Automated Partial Premature Infant Pain Profile Scoring Using Big Data Analytics

By

Tanvi Naik

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

Masters of Health Sciences

in

Health Informatics

The Faculty of Health Sciences

Program

University of Ontario Institute of Technology

April, 2017

Certificate of Approval

Abstract

Lack of valid and reliable pain assessment in the neonatal population has become a significant challenge in the Neonatal Intensive Care Unit (NICU). Current practice in the NICU involves the meticulous, time-consuming and potentially bias process of manual interpretation of pain scores. In an attempt to forego the manual scoring system, this thesis presents an initial framework to automate a partial pain score for newborn infants using big data analytics that automates the analysis of high speed physiological data. The design of the novel Artemis Premature Infant Pain Profile (APIPP) is proposed in this thesis. An ethically approved retrospective clinical research study was performed to calculate APIPP scores from premature infant data collected from the Artemis platform. Using the Premature Infant Pain Profile (PIPP) as the base gold standard scale, scoring techniques were automated to create data abstractions from the physiological streams of Heart Rate (HR) and Oxygen Saturation (SpO_2). These were then brought together to compute an automated partial pain score (APIPP) that was based on gestational age, HR and SpO_2. Through the retrospective clinical research study, and to evaluate the effectiveness and feasibility of automating the scale in the future, APIPP was retrospectively compared with the PIPP which was manually scored by nursing staff at The Hospital for Sick Children, Toronto. Furthermore, the characteristics in HR were also assessed in a thorough manner by preforming statistical tests to assess the resourcefulness of HR as a measure to identify a pain response. Future research will focus on the clinical validation of this work by carrying out prospective research to implement an algorithm based on the design proposed in this thesis that can be integrated into a clinical decision support system named Artemis.

Keywords: *neonate, neonatal pain management, big data analytics, clinical decision support system, premature infant, pain scale, physiological data streams*

Acknowledgements

"Every worthwhile accomplishment, big or little, has its stages of drudgery and triumph: a beginning, a struggle, and a victory" – Mahatma Gandhi

This research work began in 2010, with a genuine meeting of a naive student and a generous professor who was willing to give her a chance at research. When I first read about Dr. Carolyn McGregor's work with premature infants, I was greatly captivated and was eager to have an opportunity to work with her. Six years ago, I was fortunate enough to be granted an opportunity to work with Carolyn and be a part of her team. During these 6 years of working with Carolyn, I have not only grown professionally, but also personally. I would like to thank Carolyn for her support and guidance not just throughout my masters but also for the last 6 years. You recognized my potential early on and provided me with a platform to grow, gain confidence and explore the field of research.

I would also like to thank my co-supervisor, Dr. Andrew James, whose wisdom and knowledge has been a great contribution to my thesis. Your keen review of my work and detailed feedback was incredibly helpful in improving the quality of my thesis. Thank you for guiding me in the right direction and pushing me to always smile despite how difficult circumstances were. I will forever be grateful for all your support and guidance throughout this journey. I would also like to thank Nadja Bressan for her support, guidance and friendship that helped me through some of the most difficult times during my masters' journey. Thank you for encouraging me to keep going and cheering me on every step of the way.

I am most thankful to my family and friends, whose love, patience, encouragement and confidence, has sustained me throughout this journey. Thank you for helping me endure the most difficult times while making the good times all the more enjoyable. Thank you for believing in me, for recognizing my strength, and for supporting me with your love always. Even though it seemed never ending, life after 'thesis' can finally begin! My journey continues and I am confident that I can persevere through anything because I have your support.

Finally, I would like to dedicate this thesis to my late grandfather, who always encouraged me to prioritize and pursue higher education despite any odds. As a man who highly valued education, I know he would have been very proud to see me complete my masters. Wherever he is, I am certain that he will always be cheering me on the loudest for my achievements today and in the future.

I humbly desire that this thesis contributes to the lives of millions of premature infants and their families who are fighting for their child's life every day. I also hope that this thesis encourages others to pursue research in an effort to reduce pain and increase the quality of life in the neonatal population.

Publications Related to this Thesis

Naik, T., Bressan, N., James, A., & McGregor, C. (2013). Design of temporal analysis for a novel premature infant pain profile using Artemis. *Journal of Critical Care*, *28*(1), e4.

Naik, T., Thommandram, A., Fernando, S., Bressan, N., James, A., McGregor, C., (2014), "A method for a real-time novel premature infant pain profile using high rate, high volume physiological data streams", 27th International Symposium on Computer-based Medical Systems (CBMS 2014), New York, USA, pp. 34-37

Naik, T., McGregor, C., (2014) " Design of an Approach for Detection of Pain by Investigating Abnormalities in Physiological Responses in Premature Infants," Paper presented at the Canadian Student Conference on Biomedical Computing and Engineering & IEEE EMBS International Student Conference, 7th CSCBCE IEEE EMBS ISC 2014, Oshawa, Ontario.

Table of Contents

List of Figures

List of Tables

Chapter 1 - Introduction

This thesis presents an initial framework to automate a pain score for newborn infants using big data analytics. This research will also present an approach to integrate and use physiological variables to identify pain in an objective manner. Management of pain in the neonatal population is one of the most challenging problems in the Neonatal Intensive Care Unit (NICU). This research is motivated by a lack of adequate, validated pain assessment tools to assess neonatal pain. There is also a significant need for validated Clinical Decision Support Systems (CDSS) that can provide automated generation of pain scores in the neonatal population. In this thesis, an attempt will be made to design a CDSS for the automated real-time creation of a partial Premature Infant Pain Profile (PIPP) score. The elements of the PIPP score used in this design will be those that can be calculated without the need for a visual assessment of the patient at the bedside. Many gaps exist in the knowledge relating to the optimal utilization and accessibility of techniques that assess pain in a valid and reliable manner. The field of informatics has great potential for designing physiological based scales for the neonatal population, where such scales can utilize CDSS for the continuous assessment of pain.

1.1 Pain in Neonatal Population

Pain management in the neonatal population is a significant challenge within the health care community. During their stay at the Neonatal Intensive Care Unit (NICU), preterm infants are exposed to a high number of painful procedures. Specifically, infants undergo a range of 2 to 14 invasive procedures each day, for which less than one-third receive an analgesic therapy (Ranger, Johnston, & Anand, 2007). The most common procedures, heel lance or venipuncture, are used to obtain blood for screening and medical monitoring. These particular procedures are painful for infants, where such noxious stimulation can cause changes in brain activity that is

similar to that of adults (Harrison et al., 2015). The International Association for the Study of Pain defines pain as “an unpleasant sensory and emotional experience associated with actual or potential tissue damage or described in terms of such damage” ("Prevention and management of pain and stress in the neonate. ," 2000) Another definition of pain that is widely used or understood within the health care community states that “pain is whatever the experiencing person says it is, existing whenever he/she says it does” (Worley, Fabrizi, Boyd, & Slater, 2012). Until the 1990s, newborns in some clinical centers were undergoing surgery with minimal anaesthesia due to the common assumption that neonates could not perceive pain early in life. Despite the pain associated with certain surgical procedures, such as lumbar punctures or circumcisions, newborns received little or no pain management postoperatively (Slater, Cantarella, Franck, Meek, & Fitzgerald, 2008).

Since then, clinicians have become more aware of the fact that pain may be experienced from the earliest stages of postnatal life (Slater et al., 2008). Additional clinical evidence has shown further support, indicating that exposure to prolonged and repetitive pain-related stress in infants born very preterm can have potential long-term effects that can lead to altered neurobehavioral development in vulnerable infants (Anand, Palmer, & Papanicolaou, 2013; Brummelte et al., 2012; Doesburg et al., 2013; Grunau, 2013). For example, specific brain connections form during early stages of development, where such connections play a role in experiencing pain during this early stage of life. In one study, Lowery provides details regarding early brain development. He states that connections to the thalamus begin at 14 gestational weeks and are completed by 20 gestational weeks. Additionally, thalamocortical connections are present from 13 gestational weeks and undergo further development until 26 to 30 gestational weeks (Lowery et al., 2007). Additional extensive research has shown that neurons of the

cerebral cortex begin a migration from the periventricular zone at 8 weeks of gestation. At 20 gestational weeks, the cortex has acquired a full complement of neurons with glial perforation active throughout childhood (Lowery et al., 2007). Synaptic formation begins at 12 gestational weeks, accelerates rapidly during the last trimester of pregnancy, and peaks during the first year of life (Stiles & Jernigan, 2010). Electroencephalographic activity appears for the first time at 20 weeks of gestation, becomes synchronized at 26 weeks of gestation, and shows sleep/wake cycling at 30 weeks of gestation (Lowery et al., 2007). Due to such rapid changes in brain structure, there is growing concern that repeated pain in vulnerable neonates may result in various long-term consequences, including emotional, behavioural and learning disabilities (Anand & Scalzo, 2000; Bhutta & Anand, 2002; "Prevention and Management of Pain in the Neonate: An Update," 2006). These findings confirm that premature infants are able to indeed experience pain. Additional evidence involving animal models and humans reveal that early pain experiences can alter subsequent central nervous system (CNS) function (Fitzgerald, 2005). Despite these revelations, a substantial challenge remains with precisely quantifying the amount of pain that premature infants experience. By having the ability to communicate the severity of pain in a clinical setting, individuals are able to seek strategies to ease the pain by using analgesics or alternative interventions. However, the inability to effectively communicate one's distress of pain causes individuals, specifically premature infants, to be vulnerable to prolonged suffering.

Pain that is ignored and not treated can have immediate and long-term effects due to structural and physiological changes within the nervous system (Slater et al., 2008). For example, the body tends to respond to untreated pain by increasing the release of stress hormones. If the pain is left untreated, such a change can result in increased morbidity and

mortality (Slater et al., 2008). Studies have also shown that improper pain management, or unnoticed and untreated pain from an early age could potentially lead to long-term effects due to the developmental plasticity of the immature brain (Grunau, 2013). These effects may include altered pain perception, chronic pain syndromes, and somatic complaints such as sleep disturbances, feeding problems, and inability to self-regulate in response to internal and external stressors (Slater et al., 2008). A study performed by Anand et al. with neonatal rats also found that exposure to repetitive neonatal pain can result in decreased pain thresholds (Anand, Coskun, Thrivikraman, Nemeroff, & Plotsky, 1999). Furthermore, attention deficit disorders, learning disorders, and behavioural problems in later childhood may be linked to repetitive pain in the preterm infant (Slater et al., 2008). Due to such serious outcomes, the application of optimal tools and specific techniques for quantifying pain becomes vital in a vulnerable population like that of premature infants.

1.2 Pain Detection Tools

Initially, pain detection tools were designed for use in research settings and were univariate in their approach. These tools highly focused on behavioural changes such as crying and body movements. Multiple factors limit or prevent the demonstration of such behavioural changes in premature infants. For this reason, certain bio-physiological assessment clues were added to current behavioural tools in an effort to improve sensitivity. However, the incorporation of these bio-physiological clues in current scales has remained minimal. Since the 1980s, a variety of physiologic parameters have served to estimate pain intensity of acute painful procedures (Van Dijk & Tibboel, 2012). Even though more than 40 pain assessment tools are available, no single instrument has demonstrated superiority over the others for use in this vulnerable population (Ranger et al., 2007). Heart rate and blood pressure are often included in multidimensional

scales. Some scales involve the comparison of heart rate and blood pressure against their baseline value. For instance, an increase of more than 20% from baseline (N-PASS scale) or a decrease of heart rate of 10 beats per minute (MAPS) is used as an indication of pain (Van Dijk & Tibboel, 2012). Automated analysis of the various physiological changes described in Chapter 2 proves that validated automated systems can be created for measuring pain in the neonatal population.

The automatic analysis of pain has received greater attention over the last few years due to the growth in health informatics. Many research studies have outlined the need for better pain assessment tools (Gibbins & Stevens, 2001; Korhonen, Haho, & Polkki, 2013; "Prevention and Management of Pain in the Neonate: An Update," 2006; Slater et al., 2008; Van Dijk & Tibboel, 2012). Studies have shown that periodic monitoring of patient pain levels by physicians and nurses can lead to large improvements (Rudovic, Pavlovic, & Pantic, 2013). However, healthcare providers experience an increased burden of work and stress; thus such pain monitoring becomes difficult to sustain. For this reason, an automated system would be an ideal solution. To date, there has been very limited health information technology research in the neonatal population, which then presents a good opportunity for innovation. A recent study verified the advantages of objective pain assessment methods over the currently used subjective pain rating tools (Tejman-Yarden et al., 2016). Although there is vast potential for the use of computer systems for pain assessment in the neonatal population, there are very few studies in the computer systems literature addressing this particular issue. As such, there is a need for computerized tools to support pain management in neonates. However, there is a lack of empirical studies that provide guidance on how to design such tools. Another major drawback is the fact that most of the studies have been implemented in the adult population. Thus,

implementing specific concepts from these adult studies in the neonatal population is highly valuable.

1.3 Information Systems

Introducing information systems into healthcare has proven to be a challenge due to the complexities of both technology and healthcare. The neonatal space is particularly susceptible to information systems failure given that neonates are a very sensitive population, where each patient is individually different. In order for information systems to succeed in such a vulnerable population, the interconnection of the medical, technical and social contexts of healthcare within the design of these systems is highly important. Effective design starts with a solid conceptual model. The concept for these computer-based systems should be interdisciplinary and should be conceived by physicians as well as other involved healthcare practitioners to enhance the assessment, diagnosis and management of severe pain in such a vulnerable population.

Even though verbal methods (i.e., pain scales, questionnaires) and visual analogue scales are commonly used for measuring clinical pain, such tools tend to lack in reliability or validity when applied to non-verbal patients (Herr, Bjoro, & Decker, 2006). Numerous experimental studies have been conducted in such patients; however, the systems described in such studies have not been validated for use in all clinical settings. Some behavioural tools that measure facial expressions, vocalizations and body movements also exist; however, such tools may also not be entirely accurate as neonates are at times highly sedated or severely premature to provide any behavioural distress cues. Therefore, there is a need to develop a pain assessment tool that is based on physiology, and requires no communication on the part of patient. Researchers have long sought to develop a physiology-based pain assessment that does not depend on patient volitional behaviours. However, minimal advances have been performed in this area. While

several physiological variables have shown statistically significant correlations with the presence of pain or with pain intensity, no measure has provided a sufficiently high relationship with pain to be used as a valid surrogate for self-reports. Therefore, despite many years of research, there is currently no accepted technique for the physiologic assessment of pain in neonates.

1.4 Research Motivations

The challenge has been to bridge the gap between research and clinical practice by devising a pain assessment method for premature infants that is suitable in all circumstances and conditions. Despite the fact that the scientific community has disregarded misconceptions relating to pain in the neonatal population, there remains a lack of a 'gold standard' to measure pain precisely. Current pain measurement scales that are used in the NICU are highly based on behavioural indicators. The most commonly used scales are listed in Table 1. Behavioural indicators, such as facial expression, movement, and brow bulge, tend to be inadequate as the premature infants are heavily sedated or are not able to move due to the immaturity of their nervous system. Behavioural measurements are widely used for infants and nonverbal subjects of all ages (Berde & McGrath, 2009). Such measurements are sensitive to fear, anxiety as well as pain, resulting in an underestimation of pain intensity as compared to self-report measures in patients with persistent pain (Berde & McGrath, 2009). Also, some infants have limited ability to behaviorally express pain due to specific disorders, underdevelopment as a result of prematurity as well as physical exertion (Evans, McCartney, Lawhon, & Galloway, 2005). Assessing such behavioural indicators pose challenges for clinical practice and research due to subjectivity; as such, these indicators can be perceived differently by each health care professional. Additionally, as Table 1 presents, most pain scales do not adjust for gestational age when scoring. Because neonates tend to react differently to pain at varying gestational ages, accounting for the gestational age is

important, otherwise inaccuracy and misinterpretation can occur. In a study done by Johnston and Stevens, the authors compared the responses of newly-born infants at 32-week gestation with those who were currently 32-week gestational age (infants born 4 weeks earlier at 28-week gestation). When compared to the newly born 32-week gestation infant responses to heel sticks, the earlier born (28-week gestation) infants had significantly greater heart rate, significantly lower oxygen saturation, and fewer upper facial expressions of pain (Johnston & Stevens, 1996). These changes can categorically alter pain scores. For this reason, designing a pain scoring system that adjusts for gestational age becomes critical.

Table 1. Commonly used pain assessment tools

Assessment Tool	Physiologic Indicators	Behavioural Indicators	Gestational Age (GA) Tested	Scoring Adjusts for GA
Premature Infant Pain Profile (PIPP)	Heart rate & oxygen saturation	Brow bulge, eye squeeze, nasolabial furrow	28-40 week	Yes
CRIES	Heart rate & Oxygen saturation	Crying, Facial expression, Sleeplessness	32-36 week	No
Neonatal Infant Pain Scale (NIPS)	Respiratory patterns	Facial expression, cry, movement of arms and legs, state arousal	28-3 week	No
Bernese Pain Scale for Neonates (BPSN)	Heart rate, Respiratory rate, Blood pressure, Oxygen saturation	Facial expression, body posture, movements, vigilance	Term and preterm neonates	No
Neonatal Facing Coding System (NFCS)	None	Facial muscle group movement	Preterm and term neonates, infants at 4 months of age	No

The fundamental limitations in pain assessment for neonates stem from subjective assessment criteria, rather than quantifiable and measurable data, resulting in poor quality and inconsistent treatment of patient pain management. Pain is commonly known to be a subjective experience for which the gold standard of measurement is self-report (Brown, Chatterjee, Younger, & Mackey, 2011). Even though self-reported pain is clinically useful and proves to be an effective assessment approach in most situations, self-reported pain can fail for certain vulnerable populations such as neonates. Pain is an individual sensation that is difficult to interpret without any communication from the patient. Neonates, for example, are not able to provide self-reports of pain. Thus, the use of objective measures of pain assessment is imperative.

Health care professionals can also be strongly biased towards assessing these behavioural changes, resulting in inaccuracy. By using objective signs of subjective change, accuracy of pain assessments can significantly improve; however, these signs are not currently used in a precise manner. Patients would benefit from pain assessments that are performed as frequent as heart rate, temperature, respiratory rate, and blood pressure measurements. Evidently, we can use these objective physiological indicators to carefully quantify abnormalities or pain in premature infants with addition of contextual impression of the neonate as assessed by nurses.

It is difficult to implement guidelines for such a complex problem. Pain assessment guidelines today describe the preferred pain assessment tool and criteria with respect to the frequency of scoring (Van Dijk & Tibboel, 2012). However, the availability of such guidelines does not necessarily translate into proper usage by the hospital staff. A survey done among 272 paediatric nurses elicited potential barriers to optimal pain management (Czarnecki et al., 2011). Some of these included insufficient physician orders, insufficient time to pre-medicate patients before procedures and low priority given to pain management by physicians (Van Dijk &

Tibboel, 2012). This thesis proposes that by greater use of physiological indicators and more frequent assessment through automated pain assessment tools, these gaps can be eliminated. Currently, there is insufficient research relating to using physiological indicators to measure pain and its validity. In the future, there is scope for using this thesis work to create rule based-automated systems for physiological indicators, which can potentially remove many of the above-mentioned hurdles.

1.5 Research Aims and Objectives

The primary aim of this research is to design an initial framework to automate a partial pain score for newborn infants using big data analytics that automates the analysis of high speed physiological data. This study will also assess characteristics in heart rate to explore the resourcefulness of heart rate as a measure to assess pain in newborn infants. The objective is to use physiological data for creating an alternate way to generate a score, that is currently generated manually.

1.6 Research Questions

1) Can a scoring system be designed using real-time big data analytics to quantify pain in the neonatal population?
2) Can a real-time algorithm be designed using the knowledge presented in this thesis?
3) Is more frequent monitoring of pain possible by creating an automated system by using the Artemis platform?

1.7 Research Hypothesis

This research study will critically examine the following three hypotheses:

1) A scoring system can be designed for the elements of the score that do not require visual observation at bedside using real-time big data analytic techniques to quantify pain in the neonatal population.

2) A real-time algorithm can be designed using the knowledge presented in this thesis.
3) More frequent monitoring of pain is possible by creating an automated system by using the Artemis platform.

1.8 Thesis Structure

The thesis is structured as follows: Chapter 1 introduced the vital themes of this thesis including an overview on pain in the neonatal population, and the tools used to measure pain with the use of information systems. Chapter 1 also presented the research motivations as well as the research questions and hypothesis. Chapter 2 consists of the literature review, which outlines the background on the variety of machine learning/automated systems used in the past as well as a review on using physiological variables as indicators of pain. Chapter 3 presents a detailed description of the Artemis platform and how this research will use this platform to create a partial PIPP score. Chapter 4 presents two phases of the methodology: the data preparation phase and the data model phase. Chapter 5 displays in detail how the data model can be run to create an automated partial PIPP score using Artemis. Chapter 6 explores the physiological and clinical side of pain and outlines in detail the usefulness of heart rate as a physiological marker for the measurement of pain by presenting various experiments conducted with retrospective data. Chapter 7 will provide a discussion relating to the experiments completed. Finally, Chapter 8 concludes the thesis with identifying limitations of this research work as well as outlining areas for future research and addressing the feasibility of implementing this design in an NICU in the future.

Chapter 2 - Literature Review

This literature review chapter discusses the various studies conducted in the area of pain management to measure pain. The literature review begins with a discussion on how health information systems can be used to assess pain. Following is a review on how physiological parameters have been used to measure pain. It was important to review previous studies that were conducted using physiological changes to assess pain in order to create a new scale that provides more frequent monitoring. Lastly, a review was conducted on various computerized/automated pain measurement systems that have been validated till date to better understand the need for these systems to evaluate pain in the neonatal population.

2.1 Pain and Health Information Systems

Information is the critical resource around which health care organizations are designed. The utility of these information resources is what is of importance to alter the way pain is assessed. It is unfortunate that despite having an array of information resources within health care organizations, patient pain information continues to be documented in an analog format using paper and pen. It is important that we invest in and develop computerized automated systems for patient care. If this is implemented, the system can work as a safeguard for not only health care professionals but also patients. Furthermore, today's governments and health care organizations also have a strong need for automatic tools such as clinical decision support systems which can combine advances in information technology (IT) systems to help reduce the health service costs caused by patient safety incidents (Kong et al., 2012).

Computer-based automated systems or information tools are specific information technology (IT) applications that supports the needs of different types of clinicians and health care professionals. Introducing such applications and decision support systems into a health care

setting can enhance clinical tasks by providing access to information and communication technologies. It is important to turn to informatics-based solutions as healthcare is looking for new and innovative ways to deliver more services with fewer resources. However, the means by which information systems are both designed and implemented will impact how successful the system will be at enhancing care delivery. A key component to successful information system design is the methodological accuracy by which design requirements are collected and applied.

Developing computer based automated systems and informatics tools has been a challenge because despite a number of different systems design approaches, many health care IT projects end up being problematic after implementation (Kuziemsky, Weber-Jahnke, Lau, & Downing, 2008). To avoid this in the neonatal population, it is important to analyze the needs of these systems and to use an interdisciplinary approach to design the system to avoid failure. Currently, very few health information systems are used in the neonatal pain domain. Many health care organizations are still continuing to use paper based scales to measure pain in the premature infant. Additionally, The Hospital for Sick Children, which is the clinical institution used in this study, uses the Premature Infant Pain Profile (PIPP), a paper based scale to quantify pain. Hence, there is vast potential in automating pain scales in the neonatal population through big data analytics in real-time.

Since neonatal care tends to be very fragile, it is particularly important that the needs of these vulnerable patients and in many cases their families are represented in the new information systems that are applied to this domain. Furthermore, it is important to also ensure that these systems provide utility for its users such as physicians and nurses. Therefore, information system designed to enhance neonatal care requires understanding of multiple practices and settings to avoid failure or inaccuracy.

Neonatal pain management should involve a team or holistic based approach involving physicians, nurses and counsellors to ensure the system design is applicable to all areas. It is also important to consider that in healthcare, the users may not only be the health care professionals using the system but also the patients and family members whose care is being managed. Since neonatal care tends to be very sensitive, it is particularly important that the needs of these vulnerable patients and in many cases their families are represented in the new information systems that are applied to this domain.

Pain is an important indicator of medical conditions and disease processes. Health professionals are responsible for diagnosing pain, determining when pain management is necessary, and developing treatment plans. To accomplish these tasks, health professionals employ a variety of assessment tools for evaluating patient self-reports. In the case of preverbal children, methods have been devised to help them communicate their pain experiences. These children can indicate their pain levels, for example, by pointing to drawings of faces that express increasing levels of discomfort. Neonatal pain assessment, in contrast, depends exclusively on the judgment of other health care professionals. A growing body of evidence suggests that failing to diagnose and alleviate pain in newborns can have devastating and long-term effects (Brahnam, Nanni, & Sexton, 2007). A 2003 study listed several effects including immediate effects like irritability, fear, and sleep disturbances, short term effects such as diminished immune system, and long term effects like ongoing memory of the pain and also developmental delays (Mathew & Mathew, 2003). Multiple studies have shown that neonates who are given adequate pain relief consistently tend to exhibit better health outcomes.

The majority of pain assessment instruments developed for newborns incorporate observations of facial activity. Even though facial activity is easier to decipher then physiological

measures and other behavioural indicators such as crying, instruments that have relied on facial information have proven unsatisfactory primarily because of problems with observer bias. One way to reduce bias is to incorporate evaluations that have not been made by an observer. Several researchers have begun investigating machine assessment of common pain indicators. Lindh et al. for instance, have reported some success detecting pain as it relates to heart rate variability, and Petroni et al. have trained neural networks to discriminate differences in neonatal cries, including a cry in response to pain (Lindh, Wiklund, & Hakansson, 1999; Petroni, Malowany, Johnston, & Stevens, 1995). Various different health informatics tools such as these will be discussed in section 2.3 of this literature review chapter.

2.2 Physiological Variables as Indicators to Measure Pain

It is vital for researchers to bring about a change in the practices of detecting pain in neonates. Since the scales in place currently are not fully equipped to provide most accurate results, it is important to redesign the approach. One of the ways to do this will be by using the constantly changing physiological variables. To understand a neonate's body condition and to discover medical problems, caregivers continuously monitor physiological parameters such as heart rate (HR), respiratory rate (RR), blood oxygen saturation (SpO_2), and the blood pressure. The most used physiological parameter in the domain of neonatal pain is that of heart rate. Various studies have been conducted to prove the usefulness of this measure. Changes in heart rate are widely used as markers of reactivity to a painful event in preterm and term infants. Characteristic increases in heart rate following a painful event can be readily identified because heart rate data is easy and relatively inexpensive to acquire. For this reason, heart rate signal is often considered a useful measure of pain reactivity in clinical settings where distress signals are frequently nonspecific and ambiguous. Even the most premature infants have the capacity to increase their

heart rate in response to a painful or distressing event, which reflects generalized central nervous system (CNS) arousal and in particular sympathetic nervous system activation (Oberlander & Saul, 2002). Systems that tend to control cardiovascular functions are closely linked to mechanisms that modulate pain reactivity, thereby, making heart rate responses a potentially useful physiological index of reactivity to noxious events in infants (Oberlander & Saul, 2002). In one particular research study, the heart rate variability (HRV) was investigated for a group of infants (age >34 gestational weeks) with chronic pain (Faye et al., 2010). The results showed that chronic pain is associated with an increase in HR, decrease in RR, and significant decrease in HRV.

It is important to also examine the role of the homeostatic physiological systems that play a role in the mechanism of pain. For the cardiovascular system, this represents a continuous feedback system between the CNS, the autonomic nervous system, and peripheral components. These then maintain mean values of blood pressure and central venous volume within a narrow range, reflected in vascular tone and heart rate (Oberlander & Saul, 2002). Like all homeostatic functions, the greater increases and decreases of heart rate are thought to represent healthier individuals. The greater the organized patterns of rhythmic physiologic signals such as heart rate, the greater the capacity of the individual may have to respond to changing environmental demands (Oberlander & Saul, 2002).

Premature infants have significant difficulties with changing environments, which is why they are admitted to the Neonatal Intensive Care Unit (NICU), where the environment is suitable for their optimal growth and health. Heart rate is a common expression of multiple physiologic processes that reflect CNS function, autonomic control mechanisms, metabolic activity, thoracic hemodynamics, and cardiac chemoreceptors and baroreceptors, as well as levels of arousal and

levels of activity (Oberlander & Saul, 2002). In this sense, the heart rate signal can be used as a parameter that links physiologic capacities with psychological functions, providing a means to investigate stress reactivity, clinical risk, and developmental processes during childhood (Oberlander & Saul, 2002). In infants, resting mean heart rate is typically between 120 and 160 beats per minute (bpm). Heart rate is rarely constant, and its variability reflects continuous activity and interactions between a variety of central and peripheral control systems (Oberlander & Saul, 2002). Heart rate is considered an objective and easily quantifiable measure; however, its specificity as a measure of pain reactivity in premature infants has not been investigated in great detail. Overall, mainly research has found increases in mean heart rate following a noxious event, smaller heart rate changes observed in the presence of analgesics, and differences in responses between painful and non-painful conditions. These claims can help to support the use of heart rate as a physiologic index of pain for future in-depth studies on pain responses in the premature infant population.

The systems modulating the perception of pain are coupled closely with the cardiovascular system. Therefore, the most common physiological pain responses usually include those that are coupled with the stress response. These responses include increases in heart rate, respiratory rate, blood pressure, intracranial pressure, palmar sweating; and decreases in vagal tone, heart rate variability, oxygen saturation, carbon dioxide levels, and peripheral blood flow (Gibbins & Stevens, 2001).

Neonates usually tend to show increase in heart rate and decrease in heart rate variability (HRV) when exposed to a pain stimulus. HRV refers to the beat-to beat alterations in heart rate. During a study done on heel lancing, it was shown that the heart rate increased and the total HRV and the spectral power of the low frequency band decreased in preterm infants during heel

lancing and squeezing of the heel (Lindh, Wiklund, Sandman, & Hakansson, 1997). The frequency domain analysis of the HRV also discriminated a non-painful provocation of the flexor response from the noxious heel lancing and squeezing. Thus, analysis of HRV may offer a possibility of grading the level of distress caused by pain (Lindh et al., 1999). This can prove to be a useful tool in determining the intensity of pain from mild to severe. The HRV study done by (Faye et al., 2010) also showed that High Frequency Variability Index (HFVI) was able to predict the pain with sensitivity of 90% and a specificity of 75%. The background research by Stevens and Johnston in 1994 (Stevens & Johnston, 1994) also showed that the heel prick was able to give rise to an increased heart rate and decreased oxygen saturation in preterm infants.

Besides heel lances, another routine procedure for this population is that of immunizations. Johnston and Strada in 1986 stated that they noticed an initial drop in heart rate followed by a sharp increase during routine immunizations(Johnston & Strada, 1986). In adults usually when they experience intense stimuli, which they perceive as danger, it provokes a defense response with an accelerative heart rate reaction because of sympathetic activation. Infants on the other hand, react to sudden noise or painful stimuli with fear paralysis reflex, which can be characterized by temporary sympathetic inhibition and bradycardia (Padhye, Williams, Khattak, & Lasky, 2009). Thus, in infants, pain and stress can cause severe problems. The physiological changes associated with repeated painful procedures in new born infants should encourage clinicians to reduce the number of stressful events during the neonatal intensive care.

Heart rate variability as mentioned earlier is an important method that can provide insight into the interplay between sympathetic and parasympathetic activity. The high frequency band reflects respiratory-related activity almost entirely mediated by vagal tone (Marek, 1996). The

low frequency component is in some studies considered to reflect both sympathetic and parasympathetic activity and in others exclusively sympathetic modulation (Marek, 1996). In the study performed by Lindh in 1999, they recruited 25 term newborn infants on the maternity ward (Lindh et al., 1999). When they compared baseline levels, it was found that heel lancing triggered an increased total heart rate variability and power in the low frequency band, in contrast to the decrease during squeezing sequences (Lindh et al., 1999).

Blood pressure is one of the vital signals used to detect a wide range of abnormalities. Even though, in typical measurements of the blood pressure, many clinical processes are limited to measurement of only systolic and diastolic pressures—as opposed to the entire signal. However, when the entire blood pressure signal is collected, much more information can be extracted from the data (Najarian & Splinter, 2012). The analysis of the signal is done in six frequencies; they are then calculated and analyzed. These frequencies are the pulse rate (harmonics). These frequencies are ideal because most of the energy of the blood pressure signal is contained in these harmonics. Thus, the relative strength or weakness of these harmonics is often associated with certain abnormalities, which in this case can be pain (Najarian & Splinter, 2012). Increase in heart rate in preterm and term infants undergoing circumcision appears with a marked increase in blood pressure as well; which reflects increased sympathetic arousal (Oberlander & Saul, 2002). Diseases that are not directly related to the heart are also detected or diagnosed at least partially using the blood pressure signal (Najarian & Splinter, 2012).

Analgesics also play an important role as confounding factors in measuring the above-mentioned physiological indicators. The confounding effects of opioid analgesics may also influence the heart rate or heart rate variability response to a painful stimulus. In preterm and term infants, opioids clearly reduce heart rate responses to noxious events. In neonates

undergoing surgery, opioids result in decreased cardiovascular responses to surgical stress (Oberlander & Saul, 2002). Studies have shown that analgesics administered to term infants, decreased physiologic disturbances (increased heart rate, blood pressure, and intracranial pressure) and improved overall clinical outcomes including decreased incidences of sepsis (infection), metabolic acidosis, hyperglycaemia, and clotting (Gibbins & Stevens, 2001). Preterm infants who did not receive analgesia for noxious stimuli had significantly more physiologic responses and were more likely to develop intracranial haemorrhages or other complications in comparison to infants who did receive analgesia (Gibbins & Stevens, 2001).

2.3 Computerized Tools to Measure Pain

The development of new information based computerized systems and clinical parameter prediction tools in the recent past have opened up a new outlook to many areas in healthcare. Through time, the historical development of automated machine learning algorithms and its applications for medical diagnosis has led to advanced and sophisticated data analysis today. Machine learning is a subfield of artificial intelligence. Artificial intelligence is a part of computer science that tries to make computers more intelligent. One of the basic requirements for any intelligent behaviour is learning. Therefore, machine learning is one of the major branches of artificial intelligence and is rapidly growing. Machine learning algorithms were from the very beginning designed and used to analyze medical datasets (Kononenko, 2001). Today, machine learning is able to provide various vital automated tools for intelligent data analysis in the medical community. In the future, intelligent data analysis will play an even more important role due to the large amount of information produced and stored by modern technology. Current machine learning algorithms provide tools that can significantly help medical practitioners to reveal interesting relationships in their data (Kononenko, 2001). Machine learning based diagnostic instruments will be used by physicians and health care

professionals as any other tool to aid in diagnosis. These tools will be a source of possibly useful information that helps to improve diagnostic accuracy. The final responsibility and judgement whether to accept or reject this information can still remain with the physician. There have been various automated techniques that have been experimented in the pain domain.

Today, modern hospitals are well equipped with monitoring and other data collection devices, and data is gathered and shared in large information systems. Machine learning technology is currently well suited for analyzing medical data, and in particular there is a lot of work done in medical diagnosis in small specialised diagnostic problems. Even though the devices to collect data are present, it is important to utilize these resources in the right manner to make better clinical decision support systems for clinicians to solve problems such as inadequate pain management. For the functionality of these systems, the medical diagnostic knowledge can be automatically derived from the description of cases solved in the past. The derived classifier can then be used either to assist the physician when diagnosing new patients in order to improve the diagnostic speed, accuracy and/or reliability (Kononenko, 2001).

For a machine learning system to be useful in solving medical diagnostic tasks, the following features should be present: good performance, the ability to appropriately deal with missing data and with noisy data (errors in data), the transparency of diagnostic knowledge, the ability to explain decisions, and the ability of the algorithm to reduce the number of tests necessary to obtain reliable diagnosis (Kononenko, 2001).

To advance the development of a physiology-based pain measure, neuroimaging methods have been applied to pain management. It can be beneficial to incorporate machine-learning techniques, and to investigate the complex interplay of brain regions in mediating the experience

of pain (Brown et al., 2011). The use of functional MRI (fMRI) in detecting the presence of pain may be strengthened by incorporating machine learning algorithms. Machine learning algorithms such as for example a support vector machine (SVM), can allow predictive models to be trained with a known set of stimuli when detecting pain. Support vector machines and related machine learning algorithms are versatile tools that can learn complex relationships between multiple inputs. Therefore, these systems and tools are very well suited for integrating into them various factors to make classification. These are more accurate than what would result from the investigation of one data source in isolation (Brown et al., 2011).

Marquand and colleagues were the first to apply these neuroimaging fMRI techniques and machine learning algorithms to the area of pain measurement (Marquand et al., 2010). In their study, healthy individuals were exposed to thermal stimuli presented at heat perception threshold, pain perception threshold, and pain tolerance. Machine learning algorithms were trained on fMRI data and used to predict self-reported pain for each participant individually (Marquand et al., 2010). This study provided an important advancement in pain measurement, demonstrating that machine learning algorithms could be used to assess an individual's pain, if trained using fMRI data from that same individual.

To extend this work of Marquand et al., it would be useful to demonstrate that physiology-based pain assessment, using fMRI data and machine learning algorithms, can classify pain accurately without relying on self-report data from the individual that is being tested. If, for example, a SVM model could be trained on one set of neonates, and used to accurately classify pain in different premature infants, then its performance would not depend on the test neonates self-report. Therefore, while there may be considerable individual differences in the experience of pain and in patterns of brain activity induced by pain, there are nonetheless a

core set of pain-induced responses in the brain that may prove to be universal (Brown et al., 2011).

Objective pain measures are observational instruments that can be categorized as unidimensional or multidimensional. Researchers believe that a multidimensional objective measure aids better results as it evaluates two or more pain dimensions (e.g. behaviours and physiologic responses) and it has several domains within each dimension. In a comprehensive review of neonatal/paediatric objective pain measures, the authors concluded that multidimensional measures were more useful clinically and that no single domain was reliable or valid when used alone (Li, Puntillo, & Miaskowski, 2008). The goal in utilizing multidimensional measures would be to yield accuracy levels as close to 100% as possible.

2.3.1 Ontology

Ontologies are a very popular field in the pain domain. Ontological engineering, dealing with developing and using ontology, has become an important research focus in information science. In recent years, use of ontology as a mechanism for representing knowledge in clinical decision support systems (CDSS) has gained momentum and has become more common in supporting and solving decision problems as complex as pain (Noy, Rubin, & Musen, 2004).

An information system is intended to be a representation of a world as perceived by a human or a group of humans (Wand & Weber, 2004). However, since healthcare is a very diverse and complex system, it is essential that healthcare information systems represent that kind of complexity. Ontologies are representations of concepts and relationships in a domain area and are important to the information systems field in that they have been described as the best base for building theories about information systems representations. There are three different ways ontologies can be used in information systems design. Firstly, as a benchmark to

evaluate models used in systems development and secondly to provide a set of concepts to model systems and to reason about their characteristics and finally to define the meaning of information that will be available through an information system (Kuziemsky, Downing, Black, & Lau, 2007).

One particular study used the grounded theory approach, which includes three systematic coding cycles: open, axial and selective coding. The ontology represents a means of formalizing the richness obtained through the grounded theory coding by applying a systematic way of organizing the concepts as well as establishing the relationships between them (Kuziemsky et al., 2007). The study also makes a contribution to ontology information systems research by illustrating how the grounded theory approach can be used to design an ontology that contains an empirically derived vocabulary, models the concepts and relationships in a complex domain area and also details the processes and information within the ontology (Kuziemsky et al., 2007).

2.3.2 Relevance Vector Machine (RVM)

Infants are unable to directly report their level of pain, and hence, physicians and nurses are responsible for pain assessment for neonates. Pain and distress behaviours in neonates, include facial expression, cry, and body movement, and a series of methods have been suggested to objectively assess pain in neonates based on the aforementioned behaviours. Correct interpretation of the facial expressions of the patient and its correlation with pain is a fundamental step in designing an automated pain assessment management system. However, by adding other pain behaviours, including head movement and the movement of other body parts, along with physiological indicators of pain, such as heart rate, blood pressure, and respiratory rate responses should make the pain detection system more accurate.

Recent advancements in pattern recognition techniques using Relevance Vector Machine (RVM) learning techniques can assist physicians and nurses in assessing pain by constantly

monitoring the patient and providing the clinician with quantifiable data for pain management (Gholami, Haddad, & Tannenbaum, 2010). In one particular study, the researchers used the RVM classification technique to distinguish pain from non-pain in neonates as well as assess their pain intensity levels.

The RVM algorithm can potentially be useful in assessing sedation and agitation in the NICU (Gholami et al., 2010). The fundamental limitations in sedation and agitation assessment in the NICU stem from subjective assessment criteria, rather than quantifiable, measurable data for NICU sedation. An automatic sedation and pain assessment system can be used within a decision support system, which can also provide automated sedation and analgesia in the ICU (Gholami et al., 2010). However, a system such as this is yet to be implemented in the NICU. Algorithms such as RVM can aid in creating an automated system for neonatal pain measurement.

2.3.3 Statistical Systems

There are many different statistical techniques that have also been discussed in the pain literature. One such literature review looked at and tested machine-learning techniques for abdominal pain in order to improve standard statistical systems. This study highlighted various different statistical systems that can be used to measure pain. The researcher's investigation was based on a prospective clinical database with 1254 cases, 46 diagnostic parameters and 15 diagnoses (Ohmann, Moustakis, Yang, & Lang, 1996).

Following are the different automated rule induction techniques that were presented (Ohmann et al., 1996).

1. ID3: construction of a decision tree via the iterative deployment of four rules.

2. C4.5: represents a system that implements the ID3 approach but is better in that it uses decision tree simplification and or pruning.

3. NewId: represents another implementation of ID3. Understands priority scheme.

4. ITRULE: uses the information-theoretic J-measure to find the most informative set of rules from an input data set. The algorithm searches the space for possible rules and calculates the information contents of the rules (J-measure).

5. CN2: represents a star generator machine learning system.

In this study, for evaluation purposes, the database of 1254 patients was randomly split into a training set and a separate test set. The standard model and all algorithms for automatic rule generation mentioned above were trained on the training set and separately evaluated on the test set (Ohmann et al., 1996).

Detailed analyses showed that the clinical impact of computer-aided diagnosis is a combined effect of audit, structured data collection, feedback, education and the computer. Automatic knowledge acquisition techniques used to develop rule-based systems from clinical data have been applied with good success in several other clinical areas.

Furthermore, another study done by same set of authors implemented this knowledge and used it to describe a knowledge-based system for the diagnosis of acute abdominal pain, in which scores and rule sets were integrated. This system in particular was linked to a documentation program via a medical data dictionary and allowed an online application of knowledge modules to clinical data (Eich, Ohmann, & Lang, 1997).

In this study, different rule sets were generated by automatic rule generation (C4.5) from a prospective database. The rule sets and two published diagnostic scores were evaluated on a test set, resulting in a diagnostic accuracy of 57 % for a general knowledge module and between 44 and 88 % for specific knowledge modules (Eich et al., 1997). In the first approach, rule-based sets were generated by machine learning techniques from clinical databases (automatic rule generation). The algorithm used was C4.5, representing a system similar to the ID3 approach. For testing of the rule sets generated by C4.5, a prospective European database with 10233 cases was used (Eich et al., 1997). The database was split into training set and test set. The whole system consisted of 3 program modules. A data dictionary, documentation program and a Knowledge Based System (KBS) (Eich et al., 1997). The data dictionary is a controlled vocabulary by which the integration of the documentation module and the KBS can be achieved. The second module was designed to collect data in clinical studies. In this documentation program, the authors carried out several prospective evaluation studies where their aim was to build up a quality-controlled prospective database on which they can apply knowledge-based methods. The third component of this system was the KBS, which was used for diagnostic support. They integrated rule-based systems, which automatically generate rules from prospective databases and scores. Sets of rules were created through the C4.5 algorithm. The KBS has a human interface and can be used as a stand-alone system.

2.3.4 Clinical decision support/ web based systems

The first Clinical Decision Support systems (CDSS) were standalone systems that were running separately from other hospital systems. They evolved into integrated systems where decision support was embedded into hospital information systems (e.g., offering support in such clinical areas as laboratory, nurse charting, radiology or pharmacy). Then, the integrated systems evolved into separated systems with shareable information and decision support content (Farion et al.,

2009). One study focused on the design of a clinical decision support system (CDSS) that supports heterogeneous clinical decision problems and runs on multiple computing platforms. This is important as in today's health care settings; practices have become interpersonal and collaborative as opposed to homozygous decision-making.

The proposed design of this study provided a common framework that facilitated development of diversified clinical applications running seamlessly on a variety of computing platforms. It was prototyped for two clinical decision problems and settings (triage of acute pain in the emergency department and postoperative management of radical prostatectomy on the hospital ward) and implemented on two computing platforms – desktop and handheld computers (Farion et al., 2009). In a previous study of the authors, a system named MET1 (Mobile Emergency Triage) was designed to help with management of paediatric patients using information about their history, physical examination and a limited number of laboratory tests (Farion et al., 2009). MET1 included two clinical applications (supporting triage of paediatric abdominal pain and paediatric scrotal pain). The system was able to run on handheld computers. However, MET1 was designed to support set of homogeneous decision problems only in a single setting such as the emergency department (ED) and to operate on a single computing platform (a handheld computer).

In the process, the researchers then proposed a new clinical decision support system design (referred to as MET2) that represents the next generation of CDSS that can make heterogeneous decisions. The researchers used ontology driven design to represent essential components of a CDSS. The two major architectural components of this design were the application repository and the executor. The application repository managed and stored the available application models. The executor created applications according to their application

models and executed them (Farion et al., 2009). Typically, upon the ED physician's request, the executor retrieved an appropriate application model from the application repository and created the user interface according to the data ontology using components from the interface repository.

Overall in this study, they have updated the previous MET1 system to a MET2 system, which introduces two new components: interface ontology and the configuration ontology. In the MET 2 system it encompasses data, support, interface and configuration models. In this new system, they also introduced the Entity-Attribute Value (EAV) approach to structure clinical information as it allows for more flexible and effective handling of heterogeneous data (Farion et al., 2009). The authors proposed a multi-device architecture for electronic information processing and communication in the clinical setting. A similar approach is also proposed by another new paradigm: activity-based computing. This approach also highlights clinical computing that considers user activities (tasks) as first class objects in a computing environment. A new-generation CDSS should be able to support heterogeneous decision problems (in particular those that require heterogeneous decision models and solvers) at different settings and to execute seamlessly on multiple computing platforms. This is important in the neonatal pain domain, as pain can be a result of many different factors. For this reason, the system implemented should be able to make heterogeneous decisions.

Following the concept of clinical decision support systems, one paper described a prototype clinical decision support system (CDSS) for risk stratification of patients with cardiac chest pain. The researchers employed a belief Rule-base Inference Methodology using the Evidential Reasoning approach (RIMER) for developing an intelligent CDSS (Kong et al., 2012). In this RIMER approach, belief rule base was employed to model clinical domain

knowledge under uncertainties. Such a rule base was capable of capturing vagueness, incompleteness, and nonlinear causal relationships.

In this clinical decision support study, the researchers adopted a unique web-based three layer architecture in the prototype design of the clinical decision support system design (Kong et al., 2012). Three-layer architecture of a web-based intelligent clinical decision support system should include four different system components: friendly web-based user interfaces, inference engine, knowledge base, and database. The core components implemented in the belief rule-based system prototype in this study include web-based user interface, database, inference engine and knowledge base.

Benefit of a web-based decision support system is that it can deliver suggestions or recommendations generated from the system to a much broader audience. Web-based decision support system in clinical areas can have many advantages such as provision of easy access to computerized decision support for clinicians in geographically different places. It can also provide easy dissemination of clinical domain knowledge and patient data among different clinical application systems which are linked through internet (Kong et al., 2012). Increase in the use of information technology (IT) in health care, particularly the introduction of clinical decision support systems, can help simplify the health care process and substantially facilitate clinical practice and reduce medical errors. Even though clinical decision support systems are promising in helping facilitate evidence based medicine and reducing patient adverse outcomes, there are challenges in this research area that have made few clinical decision support systems widely applied in practice (Kong et al., 2012).

2.3.5 Biopotentials

One study used the expression of pain and its biopotential parameters to show how pain can be measured. The researchers collected a database using visual and biopotential signals to advance an automated pain recognition system, to determine its theoretical testing quality, and to optimize its performance (Walter et al., 2013). For this purpose, participants were subjected to painful heat stimuli under controlled conditions. This study had many unique properties such as the use of highly controlled pain stimulation, multiple camera setup, recording of depth information via a camera and most importantly multimodal detection such as simultaneous data collection on skin conductance level, electrocardiogram (ECG), electromyogram (EMG), and electroencephalography (EEG) (Walter et al., 2013).

This biopotential study's aim was to select the features and feature patterns that contribute to the highest recognition rate for pain recognition, quantification and dissociation from emotion. A range of data fusion procedures was tested for this. Overall, the study used biopotential and video analyses to measure pain (Walter et al., 2013). The authors were moving towards the vision of creating an automatic system for an objective measurement of pain, which can facilitate pain monitoring, logging and support in a clinical environment. The authors named this "pain computing".

2.3.6 Facial recognition

In the last few years in the neonatal pain area, there has been great focus on automated facial recognition scales. Various pain assessment measures (tools, instruments, etc.) based on facial expressions of pain have been developed (Yuan, Bao, & Guanming, 2008). This has proved to be a popular area of research as an automated recognition of facial expressions of pain does not require reliance on health professionals that has great medical significance. The objective of the facial recognition study done by Yuan et al. is to bypass the observational problems by

developing a machine classification system to diagnose neonatal facial expressions of pain(Yuan et al., 2008). In this particular study, they employ an expressive feature, the Gabor feature, to the representation of neonatal faces. Each neonatal facial image is convoluted with the 2D Gabor Filters to extract 412,160 Gabor features (Yuan et al., 2008). The Gabor features are obtained by convoluting the image with a series of Gabor filers. Gabor filters extract features from different orientation and different scales (Yuan et al., 2008). Image analysis with Gabour filters is thought to be similar to perception in the human visual system, which is why it is used in this case.

An automated facial expression analysis system usually consists of three parts: face detection, facial feature extraction/representation and classification. In this experiment by Yuan and colleagues (2008), they attempted to distinguish cry expressions that were in response to pain from those cry expressions that were in response to a less noxious stimulus. Thus, two stimuli were included in this study: heel puncture and transporting the neonate from one crib to another. In the feature extraction stage, 2D Gabor filter were applied to extract the expression features from facial images. Finally, in the feature selection and classification stage, the proposed HybridBoost is applied to select the most informative features/weak classifiers, and by this a hierarchy of strong classifiers are constructed (Yuan et al., 2008). Experiments with 510 neonatal expression images showed that the proposed method in this study was effective and only 30 Gabor features were enough to achieve good classification performance (Yuan et al., 2008). The recognition rate of pain versus non-pain was up to 88%. Over the last decades, such automatic facial expression analysis has become an active research field that shows high potential in many areas, such as the automated assessment of neonatal pain.

Another facial recognition study describes the Infant Classification Of Pain Expressions (COPE) Project (Brahnam et al., 2007). A short-term goal of this project was to investigate the

feasibility of using holistic face recognition techniques to detect pain signals in newborn facial displays. The long-term goal of the authors was to develop working systems that can be implemented in neonatal units (Brahnam et al., 2007). Selection of stimuli used to provoke facial displays in the neonates was of critical importance in this study. For the initial infant study, four noxious stimuli were selected: the puncture of a heel lance, friction, produced by swabbing on the external lateral surface of the heel, an air stimulus on the nose and lastly transport from one crib to another (Brahnam et al., 2007). The objective of performing these procedures was to obtain a representative and challenging set of images for evaluating face classification systems of pain.

2.4 Neurological Indicators

Pain causes detectable biochemical, physiological, and behavioural changes during the rapid development phase of the central nervous system. Repeated and long-term experiences of pain have adverse effects on brain development and cognitive development as well as on later pain behaviour manifestation and pain sensations. Thus, it is important to alleviate pain during hospitalisation in order to promote the optimal neurological and functional development of preterm infants (Korhonen et al., 2013).

Brain imaging is a very active area of pain research that can aid in leading to improved pain measurement in the future (Apkarian, Bushnell, Treede, & Zubieta, 2005; Borsook & Becerra, 2006; Slater et al., 2006). Some imaging methods that can be used for this purpose are positron-emission tomography, single-photon emission computed tomography, near infrared spectroscopy, and functional magnetic resonance imaging (Berde & McGrath, 2009). It can also be possible to detect signals reflecting regional brain glucose use, blood flow, or regional ratios of oxy- to deoxy-hemoglobin, respectively, as surrogate measures of regional neuronal metabolic

activity (Berde & McGrath, 2009). Furthermore, magnetic or electric source potential mapping or processed electroencephalographic measures can also be used as surrogate measures of regional brain electrical activity (Berde & McGrath, 2009). In the mid-to- late 1980s, infant behavioural researchers used cortisol levels to assess behavioural responses to various stressors. Stress response to handling, pain response to heel stick, and circumcision, as well as the effects of biomedical status on response were key issues. Cortisol responses were noted to be greater following painful procedures than routine handling, and behavioural responses did not correlate well with peak cortisol levels (Herrington, Olomu, & Geller, 2004).

A research study done by Slater and colleagues in 2008 assessed the association between cortical pain responses in young infants and currently used tools for the assessment of pain in these infants (Slater et al., 2008). They were able to record infant cortical activity in response to noxious stimulation, which provided for a first of its kind opportunity to look at the relationship between clinical pain assessment scores, on the basis of behavioural and physiological responses, with measurements of pain processing in the brain. The researchers presented an analysis of the association between the cortical haemodynamic activity and the components of a clinical pain assessment tool (PIPP) (Slater et al., 2008).

Behaviours to communicate pain require motor responses to sensory and emotional stimuli (Slater et al., 2008). The maturity of the premature infant to this complex system is not clearly understood currently. The results of the study by Slater (2008) raise further awareness of the ability of infants to experience pain and highlight the possibility that pain assessment based on behavioural tools alone may underestimate the pain response in infants. The researchers also concluded that positive brain response to painful stimulus could occur even in the absence of any

facial expression. These conclusions are important as they obtain evidence for the urgent need to provide physiological changes in a greater focus in the detection of pain in neonatal infants.

2.5 Conclusion

This chapter has presented various studies that have outlined that machine-based systems can be used to detect and analyze physiological changes associated with pain such as increase in heart rate. Numerous automated approaches have been presented to assess pain based on analysis of physiological indicators (Brown et al., 2011; Lindh et al., 1999). Studies have shown that there is high correlation between pain/discomfort and changes in vital signs (e.g., heart rate increasing) (Janig, 1995). Many of the automated systems already in place such as the traditional assessment of infant's crying and facial recognition systems are biased and depend totally on the observer's subjective judgement (Riddell & Craig, 2007). Therefore, using the knowledge from this chapter and developing a quantitative and minimally biased pain assessment system that can continuously provide feedback is important. It is important to use the knowledge and techniques presented in this chapter to design a pain assessment scale that can be standardized. There is tremendous potential to use these techniques to automate a calculable pain score.

Chapter 3 - Artemis

This chapter presents The Artemis platform used in this thesis. The Artemis framework is a framework for concurrent multi-patient, multi-diagnosis and multi-stream temporal analysis in real-time for clinical decision support as well as prospective and retrospective clinical research (Blount et al., 2010). The Artemis framework relies on the acquisition of physiological data from existing bedside medical devices contained within the NICU as the data are generated. These devices, which are already enabled to output their data for collection, can be connected via an Ethernet and/or serial port. Sources of data include bedside physiological monitoring devices, medical equipment such as ventilators and infusion pumps, and clinical information management systems (CIMS), which house the patient's electronic medical record and laboratory results (Blount et al., 2010; McGregor, 2013; McGregor, Catley, James, & Padbury, 2011). The analytic results provide clinicians with integrated temporal summaries of events (McGregor, Catley, James, et al., 2011), which delivers advanced clinical decision support. Since 2009, the Artemis system has been deployed in several NICUs of various hospitals. In such clinical settings, the Artemis system has the ability to acquire, collect, analyze, and store data containing electrocardiogram (ECG), heart rate (HR), respiratory rate (RR), oxygen saturation (SpO2), and blood pressure (BP) streams as well as Clinical Information Management System (CIMS) observations for clinical research (Blount et al., 2010).

The Artemis framework is outlined in Figure 1 (McGregor, Catley, James, et al., 2011). The data acquisition component enables the provision of real-time synchronous medical device data and asynchronous CIMS data to be refined, formatted and standardized. Subsequently, this data is forwarded for analysis within the Online Analysis component that operates in real-time (McGregor, Catley, James, et al., 2011).

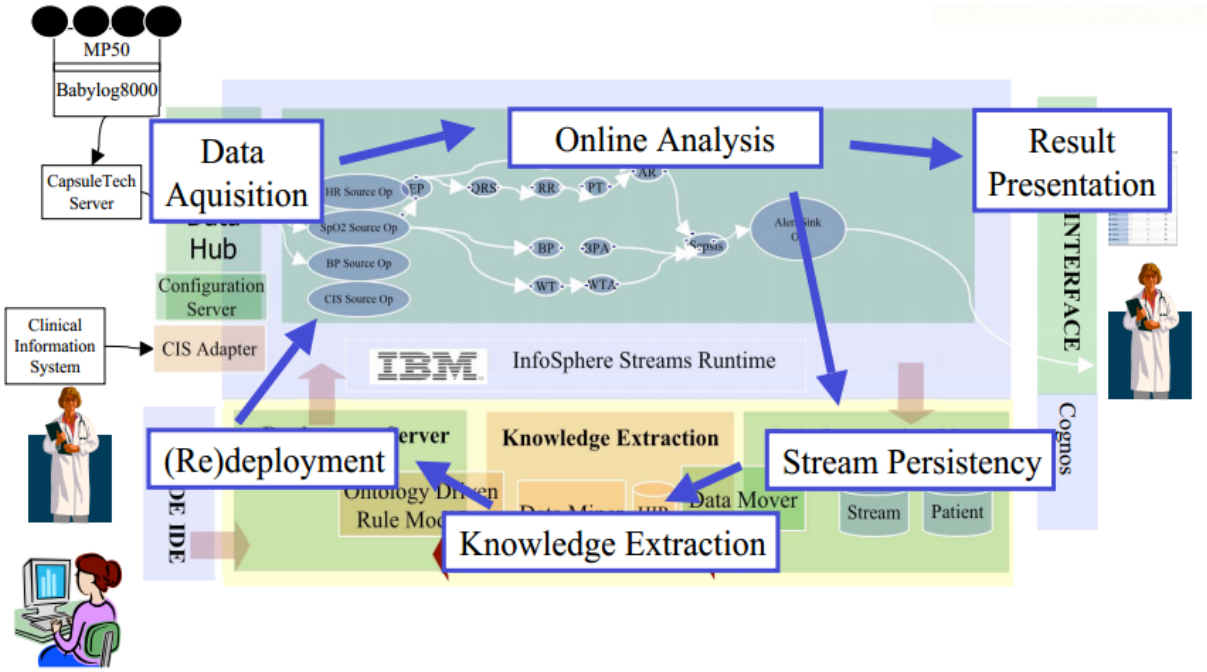


Figure 1. Artemis Framework

For this real-time component, Artemis employs IBM's InfoSphere Streams, a novel streaming middleware system that processes data in real-time and then enables data storage within the Data Persistency component (McGregor, Catley, & James, 2011). Consequently, the system is capable of processing and storing both the raw data and derived data from multiple infant sources at the rate at which the data are generated (Blount et al., 2010). Stream processing is supported by IBM's Stream Processing Language (SPL), which is the system specific programming language for IBM's InfoSphere Streams middleware (Blount et al., 2010). For the Knowledge Extraction component, Artemis utilizes a newly proposed temporal data mining approach (McGregor, Catley, & James, 2012). This component supports the discovery of condition onset behaviours in physiological data streams and associated clinical data. New knowledge, once tested and derived from rigorous clinical research techniques, is transferred for use within the Online Analysis component through the Re-deployment component, which translates the knowledge to an SPL representation.

Overall, Artemis can be defined as an extensive and cultivated Big Data informatics platform within which compatible algorithms, such as the one proposed in this thesis for pain assessment, can be functionally deployed. "Big Data" refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze (Manyika et al., 2011). Online health analytics have significant relevance in the critical care domain. The enormous quantities of complex physiological data continuously produced by critical care monitors and equipment exceeds the clinician's capacity for processing (McGregor, 2013; McGregor, Catley, James, et al., 2011). The use and adoption of online health analytics has great potential to enable quality improvement. Such analytics enable the real-time processing of early prognosticators of impending clinical deterioration (McGregor, 2013; McGregor, Catley, James, et al., 2011) and perform as an early warning system. The application of online health analytics promotes timely intervention and improved outcomes for patients (McGregor, 2013; McGregor, Catley, James, et al., 2011). The utilization of a similar system can be useful in additional research areas and future retrospective research opportunities such as that presented in this thesis. The work presented in this thesis, is part of a larger Artemis project, which is ongoing at the Health Informatics Research Lab at University of Ontario Institute of Technology (UOIT). Within the Artemis project, many projects have been completed within varied areas of research, which focus on a real-time algorithm design. For example, a similar algorithm design has been used for monitoring conditions related to retinopathy of prematurity (Cirelli, McGregor, Graydon, & James, 2013). Additionally, such algorithms have been incorporated into the automated monitoring techniques for apnoea of prematurity (Catley, Smith, McGregor, James, & Eklund, 2011) and detection of sleep-wake cycling patterns in neonates (Eklund et al., 2014). Such projects depict the optimistic possibility of Artemis in providing a significant contribution to clinical diagnostics with physiological data. Similarly, this thesis work will demonstrate the utility of the Artemis platform for the effective

detection and monitoring of pain within the neonatal population in the future. The Artemis platform has been used previously for preliminary work on this subject. Such work focused on designing rules to create a novel pain profile, and also includes a pilot retrospective study compared between Artemis generated pain score and a pain score that was in use at The Hospital for Sick Children, Toronto, Canada (Naik, Bressan, James, & McGregor, 2013; Naik & McGregor, 2014; Naik et al., 2014). Using this background work with Artemis capabilities, it is possible to design an algorithm unique to this research that can be deployed within Artemis to create a pain detection system that can provide diagnostic support to health care professionals.

Chapter 4 - Methodology

Literature has demonstrated that untreated pain in premature infants may cause long-term effects associated with changes in the nervous system, pain perception, sleep disturbances, feeding problems, and chronic pain syndromes (Slater et al., 2008). The inexistence of a gold standard within pain management, the complexity of the premature infant and the rapid deterioration of the fragile neonatal population poses challenges that must be overcome in order to achieve optimal pain management tools. For this reason, this chapter will introduce several methods to achieve such goals.

The methodology in this chapter is presented in two phases. In Phase I, a data preparation phase is introduced to demonstrate the work in which retrospective analysis is completed using the data collected from the Artemis platform in the DB2 system. In this data preparation phase, the abstractions were put into individual streams to look for features in the physiological streams. The data preparation results were used during the second phase, where a data model was created using Microsoft Excel. Phase II involved combining the, abstractions and features from the data preparation phase in order to compute an automated partial pain score based on big data analytics and quantifiable scoring from the streams prepared in data preparation phase (Heart Rate (HR), Oxygen Saturation (SpO_2) and Gestational Age (GA)). The steps outlined in this chapter are completed in order to carry out experiments in chapter 5.

The population for this study was twenty-three premature infants; thirteen males and ten females; gestational age 33.2 ± 5.41 weeks and birth weight 2060 ± 910 grams. Data are presented as mean ± SD.

4.1 Phase I: Data preparation

Demographic data were extracted from the medical records such as GA, birth weight, gender, admission diagnosis, PIPP and surgical information. This data was collected for twenty-three subjects, and physiological data for heart rate (HR) and oxygen saturation (SpO_2) were streamed from The Artemis Platform. Physiological data from the NICU bedside devices were captured in real-time. The approach to automation involved blocking the second-by-second data into a one-hour window to construct an hourly partial PIPP score.

Excel was used for the construction of a series of queries that would each create a temporal abstraction for a given hour. As noted in the discussion in chapter 7, this process would be replaced by Infosphere streams code that can run in real-time and has windowing capabilities. For each subject, the following steps were performed to extract and organize the data for analysis to enable hour-by-hour analysis for the physiological data components of HR and SpO_2:

1. The start and end date and time to be used for each subject was identified and was then converted to GMT epoch timestamp to run the query. The identification of start and end epochs for every hour was essential. Data preparation file was created in Microsoft Excel® for each subject separately.

2. The epochs were put in a row for each hour to match the original date and time. The date and time was converted to epoch with the following equation in Microsoft Excel®:

$$((Date\ \&\ time - 25569) \times 86400) \qquad \textit{Equation 1}$$

Table 2. Date & time conversion to EPOCH

Date & Time (DD/MM/YY HH:MM)	EPOCH
09/03/2010 20:00	1268164800
09/03/2010 21:00	1268168400
09/03/2010 22:00	1268172000
09/03/2010 23:00	1268175600
10/03/2010 0:00	1268179200

3. Since there are 3600 seconds in one hour, generation of the start of the hour epoch as well as the end of the hour epoch is important. One column was created for the start of the hour epoch, using the following formula in Microsoft Excel®:

$$(epoch\ (original\ time) - 3601) \qquad \textit{Equation 2}$$

Another column was created for end of the hour epoch for which the following formula was used in Microsoft Excel®:

$$(epoch\ (original\ time) + 1) \qquad \textit{Equation 3}$$

This ensured that full data for within the hour was collected. Table 3 shows an overview of the preparation sheet for data extraction.

Table 3. Data preparation sheet for EPOCH

Patient ID	Date & Time	EPOCH (original time)	Start Epoch (-3601)	End Epoch (+1)
Subject 1	09/03/2010 20:00	1268164800	1268161199	1268164801
Subject 1	09/03/2010 21:00	1268168400	1268164799	1268168401
Subject 1	09/03/2010 22:00	1268172000	1268168399	1268172001
Subject 1	09/03/2010 23:00	1268175600	1268171999	1268175601
Subject 1	10/03/2010 0:00	1268179200	1268175599	1268179201

4. The query was run in STDM framework instantiated within The Artemis Platform in the knowledge discovery component to extract results for HR and SpO_2 for each hour. The following formula was used in Microsoft Excel®:

=CONCATENATE ("select patientid, instancestart, avg(hrvalue) as AVG, min(hrvalue) as MIN, max(hrvalue) as MAX, stddev(hrvalue) as STDDEV, count(hrvalue) as COUNT from (select patientid, '", D2, "' as instancestart, timestamp, hrvalue from db2inst1.rawhr where patientid = 'N",A2, "' and timestamp > ", E2, " and timestamp < ", F2, ") group by patientid, instancestart ")

5. The query was run for each hour for each subject. This extracted per second data for every hour leading up to the PIPP score recorded.

Table 4. Output sheet including artifacts for Heart Rate data

Subject ID	DATE & TIME	EPOCH	AVG	MIN	MAX	STD DEV	COUNT
Subject 1	09/03/2010 16:00	1268150400	162.49	102	199	15.50	3516
Subject 1	09/03/2010 17:00	1268154000	24004.3	74	8388607	446660	3517
Subject 1	09/03/2010 18:00	1268157600	8388607	8388607	8388607	0	161
Subject 1	09/03/2010 20:00	1268164800	131.1	119	138	2.97	902
Subject 1	09/03/2010 21:00	1268168400	139.29	128	166	5.34	3517
Subject 1	09/03/2010 22:00	1268172000	142.50	83	173	8.45	3516
Subject 1	09/03/2010 23:00	1268175600	140.93	115	155	3.68	3511

6. To clean the data, the query was updated and changed to exclude the artifacts. If any artifacts were found, they were removed for each hour for both HR and SpO_2 and the total counts of the rows was recorded. Following query was used to remove artifacts in Microsoft Excel®:

=CONCATENATE("select patientid, instancestart, avg(hrvalue) as AVG, min(hrvalue) as MIN, max(hrvalue) as MAX, stddev(hrvalue) as STDDEV, count(hrvalue) as COUNT from (select patientid, '", D2, "' as instancestart, timestamp, hrvalue from db2inst1.rawhr where patientid = 'N",A2, "' and timestamp > ", F2, " and timestamp < ", G2, " and hrvalue <> 8388607) group by patientid, instancestart")

7. After removal of artifacts (if any), the final output provided per second data for HR, which was then summarized into average, minimum, maximum, standard deviation and number of rows (counts) for each hour. Table 5 shows the output sheet for HR.

Table 5. Output sheet for Heart Rate data without artifacts

Subject ID	DATE & TIME	EPOCH	AVG	MIN	MAX	STD DEV	COUNT
Subject 1	09/03/2010 16:00	1268150400	162.49	102	199	15.499	3516
Subject 1	09/03/2010 17:00	1268154000	153.10	74	177	8.066	3507
Subject 1	09/03/2010 20:00	1268164800	131.10	119	138	2.968	902
Subject 1	09/03/2010 21:00	1268168400	139.29	128	166	5.336	3517
Subject 1	09/03/2010 22:00	1268172000	142.50	83	173	8.449	3516
Subject 1	09/03/2010 23:00	1268175600	140.93	115	155	3.682	3511
Subject 1	10/03/2010 0:00	1268179200	139.57	122	156	3.893	3451

8. Similarly, an output was provided for per second data for SpO_2, which was then summarized into average, minimum, maximum, standard deviation and number of rows (counts) for each hour. Table 6 shows the output sheet for SpO_2.

Table 6. Output sheet for SpO_2 without artifacts

PATIENT ID	DATE & TIME	EPOCH	AVG	MIN	MAX	STD DEV	COUNT
Subject 1	09/03/2010 16:00	1268150400	94.92	78	100	4.008	3516
Subject 1	09/03/2010 17:00	1268154000	95.12	83	98	2.162	3504
Subject 1	09/03/2010 20:00	1268164800	95.24	77	99	5.042	902
Subject 1	09/03/2010 21:00	1268168400	94.16	78	99	2.712	3516
Subject 1	09/03/2010 22:00	1268172000	94.56	71	97	3.859	3506
Subject 1	09/03/2010 23:00	1268175600	94.13	86	96	1.228	3511
Subject 1	10/03/2010 0:00	1268179200	92.23	85	95	1.639	3453

9. After this, the PIPP scores, when available, were manually inputted into the rows (hours) (PIPP scores were collected by nurses and extracted for research by a research assistant from the CIMS system at SickKids Hospital).
10. GA was also calculated based on the age of the infant at the time of first collected PIPP score. Based on this calculation, the GA was changed over time. Table 7 shows the final data set view.

Table 7. Final dataset view

Subject ID	DATE & TIME	EPOCH	AVG	MIN	MAX	STD DEV	N	PIPP	GA
Subject 1	09/03/2010 16:00	1268150400	162.49	102	199	15.499	3516	8	28
Subject 1	09/03/2010 17:00	1268154000	153.10	74	177	8.066	3507		28
Subject 1	09/03/2010 20:00	1268164800	131.10	119	138	2.968	902	5	28
Subject 1	09/03/2010 21:00	1268168400	139.29	128	166	5.336	3517		28
Subject 1	09/03/2010 22:00	1268172000	142.50	83	173	8.449	3516		28
Subject 1	09/03/2010 23:00	1268175600	140.93	115	155	3.682	3511		28
Subject 1	10/03/2010 0:00	1268179200	139.57	122	156	3.893	3451	10	28

4.2 Phase II: Data Model

Following the data preparation phase, a data model was created based on the Premature Infant Pain Profile (PIPP) using Artemis data. The PIPP scale was chosen as the base comparison scale for this thesis because the clinical institution (The Hospital for Sick Children, Toronto, Canada) that was associated with this study used this particular scale in their clinical setting. The PIPP scores were collected from The Hospital for Sick Children's NICU records. Nurses collected PIPP scores in a handwritten format. These records were retrospectively examined and the scores were recorded for research purposes.

The PIPP scale was used in correlation with Artemis data to compute a total score for each hour using maximum HR, minimum SpO_2 and GA. Since PIPP is the base scale used in this study, the scale was used to explore and demonstrate a new scoring system. Following this, Artemis Premature Infant Pain Profile (APIPP) was created. By depicting the parameters of scoring from the already in place PIPP scale, an attempt was made to create a model scoring system using The Artemis Platform.

4.2.1 Premature Infant Pain Profile (PIPP)

The Premature Infant Pain Profile (PIPP) is a behavioural and physiological multidimensional assessment tool, which provides a measure of the premature infant's response to pain (Stevens, Johnston, Petryshen, & Taddio, 1996). Scoring indicators include gestational age, behavioural states and physiological factors. The presence or degree of change/pain is rated on a four-point scale, which includes seven indicators. The indicators include upper facial activity, physiological activity and behavioural state. The scale ranges from zero (minimum score) to 21 (maximum score). The PIPP scale was the first of its kind multidimensional premature infant pain assessing scale.

The PIPP is one of the very few scales that accounts for the infant's gestational age, thus allowing the distinction among mature, full-term and, preterm infants (Gallo, 2003). PIPP was developed using data from four studies involving 238 neonates undergoing heel stick, where such studies identified the indicators of PIPP while establishing the validity of the instrument (Franck & Miaskowski, 1997).

When Stevens et al. created the PIPP scale, they reviewed previous literature and concluded that previous scales were mostly one-dimensional and only included behavioural responses to pain (Stevens et al., 1996). These scales did not include the physiological indicators

or factors to modify the pain response. The investigators highlighted a need for multidimensional pain measure for premature infants. Hence, the PIPP scale was developed using multiple indicators. The PIPP scale was developed and validated using a prospective and retrospective design. Indicators of pain were identified by clinical experts and using literature. Indicators were retrospectively tested using four existing data sets.

PIPP provided a sufficient background for this thesis to carry out a case study that explores the hypothesis of designing a scoring system using real-time big data analytic techniques to quantify pain in the neonatal population. To investigate the research questions for this thesis, the election of parameters that will be used for the case study was important. Since The Artemis Platform contains the data for both physiological parameters contained within the PIPP (HR, SpO_2), these particular parameters were included in the design of the novel APIPP scale.

Table 8. Premature Infant Pain Profile (PIPP)

Process	Indicator	0	1	2	3	Total Score
Chart	**Gestational Age**	36 weeks or more	32-35 weeks, 6 days	28-31 weeks, 6 days	Less than 28 weeks	
Observe infant for 15 seconds	**Behavioural State**	Active, awake, eyes open, facial movement	Quiet awake, eyes open, no facial movements	Active sleep, eyes closed, facial movements	Quiet sleep, eyes closed, no facial movements	
Observe baseline heart rate	**Heart Rate Maximum**	0-4 beats per min increase	5-14 beats per minute increase	15-24 beats per min increase	25 beats per min or more	

and oxygen saturations for 30 seconds					increase	
	Oxygen saturation minimum	Decrease of 0-2.4%	Decrease of 2.5-4.9%	Decrease of 5-7.4%	Decrease of 7.5 or more	
Observe infant's facial actions for 30 seconds	**Brow Bulge**	None	Minimum	Moderate	Maximum	
	Eye Squeeze	None	Minimum	Moderate	Maximum	
	Nasolabial furrow	None	Minimum	Moderate	Maximum	
TOTAL SCORE						

Since this thesis presents an approach to automate a physiological scale, the two physiological parameters (HR, SpO_2) contained within the PIPP scale were used for analysis in the design of scoring the new scale APIPP. Gestational age was also included in the APIPP analysis as Gestational age has proven to have varying effects on a neonates' pain perception. Additionally, gestational age is a calculable field that is independent of the infant's state at the bedside.

4.2.2 Artemis Premature Infant Pain Profile (APIPP) scoring analysis

APIPP scoring is an attempt to assess the possibility of automation in pain assessment of premature infants. For this reason, the two physiological indicators contained within the PIPP scale were included in the APIPP scale. The physiological data for HR and SpO_2 were also readily available in The Artemis platform. Gestational age was also used as a factor for this scale as gestational age can have varied effects on how a neonate responds to pain. The APIPP scale displayed in Table 9 ranges from zero (minimum score) to 9 (maximum score).

Table 9. Artemis Premature Infant Pain Profile (APIPP)

Indicator	0	1	2	3	Total Score
Gestational Age	36 weeks or more	32-35 weeks, 6 days	28-31 weeks, 6 days	Less than 28 weeks	
Heart Rate Maximum	0-4 beats per min increase	5-14 beats per minute increase	15-24 beats per min increase	25 beats per min or more increase	
Oxygen saturation minimum	Decrease of 0-2.4%	Decrease of 2.5-4.9%	Decrease of 5-7.4%	Decrease of 7.5 or more	
TOTAL SCORE					

Scoring analysis:

For this generation of the APIPP Temporal Abstraction, the hourly summary data was used from section 4.1. All individual physiological streams data and gestational age data was scored individually and then combined in one Microsoft Excel sheet to compute a total APIPP score (based on the PIPP scoring criteria) that can be compared to the PIPP score.

1. First, maximum heart rate, minimum SpO2 and gestational age scores were computed based on the PIPP scoring criteria for each hour of data extracted from the Artemis platform.

Table 10. Individual scoring of max HR, min SpO_2 and GA

Subject ID	Date & Time	Max HR	MAX HR Score	Min SpO2	MIN SpO2 Score	GA	GA Score
Subject 1	09/03/2010 16:00	199		78		28	

Subject 1	09/03/2010 17:00	177	0	83	0	28	2
Subject 1	09/03/2010 20:00	138	0	77	2	28	2
Subject 1	09/03/2010 21:00	166	3	78	0	28	2
Subject 1	09/03/2010 22:00	173	1	71	2	28	2
Subject 1	09/03/2010 23:00	155	0	86	0	28	2
Subject 1	10/03/2010 0:00	156	0	85	0	28	2
Subject 1	10/03/2010 1:00	180	2	70	3	28	2

2. Subsequently, maximum HR, minimum SpO_2 and gestational age were summed together to calculate a total APIPP score. This score was then compared to the PIPP score recorded by nurses at The Hospital for Sick Children, Toronto, Canada. This APIPP score was computed for all 23 subjects and compared to PIPP score.

Table 11. Final output for APIPP scoring

MAX HR	MIN SpO_2	GA	Total APIPP Score	PIPP Score
0	0	2	2	8
0	2	2	4	
3	0	2	5	5
1	2	2	5	
0	0	2	2	
0	0	2	2	
2	3	2	7	10

Overall, this chapter presented the preparation of the dataset that was used to carry out experiments in chapter 5. A data model was also created based on the PIPP scale using Artemis data. Depicting the scoring rules from the PIPP scale and using the data from The Artemis

Platform, a model scoring system was presented (APIPP). The following chapter will demonstrate the data model presented in this chapter by presenting a case study analysis.

Chapter 5 - Case study & Evaluation of Data model

This chapter presents the implementation of methodology described in chapter 4. This study is a retrospective quantitative study conducted with the data collected at The Hospital for Sick Children, Toronto, Canada in the Neonatal Intensive Care Unit (NICU). This study was reviewed and approved by the Research Ethical Board under REB# 1000036025. For this analysis, physiological data was captured continuously from NICU bedside monitoring devices and was streamed for temporal analysis using The Artemis Platform that was discussed earlier. Data was extracted between 2010 and 2013 from surgical and non-surgical patients during their stay in the NICU. After closely inspecting the data for case study analysis, 13 out of the 23 subjects were excluded due to insufficient data. Reasons for exclusion are discussed in more detail in section 5.2. Overall, total of 10 subjects data was used for experiments.

Following the creation of the data model and scoring of APIPP, the APIPP score was compared to the currently in place PIPP data collection for each subject. An in-depth analysis was conducted with the PIPP scores to examine if the APIPP score produces greater, lesser or equal results.

The data for the three main variables analyzed in chapter 5 and 6 are significantly spread out with great variability in their data. Following is the breakdown for the ten subjects' data for GA, surgery data and PIPP scores.

a. Gestational Age:

   Out of ten subjects, three subjects were 'term infants' from 37 to 41 weeks; three subjects were 'moderate preterm to late preterm infants' from 32 to 36 weeks; two were 'very preterm infants' from 28 to 31 weeks and one was 'extremely preterm' with less than 28

weeks of gestational age. Considering the wide distribution of the subjects through the spectrum, there is a small number of subjects per gestational age.

b. Surgery:

Seven subjects underwent a surgical procedure during their stay in the NICU with a total of 2105 lines of data, an average of 301 lines. Three subjects did not undergo surgery with a total of 295 lines, an average of 98 lines. In addition, only two patients presented the same surgical procedure: colostomy.

c. PIPP:

Ten subjects had a total of 478 PIPP scores collected. Out of this, following are the distributions for each PIPP scores: PIPP 0 = 15, PIPP 1 = 22, PIPP 2 = 38, PIPP 3 = 89, PIPP 4 = 84, PIPP 5 = 80, PIPP 6 = 52, PIPP 7 = 35, PIPP 8 = 25, PIPP 9 = 18, PIPP 10 = 6, PIPP 11 =5, PIPP 12=5, PIPP 14=4. These scores were not regularly collected but are rather collected at different intervals, posing a challenge, as the PIPP scores were not continuous for every hour. PIPP scores were missing for 1651 out of a total of 2400 hours that were analyzed for the ten subjects. For this reason, only 749 hours of data could be used for analysis.

5.1 Case Study

For the preliminary analysis, the data was analysed as a case study for subject 1. A total of 1910 hours of data were analysed for subject 1. For comparison purposes, the data for both scores for the subject were inputted into a pivot table, which was later plotted on a line graph. The data was filtered to exclude null values (blank) that were present within the PIPP data for an accurate depiction and comparison. Figure 2 displays the data for subject 1.

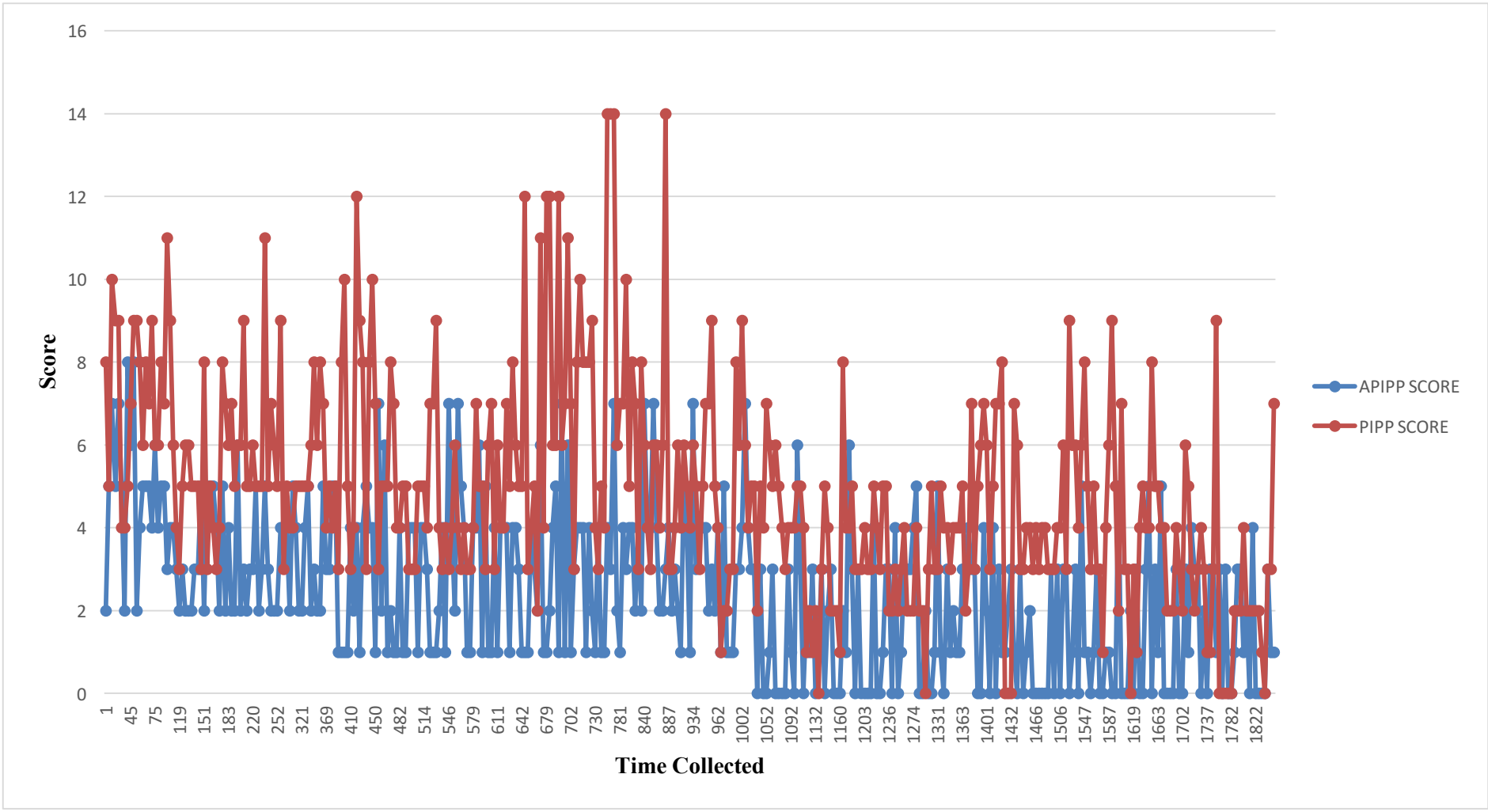


Figure 2. PIPP VS. APIPP Comparison for Subject 1

After the analysis, subject 1 was found to have greater PIPP scores then the APIPP scores. The PIPP maximum score is 21 and APIPP maximum score is 9; as such, investigating the difference between the two scores is imperative in order to understand the potential clinical efficacy. The following steps were carried out to create a data model:

1. The difference between APIPP and PIPP score was computed in Excel for all values except null. Following this, four categories were created to analyse the difference between these two scores:

   Greater = PIPP score is greater than APIPP score

   Lesser = PIPP score is less than APIPP score

   Equal = PIPP score and APIPP score have a similar score

   Null= PIPP score was not collected when APIPP score was present

2. Another column was created to display whether the PIPP score was greater, lesser, equal or null. Table 12 displays the output.

Table 12. Difference comparison between PIPP and APIPP scores

APIPP score	PIPP Score	Difference between scores	PIPP vs. Artemis
2	8	6	Greater
4			Null
5	5	0	Equal
5			Null
2			Null
2			Null
7	10	3	Greater

3. Following this, a pivot table was created to compare the percentage difference for each of the four categories. Results were summarized in a table and displayed in graphical form. Following result was outputted:

Category	% of Time outputted
Lesser	3.46%
Equal	1.83%
Greater	14.76%
Null	79.95%
Total	**100.00%**

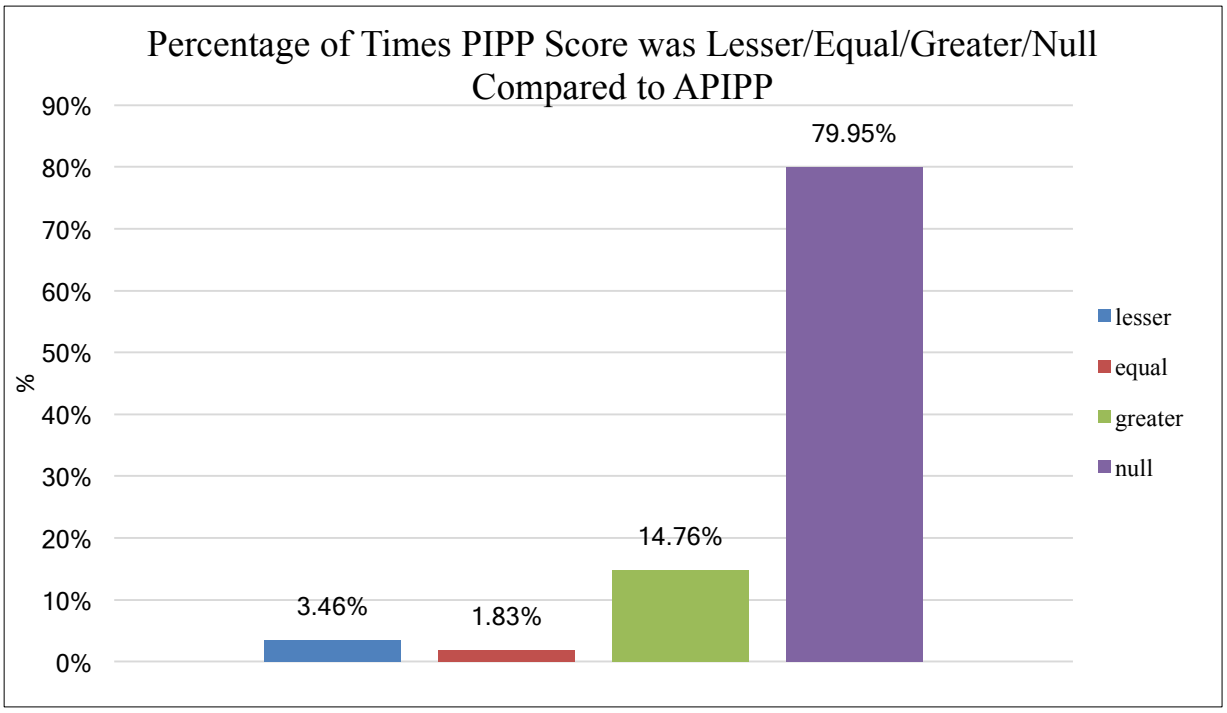


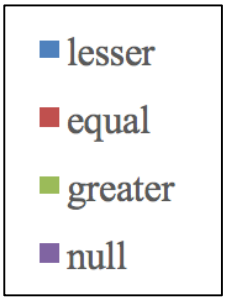
Figure 3. Percentage of times PIPP score was lesser/equal/greater/null compared to APIPP score for Subject 1

The results showed a higher percentage of null values (79.95%). Such a result demonstrates that among 79.95% of the hours that the infant was in the NICU, pain scores were not recorded on an hourly basis. This result shows that 79.95% of the time, the PIPP values were not collected or available when the APIPP values were. Since the APIPP score was consistently scored for every hour, these scores produced a more frequent result in comparison to the PIPP score, while the PIPP scores that were collected by the nurses at The Hospital for Sick Children, Toronto, Canada were collected in an intermittent manner. The data analysed showed that these values were not collected consistently every hour, but rather, were collected infrequently at random time points.

The second highest percentage was of the greater category. For 14.76% of the time, the PIPP score was greater than the APIPP score. This result was predictable as the PIPP scoring is a 21-point scale whereas the APIPP is a 9-point scale. However, understanding the degree of difference between the PIPP score and APIPP score is important. Table 13 shows the differences that were recorded for each of the four categories.

Table 13. List of # of times a difference was recorded in each of the four categories (Subject 1)

Difference	# of Times Difference was Recorded
-4	2
-3	7
-2	15
-1	42
0	35
1	36
2	63
3	56
4	52
5	24
6	24
7	11
8	6
9	3
10	2
11	5
(Null)	1527
Grand Total	**1910**



In total, 1910 hours of data were analysed for subject 1. In 66 instances, the PIPP score was less than APIPP score. In 35 instances, the PIPP score was equal to APIPP score. In 282 instances, the PIPP score was greater than APIPP score. For 1527 hourly instances of the APIPP score, the PIPP score was not recorded.

To better understand and analyse the difference between APIPP and PIPP, the hours where a PIPP score was not recorded were removed. Within this context, for 73.63% of the hours studied, the PIPP was greater than APIPP. This states that if lesser, equal and greater categories were compared, for subject 1, for 73.63% of the time the PIPP score was greater than the APIPP score. It is also noteworthy to discover that for 17.23% of time the PIPP score was found to be lower than the APIPP score.

Category	**% of times outputted**
Lesser	17.23%
Equal	9.14%
Greater	73.63%
Total	**100.00%**

Despite a 12-point difference between the two scales, it is important to investigate the lesser category findings as the generation of higher scores by the APIPP scale is a significant finding that requires further investigation. Various reasons can be associated with such a result. For example, because the automated scoring of APIPP provides an output based on physiological changes, the APIPP score may be more sensitive than the PIPP score. Because the APIPP score incorporates scoring of physiological changes with per second data into hourly chunks, the automated scoring of APIPP may have detected signs that were missed by the nurses' PIPP scoring. Such a finding will be important to investigate in the future. Figure 4 shows the difference for the lesser, equal and greater categories.

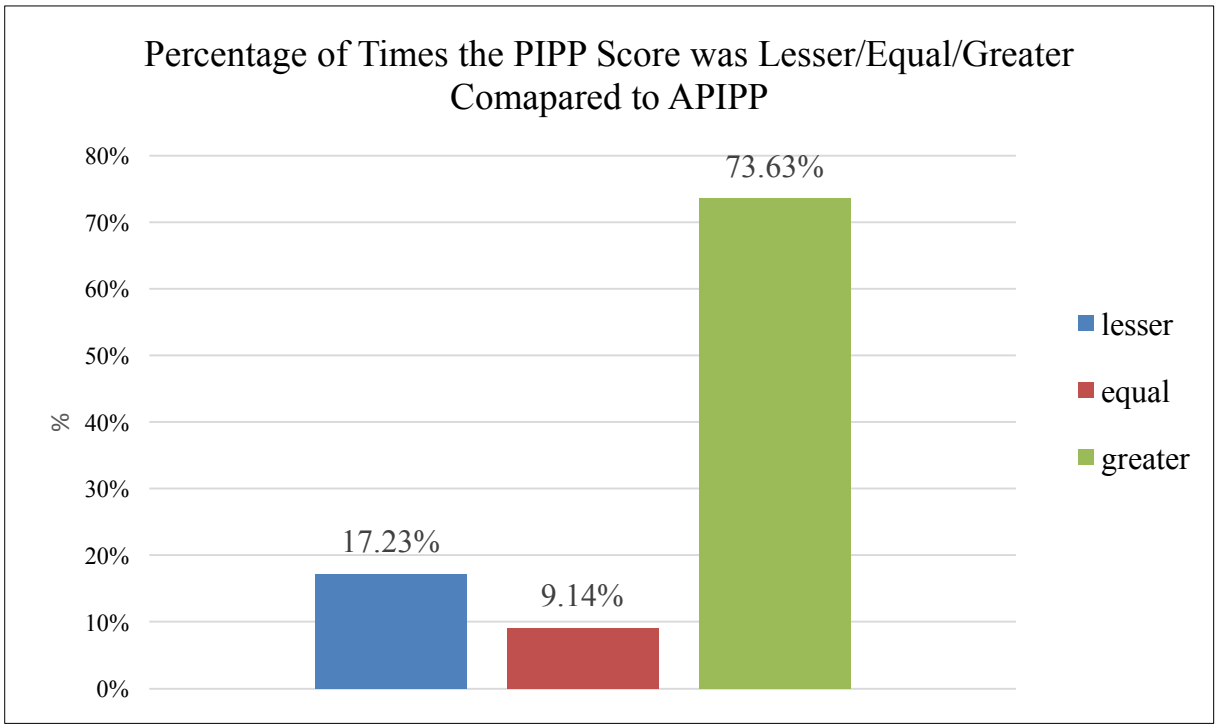


Figure 4. Percentage of times PIPP score was lesser/equal/greater compared to APIPP score for Subject 1 (excluding null)

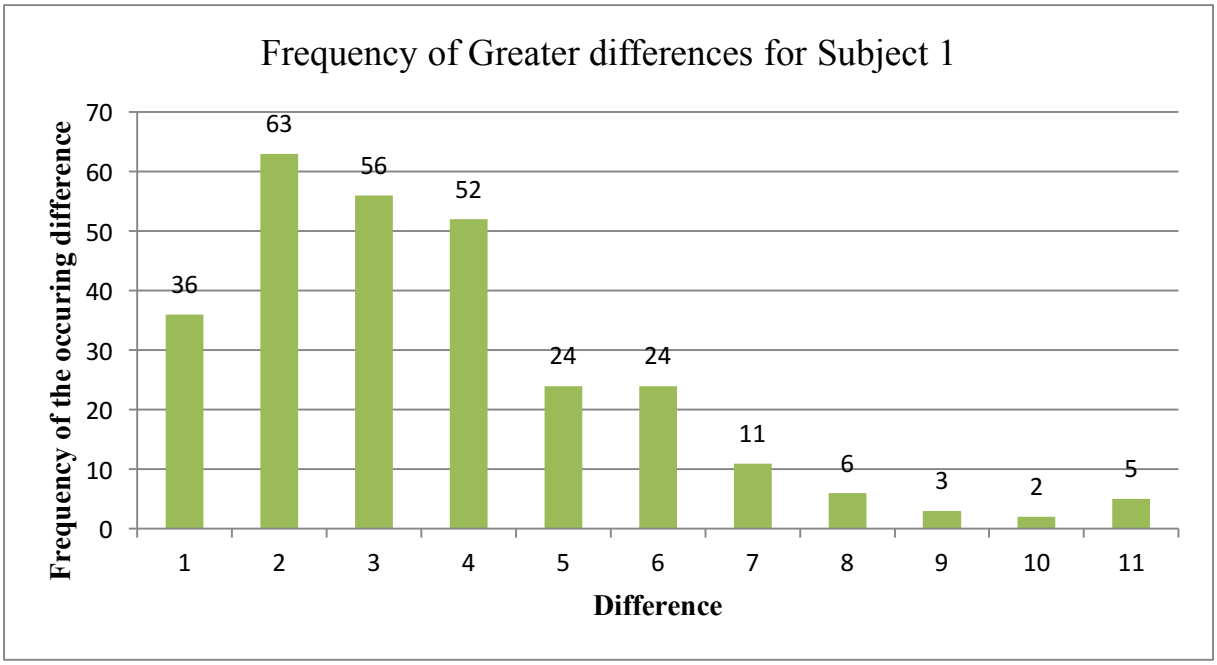


Figure 5. Frequency of the greater category difference for Subject 1

While analysing the greater category of differences, the difference between the PIPP and APIPP score was minimal. As Figure 5 shows, the PIPP score is not greater by a colossal number; the highest occurring difference is a difference of 2. Furthermore, the most occurring differences are between the range of 1-4. The difference of 1-4 is made up of 73% of the data, whereas, the remaining 5-11 difference only makes up 26%. It is important to conclude that most of the hours analysed, only reported a difference in the lower range of 1-4 when compared between PIPP and APIPP. In a 12-point difference between two scales, a difference of 1-4 does not account for a significant difference.

5.2 Evaluation of Artemis Premature Infant Pain Profile (APIPP) Data Model

Testing the APIPP model on a larger cohort of subjects is important in order to evaluate the model. After closely inspecting the data for case study analysis, 13 out of the 23 subjects were excluded due to insufficient data.

The exclusion was mainly due to discontinuous data. Since the data was collected for every hour, it was important to have continuous data for accurate analysis purposes. Thirteen subjects had multiple hours missing throughout the data set, which could have led to inaccurate results. Overall, following exclusion, a total of 2400 hours were analysed, 240 ± 558.5 hours for ten patients. Data are presented as mean ± SD. Table 14 shows the percentage of missing data in this population.

Table 14. Summary of missing data for each subject

Subj_No	DataStart	DataBreak	DataContinues	DataEnds	DataMissing[%]
Subject 1	09/03/2010 20:00			11/03/2010 22:00	0%
Subject 2	08/05/2010 1:00			08/05/2010 12:00	0%
Subject X_1	06/06/2010 20:00	07/06/2010 18:00	08/06/2010 3:00	09/06/2010 19:00	11%
Subject 3	31/10/2010 22:00			08/11/2010 7:00	0%
Subject 4	17/12/2010 8:00			18/12/2010 16:00	0%
Subject X_2	Missing data at multiple points				9%
Subject X_3	Missing data at multiple points				24%
Subject X_4	Missing data at multiple points				26%
Subject 5	03/08/2011 23:00			05/08/2011 5:00	0%
Subject 6	10/10/2011 0:00			10/10/2011 23:00	0%
Subject 7	13/10/2011 22:00			17/10/2011 7:00	0%
Subject X_5	Missing data at multiple points				73%
Subject X_6	02/05/2013 16:00	02/05/2013 22:00	03/05/2013 10:00	05/05/2013 0:00	14%
Subject 8	14/05/2013 21:00			16/05/2013 10:00	0%
Subject X_7	Missing data at multiple points				16%
Subject X_8	17/05/2013 22:00	17/05/2013 22:00	18/05/2013 2:00	18/05/2013 8:00	27%
Subject X_9	Missing data at multiple points				5%
Subject 9	19/05/2013 10:00			22/05/2013 9:00	0%
Subject X_{10}	13/06/2013 21:00			19/06/2013 14:00	1%
Subject X_{11}	Missing data at multiple points				3%
Subject 10	31/05/2013 22:00			02/06/2013 2:00	0%
Subject X_{12}	08/06/2013 12:00	09/06/2013 0:00	09/06/2013 7:00	09/06/2013 7:00	30%
Subject X_{13}	Missing data at multiple points				53%

Subjects with continuous hourly data were considered for analysis. After exclusion, ten subjects: six males and four females, where gestational age was 34 ± 4.8 weeks and birth weight was 2364 ± 870.5 grams, were considered for analysis. Data are presented as mean ± SD. Table 15 shows the background information for ten subjects used in this experiment including gestational age (GA), birth weight (BW) in grams, sex of the subject, year of admission; admission diagnosis and the type of surgery performed if any.

Table 15. Background information for subjects 1-10

Subject ID	GA	BW (g)	Sex	Year	Admission Diagnosis	Comments	Surg
Subject 1	28	680	M	2010	TEF/OA	VAP [March 30, May 2]	1
Subject 2	35	2620	F	2010	Late preterm, gastrochisis, ileal atresia	Repair, resection [May 5]	1
Subject 3	36	2690	M	2010	Hypoxic-ischaemic encephalopathy	Therapeutic hypothermia, no surgery	0
Subject 4	31	2655	F	2010	Prematurity, posthaemorrhagic hydrocephalus	VP sunt insertion [December 16]	1
Subject 5	24	700	F	2011	Haemodynamically significant ductus arteriosus	PDA ligation [August 3]	1
Subject 6	40	3285	F	2011	Anorectal malformation	Colostomy [October 9]	1
Subject 7	36	3100	M	2011	Trisomy 21, ASD, VSD, PPHN	No surgery	0
Subject 8	39	2900	M	2013	TTN, PPHN	No surgery	0
Subject 9	38	2500	M	2013	Anorectal malformation, hypospadias	Colostomy [May 19]	1
Subject 10	33	2510	M	2013	Posterior urethral valves	Valve ablation [May 31]	1

Following the analysis for Subject 1, a similar analysis was conducted with ten subjects to evaluate the APIPP data model. Similar to subject 1 analysis, the difference between the two scores were investigated. Table 16 shows the data for all 10 subjects for all four categories.

Table 16. Ten subject's data for four difference categories

	Lesser (%)	Equal (%)	Greater (%)	Null (%)
Subject 1	3.46	1.83	14.76	79.95
Subject 2	9.09	9.09	9.09	72.73
Subject 3	1.69	1.13	11.30	85.88
Subject 4	6.25	3.13	15.63	75.00
Subject 5	6.67	3.33	23.33	66.67
Subject 6	0	4.17	4.17	91.67
Subject 7	2.47	0	12.35	85.19
Subject 8	5.41	5.41	2.70	86.49
Subject 9	1.43	4.29	25.71	68.57
Subject 10	3.57	3.57	21.43	71.43
AVG	**4.00**	**3.60**	**14.05**	**78.36**
STDEV	2.66	2.38	7.36	8.20

Similar results as subject 1 analysis were found after evaluating the data. For 78.36% ± 8.2 of the time, the PIPP score was not available to be analysed while the APIPP score was available. For 14.05% ± 7.4 of the time, the PIPP score was higher than the APIPP score. For 4% ± 2.7 of the time, the PIPP score was less than the APIPP score. Lastly, for 3.60% ± 2.4 of the time, the PIPP

score was equal to the APIPP score. Data are presented as mean ± SD. The results presented a similar pattern to the results of subject 1 as each category of differences resulted in a similar classification.

Table 17. Classification of percentages for four difference categories for Subject 1 vs. 10 subjects

Difference Category	Subject 1	10 subjects (Avg)
Null	79.95%	78.36%
Greater	14.74%	14.05%
Lesser	3.46%	4.00%
Equal	1.83%	3.60%

The following figure 6 displays the data for 10 subjects for all four categories.

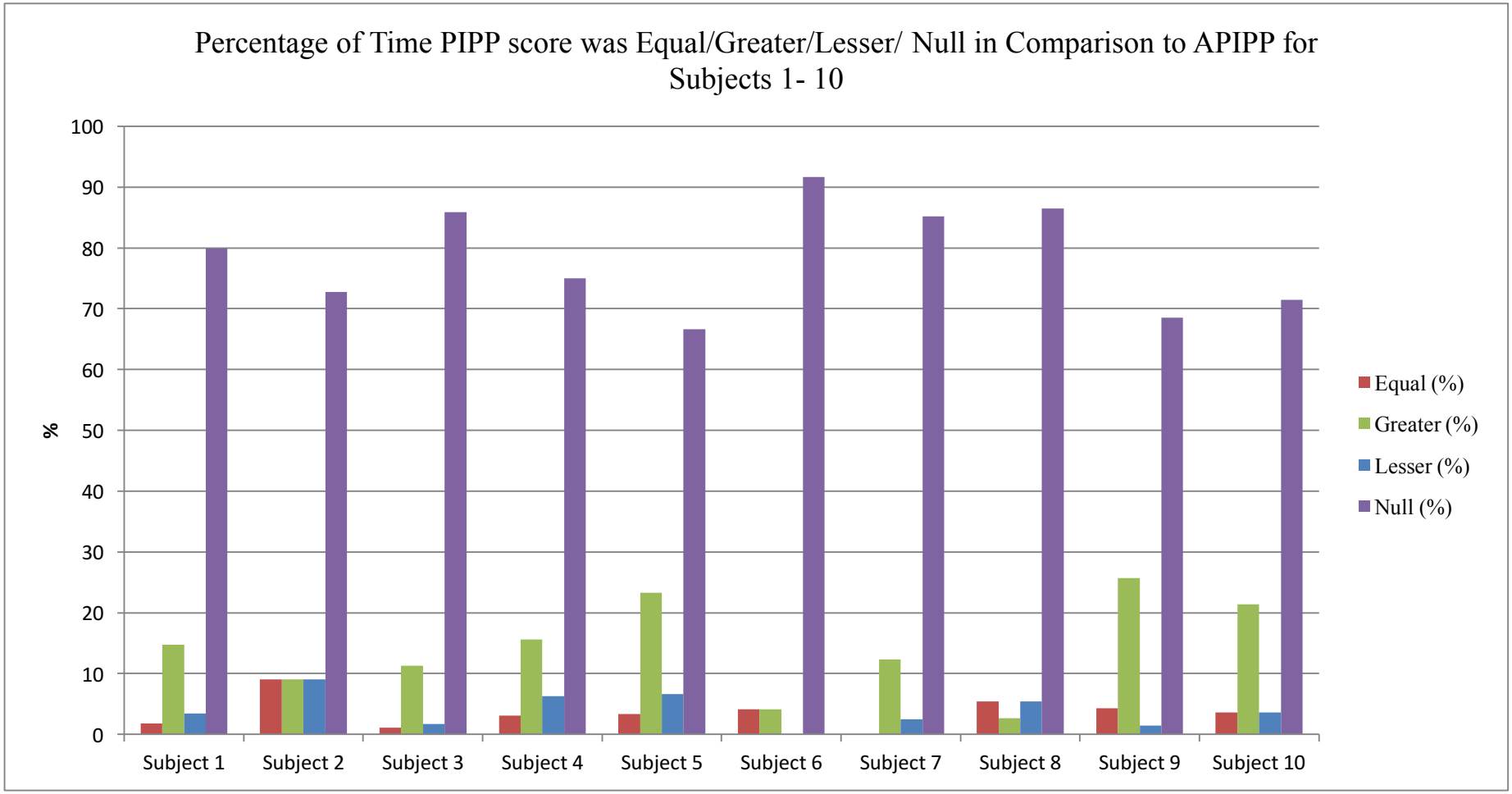


Figure 6. Percentage of times PIPP score was equal/greater/lesser/null in comparison to APIPP score for Subjects 1-10

Additionally, it was important to exclude the null (blank) data and analyse the subjects based on the greater, lesser and equal categories. Table 18 displays the results.

Table 18. Ten subject data for greater, lesser and equal categories (excluding null category)

	Greater (%)	Lesser (%)	Equal (%)
Subject 1	73.63	17.23	9.14
Subject 2	33.33	33.33	33.33
Subject 3	80.00	12.00	8.00
Subject 4	62.50	25.00	12.50
Subject 5	70.00	20.00	10.00
Subject 6	50.00	0.00	50.00
Subject 7	83.33	16.67	0.00
Subject 8	20.00	40.00	40.00
Subject 9	81.82	4.55	13.64
Subject 10	75.00	12.50	12.50
AVG	**62.96**	**18.13**	**18.91**
STDEV	20.64	11.60	15.43

Table 19. Classification of percentages for three difference categories for Subject 1 vs. 10 subjects (excluding null category)

Difference Category	Subject 1	10 Subjects (Avg)
Greater	73.63%	62.96%
Lesser	17.23%	18.13%
Equal	9.14%	18.91%

These results propose a similar pattern to that of subject 1, where the PIPP score was greater than the APIPP score for majority of the hours. Of 62.96% ± 20.6 of the study population, the PIPP score is greater. For 18.13% ± 11.6 of the population, the PIPP score is less than APIPP. Lastly, for 18.91% ± 15.4 of the population, the PIPP score is equal to APIPP. Data are presented as mean ± SD. Figure 7 displays the data for each subject based on these three categories.

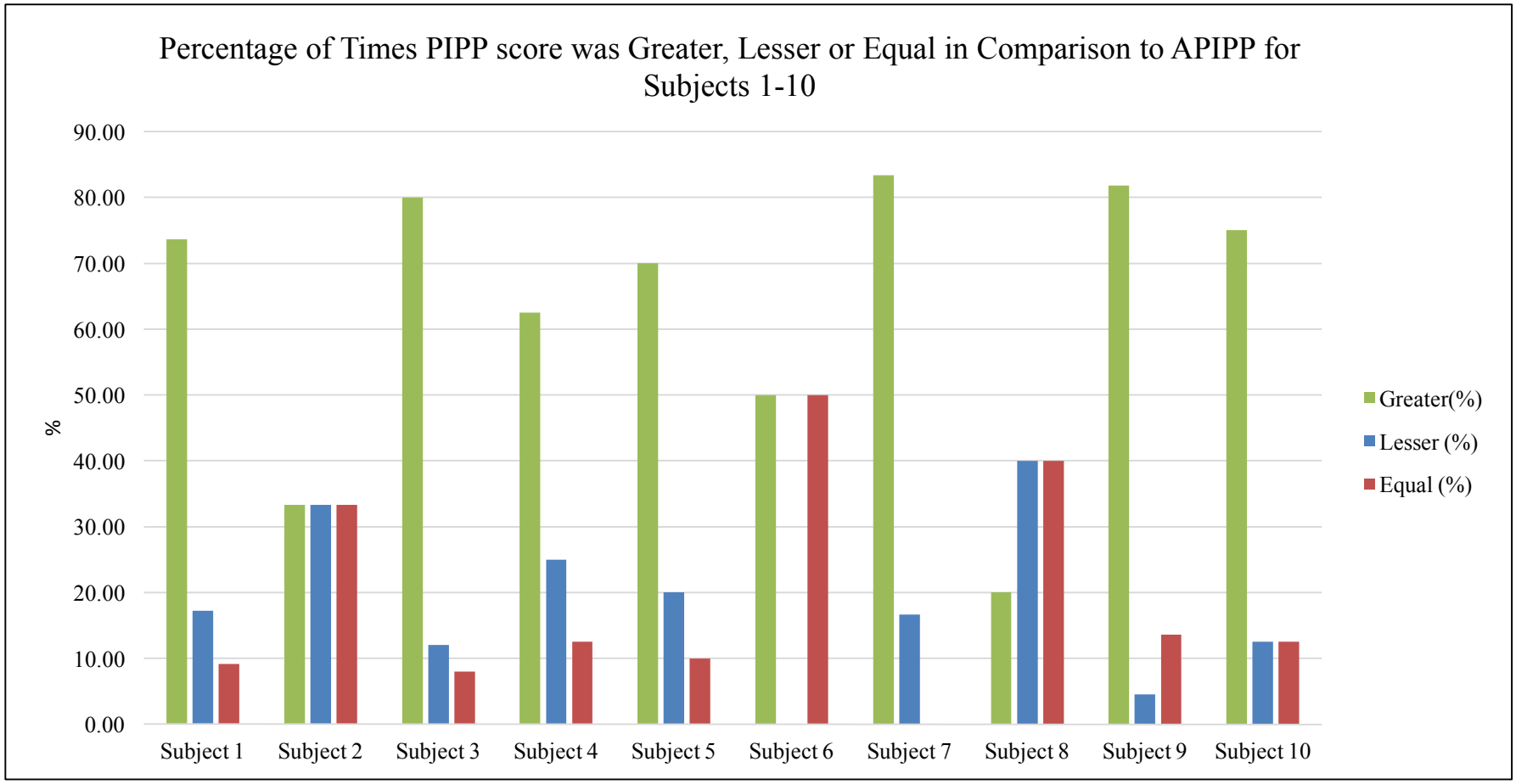


Figure 7. Percentage of Times PIPP score was Greater, Lesser or Equal in Comparison to APIPP for Subjects 1-10

Once again, analysing the greater differences between PIPP and APIPP scores for all 10 subjects was imperative. Table 20 displays the sum of times each difference occurred across all 10 subjects in the greater category.

Table 20. Sum of times each greater difference occurred for subjects 1-10

Difference	S 1	S 2	S 3	S 4	S 5	S 6	S 7	S 8	S 9	S 10	TOTAL
1	36	1	7	1			1		1	1	**48**
2	63		3	2	2		3		1	4	**78**
3	56		4	2	3	1	1		5		**72**
4	52		4					1	5		**62**
5	24		2		1		3		3	1	**34**
6	24						2		2		**28**
7	11				1				1		**13**
8	6										**6**
9	3										**3**
10	2										**2**
11	5										**5**

As can be seen by Table 20 and Figure 8, the data for subject 1-10 presents a similar result to that of subject 1 alone. Similar to subject 1, subjects 1- 10 have a high frequency of times occurring differences in the lower range of 1-4. Similar to subject 1, the most occurring differences are two and three. The greater occurring differences are mostly below the difference of six.

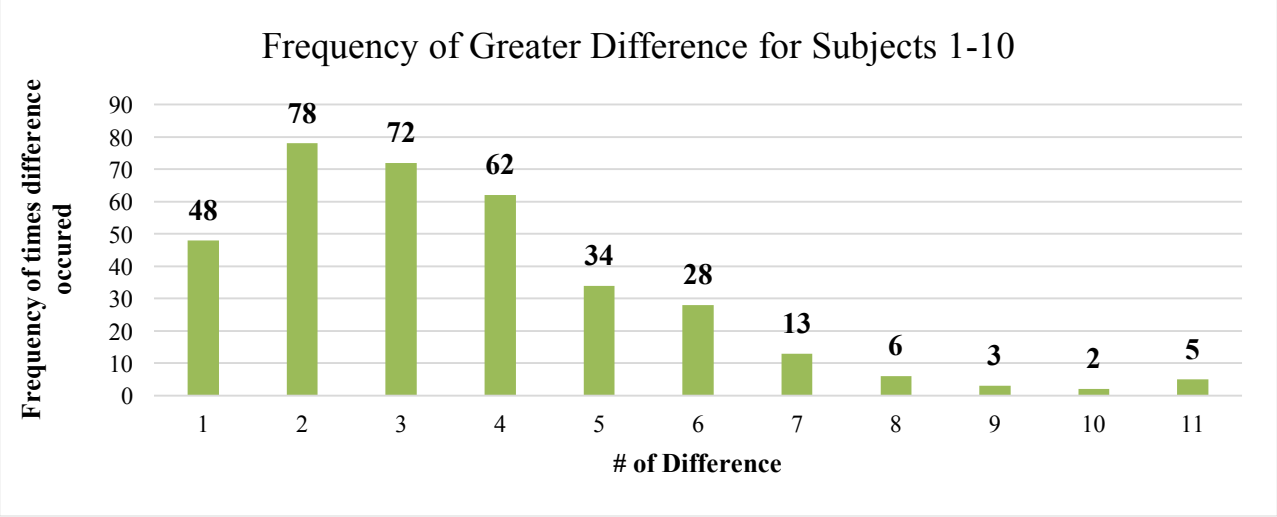


Figure 8. Frequency of Greater Difference for Subjects 1-10

In this case, it was important to note the number of hours analyzed for each subject as this can have a great impact on the results. A total of 2400 hours were analysed for 10 subjects. The average for this population was 240 hours and the standard deviation was 558. hours. The standard deviation is very large, indicating that the number of hours between these subjects is very spread out. Since Subject 1 had a large number of hours (1910) for which data was available, analyzing the data without subject 1 is important. After excluding subject 1 from the dataset, the total number of hours analyzed was 490. The average number of hours for this population of 9 subjects was 54.4 hours and the standard deviation was 48.17 hours. In this case, since the standard deviation is closer to the mean, the number of hours analysed for all 9 subjects were closer to the mean.

The data for 9 subjects was analyzed and graphed. Figure 9 displays the results for the frequency of differences without subject 1. Once again, the most occurring differences were those on the lower side, with the difference of three being the highest occurring difference with a total of sixteen times. Excluding subject one also eliminated the differences ranging between 8 and 11.

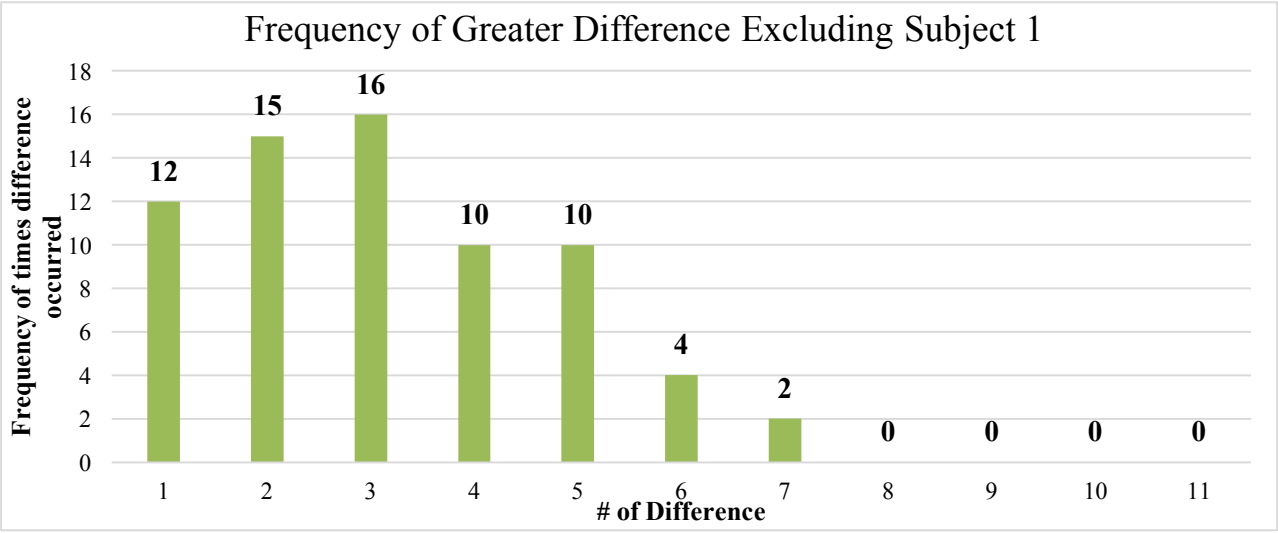


Figure 9. Frequency of Greater Differences Excluding Subject #1

Overall, in this chapter, APIPP was calculated and compared to PIPP in order to reflect on the potential efficacy of the APIPP model. Following this, the model was evaluated to examine how the APIPP compared to the already established scoring system of PIPP. The model was evaluated by conducting an analysis on both the scores to compare if the PIPP score was greater, lesser or equal to that of APIPP.

This analysis was first computed on subject 1 as a case study. Following this, for the purposes of evaluating the model, 10 subjects' data was also analyzed by implicating similar analysis on this subset of subjects to assess the model. Both analysis presented similar results, where majority of hours were presented as null (blank) for PIPP scores when APIPP score was present. Such results showed that the PIPP score was not collected for the majority of the time during which the APIPP score was available and scored. These null values were excluded and only the available values for PIPP were compared to APIPP score for a more reliable comparison. Additionally, for the majority of the time, the PIPP score was found to be greater than the APIPP score. Such a finding was expected as the PIPP score was scored out of 21 whereas the APIPP score was scored out of 9. For this reason, computing the difference between the two scores was important. On average, for both analyses, the difference between scores was less than six with a difference of two and three being more prominent in both analyses. Such differences between the two scores were negligible. Such a finding is noteworthy as PIPP scoring has a much higher score compared to the APIPP score. Such an analysis shows that the PIPP score must be highly scored based on physiological parameters as the difference between the two scores is minimal despite the difference between them. Conducting future research to investigate these findings is important. Chapter 6 presents a discussion from the clinical perspective in the context of the literature review preformed and presented in chapter 2. Chapter

6 will discuss the clinical side of using physiological parameters as indicators for pain. Since heart rate is the most commonly used physiological indicator, heart rate will be used for clinical analysis with the PIPP scale.

Chapter 6 - Physiological Data Analysis

Following the case study and evaluation of data model in chapter 5, the relationship between physiological parameter (heart rate) and clinical data such as gestational age, admission diagnosis and surgery, will be explored in this chapter. This chapter attempts to investigate the clinical aspect of pain management. Heart rate will be used as a base physiological parameter as heart rate is widely used in the clinical field. Heart rate will be analyzed in accordance with various other clinical indicators and PIPP scale to explore it as a reliable measure for designing a computational automatic score in the future to quantify pain in the premature infant. By exploring this clinical aspect through various experiments, such an analysis will provide an overview to create objective tools to measure pain in the neonatal population in the future. Additionally, this analysis will help in the assessment as well as management of pain relief, avoiding the consequences that occur due to untreated pain.

Similar to chapter 5, the data of ten subjects was used to conduct the experiments in this chapter. The subjects' consisted of six males and four females, whose mean gestational age was 34 ± 4.8 weeks and mean birth weight was 2364 ± 870.5 grams. Data are presented as mean ± SD. Table 21 outlines the background information for each patient, including gestational age (GA), sex of the subject, year of admission; admission diagnosis and the type of surgery performed if any.

Table 21. Background information for Subject 1-10

Subject ID	GA	BW (g)	Sex	Year	Admission Diagnosis	Comments	Surg
Subject 1	28	680	M	2010	TEF/OA	VAP [March 30, May 2]	1
Subject 2	35	2620	F	2010	Late preterm, gastrochisis, ileal atresia	Repair, resection [May 5]	1
Subject 3	36	2690	M	2010	Hypoxic-ischaemic encephalopathy	Therapeutic hypothermia, no surgery	0
Subject 4	31	2655	F	2010	Prematurity, posthaemorrhagic hydrocephalus	VP sunt insertion [December 16]	1
Subject 5	24	700	F	2011	Haemodynamically significant ductus arteriosus	PDA ligation [August 3]	1
Subject 6	40	3285	F	2011	Anorectal malformation	Colostomy [October 9]	1
Subject 7	36	3100	M	2011	Trisomy 21, ASD, VSD, PPHN	No surgery	0
Subject 8	39	2900	M	2013	TTN, PPHN	No surgery	0
Subject 9	38	2500	M	2013	Anorectal malformation, hypospadias	Colostomy [May 19]	1
Subject 10	33	2510	M	2013	Posterior urethral valves	Valve ablation [May 31]	1

6.1 Statistics

For statistical analysis within the heart rate domain, heart rate of ten subjects was plotted based on three different categories on three different graphs for heart rate average, maximum and minimum using Graphpad Prism®. For each subject, an average was computed within the hourly data set of each subject. After this, each average for the three categories was plotted on a graph for all ten subjects. The graph displays the heart rate average for each subject and their standard deviation. Figure 10, 11 and 12 displays the results for heart rate average, maximum and minimum, respectively.

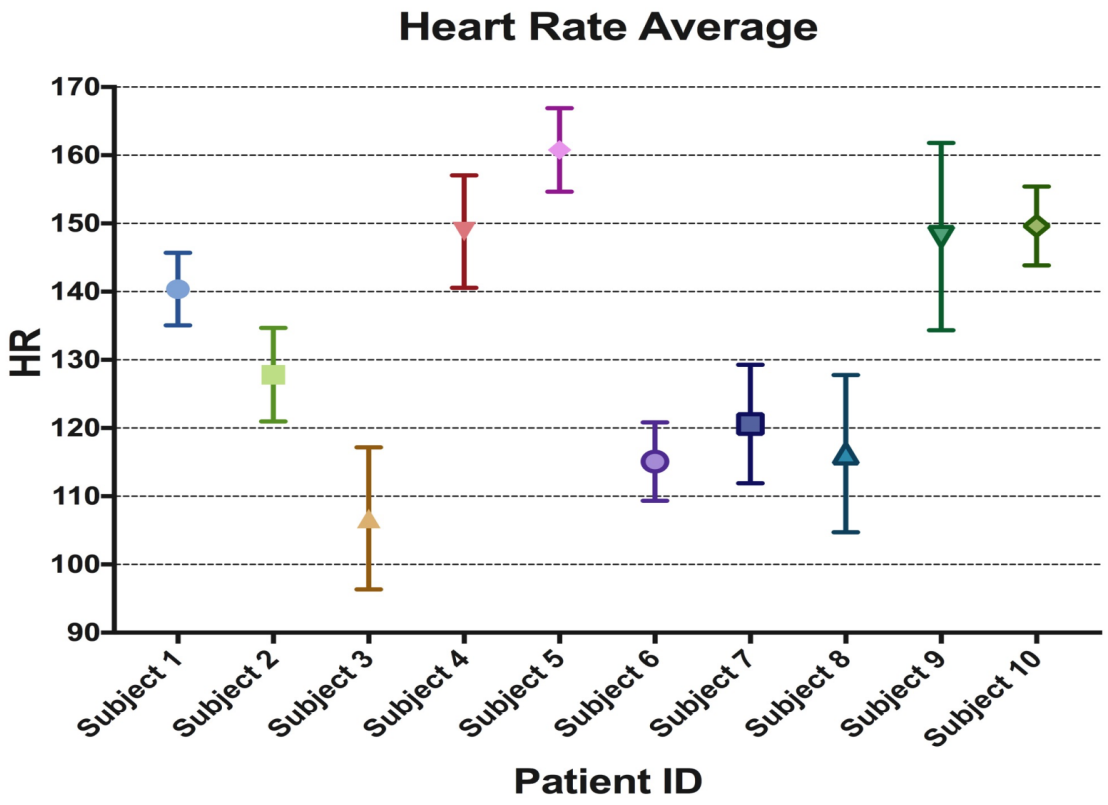


Figure 10. Heart Rate Average data for subjects 1-10

After analysing the heart rate average for ten patients, the non-surgical patients were found to have lower mean heart rate compared to the subjects who had surgery and presented with a severe admission diagnosis. As seen in Figure 10, Subject 3, 6, 7 and 8 are in the lower heart rate range. Three out of four of these subjects did not undergo surgery and hence have a lower heart rate average. Subject 6 was the only subject to have a surgery (colostomy) in this group. This subject is also the oldest subject in this data set, who is 40 weeks of gestational age. The subject's gestational age may be associated with the heart rate being lower despite the subject having a previous surgery. Subjects 1, 2, 4, 5, 9 and 10 underwent surgeries, and therefore their heart rate is within the higher range, as the heart rate is over 130. Subject 5, who had an admission diagnosis of significant ductus arteriosus and had a corresponding PDA ligation surgery, shows the highest heart rate. Subject 5 was one of the subjects with serious

comorbidities that required an intensive surgery. In addition, Subject 5 was also extremely preterm with a gestational age of 24, which is the lowest in this group. Such findings show a pattern between heart rate, gestational age, surgical procedures and illness. This pattern of heart rate averages demonstrates the potential utilization for heart rate as a measure within the pain domain.

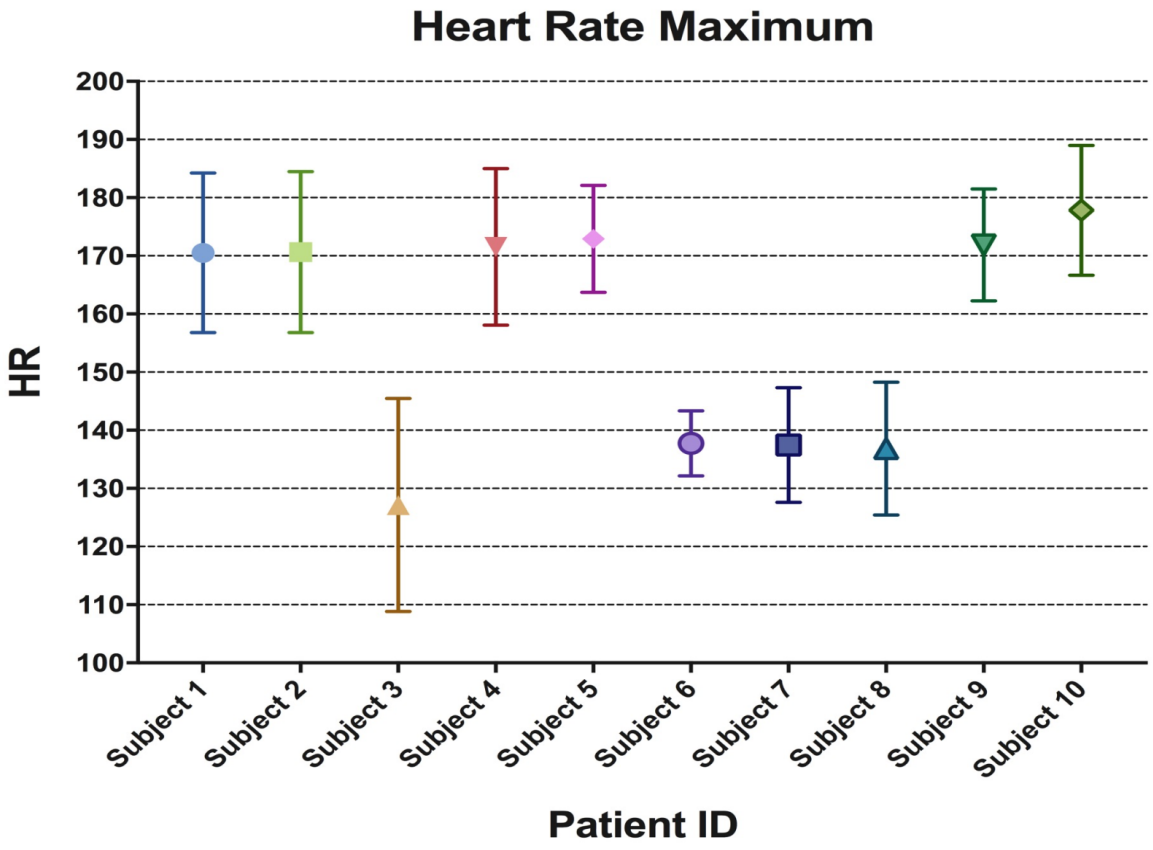


Figure 11. Maximum Heart Rate Data for Subjects 1-10

Figure 11 displays the results for maximum heart rate of the 10 subjects. This figure depicts a similar result to the figure for heart rate average. However, compared to the average heart rate data, the heart rate maximum data is more consistent. Once again, subjects 3, 6, 7 and 8 have a lower maximum heart rate, where most subjects have a heart rate of under 140. Among the four subjects, three subjects did not have a surgery, while subject 6, who is the oldest subject, had a

surgery. Hence the heart rates of these four subjects were not very elevated in comparison to the other subjects. Subjects 1, 2, 4, 5, 9, and 10 are displaying a much higher maximum heart rate, with most subjects having a heart rate of more than 170. A consistent trend is observed with the maximum heart rate data, which displays maximum heart rate of 170 for most of the subjects. As such, a pattern is found between heart rate and illness.

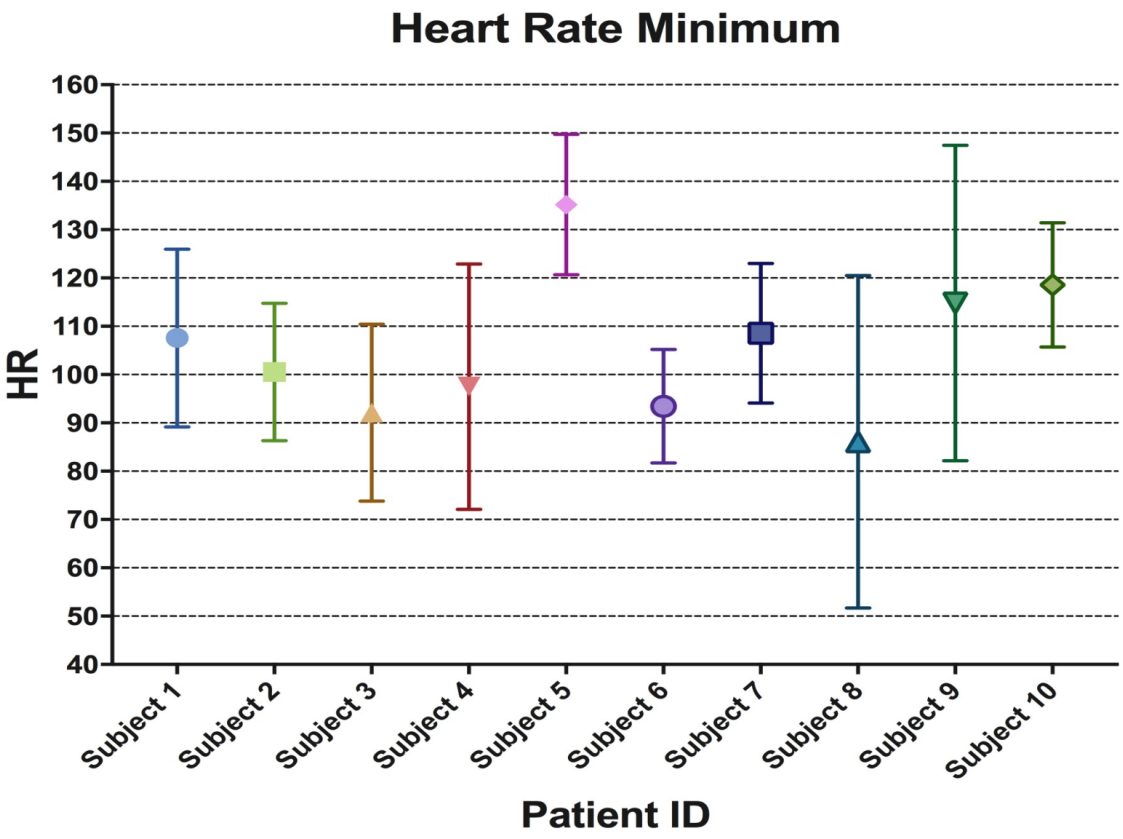


Figure 12. Minimum Heart Rate Data for Subjects 1-10

Figure 12 displays the results for minimum heart rate. Most of the subjects have a fairly similar minimum heart rate. However, subject 5 shows a much higher minimum heart rate compared to other subjects. Subject 5 is also the youngest and most premature neonate with a gestational age of 24 who underwent a serious surgery (PDA ligation) for a serious comorbidity of significant

ductus arteriosus. In common clinical practice, the nurses monitor the heart rate in a manner where heart rate does not decline to an anomalistic rate as such a heart rate can hinder the stability of the neonate, which is reflected in the minimum heart rate data presented in Figure 12. Figure 12 shows that the minimum heart rate is similar for most subjects. Thus, the minimum heart rate is shown to be a less useful marker for assessing pain.

6.2 ANOVA

In this experiment, a one-way analysis of variance ANOVA was used to determine if any significant differences existed between the heart rate average, minimum and maximum among the ten subjects. One may assume that heart rate does differ from individual to individual; however patterns can be observed for resting, normal and exercising, which in turn can help to establish heart rate patterns for healthy individuals. Similarly, this test aims to explore the hypothesis that a pattern exists among premature infants based on gestational age, illness or pain. This test was also used to determine whether any significant differences exist between the means of three or more patients.

Population: Ten subjects. One-way ANOVA was performed using Microsoft Excel®. Specifically, this test tests the null hypothesis, which states that there are no differences between the means of this group of subjects. If the null hypothesis is rejected, the alternative hypothesis is accepted, which would show that at least two subjects are significantly different from each other. After conducting the one-way ANOVA test with heart rate averages, a statistically significant difference was found between groups as determined ($F(9, 531) = 220.33$, $p = 6.13E\text{-}173$). Table 22 displays the results.

Table 22. ANOVA heart rate average

SUMMARY						
Groups	**Count**	**Sum**	**Average**	**Variance**		
Subject 1	51	7183	140.85	28.722		
Subject 2	11	1412	128.36	46.486		
Subject 3	177	18979	107.23	108.571		
Subject 4	32	4776	149.25	68.252		
Subject 5	30	4838	161.28	38.021		
Subject 6	24	2772	115.51	32.072		
Subject 7	81	9806	121.06	75.933		
Subject 8	37	4301	116.24	132.911		
Subject 9	70	10365	148.07	188.908		
Subject 10	28	4190	149.64	33.423		
ANOVA						
Source of Variation	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***
Between Groups	185835.8	9	20648.43	220.3351	6.13E-173	1.897504
Within Groups	49762	531	93.71375			
Total	235597.8	540				

After conducting the one-way ANOVA test with heart rate maximum, a statistically significant difference was found between groups (F (9, 531) = 131.1, p = 8.80E-129). The results are presented in Table 23.

Table 23. ANOVA heart rate maximum

SUMMARY				
Groups	**Count**	**Sum**	**Average**	**Variance**
Subject 1	51	8696	170.51	188.495
Subject 2	11	1877	170.64	191.255
Subject 3	177	22506	127.15	335.891
Subject 4	32	5489	171.53	181.289
Subject 5	30	5188	172.93	84.478
Subject 6	24	3306	137.75	31.065
Subject 7	81	11135	137.47	97.302
Subject 8	37	5064	136.86	130.342

Subject 9	70	12030	171.86	92.559		
Subject 10	28	4979	177.82	125.041		
ANOVA						
Source of Variation	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***
Between Groups	225532.89	9	25059.21	131.1268	8.80E-129	1.897504
Within Groups	101477.68	531	191.11			
Total	327010.58	540				

After conducting the one-way ANOVA test with heart rate minimum, a statistically significant difference was found between groups as determined by one-way ANOVA (F (9, 531) = 120.7, *p* = 4.61E-30). The results are presented in Table 24.

Table 24. ANOVA heart rate minimum

SUMMARY						
Groups	**Count**	**Sum**	**Average**	**Variance**		
Subject 1	51	5486	107.57	338.53		
Subject 2	11	1106	100.55	202.473		
Subject 3	177	16307	92.13	335.455		
Subject 4	32	3120	97.5	643.677		
Subject 5	30	4056	135.2	210.786		
Subject 6	24	2243	93.46	137.824		
Subject 7	81	8795	108.58	208.697		
Subject 8	37	3186	86.11	1183.21		
Subject 9	70	8037	114.81	1064.588		
Subject 10	28	3320	118.57	164.995		
ANOVA						
Source of Variation	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***
Between Groups	85933.96	9	9548.22	20.74249	4.61E-30	1.897504
Within Groups	244430.75	531	460.32			
Total	330364.7	540				

The results for each of the three categories (average HR, maximum HR, and minimum HR) show that a significant difference exists. Although a one-way ANOVA can indicate that a significant

difference exists between two groups, this test is unable to specify which specific groups were significantly different from each other. To determine which specific groups differed from each other, the Tukey Kramer test was conducted. This test is performed following an analysis of variance (ANOVA) test.

6.3 Tukey Kramer Test

Following the one-way ANOVA, the Tukey Kramer test was performed as the one-way ANOVA test for all three experiments showed that there are significant differences between subjects. The purpose of Tukey Kramer test is to determine which groups in the sample differ. This test compares all possible pairs of means and is based on a studentised range distribution (q). If the results of ANOVA are positive (i.e. a significant difference is shown among groups), a Tukey Kramer test can be performed. It is not likely that all groups differ when compared to each other, but that only some have significant differences. The Tukey Kramer test helps to clarify which specific groups among the sample have significant differences. This test revealed where the significant differences were between these subjects. Three different Tukey Kramer test were manually preformed in Microsoft Excel® with the data for ten subjects based on heart rate average, minimum and maximum.

To compute the Tukey test, the critical value is very important. The critical value is used to evaluate whether differences between any two pairs of means are significant. The critical value also involves the absolute difference that has to be exceeded to achieve significance. For this, all possible pairs of comparisons were listed. The # of comparisons are calculated through equation 5 (n= # of groups):

$$\frac{n(n-1)}{2} \qquad \textit{Equation 4}$$

$$\frac{10(10-1)}{2} = 45$$

In this experiment, forty-five comparisons were required to compute the test. After this, the absolute difference is calculated through equation 6 (x = mean):

$$\bar{x}_{subject\ 1} - \bar{x}_{Subject\ 2} \qquad \textit{Equation 5}$$

Following this, the critical value was calculated using Equation 7. For this, the q value was found through the critical values of the studentised range table with 0.05. *MS* value was taken from the ANOVA table. *n* represented the total # of observations for each subject. Since the data for all subjects were unequal, the lesser observations out of the two subjects were used.

$$q\frac{\sqrt{MS}}{n} \qquad \textit{Equation 6}$$

Lastly, if the absolute difference was larger than the critical value, the comparison was found to be significantly different. When the Tukey Kramer test was conducted for heart rate average, a total of 45 comparisons were outputted, of which 11 comparisons showed a non-significant result and 34 showed significantly different results. Table 25 presents the results.

Table 25. Tukey Kramer Test- heart rate average

Tukey Kramer Test for Average Heart Rate			
Comparison	**Absolute Difference**	**Critical Range**	**Results**
Subject 1 vs. Subject 2	12.49	13.0587	not significantly different
Subject 1 vs. Subject 3	33.63	6.0647	significantly different
Subject 1 vs. Subject 4	8.4	7.6564	significantly different
Subject 1 vs. Subject 5	20.43	7.9075	significantly different
Subject 1 vs. Subject 6	25.34	8.8408	significantly different
Subject 1 vs. Subject 7	19.79	6.0647	significantly different
Subject 1 vs. Subject 8	24.61	7.1203	significantly different
Subject 1 vs. Subject 9	7.22	6.0647	significantly different
Subject 1 vs. Subject 10	8.79	8.185	significantly different

Subject 2 vs. Subject 3	21.13	13.0587	significantly different
Subject 2 vs. Subject 4	20.89	13.0587	significantly different
Subject 2 vs. Subject 5	32.92	13.0587	significantly different
Subject 2 vs. Subject 6	12.85	13.0587	not significantly different
Subject 2 vs. Subject 7	7.3	13.0587	not significantly different
Subject 2 vs. Subject 8	12.12	13.0587	not significantly different
Subject 2 vs. Subject 9	19.71	13.0587	significantly different
Subject 2 vs. Subject 10	21.28	13.0587	significantly different
Subject 3 vs. Subject 4	42.02	7.6564	significantly different
Subject 3 vs. Subject 5	54.05	7.9075	significantly different
Subject 3 vs. Subject 6	8.29	8.8408	not significantly different
Subject 3 vs. Subject 7	13.83	4.8123	significantly different
Subject 3 vs. Subject 8	9.02	7.1203	significantly different
Subject 3 vs. Subject 9	40.84	5.1766	significantly different
Subject 3 vs. Subject 10	42.42	8.185	significantly different
Subject 4 vs. Subject 5	12.03	7.9075	significantly different
Subject 4 vs. Subject 6	33.74	8.8408	significantly different
Subject 4 vs. Subject 7	28.19	7.6564	significantly different
Subject 4 vs. Subject 8	33.01	7.6564	significantly different
Subject 4 vs. Subject 9	1.18	7.6564	not significantly different
Subject 4 vs. Subject 10	0.39	8.185	not significantly different
Subject 5 vs. Subject 6	45.77	8.8408	significantly different
Subject 5 vs. Subject 7	40.22	7.9075	significantly different
Subject 5 vs. Subject 8	45.04	7.9075	significantly different
Subject 5 vs. Subject 9	13.21	7.9075	significantly different
Subject 5 vs. Subject 10	11.64	8.185	significantly different
Subject 6 vs. Subject 7	5.55	8.8408	not significantly different
Subject 6 vs. Subject 8	0.73	8.8408	not significantly different
Subject 6 vs. Subject 9	32.56	8.8408	significantly different
Subject 6 vs. Subject 10	34.13	8.8408	significantly different
Subject 7 vs. Subject 8	4.82	7.1203	not significantly different
Subject 7 vs. Subject 9	27.01	5.1766	significantly different
Subject 7 vs. Subject 10	28.58	8.185	significantly different
Subject 8 vs. Subject 9	31.83	7.1203	significantly different
Subject 8 vs. Subject 10	33.4	8.185	significantly different
Subject 9 vs. Subject 10	1.57	8.185	not significantly different

The results of Tukey Kramer Test for heart rate average showed eleven comparisons as 'not significantly different'. Table 26 presents the results for the 'not significantly different' results. The results illustrate a similarity between the average heart rate of these subjects. Based on this evidence, the clinical record and notes of the subjects were further analysed to investigate the similarities and/or pattern.

The clinical record of the subjects showed a similarity regarding the surgical procedures that were performed during their stay in the NICU. Seven pairs of subjects who underwent surgical procedures presented 'not significantly different result' out of which one group of subjects did not undergo any surgical procedure. The analysis found that the subjects who showed a 'not significantly different' result were the subjects who presented with a severe illness in their admission diagnosis.

Table 26. 'Not significantly different' subject group

Comparison	Absolute Difference	Critical Range	Results	Surgery (1=Y 0=N)
Subject 1 vs. Subject 2	12.49	13.0587	Not significantly different	1,1
Subject 2 vs. Subject 6	12.85	13.0587	Not significantly different	1,1
Subject 2 vs. Subject 7	7.3	13.0587	Not significantly different	1,0
Subject 2 vs. Subject 8	12.12	13.0587	Not significantly different	1,0
Subject 3 vs. Subject 6	8.29	8.8408	Not significantly different	0,1
Subject 4 vs. Subject 9	1.18	7.6564	Not significantly different	1,1
Subject 4 vs. Subject 10	0.39	8.185	Not significantly different	1,1
Subject 6 vs. Subject 7	5.55	8.8408	Not significantly different	1,0
Subject 6 vs. Subject 8	0.73	8.8408	Not significantly different	1,0
Subject 7 vs. Subject 8	4.82	7.1203	Not significantly different	0,0

Subject 9 vs. Subject 10	1.57	8.185	Not significantly different	1,1

a) 1,0: One patient with surgery and one without having underwent surgery

The subject groups in this category involved a subject who had a surgery and the other who did not undergo surgery. However, after analysis, subjects who did not have surgery were found to have severe illnesses. As such, the subjects who did not have surgery presented similar results to those that did have surgery. There were five subject groups who were in this category.

Subject 2 vs. Subject 8: Subject 2 is a late preterm infant at 35 GA with admission diagnosis of gastrochisis and ileal atresia who had a repair and resection surgery. Subject 8 was 39 GA who did not undergo surgery. However, his admission diagnosis was Transient Tachypnea of the Newborn (TTN) and Persistent Pulmonary Hypertension of the Newborn (PPHN). Such a diagnosis suggests that the subject was very sick.

Subject 2 vs. Subject 7: Subject 2 is a late preterm infant at 35 GA with admission diagnosis of gastrochisis and ileal atresia who had a repair and resection surgery. Subject 7 was 36 GA who did not undergo surgery; however, his admission diagnosis illustrated severe sickness such as Trisomy 21, Atrial Septal Defect (ASD), Ventricular Septal Defect (VSD) and Persistent Pulmonary Hypertension of the Newborn (PPHN).

Subject 3 vs. subject 6: Subject 3 is a late preterm infant at 36 GA who did not undergo surgery; however, the subject presented with hypoxic-ischaemic encephalopathy (HIE), which is a severe illness where the infant's brain does not receive enough oxygen and

blood. Studies have found that patients with HIE present lower heart rate variability compared to healthy infants. Subject 6, on the other hand, did have a colostomy surgery.

Subject 6 vs. subject 7: Subject 6 was a term infant at 40 GA who had a colostomy surgery. Subject 7 was a moderate to late preterm infant at 36 GA who did not undergo surgery; however, his admission diagnosis indicated severe comorbidities including Trisomy 21, Atrial Septal Defect (ASD), Ventricular Septal Defect (VSD) and Persistent Pulmonary Hypertension of the Newborn (PPHN).

Subject 6 vs. subject 8: Subject 6 was a term infant at 40 GA who had a colostomy surgery. Subject 8 was also a term infant at 39 GA who did not undergo surgery. However, his admission diagnosis was Transient Tachypnea of the Newborn (TTN) and Persistent Pulmonary Hypertension of the Newborn (PPHN). This is suggestive of severe illness.

b) 1,1: Both subjects with surgery

The subject groups involved in this category all had surgeries; due to this they outputted a 'not significantly different result' amongst them. There were five subject groups involved in this category that include: Subject 1 vs. subject 2; Subject 2 vs. Subject 6; Subject 4 vs. subject 9; Subject 4 vs. subject 10; Subject 9 vs. subject 10.

c) 0,0: No subjects with surgery

Subject 7 vs. subject 8: Even though both these subjects did not have surgery, they were both significantly sick compared to other subjects. Subject 7 was 36 GA who did not undergo surgery; however, his admission diagnosis showed severe comorbidities such as

Trisomy 21, Atrial Septal Defect (ASD), Ventricular Septal Defect (VSD) and Persistent Pulmonary Hypertension of the Newborn (PPHN). Subject 8 was 39 GA with an admission diagnosis of Transient Tachypnea of Newborn (TTN) and Persistent Pulmonary Hypertension of the Newborn (PPHN). Both subjects had PPHN, which could be a potential reason as to why differences between the subjects were not significant.

In conclusion, the heart rate average Tukey test shows that surgery, severity of illness, and gestational age are factors that can contribute to a similar pattern of heart rate. This analysis shows that the subjects with serious comorbidities are similar in their heart rate averages and therefore show a 'not significantly different' result between them. Severity of illness has been explored as an influence on pain scores among preterm infants in many studies (Johnston et al., 1999; Stevens et al., 1999; Stevens, Johnston, & Horton, 1994).

Table 27. Tukey Kramer Test- maximum heart rate

Tukey Kramer Procedure for Maximum Heart Rate			
Comparison	**Absolute Difference**	**Critical Range**	**Results**
Subject 1 vs. Subject 2	0.13	18.6482	not significantly different
Subject 1 vs. Subject 3	43.36	8.6606	significantly different
Subject 1 vs. Subject 4	1.02	10.9335	not significantly different
Subject 1 vs. Subject 5	2.42	11.2921	not significantly different
Subject 1 vs. Subject 6	32.76	12.6249	significantly different
Subject 1 vs. Subject 7	33.04	8.6606	significantly different
Subject 1 vs. Subject 8	33.64	10.1679	significantly different
Subject 1 vs. Subject 9	1.35	8.6606	not significantly different
Subject 1 vs. Subject 10	7.31	11.6884	not significantly different
Subject 2 vs. Subject 3	43.48	18.6482	significantly different
Subject 2 vs. Subject 4	0.89	18.6482	not significantly different
Subject 2 vs. Subject 5	2.3	18.6482	not significantly different
Subject 2 vs. Subject 6	32.89	18.6482	significantly different
Subject 2 vs. Subject 7	33.17	18.6482	significantly different

Subject 2 vs. Subject 8	33.77	18.6482	significantly different
Subject 2 vs. Subject 9	1.22	18.6482	not significantly different
Subject 2 vs. Subject 10	7.19	18.6482	not significantly different
Subject 3 vs. Subject 4	44.38	10.9335	significantly different
Subject 3 vs. Subject 5	45.78	11.2921	significantly different
Subject 3 vs. Subject 6	10.6	12.6249	not significantly different
Subject 3 vs. Subject 7	10.32	6.8721	significantly different
Subject 3 vs. Subject 8	9.71	10.1679	not significantly different
Subject 3 vs. Subject 9	44.7	7.3924	significantly different
Subject 3 vs. Subject 10	50.67	11.6884	significantly different
Subject 4 vs. Subject 5	1.4	11.2921	not significantly different
Subject 4 vs. Subject 6	33.78	12.6249	significantly different
Subject 4 vs. Subject 7	34.06	10.9335	significantly different
Subject 4 vs. Subject 8	34.67	10.9335	significantly different
Subject 4 vs. Subject 9	0.33	10.9335	not significantly different
Subject 4 vs. Subject 10	6.29	11.6884	not significantly different
Subject 5 vs. Subject 6	35.18	12.6249	significantly different
Subject 5 vs. Subject 7	35.46	11.2921	significantly different
Subject 5 vs. Subject 8	36.07	11.2921	significantly different
Subject 5 vs. Subject 9	1.08	11.2921	not significantly different
Subject 5 vs. Subject 10	4.89	11.6884	not significantly different
Subject 6 vs. Subject 7	0.28	12.6249	not significantly different
Subject 6 vs. Subject 8	0.89	12.6249	not significantly different
Subject 6 vs. Subject 9	34.11	12.6249	significantly different
Subject 6 vs. Subject 10	40.07	12.6249	significantly different
Subject 7 vs. Subject 8	0.6	10.1679	not significantly different
Subject 7 vs. Subject 9	34.39	7.3924	significantly different
Subject 7 vs. Subject 10	40.35	11.6884	significantly different
Subject 8 vs. Subject 9	34.99	10.1679	significantly different
Subject 8 vs. Subject 10	40.96	11.6884	significantly different
Subject 9 vs. Subject 10	5.96	11.6884	not significantly different

Table 27 presents the results for the Tukey Kramer test performed for maximum heart rate. The maximum heart rate Tukey test showed 20 ‘not significantly different’ results and 25 ‘significantly different’ results out of the 45 different comparisons performed. The subject groups are both divided into almost equal sides with a difference of five.

Table 28. Tukey Kramer test- minimum heart rate

Tukey Kramer Procedure for Minimum Heart Rate			
Comparison	**Absolute Difference**	**Critical Range**	**Results**
Subject 1 vs. Subject 2	7.02	28.9421	not significantly different
Subject 1 vs. Subject 3	15.44	13.4413	significantly different
Subject 1 vs. Subject 4	10.07	16.9688	not significantly different
Subject 1 vs. Subject 5	27.63	17.5253	significantly different
Subject 1 vs. Subject 6	14.11	19.5939	not significantly different
Subject 1 vs. Subject 7	1.01	13.4413	not significantly different
Subject 1 vs. Subject 8	21.46	15.7807	significantly different
Subject 1 vs. Subject 9	7.25	13.4413	not significantly different
Subject 1 vs. Subject 10	11	18.1404	not significantly different
Subject 2 vs. Subject 3	8.42	28.9421	not significantly different
Subject 2 vs. Subject 4	3.05	28.9421	not significantly different
Subject 2 vs. Subject 5	34.65	28.9421	significantly different
Subject 2 vs. Subject 6	7.09	28.9421	not significantly different
Subject 2 vs. Subject 7	8.03	28.9421	not significantly different
Subject 2 vs. Subject 8	14.44	28.9421	not significantly different
Subject 2 vs. Subject 9	14.27	28.9421	not significantly different
Subject 2 vs. Subject 10	18.03	28.9421	not significantly different
Subject 3 vs. Subject 4	5.37	16.9688	not significantly different
Subject 3 vs. Subject 5	43.07	17.5253	significantly different
Subject 3 vs. Subject 6	1.33	19.5939	not significantly different
Subject 3 vs. Subject 7	16.45	10.6656	significantly different
Subject 3 vs. Subject 8	6.02	15.7807	not significantly different
Subject 3 vs. Subject 9	22.68	11.473	significantly different
Subject 3 vs. Subject 10	26.44	18.1404	significantly different
Subject 4 vs. Subject 5	37.7	17.5253	significantly different
Subject 4 vs. Subject 6	4.04	19.5939	not significantly different
Subject 4 vs. Subject 7	11.08	16.9688	not significantly different
Subject 4 vs. Subject 8	11.39	16.9688	not significantly different
Subject 4 vs. Subject 9	17.31	16.9688	significantly different
Subject 4 vs. Subject 10	21.07	18.1404	significantly different
Subject 5 vs. Subject 6	41.74	19.5939	significantly different
Subject 5 vs. Subject 7	26.62	17.5253	significantly different
Subject 5 vs. Subject 8	49.09	17.5253	significantly different
Subject 5 vs. Subject 9	20.39	17.5253	significantly different
Subject 5 vs. Subject 10	16.63	18.1404	not significantly different

Subject 6 vs. Subject 7	15.12	19.5939	not significantly different
Subject 6 vs. Subject 8	7.35	19.5939	not significantly different
Subject 6 vs. Subject 9	21.36	19.5939	significantly different
Subject 6 vs. Subject 10	25.11	19.5939	significantly different
Subject 7 vs. Subject 8	22.47	15.7807	significantly different
Subject 7 vs. Subject 9	6.23	11.473	not significantly different
Subject 7 vs. Subject 10	9.99	18.1404	not significantly different
Subject 8 vs. Subject 9	28.71	15.7807	significantly different
Subject 8 vs. Subject 10	32.46	18.1404	significantly different
Subject 9 vs. Subject 10	3.76	18.1404	not significantly different

This Tukey test revealed 25 ‘not significantly different’ results and 20 ‘significantly different’ results out of the 45 different comparisons performed. In this case, there are more subject groups who outputted ‘not significantly different’ results compared to the subject groups who outputted ‘significantly different’ results. This result may be indicative of the fact that most subjects presented a similar average minimum heart rate. Figure 12 previously showed that this result presented an average minimum heart rate of 105 beats/minute and a standard deviation of 14.3 beats/minute. Additionally, this result showed that the minimum heart rates are similar to the mean of the 10 subjects. For this reason, the test outputted more ‘not significantly different’ result in comparison to those that were ‘significantly different’. This result further reiterates the fact that heart rate minimum is not a sensitive marker to be used for pain assessment.

6.4 Curve Fitting

The population of ten subjects were introduced to the Curve Fitting Toolbox, using Matlabworks®, fitting gestational age versus maximum, average and minimum heart rate data in this population. The goodness of fit of a statistical model describes how well the data fits a set of observations. The data was fitted to a regression polynomial linear model, Equation 8, to test if

the heart rate in this subset of patients was following a pattern that was dependent on gestational age.

$$f(x) = p1 * x + p2 \quad Equation\ 8$$

Table 29. Goodness of fits

Fit name	Data	Fit type	SSE	R-square	DFE	Adj R-sq	RMSE
GA vs Average Heart Rate	MAX vs. GA	poly1	2.6491e+05	0.1982	541	0.1967	22.1285
GA vs Max Heart Rate	AVG vs. GA	poly1	1.7582e+05	0.2599	541	0.2585	18.0276
GA vs Min Heart Rate	MIN vs. GA	poly1	3.0621e+05	0.0755	541	0.0737	23.7908

SSE, R-square and RMSE demonstrated that the regression line does not fit the data, thus the heart rate in this population is not following a particular pattern that was dependent on gestational age. Figure 13, 14 and 15 illustrate the polynomial fit against the data of ten subjects.

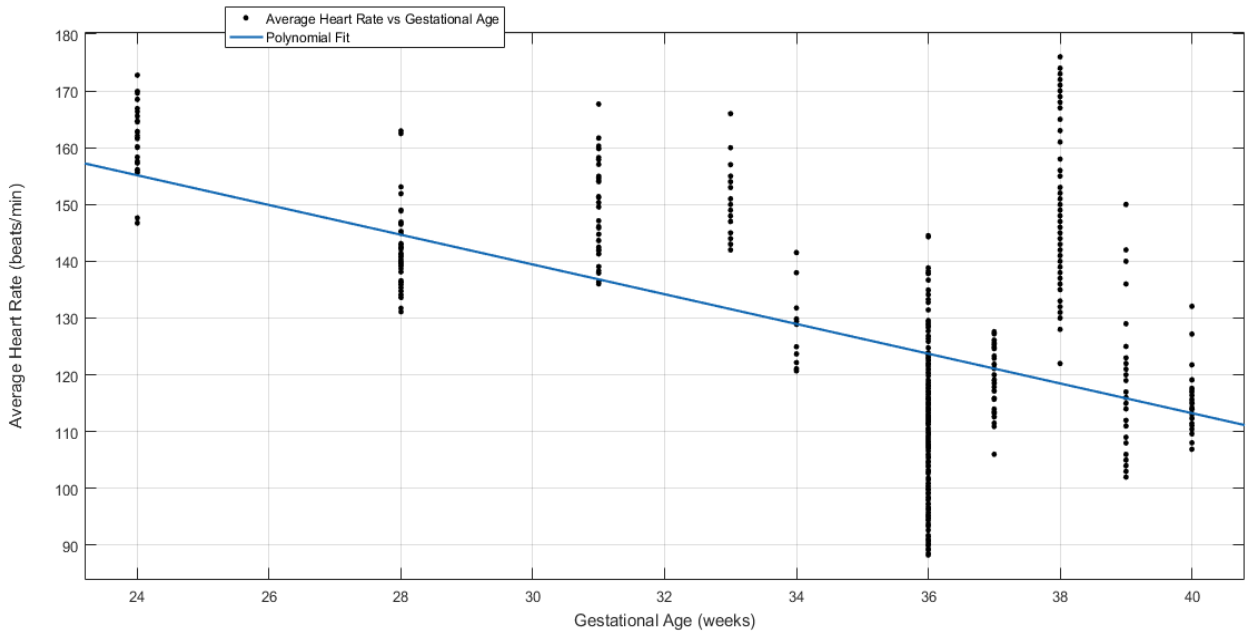


Figure 13. Gestational Age vs. Average Heart Rate for Subject 1-10

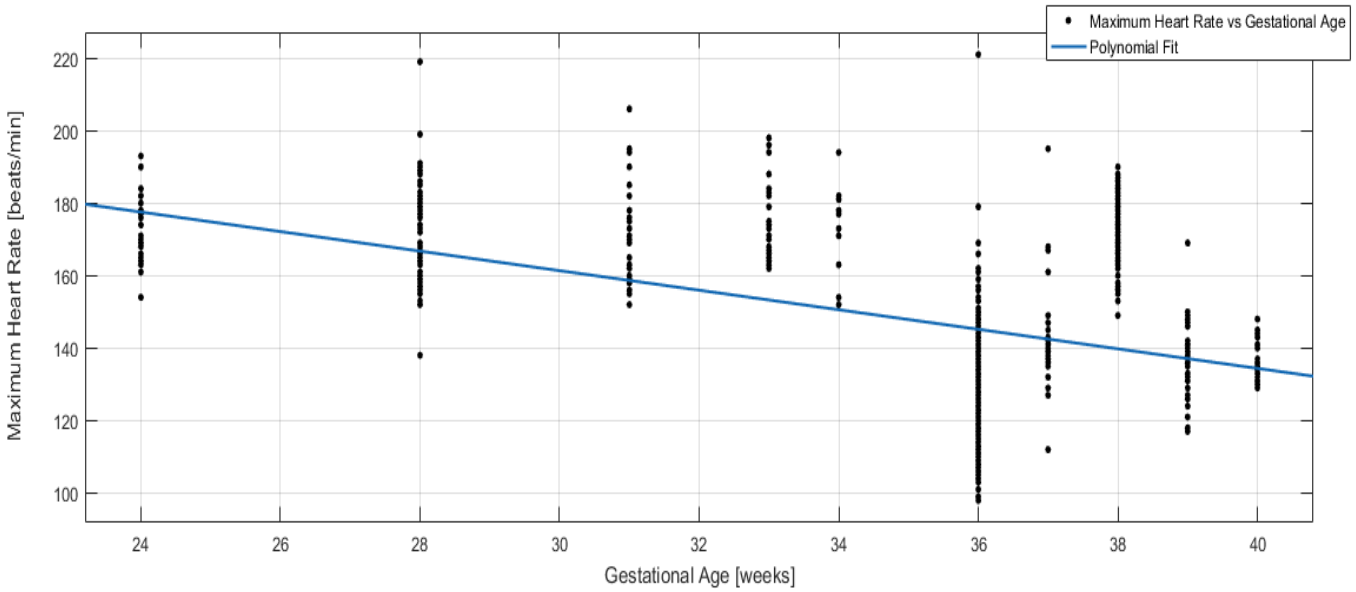


Figure 14. Gestational Age vs. Maximum Heart Rate for Subject 1-10

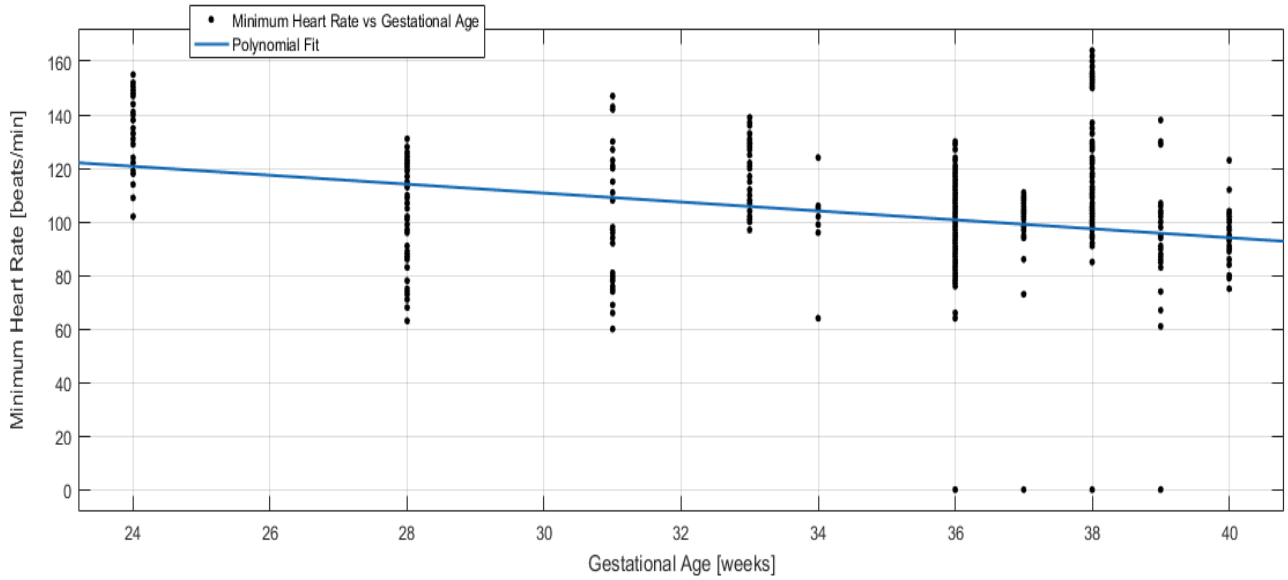


Figure 15. Gestational Age vs. Minimum Heart Rate for Subjects 1-10

The figures highlight the complex distribution of the heart rate data through the gestational age spectrum without presenting a consistent pattern. Since no apparent pattern was observed within the gestational age spectrum, clinical notes were analysed to look into the subjects' surgery data. In this case, as the gestational age of 36 is showing a very low heart rate average as well as very low maximum heart rate, it was reported that both the subjects who reported being 36 gestational

age, did not have a surgery. Hence, it could be concluded that due to perhaps less pain or discomfort in this gestational age group, the heart rate was reported to be lower compared to other subjects. The regression line did not fit the data as every subject is unique and has different clinical implications. A particular premature infants heart rate is driven by what clinical context he/she are in.

6.5 Surface model comparing PIPP vs. AVG HR vs. GA

Utilizing Curve Fitting Toolbox in Matlabworks®, the population of ten subjects was plotted using biharmonic interpolation to scatter the data in the surface model, as shown in Figure 16. This analysis was performed to analyse and find the correlations between PIPP scores taken from Hospital for Sick Children, Toronto, Canada, the average heart rate data from The Artemis Platform and the GA.

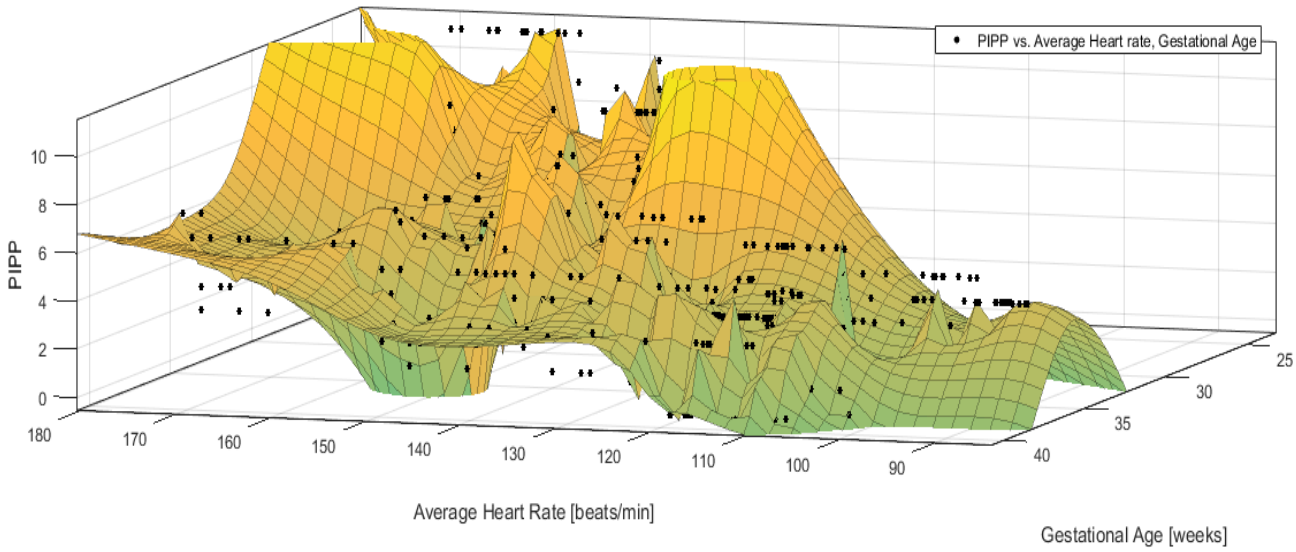


Figure 16. Biharmonic Interpolation

Figure 16 illustrates the PIPP scores in a very scattered pattern where a particular PIPP score is recorded for multiple heart rate values and GA. Thus, showing that specific criteria are not followed for heart rate and pain scoring. Based on this result, inconsistency is observed, as PIPP

scores do not match consistently with heart rate thresholds. There can be many reasons for the inconsistent result. Nonetheless, designing a more reliable, accurate and consistent automated system of scoring is important in order to eliminate these problems and discrepancy in scoring.

6.6 Conclusion

In conclusion, the results presented in this chapter may help to better understand the heart rate indicator and its constantly changing complexity due to painful stimuli. It is evident from the results presented in this chapter that heart rate average can prove to be a useful measure in providing insight into detecting painful events. In contrast, through the statistical analysis completed (ANOVA, Tukey Kramer Test, and Curve Fitting), heart rate minimum is not found to be a useful marker since heart rate minimum presented a standard consistent pattern between all subjects as in clinical practice, heart rate is monitored carefully by health care professionals to ensure it does not go below the baseline standards set. Based on the overall analysis presented in this chapter, heart rate can be a useful marker when used in accordance with gestational age to classify pain in the premature infant. These results also reiterate the need for a pain assessment tool that adjusts for gestational age when scoring. The completed analysis showed a clear correlation between gestational age and heart rate. Premature infants baseline heart rate differs at varying gestational ages, which in turn, can affect their pain scores in a substantial way. Analysing continuous physiological data helped to identify and examine areas, which may have been missed with the manual collection of PIPP scores at inconsistent time points. Manual collection of physiological indicators at inconsistent time points loses information relevant to the neonates' physiology. As such, the creation of an automated system using big data analytics is important for the collection of physiological changes and the accurate detection of pain by avoiding misreporting of pain.

Chapter 7 - Discussion

In this chapter, the findings of the evaluation of data model are discussed and how they are linked to the existing literature on pain management. Limited understanding on infant pain has led to its lack of recognition in clinical practice. As a result, there is still a lack of a gold standard per say for assessing pain in the neonatal population. Many infants still don't receive any pain treatment during commonly performed painful procedures (Cignacco et al., 2009; Harrison et al., 2014; Johnston, Barrington, Taddio, Carbajal, & Filion, 2011). These studies highlight an important knowledge to practice gap, which can negatively impact the health of newborn infants. In the past, a common presumption was that neonatal infants were not capable of fully perceiving pain. However, such a claim was contradicted by evidence from subsequent studies. For example, a recent study by Goksan et al. incorporated an fMRI to study newborn infant pain. The fMRI identified a network of brain regions that are active following acute noxious stimulation in newborn infants, where such activity was compared to that observed in adults. After analyzing this, the study discovered significant infant brain activity in 18 of the 20 active adult brain regions (Goksan et al., 2015). Brain regions that encode sensory and affective components of pain are active in infants, suggesting that the infant pain experience closely resembles that seen in adults (Goksan et al., 2015). These findings and the ones outlined in this thesis highlight the importance of developing effective and reliable pain assessment tools in this vulnerable population. Hence, an attempt was made in this thesis to design a clinical DSS model for automated partial PIPP scoring based on the elements that can be determined without the need for clinical observation. The clinical DSS was designed depicting the current standard of pain assessment to outline the possibility of automating a pain scoring system using big data analytics.

The evaluation of the data model presented noteworthy results where one of the important finding was that the PIPP score was not scored each hour APIPP was scored, and on average, PIPP was only scored 20% of the hours within this study. Such an issue outlines the limitation of the manual scoring systems being used at hospitals. The analysis completed in this thesis between the two scales also explored the difference between the two scales since the PIPP was scored out of 21 and the APIPP was scored out of 9 points. There was a 12-point difference between these two scales. For the case study on subject 1, the PIPP score was not available when the APIPP was available for 79.95% of the time. A similar result was found for the analysis done with 10 subjects, where the PIPP score was not available when the APIPP was available for 78.36% of the time. Such a finding is significant, as the PIPP data could not be compared with the hourly APIPP score for nearly 80% of the time. Such a result suggests that the nurses did not score the PIPP consistently as many hours of data were missing. This is worrisome as significant changes in the premature infant's physiology can take place during the hours in which the PIPP is not scored. The second highest percentage was found for the greater category. Specifically, for case study for subject 1, PIPP score was greater than the APIPP score for 14.76% of the time. Similarly, for the data analysis with the 10 subjects, PIPP score was greater than the APIPP score for 14.05% of the time. Such a result was expected as PIPP score is scored out of a higher value in comparison to APIPP. Thus, the percent difference between the two scales was explored. For both analyses, high frequency of occurring differences were in the lower range with most differences ranging between one and four. More specifically, a difference of two and three had the highest occurrence for both analyses. This was concluded to be a non-significant difference acknowledging the fact that there was a 12-point difference between the two scales. This is a significant finding as PIPP is scored out of a much higher number (21) compared to APIPP (9).

The highest frequency of differences being two or three concludes that APIPP score was fairly similar to the PIPP score that was scored by the nurses. It can be inferred from this finding that perhaps the PIPP score tends to be highly based on the physiological changes and GA, which the APIPP was constructed of. In this case, it can also be inferred that perhaps the nurses were not able to score the various behavioural parameters due to sedation or inactivity in the infant due to prematurity. Additionally, it could also be speculated that perhaps APIPP score was more sensitive as it consisted of per second physiological data, which the PIPP perhaps missed. It is important to further investigate these findings. To further investigate these findings, a physiological analysis using HR data was conducted in chapter 6. Chapter 6 presented various experiments that presented a clinical perspective on pain management. The results presented in chapter 6 helped in understanding the complexity of the heart rate indicator and its resourcefulness as a marker for pain detection. These results indicate the importance of incorporating a clinical context within pain assessment tools. Factors such as GA, severity of illness and severity of surgery can present fluctuating changes within heart rate. The experiments discussed in this thesis are preliminary in nature and were run in DB2. Executing these experiments in InfoSphere streams is possible. In the future, this concept can be implemented in real-time in streams to run every hour and produce a pain score that could assist physicians and nurses. It is important to carry out prospective research to experiment with real-time stream processing to integrate a scale such as APIPP. An exploratory approach presented in this thesis helps to recognize the need and the resourcefulness of designing an automated pain scoring system using a platform such as Artemis, which provides real-time physiological data. A stand-alone automated system can be designed based on this work, which will provide continuous monitoring of pain in the NICU with the contextual support of the physicians and nurses.

The various different pain management systems described in chapter 2 can be used as background knowledge to implement a well-rounded, reliable pain management system. The richness of information can be extrapolated from the neonatal information sources in the form of concepts and categories, and these concepts and categories can be used to understand the context of neonatal pain management, and finally interconnect the concepts and categories as an ontology to severe pain management. One of the main areas that has been extensively implemented in the neonatal population is facial recognition systems. Recent research has provided evidence of the usefulness of facial cues for automatic pain analysis; however, such research has mainly focused on detection of presence/absence of pain. Such facial cues can become inadequate during circumstances where a neonate is recovering from surgery, or is highly sedated and premature to provide facial cues. For this reason, such systems should be integrated into more objective-based systems that can provide reliable results despite these hurdles. Computer-aided decision support offers help, but the existing systems are not user-friendly or do not support an on-line application from clinical documentation. Thus, these applications are not available in everyday clinical practice for health care providers to use (Eich et al., 1997). The pain detection system should be designed so that the system is at the central point-of care where the clinicians have easy accessibility.

Currently, one central problem is the fact that a simple method is not available for the direct measurement of pain. In most cases of pain assessment, the examining physician must rely on the patient's qualitative description about the location, quality and intensity of the pain sensation. The quantification of pain is possible with the help of the visual analog scale (VAS) or the numeric rating scale (NRS). However, these methods are only plausible during situations where the patient is sufficiently alert and cooperative, which is not always possible in the

medical field, including times where the patient is in post-surgery phases or in this case a neonate who is not able to communicate the severity of pain (Walter et al., 2013). Overall, the methods are either considered inadequate or still in development. If the conditions do not allow for a sufficiently valid measurement of pain, treating the pain may lead to an over- or underestimation of analgesic administration as well as long-term effects such as alteration in response to subsequent painful experience. Current hospital practices require the nursing staff to apply validated pain scoring methods before taking appropriate actions to ameliorate newborn pain or discomfort. However, current nursing workload in the NICU does not allow bedside nurses to assess neonatal pain accurately (Hall & Anand, 2014). Additionally, if the validated pain scales are not working properly, the pain is under rated or over rated frequently. Many pain scales such as PIPP combine behavioural, physiological, and other variables, but these variables may not respond to neonatal pain in similar or specific ways. The inter-rater reliability and subjectivity of human assessments are further limiting factors in their prevalent use (Hall & Anand, 2014). The use of qualitative or subjective methods, rather than quantifiable data for neonatal pain assessment, results in inconsistencies. Due to a large pharmacokinetic variability of analgesic drugs in neonates, their pain management is often of poor quality and inconsistent from shift to shift (Guedj et al., 2014). Adopting an objective pain assessment method will greatly enhance the quality of pain management in NICUs by avoiding untreated pain or excessive analgesia. The case study provided in this thesis verifies that the PIPP scores were not consistently collected compared to the automated score of APIPP. Pain assessment methods should be designed to reduce nursing workload and the side effects of under- or overdosing analgesics. Studies that aim at a practical application of findings in the field of automatic pain recognition, specifically within the neonatal domain, are virtually non-existent today. The results

reported in this thesis as well as in earlier studies indicate a high potential for developing machine detection systems that could be implemented in the neonatal pain domain.

Chapter 8 - Conclusion

In summary, identification of pain in neonates continues to challenge nurses and physicians who hold the responsibility to provide optimal care to this vulnerable population. The inability to affectively quantify and relieve pain leaves health care providers with limited knowledge to guide practice. It remains a clinical art to combine patients' reports, behavioural observation, and physiologic measurement with the history, physical exam, laboratory information, and overall clinical context in guiding clinical judgments and therapeutic interventions (Berde & McGrath, 2009). The complexity of pain in newborns was not recognized until the 1980s. Infants were believed to have no capacity to experience or remember pain; as such infants were not treated for pain in circumstances where pain would be anticipated. Since then, assessment and management of pain have advanced and has become the focus of substantial research. Despite this, we are far from having a standardized practice for managing pain in premature infants. There are many variations in the methods and scales used across different healthcare organizations. Each health care organization uses the pain management practice that best suits their organization. It is important to initiate the inclusion of a wide variety of physiological indicators and contextual information, such as gestational age, severity of illness and severity of surgery to increase the sensitivity and validity of the pain scales. This thesis outlines various studies that have displayed the importance and usefulness of these physiological and neurological techniques in Chapter 2. It is important that these findings are effectively understood and researched to create tools that can provide absolutely accurate detection of pain in neonates.

The objective measurement of subjective, multidimensionality-experienced pain is still a key problem in the neonatal population that has yet to be adequately solved. Due to inconsistencies in pain measurement, opiates are being used at an increasing rate despite

concerns about patient safety and misuse (Midboe et al., 2011). Various automated systems have attempted to resolve shortcomings by using information systems to make the pain measurement system more reliable and time-efficient. The main goal or objective of information systems is to improve the performance of people in organizations through the use of information technology. Computer-based clinical decision support systems show promise for improving clinical decision-making, evidence-based guideline adherence, care coordination across providers and disciplines, as well as patient education and communication (Midboe et al., 2011).

Almost a decade ago, a question was asked relating to the ethical imperative to treat pain in infants and whether the pain management was fully effective at the time (Franck, 1997). Based on the findings in this thesis, as well as other similar findings of studies conducted in the recent past (Harrison et al., 2014; Johnston et al., 2011; Roofthooft, Simons, Anand, Tibboel, & van Dijk, 2014), the answer is unfortunately still "no". As a community of clinicians, health care researchers, and funders of research, greater attempts should be made to ensure that a newborn infant does not need to undergo an unnecessary painful procedure without provision of effective pain reduction. This cannot be done until we have the tools to assess pain in an effective manner. Misreported pain is the cause of majority of the concerns reported in this thesis.

8.1 Research and Findings

To explore the research questions of this thesis, firstly, a literature review was completed in chapter 2 to present various studies conducted in the area of pain management to uncover the need for an automated real-time based pain scoring system in the NICU. Literature review was structured based on three significant themes of the thesis. Firstly, a discussion on how health information systems can be used to assess pain was explored. Secondly, a review was presented on how physiological parameters can be used for the assessment of pain. It was determined that

physiological cues can prove to be very effective in the assessment of pain in premature infants. However, their specificity as a measure of pain reactivity in premature infants has not been investigated in great detail. Lastly, a review was conducted on various computerized/ automated pain measurement systems to aid in designing one for the NICU. Upon completion of the review, it was determined that there is a lack of automated pain management systems that are designed specifically for the NICU. Hence, this provided a great scope for exploring this area of research and proposing the hypothesis to design a pain scoring system using big data analytic techniques that can provide frequent monitoring of pain. The Artemis platform is presented in detail in chapter 3 to outline its usability within this thesis. The study design presented in this thesis has demonstrated the utility of the Artemis platform for the effective detection and monitoring of pain within the neonatal population. It is possible to deploy an algorithm unique to this research within Artemis to create a pain detection system that can be automated.

Chapter 4 presented the methodology in two phases outlining the preparation of data and the creation of the data model. The data model involved combining the abstractions and features from the data preparation phase to compute an automated partial pain score based on big data analytics and quantifiable scoring using HR, SpO_2 and GA. Following this, the data model/ APIPP scale was evaluated in chapter 5 by comparing it with the PIPP scores collected manually by nurses. A thorough analysis was conducted for evaluation purposes to examine if the APIPP score produced greater, lesser or equal results as the PIPP score. One of the main findings from this evaluation was that PIPP score was not available almost 80% of the time when compared to APIPP score which was generated automatically using the Artemis platform. Hence, the hypothesis was proved that more frequent monitoring of pain is possible by creating an automated system using Artemis platform. Furthermore, when evaluating the difference between

two scores, a minimal difference of two and three was prominent. Such a finding is noteworthy as it is significant to note that PIPP scoring is scored on a 21-point scale whereas APIPP is scored on a 9-point scale. It can be inferred that the PIPP was perhaps highly scored on physiological parameters and GA due to its score being very close to the APIPP score. This can be due to many reasons such as perhaps the infant's inability to give behavioural cues for scoring purposes. Hence the score being highly scored on only the available physiological cues such as HR and SpO_2. Such a negligible difference between two scales proposes further investigation.

To create a physiologic based pain scale, it was imperative to investigate how physiological changes can be useful in detecting a pain response. Hence, Chapter 6 explored the physiological and clinical side of pain by outlining in detail the resourcefulness of HR as a physiological marker for detecting pain. Various statistical experiments were conducted to review HR in accordance with GA, surgical details and PIPP scoring. The results showed that HR average can prove to be a useful measure in detecting painful events as this correlated with the surgery each subject underwent. Subjects who underwent severe surgeries, presented with a high HR compared to those who did not undergo a surgery. Similarly, in pain detection, high HR will show abnormalities or pain that is present in an infant at a given time the score is being collected. In an automated pain detection system, this has the potential to give alerts to health care professionals when the HR goes above the set threshold for a particular subject. These findings also reiterate the need to design a pain assessment tool that adjusts for gestational age as the baseline heart rate of each subject is different at varying gestational ages. Chapter 7 presented a discussion which brings together the concepts presented in each chapter.

Upon completion of the study and using the knowledge presented in this thesis, it has been confirmed that despite the use of the DB2 system in this study, the algorithm logic can be

converted to the Streams Processing Language and can be applied to the Artemis platform to create an automated pain detection system that can run in real-time. Analyzing continuous physiological data in this thesis, helped to identify areas which may have been overlooked due to manual collection of PIPP scores at inconsistent time points. Due to having retrospective per second data for each subject, more information was available to predict the changes that occurred within the subject's physiology that potentially would have been missed due to manual interpretation at inconsistent time points. This research explores an avenue that has not been attempted before in this domain. Using the findings presented in this thesis, this novel research has the potential to make significant contributions to the medical and informatics field.

8.2 Limitations

Various limitations have occurred while conducting the experiments in this thesis. One of the first limitations of the study was that the data used was retrospective. Due to retrospective data collection, contemplating contextual details about the subject, such as dosage of analgesia given or the time it was administered, was difficult. Because of this, analyzing PIPP scores that were collected by nurses was also difficult. Since the population for this data set was that of surgical patients, many were administered analgesic medications that could prevent accurate pain scoring. In the future, prospective data collection is imperative, where subjects will be recruited based on specific eligibility criteria. An example of such criteria include those who are not using medications that could alter their response to pain. In this data set, most PIPP scores were less than 10, while a PIPP score of 2 and 3 was the most frequent. This is not a significant change as the PIPP score is scored out of 21. Due to a larger proportion of the study population undergoing surgeries and under analgesia, the recorded PIPP scores were not very high. Also, each subject

used in this research was unique as subjects had different surgical procedures, making it almost impossible to standardize for testing purposes.

Furthermore, the PIPP scores were not collected continuously but rather collected at random time points. On the other hand, APIPP score data was available for every hour for each patient. As a result, continuous data analysis was not possible, as the availability of APIPP and PIPP scores did not match. As such, comparing such data was difficult as more than half the data could not be used for scoring analysis. The quality of the Artemis data for the timeframe chosen for this thesis was also poor as many hours of data were missing due to artifacts. As a result, some subjects had to be excluded from the study analysis and only the subjects who had continuous data were used.

8.3 Future Work

In the future, this work will be used to design a physiological indicator based scale that can be integrated into a decision support system named Artemis. Using various physiological data streams, the novel scale can be integrated into the Artemis platform to predict nociceptive events. With such an informatics tool, the identification of nociceptive stimuli can be improved, and therefore, improve the use of drugs and non-pharmacological interventions for pain relief. Such a pain assessment system can provide continuous and minimally biased assessment of pain. Future work can attempt to implement this theoretical design in streams and test the implementation in SPL, which is the primary language used by the Artemis platform. This avenue of research presents future opportunities, a few of which could include the use of full 24-hour data sets for all patients as well as the continuous real-time analysis of the data. Through this design of a pain scale, alert systems can also be designed in the CDSS to assist the nurses or other health care professionals. These alert systems can be based on threshold changes in the physiological

parameters that are detected by the system. For this, it is important to set baseline parameters for each neonate within the system. By doing this, the CDSS can provide alerts when it observes an unusual physiological change. This work can enable, not only detection, but also quantification of pain in neonates as mild, moderate or severe by examining and deriving patterns within the physiological data streams.

Future work should also carry out similar research in a prospective manner in order to reduce limitations. Following this thesis work, a prospective observational study can be conducted with two cohort of neonatal infants, with one group being surgical and the other being the non-surgical group. To validate this approach, APIPP can be run and scored for each patient and simultaneously, nurses can collect PIPP scores as frequently as the APIPP score is generated.

Incorporation of contextual information with other pain indicators is essential for the refinement of the assessment process. This thesis has presented a need for developing a continuous, context-sensitive, and multimodal system, which will provide best practice to assess neonatal pain. The knowledge from this thesis can be used to research and implement this design in an NICU in the future. Furthermore, this thesis will provide a valid insight into the importance of the assessment of physiological parameters in the area of neonatal pain management. In turn, better and more accurate pain management strategies can be created, improving the health of the youngest members of society.

8.4 Concluding Remarks

The research in automated continuous detection of pain using physiological parameters has not been previously attempted. Most work in this domain has involved great emphasis on behavioural parameters for scoring pain. This thesis provides extensions in the form of research contributions and presents a novel concept of using automated real-time scoring of pain

parameters. This thesis also has presented a framework that will facilitate the implementation of automated pain detection system that can be executed within the Artemis platform to assess pain in a continuous and reliable manner. This has the potential to provide regular pain assessment in the NICU which can lead to better pain assessment and management that can otherwise go unnoticed and result in eventual reduced mortality and morbidity of patients.

The primary objective of this thesis was to investigate how big data analytic techniques can be used to generate a neonatal pain score in the future. In parallel with this objective, another aim of this research was to assess physiological parameters in order to explore the possibility of generating an automated pain score using physiological elements. Overall, using retrospective physiological data, these objectives were fulfilled to create an alternate way to design and generate a score that was being generated manually. Current practice in an NICU involves the meticulous, time-consuming and bias process of manual interpretation of pain scores by nurses, subsequently the automation of this process has the potential to increase the time frequency at which pain scores are recorded, eliminate bias and inconsistency of scoring as well as improve the timing of medical intervention to allow for healthy development in neonates. Through this retrospective analysis of pain scoring, it was determined that this approach can be applied to a real-time automated environment enabled by stream computing. Applicability and future work relating to what has been described in this thesis will lead to innovative automated systems integrated into clinical practice. This design of the APIPP scoring as a clinical decision support tool can fundamentally change the notion of pain monitoring in the future.

References

Anand, K. J., Coskun, V., Thrivikraman, K. V., Nemeroff, C. B., & Plotsky, P. M. (1999). Long-Term Behavioral Effects of Repetitive Pain in Neonatal Rat Pups. *Physiology & behavior, 66*(4), 627-637.

Anand, K. J., Palmer, F. B., & Papanicolaou, A. C. (2013). Repetitive neonatal pain and neurocognitive abilities in ex-preterm children. *Pain, 154*(10), 1899-1901. doi:10.1016/j.pain.2013.06.027

Anand, K. J., & Scalzo, F. M. (2000). Can adverse neonatal experiences alter brain development and subsequent behavior? *Biol Neonate, 77*(2), 69-82. doi:14197

Apkarian, A. V., Bushnell, M. C., Treede, R. D., & Zubieta, J. K. (2005). Human brain mechanisms of pain perception and regulation in health and disease. *Eur J Pain, 9*(4), 463-484. doi:10.1016/j.ejpain.2004.11.001

Berde, C., & McGrath, P. (2009). Pain measurement and Beecher's challenge: 50 years later. *Anesthesiology, 111*(3), 473-474. doi:10.1097/ALN.0b013e3181b27c4c

Bhutta, A. T., & Anand, K. J. (2002). Vulnerability of the developing brain. Neuronal mechanisms. *Clin Perinatol, 29*(3), 357-372.

Blount, M., Ebling, M. R., Eklund, J. M., James, A. G., McGregor, C., Percival, N., . . . Sow, D. (2010). Real-time analysis for intensive care: development and deployment of the artemis analytic system. *IEEE Eng Med Biol Mag, 29*(2), 110-118. doi:10.1109/memb.2010.936454

Borsook, D., & Becerra, L. R. (2006). Breaking down the barriers: fMRI applications in pain, analgesia and analgesics. *Molecular Pain, 2*, 30-30. doi:10.1186/1744-8069-2-30

Brahnam, S., Nanni, L., & Sexton, R. (2007). Introduction to Neonatal Facial Pain Detection Using Common and Advanced Face Classification Techniques. In H. Yoshida, A. Jain, A. Ichalkaranje, L. Jain, & N. Ichalkaranje (Eds.), *Advanced Computational Intelligence Paradigms in Healthcare – 1* (Vol. 48, pp. 225-253): Springer Berlin Heidelberg.

Brown, J. E., Chatterjee, N., Younger, J., & Mackey, S. (2011). Towards a physiology-based measure of pain: patterns of human brain activity distinguish painful from non-painful thermal stimulation. *PLoS One, 6*(9), e24124. doi:10.1371/journal.pone.0024124

Brummelte, S., Grunau, R. E., Chau, V., Poskitt, K. J., Brant, R., Vinall, J., . . . Miller, S. P. (2012). Procedural pain and brain development in premature newborns. *Ann Neurol, 71*(3), 385-396. doi:10.1002/ana.22267

Catley, C., Smith, K., McGregor, C., James, A., & Eklund, M. (2011). A Framework for Multidimensional Real-Time Data Analysis: A Case Study for the Detection of Apnoea of Prematurity. *International Journal of Computational Models and Algorithms in Medicine (IJCMAM), 2*(1).

Cignacco, E., Hamers, J., van Lingen, R. A., Stoffel, L., Buchi, S., Muller, R., . . . Nelle, M. (2009). Neonatal procedural pain exposure and pain management in ventilated preterm infants during the first 14 days of life. *Swiss Med Wkly, 139*(15-16), 226-232. doi:smw-12545

Cirelli, J., McGregor, C., Graydon, B., & James, A. (2013). Analysis of continuous oxygen saturation data for accurate representation of retinal exposure to oxygen in the preterm infant. *Stud Health Technol Inform, 183*, 126-131.

Czarnecki, M. L., Simon, K., Thompson, J. J., Armus, C. L., Hanson, T. C., Berg, K. A., . . . Malin, S. (2011). Barriers to pediatric pain management: a nursing perspective. *Pain Manag Nurs, 12*(3), 154-162. doi:10.1016/j.pmn.2010.07.001

Doesburg, S. M., Chau, C. M., Cheung, T. P., Moiseev, A., Ribary, U., Herdman, A. T., . . . Grunau, R. E. (2013). Neonatal pain-related stress, functional cortical activity and visual-perceptual abilities in school-age children born at extremely low gestational age. *Pain, 154*(10), 1946-1952. doi:10.1016/j.pain.2013.04.009

Eich, H. P., Ohmann, C., & Lang, K. (1997, 11-13 Jun 1997). *Decision support in acute abdominal pain using an expert system for different knowledge bases*. Paper presented at the Computer-Based Medical Systems., 1997. Proceedings., Tenth IEEE Symposium on Computer-Based Medical Systems (CBMS '97).

Eklund, J. M., Fontana, N., Pugh, E., McGregor, C., Yielder, P., James, A., . . . McNamara, P. (2014). *Automated Sleep-Wake Detection in Neonates from Cerebral Function Monitor Signals*. Paper presented at the 27th International Symposium on Computer-based Medical Systems (CBMS 2014), New York, USA.

Evans, J. C., McCartney, E. M., Lawhon, G., & Galloway, J. (2005). Longitudinal comparison of preterm pain responses to repeated heelsticks. *Pediatr Nurs, 31*(3), 216-221.

Farion, K., Michalowski, W., Wilk, S., O'Sullivan, D., Rubin, S., & Weiss, D. (2009). Clinical Decision Support System for Point of Care Use. *Methods of Information in Medicine, 48*, 381-390.

Faye, P. M., De Jonckheere, J., Logier, R., Kuissi, E., Jeanne, M., Rakza, T., & Storme, L. (2010). Newborn infant pain assessment using heart rate variability analysis. *Clin J Pain, 26*(9), 777-782.

Fitzgerald, M. (2005). The development of nociceptive circuits. *Nat Rev Neurosci, 6*(7), 507-520.

Franck, L. S. (1997). The ethical imperative to treat pain in infants: are we doing the best we can? *Crit Care Nurse, 17*(5), 80-87.

Franck, L. S., & Miaskowski, C. (1997). Measurement of neonatal responses to painful stimuli: a research review. *J Pain Symptom Manage, 14*(6), 343-378.

Gallo, A. M. (2003). The fifth vital sign: implementation of the Neonatal Infant Pain Scale. *J Obstet Gynecol Neonatal Nurs, 32*(2), 199-206.

Gholami, B., Haddad, W. M., & Tannenbaum, A. R. (2010). Relevance vector machine learning for neonate pain intensity assessment using digital imaging. *IEEE Trans Biomed Eng, 57*(6), 1457-1466. doi:10.1109/tbme.2009.2039214

Gibbins, S., & Stevens, B. (2001). State of the art: Pain assessment and management in high-risk infants. *Newborn and Infant Nursing Reviews, 1*(2), 85-96. doi:10.1053/nbin.2001.24558

Goksan, S., Hartley, C., Emery, F., Cockrill, N., Poorun, R., Moultrie, F., . . . Slater, R. (2015). fMRI reveals neural activity overlap between adult and infant pain. *eLife Sciences Publications Ltd., 2015*(4), 1-13.

Grunau, R. E. (2013). Neonatal Pain in Very Preterm Infants: Long-Term Effects on Brain, Neurodevelopment and Pain Reactivity. *Rambam Maimonides Medical Journal, 4*(4), e0025. doi:10.5041/RMMJ.10132

Guedj, R., Danan, C., Daoud, P., Zupan, V., Renolleau, S., Zana, E., . . . Carbajal, R. (2014). Does neonatal pain management in intensive care units differ between night and day? An observational study. *BMJ Open, 4*(2), e004086. doi:10.1136/bmjopen-2013-004086

Hall, R. W., & Anand, K. J. S. (2014). Pain Management in Newborns. *Clin Perinatol, 41*(4), 895-924. doi:10.1016/j.clp.2014.08.010

Harrison, D., Joly, C., Chretien, C., Cochrane, S., Ellis, J., Lamontagne, C., & Vaillancourt, R. (2014). Pain prevalence in a pediatric hospital: raising awareness during Pain Awareness Week. *Pain Res Manag, 19*(1), e24-30.

Harrison, D., Reszel, J., Wilding, J., Abdulla, K., Bueno, M., Campbell-Yeo, M., . . . Stevens, B. (2015). Neuroprotective Core Measure 5: Minimizing Stress and Pain—Neonatal Pain Management Practices During Heel Lance and Venipuncture in Ontario, Canada. *Newborn and Infant Nursing Reviews, 15*(3), 116-123.

Herr, K., Bjoro, K., & Decker, S. (2006). Tools for assessment of pain in nonverbal older adults with dementia: a state-of-the-science review. *J Pain Symptom Manage, 31*(2), 170-192. doi:10.1016/j.jpainsymman.2005.07.001

Herrington, C. J., Olomu, I. N., & Geller, S. M. (2004). Salivary cortisol as indicators of pain in preterm infants: a pilot study. *Clin Nurs Res, 13*(1), 53-68. doi:10.1177/1054773803259665

Janig, W. (1995). The sympathetic nervous system in pain. *Eur J Anaesthesiol Suppl, 10*, 53-60.

Johnston, C., Barrington, K. J., Taddio, A., Carbajal, R., & Filion, F. (2011). Pain in Canadian NICUs: have we improved over the past 12 years? *Clin J Pain, 27*(3), 225-232. doi:10.1097/AJP.0b013e3181fe14cf

Johnston, C., & Stevens, B. J. (1996). Experience in a neonatal intensive care unit affects pain response. *PEDIATRICS, 98*(5), 925-930.

Johnston, C., Stevens, B. J., Franck, L. S., Jack, A., Stremler, R., & Platt, R. (1999). Factors explaining lack of response to heel stick in preterm newborns. *J Obstet Gynecol Neonatal Nurs, 28*(6), 587-594.

Johnston, C., & Strada, M. E. (1986). Acute pain response in infants: a multidimensional description. *Pain, 24*(3), 373-382.

Kong, G., Xu, D.-L., Body, R., Yang, J.-B., Mackway-Jones, K., & Carley, S. (2012). A belief rule-based decision support system for clinical risk assessment of cardiac chest pain. *European Journal of Operational Research, 219*(3), 564-573. doi:http://dx.doi.org/10.1016/j.ejor.2011.10.044

Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine, 23*(1), 89-109. doi:http://dx.doi.org/10.1016/S0933-3657(01)00077-X

Korhonen, A., Haho, A., & Polkki, T. (2013). Nurses' perspectives on the suffering of preterm infants. *Nurs Ethics, 20*(7), 798-807. doi:10.1177/0969733012475251

Kuziemsky, C. E., Downing, G. M., Black, F. M., & Lau, F. (2007). A grounded theory guided approach to palliative care systems design. *Int J Med Inform, 76 Suppl 1*, S141-148. doi:10.1016/j.ijmedinf.2006.05.034

Kuziemsky, C. E., Weber-Jahnke, J. H., Lau, F., & Downing, G. M. (2008). An Interdisciplinary Computer-based Information Tool for Palliative Severe Pain Management. *Journal of the American Medical Informatics Association : JAMIA, 15*(3), 374-382. doi:10.1197/jamia.M2519

Li, D., Puntillo, K., & Miaskowski, C. (2008). A review of objective pain measures for use with critical care adult patients unable to self-report. *J Pain, 9*(1), 2-10. doi:10.1016/j.jpain.2007.08.009

Lindh, V., Wiklund, U., & Hakansson, S. (1999). Heel lancing in term new-born infants: an evaluation of pain by frequency domain analysis of heart rate variability. *Pain, 80*(1-2), 143-148.

Lindh, V., Wiklund, U., Sandman, P. O., & Hakansson, S. (1997). Assessment of acute pain in preterm infants by evaluation of facial expression and frequency domain analysis of heart rate variability. *Early Hum Dev, 48*(1-2), 131-142.

Lowery, C. L., Hardman, M. P., Manning, N., Hall, R. W., Anand, K. J., & Clancy, B. (2007). Neurodevelopmental changes of fetal pain. *Semin Perinatol, 31*(5), 275-282. doi:10.1053/j.semperi.2007.07.004

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. Retrieved from http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Marek, M. (1996). Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Circulation, 93*(5), 1043-1065.

Marquand, A., Howard, M., Brammer, M., Chu, C., Coen, S., & Mourão-Miranda, J. (2010). Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. *NeuroImage, 49*(3), 2178-2189. doi:http://dx.doi.org/10.1016/j.neuroimage.2009.10.072

Mathew, P. J., & Mathew, J. L. (2003). Assessment and management of pain in infants. *Postgraduate Medical Journal, 79*(934), 438-443. doi:10.1136/pmj.79.934.438

McGregor, C. (2013). Big Data in Neonatal Intensive Care. *Computer, 46*(6), 54-59. doi:10.1109/MC.2013.157

McGregor, C., Catley, C., & James, A. (2011). *A Process Mining Driven Framework for Clinical Guideline Improvement in Critical Care* Paper presented at the Learning from Medical Data Streams 13th Conference on Artificial Intelligence in Medicine (LEMEDS).

McGregor, C., Catley, C., & James, A. (2012). *Variability analysis with analytics applied to physiological data streams from the neonatal intensive care unit*. Paper presented at the Computer-Based Medical Systems (CBMS).

McGregor, C., Catley, C., James, A., & Padbury, J. (2011). Next generation neonatal health informatics with Artemis. *Stud Health Technol Inform, 169*, 115-119.

Midboe, A. M., Lewis, E. T., Cronkite, R. C., Chambers, D., Goldstein, M. K., Kerns, R. D., & Trafton, J. A. (2011). Behavioral medicine perspectives on the design of health information technology to improve decision-making, guideline adherence, and care coordination in chronic pain management. *Transl Behav Med, 1*(1), 35-44. doi:10.1007/s13142-011-0022-6

Naik, T., Bressan, N., James, A., & McGregor, C. (2013). Design of temporal analysis for a novel premature infant pain profile using artemis. *Journal of Critical Care, 28*(1), e4.

Naik, T., & McGregor, C. (2014). *Design of an Approach for Detection of Pain by Investigating Abnormalities in Physiological Responses in Premature Infants*. Paper presented at the Canadian Student Conference on Biomedical Computing and Engineering & IEEE EMBS International Student Conference, 7th CSCBCE IEEE EMBS ISC 2014, Oshawa, Ontario.

Naik, T., Thommandram, A., Fernando, K. E. S., Bressan, N., James, A., & McGregor, C. (2014). *A Method for a Real-Time Novel Premature Infant Pain Profile Using High Rate, High Volume Physiological Data Streams*. Paper presented at the Computer-Based Medical Systems (CBMS), New York, USA.

Najarian, K., & Splinter, R. (2012). *Biomedical signal and image processing* (2 ed.). Boca Raton: CRC/Taylor & Francis.

Noy, N. F., Rubin, D. L., & Musen, M. A. (2004). Making biomedical ontologies and ontology repositories work. *Intelligent Systems, IEEE, 19*(6), 78-81. doi:10.1109/MIS.2004.67

Oberlander, T., & Saul, J. P. (2002). Methodological considerations for the use of heart rate variability as a measure of pain reactivity in vulnerable infants. *Clin Perinatol, 29*(3), 427-443.

Ohmann, C., Moustakis, V., Yang, Q., & Lang, K. (1996). Evaluation of automatic knowledge acquisition techniques in the diagnosis of acute abdominal pain. Acute Abdominal Pain Study Group. *Artif Intell Med, 8*(1), 23-36.

Padhye, N. S., Williams, A. L., Khattak, A. Z., & Lasky, R. E. (2009). Heart rate variability in response to pain stimulus in VLBW infants followed longitudinally during NICU stay. *Dev Psychobiol, 51*(8), 638-649. doi:10.1002/dev.20399

Petroni, M., Malowany, S. A., Johnston, C., & Stevens, B. J. (1995). Identification of Pain from Infant Cry Vocalizations using Artificial Neural Networks (ANNs). *The International Society for Optical Engineering*(2).

Prevention and management of pain and stress in the neonate. . (2000). *Pediatrics (American Academy of Pediatrics. Committee on Fetus and Newborn), 105*(2), 454-461.

Prevention and Management of Pain in the Neonate: An Update. (2006). *PEDIATRICS, 118*(5), 2231-2241. Retrieved from doi:10.1542/peds.2006-2277

Ranger, M., Johnston, C. C., & Anand, K. J. (2007). Current controversies regarding pain assessment in neonates. *Semin Perinatol, 31*(5), 283-288. doi:10.1053/j.semperi.2007.07.003

Riddell, R. R. P., & Craig, K. D. (2007). Judgments of infant pain: the impact of caregiver identity and infant age. *J Pediatr Psychol, 32*(5), 501-511. doi:10.1093/jpepsy/jsl049

Roofthooft, D. W., Simons, S. H., Anand, K. J., Tibboel, D., & van Dijk, M. (2014). Eight years later, are we still hurting newborn infants? *Neonatology, 105*(3), 218-226. doi:10.1159/000357207

Rudovic, O., Pavlovic, V., & Pantic, M. (2013). Automatic pain intensity estimation with heteroscedastic conditional ordinal random fields. *Lecture notes in computer science*, 1-10.

Slater, R., Cantarella, A., Franck, L., Meek, J., & Fitzgerald, M. (2008). How well do clinical pain assessment tools reflect pain in infants? *PLoS Med, 5*(6), e129. doi:10.1371/journal.pmed.0050129

Slater, R., Cantarella, A., Gallella, S., Worley, A., Boyd, S., Meek, J., & Fitzgerald, M. (2006). Cortical pain responses in human infants. *J Neurosci, 26*(14), 3662-3666. doi:10.1523/jneurosci.0348-06.2006

Stevens, B., & Johnston, C. (1994). Physiological responses of premature infants to a painful stimulus. *Nurs Res, 43*(4), 226-231.

Stevens, B., Johnston, C., Franck, L., Petryshen, P., Jack, A., & Foster, G. (1999). The efficacy of developmentally sensitive interventions and sucrose for relieving procedural pain in very low birth weight neonates. *Nurs Res, 48*(1), 35-43.

Stevens, B., Johnston, C., & Horton, L. (1994). Factors that influence the behavioral pain responses of premature infants. *Pain, 59*(1), 101-109.

Stevens, B., Johnston, C., Petryshen, P., & Taddio, A. (1996). Premature Infant Pain Profile: development and initial validation. *Clin J Pain, 12*(1), 13-22.

Stiles, J., & Jernigan, T. L. (2010). The Basics of Brain Development. *Neuropsychol Rev, 20*(4), 327-348. doi:10.1007/s11065-010-9148-4

Tejman-Yarden, S., Levi, O., Beizerov, A., Parmet, Y., Nguyen, T., Saunders, M., . . . Moeller-Bertram, T. (2016). Heart rate analysis by sparse representation for acute pain detection. *Med Biol Eng Comput, 54*(4), 595-606. doi:10.1007/s11517-015-1350-3

Van Dijk, M., & Tibboel, D. (2012). Update on pain assessment in sick neonates and infants. *Pediatr Clin North Am, 59*(5), 1167-1181. doi:10.1016/j.pcl.2012.07.012

Walter, S., Gruss, S., Ehleiter, H., Junwen, T., Traue, H. C., Werner, P., . . . Moreira da Silva, G. (2013, 13-15 June 2013). *The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system.* Paper presented at the Cybernetics (CYBCONF), 2013 IEEE International Conference on Cybernetics.

Wand, Y., & Weber, R. (2004). Reflection: Ontology in Information Systems - Foreword. *Journal of Database Management, 15*(2).

Worley, A., Fabrizi, L., Boyd, S., & Slater, R. (2012). Multi-modal pain measurements in infants. *J Neurosci Methods, 205*(2), 252-257. doi:10.1016/j.jneumeth.2012.01.009

Yuan, L., Bao, F. S., & Guanming, L. (2008). *Recognition of Neonatal Facial Expressions of Acute Pain Using Boosted Gabor Features*. Paper presented at the 20th IEEE International Conference on Tools with Artificial Intelligence