 'A study on ECG signal characterization and practical implementation of some ECG characterization techniques', *Meas. J. Int. Meas. Confed.*, vol. 147, p. 106384, 2019, doi: 10.1016/j.measurement.2019.02.040.

[68] S. Kaplan Berkaya, A. K. Uysal, E. Sora Gunal, S. Ergin, S. Gunal, and M. B. Gulmezoglu, 'A survey on ECG analysis', *Biomed. Signal Process. Control*, vol. 43, pp. 216–235, 2018, doi: 10.1016/j.bspc.2018.03.003.

[69] J. Ribeiro Pinto, J. S. Cardoso, and A. Lourenco, 'Evolution, current challenges, and future possibilities in ECG Biometrics', *IEEE Access*, vol. 6, pp. 34746–34776, 2018, doi: 10.1109/ACCESS.2018.2849870.

[70] A. Fratini, M. Sansone, P. Bifulco, and M. Cesarelli, 'Individual identification via electrocardiogram analysis', *Biomed. Eng. Online*, vol. 14, no. 1, pp. 1–23, 2015, doi: 10.1186/s12938-015-0072-y.

[71] A. Subasi, *Practical Machine Learning for Dara Analysis Using Python*. 2020.

[72] F. Shaffer and J. P. Ginsberg, 'An Overview of Heart Rate Variability Metrics and Norms', *Front. Public Heal.*, vol. 5, p. 258, 2017, doi: 10.3389/fpubh.2017.00258.

[73] R. Castaldo, L. Montesinos, P. Melillo, C. James, and L. Pecchia, 'Ultra-short term HRV features as surrogates of short term HRV: A case study on mental stress detection in real life', *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 1, pp. 1–13, 2019, doi: 10.1186/s12911-019-0742-y.

[74] M. Jiang, R. Mieronkoski, A. M. Rahmani, N. Hagelberg, S. Salantera, and P. Liljeberg, 'Ultra-short-term analysis of heart rate variability for real-time acute pain monitoring with wearable electronics', *2017 IEEE Int. Conf. Bioinforma. Biomed.*, pp. 1025–1032, 2017, doi: 10.1109/BIBM.2017.8217798.

[75] Y. Chen, L. Zhang, B. Zhang, and C. A. Zhan, 'Short-term HRV in young adults for momentary assessment of acute mental stress', *Biomed. Signal Process. Control*, vol. 57, p. 101746, 2020, doi: 10.1016/j.bspc.2019.101746.

[76] J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey, *Graphical Methods for Data Analysis*. CRC Press, 1983.

[77] E. Alpaydin, *Introduction to Machine Learning*, 3rd ed. MIT Press, 2014.

[78] A. C. Müller and S. Guido, *Introduction to Learning Machine with Python*, 1st ed. O'Reilly Media, 2017.

[79] D. Chicco and G. Jurman, 'The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation', *BMC Genomics*, vol. 21, no. 6, pp. 1–13, 2020, doi: 10.1186/s12864-019-6413-7.