The Dynamic Relationship between Photoplethysmography Features and Intracranial Pressure in Patients with Traumatic Brain Injury

---

School of Science and Technology
Research Centre for Biomedical Engineering

A thesis submitted in partial fulfilment for the degree of
Doctor of Philosophy

City, University of London
2024

Supervisor:
Prof. Panicos Kyriacou

Candidate:
George R.E. Bradley

Declaration of Authorship

I, George R.E. Bradley, hereby affirm that the research conducted and presented in the thesis entitled *"The Dynamic Relationship between Photoplethysmography Features and Intracranial Pressure in Patients with Traumatic Brain Injury"* is entirely my own, with the exception of cited contributions. This work was completed during my candidature at City, University of London, and has not been submitted for any other academic degree or award.

Signature: **Date:**

George R.E. Bradley 01/04/2024

Acknowledgements

Firstly, I extend my thanks and deep appreciation to my first supervisor, Professor Panicos Kyriacou, for granting me the opportunity to undertake this research. His unwavering support, guidance, encouragement, and reassurance have been instrumental throughout.

Additionally, my sincere gratitude goes to the *George Daniels Educational Trust* for their generous support through the *George Daniels Doctoral Studentship*, which afforded me the ability to pursue this research.

My thanks also extend to my colleagues at the Research Centre for Biomedical Engineering, who have become close friends. Their motivation, support, and friendship have both deeply enriched me and my PhD journey, making it an immensely enjoyable experience.

"In all chaos there is a cosmos, in all disorder a secret order" - Carl Jung

Contents

List of Tables

List of Figures

Glossary of Abbreviations

ABI: Acquired Brain Injury

AC: Alternating current

ACRM: American Congress of Rehabilitation Medicine

BTF: American Brain Trauma Foundation

CBF: Cerebral blood flow

CDC: US Centers for Disease Control and Prevention

CPP: Cerebral perfusion pressure

CSF: Cerebrospinal fluid

CT: Computed tomography

DC: Direct current

DCS: Diffuse correlation spectroscopy

EPDA: Envelope Photoplethysmography Denoising Algorithm

EVD: Extraventricular drain

GCS: Glasgow Coma Scale

GBM: Gradient Boosting Machine

GOS-E: Glasgow Outcome Scale - Extended

HICs: High-income countries

ICH: Intercranial hypertension

ICP: Intracranial pressure

ICU: Intensive Care Unit

IPS: Instances per second

IQR: Inter-quartile range

ITU: Intensive therapy unit

KNN: K-Nearest Neighbours

LGBM: Light Gradient Boosting Machine

LMICs: Low- and middle-income countries

LOPOCV: Leave-one-patient-out cross-validation

MAE: Mean absolute error

MAP: Mean arterial pressure

MIC: Maximum Information Coefficient

ML: Machine learning

MRI: Magnetic resonance imaging

MSE: Mean square error

mmHg: Millimeters of mercury

mTBI: Mild TBI

NIR: Near-infrared

NIR-PPG: Near-infrared photoplethysmography

NICE: The National Institute for Health and Care Excellence

NIRS: Near-infrared spectroscopy

nICP: Non-invasive intracranial pressure sensor

ONSD: Optical nerve sheath diameter

PPG: Photoplethysmography

RF: Random Forest

RMSE: Root mean square error

RFE: Recursive Feature Elimination

rSO2: Regional cerebral oxygen saturation

SNR: Signal-to-noise ratio

SpO2: Oxygen Saturation

SVM: Support Vector Machine

SVR: Support Vector Regression

TBI: Traumatic brain injury

TCD: Transcranial Doppler

TV: Total variation

USA: United States of America

WHO: World Health Organization

XGB: eXtreme Gradient Boosting

Publications

Journals

1. G. R. E. Bradley and P. A. Kyriacou, "Exploring the Dynamic Relationship: Changes in Photoplethysmography Features Corresponding to Intracranial Pressure Variations" [In Review].

2. G. R. E. Bradley and P. A. Kyriacou, "Evaluating the effectiveness of non-invasive intracranial pressure monitoring via near-infrared photoplethysmography using classical machine learning methods," Biomedical Signal Processing and Control, vol. 96, no. 14, pp. 106517, 2024, doi: https://doi.org/10.1016/j.bspc.2024.106517.

3. G. R. E. Bradley and P. A. Kyriacou, "Opening the envelope: Efficient envelope-based PPG denoising algorithm," Biomedical Signal Processing and Control, vol. 88, pp. 105693–105693, Feb. 2024, doi: https://doi.org/10.1016/j.bspc.2023.105693.

4. M. Roldan, G. R. E. Bradley, E. Mejía-Mejía, T. Y. Abay, and P. A. Kyriacou, "Non-invasive monitoring of intracranial pressure changes: healthy volunteers study," Frontiers in physiology, vol. 14, Aug. 2023, doi: https://doi.org/10.3389/fphys.2023.1208010.

5. G. R. E. Bradley, M. Roldán, and P. A. Kyriacou, "Machine Learning Approaches to Intracranial Pressure Prediction in Patients with Traumatic Brain Injury: A Systematic Review," Applied sciences, vol. 13, no. 14, pp. 8015–8015, Jul. 2023, doi: https://doi.org/10.3390/app13148015.

Abstract

Traumatic brain injury (TBI) is as an alteration in brain function pathology by a sudden trauma, causing damage to the brain. Symptoms vary from mild to severe, depending on the extent of the damage to the brain. Intracranial pressure (ICP) monitoring is a "gold standard" for severe TBI patients, measuring pressure inside the skull. An ICP crisis is a sustained ICP value above a threshold of 20-25 mmHg. Effective ICP monitoring and intervention at defined thresholds can reduce mortality and secondary brain injury. The gold standard for ICP monitoring involves invasive neurosurgical intervention to implant a pressure sensor into the brain, which is expensive, carries risk, requires expertise and is only accessible within a hospital setting. Consequently, timely and effective monitoring is restricted, exacerbating adverse outcomes, and excluding millions of mild and moderate TBI cases from continuous and quantitative assessment.

There is a nascent body of research investigating non-invasive ICP monitoring aiming to reduce the barrier to entry to efficacious monitoring and intervention for patients and healthcare systems. A growing body of research is exploring the use of Photoplethysmograph (PPG) for the non-invasive estimation of ICP. This thesis makes novel contributions to knowledge by evaluating the relationship between PPG derived features and variations in ICP using the largest clinically collected, labelled PPG dataset, to-date. The novel PPG data, produced by an in-house, Near-infrared spectroscopy, reflectance, non-invasive optical ICP sensor, was collected from 40 TBI patients at The Royal London Hospital. Data is categorised based on "proximal" and "distal" photodiodes, hypothesised to correspond to extracerebral and cerebral data, respectively.

The research makes novel contributions to the field via the testing of three main hypotheses: (i) PPG feature alterations correlate with ICP changes, (ii) distal PPG data shows stronger correlations with ICP than proximal data, and (iii) PPG features can estimate ICP non-invasively.

A total of 141 features were extracted for each one-minute window of PPG data, including the original waveform and its first and second derivatives. Spearman's correlation and the Kruskal-Wallis test evaluated the first two hypotheses. Results indicated significant correlations between PPG features and ICP levels, with 77.30% and 79.43% of features significantly correlated ($p < 0.05$) for distal and proximal datasets, respectively. Group analysis revealed significant changes across ICP groups (0-10, 10-20, 20-39 mmHg) in 81.56% and 75.89% of features. The mean absolute correlation of all features and significantly correlated features was 25.76% and 24.24% higher for distal than for proximal features, supporting the potential of PPG-based ICP monitoring.

To test the third hypothesis, five classical machine learning models were implemented, optimised, and assessed across six key metrics using a leave-one-patient-out cross-validation approach. Distal models outperformed proximal ones, with the best model, a Random Forest, achieving a mean RMSE of 5.030 mmHg, MAE of 4.067 mmHg, and Bland-Altman limits of agreement around 8.5 mmHg, and a low correlation coefficient of -0.007.

There is a need for the development of a continuous noninvasive, easy-to-use, inexpensive monitoring device. Such a device would reduce the barrier to entry for ICP monitoring across all severities, providing timely diagnosis whilst serving the currently undeserved majority of TBI cases. This work provides a credible foundation for further research in this domain.

Chapter 1

Introduction

1.1 Research Motivation

Traumatic brain injury (TBI) is an alteration in brain function pathology resulting from a sudden trauma, causing damage to the brain. The severity of symptoms are categorised as mild, moderate, or severe, depending on the extent of the damage to the brain [1].

TBI constitutes a significant burden to healthcare systems with an estimated global incidence rate of up to 69 million cases per annum of which 5.48 million are estimated to be severe [2]. The majority of TBI cases are mild, which despite being low in severity can have serious, long term impact on patients. TBI not only causes a significant burden for healthcare systems and nations but also to individuals and their families due to the complex and expensive medical care the condition necessitates. The total annual economic cost of TBI was estimated at US $400 billion [3].

Currently, after head injury, patients undergo assessment using the Glasgow Coma Scale (GCS) which evaluates verbal, eye and motor responses. Upon admission to hospital patients considered to be at high risk for clinically important TBI undergo a clinical examination to establish any need for evaluation using computed tomography (CT) scans or magnetic resonance imaging (MRI). In severe cases, if the resulting scans are abnormal, intracranial pressure (ICP) monitoring is initiated.

Within neurocritical care, ICP monitoring is a clinical cornerstone for the management of severe TBI. ICP monitoring measures the pressure inside the skull. Elevated ICP, often defined as sustained ICP above 20-25 millimeters of mercury (mmHg), can precipitate further brain injury and increase mortality rates. Studies consistently link elevated ICP

with poor neuropsychological performance and functional outcomes [4]–[6]. Efficacious ICP monitoring and intervention at clinically defined thresholds may reduce mortality and secondary injury to the brain. The gold standard for ICP monitoring is invasive, necessitating neurosurgical intervention to implant a pressure sensor via a burr hole in the skull. The gold standard is expensive, high risk, requires a high level of expertise and is only accessible within a hospital setting. Consequently, this restricts access to timely and effective monitoring, exacerbating the potential for adverse outcomes, particularly in severe TBI cases and leaves the majority of cases, constituted by mild TBI under-served.

No method to accurately, non-invasively and continuously monitor ICP currently exists. This work is motivated by the clinical need and desire for an easy-to-use, non-invasive, continuous ICP monitor to cost effectively facilitate, accurate and timely triage and intervention for TBI patients, helping reduce the barrier to entry for patients and healthcare systems whilst increasing patient care outcomes.

1.2 Research aim & objectives

The aim of this research is to evaluate the dynamic relationship between photoplethysmography (PPG) features and ICP in patients with TBI using a labeled dataset of PPG signals collected from patients' foreheads (below the hairline), with labels derived from invasive, gold-standard ICP data. This aim is fulfilled via the testing of three main hypotheses: (i) if alterations in PPG features correlate with changes in ICP levels, (ii) if the association between PPG features and variations in ICP levels is stronger in long-distance PPG data compared to short-distance PPG data (PPG data was collected from two photodetectors at two different distances from the photodiodes, data is referred to as either "short-distance" or "long-distance" data based on photodiode-photodetector separation), and (iii) if PPG derived features can accurately estimate ICP non-invasively. In order to test these hypotheses and fulfill the aim the following objectives are set:

- Evaluate and identify gaps within the existing research on PPG driven non-invasive ICP monitoring.
- Format and preprocess raw clinical data, including both non-invasive PPG and invasive ICP measurements, ensuring temporal synchronicity between the datasets.
- Separate the data into two datasets of short-distance and long-distance PPG data.

- Implement, apply and evaluate a denoising algorithm to effectively identify and remove noise within the non-invasive PPG signal.
- Implement an algorithm for the detection of fiducial points of each cardiac cycle within the PPG data.
- Develop and apply algorithms for the extraction of morphological and time-series features from the PPG's original waveform, as well as its first and second derivatives.
- Refine the dataset to create a subset facilitating a robust analysis of the relationship between PPG features and ICP levels.
- Statistically evaluate the relationship between short and long-distance PPG derived features and variations in ICP levels.
- Divide each set of short and long-distance PPG features into separate "evaluation" and "optimisation" datasets.
- Conduct feature selection on the optimisation feature sets using a combination of collinearity testing, Maximum Information Coefficient (MIC) ranking, and Recursive Feature Elimination (RFE).
- Implement, optimise and train five classical machine learning models for estimating ICP values using the derived feature sets.
- Evaluate and compare the performance of the five models on both the short and long distance PPG datasets using commonly employed statistical metrics.

1.3 Contribution to knowledge

The main contributions to knowledge of this work include:

- The development of a novel, effective and efficient PPG denoising algorithm which handles labelled datasets.
- To the knowledge of the author, the largest group of features to be extracted from the PPG signal for the investigation of the relationship between PPG features and ICP.

- A novel statistical analysis of the dynamical relationship between PPG features using the largest, clinically collected, labelled dataset of its kind, to-date.
- To the knowledge of the author, the largest group of machine learning models to be implemented and evaluated for the predictive task of estimating ICP via PPG derived features.
- Findings from the statistical analysis and evaluation of machine learning models using data from both photodetectors of the PPG sensor can be used to inform future sensor design and development.

1.4 Thesis outline

Chapter 2: Background and Literature Review

Chapter 2 provides a comprehensive exploration of TBI and its implications on cerebral physiology and ICP. It begins with an in-depth examination of TBI, covering its epidemiology, demographics, economic burden, and assessment methods, such as the GCS. Following this, the foundational concepts of cerebral physiology and hemodynamics are introduced, with a focus on the relationship between TBI and variations in ICP. Both invasive and non-invasive approaches to ICP monitoring are then reviewed, highlighting their methodologies, strengths and limitations. The chapter concludes with an introduction to PPG, detailing its principles and the potential utility of PPG waveforms in non-invasive ICP estimation, including the sensor central to this research.

Chapter 3: Dataset and Pre-processing

Chapter 6 discusses the core dataset driving this research, describing its collection, structure, and challenges. It details the formatting, denoising, and pre-processing of the raw data.

Chapter 4: Feature Engineering

Chapter 7 introduces and explains the development of algorithms for identifying fiducial points and conducting feature engineering.

Chapter 5: Statistical analysis

Chapter 8 outlines the statistical analysis employed to assess the relationship between morphological and time-series PPG features and variations in ICP levels. This chapter evaluates the two hypotheses: (i) if alterations in PPG features correlate with changes in ICP levels, and (ii) if the association between PPG features and variations in ICP levels is stronger in long-distance PPG data compared to short-distance PPG data.

Chapter 6: Classical Machine Learning Models

Chapter 9 introduces and explains the workings of the five classical machine learning models implemented within this research, drawing attention to the main parameters of each algorithm and their effects on model performance.

Chapter 7: Evaluation of Machine Learning Models

Chapter 10 presents the methodology and results for feature selection, hyperparameter optimisation and evaluation and comparison of results obtained from the implemented models across datasets. It employs various performance evaluation metrics to assess the models' efficacy in utilising PPG features for non-invasive ICP estimation.

Chapter 8: Discussion and Conclusion

Chapter 11 serves as the culmination of the thesis, offering a summary of key findings, a discussion, and comparison with related works. It presents the strengths and limitations of the research providing recommendations for future research endeavours in this domain.

Chapter 2

Background and Literature Review

This chapter aims to provide a comprehensive understanding of the epidemiology, clinical principles, and cerebral physiology relevant to TBI and its monitoring. It begins by defining TBI, describing its types, epidemiology, and the societal and economic burden associated with the condition. Next, it delves into the fundamental cerebral physiological and hemodynamic components, highlighting how fluctuations in ICP can impact cerebral dynamics and lead to additional brain injuries or adverse patient outcomes. The chapter then discusses the gold standard of invasive ICP monitoring, alongside other invasive and non-invasive methods, evaluating the advantages and limitations of each approach. Finally, it introduces PPG and its basic principles, focusing on the non-invasive ICP monitoring sensor central to this research, and explores the theoretical relationship between PPG and ICP in the context of cerebral physiology and hemodynamics.

2.1 Traumatic Brain Injury

2.1.1 Definition, types and severity

Acquired Brain Injury (ABI) characterises any damage to the brain that occurs after birth, which can be categorised into traumatic and non-traumatic causes. Non-traumatic ABI may include, stoke, infectious disease, oxygen deprivation, or degenerative neurological conditions. TBI is a form of ABI caused by external mechanical forces. TBI occurs when a sudden trauma causes damage to the brain. It can result from "a forceful bump, blow, or jolt to the head or body, or from an object that pierces the skull and enters the brain" [7]. Depending on the extent of the damage to the brain, symptoms can be mild, moderate,

or severe [1]. TBI can be categorised into two broad injury types: (i) closed TBI and (ii) open TBI. Closed brain injury occurs when there is a non-penetrating injury to the brain with no break in the skull but with a force strong enough to move the brain within the skull. Open TBI also referred to as penetrating brain injury occurs when an object pierces the skull and enters the tissue of the brain.

The GCS is a widely used clinical tool designed to assess a patient's level of consciousness, guide treatment following a TBI and to gauge the severity of the injury. GCS is discussed in detail in section 2.1.3 but referenced here for use in description of TBI severity. GCS scores patients based on their verbal, motor, and eye-opening responses to stimuli, with the total score ranging from 3 (indicating deep unconsciousness) to 15 (indicating full consciousness). A lower score suggests a more severe injury.

Mild TBI

There is no consensus definition of mild TBI (mTBI) despite it making up the majority of TBI cases (81% to $>$ 90%) [2], [8]. The three leading definitions of mTBI are outlined by the American Congress of Rehabilitation Medicine (ACRM), the US Centers for Disease Control and Prevention (CDC), and the World Health Organization (WHO) [9]. While there is variability in these definitions, there are commonalities which can be used to create a general definition of mTBI. These include: (i) an alteration in mental state or consciousness, and normal structural imaging, (ii) a loss of consciousness lasting less than 30 minutes, (iii) a GCS score ranging from 13 to 15, and (iv) post-traumatic amnesia not exceeding 24 hours. The ACRM's expert panel, with a 93.8% agreement, posits that the term "concussion" can be used synonymously with mTBI provided that neuroimaging results are normal or such imaging is not deemed clinically necessary[10].

One of the primary symptoms of mTBI is an alteration in mental state, manifesting as headaches, confusion, irritability, and dizziness, among others. The subjective nature of assessment of these symptoms and the absence of objective markers significantly complicate the diagnosis, leading to the under-reporting of the condition. Mild TBI is particularly prevalent among military personnel and contact sport athletes, with studies indicating that up to 56% of concussions in military contexts and between 30% to 50% in sports settings go unreported [11]–[14].

The implications of mTBI extend beyond immediate health concerns, posing serious

long-term consequences. An unselected observational cohort study across three centers in the United States evaluated the 3-month and 6-month month outcomes for 485 patients with mTBI (GCS scores between 13 and 15). The findings revealed that at 3 months post-injury, 34% of the cohort had either died (2%), or were experiencing moderate (28%) to severe (4%) disabilities. The 6-month data, of the patients who continued in the study indicate that, 3% had died, while 30% and 3% were found to be moderately and severely disabled, respectively [15]. In a separate cohort study of 1104 mTBI patients assessed with the Glasgow Outcome Scale - Extended (GOSE) at similar intervals, 23% and 16% of patients in the uncomplicated mTBI group (n=569) were considered impaired at 3 and 6 months post-injury, respectively. This rate of impairment notably increased among patients with complicated mTBI (n=535), with 41% and 35% of patients being considered impaired at 3 and 6 months post-injury, respectively [16].

Moderate and severe TBI

Moderate and severe TBIs are distinguished using similar diagnostic criteria. Specifically, moderate TBI is identified by a loss of consciousness lasting from 30 minutes to 24 hours, alterations in consciousness or mental state extending beyond 24 hours, post-traumatic amnesia persisting for 1 to 7 days, and a GCS score ranging from 9 to 12. Severe TBI, in contrast, involves a loss of consciousness for more than 24 hours, post-traumatic amnesia exceeding 7 days, and a GCS score between 2 and 8. Both conditions may present with either normal or abnormal structural imaging findings [17]. A moderate to severe TBI can significantly increase ICP beyond the normal range of 10 to 15 mmHg for adults and 3 to 7 mmHg for young children [18]. Elevated ICP can precipitate multiple adverse events, including cerebral herniation, ischemia, and or death.

2.1.2 Epidemiology

Global incidence

The global annual incidence of all-cause, all-severity TBI is estimated to range between 50 to 69 million cases [2], [8]. TBI has the highest incidence rate of all common neurological disorders. Figure 2.1 depicts the global incidence rate of TBI compared with other common neurological disorders. The distribution of TBI severity is predominantly mild (81%) and moderate (11%), with an estimated 5.48 million cases annually classified as

severe [2]. Regions with the lowest resources bear the highest disease burden, with low- and middle-income countries (LMICs) experiencing three to four times as many TBIs as high-income countries (HICs) [19]. Southeast Asia and the Western Pacific report the greatest incidence rates, with 18.3 million and 17.3 million cases per year, respectively [2]. In contrast, Europe records more than 2 million hospital admissions due to TBI each year, resulting in approximately 82,000 fatalities [8]. Statistics published by the CDC indicate that 3.5 million people sustain a TBI in the United States of America (USA) annually [20], where at least 25% of these cases are diagnosed as either moderate or severe [21].

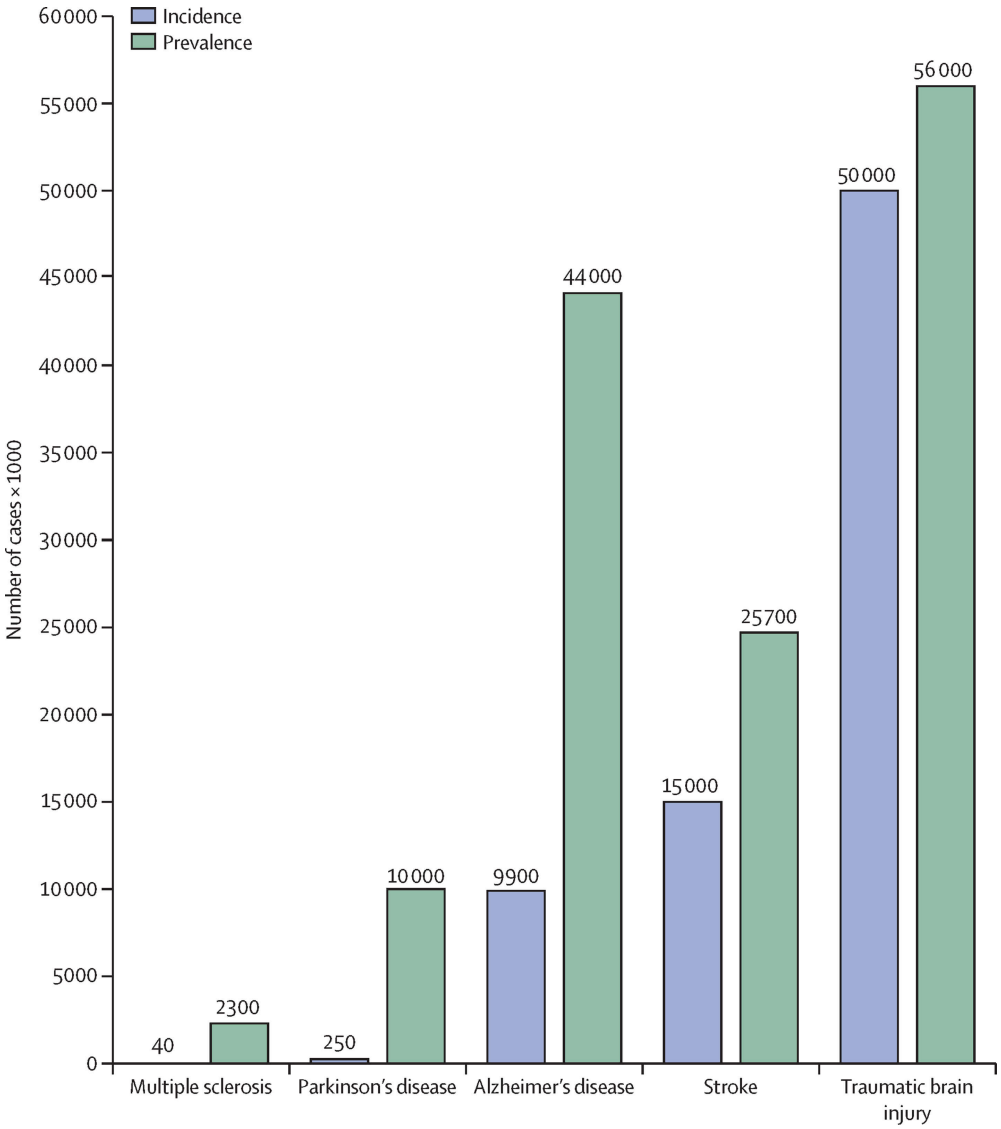

Figure 2.1: Global incidence and prevalence of traumatic brain injury compared with other common neurological diseases [8].

Trends

The incidence rate of TBI is increasing. A retrospective cohort study in Nova Scotia, evaluating severe TBI cases from 2002 to 2018, reported a 39% increase in the incidence of severe TBI over the study period, totaling 5,590 patients [22]. Similarly, data from California's Office of Statewide Health Planning and Development between 2005 and 2014 showed a 57.7% increase in TBI-related emergency department visits, equating to a 40.5% rise in TBI visit rates over a decade [23].

Mechanism of injury

The primary cause of TBI varies significantly between LMICs and HICs. In LMICs, road traffic collisions are the leading cause of TBI [24]. Conversely, in HICs, falls constitute the majority of TBI cases. The CENTER-TBI registry, which includes data from 56 acute trauma receiving hospitals across 17 countries in Europe and Israel, found that 56% of TBI cases resulted from falls, 71% of which were ground-level falls [25]. The study by Kureshi et al. further supports this finding, suggesting falls as the predominant mechanism of injury among severe TBI patients, where the rate of fall-related TBIs more than doubled from 2002 to 2017 [22].

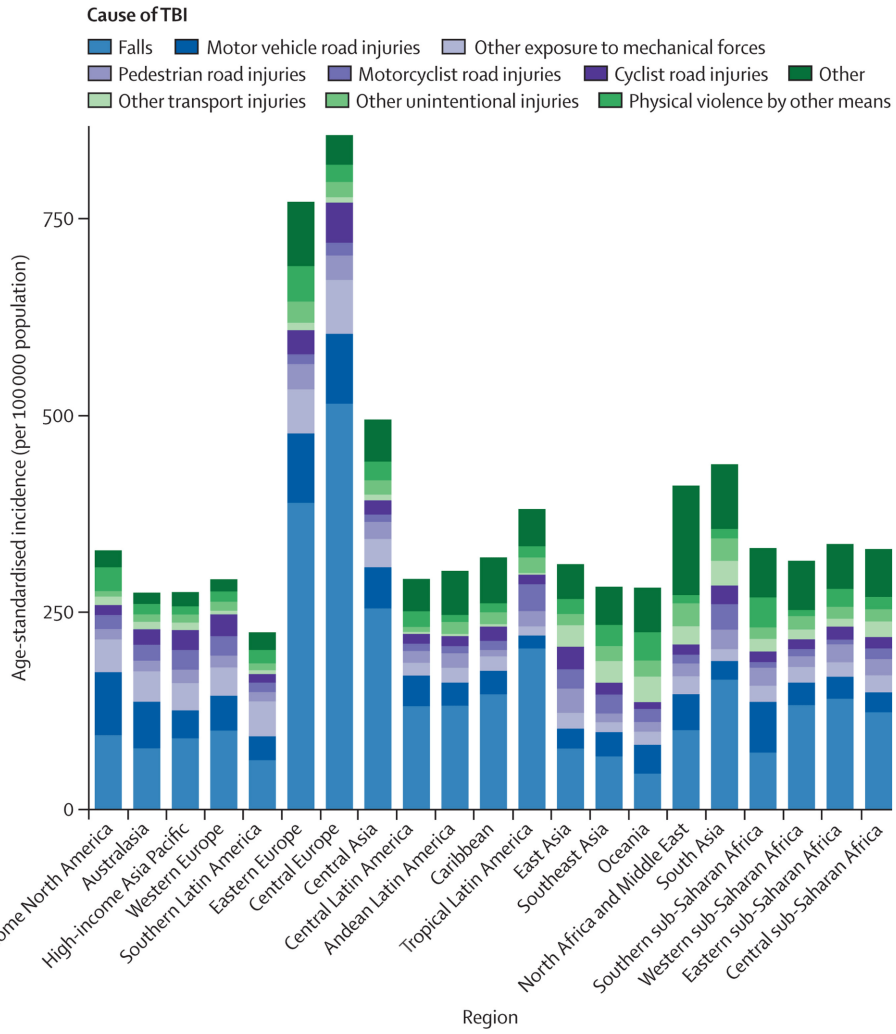

Figure 2.2: Cause composition of age-standardised incidence of traumatic brain injury by Global Burden of Disease region for both sexes, 2016 [24].

Risk factors

The growing and aging population is reflected in the risk factors of TBI. Hospital admissions for TBI predominantly affect older individuals (aged $\geq$ 65 years), with the incidence notably higher in this demographic compared to any other age groups [26]. The CDC report that older adults constitute 43.9% of all TBI-related hospital admissions in the USA. Similarly, data from China indicate 18.3% of TBI patients are from an older cohort [27].

Conversely, children and adolescents (0–19 years) represent the second-highest incidence group for TBI hospital admissions, with the European Union (EU) reporting approximately 345 hospital admissions per 100,000 children or adolescents annually [28]. In the United States, the annual incidence of mild TBI in this age group ranges between 1 million and 2 million [29].

Violence emerges as the third-most common cause of TBI, as illustrated by findings from the CENTER-TBI and TRACK-TBI studies, with approximately 6.7 to 13% of TBIs attributed to violence [30].

A cohort study examining historical data found that former professional football players have a neurodegenerative disease mortality rate approximately triple that of general population controls, suggesting significantly elevated risks within this group [31]. Furthermore, evidence indicates that the likelihood of experiencing a sport-related concussion is markedly higher in contact sports, with rugby and American football identified as particularly high-risk activities [32].

Economic burden

The in-hospital treatment of patients with TBI is considered to be expensive, especially in patients with severe TBI. The estimated total global annual burden being US $400 billion [3]. In 2010 in the USA, the TBI-related in-hospital charges totalled US 21.4 billion [33]. These patients also have the longest hospital or intensive care unit (ICU) length of stay (LOS) and the highest number of surgical and medical interventions [34], [35].

2.1.3 Severity assessment and classification

The National Institute for Health and Care Excellence (NICE) provide gold standard guidelines for the assessment of TBI severity and injury classification [36]. According to these guidelines initial physical examination is conducted using the GCS followed by a computerised tomography (CT) scan for the classification of brain injury if indicated by the initial examination results.

Glasgow Coma Scale

The GCS was developed in 1974, with the purpose of providing a tool for the assessment of level of consciousness and coma [37]. A GCS score is calculated by summing the scores of three categories: (i) eye opening response (score 1–4), (ii) verbal response (score 1–5) and (iii) motor response (score 1–6). Scores between 3–8 are defined as severe, scores between 9–12 as moderate and scores between 13–15 as mild. Table 2.1 depicts the GCS score matrix.

Glasgow Coma Scale (Categories and Scores)

Eye opening response	Verbal response	Motor Response
4: Spontaneous	5: Oriented	6: Obeys commands
3: To verbal stimuli	4: Confused	5: Localizes pain
2: To pain	3: Inappropriate words	4: Withdraws from pain
1: None	2: Incoherent	3: Flexion to pain or decortricate
	1: None	2: Extension to pain or decerebrate
		1: None

Table 2.1: Table of Glasgow Coma Scale categories and scoring.

The GCS is the gold standard clinical examination tool which categorises TBI as mild, moderate, or severe. Genenarelli et al's study which analysed data from 174,160 patients submitted from 165 trauma centres demonstrated a progressive association between increasing mortality after TBI and decreases in GCS score [38]. Once the severity of a patient has been assessed using the GCS, if the resultant GCS score indicates, further examination is carried out using brain imaging techniques which provides further information on the extent and type of injury to the brain.

Computerised tomography head scan

The gold standard imaging technique is CT [39], CT remains the technique of choice for initial assessment due to its speed, accessibility, and sensitivity for depicting brain injuries that require neurosurgical intervention, such as large volume haemorrhage [40], [41]. Additionally unlike MRI there is no need to screen for safety when using CT imaging[42].

The criteria for conducting a CT head scan vary based on age groups and specific risk factors following head injuries. For individuals aged 16 and over, a CT head scan is recommended within one hour if they exhibit certain risk factors such as a GCS score of 12 or less, signs of skull fracture, post-traumatic seizure, or other neurological deficits. For those under 16, the recommendation is also within one hour for suspicion of non-accidental injury, post-traumatic seizure, or specific GCS scores. Further, for this age group, if multiple risk factors are present, a CT head scan is advised within one hour. Additionally, observation is suggested for those with one identified risk factor, with further scans indicated if specific symptoms arise during observation [36].

2.1.4 Summary

TBI represents a significant global health concern with an estimated annual incidence ranging between 50 to 69 million cases worldwide. The burden of TBI disproportionately affects LMICs, where incidences are three to four times higher than in high-income countries, primarily due to the prevalence of road traffic collisions in LMICs versus falls in higher-income regions. The majority of these cases are mTBI, accounting for 81% to over 90% of all TBI incidents. However, the diagnosis of mTBI remains challenging due to its subjective nature and the reliance on self-reported symptoms, making under-reporting a significant issue, especially among military personnel and athletes. Although the GCS serves as a crucial tool in assessing TBI severity, its lack of precision highlights the necessity for more refined diagnostic methods. Given the vast scale of TBI incidence, the specific challenges posed by mTBI, and the limitations of current assessment tools like the GCS, there's a compelling need for the development of a simple, inexpensive, and non-invasive monitoring device. Such a device should be universally accessible and capable of effectively serving the majority of TBI cases, including settings with limited healthcare resources, to address both the acute and long-term consequences of all severities of TBI.

2.2 Intracranial Pressure, Cerebral Physiological and Haemodynamic Components

The brain's function and structural integrity are directly influenced by its physiological and hemodynamic environment, including factors such as cerebral blood flow (CBF), cerebral perfusion pressure (CPP), mean arterial pressure (MAP), and the mechanism of cerebral autoregulation. These elements are interdependent and crucial for maintaining the balance within the cranial vault. Disruptions in these parameters, especially following TBI, can lead to significant changes in ICP, impacting patient outcomes.

2.2.1 The Monro-Kellie Doctrine and ICP

ICP is the pressure within the cranial vault, governed by the Monro-Kellie hypothesis, which posits that the cranial compartment's volume is fixed, comprising brain tissue, cerebral blood, and cerebrospinal fluid (CSF) [43]. Variations in ICP arise from changes

in these components due to pathological conditions such as mass lesions, venous sinus obstructions, and cerebral oedema. Figure 2.3 illustrates the progressive relationship between the increase in volume of the intracranial pathologies within the cranial vault (depicted in red) which can be caused by TBI and the corresponding rise in ICP.

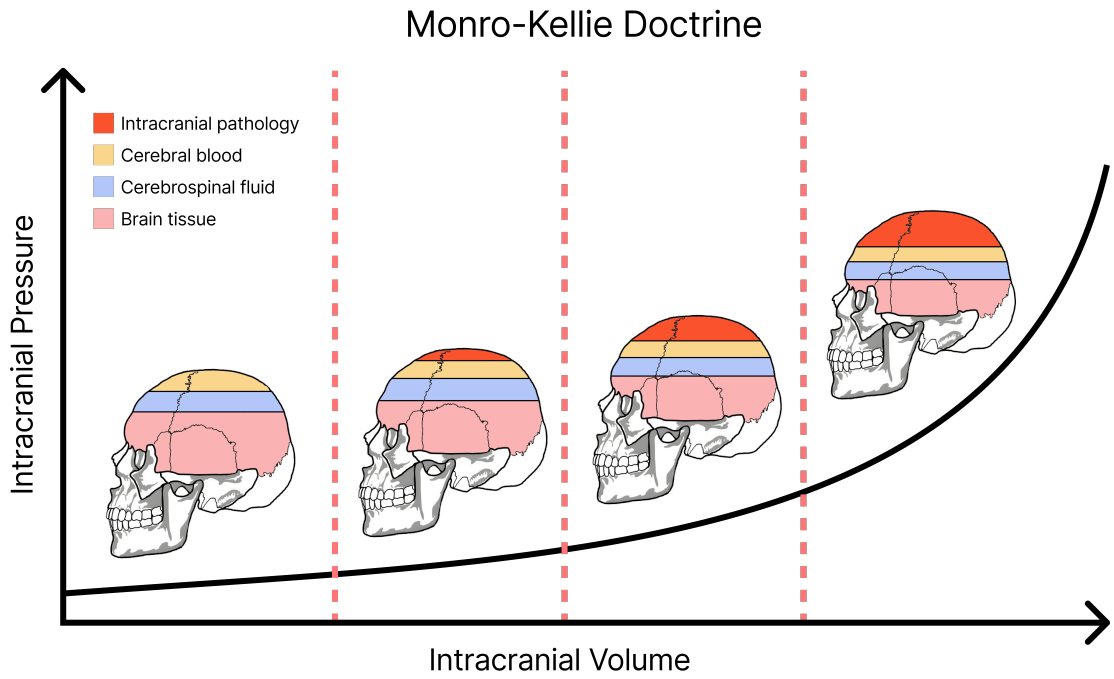

Figure 2.3: Illustration of the relationship between increasing intracranial pathologies and consequential increases in intracranial pressure. The four skull images represent different stages of volume increase within the cranial compartments due to the increasing volume of intracranial pathologies mirrored by a subsequent increase in intracranial pressure

These changes can trigger compensatory mechanisms aimed at preserving normal ICP levels, which are typically between 10 to 15 mmHg in adults and 3 to 7 mmHg in children [18]. Within this framework, cerebral compliance, or the brain's ability to accommodate volume changes without significant variations in ICP, plays a crucial role. It reflects the adaptability of the intracranial contents and their capacity to compensate for volume increases, thereby maintaining a stable ICP within physiological limits. Reduced cerebral compliance indicates a diminished capacity for such compensation, making the brain more susceptible to ICP fluctuations and the potential for secondary injuries, underlining the importance of monitoring these parameters closely in patients with potential or actual brain injuries.

2.2.2 Cerebral Perfusion Pressure

CPP is defined as the difference between MAP and ICP (or central venous pressure (CVP) if greater than ICP), expressed in millimeters of mercury (mmHg). CPP is calculated as:

$$CPP = MAP - ICP.$$

This differential is fundamental for ensuring adequate blood flow to the brain, providing the oxygen and nutrients necessary for cerebral metabolism. Normal CPP values range from 60 to 100 mmHg, maintaining adequate cerebral blood flow and ensuring brain health [44].

2.2.3 Cerebral Blood Flow

CBF represents the volume of blood flowing through a given amount of brain tissue over time, measured in milliliters per 100 grams of brain tissue per minute (ml/100g/min). Cerebral autoregulation plays a key role in maintaining CBF by adjusting cerebral vascular resistance to ensure consistent blood flow despite variations in CPP, within the autoregulatory limits of approximately 50 to 150 mmHg [44].

2.2.4 Mean Arterial Pressure

MAP is a critical determinant of CPP and hence, indirectly affects CBF. It is the average pressure in a patient's arteries during one cardiac cycle and is influenced by factors such as blood volume, cardiac output, and vascular resistance [45].

2.2.5 Autoregulation

Cerebral autoregulation refers to the brain's ability to maintain a relatively constant CBF despite changes in CPP [46]. This mechanism works through the dilation and constriction of cerebral blood vessels in response to changes in blood pressure, ensuring stable oxygen and glucose delivery to brain tissue.

2.2.6 Impact of TBI on ICP

TBI can precipitate elevated ICP through mechanisms such as cerebral edema, hemorrhage, and disruption of CSF clearance [43], [47]. The consequential increase in brain

volume or intracranial blood exerts additional pressure against the skull directly elevating ICP. This scenario challenges the capacity of cerebral compliance. When compliance reaches its limit, the rise in ICP can consequently reduce CPP. The elevation in ICP can reduce CPP by increasing the resistance against arterial blood entering the brain, thus reducing the net driving force for blood flow. In response, cerebral autoregulation endeavours to maintain constant CBF despite variations in CPP by adjusting the diameter of cerebral vessels, vasodilation to increase blood flow when CPP is low and vasoconstriction to decrease flow when CPP is high. However elevated ICP may reduce CPP below a critical threshold causing cerebral dysregulation. Under these circumstances, the cerebral vessels cannot dilate sufficiently to counteract the reduced driving pressure, resulting in a sustained decrease in CBF. This persistent reduction in blood flow initiates a cascade of secondary brain injuries, through neuronal injury, herniation, and brain death [48]–[50]. These secondary injuries exacerbate the primary damage caused by the TBI, leading to outcomes such as cognitive deficits, physical impairments, and in severe cases, increased mortality risk [15], [16]

2.2.7 Summary

This chapter delves into the cerebral physiological and hemodynamic components and their roles in regulating ICP and their alteration due to TBI. It explains how TBI-induced disruptions in CBF, CPP, MAP, and cerebral autoregulation can precipitate significant elevations in ICP. Sustained elevations in ICP can instigate secondary brain injuries by compromising blood flow to the brain, thereby reducing the essential supply of oxygen and nutrients. The chapter underscores the need for accurate ICP monitoring for the timely intervention and maintenance of normal ICP levels reducing the risk of secondary injury to the brain.

2.3 Intracranial Pressure Monitoring

TBI can impact cerebral physiology, potentially causing elevations in ICP. Without effective monitoring, these changes in ICP can result in secondary brain injuries. This chapter discusses the gold standard of invasive ICP monitoring, alongside other invasive and non-invasive methods. It evaluates the advantages and limitations of each approach.

2.3.1 ICP monitoring

ICP monitoring serves either as a therapeutic guideline or as a diagnostic tool in various pathological conditions that lead to neurological damage. The guidelines of the American Brain Trauma Foundation (BTF) specify criteria for implementing ICP monitoring in patients with severe TBI who present with a normal CT scan. According to these guidelines, ICP monitoring is indicated under the presence of any two or more of the following conditions at admission: (i) age exceeding 40 years, (ii) the occurrence of unilateral or bilateral motor posturing, or (iii) a systolic blood pressure < 90 mmHg. The guidelines recommend treating ICP above 22 mmHg [51]. Elevated levels of ICP is referred to as intracranial hypertension (ICH).

2.3.2 Invasive monitoring

Intraventricular pressure monitoring

Intraventricular pressure monitoring is the monitoring of ICP by the insertion of a catheter into a ventricle of the brain through a small hole in the skull made by a neurosurgeon, this hole is referred to a as burr hole. Intraventricular pressure monitoring is achieved through the use of an extraventricular drain (EVD) connected to an external strain gauge fluid device which measures ICP the value. Although it is the oldest method, intraventricular pressure monitoring via an EVD remains the gold standard for ICP monitoring techniques [52], [53]

EVD monitoring involves the placement of a catheter into one of the brain's four ventricles. This method provides several benefits beyond mere ICP monitoring. By facilitating the insertion of an EVD, it enables not only the therapeutic drainage of CSF to assist in controlling ICP but also allows for the intrathecal delivery of medications and the removal of intraventricular hemorrhages [54]. The conventional technique for EVD insertion employs the freehand pass method, guided by surface anatomical landmarks [55]. Given the predominance of the left frontal cerebral hemisphere for language functions in over 90% of individuals, the right frontal hemisphere is generally chosen as the entry point [56]. A burr hole is created at Kocher's point, carefully positioned to bypass the superior sagittal sinus and the motor strip of the frontal cortex [57]. Following the application of local anesthesia, a skin incision is made down to the bone, and the periosteum is cleared.

Subsequently, a twist drill creates an opening in the skull along the predetermined path for ventricular access, piercing through the pia and dura mater. The ventricular catheter, is then navigated towards the ipsilateral Foramen of Monro [54]. With the catheter in place and the stylet removed, the initiation of CSF flow through the EVD confirms proper placement, at which point the catheter is connected to an external strain gauge fluid measurement device. Figure 2.4 depicts the placement of a intraventricular ICP monitoring device.

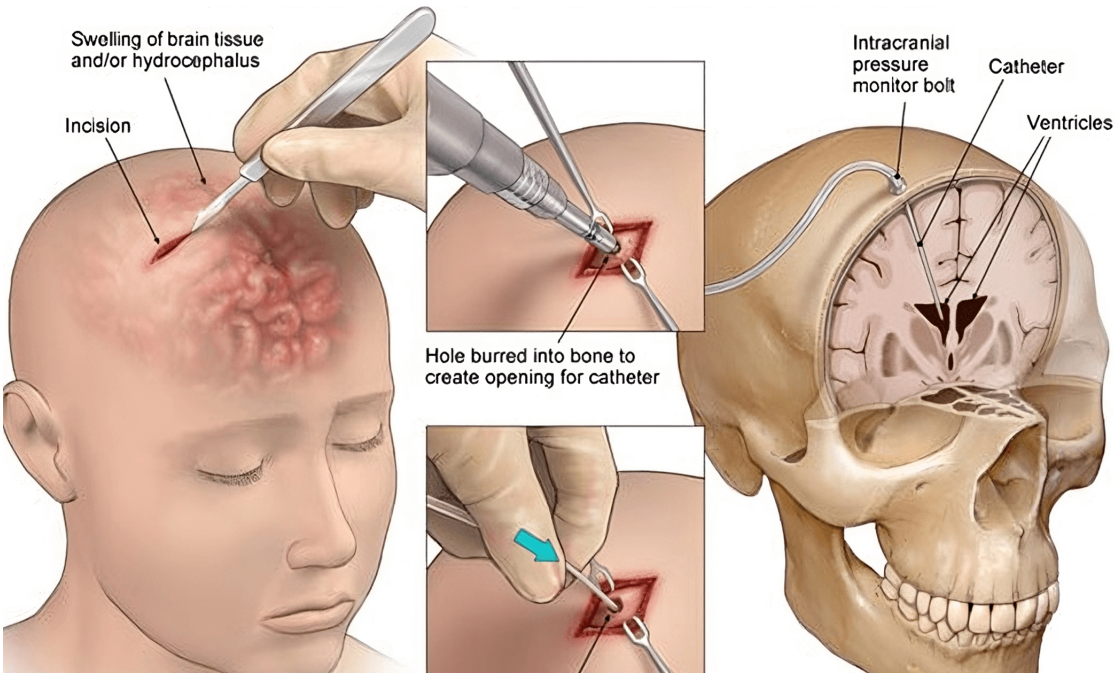

Figure 2.4: An illustration depicting the placement of an intraventricular, intracranial pressure monitoring device [58].

Despite being the gold standard method of monitoring ICP, EVD monitoring is expensive and may lead to a series of complications. The most common complications being: (i) intracranial hemorrhage, (ii) infection and (iii) failure to tap ventricle or misplacement.

Miller et al. observed in their study that hemorrhage occurred in 94 (21.6%) of the patients subjected to imaging following an EVD placement procedure [59]. Saladeno et al. found a misplacement rate of 12.3%, with ventriculostomy-related hemorrhages occurring in 7.1% of patients [60]. Hagel et al. reported an 8.3% cumulative incidence rate of EVD-related infections [61], while Wright et al. documented an infection rate of 23.5% [62]. The estimated diagnostic, procedural, and material cost of EVD placement is in the range of US $1300–$3200 [63], [64].

Intraventricular ICP monitoring is the oldest and gold standard method of monitoring

ICP. Despite this it suffers from a high level of complexity, infection, hemorrhage and cost to patients and healthcare systems.

Transducer-based Pressure Monitoring

Transducer-based ICP monitoring technologies can be categorised into three main types: (i) fiber optic sensors, (ii) strain gauge sensors, and (iii) pneumatic sensors.

Fiber optic-based ICP monitors operate by transmitting light through a fiber optic cable to a pressure-sensitive reflective diaphragm at the sensor's tip. Variations in ICP cause movements of this diaphragm, which in turn alter the intensity of the light reflected back. These changes in light intensity are then converted into ICP readings [65]. A study was conducted between 1992 and 2004, involving 1,000 patients who had a fitted Camino fiberoptic intraparenchymal ICP monitor for an average monitoring time of 184.6 $\pm$ 94.3 hours. The study reported that of the 574 probe tips examined, 8.5% were positive for bacterial growth. Of the 92.2% of patients which underwent a control CT scan, the scans indicated a 2.5% incidence rate of haemorrhage. Technical complications were noted in 4.5% of cases, predominantly associated with the fibre optic cable. Further analysis by Münch et al. based on data from 136 Camino devices, indicated an infection rate of 0.7%, a 5.1% occurrence of intraparenchymal hematoma, and a 23.5% rate of technical complications [66]. Another study by Bekar et al., which reviewed 328 patients with Camino monitors, reported hemorrhage in 1.1% of cases, infections in 4.75%, and technical errors in 3.14% [67].

Strain gauge devices measure ICP through a transducer that alters resistance in response to bending induced by ICP changes. This change in resistance translates to an ICP measurement. Hong et al's investigation into the Codman Microsensor, a subdural transducer-tipped catheter, reported no complications related to infection or hemorrhage [68]. A more extensive prospective study involving 128 patients highlighted three instances of surgically related hemorrhages but, aligning with Hong et al's findings, no infections associated with the sensor's placement [69]. Similarly, an evaluation of the Raumedic Neurovent-P ICP monitor by Citerio et al. found no infections and only two instances of hemorrhage among 99 patients, neither requiring intervention.

The Spiegelberg 3-PN sensor, a pneumatic strain gauge ICP monitor, employs an air pouch at the sensor tip. This pouch maintains a constant pressure equivalent to the

surrounding ICP. Through a transducer linked to the catheter tip, an ICP value is derived [70]. Lang et al's study, involving 87 patients using the Spiegelberg sensor, reported no hemorrhages and identified incorrect measurements in three cases due to sensor leaks. Unlike the previously mentioned fiber optic and strain gauge sensors, the Spiegelberg sensor features the capability for hourly recalibration [71].

Transducer-based devices have shown to monitor ICP with a precision comparable to that of EVD-based monitoring systems [72]. They offer the advantage of not requiring a fluid coupling system, which reduces the risk of infection, as evidenced by the findings of the aforementioned studies. However, unlike EVD-based systems, transducer-based devices, with the exception of the Spiegelberg 3-PN sensor, cannot be recalibrated once placed. They also do not offer the benefits of therapeutic CSF drainage, intrathecal drug administration, or drainage of intraventricular hemorrhage.

2.3.3 Non-invasive monitoring

Transcranial Doppler

Transcranial Doppler (TCD) ultrasonography provides a noninvasive real-time measurement of blood flow characteristics and cerebrovascular haemodynamics within the basal arteries of the brain. TCD is predicated on the principles of the Doppler effect. According to the Doppler effect ultrasound waves emitted from the probe are transmitted through the skull and reflected by moving red blood cells within the intracerebral vessels. The difference in the frequency between the emitted and reflected waves is referred to as the "doppler shift frequency". The Doppler shift frequency is directly proportional to the blood flow velocity [73]. Figure 2.5 depicts the application of TCD for measuring cerebrovascular haemodynamics.

Through the use of spectral analysis, parameters which include the pulsatility index can be extracted from the doppler signal [74].

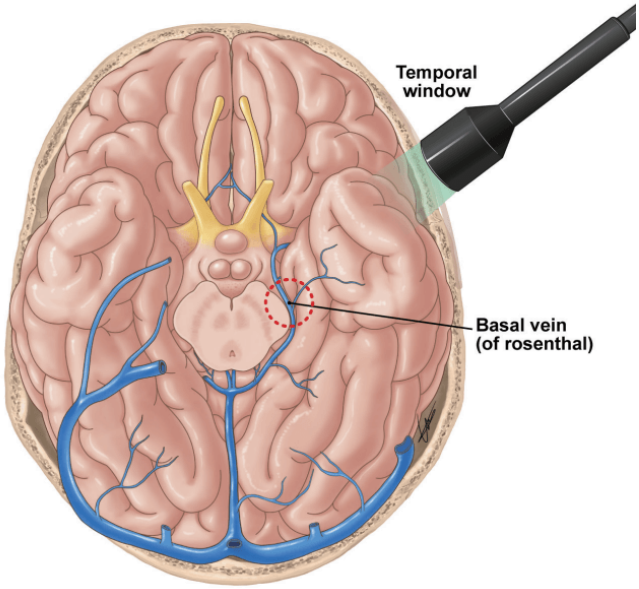

Figure 2.5: An illustration depicting measurement of venous transcranial doppler by using the temporal window to insonate the vein of Rosenthal [75].

The pulsatility index has been found to correlate with invasively measured ICP. Voulgaris et al. who investigated 37 patients with severe TBI (GCS $>$ 8 on admission) found a strong correlation between ICP and PI (r=0.82, p $>$ 0.0001) for ICP values $>$ 20 mmHg. However, Voulgaris et al. report that at ICP values below 20 mmHg there was no significant correlation between ICP and PI (p $>$ 0.05) [76]. Similarly to Voulgaris et al's study, Bellner et al. also reported a significant correlation (p $>$ 0.0001) between the ICP and the PI with a correlation coefficient of 0.938. In the ICP interval between 5 to 40 mmHg the correlation between ICP and PI enabled an estimation of ICP from the PI values with an SD of 2.5 ($\pm$4.2 mmHg). This small deviation was only applicable to ICP values lower than approximately 30 mmHg. As ICP values increased so did the deviation. Consequently accurate ICP measurements at elevated ICP values using TCD were not possible [77].

TCD based ICP measurement as shown in the above studies is an unreliable technique of ICP measurement across the pressure range. In addition to this TCD is prone to intra and inter-observer variability [78]–[80]. Shen et al's study concludes that without regular practice, measurement accuracy and observer agreement is negatively impacted [81]. Additionally TCD is non efficacious in 10-15% of patients due to the ultrasound waves not being able to penetrate through the skull [82].

Optic nerve sheath diameter

The optic nerve, is a component of the central nervous system. It traverses from the intracranial space into the orbit, surrounded by subarachnoid space, specifically the dural sheath. This sheath surrounding the optic nerve, contains CSF, linking the subarachnoid space around the nerve closely with the intracranial subarachnoid space [77]. Elevated ICP can increase pressure around the optic nerve, influencing the optic nerve sheath diameter (ONSD) and leading to the development of papilledema. Dilation of the ONSD has been shown to be an early indication of a rise in ICP [83], [84].

Several studies have established a correlation between invasive ICP measurements and ultrasonographically gathered ONSD data [83]–[86]. Dubourg et al's systematic review and meta-analysis, which examined ultrasonography of the ONSD in 231 patients for detecting raised ICP, reported a pooled sensitivity of 0.90 and a specificity of 0.85 [87]. Similar to the TCD method of ICP assessment, ONSD measurement is susceptible to intra-observer variability. However, this variability is reportedly lower than that associated with TCD measurements [88]–[90]. Nonetheless, various conditions, including tumors and Graves disease, can influence the ONSD, potentially compromising the efficacy of the ONSD method in monitoring ICP [91].

The ONSD is a non-continuous monitoring technique that although able to discern between normal and increased ICP is not accurate enough to be used in place of an invasive ICP monitoring method. Additionally, similarly to the TCD method, there is reported inter-observer variance and similarly to the TCD monitoring technique requires regular operator training in order to yield accurate results.

Optical & PPG ICP estimation

A growing body of research is exploring methods that leverage cardiac waveforms, such as diffuse correlation spectroscopy (DCS) and near-infrared spectroscopy (NIRS) data for the non-invasive estimation of ICP.

An initial pilot study in 1997 investigated the use of NIRS to asses variations in regional cerebral oxygen saturation (rSO2) in 8 patients with head injury. Half of the patients assessed had an ICP $>$ 25mmHg, the other half had an ICP $<$ 25mmHg. The study's findings indicated that NIRS values in the group of patients with elevated ICP were significantly lower than in the low ICP cohort. Furthermore, the rSO2 values where

significantly lower in patients with elevated ICP compared to patients within the lower ICP group and did not display a significant increase after hyperoxygentation. These results indicated the possible utility of NIRS as a modality of assessing elevated ICP [92]

Further investigations by Ruesch et al. corroborated the findings of the initial pilot study of 1997. Ruesch et al. induced variations in ICP via fluid injections using an intraventricular catheter, mimicking a form of hydrocephalus within non-human primates. The ICP was monitored using an intraparenchymal ICP sensor, and hemoglobin concentration changes were measured non-invasively using a NIRS optical sensor. The ICP values were varied from 3 to 40 mmHg covering a significant range of ICP values from what is considered a normal range to severely elevated values. The findings of this research demonstrated a good correlation between hemoglobin levels collected using NIRS and induced ICP, further suggesting the capability of NIRS for non-invasive ICP monitoring [93].

In subsequent research, Ruesch et al. utilised PPG features from cardiac DCS pulsations along with MAP to predict ICP non-invasively within five non-human primates. Similarly to their previous study an intraventricular catheter was used to induce changes in ICP through the introduction of a saline solution. Elevated ICP levels were induced reaching approximately 30 mmHg, each induced ICP level was maintained for approximately 90 minutes. From the cardiac pulsatile waveforms a number of morphological and time series features which included peak height, prominence and half peak width were extracted amongst others. These extracted features along with MAP were then used to train a Random Forest (RF) regression model. The training and testing data was randomly split in a 80:20 ratio. The predictive model yielded results achieving a Coefficient of Determination (R^2) of 0.92 and a root mean squared error (RMSE) of 3.3 mmHg [94].

This group expanded their work by investigating features extracted from near-infrared PPG (NIR-PPG) in combination with MAP for ICP estimation. Inline with their previous studies ICP was induced within eight non-human primates using a saline solution via a intraventricular catheter. A NIRS system operating at 690 and 830 nm, was used to measure cerebral hemoglobin concentration changes. Arterial blood pressure (ABP) was recorded using an arterial line placed in the carotid artery. Induced ICP ranged between 5 and 60 mmHg. Predominantly morphological features such as peak height, width and prominence amongst others were extracted. These features in combination with MAP were used as features for their ICP estimation model. A RF model was trained and

evaluated using 5 fold cross validation with a randomly sampled training and testing dataset in a ratio of 80:20. The model resulted in a R^2 to 0.937 and an RMSE of 2.703 mmHg [95].

Further studies have also shown the potential of using PPG waveform features to correlate with ICP levels. A study simultaneously collected ICP and brain PPG signals collected at 660 nm over each hemisphere of the brain from 12 patients. Invasive ICP values were recorded using an extra-ventricular drain. From the PPG waveform, a number of morphological features were extracted. A unspecified algorithm was developed using an underscribed training and testing approach with findings indicating a R^2 of 0.66 [96].

2.3.4 Summary

Invasive ICP monitoring techniques remain the most accurate method of monitoring ICP. The EVD method of monitoring ICP remains the gold standard. Invasive ICP monitoring techniques are expensive, require a high level of expertise to implant and are prone to infection. There are no existing continuous, non-invasive ICP monitoring techniques that are accurate across the pressure range in relation to invasive monitors.

Recent research efforts have focused on leveraging morphological and time-series features derived from cardiac waveforms to estimate ICP non-invasively. While these studies have yielded some encouraging results, their scope has predominantly been confined to animal models, incorporating MAP as an additional variable. These studies utilised random sampling cross-validation methods for training and evaluation. The relationship between MAP and ICP is highly sensitive. MAP was found to be the strongest predictive feature across all the studies which used it. The inclusion of MAP in combination with the training and evaluation approach we suggest beneficially biases the results of these studies, although the limited number of non-human primates within these studies is appreciated. To the knowledge of the author one study aims to estimate ICP using cardiac waveforms collected from patients however this study neither describes the model used nor the training and testing approach.

This research, using the largest clinically collected ICP labelled dataset to date aims to robustly investigate the relationship between NIR-PPG derived features and ICP and the efficacy of predicting ICP using said features exclusively.

2.4 Estimating Intracranial Pressure using Photoplethysmography

2.4.1 Introduction to PPG

PPG signals are non-invasive, cost-effective optical signals, easily obtained from the skin's surface using a light source and photodetector. PPG signals capture information about blood volume fluctuations by detecting variations in light absorption or reflection, induced by changes in blood volume. The growing interest in non-invasive physiological monitoring, coupled with the paradigm shift in healthcare towards continuous and ubiquitous patient monitoring rather than exclusive in-hospital care has triggered a substantial surge in the adoption and exploration of PPG technology [97]–[99]. The magnitude of this interest is underscored by the significant rise in search results related to PPG over the past two decades which has seen an increase of 2392.72%, as evidenced by the search results of the terms "Photoplethysmography" or "Photoplethysmogram" in PubMed, Europe PubMed Central and Scopus from 2000 to 2022, as illustrated in Figure 2.6.

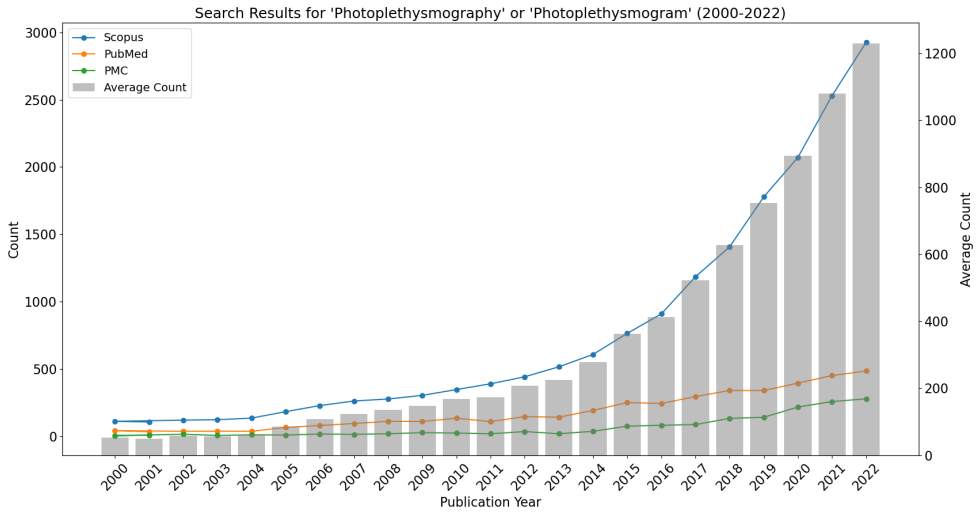

Figure 2.6: Bar chart depicting the annual number of articles indexed in PubMed from 2000 to 2022, using the keywords Photoplethysmography or Photoplethysmogram.

This same trend has led to the proliferation of non-invasive medical devices as well as consumer-oriented wearable devices such as smartwatches and fitness trackers. Many of these devices use PPG technology to enable non-invasive and continuous monitoring of various physiological markers, including heart rate, blood pressure, oxygen saturation, and sleep patterns [3].

2.4.2 Basic working principle of PPG

PPG is a non-invasive optical technique that detects blood volume changes with each cardiac cycle based upon the absoptivity of light. PPG sensors are comprised of two main components: (i) photodiode and (ii) photodetector. In order to obtain a PPG signal the tissue is irradiated by light emitted by the photodiode of which a certain amount is absorbed by the tissue, the unabsorbed light being collected by the photodetector.

A cardiac cycle consists of two stages: (i) systole and (ii) diastole, representing the periodic contraction and relaxation of the heart, respectively. This cycle is essential for the effective pumping of blood throughout the body, ensuring the delivery of oxygen and nutrients to tissues while facilitating the removal of carbon dioxide and other metabolic wastes. Systole is the phase of the cardiac cycle during which the heart muscle contracts. This contraction results in the ejection of blood from the heart chambers and increase in blood volume throughout the body, including all the peripheral tissue sites. Diastole is the phase during which the heart muscle relaxes after contraction. During this relaxation phase blood volume decreases and the heart chambers fill with blood, preparing for the next cycle of contraction.

The pulsatility of the PPG signal emerges as a consequence of the blood volume changes between systolic and diastolic phases of the cardiac cycle. The systolic increase in blood volume results in increased absorbance of light in tissue compared to the diastolic state changing the amount of unabsorbed light being recorded by the photodetector. This variation in light absorption and the corresponding detection by the photodetector gives rise to the PPG pulsatile waveform which is synchronous with each heartbeat [99], [100].

Transmittance PPG sensors emit light through tissues such as a fingertip and measure the amount of light that emerges on the opposite side. Conversely, reflectance PPG sensors, emit light that reflects off the surface of thicker tissues such as the wrist or forehead.

PPG waveform

The PPG waveform is composed of two elements: (i) the alternating current (AC) and (ii) the direct current (DC) components. The DC component arises from the light absorption by non-pulsatile tissues, such as muscles and bones. This component changes gradually over time and is subject to variations influenced by physiological factors such as respiratory

patterns and activity within the sympathetic nervous system.

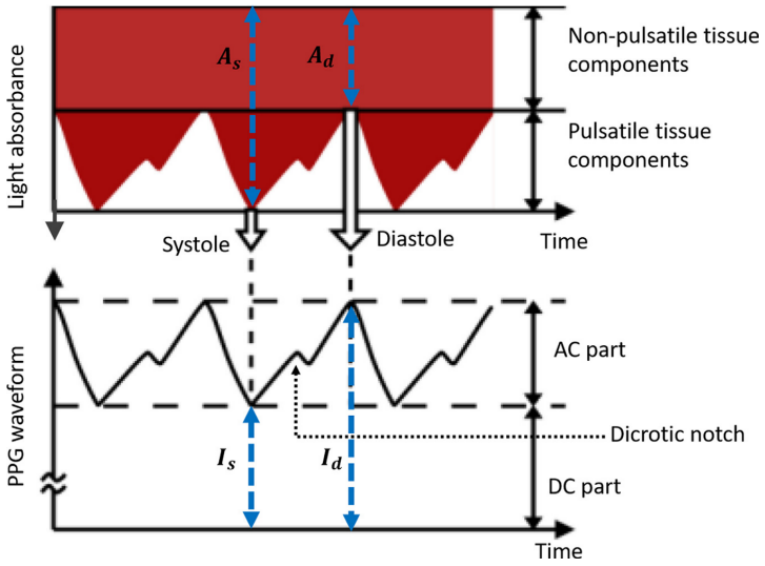

Figure 2.7: The schematic of the PPG waveform produced due to the absorbance of light in tissue, as a function of time. Light absorbance in the pulsatile tissue components produces the pulsatile AC part in the PPG waveform. Absorbance in the nonpulsatile tissue components contributes to the slowly varying DC part of the PPG waveform [99].

The pulsatile AC component which is superposed on the DC component is created by the aforementioned variations in light absorption due to changes in blood volume between the systolic and diastolic phases of the cardiac cycle. The repeated systole and diastole of the heart result in a periodic PPG signal which for every cardiac cycle includes one systolic and one diastolic peak. Figure 2.7 depicts the inverse relationship between the intensity of the recorded PPG (I) and light absorbance (A) [99]. During systole and the subsequent increased blood volume, more light is being absorbed and less light is reaching the photodetector resulting in a lower light intensity represented as a higher pulse amplitude on the PPG waveform. Conversely, during diastole, the blood volume reduction leads to less light absorption and more light reaching the photodetector, indicating higher light intensity but corresponding to a lower pulse amplitude on the waveform. Thus, variations in light intensity detected by the PPG sensor inversely affect the pulse amplitude observed in the PPG signal.

2.4.3 Non-invasive intracranial pressure sensor

Monte-Carlo light-tissue interaction modelling

Roldan et al. performed a Monte-Carlo simulation of the interaction between near-infrared (NIR) light and a multi layered tissue model of an adult human head. The simulation aimed to investigate the relationship between emitter and photodetector separations and subsequent variations in light absorption and penetration of the tissue layers. The study investigated emitter-photodetector separations between 1 to 5 cm.

At all simulated separations the maximum and minimum absorbance were in the skull and the subarachnoidal space, respectively. Additionally, as separation increased the relative absorption of light decreased within the extracerebral layers which include the scalp and skull. At the deeper layers of the tissue model which include the grey and white matter, the inverse was apparent, the larger the separation the greater the resulting absorption of light. The findings of this study suggest that a larger emitter-photodetector separation especially above 3 cm facilitates a deeper interrogation of the head tissues allowing for the collection of data from the white matter of the brain [101].

Description of non-invasive intracranial pressure sensor

This research is based upon data produced by an in-house, NIRS, reflectance, PPG sensor for non-invasive estimation of ICP hereinafter referred to as the "nICP" sensor [102]. The nICP sensor consists of four LEDs at four different wavelengths (770, 810, 855 and 880 nm) and two photodiodes "*proximal*" and "*distal*" positioned at 10 mm and 35 mm from the light source representing short-distance and long-distance data respectively. The sensor is arranged as shown in Figure 2.8. The research focuses upon the 810 nm wavelength data as it represents the isosbestic point. At the isosbestic point the absorption properties of the oxy- and deoxyhemoglobin are the same allowing for an optical signal independent of blood oxygentation, eliminating it as a confounding factor [103].

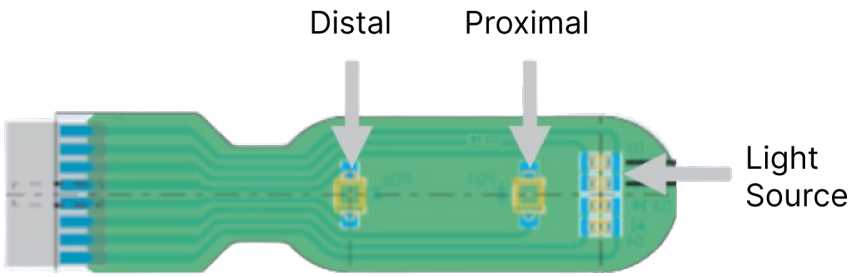

Figure 2.8: A diagram showcasing the design of the non-invasive intracranial pressure sensor, highlighting the placement of the photodiodes and the light source, as well as depicting the spatial arrangement and distances between the proximal and distal photodiodes relative to the light source.

Findings from the montecarlo simulation of the light-tissue interaction propose that the data from the proximal photodiode corresponds to extracerebral data, while the data from the distal photodiode represents a combination of extracerebral and cerebral data [101].

Figure 2.9 shows the indicative placement of the non-invasive nICP sensor (indicated by a green arrow) and the invasive sensor (indicated by a red arrow) on a patient. The nICP sensor is positioned on the patient's forehead, below the hairline.

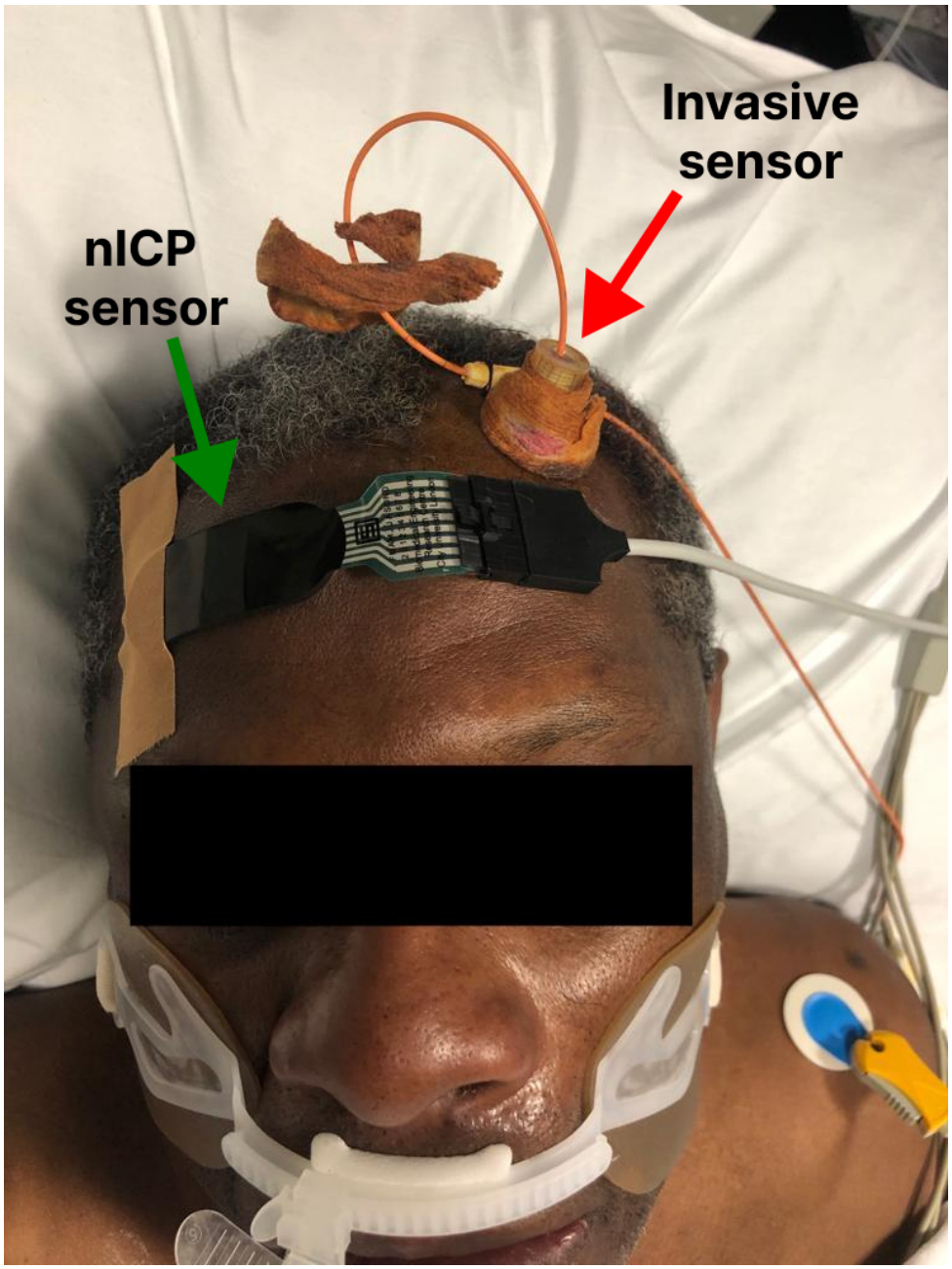

Figure 2.9: An image depicting the indicative placement of the non-invasive nICP sensor (indicated by a green arrow) and the invasive sensor (indicated by a red arrow) on a patient.

2.4.4 Theory: Relationship between photoplethysmography and ICP

Building upon the description of cerebral physiology and haemodynamics described in Chapter 3, ICP is the pressure within the cranial vault, governed by the Monro-Kellie hypothesis, which suggests that the cranial compartment's volume is fixed, comprising brain tissue, cerebral blood, and CSF [43]. Variations in ICP arise from changes in

these components due to pathological conditions. Such changes can trigger compensatory mechanisms aimed at preserving normal ICP levels.

Elevation in ICP can reduce CPP by increasing the resistance against arterial blood entering the brain. This elevation in ICP and consequential reduction in CPP triggers cerebral autoregulation, which adjusts the diameter of cerebral vessels. However, elevated ICP may reduce CPP below a critical threshold causing cerebral dysregulation where blood vessels cannot dilate sufficiently to counter the reduced driving pressure, leading to a sustained drop in CBF.

The PPG morphology mirrors blood volume changes within each cardiac cycle based upon the relative change in light absorbance between the systolic and diastolic phase. Given the aforementioned influence of variations in ICP on CPP and its subsequent impact on cerebral autoregulation mechanisms. It seems reasonable to suggest that increases in ICP and the suspected change in flow dynamics and or decrease in cerebral blood volume stemming from cerebral dysregulation, diminished CPP, and CBF alongside resultant variations in light absorbance, will manifest in identifiable changes to PPG waveform morphology.

2.4.5 Summary

PPG data reflects the changes in blood volume within each cardiac cycle. The presented theory linking PPG and ICP describes how variations in ICP levels can affect cerebral blood volume and general flow dynamics in the brain. The Monte-Carlo simulation indicated that NIR light with certain photodiode-photodetector separations allows for the collection of back scattered light from brain matter. Based upon these findings and design of the nICP sensor it is suggested that the near infrared PPG data collected from the distal photodiode of the sensor will provide information reflecting haemodynamic changes in the brain which can be associated with ICP.

Chapter 3

Dataset and Pre-processing

This chapter introduces the dataset of which this research is based upon, it explains the data collection protocol, recruitment of patients and the loading and structuring of the data. In addition to this the chapter describes the data preprocessing including the novel denoising algorithm which was designed and applied to the dataset. The denoising algorithm detailed in this chapter resulted in a publication that describes and evaluates its performance [104].

3.1 Dataset

3.1.1 Data collection

The non-randomised data collection was performed over a 78-week period between January 2020 and July 2021 (ClinicalTrials No. NCT05632302). Each patient had an implanted, invasive, intra-parenchymal (Raumedic® Neurovent-P) pressure probe which was interfaced with a GE® monitor as part of their normal medical treatment. The nICP sensor was affixed to the patient's forehead below the hairline. Both the invasive and non-invasive data was collected with a sampling frequency of 100 Hz. The non-invasive and invasive data were collected synchronously. The nICP monitor was calibrated for each patient through the adjustment of the LED intensity and amplification gain according to patients' characteristics and ambient light. Calibration was performed before recording started. If the patient left the intensive therapy unit (ITU) for a scan or surgery, the nICP monitor probe was disconnected and left in situ. We hereinafter refer to the data collected from the nICP sensor as "NIR-PPG" and the data collected from the invasive

ICP monitor as "ICP data".

3.1.2 Participants

The dataset consists of data from 40 patients with a severe TBI diagnosis, recruited from the ITU of the Royal London Hospital in the United Kingdom. Except for one individual, all participants were in an unconscious state due to head injuries. The study also included a conscious patient who was undergoing ICP monitoring as part of an assessment for normal pressure hydrocephalus. Candidates were not considered for the study if they were not expected to survive beyond 48 hours or if a personal consultee recommended against their involvement. Individuals who had received a decompressive craniectomy were not included due to the procedure's adverse affects on signal quality. The average age of the patients was 43.92 years, and the gender distribution was predominantly male, with a ratio of 14:2.

Conscious participants signed written consent forms before joining the study. For those unable to give consent themselves due to incapacitation, permission was obtained from either a personal consultee or an independent healthcare professional when no consultee was available, in alignment with the UK Mental Capacity Act of 2005.

Participants' personal information was protected and handled according to the Data Protection Act, NHS Caldecott Principles, The Research Governance Framework for Health and Social Care, and the conditions of the Research Ethics Committee Approval.

3.1.3 Data loading and structure

The raw ICP and NIR-PPG data were loaded, resulting in three structured datasets for the ICP measurements and NIR-PPG signals from both the proximal and distal photodetectors, respectively termed as "proximal data" and "distal data". Each dataset was structured with columns uniquely representing each patient's data.

During the data collection phase, two inconsistencies emerged: (i) the lengths of the NIR-PPG data differed from those of the ICP data, due to variations in the recording durations and (ii) some patients had multiple data files, each indicating the start of a new recording session. This situation arose primarily due to patients leaving ITU for required interventions or diagnostic scans.

To address these challenges, an assumption was adopted that the NIR-PPG and ICP

recordings for each patient commenced simultaneously. This assumption guided the initial data processing steps, where the raw data files for each patient were first synchronised based on their lengths. The process involved trimming the longer of the two datasets (NIR-PPG or ICP) to match the length of the shorter dataset, ensuring alignment.

Ensuring the datasets are trimmed for alignment is essential for the subsequent stages of data analysis, especially when labeling the datasets. Alignment guarantees mutual length and temporal correspondence between both the NIR-PPG and ICP signals is maintained. This alignment is vital for subsequent, accurate feature extraction and labeling, as it ensures that the NIR-PPG data used to generate features corresponds accurately to the same point in time in the ICP signal which is used to derive each label. Misalignment could lead to incorrect associations between the features and their corresponding labels, thereby compromising the integrity of the analysis.

For patients with multiple data files, each file was treated independently, undergoing the length adjustment process to ensure consistency across all sessions. Subsequently, the adjusted datasets from different files were concatenated, creating a unified dataset that represented the entirety of a patient's recorded data. This processing ensured that the final data had a uniform length across datasets. Figures 3.1 and 3.2 demonstrate two examples of the raw patient data.

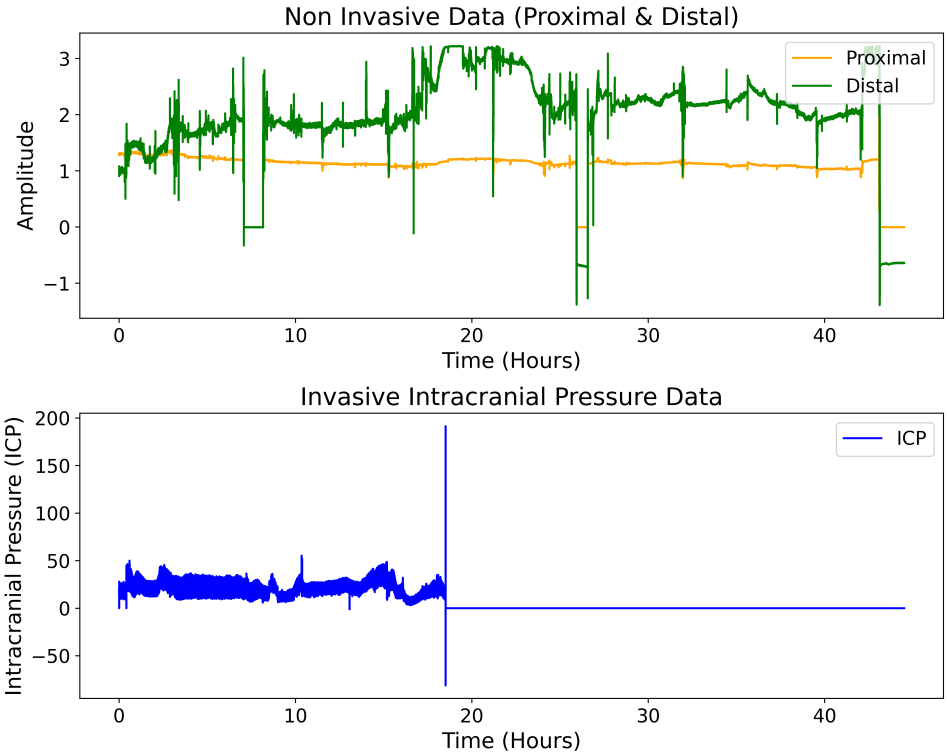

Figure 3.1: A figure consisting of two subplots depicting the raw data of a individual patient. The top subplot depicts the loaded raw non invasive data collected from both the proximal and distal photodiodes. The bottom subplot depicts the invasive intracranial pressure data.

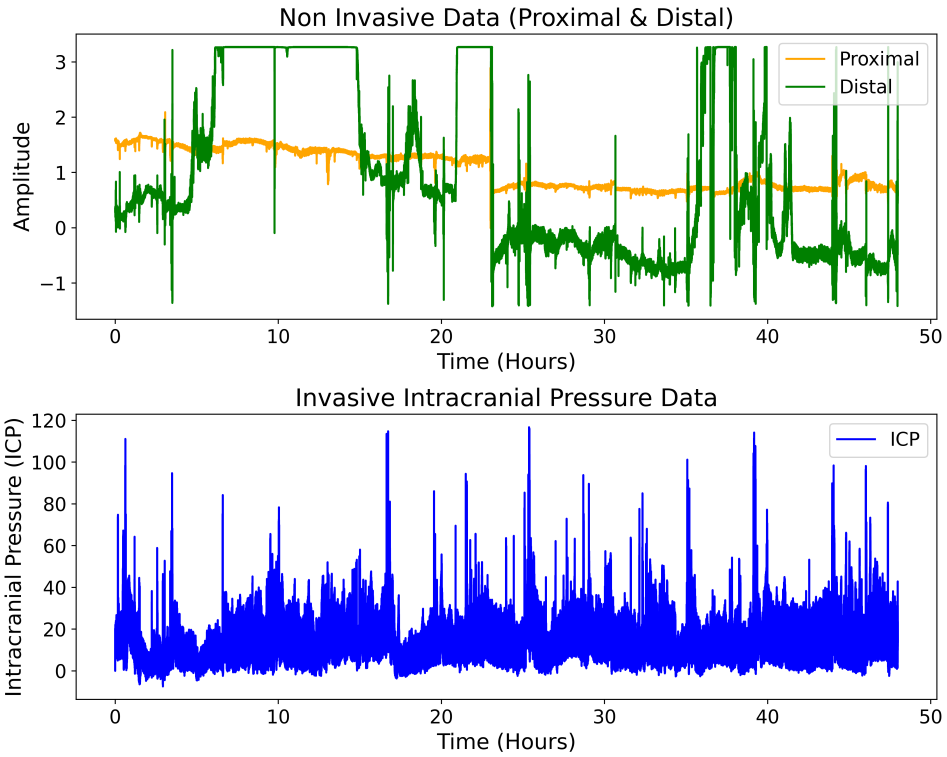

Figure 3.2: A figure consisting of two subplots depicting the raw data of a individual patient. The top subplot depicts the loaded raw non invasive data collected from both the proximal and distal photodiodes. The bottom subplot depicts the invasive intracranial pressure data.

3.1.4 Handling of erroneous intracranial pressure values

Prior to the denoising of the dataset, an additional step was taken in order to handle erroneous ICP values. The assumption was made that values of ICP $\geq$ 60 mmHg or $<$ 0 mmHg, could be erroneous or biologically implausible. Consequently, these instances were removed from both the ICP and NIR-PPG data. Values above 60 mmHg may indicate potential measurement errors, while negative values are physiologically implausible. Figure 3.3 depicts an example of recorded ICP with erroneous data points both above 60 mmHg and below 0 mmHg. In cases which Figure 3.4 illustrates were there are sustained ICP values outside of the biologically feasible bounds it seemed reasonable assume there was a invasive measurement error and subsequently the data of these patient's was excluded.

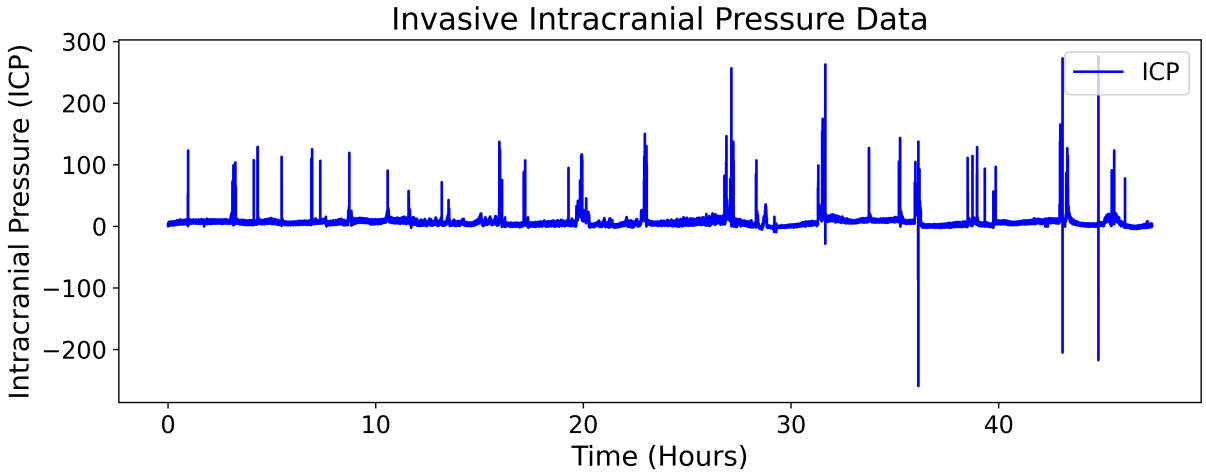

Figure 3.3: A figure depicting the invasive intracranial pressure data recorded from a patient which includes assumed erroneous values above 60 mmHg and below 0 mmHg.

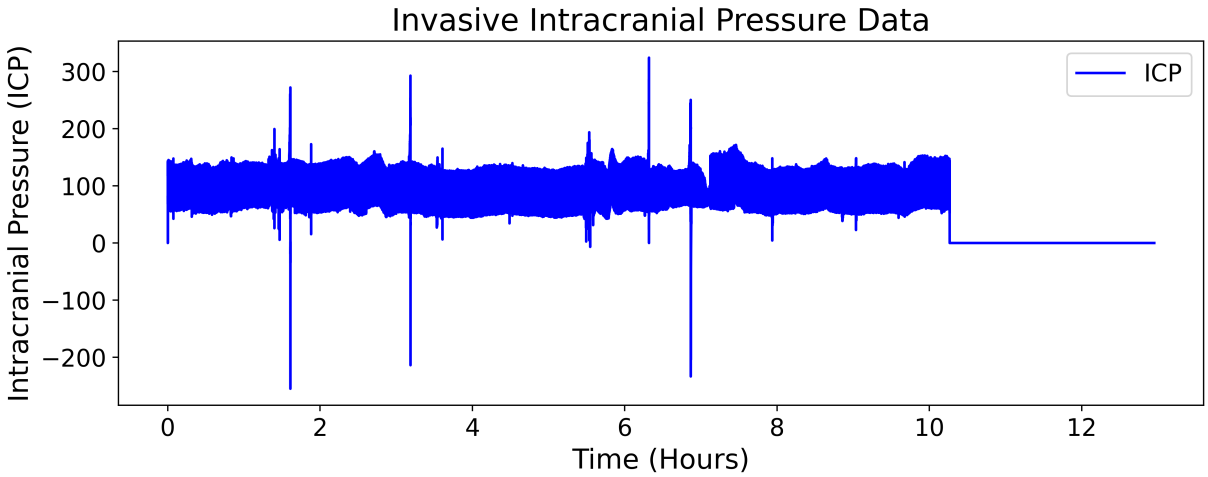

Figure 3.4: A figure depicting the invasive intracranial pressure data recorded from a patient which includes assumed erroneous elevated values with a mean ¿ 100 mmHg.

3.1.5 Data exclusion

For each patient, ICP data was saved in a ASCII (.asc) file format. The LabVIEW software was used for NIR-PPG data acquisition. The NIR-PPG data was stored in a LabVIEW (.lvm) file for each patient. To facilitate data analysis, both .asc and .lvm files were converted to Comma-Separated Values (.csv) format.

Initially, the dataset consisted of 40 patients. During the conversion process, the ICP data of 4 patients were identified as missing, and the NIR-PPG data for 2 patients was absent after conversion. Consequently, the data of these 6 patients were excluded from further analysis. Subsequent analysis of the ICP data identified that the ICP data of an additional 4 patients was erroneous. Characterised by an average ICP value exceeding 60 mmHg. These patients were also excluded.

Furthermore, visual inspection of the NIR-PPG data revealed that 4 patients exhibited unreliable pulsatility. This was characterised by data where the majority did not display a clearly identifiable pulsatile component. Since this research relies on features extracted from the pulsatile aspect of the NIR-PPG signal, these patients were excluded to ensure the reliability of the analysis. Figure 3.5 provides an illustrative comparison between a reliable and unreliable pulsatile NIR-PPG signal. The unprocessed data from the 27 patients used in this research totaled 1190.73 hours of NIR-PPG recordings, with an average of 44.10 hours per patient.

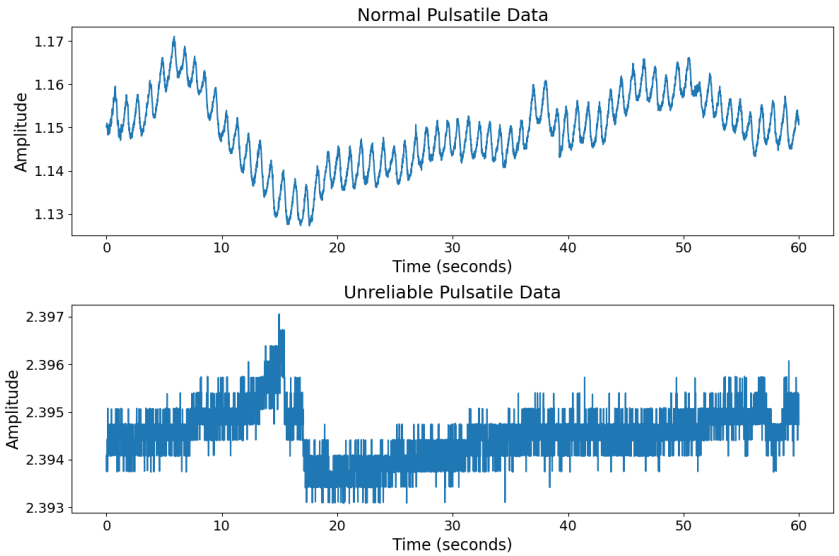

Figure 3.5: Comparison of pulsatile data: A figure containing two subplots. The subplot above demonstrates normal pulsatile data, characterised by consistent PPG waveform patterns. In contrast, the subplot below shows a segment of data from a patient where pulsatile signals were deemed unreliable for further analysis.

3.2 Signal pre-processing

3.2.1 *Envelope-Based PPG Denoising Algorithm*

Introduction

PPG signals are inherently vulnerable to various forms of noise that can distort the signal, posing challenges in obtaining reliable and accurate information. Among the sources of noise are power line interference, low and high-frequency noise, baseline drift, motion artifacts, and saturation of the photodiode by light [105]. The prevalence of such noise underscores the importance of data denoising approaches for both retrospective and real-time signal processing and analysis. Addressing these challenges, a novel envelope-based denoising algorithm, hereinafter referred to as the "Envelope PPG Denoising Algorithm" (EPDA), was developed, designed to accomplish two objectives: (i) the identification and removal of anomalous data while preserving the indices of the removed data, and (ii) the assurance of computational efficiency, recognising the practical constraints of real-time processing and resource utilisation. The EPDA was applied to the NIR-PPG dataset, to identify and remove anomalous regions in the non-invasive data, the same anomalous regions were also removed from the ICP data.

Existing PPG denoising algorithms

We distinguish between two main approaches to denoising: (i) anomaly detection denoising and (ii) filter-based denoising. Anomaly detection denoising refers to methods that identify and eliminate unwanted anomalous data points or indexes within a dataset. On the other hand, filter-based denoising involves modifying the frequency content of a signal by selectively allowing or attenuating specific frequency components.

Within these definitions, the EPDA is classified as an anomaly detection denoising approach. Anomaly detection-based denoising proves particularly valuable when dealing with continuously labelled datasets, where maintaining synchronisation between the data and the label source is crucial. This approach involves storing the indexes of anomalous data, facilitating the removal of anomalous data points from both the data being denoised and the corresponding indexes from the label source, thus preserving data-label integrity.

There is a dearth in research focusing on non-filter based PPG denoising. Recently there has been a rise in the use of machine learning (ML) techniques for denoising. Lee

et al. introduced a Bidirectional Recurrent Auto-Encoder (BRDAE) for PPG denoising, showcasing a 7.9 dB improvement in signal-to-noise ratio for noise-augmented data during validation [106]. Kwon et al. proposed a transformer-based deep generative model to eliminate noise within PPG signals [107]. Mohagheghian et al. developed a convolutional autoencoder approach for noise reduction, reducing the average detected heart rate (HR) and RMSE by 45.74% and 23%, respectively, on arterial fibrillation and non-arterial fibrillation data [108]. Xu et al. created a motion artifact removal Time-Delay Neural Network (TDNN) that uses the signal envelope to normalise PPG signal amplitudes while preserving other waveform features such as dominant frequency and pulse width [109]. There have also been alternative statistical approaches to denoising. Lin et al. introduced a denoising method which characterises the signal through the calculation of features using fiducial points (onset, peak, and offset points) and employs adaptive thresholds to classify and remove anomalous segments of the signal [110]. Dao et al. presented a motion artifact detection algorithm that utilises time-frequency spectral (TFS) features [111]. These TFS features help distinguish between motion artifact-corrupted segments of data and clean data segments. Substantial distinctions arise when comparing the EPDA to other denoising algorithms. The EPDA, as a statistical approach, offers distinct advantages over ML methods, primarily in terms of enhanced transparency and interpretability, aspects inherently embedded within statistical methods but often lacking in ML approaches. ML techniques for denoising often require a large amount of labelled data for effective training, a resource that may not be readily accessible or easily generated. Furthermore, certain denoising approaches attempt to reconstruct noise-contaminated PPG data, potentially resulting in the loss of important features present in the uncorrupted PPG signal. Some algorithms within this domain exclusively target the capture of HR and oxygen saturation (SpO2) information, sidelining the importance of preserving the signal's morphology and amplitudes, which may be of significance for other physiological markers [112], [113].

While certain research utilises the signal envelope this usage typically occurs as a step not central to the denoising itself. For instance, it is has been employed as a preprocessing step for data normalisation or for identifying fiducial points within noisy data [109], [114]. To the best of the authors' knowledge, there is no existing denoising approach that leverages the envelope difference (calculated as the absolute difference between the upper and lower envelope of the signal), as the foundational basis for denoising. Another attribute of the EPDA, not prevalent in many denoising approaches is the preservation of the in-

dices of identified anomalous data. This preservation becomes particularly valuable when dealing with synchronised labelled data. Existing statistical approaches to denoising tend to incur large computational costs through the calculation of spectral features, transformation of the data or the calculation of morphological features for each pulse cycle. It is hypothesised that the EPDA through the efficient calculation of the envelope difference and use of the envelope difference to characterise the signal incurs a lower computational expense with respect to other denoising approaches which becomes important when considering the application of algorithms to large datasets retrospectively or for real time use.

It therefore seems reasonable to suggest that the EPDA helps to address a research gap in the development of easily implementable, computationally efficient, and interpretable denoising approaches which are data efficient and which can preserve the indexes of identified anomalous data for use in synchronised, labelled datasets.

Manifestation of noise

Within the non-invasive, NIR-PPG data, two main sources of noise are hypothesised to be present: (i) motion artifacts, which are identified by irregular signal morphology or high amplitude variance, and (ii) photodetector saturation, characterised by areas with little or no amplitude variance referred to as "flat lines". Figure 3.6 illustrates the presence of suspected motion artifacts and photodetector saturation within the dataset.

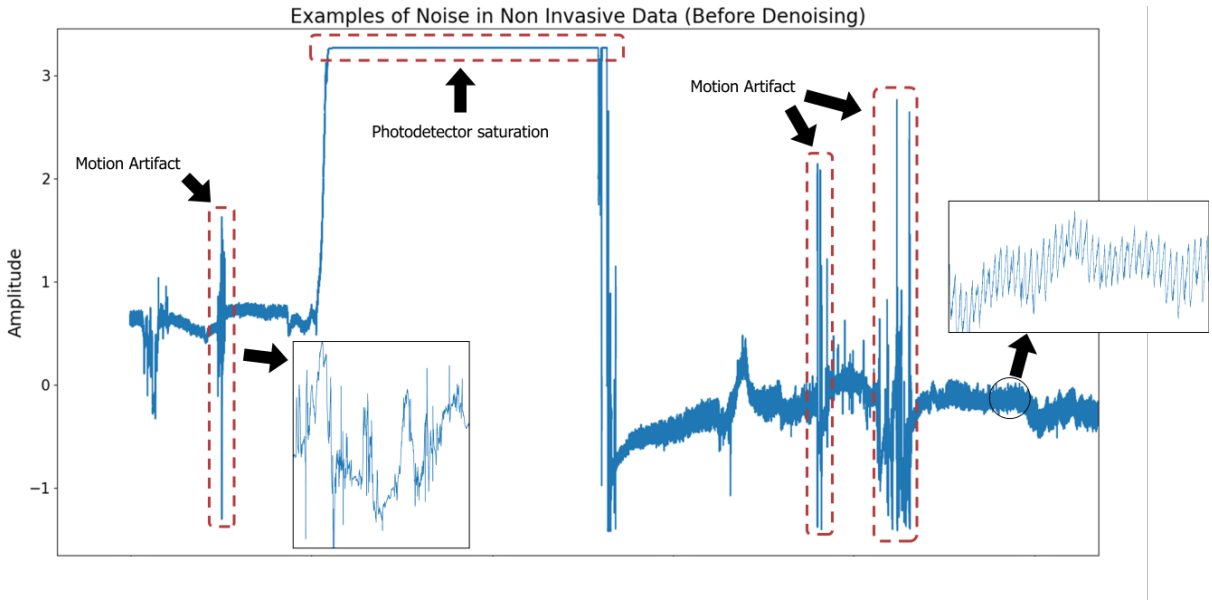

Figure 3.6: PPG signal data with highlighted noise attributed to motion artifact and photodetector saturation. Visual aids emphasise the presence of noise within the window.

Envelope PPG denoising algorithm

The EPDA operates through a 6-step process: (i) the filtering and calculation of the upper and lower envelopes of the signal, (ii) the calculation of the *envelope difference*, (iii) the calculation of thresholds, (iv) detection of anomalous indexes, (v) the calculation of segmentation points, and (vi) the removal of identified anomalous data. In the initial step, a fourth-order bandpass Butterworth filter with cutoff frequencies set at 0.5 and 12 is applied to the data removing undesired frequency components from the signal. Once the data has been filtered, the upper and lower envelopes of the signal are calculated. To calculate the envelopes, the peaks and troughs of the signal are detected using Scipy's "find peaks" function [115]. The "distance" input parameter of the "find peaks" function is set to 20% of the total length of the data window represented in seconds. This distance definition optimises computational efficiency by reducing the number of detected peaks and troughs. Consequently, fewer interpolation points are needed when computing the envelopes whilst preserving the ability to effectively characterise the anomalous regions of the signal. Once the peaks and troughs are identified, linear interpolation is used to calculate the upper and lower envelopes of the signal. Linear interpolation is preferred over other methods, such as spline interpolation, due to its computational efficiency. It estimates values by creating a straight line between two adjacent data points, capturing the relationships between points, and efficaciously representing anomalous segments of the data. The *envelope difference* is obtained by taking the absolute difference between the upper and lower envelopes. The *envelope difference* provides a simplified representation of the signal while preserving and emphasising significant changes in amplitude aiding in the characterisation of noise. Figure 3.7 visually demonstrates the detected peaks and troughs, as well as the upper and lower envelopes, and the envelope difference. To demonstrate the step wise progression of the EPDA, the same window of data is used from Figure 3.7 to Figure 3.10.

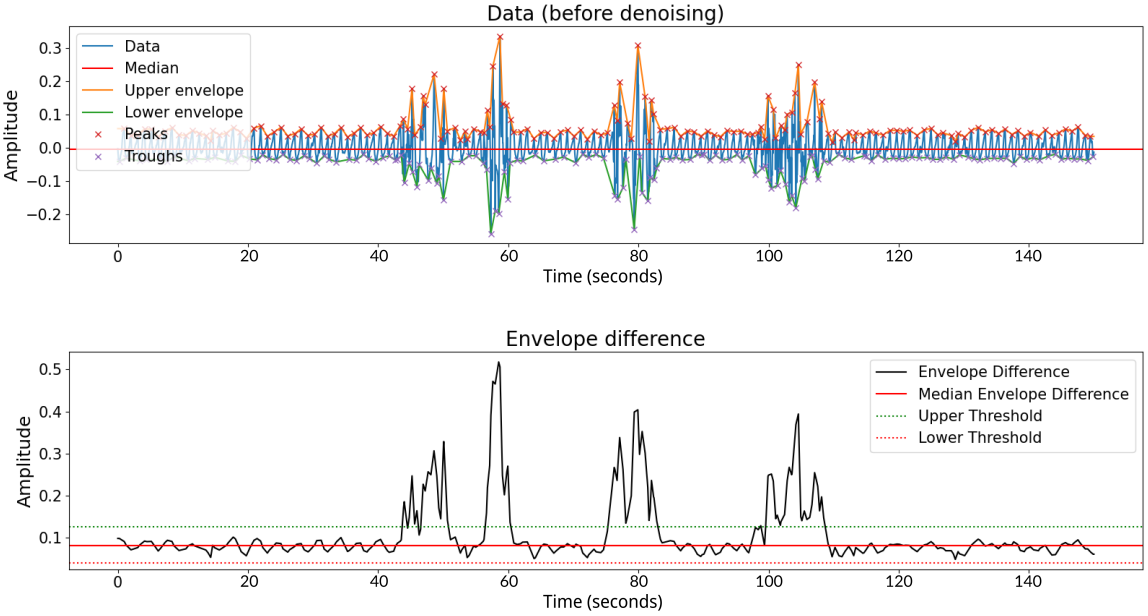

Figure 3.7: Figure with two subplots: The top subplot displays the raw data, peaks, troughs, median of the raw data, and upper and lower envelope. The bottom subplot illustrates the envelope difference, median of the envelope difference, and upper and lower thresholds.

The EPDA incorporates two distinct functions to effectively detect different types of noise. One function targets the identification of motion artifacts, while the other focuses on detecting flat line segments in the data. For motion artifact detection, the EPDA dynamically establishes upper and lower thresholds based on the interquartile range (IQR) of the data. The thresholds are calculated using the equations 3.1 and 3.2

$$\boldsymbol{upper} = 75^{th}\ percentile\ +\ (threshold\ *\ interquartile\ range) \tag{3.1}$$

$$\boldsymbol{lower} = 25^{th}\ percentile\ -\ (threshold\ *\ interquartile\ range) \tag{3.2}$$

The IQR is robust to outliers and provides a measure of the spread of data that is less influenced by extreme values than measures like the standard deviation. By using the IQR, we ensure that our thresholds are based on the central 50% of the data, making them less sensitive to outliers and more representative of the typical data distribution. The choice of percentiles, specifically the 25^{th} (lower quartile) and 75^{th} (upper quartile) percentiles, is a common approach in statistics. These percentiles divide the data into

quartiles, with the IQR representing the spread of the middle 50% of the data. By setting the lower threshold at the 25^{th} percentile and the upper threshold at the 75^{th} percentile, we are effectively capturing the central range of data where most observations are expected to fall. This ensures that our thresholds are meaningful and relevant to the majority of data points. By introducing a user-defined threshold multiplier we aim to enable flexibility and adaptability in the anomaly detection process. By multiplying the IQR with the user-defined threshold value, we enable users to adjust the strictness of the thresholds according to their specific needs. A higher multiplier makes the thresholds more permissive, while a lower multiplier makes them more stringent. This parameter allows users to fine-tune the sensitivity of the algorithm to anomalies.

The user can define the *threshold value*, with a default value of 2 (within this study, the default value of 2 was maintained across patient data). When detecting flat line segments, both the *line height threshold* and *line temporal threshold* are user-defined. The *line height threshold* determines the minimum required amplitude to be considered a flat line, while the *line temporal threshold* specifies the duration for which data must remain below the amplitude threshold to be identified as a flat line. Thus, if a segment of data remains below the amplitude threshold for a duration equal to or longer than the temporal threshold, it is identified as a flat line.

Within the function for motion artifact detection, all instances where a slope change occurs within the envelope difference is calculated. Any instance of data with a slope change that exceeds the upper threshold or falls below the lower threshold is identified as an anomaly. Pseudocode 1 outlines the motion artifact anomaly detection algorithm.

Pseudocode 1 Motion artifact detection

```
slope_change_indexes ← empty list                ▷ Initialise an empty list to store slope change indexes
for i ← 1 to length(envelope_difference) − 1 do
    diff_prev ← envelope_difference[i] − envelope_difference[i − 1]
    diff_next ← envelope_difference[i + 1] − envelope_difference[i]
    if (diff_prev ≥ 0 and diff_next < 0) or (diff_prev < 0 and diff_next ≥ 0) then
        slope_change_indexes.append(i)                ▷ Add current index to the list
    end if
end for
anomalies ← empty list                ▷ Initialise an empty list to store anomalies
for index in slope_change_indexes do
    if envelope_difference[index] > Upper Threshold or envelope_difference[index] < Lower Threshold then
        anomalies.append(index)
    end if
end for
```

In order to increase the computational efficiency of the denoising algorithm the motion artifact detection ensures that not all instances of data between the first and last threshold crossing points are defined as an anomaly. Instead, only instances of data where a slope change occurs within this range are considered anomalies. This logic minimises unnecessary iterations when calculating the segmentation points for each detected anomaly and substantially reduces the computational expense of the algorithm. Figure 3.8 depicts the difference between the two approaches of detecting anomalies which amounts to a 5400% difference between approaches for the example data.

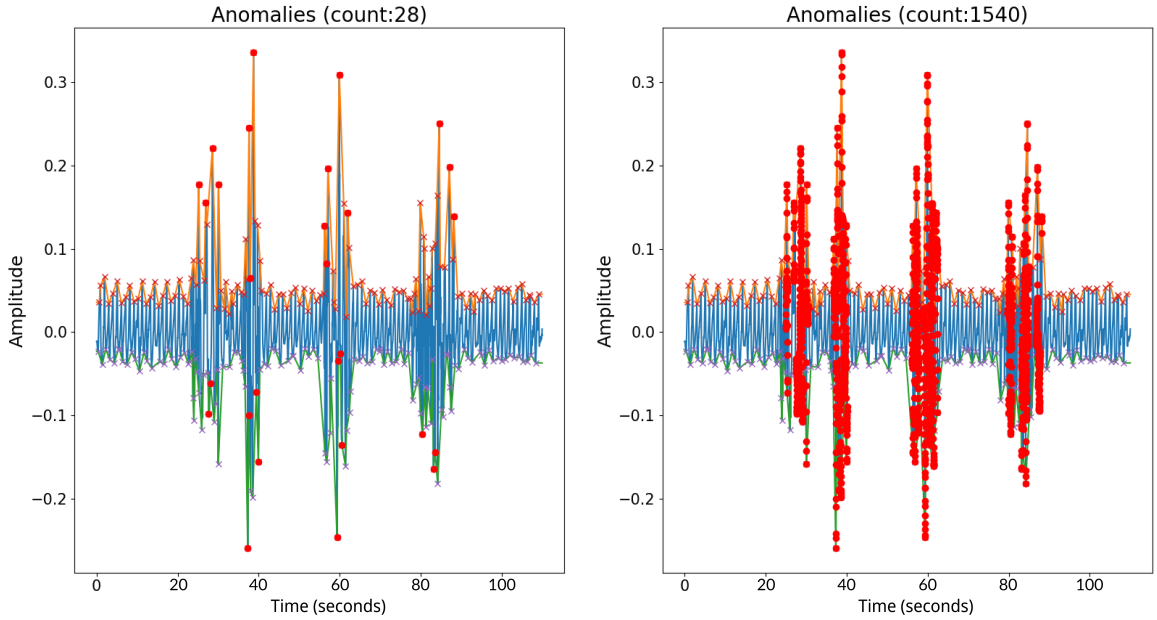

Figure 3.8: Figure with two subplots: The left subplot shows the results of the optimised anomaly detection approach (count: 28). The right subplot displays the results of the unoptimised anomaly detection approach (count: 1540). A 5400% difference between the two approaches

The logic for the detection of flat line segments of the data can be divided into two steps. The first step involves applying the user-defined *line height threshold* and representing the data as 0s and 1s. Values in the envelope difference that exceed the *line height threshold* are set to 1. All values less than the *line height threshold* are set to 0. This representation allows for the simple characterisation of the start and end points of potential flat line segments in the data. Once the data is represented as 0s and 1s, flat line segments are identified using the algorithm described in Pseudocode 2.

Pseudocode 2 Flat line detection

```
flat_line_segments ← empty list ▷ Defining a list to store the flat line segments of the signal
diffs ← list ▷ A list containing the differences between adjacent elements in envelope_difference
line_height_threshold ← float ▷ A user defined height threshold
line_temporal_threshold ← float ▷ A user defined temporal threshold
height_detection ← where(abs(data_chunk_filt) < line_height_threshold, 0, 1) ▷ Representing the data as either 0 or 1
segment_indices ← find_non_zero_segments(diffs) ▷ Indices where the difference between adjacent elements is non-zero
segment_indices ← [0] + segment_indices + [length(height_detection)] ▷ Segment indices with the start and end indices
for i ← 0 to length(segment_indices) − 1 step 2 do
    start_index ← segment_indices[i]
    end_index ← segment_indices[i + 1] − 1
    if not any(height_detection[start_index : end_index]) then
        if (end_index − start_index + 1) ≥ (fs × line_temporal_threshold) then
            if start_index < 0 then
                start_index ← 0
            end if
            if end_index > length(envelope_difference) then
                end_index ← length(envelope_difference) − 1
            end if
            flat_line_segments.append((start_index, end_index))
        end if
    end if
end for
```

The flat line detection code searches for 1s within the augmented data which indicate the possible start and end points of flat line segments. If the distance between a start and end point is $\geq$ to the user defined *line temporal threshold*, the segment is identified as a flat line. The output of the flat line detection code is a list of indexes representing the start and end indexes of the flat line segments in the data.

Once any motion artifacts and flat line segments are detected and their indexes recorded, suitable segmentation points are calculated for each identified anomaly. These segmentation points are determined based on the indexes where the envelope difference and the median of the envelope difference intersect or closely align. Once these indexes have been calculated the segmentation points for each anomaly is derived. The result of the segmentation code is a list of indexes that represent the start and end points of anomalous data. In the motion artifact detection algorithm, the crucial step of only considering instances of data that surpass a threshold and are indexes where slope change occurs as anomalies plays a significant role in the calculation of segmentation points. By focusing only on the instances where the slope change exceeds the threshold, the algorithm can highlight potential areas of interest, which are most likely to contain anomalies caused by motion artifacts. This approach reduces the number of iterations required

and eliminates the computation of redundant segmentation indexes for the same identified anomaly. As a result, the computational complexity of the algorithm is significantly reduced. The algorithm can process the data more efficiently, leading to faster anomaly detection. This optimisation is important when dealing with large datasets, real-time processing, or resource-constrained environments, as it helps improve the overall performance of the motion artifact detection process.

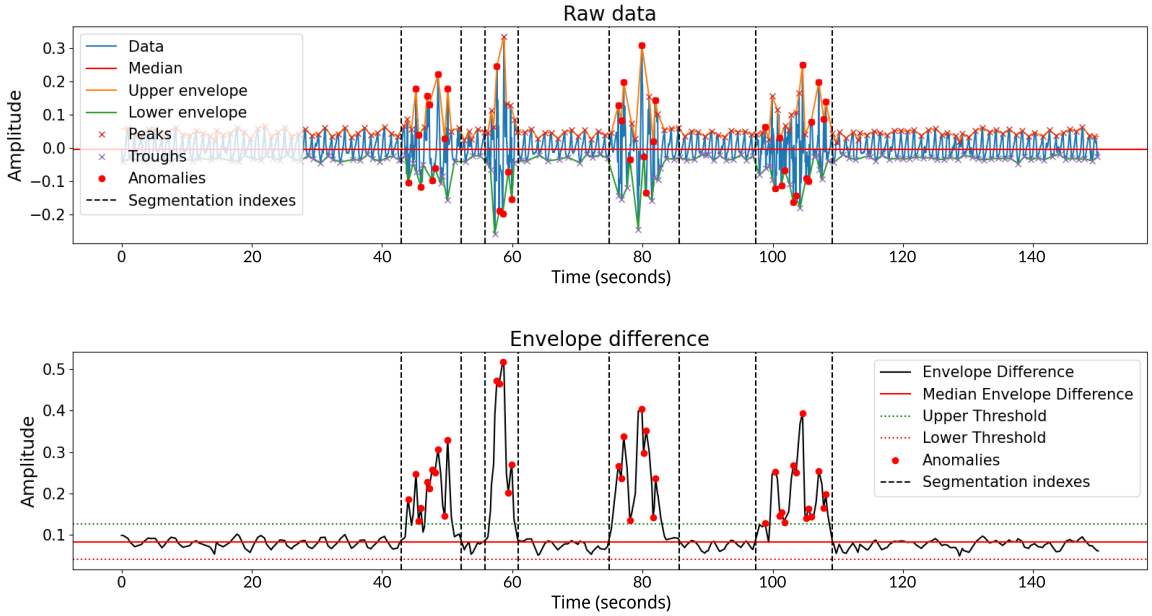

Figure 3.9: Figure with two subplots: The top subplot presents the raw data, including peaks, troughs, median values, upper and lower envelopes, anomalies, and calculated segmentation points. The bottom subplot illustrates the envelope difference, median difference, upper and lower thresholds, along with anomalies and calculated segmentation points

Before removing the index ranges between the calculated segmentation points for each anomaly, the algorithm combines adjacent "segmentation sections" that are in close proximity to create larger segmentation sections. This process serves two purposes, it: i) reduces the computational complexity of the algorithm by minimising the number of segmentation sections that need to be iterated, and ii) incorporates potential data instances between segmentation sections that may not have been explicitly identified as anomalies but exhibit poor morphological quality. Once the anomalous regions of the data have been removed, the data is filtered using a lowpass 2nd order Butterworth filter with a cutoff frequency of 10hz. Figure 3.10 present the data before and after denoising.

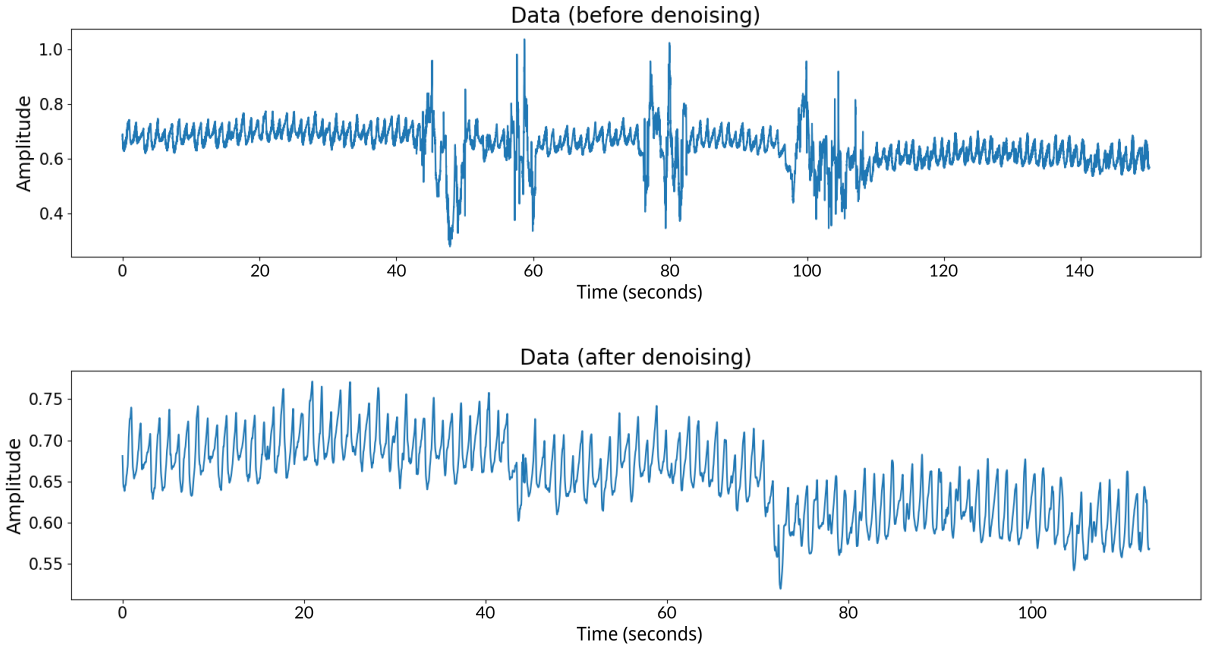

Figure 3.10: Figure with two subplots: The top subplot presents the raw data. The bottom subplot illustrates the denoised data

Algorithm evaluation

To evaluate the performance of the EPDA, an assessment was conducted using the NIR-PPG dataset comprised of approximately 81,015.99 minutes or 1360.27 hours of data from 31 patients. In the absence of a "clean" reference dataset, the evaluation was carried out through the calculation and analysis of 5 metrics. These included: (i) signal-to-noise ratio (SNR), (ii) variance, (iii) total variation (TV), (iv) Shannon Entropy, and instances-per-second (IPS)

SNR is used to quantify the strength of the signal relative to noise through the ratio of the power of the signal to the power of the noise. Throughout the data collection process, the patients were situated in hospital beds, the majority of patients were sedated. Given this, the patients remained predominantly motionless during data collection, occasionally being repositioned by nursing staff. As a result of this environment, it was inferred that the primary sources of noise within the data were likely to be of higher frequency outside of the typical frequency bands of the PPG. Given this context and the absence of a standardised method to calculate the SNR a cutoff frequency of 10 Hz was deemed appropriate inorder to isolate the morphology of the PPG using a lowpass filter and to isolate higher frequency noises using a highpass filter. When calculating the SNR of the data, the signal was filtered using a 2nd order, lowpass Butterworth filter with a cutoff

frequency of 10 Hz. The noise within the signal was isolated using a 2nd order, highpass Butterworth filter with a cutoff frequency of 10 Hz. The SNR equation:

$$\text{SNR} = \frac{P_{\text{signal}}}{P_{\text{noise}}} \tag{3.3}$$

Where P_{signal} is the power of the signal, calculated as the mean of the squared signal and P_{noise} is the power of the noise within the signal, calculated as the mean of the squared isolated noise. A higher SNR value indicates a stronger, more distinguishable signal compared to noise, whereas a lower SNR value indicates that the signal is relatively weak compared to the noise. A desirable outcome after denoisisng would be the increase in SNR, indicating effective removal or reduction of noise while retaining signal content.

Variance captures the spread or dispersion of data points around the mean, reflecting the average squared difference of each data point from the mean value, formally represented as:

$$\text{Var}(X) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{3.4}$$

Where N represents the total number of data points in the dataset and x represents each individual data point in the dataset. TV measures the overall variations in signal intensity across the entire duration by summing the absolute differences between adjacent data points:

$$\text{TV} = \sum_{i=1}^{N-1} |x_{i+1} - x_i| \tag{3.5}$$

Where N represents the total number of data points in the dataset and x represents each individual data point in the dataset. We hypothesise that a reduction in variance and TV values before and after denoising signify smoother and less variable signals, indicating efficacious noise suppression and enhanced signal regularity.

Shannon entropy characterises the complexity or randomness of a signal by quantifying the average information required to describe the signal's values:

$$H(X) = -\int_{-\infty}^{\infty} f(x)\log_2(f(x))\,dx \tag{3.6}$$

Where $f(x)$ denotes the probability density function (PDF) of the continuous random

variable x, $\log_2(f(x))$ represents the logarithm base 2 of the probability density function. A higher Shannon entropy value indicates greater complexity or randomness, while a lower value suggests more predictability or regularity. We posit that a decrease in Shannon entropy post-denoising indicates successful noise reduction.

These 4 metrics, which can be calculated without a "clean" reference dataset, capture the dispersion, randomness and strength of a signal and thereby provide a credible means to evaluate the efficacy of a denoising algorithm. To assess the statistical significance of the differences in metrics calculated before and after denoising, a statistical testing approach was employed. Firstly, the normality of the calculated metrics was evaluated using the Shapiro-Wilk test. If the data in both groups exhibited a normal distribution ($p > 0.05$), a parametric paired t-test ($a = 0.05$) was chosen for further analysis. In cases where one or both groups displayed a non-normal distribution ($p > 0.05$), the non-parametric Wilcoxon signed-rank test ($a = 0.05$) was opted for. This dual testing strategy ensured a robust assessment of the statistical significance of the observed differences in metrics. The choice of parametric or non-parametric test was based on the underlying distribution of the data, enhancing the reliability and validity of the statistical analysis.

The computational speed of the EPDA was assessed by recording the run time of the algorithm on each data sample. From this a 5th metric, instances-per-second was derived which gives a representation of the speed of the EPDA. The analysis of the algorithm's speed enables an assessment of the EPDA's suitability for big datasets and real-time applications. This evaluation approach allows for a holistic understanding of the denoising algorithm's effectiveness.

Results

Figure 3.11 provides a representative example of the complete data of one patient, both before and after denoising. Upon visual inspection, the data after denoising exhibits reduced extreme amplitude variations, which are believed to be associated with motion artifacts. Additionally, the range and dispersion of data following denoising appear narrower, suggesting a more uniform distribution. These observations suggest that the denoising process effectively mitigates motion-related artifacts and leads to a more consistent and refined dataset.

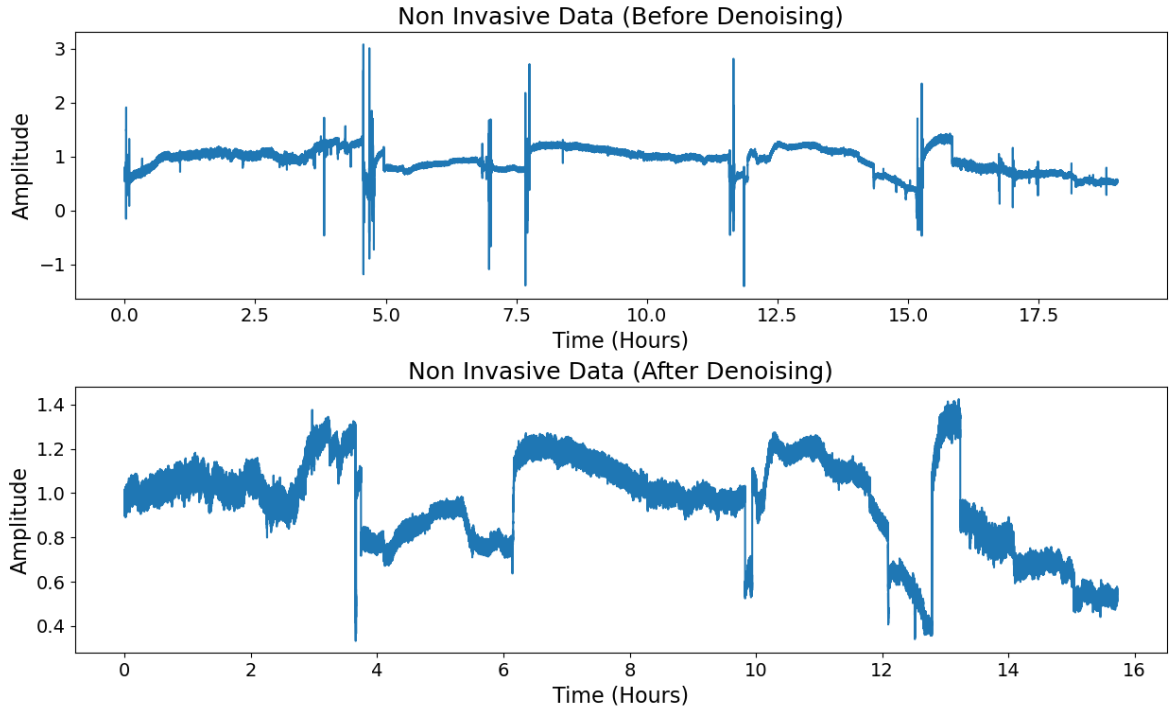

Figure 3.11: Figure with two subplots: The top subplot presents the raw data of an entire patient. The bottom subplot illustrates the denoised data

Table 1 provides an overview of the EPDA's impact on various metrics, namely the SNR, variance, total variation, and entropy, for each patient. The data includes measurements before and after denoising, along with the computation time of the denoising algorithm and the derived metric IPS.

Examining the results, we observe the affects of the denoising algorithm on the data. The SNR indicates an average improvement from 53.3 (SD: 8.62) to 74.69 (SD: 7.63) after denoising. This represents a 40.16% increase in SNR suggesting that denoising has significantly enhanced the signal power, leading to a more accurate representation of underlying information.

Moreover, the variance, total variation, and entropy, which respectively capture the amount of variability, structural complexity, and information content in the data, exhibit a reduction of approximately 45.59%, 85.84%, and 2.16% on average after denoising. This signifies a successful reduction in unwanted noise and random fluctuations in the data.

The observed reduction of approximately 85.84% in total variation after denoising indicates a success in simplifying the data distribution and eliminating unnecessary variations.

Motion artifacts often introduce irregularities into the data usually manifesting within

the signal as large amplitude variance. The decrease in total variation observed in our results suggests that the EPDA has effectively reduced the impact of motion artifacts, leading to a more uniform and smoother representation of the data. Motion artifacts which cause sudden changes in the amplitude of the data lead to higher values of total variation. When these artifacts are successfully removed, the abrupt changes caused by the motion artifacts are replaced with smoother transitions, resulting in a reduction in total variation.

Patient	SNR before	SNR after	Variance before	Variance after	Total variation before	Total variation after	Entropy before	Entropy after	Computation time (s)	Instances per second
1	31.189	65.903	0.048	0.013	140773.859	10909.716	16.527	16.352	36.899	411135.192
2	47.007	65.737	0.386	0.235	21739.31	4100.025	16.506	16.2	37.388	411730.976
3	46.194	72.751	0.051	0.046	37075.196	5405.411	15.734	15.601	16.071	425686.583
4	61.008	78.478	0.343	0.084	9166.518	3364.145	16.16	15.897	31.84	339511.046
5	48.193	66.293	2.982	0.601	71059.79	6086.925	16.032	15.444	31.152	422664.52
6	60.696	73.918	1.036	0.795	23131.757	5004.932	16.575	16.194	41.182	409219.627
7	41.05	67.459	0.423	0.256	192933.432	19756.286	16.689	16.501	43.945	417796.334
8	54.202	73.708	0.324	0.277	10639.948	1929.227	15.449	15.312	14.546	417460.696
9	62.55	83.328	0.16	0.163	17511.129	7775.109	16.695	16.588	45.844	400679.008
10	44.915	72.867	0.445	0.394	83002.505	18462.602	16.642	16.458	38.469	449186.729
11	50.319	68.012	1.405	0.553	72820.156	10622.558	16.592	15.763	36.006	479913.608
12	44.613	69.338	1.009	0.551	171845.39	14079.652	16.568	16.259	38.774	441844.878
13	59.473	77.376	0.656	0.57	35550.527	6366.344	16.553	16.337	39.746	403186.178
14	56.559	72.382	1.719	0.956	31979.635	4606.092	16.558	16.004	47.846	359541.675
15	51.393	67.601	1.497	0.885	36806.149	10484.826	16.54	16.289	47.248	364712.38
16	45.931	62.772	2.355	0.702	144723.747	14598.065	16.459	15.975	36.288	476012.512
17	55.117	74.773	0.57	0.153	17258.795	3767.316	16.646	16.381	47.626	379509.037
18	62.904	84.446	0.026	0.026	17806.375	4194.915	16.639	16.512	43.365	390169.125
19	66.553	89.476	0.015	0.009	6909.931	1307.428	16.614	15.562	40.313	407831.278
20	61.429	91.887	0.016	0.008	28334.381	1578.581	16.668	16.268	42.294	410127.987
21	54.656	72.106	0.299	0.292	34702.886	11940.3	16.625	16.376	38.504	448506.658
22	43.825	71.005	0.215	0.185	192317.583	16779.758	16.652	16.501	40.587	425845.848
23	62.883	88.807	0.015	0.012	18877.106	2112.399	16.659	16.128	37.575	457268.441
24	51.973	70.694	0.356	0.159	29089.716	4875.059	16.633	15.723	45.736	372933.237
25	40.743	67.083	0.022	0.022	280400.591	13161.378	16.598	16.038	34.196	473043.35
26	62.51	74.459	0.223	0.208	35619.511	23093.016	16.628	16.257	30.057	559857.53
27	60.102	75.865	0.351	0.327	13403.344	7517.603	16.169	15.973	25.66	417419.782
28	59.835	86.582	0.027	0.021	18288.607	3848.823	16.663	16.535	42.04	410817.225
29	44.86	73.132	0.505	0.498	63663.937	5187.896	15.949	15.612	19.78	438663.051
30	59.41	76.07	0.122	0.085	31772.232	20243.942	16.656	16.476	46.778	371073.191
31	60.352	80.996	0.717	0.879	28295.027	8401.34	16.826	16.344	43.524	483475.432

Table 3.1: Table presenting the Signal-to-noise Ratio (SNR), Variance, Total variation, and entropy of the data for each patient before and after denoising. The table also includes the computation time of the algorithm for each patient and the derived metric Instances-per-second.

To assess the impact of denoising on the calculated metrics, a statistical analysis was conducted. The normality of the data distribution for each metric was first evaluated using the Shapiro-Wilk test. The results revealed that the data of the SNR was normally distributed with a p-value of 0.057 and 0.055 for the "before denoising" and "after denoising" groups respectively. The data of the remaining 3 metrics was found to be non-normally distributed. Therefore a non-parametric Wilcoxon signed-rank test was chosen to determine if there were significant differences in all 4 metrics before and after denoising, using a significance level of 0.05. By using a single non-parametric test for all the metrics, despite the normality of one we ensure consistency across the analysis. Allowing for a reliable comparison of the denoising effect across all the metrics while accommodating potential data non-normality.

The Wilcoxon signed-rank test which compares the distribution of the differences between paired observations suggests that there was a significant difference in the SNR (p-value $= 9.313225746154785e-10$), variance (p-value $= 9.955838322639465e-07$), total variation (p-value $= 9.313225746154785e-10$) and entropy (p-value $= 9.313225746154785$e-10) before and after denoising. Showing a significant difference in all 4 metrics. Figure 3.12 contains the boxplots representing the values for each of the metrics before and after denoising.

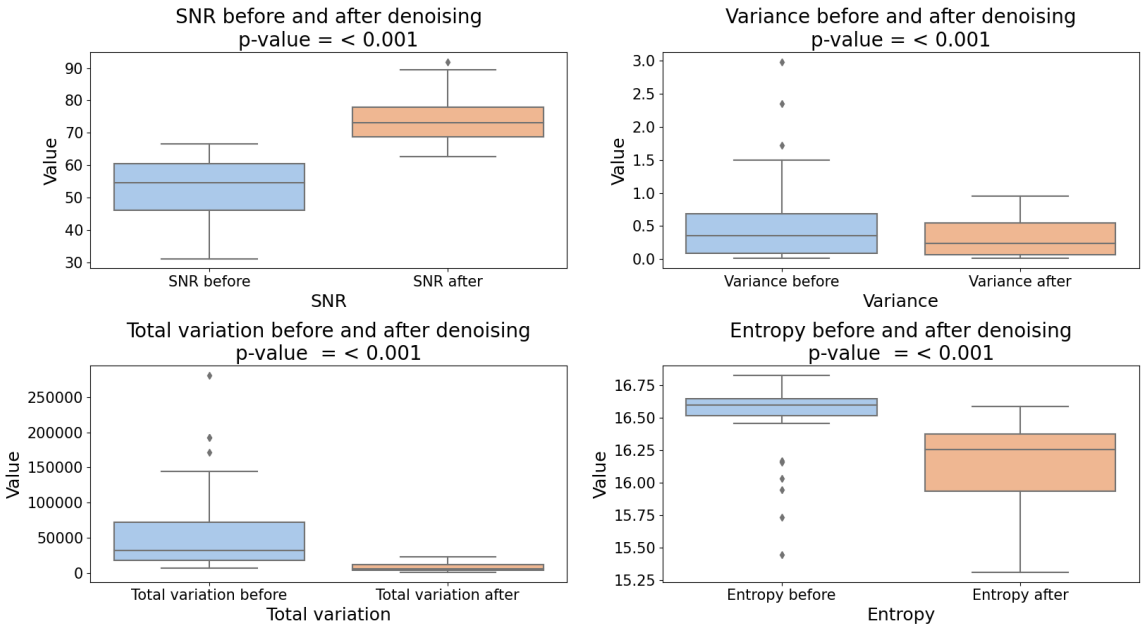

Figure 3.12: Figure with four subplots, each subplot contains 2 boxplots representing the values for the signal-to-noise ratio, variance, total variation and entropy of the data before and denoising.

In addition to the analysis and interpretation of the 4 metrics, the IPS was also calculated for each patient's data. The calculated average IPS value across patients amounted to 421,833.003, signifying a substantive computational speed. This IPS can be presented as processing 70.3 minutes of data per second which suggests the EPDA may have potential for real-time applications and scenarios necessitating efficient data processing such as large retrospective studies.

This metric, when considered in conjunction with the analysis of the four other evaluation metrics, provides a comprehensive assessment of the algorithm's overall effectiveness. These findings demonstrate significant improvements in SNR, variance, total variation, and entropy metrics following denoising, as supported by both statistical summary of the results, statistical testing through the use of Wilcoxon signed-rank test and visual analysis of the boxplots.

Discussion

The evaluation of denoising algorithms in the absence of clean reference data is challenging. The evaluation through the use of SNR, variance, total variance, entropy and IPS endeavours to capture the dispersion, randomness and strength of the signal before and after denoising in addition to its computation speed to help evaluate the EPDA and its utility.

While searching for applicable evaluation metrics we noted that there is a dearth of research which focuses on PPG denoising evaluation, with some work on the quality assessment of pulses through the use of a number of signal quality indexes [116]–[118]. These indices include, skewness, kurtosis and perfusion index (PI). We suggest that these indices while they appear to be credible for the assessment of individual PPG pulses may not directly capture the impact of denoising. We propose that while skewness, which measures data asymmetry, and kurtosis, which measures peakedness or flatness, are relevant for characterising the morphology of individual PPG pulses, they may not offer meaningful insights into the effectiveness of denoising methods. Denoising techniques aim to reduce noise in the PPG signal, but their success may not always be easily interpretable using skewness and kurtosis alone. For instance, if a dataset contains substantial high-amplitude noise at the beginning, causing overall skewing, and this noise is successfully removed by denoising, it might appear as if the denoising was effective based on the

restoration of symmetry. Conversely, if noise is distributed evenly throughout the dataset, denoising might effectively reduce the noise without significantly impacting the overall skewness. As a result, skewness and, similarly, kurtosis do not appear to be reliable indices for evaluating the efficacy of denoising algorithms.

Motion artifacts in PPG signals can introduce high-amplitude variance, leading to fluctuations in the signal that are not related to changes in perfusion. We posit that these motion-related fluctuations may artificially increase the amplitude of the PPG waveform, positively biasing the calculations of perfusion index (PI). PI is often calculated as the ratio of the pulsatile blood flow to the non-pulsatile blood flow, and motion artifacts can falsely elevate the pulsatile component, resulting in an inflated PI value. Therefore, when evaluating denoising algorithms on PPG datasets containing motion artifacts, denoising can effectively remove the noise and reduce the artificially inflated amplitude variance caused by motion, leading to a counterintuitive reduction in PI. This reduction may be misinterpreted as a deterioration in the signal quality, while in fact, it reflects the removal of motion artifacts and the restoration of a more accurate representation of the underlying physiological information.

The ability of the EPDA to return the indexes of identified anomalous data is important when dealing with two datasets that are synchronously collected. In scenarios where two datasets are collected simultaneously for purposes such as implementing supervised predictive modeling, maintaining temporal synchronicity between the datasets is crucial to ensure the production of reliable and accurate results. The EPDA's capability to identify and provide the indexes of anomalous data allows users to remove the corresponding indexes in the second dataset. By doing so, the temporal synchronicity between the two datasets is preserved. This is particularly useful when using PPG data to predict physiological biomarkers and constructing labeled datasets for supervised learning tasks to help ensure the accuracy and reliability of the results obtained.

The proposed algorithm demonstrates potential although we acknowledge there are avenues for improvement. One of these is the implementation of a sliding window approach. The sliding window approach in place of a fixed window size may provide the calculation of more contextually relevant thresholds by considering short-term fluctuations within each window. This is likely to facilitate the algorithm's ability to changes in signal characteristics and morphological quality resulting in a more robust approach.

A limitation of the study is the calculation of SNR without a "clean" reference and in

the absence of a standardised approach the SNR calculations of this study were predicated on a cutoff frequency of 10Hz under the assumption that due to the sedated or low motion state of the patients during data collection the majority of noise present within the data would be of higher frequency outside of the expected frequency band of the PPG. Consequently, the proposed SNR calculation is suitable when comparing what is expected to be within a normal PPG frequency range to higher-frequency noise. Which although may be suitable for our dataset but may not be generalisable. We suggest that future studies should be conducted to validate the EPDA on different datasets with a greater presence of noise within the PPG frequency band. The collection of bilateral PPG signals, with body movement added on one side could be collected to introduce motion artifacts into the signal and offer a reliable reference for SNR calculations.

The denoising of PPG signals plays a pivotal role in advancing non-invasive physiological monitoring. The "Envelope PPG denoising algorithm" is designed to accomplish the dual objectives of identifying and removing anomalous data while ensuring computational efficiency. This algorithm contributes to the advancement of PPG signal processing and its application in diverse healthcare domains, ultimately enhancing the potential for accurate patient monitoring.

Conclusion

This study aimed to contribute to the field of PPG signal denoising, acknowledging its importance in the context of non-invasive physiological monitoring. The investigation focused on the development and evaluation of the "Envelope PPG Denoising Algorithm". The approach was shown to be efficacious in the denoising of PPG data, reporting significant differences in the SNR, variance, total variation and entropy after denoising. The IPS of the EPDA was calculated as 421,833.003, which can be represented as processing 70.3 minutes of data per second.

The research has contributed a denoising algorithm, the EPDA, to the field of PPG signal processing. The work augments the potential for non-invasive accurate patient monitoring across various healthcare domains, ultimately helping to improve the impact of non-invasive monitoring technologies, such as wearables, in enhancing patient care and wellbeing.

3.3 Summary

From the clinically collected data, both the invasive ICP and non-invasive NIR-PPG data were significantly impacted from the presence of erroneous data and noise. The erroneous ICP data was handled via the definition of reasonable and physiological bounds. Data which fell outside of these bounds were subsequently removed from both the reference and NIR-PPG data.

The PPG signal is affected by numerous sources of noise, which include motion artifact, power line interference and photodetector saturation amongst others. Given this research is centered around the exploration of the relationship between PPG derived features and ICP, the efficacious denoising of the NIR-PPG data was essential for the removal of anomalous data which would enable the extraction of more reliable features and consequentially more robust results. The EPDA was developed and applied to the dataset as a reasonable and computationally efficient approach to denoising.

Chapter 4

Feature Engineering

Following the removal of erroneous ICP data and the application of the EPDA, the resulting, final dataset consisted of 231,267,378 samples of data, represented as 38,544.56 minutes or 642.41 hours of data collected from 27 patients. This chapter introduces the feature extraction, providing details on the window definition, fiducial point detection and calculation of derivatives. All implementation of feature extraction was conducted using the Python programming language.

4.1 Window definition

The NIR-PPG dataset was segmented into 1-minute windows, with the median ICP value over each window being used as a label. This 1-minute window size was chosen considering the possible clinical relevance and practicality of the prediction frequency. Longer window durations may be less favourable in a clinical setting, and the goal was to use a window size that maximises the amount of relevant data available for prediction while still being clinically manageable.

4.2 Baseline wandering removal and data normalisation

In order to attempt to diminish the possible affect of calibration and baseline wandering on the data collected from multiple patients the AC and DC components of each patient's signal data were isolated using a bandpass and lowpass filter respectively. The bandpass

frequency band ranged from 0.4 to 10 Hz, the low pass cutoff frequency was set to 0.4 Hz. By dividing the AC component of the signal by the DC component the aim was to minimise the effects of the patient level calibration and provide a standardised representation of the signal that is less influenced by calibration variations. Following this, each window's data were normalised to a range of 0 to 1 to account for inter-patient amplitudinal variance and to aid in the creation of a shared distribution across patients in the feature space.

4.3 Fiducial point detection algorithm

Due to the presence of noise within the dataset and the lack of a reliable diacrotic notch across patient and between patient's data and in order to maximise the reliable detection of pulses, the fiducial point algorithm was focused upon three fiducial points specifically the pulse onset, peak and end.

The detection of the fiducial points is orientated around the moving average of the input data. The moving average was calculated over a window size which matched the sampling frequency. The crossing points between the moving average and the filtered data were calculated.

These crossing points were used in pairs to detect either peaks or troughs within the data. The mean of the data between two crossing points was compared against the mean value of the moving average within the same interval. If the mean of the data segment was higher than the moving average the location of the maximum value between crossing points was classified as a peak. Vice versa if lower, the index of the data which had the lowest value between points was classified as a trough. Figure 4.1 illustrates and described the three major steps of the peak detection algorithm.

The derivatives of the signal were calculated using a Savitzky-Golay filter using a polynomial of 7 and a window size of 101. Figures 4.2, 4.3 and 4.4 depict the pulse detection on the original, first and second derivative of a one minute window of data.

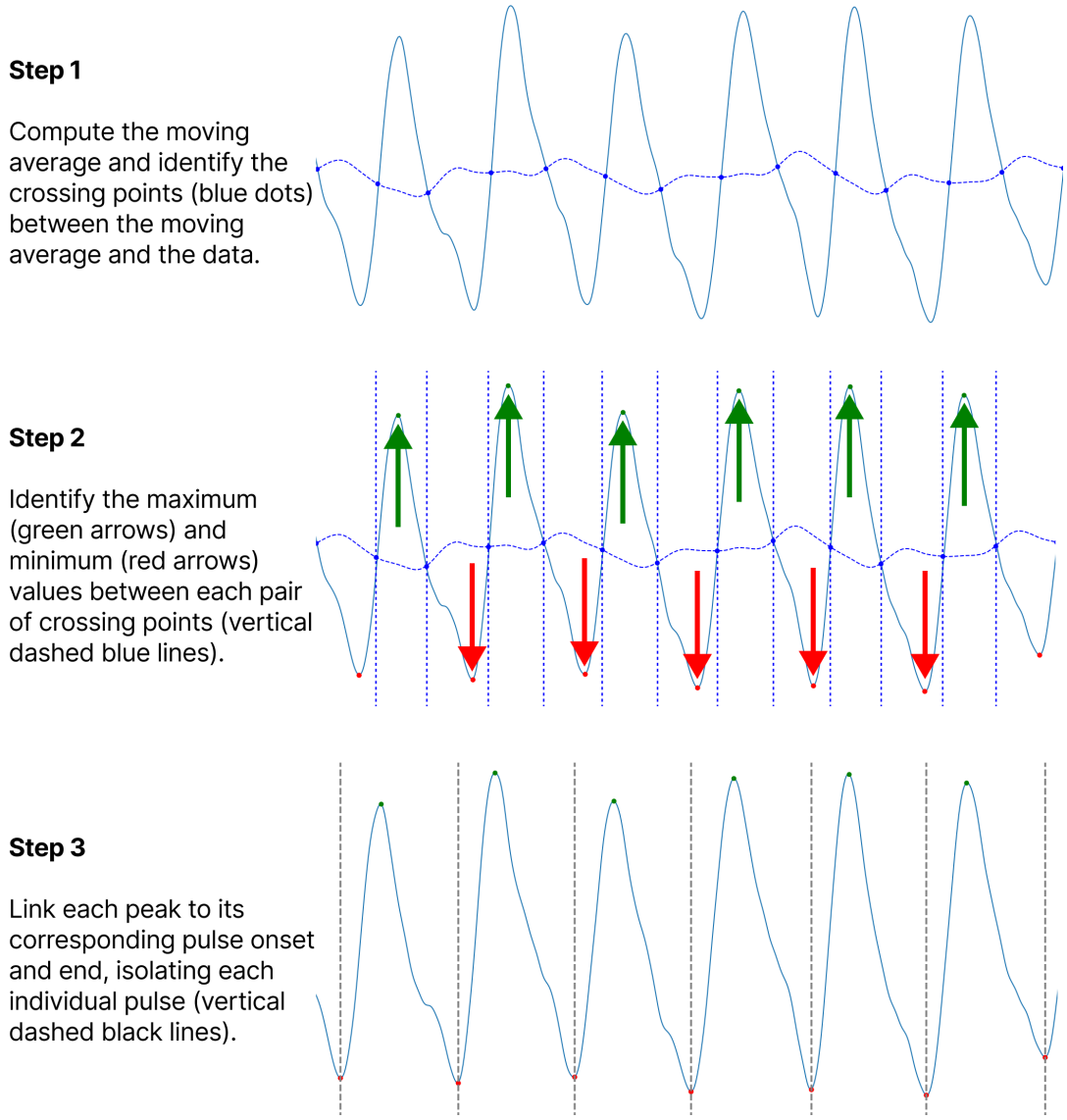

Figure 4.1: This figure illustrates the major steps of the peak detection algorithm. The process begins with the calculation of the moving average and identification of the crossing points between the moving average and the raw data. Next, the algorithm determines the maximum and minimum values between each pair of crossing points, highlighted by green and red arrows, respectively. Finally, each detected peak is linked to its corresponding pulse onset and end, isolating individual pulses.

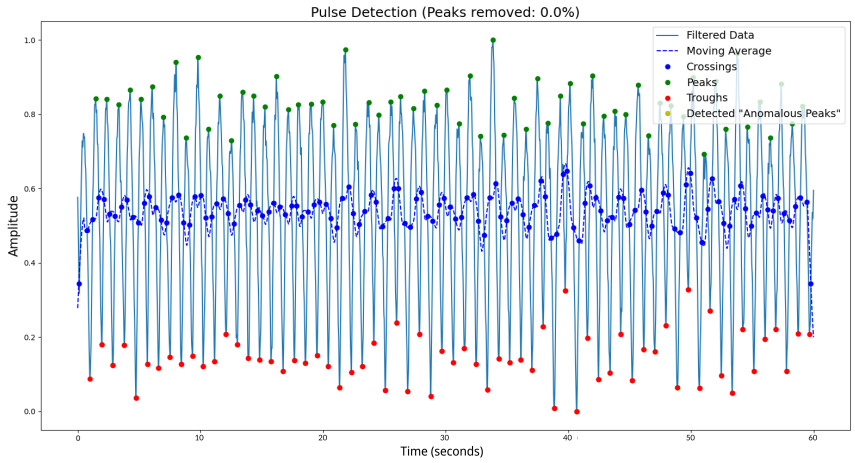

Figure 4.2: A figure depicting the detected fiducial points on a one minute window of signal data.

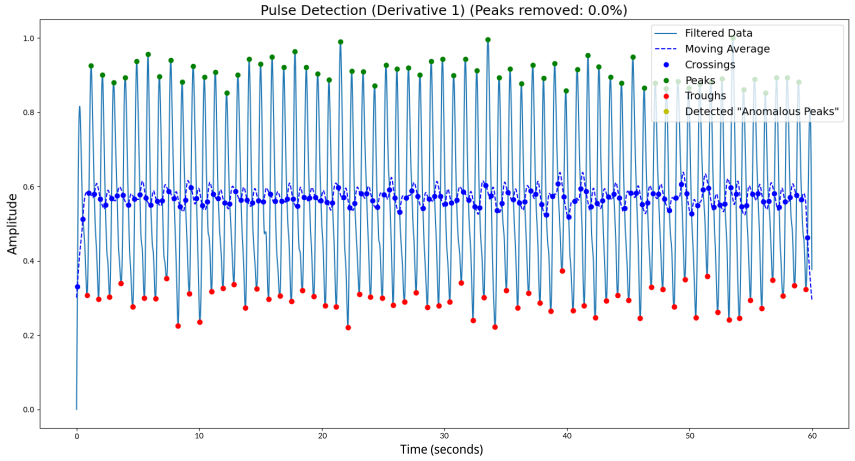

Figure 4.3: A figure depicting the detected fiducial points on a one minute window of first derivative signal data.

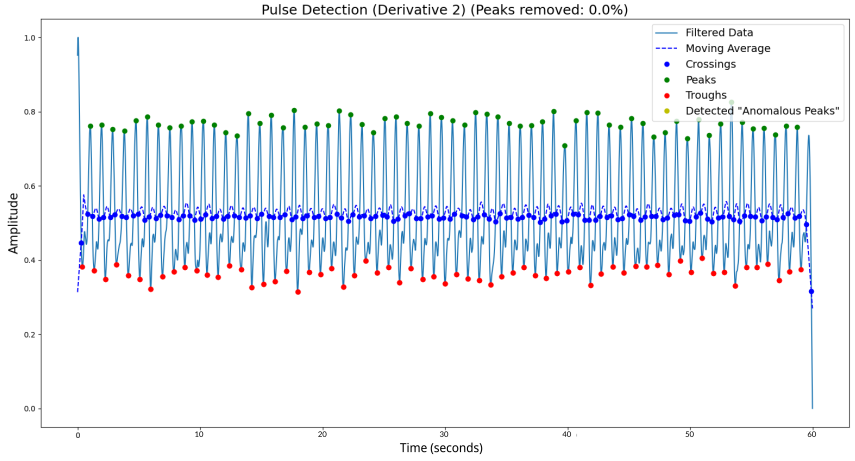

Figure 4.4: A figure depicting the detected fiducial points on a one minute window of second derivative signal data.

In order to handle the presence of possible residual anomalous pulses within each window a z-score outlier detection approach was developed and applied using three pulse characteristics: (i) difference between consecutive peaks, (ii) pulse onset-peak difference and (iii) peak-end difference. A z-score threshold of 3 was used. No features were extracted from pulses which were detected as anomalous. Figure 4.5 depicts an example of pulse detection with the identification of anomalous pulses.

Figure 4.5: A figure depicting identification of anomalous pulses within a 1 minute window of signal data.

4.4 Extracted features

The aim and purpose of feature engineering and the subsequent feature extraction, was to extract as much information from the pulsatile AC component of the signal aiming to identify a broad spectrum of features potentially linked to haemodynamic changes associated with variations in ICP. To accomplish this, features were derived not only from the original signal (which captures blood volume fluctuations over time due to systolic and diastolic phases) but also from its first and second derivatives. The first derivative provides insight on the velocity of changes, and the second derivative contains information describing the acceleration of these changes. For each window of data 141 morphological and time-series features were extracted across the original signal and its first and second derivative. A total list of the features and their description can be found in Table A1 in the appendix. Figures 4.6 and 4.7 depict the primary time series and morphological

features extracted from the cardiac pulses. The feature representation for each window was determined by computing the median of each feature across all pulses within that window.

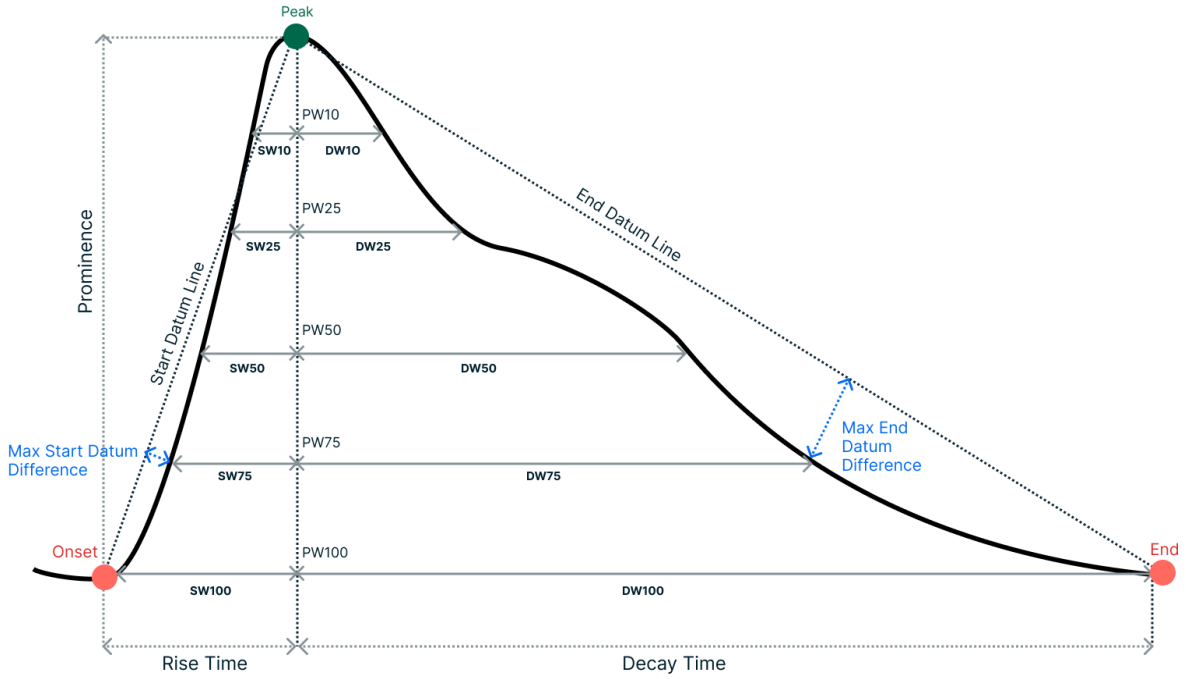

Figure 4.6: A figure illustrating the key time-series and morphological features extracted from individual cardiac pulses. Within the figure, SW, DW, and PW represent systolic width, diastolic width, and pulse width, respectively. The numerical value associated with SW, DW, or PW indicates the position along the pulse prominence where the corresponding measurement is taken.

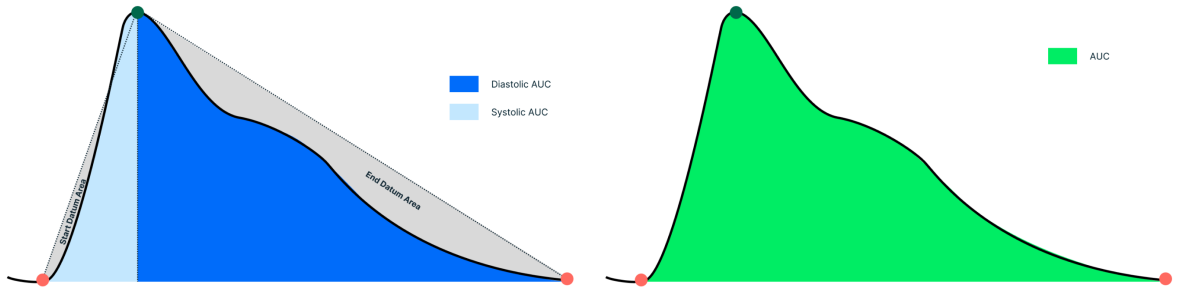

Figure 4.7: A figure depicting the Area Under the Curve (AUC), along with the Diastolic AUC and Systolic AUC. Additionally, the start and end datum areas are identified.

Linking back to Section 2.2 and Section 2.4.4, which introduce cerebral physiological and haemodynamic components and the theory describing the relationship between PPG and ICP, respectively, there are three major components that the extracted features aim to characterise, which are closely associated with changes in ICP. These components are:

(i) changes in blood volume, (ii) changes in the mechanistic properties of the vessels as a consequence of variations in cerebral compliance, and (iii) features associated with blood pressure. As noted in Section 5.4, as ICP increases and cerebral compliance reduces, vascular resistance increases, triggering autoregulation and causing CPP to increase to maintain CBF.

The first component the features aim to represent is possible blood volume changes due to increases in ICP and subsequent drops in CBF as autoregulation maintains CPP. It is well understood that the amplitude of the systolic peak of the PPG signal represents relative levels of blood volume. During systole, the increased blood volume absorbs more light, reducing the light reaching the photodetector, which is represented as a higher pulse amplitude on the PPG waveform [119]–[123]. This change in blood volume may also manifest in changes to the pulse width and area under the curve of the pulse. Pulse width has been associated with vascular compliance [124]. A wider pulse width suggests increased blood volume or vessel dilation, while a narrower pulse width indicates decreased blood volume or vessel constriction. The pulse area indicates the total volume of blood pumped by the heart during a cardiac cycle and has been linked to cardiac output [125]. The total area of the PPG waveform over time represents the total blood volume change. Since cardiac output is the product of the blood ejected by the heart per beat and the heart rate, a higher AUC typically corresponds to a higher blood volume. To characterise these volumetric changes, the total, systolic, and diastolic pulse widths at 10, 25, 50, 75, and 100 percent of the pulse amplitude were extracted, along with the total pulse area, systolic and diastolic pulse areas, and the ratio between the systolic and diastolic areas.

The increase in ICP exerts pressure on the cerebral vessels, leading to a decrease in cerebral compliance. This reduction in compliance is characterised by the vessels' diminished ability to deform, becoming less elastic and "stiffer" as ICP increases. Pulse width has been shown to be associated with vascular compliance [124] and increased arterial stiffness [126]. In this research, it is proposed that aterial stiffness has a similar effect to the reduction in vessel compliance due to increased ICP. Awad et al concluded that the pulse width of finger and ear PPG pulses is sensitive to changes in vascular resistance. Ferizoli et al's results supported these findings, showing that pulse width features and area-related features are closely associated with stiffness.

Ahn et al. investigated the concept of a vascular ageing index, where vascular ageing is linked to atherosclerotic disorders, resulting in vessel stiffening and altered mechanistic

properties [127]. Ahn et al.'s work utilised rise time and decay time (the time between pulse onset and systolic peak, and the time between systolic peak and pulse end, respectively) which has also provided insights into reactive hyperemia [128], along with area-related features such as the ratio between the areas of the systolic and diastolic pulse elements. Additionally, research by Takazawa et al. [129] and others [130] used the second derivative of the PPG signal to measure vascular ageing and stiffness.

Therefore, the features previously described for characterising blood volume changes, combined with rise and decay times, may also be useful for representing changes in cerebral compliance. To describe possible changes in compliance, these features were extracted across the first and second derivatives of the PPG signal, capturing the rate of change or acceleration of blood volume due to the diminished ability of the vessels to deform during systole and diastole. In addition, skew and kurtosis were also extracted, as they may be associated with characterising effects on the mechanistic properties of vessels. The work of Ferizoli et al demonstrated that skewness and kurtosis show stable increases with increases in arterial stiffness.

MAP and consequently blood pressure (BP) are physiological markers closely associated with variations in ICP. As described in Section 2.2, MAP is a critical determinant of CPP, which is defined as the difference between MAP and ICP. CPP is fundamental to maintaining adequate blood flow to the brain, ensuring that cerebral tissue receives sufficient oxygen and nutrients. Therefore, fluctuations in BP can directly impact CPP, thereby influencing CBF. Variations in ICP can cause increases in vascular resistance, which necessitates an increase in BP to maintain adequate CBF. Consequently, changes in BP and subsequently CPP and CBF may be associated with changes in ICP, as autoregulation attempts to regulate cerebral perfusion in response to varying ICP.

Highlighting the association between BP dynamics and ICP variations, related research has attempted to estimate ICP non-invasively using PPG signals in combination with BP-derived features. These studies have consistently shown that continuous BP is among the strongest predictive features for ICP estimation across various methodologies [94], [95]. Unlike these previous efforts, which focused on combining PPG signals with BP-derived features, this work investigates the direct relationship between the PPG signal and ICP.

Extensive research has been conducted into the use of PPG for blood pressure estimation [131]–[136]. Consequently, numerous features have been developed to estimate BP from PPG signals, many of which have been utilised in this research. In particular, the

studies by El Hajj et al. and Kurylyak et al. utilised features such as pulse width, systolic and diastolic pulse widths across multiple percentages of the pulse amplitude, pulse rise time, and decay time for BP estimation [136], [137]. Additionally, the "upslope", defined as the slope of a straight line from the pulse onset to peak, has been introduced as a useful feature for BP estimation [138]. Features derived from datum lines, such as straight lines from the pulse onset to peak and peak to pulse end, have been linked to systolic and diastolic blood pressure variations [139]. These features have been expanded upon; for example, the "downslope", the slope of a straight line from the pulse peak to end, was also extracted. This was combined with several ratios such as the ratio between the upslope and downslope, the ratio between the rise time and decay time, ratios of the diastolic and systolic pulse width across the percentages of the pulse amplitude, and the ratio between the pulse length and height, among others.

Given their associations with haemodynamics, it is hypothesised that these features will demonstrate an association with ICP and prove valuable for non-invasive ICP estimation.

4.5 Summary

This research aims to evaluate the relationship between NIR-PPG derived features and ICP. Therefore, one of the principal components is the extraction of morphological and time-series features from the signal. Previously we described the pre-processing of the signal, the removal of anomalous ICP values and the application of the EPDA which resulted in approximately 642 hours of data collected from 27 patients. In order to effectively characterise the pulsatile element of the NIR-PPG signal, 141 features were extracted for every one minute window of data across the original, first and second derivative.

To manage any residual noise within our data and to maximise the reliability of feature extraction the fiducial point algorithm which the features are predicated upon focused on the identification of three fiducial points, the pulse onset, peak and end. The efficacy of the pulse detection is crucial for the extraction of reliable features which will subsequently be utilised to determine the relationship between NIR-PPG derived features and ICP.

Chapter 5

Statistical Analysis

This chapter utilises the extracted features to statistically investigate the relationship between NIR-PPG derived features and variations in ICP.

5.1 Introduction and hypotheses

As described in Chapter 3, elevation in ICP can reduce CPP by increasing the resistance against arterial blood entering the brain. This elevation in ICP and consequential reduction in CPP triggers cerebral autoregulation. However, elevated ICP may reduce CPP below a critical threshold causing cerebral dysregulation leading to a sustained drop in CBF.

It was hypothesised in Chapter 5 that PPG may provide identifiable changes in its morphology which mirror variations in haemodynamics caused by increases in ICP.

This statistical analysis investigates the relationship between NIR-PPG features and variations in ICP and tests two primary hypotheses. Firstly, it examines whether alterations in NIR-PPG derived features correspond to changes in ICP levels and secondly, it explores whether these changes are more pronounced in features derived from long-distance NIR-PPG data compared to short-distance NIR-PPG data.

5.2 Data preparation

To effectively evaluate our hypotheses, we refined the denoised dataset to only include high-quality pulsatile data for statistical analysis, a procedure we termed the "6 hour protocol". As outlined in section 4.1, we performed feature extraction on 1-minute windows

of data, we therefore oriented our data selection around this window size. We developed a "window selector" tool in Python using Tkinter, enabling the inspection of patient data in 1-minute windows to categorise each as either "good" or "bad" quality, with instances classified as "good" being recorded. Figure 5.1 illustrates the *window selector* interface. This selection process entailed commencing at the beginning of the recording and labelling 6 hours of data per patient as "good", or using all available "good" data if less than 6 hours were classified as "good". ICP changes tend to occur gradually over time due to the slow adaptation of the brain's compartments to alterations in volume and pressure [140]. The period of 6 hours was deemed a reasonable duration which would encompass the natural fluctuations and trends in ICP. This assessment was conducted by the same person across all patients using the distal NIR-PPG data, with the identified "good" instances subsequently being extracted from the both proximal and reference datasets to ensure temporal synchronicity. Although artifacts were present in both the proximal and distal data, the distal data consistently exhibited more severe artifacts during these periods. Therefore, the distal data was selected as the primary reference for guiding data labeling.

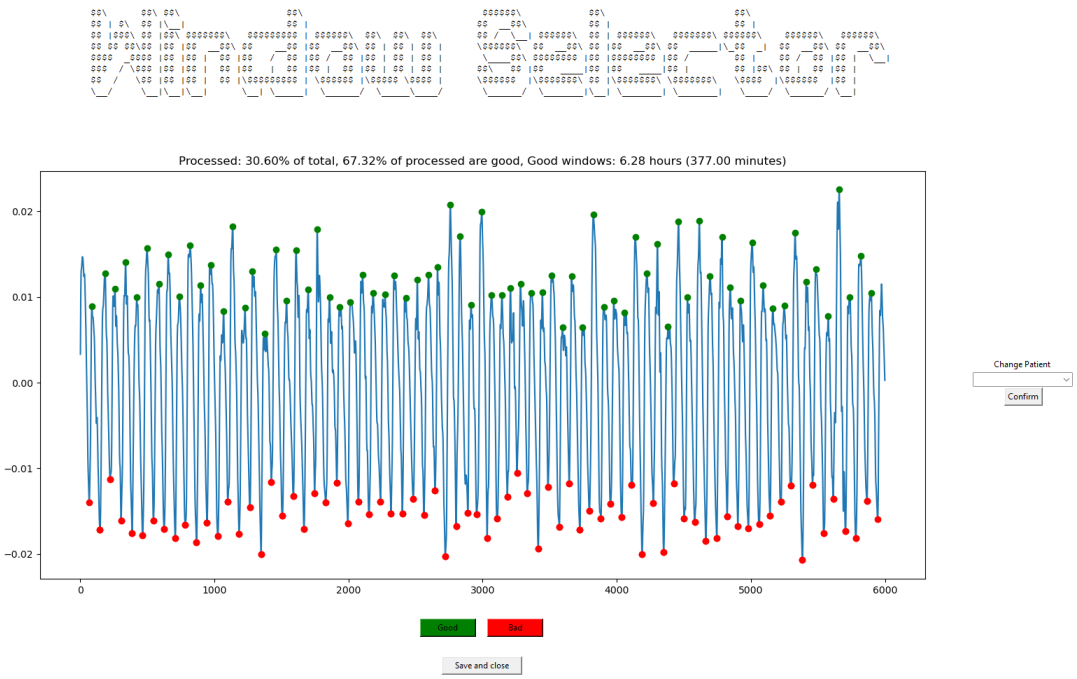

Figure 5.1: A figure showcasing the *window selector* interface, illustrating a one-minute window of distal data. It displays the buttons provided for the user to categorise the data segment as good or bad quality. Additionally, the interface presents an overview at the patient level, indicating the volume of data analysed and the proportion of data classified as good.

5.3 Statistical analysis

The study conducted two types of statistical analyses: correlation analysis and group analysis, to explore the two hypotheses of this study: i) the relationship between alterations in NIR-PPG derived features and changes in ICP levels, and ii) the comparative prominence of these changes in features obtained from long-distance versus short-distance NIR-PPG data. This analysis included data from both proximal and distal photodiodes, referred to as "proximal features" and "distal features", respectively.

To examine the relationship between individual features and the label, the Spearman correlation method was utilised. This choice was made over Pearson correlation as Spearman correlation is capable of identifying linear, non-linear, and non-monotonic relationships without assuming data linearity.

A group analysis was conducted to investigate the null hypothesis that there are no significant differences in feature values across ICP groups. For the group analysis, the study assessed the sampled feature data for normality using the Shapiro-Wilk Test and for homogeneity of variance using Levene's Test. If the data met the necessary assumptions, a parametric, One-way ANOVA was employed. If not, a Kruskal-Wallis non-parametric statistical test was applied.

If a significant difference was detected for any feature in the across group analysis, a post hoc pairwise analysis was conducted to investigate the second null hypothesis that there are no significant differences in feature values between any two ICP groups. The post-hoc, pairwise analysis was conducted between groups using a Mann-Whitney U test with a Bonferroni correction to control for multiple comparisons.

5.3.1 Group construction and data sampling protocol

The group analysis necessitated the formation of groups and the sampling of data for each. Grouping started at an ICP of 0 mmHg, increasing in increments of 10 mmHg. If the highest recorded ICP value was above 20 mmHg and the gap between this maximum value and 20 mmHg exceeded 10 mmHg, an additional group was established, with its upper boundary set at this maximum ICP level.

The criteria for group formation specified that groups must span at least 10 mmHg to facilitate a meaningful comparison of feature variations across different ICP levels. This threshold was based on the premise that meaningful physiological changes are likely to

be observed across ICP intervals exceeding 10 mmHg. Moreover, the larger group size allowed for the inclusion of data from more patients, especially in the underrepresented tails of the ICP dixstribution.

Elevated levels of ICP is referred to as intracranial hypertension which is typically defined as sustained ICP values greater than 20-25 mmHg [141]. Guidelines recommend treating elevated ICP above 22 mmHg [51]. An interest of this study was to examine feature changes across low, normal to moderate and elevated ICP values. In adults, normal ICP levels are typically between 10 to 15 mmHg [18]. The employed group construction methodology establishes reasonable bounds for low, normal to moderate, and high ICP ranges.

ICP dynamics vary between patients, particularly at the clinically scarce elevated ICP values. To strike a balance between the stratified sampling of patients' data and the power of the statistical test being conducted, a "bounded stratified sampling" approach was performed. The size of each group was determined based on the availability and eligibility of patient data within each group. Within each group, the number of data points available per patient was counted. To ensure meaningful statistical power and maintain a reasonable balance of representation between patients, only patients with a minimum of 20 data points (equivalent to 20 minutes of data) were considered eligible for sampling. The maximum number of samples that could be drawn from a single patient was limited to 60 data points, preventing any one patient from being disproportionately over-represented. The size of each group was then calculated by summing the contributions of eligible patients. To maintain consistency, the number of samples in each group was adjusted to match the size of the smallest group, ensuring uniformity in the analysis.

Once a group size had been defined, the sampling procedure involved iteratively and randomly drawing samples from the eligible patient data until the required number of samples for each group was reached. Figures 5.2 and 5.3 depict the number of patients sampled for each group and the number of samples collected from each patient for each group, respectively.

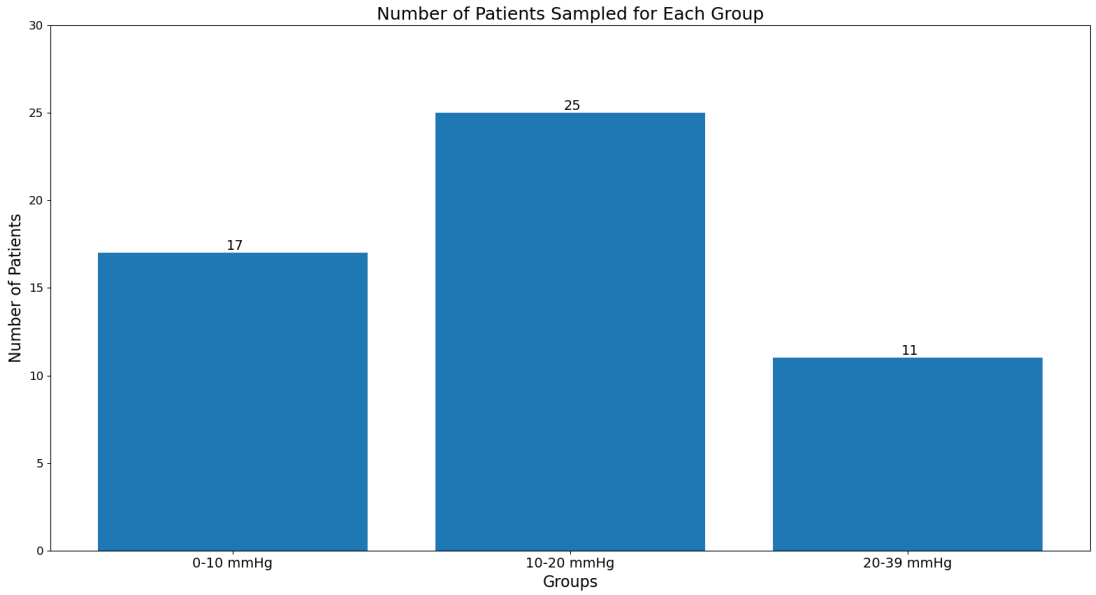

Figure 5.2: This bar chart depicts the number of patients sampled for each intracranial pressure (ICP) group, categorised by the ranges of ICP values: 0-10 mmHg, 10-20 mmHg, and 20-39 mmHg. The chart highlights the distribution of patients within each group, with 17 patients in the 0-10 mmHg group, 25 patients in the 10-20 mmHg group, and 11 patients in the 20-39 mmHg group.

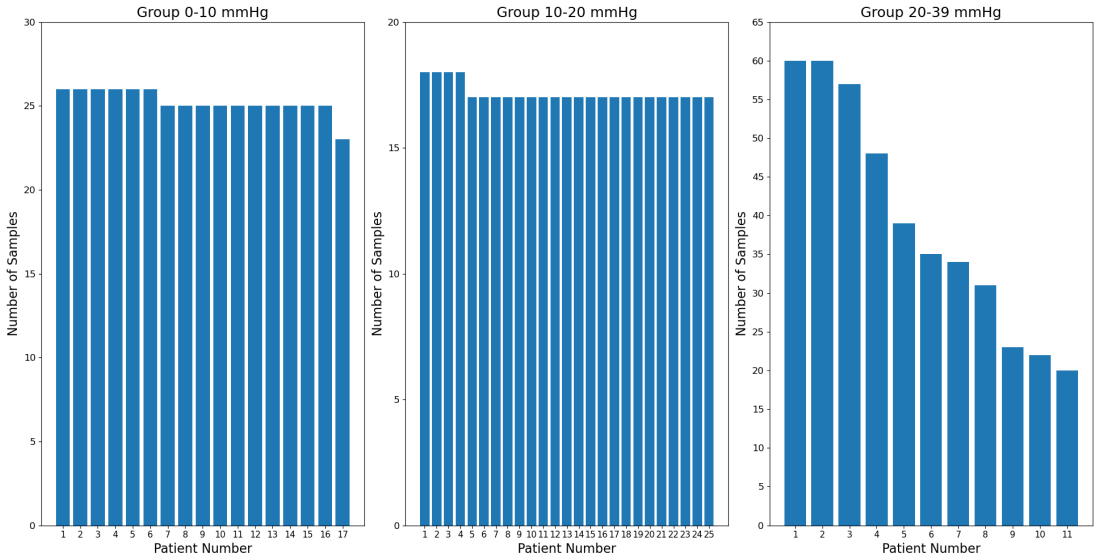

Figure 5.3: A figure containing 3 bar charts depicting the number of samples collected from each patient within the defined ICP groups. The charts illustrate the distribution of samples across individual patients for the groups: 0-10 mmHg, 10-20 mmHg, and 20-39 mmHg.

5.4 Results

The findings affirm the two hypotheses under investigation, which we now examine sequentially. Initially, the outcomes from both the correlation and group analyses reinforce the hypothesis that variations in features derived from NIR-PPG correspond to changes in ICP levels. The correlation analysis revealed that the majority of the 141 features exhibited significant correlations ($p < 0.05$) with the label in both datasets. Notably, 109 (77.30%) and 112 (79.43%) of the features demonstrated significant correlations with the label for proximal and distal features, respectively. Table 5.1 presents the ten features with the smallest p-values for both proximal and distal feature sets. Due to the extensive number of features analysed, this discussion is concentrated on the top 10 most significant features, alongside aggregate metrics summarising the overall results.

PROXIMAL		**DISTAL**	
Feature	**Coefficient**	**Feature**	**Coefficient**
start datum difference	-0.1943	ds ratio 75 (deriv 1)	-0.2149
max start datum difference	-0.1699	kurt	-0.2132
skew	0.1527	systolic width 50	-0.2064
systolic width 25 (deriv 2)	0.1499	diastolic width 75 (deriv 1)	-0.2051
ds ratio 75 (deriv 1)	-0.1472	ds ratio 50 (deriv 1)	-0.2051
pulse width 25 (deriv 2)	0.1466	skew	0.2038
kurt	-0.1337	ds ratio 75	0.2033
pulse width 50 (deriv 2)	0.1319	skew (deriv 1)	0.2003
ds ratio 75 (deriv 2)	-0.1276	pulse width 75 (deriv 1)	-0.2001
systolic width 25	-0.1275	ds ratio 50	0.1941

Table 5.1: Table containing the top 10 correlated features sorted by ascending p-value for both short-distance, proximal and long-distance, distal NIR-PPG derived features. The Coefficient column represents the Spearman correlation coefficient, which ranges from -1 to 1. A value of 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation. Larger absolute values signify stronger relationships between features and the label.

For the significant features, the mean absolute correlation coefficients were 0.0729 and 0.0930 for proximal and distal features, respectively. Within the subset of the 10 most correlated features, these averages increase to 0.1481 for proximal and 0.2046 for distal

features. Excluding seven features (start datum difference, max start datum difference, skew, skew of the first derivative, and kurtosis) from the top 20 features across both proximal and distal datasets the remaining 65% of the features are related to pulse width.

The group analysis reinforces the findings of the correlation analysis and provides additional support for the initial hypothesis. The sampled data for all features did not meet the assumptions of normality and/or homogeneity of variance, as assessed by the Shapiro-Wilk and Levene's tests, respectively. Consequently a non-parametric Kruskal-Wallis test was employed to identify changes in the distributions of feature values across the ICP range. Figure 5.4 illustrates the distribution of ICP values within the dataset.

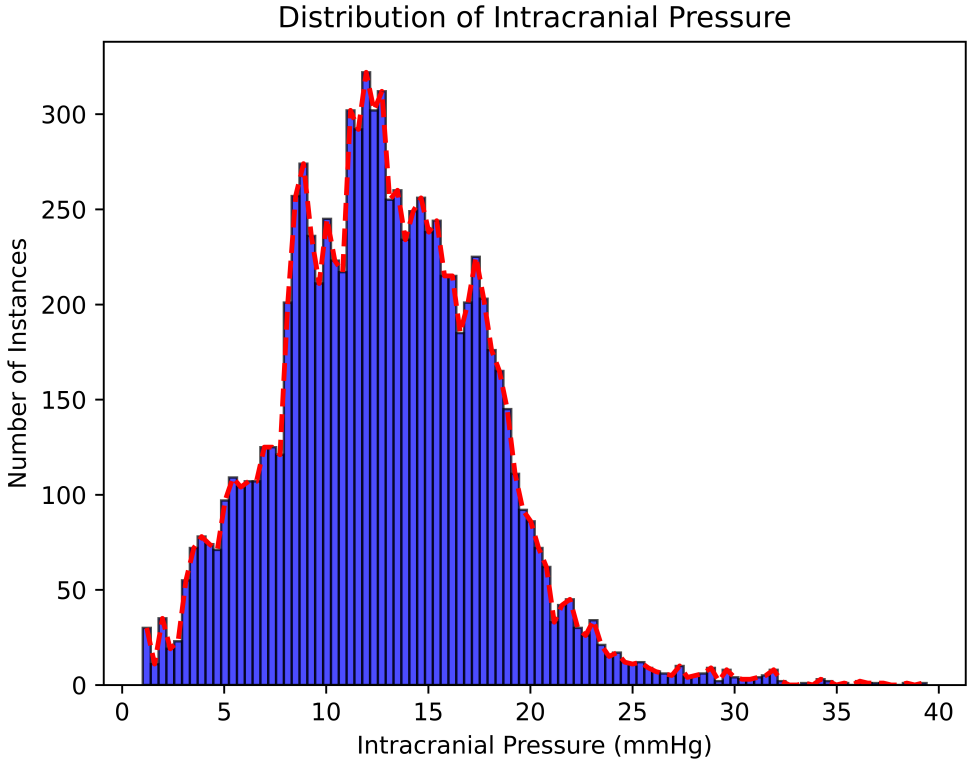

Figure 5.4: A histogram illustrating the distribution of intracranial pressure values within the dataset.

The dataset's maximum ICP value was 39 mmHg. Following the logic for constructing groups, three groups were formed: (0-10), (10-20), and (20-39) mmHg. The smallest group, encompassing data from 20 to 39 mmHg, comprised 429 instances, equating to approximately 7.15 hours of data. To ensure uniformity, the other groups were sampled to match this size. The results of the group analysis indicate that 115 (81.56%) and 107 (75.89%) features exhibited significant changes ($p < 0.05$) across ICP groups for proximal and distal features, respectively. Table 5.2 presents the top 10 features sorted by lowest p-value for both proximal and distal datasets.

PROXIMAL			DISTAL		
Feature	**P-Value**	η^2	**Feature**	**P-Value**	η^2
skew (deriv 1)	7.14E-37	0.128	skew (deriv 1)	1.04E-32	0.113
skew	7.51E-30	0.103	ds ratio 75 (deriv 1)	3.51E-30	0.104
ds ratio 75 (deriv 2)	7.20E-29	0.099	ds ratio 50 (deriv 1)	1.08E-28	0.099
ds ratio 50	9.15E-28	0.095	skew	6.61E-28	0.096
ds ratio 75	1.01E-25	0.088	diastolic width 75 (deriv 1)	1.15E-27	0.095
kurt	1.47E-25	0.088	kurt	1.09E-25	0.088
AUC Ratio	1.22E-24	0.084	diastolic width 50 (deriv 1)	1.30E-25	0.088
datum area ratio	8.40E-24	0.081	ds ratio 100	1.53E-25	0.087
end datum difference (deriv 1)	1.34E-23	0.080	pulse width 75 (deriv 1)	2.00E-25	0.087
ds ratio 100	5.76E-23	0.078	ds ratio 75	1.39E-24	0.084

Table 5.2: Table containing the top 10 features which changed most significantly with variations in ICP values. The NIR-PPG derived features were sorted by ascending p-value for both short-distance, proximal and long-distance, distal data. The associated p-value and η^2 for each feature is also presented.

Figure 5.5 illustrates the top 10 distal features arranged by descending p-value returned by the Kruskal-Wallis analysis. The presented box plots provide a comparative analysis of feature values across ICP ranges for proximal (orange box plots) and distal (blue box plots) data. The presented box plots provide a comparative analysis of feature values across ICP ranges for proximal (orange box plots) and distal (blue box plots) data. Each box plot represents the distribution of a specific feature within a given ICP range. The boxes illustrate the interquartile range (IQR), containing the middle 50% of the data points. Inside each box, the red line indicates the median of the data, providing a measure of central tendency. The whiskers denote the range within 1.5 times the IQR from the quartiles. The box plots include annotations indicating the results of the overall group Kruskal-Wallis tests, which compare the distributions across different ICP groups. Additionally, the associated results of the post hoc pairwise analysis are annotated, indicating which pairs showed significant changes.

A visual examination of the boxplots indicate that the features derived from NIR-PPG data show statistically significant variations with changes in ICP. Across ICP groups, most of the examined features demonstrate significant changes in median values and IQR.

Notably, the majority (approximately 70%) of the top 10 most significant features exhibit an inverse relationship with ICP, indicating that as ICP increases, the values of these features tend to decrease.

Focusing specifically on features within the top 10 that directly describe aspects of pulse width without referring to ratios of systolic to diastolic pulse widths, we observe consistent trends. Features such as the pulse width at 75% of pulse prominence (1st derivative), diastolic width at 50% of pulse prominence (1st derivative), and diastolic width at 50% and 75% of pulse prominence (1st derivative) all display a robust inverse relationship with ICP. These findings suggest that as ICP values rise, the width of the cardiac pulse decreases. Notably, the diastolic width emerges as one of the features undergoing the most significant changes, implying that the reduction in width may primarily stem from a decrease in the diastolic segment of the pulse.

Furthermore, along with the narrowing of width, a significant inverse correlation between skewness, kurtosis, and the ratio of diastolic to systolic width at different percentages of pulse prominence with ICP is observed.

Furthermore, along with the narrowing of pulse width, a significant relationship between skewness, kurtosis, and the ratio of diastolic to systolic width at different percentages of pulse prominence with ICP is observed. The observed decrease in kurtosis, increase in skew, and decrease in the ratio between the systolic and diastolic pulse width of the pulse may collectively indicate that the pulse is becoming more rounded. A decrease in kurtosis suggests that the pulse distribution is becoming less peaked and more flattened, implying a broader, more rounded shape. An increase in skew indicates a shifting or diminishing of the systolic peak, suggesting that the peak is either becoming less pronounced or moving away from the left hand side of the pulse, contributing to a more symetrical and rounded pulse shape. Additionally, a decrease in the ratio between the systolic and diastolic pulse indicates that the difference between these two phases is reducing, further supporting the idea that the pulse profile is losing its sharp systolic peak and becoming more uniformly rounded.

The majority (60%) of the top 10 features of the distal dataset which we posit includes "cerebral" information were first derivative features. The first derivative representing the velocity/rate of change over time. Within this context the first derivative may provide information about the velocity of blood volume changes happening within the cerebral vasculature.

A post hoc pairwise analysis was conducted using a Mann-Whitney U test with a Bonferroni correction if a significant change was detected across groups from the Kruskal-Wallis test. This approach led to the post hoc analysis of 115 and 107 features for the proximal and distal data, respectively.

In the proximal dataset, 80.87% and 86.09% of the features changed significantly between the 0-10 and 20-39 mmHg groups and the 10-20 and 20-39 mmHg groups, respectively. Similarly, in the distal dataset, 85.047% of features changed significantly between these same pairs. A minority of pairwise statistical changes were identified between the 0-10 mmHg and 10-20 mmHg groups. In the proximal and distal datasets, these changes accounted for 40% and 21.5%, respectively. These findings align with our expectation that the majority of features would show significant differences when comparing groups with low ICP (0-10 mmHg) and normal/moderate ICP (10-20 mmHg) to those with high ICP (20-39 mmHg).

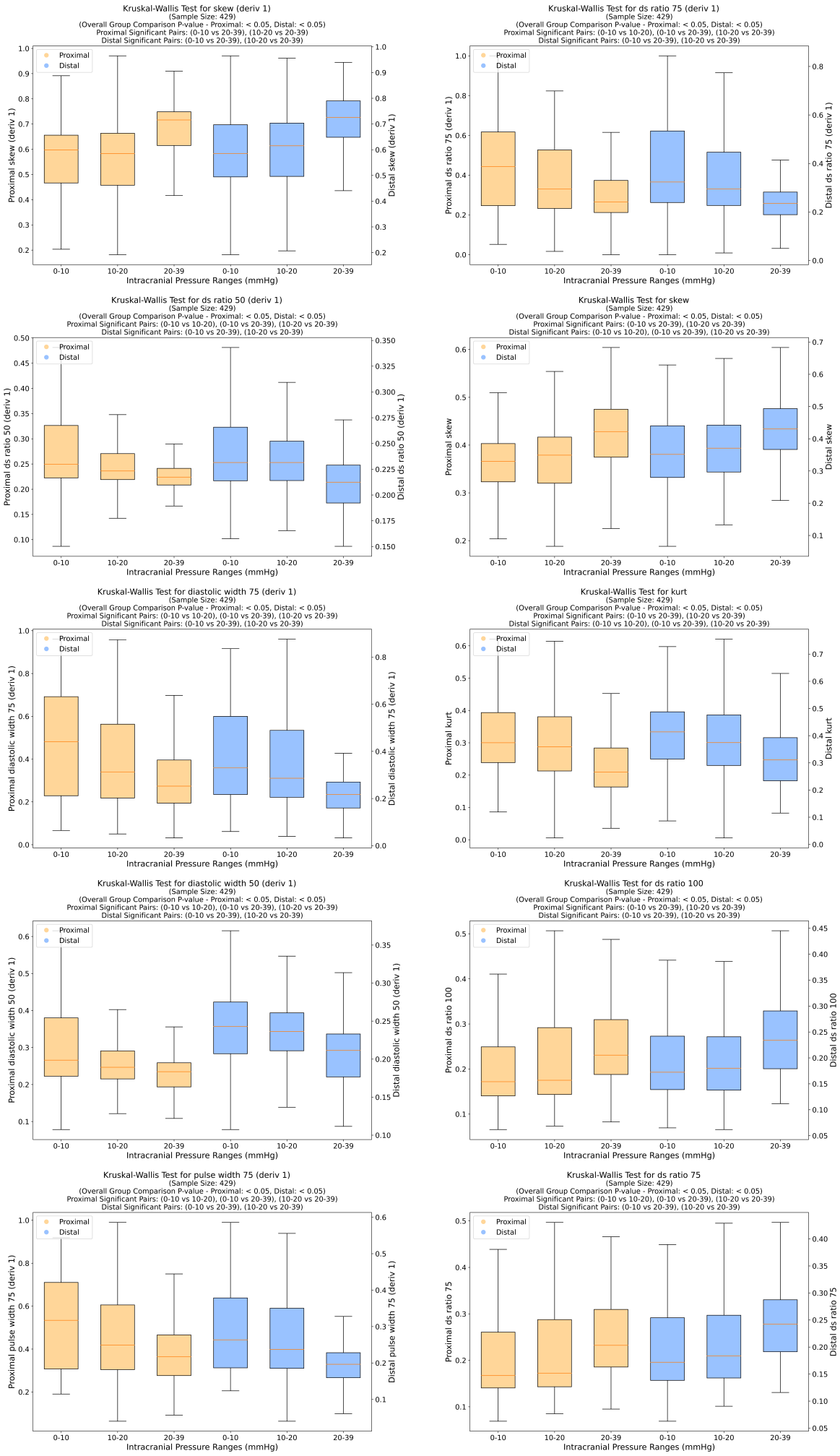

Figure 5.5: Boxplots illustrating the the variation in feature values of the 10 features with the most significant differences in a Kruskal-Wallis analysis on distal feature data. They display distributions for both proximal and distal feature data. The central line represents the median, the box encompasses the interquartile range (IQR), and the whiskers denote the range within 1.5 times the IQR from the quartiles.

The findings partially validate the second hypothesis, indicating that alterations in features are more prominent in long-distance NIR-PPG data compared to short-distance NIR-PPG data. Although there is a non-significant difference in the number of significantly correlated features (2.75%) and significant differences across groups (7.21%) between the proximal and distal datasets, the correlation analysis shows that the absolute average correlation coefficient of all features and significantly correlated features in the distal dataset is 25.76% and 24.24% higher, respectively, than in the proximal dataset. This suggests that distal, long-distance NIR-PPG pulse features may be more closely related to changes in ICP compared to proximal, short-distance features.

5.5 Discussion

To the best of the authors' knowledge, this paper represents the first attempt to investigate variations in time-series and morphological features derived from NIR-PPG in response to changes in ICP, using clinically collected data from 27 patients. The initial hypotheses of this study were twofold: first, to ascertain whether alterations in NIR-PPG derived features correlate with fluctuations in ICP levels, and second, to determine whether these changes are more pronounced in features derived from long-distance NIR-PPG data compared to short-distance NIR-PPG data. The results of this investigation provide support for both hypotheses. The findings indicate that a majority of features derived from both short and long-distance NIR-PPG data exhibit significant changes in response to variations in ICP. Although there is a non-significant difference in the number of significantly correlated features and significant differences across groups between the proximal and distal datasets the distal dataset yielded a meaningfully higher absolute average correlation coefficient of all features and significantly correlated features in-comparison to the proximal dataset which supports the proposition by Roldan et al. that data from the proximal photodiode corresponds predominantly to extracerebral data, whereas data from the distal photodiode represents a combination of extracerebral and cerebral data [101].

In section 2.4.4, we introduced our hypothesis regarding the relationship between cerebral hemodynamics and changes in NIR-PPG morphology. We describe how elevations in ICP correspond to compression of cerebral vessels altering their mechanical properties, causing a reduction in cerebral compliance, increase in vascular resistance and subsequent reductions in CPP. This process initiates cerebral autoregulatory mechanisms aimed at

maintaining CPP and normalising CBF to the brain. However, when elevated ICP diminishes CPP below a critical threshold, cerebral autoregulation may fail, resulting in blood vessels being unable to vasodilate sufficiently to counter the reduced CPP, leading to a sustained drop in CBF. We contend that the NIR-PPG signal may offer insights into these variations in cerebral compliance and CBF. Our findings indicate that the majority of the most significant changes in distal features affected the pulse width. We hypothesise that the recorded changes in pulse width may reflect alterations in blood flow and vessel elasticity. As ICP increases, cerebral compliance decreases reducing CPP below the critical threshold and consequently diminishing CBF, less blood volume per cardiac cycle reaches the brain. We acknowledge that the normalisation of the NIR-PPG data prior to analysis does not enable us to determine if the changes in pulse width are as a consequence of shorter cycle duration or changes in amplitude. Despite this, we hypothesise a narrower pulse width may signify a shorter duration of blood volume change, potentially associated with decreased compliance or lower CBF. Decreased cerebral compliance may result in an increase in the velocity of blood volume changes due to the vessels' reduced ability to accommodate changes in blood volume, reducing the first derivative's pulse width.

The results also suggest that as ICP increases, pulses become more rounded, indicated by a decrease in kurtosis and the ratio of diastolic to systolic width at different percentages of pulse prominence coupled with an increase in skewness. This rounding of the pulse may also be associated with a dampening effect caused by reduced compliance. We posit that the diminished ability of vessels to deform during systole and diastole with decreased compliance attenuates the pulse wave creating the rounding effect.

The study conducted by Cardoso et al [142] analysed the CBF pulse wave and its relationship with ICP in 15 patients suffering from hydrocephalus, benign intracranial hypertension or head injury. Our findings and hypothesis are in line with the outcomes of this study which found that increases in ICP is accompanied by an increase in the amplitude of the cerebrospinal fluid pulse wave which in values above 20 mmHg first become rounded and, at higher ICP values acquires a pyramidal shape.

Additionally, the majority of the most significant distal feature changes were observed in the first derivative, particularly within the distal dataset, which we believe encompasses "cerebral" information. The first derivative, representing flow velocity. An increase in ICP can compress cerebral vessels, thereby diminishing CBF. This constriction not only reduces the volume of blood that can flow through these vessels but may also affect the

velocity of blood flow, likely captured by changes in the first derivative of the NIR-PPG signal. Consequently, a reduction in the width of the first derivative may signify a faster transit time of blood perhaps indicating compensatory mechanisms reacting to maintain cerebral perfusion under elevated ICP. Understanding these changes in features related to blood flow velocity appear crucial in describing ICP dynamics via NIR-PPG data.

While acknowledging the preliminary nature of these findings, we believe that they offer a credible foundation for further exploration of the dynamic interplay between ICP and NIR-PPG morphology.

5.6 Summary

The results of this study support both hypotheses, demonstrating significant correlations between the extracted features and ICP levels. Significantly 77.30% and 79.43% of the features exhibited significant correlation with the label for the distal and proximal datasets, respectively. Moreover, the Kruskal-Wallis group analysis underscores the consistency of these findings, indicating that 81.56% and 75.89% of the features showcased significant changes across the ICP range. Notably, the mean absolute correlation of significant long-distance derived features surpassed short-distance features by 24.4%. These findings underscore the potential of NIR-PPG based devices as non-invasive tools for estimating dynamic changes in ICP. These findings are clinically relevant and support the work within this domain which strives to utilise NIR-PPG data and derived features to estimate ICP non-invasively providing a reference for demonstrating the sensitivity of NIR-PPG waveform features to alterations in ICP. The positive findings indicate promise in the area of NIR-PPG driven non-invasive ICP monitoring.

Chapter 6

Classical Machine Learning Algorithms

Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed. A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E [143]. One of the broad methods of grouping machine learning approaches is whether or not they are trained with supervision. The type of supervision can be broken into two major categories: (i) supervised and (ii) unsupervised learning. In supervised learning, the algorithm also referred to as the "model" is trained on labeled data, meaning the desired outputs are predefined in the dataset, allowing the model to learn a function that maps input variables (independent variables) to the desired outputs (dependent variables). Conversely, unsupervised learning involves unlabeled data, where the model must identify patterns and derive new representations without supervision.

This research is predicated upon a supervised learning dataset. As mentioned in section 4.1 the NIR-PPG data was collected synchronously with reference ICP data. The dataset was segmented into 1-minute windows and the median ICP value over each window was used as a label. For each window of data 141 feature were extracted and the feature representation for each window was determined by computing the median of each feature across all pulses within the window. In this research, each 141 input features is associated with an output label which is the median ICP value over the same duration.

Supervised learning addresses either classification or regression tasks. Classification

involves predicting discrete outcomes. In contrast, regression aims to predict continuous values. In this project, the ICP values are continuous, defining the task as a regression problem under supervised learning. In this chapter the classical machine learning regression models applied to predict ICP values are introduced and explained.

6.1 K-Nearest Neighbours (KNN)

K-Nearest Neighbours (KNN) [144] regression is a non-parametric method that, approximates the association between input features by looking directly at the observed, training data. When a KNN model is defined, there a number of key parameters, including: number of neighbours, weighting method, search algorithm and distance metric. These parameters define how many neighbours will be used to make predictions, how their influence is weighted, how neighbours are searched for and how distance between points is measured.

Unlike other models, the KNN model doesn't learn a mapping function from input to output. Instead, the training data is stored in a structured way to enable the efficient searching for neighbours for a given input feature set. KNN calculates the distance between the input features and those of every training instance, using the chosen distance metric in order to find the "nearest" instances of the training data referred to as neighbours. A common distance metric is the Euclidean distance which measures the straight-line distance between two points in Euclidean space. The formula for Euclidean distance between two points A and B, where A has coordinates $(a_1, a_2, \ldots, a_n)$ and B has coordinates $(b_1, b_2, \ldots, b_n)$ in an n-dimensional space is given by:

$$\mathrm{d} = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \ldots + (a_n - b_n)^2} \tag{6.1}$$

Based on the derived distances, the model identifies the n number of nearest neighbours in the training dataset to the new data point. Once the n nearest neighbours are identified, the prediction for the new input is then based on these neighbours' output values. The derived output for the new input values is dependent upon the weighting method. Usually the weighting method is either "uniform" or "distance" based. The final predicted output is an average of the neighbours' output value. If the weighting method is uniform, each neighbour's output value contributes equally to the final prediction. If it is distance based

the contributions of the neighbours are weighted by the inverse of their distance to the new input, giving more influence to closer neighbours.

A significant advantage of KNN is its adaptability to new data, as it requires no explicit training. However, this also makes it computationally demanding and memory-intensive, especially as the size of the dataset grows, since it must compute distances to all training samples for each new instance. Additionally, KNN struggles with the "curse of dimensionality", where high-dimensional feature spaces lead to data sparsity. In such sparse spaces, it becomes challenging for the KNN algorithm to find close and relevant neighbours as all points appear almost equally distant from each other leading to the loss of the concept of "nearest" which subsequently affects the effectiveness of the algorithm.

6.2 Support Vector Regression

Support Vector Regression (SVR) [145] extends upon the principles of Support Vector Machines (SVM). SVR aims to fit the best possible function within a tolerance level from the actual data points. The SVR algorithm uses kernel functions to project training data into a higher dimensional transformed feature space in which linear regression can be applied to solve non-linear problems. Among these kernels, the Radial Basis Function (RBF) kernel is common for non-linear datasets. In this research, the SVR model was implemented using an RBF kernel function.

Assuming the use of a RBF kernel, when initialising an SVR model, there are three key parameters which are typically defined: (i) C, (ii) ϵ and (iii) γ. The parameter C is a regularisation parameter which balances the trade-off between achieving a low error on the training data and maintaining a simple model to avoid overfitting. A higher C value tries to fit the training data as well as possible (at the risk of overfitting), while a lower C pushes for a less complex model. The epsilon parameter (ϵ) sets a margin of tolerance where predictions within this margin are not penalised. The gamma (γ) influences the curvature of the decision boundary. A high γ value creates a more complex, tightly fitting model, while a low γ results in a smoother, broader fit.

The SVR algorithm can handle both linear and non-linear data. In order to handle non-linear tasks and to fit non-linear decision boundaries a kernel function (such as RBF) enable the SVR to operate as if the data has been transformed into a higher-dimensional space without explicitly performing the transformation. This "kernel trick" is performed

by calculating the similarity between pairs of data points in the original input space allowing the algorithm to find a linear separating hyperplane in the new higher-dimensional space representation, even though the data in the original space is not linearly separable. The RBF kernel function is defined as:

$$\mathrm{K}(x_i, x_j) = e^{-\gamma \|x_i - x_j\|^2} \tag{6.2}$$

Where $K(x_i, x_j)$ represents the kernel function's output applied to pairs of input vectors x_i and x_j. The parameter γ affects how quickly the similarity measure decreases as the distance between the points increases. A larger γ makes the similarity drop off faster with distance, meaning the model pays more attention to smaller differences. The Euclidean distance between points x_i and x_j is represented as $(\|x_i - x_j\|)$. The exponential function transforms the distance into a similarity score. By using the exponential function, distances are converted such that closer points have a high similarity score (close to 1), and points further away have a decreasing similarity score (close to 0).

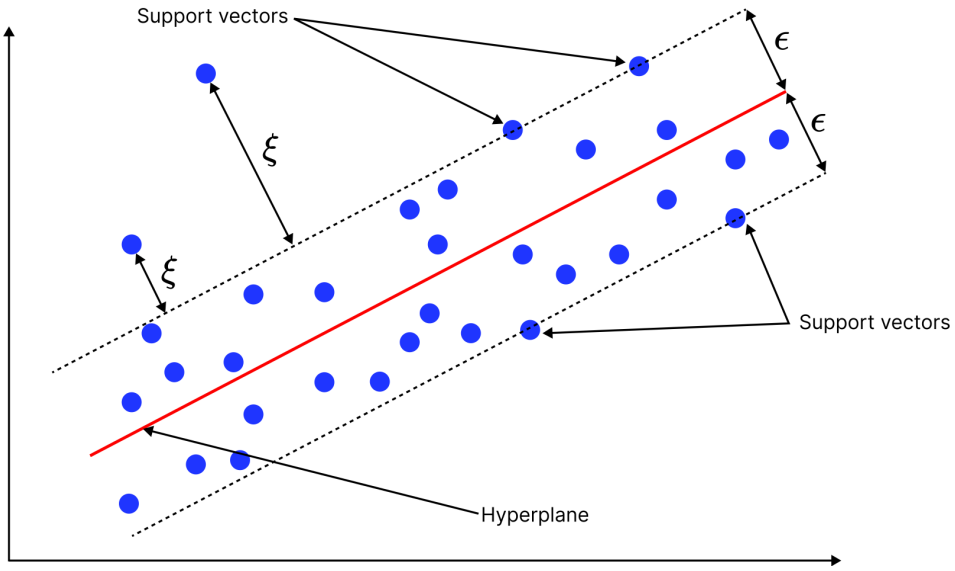

Figure 6.1: A diagram of Support Vector Regression where the data points are plotted in space. The hyperplane (in red) best fits the data with a certain margin (epsilon, ϵ), and the support vectors are the data points closest to the hyperplane. The points outside of the margin are marked with variables (ξ), indicating the tolerance for errors in the model.

Figure 6.1 illustrates the basic functioning of the SVR algorithm. The red line represents the optimal hyperplane, which is the function that SVR has learnt to predict the label. On either side of the hyperplane, represented as two dashed lines is the margin of tolerance defined by the ϵ parameter. The data points, that fall outside of these margins

are considered as errors and are associated with a cost, proportionate to their distance from the margin indicated as ξ. The data points that fall on the margin or beyond contribute to the determination of the hyperplane and are termed support vectors.

SVR, with the use of the kernel trick and the ability to choose different kernels, offers a flexible approach to regression. It can handle complex, non-linear relationships without explicitly transforming the data, which saves computational resources. However SVR can be computationally intensive, which limits its scalability and practicality for use with large datasets.

6.3 Random Forest

Random Forest (RF) [146] is an ensemble learning method which builds multiple decision trees and combines the predictions of each tree to produce a more robust overall prediction and mitigate overfitting. Each tree in the "forest" is created using a random subset of the training data, referred to as a bootstrap sample. This randomness helps in building diverse trees which aims to produce more reliable and stable overall predictions.

The construction of each decision tree in a RF begins at the root node. Here, the algorithm chooses the best feature (the features that best describes the data points, such as pulse amplitude or width) to split the data into groups that are as homogeneous as possible within each group and as heterogeneous as possible between groups. This splitting is done to organise the data into "branches", which lead to more nodes. At each node, this splitting process is repeated until a terminal node or criteria is reached. This could be a defined maximum number of splits allowed or when further splitting is no longer meaningful.

Each end of the branch, known as a leaf node, represents a group of training instances that are similar to each other. When a prediction is being made on a new data point, it travels down the tree branches, dictated by the features it exhibits, until it reaches a leaf. The prediction for this new instance is then made by averaging the label values of the training instances within that leaf.

The overall performance and behavior of a Random Forest can be adjusted using various parameters. These parameters include, the "maximum depth" of the trees, which prevents them from becoming overly complex and learning noise from the training data. The number of "estimators" which controls how many trees the forest has. More trees

generally lead to better performance because errors in individual trees are likely to be averaged out. However, having many trees can make the model computationally expensive. Lastly, parameters that define the minimum number of samples required to split the data further or to define a leaf node help in ensuring that each decision made by the tree captures enough information to be relevant whilst avoiding overly specific rules that prevent model generalisation. These parameters help in achieving a balance, ensuring the model is neither too simple nor too complex.

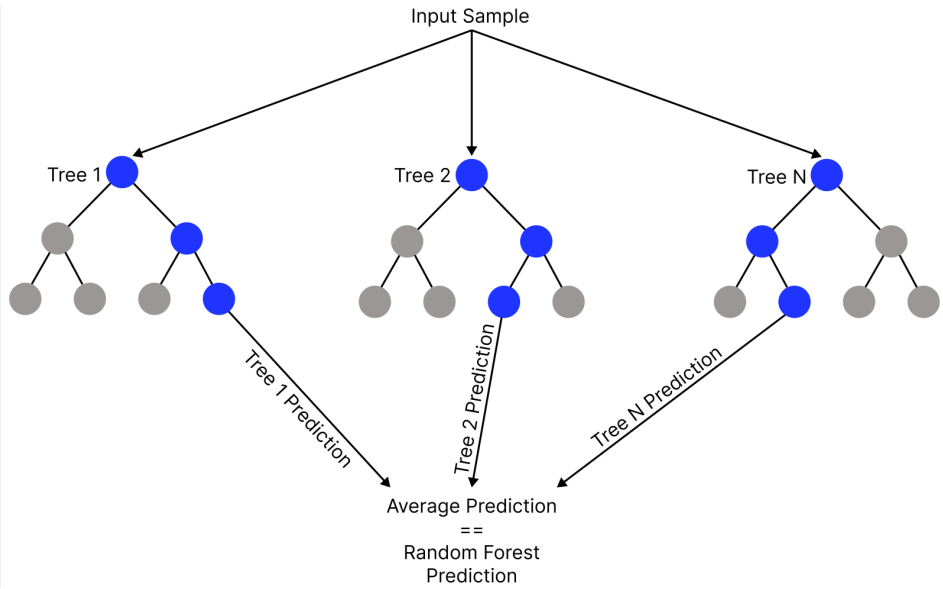

Figure 6.2: An illustration of a Random Forest model showing how an input sample is processed through multiple decision trees (Tree 1 to Tree N). Each tree independently makes a prediction, and the final output is the average of these predictions, forming the Random Forest regression result.

As illustrated in Figure 6.2, when a new input instance is introduced to the RF model, it is evaluated by each tree in the "forest". Each tree independently provides a prediction and the final prediction from the model is derived by averaging all the predictions from individual trees. This method of averaging acts to smooth out the variations among the predictions but also offers a more generalised estimate, avoiding the dependence on a single decision tree.

The RF mitigates the risk of overfitting and increases model robustness by averaging predictions across multiple trees. Additionally, it can quantify the importance of different features in predicting the target variable and does not assume a specific data distribution allowing it to handle both linear and nonlinear datasets. Despite these strengths, the RF algorithm due to its structural complexity can lead to significant computational demands.

6.4 Gradient Boosting Machines

Gradient Boosting Machines (GBM) [147] are ensemble machine learning models that, like Random Forests, use decision trees to make predictions. However, GBMs construct their ensemble sequentially, with each new tree focused on correcting the errors made by the previous ones. This approach is based on the principle of "boosting", wherein multiple simple models (referred to as "weak learners") usually shallow decision trees are combined to create an accurate ensemble model.

The process of gradient boosting is centered around three main components: a loss function (which in the case of regression is usually MSE) that the algorithm aims to minimise, weak learners that individually make predictions, and an additive approach that incrementally builds the model by incorporating one learner at a time, each addition aims to reduce the overall prediction error.

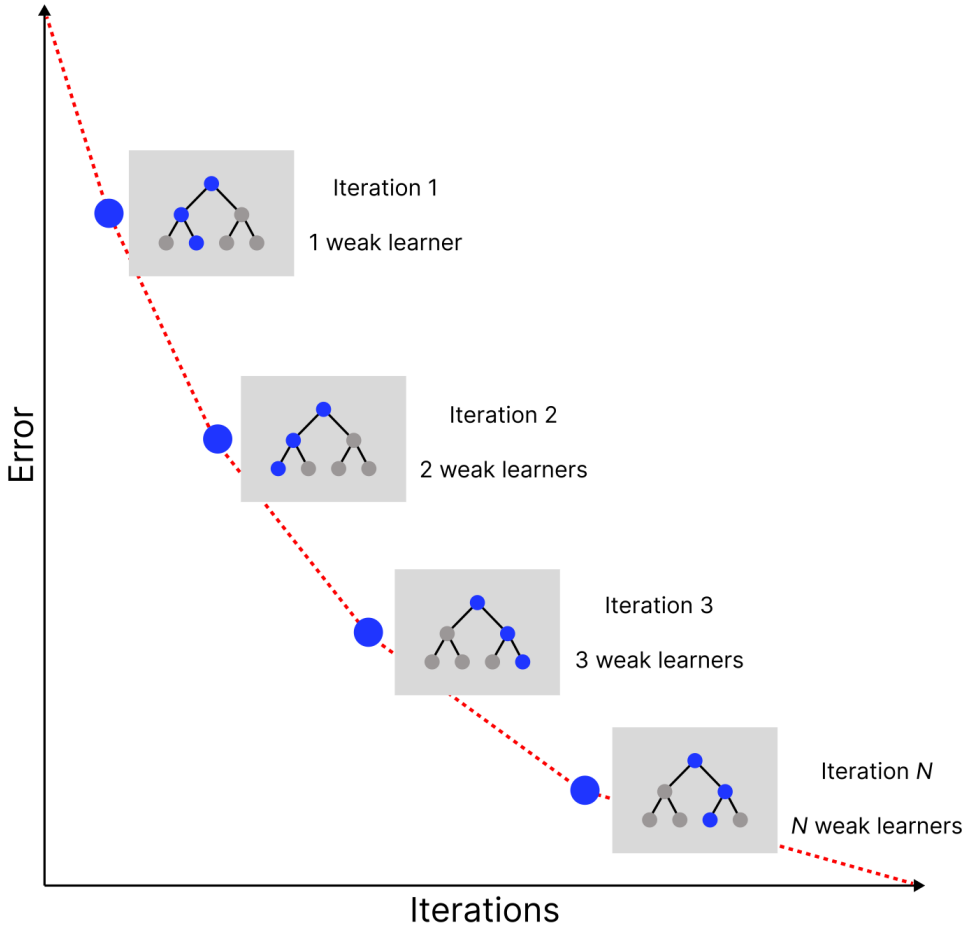

Figure 6.3: A graph illustrating the iterative process of boosting in Gradient Boosting Machines. Each point represents a stage of the algorithm, where the addition of a weak learner aims to reduce the overall prediction error. The error decreases with each iteration, as more weak learners are added, demonstrating the model's improvement over time.

As the model progresses, each new tree focuses on the residual errors (the differences between the predicted and actual values). These errors guide the training of each new tree, ensuring that the ensemble pays more attention to the harder-to-predict instances of data. The final prediction for any given instance is the sum of the predictions from all individual trees. As depicted in Figure 6.3 the theoretical aim of GBMs is that as the algorithm adds trees, each one adjusts according to the gradient of the loss function minimising the error as the number of trees increases.

Gradient boosting has been extended into more specialized forms with Light Gradient Boosting Machine (LGBM) and Extreme Gradient Boosting (XGB), each introducing unique features that enhance performance and efficiency.

In both LGBM and XGB, several parameters play a key role in tuning the models for optimal performance. Among others these include the maximum tree depth, the number of trees and the learning rate. Similarly to the RF algorithm, the maximum depth of each tree in the model can be set to controls the complexity of the decision boundaries. A deeper tree can model more intricate patterns but risks overfitting if too deep. Number of iterations or estimators, refers to the the number of trees used in the model and indicates how many times the boosting process is repeated. More iterations allow the model to continually improve its accuracy at the risk of overfitting if too many trees are used. Learning rate affects the contribution of each tree to the final model. A lower learning rate requires more trees to converge to a robust model, typically leading to a more gradual but stable learning process. Each of these parameters helps manage the trade-off between model complexity and overfitting.

6.4.1 Light Gradient Boosting Machine

The LGBM algorithm [148] aims to enhance GBM by implementing two main optimisations: (i) Gradient-based One-Side Sampling (GOSS) and (ii) Exclusive Feature Bundling (EFB). The GOSS sampling strategy selectively samples instances with large errors during training to optimise memory usage and focus model training on more challenging instances of data. EFB groups features that are exclusive or nearly exclusive. This bundling reduces the feature dimensionality thus lowering the computational complexity, allowing the algorithm to run faster and more efficiently.

These enhancements make LGBM particularly suited for handling large datasets and

complex feature spaces more efficiently than traditional GBM.

6.4.2 eXtreme Gradient Boosting

XGB [149] is another optimised implementation of gradient boosted trees. Firstly, XGB implements parallel processing, which enables the ability to build trees concurrently during training, rather than sequentially. This parallelisation can be leveraged using multi-core computing architectures, significantly increasing the training speed of the model.

Secondly, XGB includes a regularisation in its objective function helping to mitigate overfitting by penalising overly complex models aiding in the models ability to generalise to unseen data.

6.5 Summary

This chapter provides a foundational overview of the classical machine learning models that have been used within this research to predict ICP values using features derived from NIR-PPG data. The basic functionality of each model has been outlined and their respective strengths and weaknesses discussed. Key parameters crucial for tuning these models have also been detailed. This introduction is important as the model optimisation, evaluation, and the analysis of results is presented in the the forthcoming chapter.

Chapter 7

Evaluation of Machine Learning Models

Following the introduction of the classical machine learning models used within this research. This chapter introduces the methodology of the hyperparameter optimisation and model training and evaluation in addition to an analysis of results The work outlined in this chapter culminated in a publication that describes the implementation, optimisation, and evaluation of the models [150].

7.1 Data partitioning

As depicted in Figure 7.1 the proximal and distal NIR-PPG datasets were both segmented into two distinct datasets: (i) *evaluation* and (ii) *optimisation*. In order to create the optimisation and evaluation datasets, the aggregate dataset of both the proximal and distal NIR-PPG data was split by patient in a ratio of 70:30. Patients were randomly sampled, the patient data of 19 patients and 8 patients made up the evaluation and optimisation sets respectively. In order to effectively evaluate the models' ability for continuous monitoring the temporal order of the windows was maintained. The data of the same 19 and 8 patients comprised the evaluation and optimisation datasets for both the proximal and distal data.

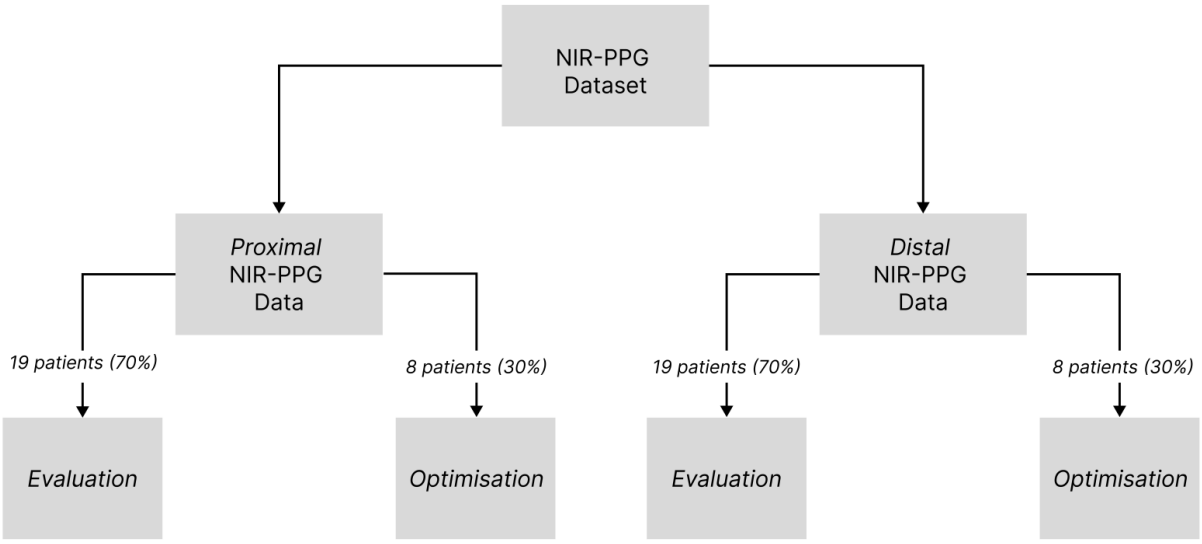

Figure 7.1: Partitioning of NIR-PPG Dataset for model optimisation and evaluation. The complete NIR-PPG dataset was divided into two subsets for both proximal and distal data: one for optimisation and the other for evaluation. The data was split according to a 30:70 ratio by patient count, with 8 patients' data allocated for optimisation and 19 for evaluation. This partitioning was consistent across both proximal and distal datasets to maintain comparability.

7.2 Feature selection

7.2.1 Collinearity test and feature elimination

A significant amount of the features within the feature sets were colinearly related due to being derived from similar or the same morphological elements of the pulse. In order to perform feature selection whilst dealing with the presence of colinearly related features a three step feature selection process was performed. Initially the maximum information coefficient (MIC) was calculated for each feature against the label. MIC captures linear, non-linear and non-monotonic relationships offering a more precise representation of complex dependencies between features and the label.

To address the issue of collinearity among variables, pairs of features were identified which exhibited a Pearson correlation coefficient ≥ 0.9. Within each of these pairs, the feature with the lower MIC score was removed, prioritising the retention of features with higher potential predictive power.

7.2.2 Feature selection via Recursive Feature Elimination

Following the identification and removal of colinearly related features, Recursive Feature Elimination with Cross-Validation (RFECV) was implemented to further refine the selection of predictive features. When conducting the RFECV a group-wise holdout method was employed in the cross-validation. This approach involves dividing the data into distinct groups where each group corresponds to a single patient's data. During each iteration of the RFECV process, entire groups (patients) are held out from the training set and used for validation. This method prevents data leakage, as the model is trained and validated on entirely separate patient groups, preserving the independence of the validation process, preventing any patient's data from simultaneously appearing in both training and validation sets. The group-wise holdout method was used to prevent bias in the feature selection process whilst attempting to maximise the generalisability of selected features to new, unseen patient data.

The RFECV technique employed a random forest regression model as the estimator. The minimum number of features to be selected was set to 1, with a step size of 1, meaning that RFECV iteratively removes features one-by-one until only one feature remains. For each set of features, the random forest regression model is evaluated. Using a group-wise holdout approach, the model's performance is assessed in a cross-validation scheme, ensuring that the model is validated against each patient's data separately. The evaluation of model performance during feature selection was conducted using the negative root mean squared error (neg_RMSE). By maximising the neg_RMSE, RFECV identifies feature sets which effectively minimise the RMSE, thereby reducing prediction errors.

7.2.3 Feature selection results

Within the proximal and distal datasets, 377 and 162 correlated feature pairs were identified. Once the features with the lower MIC score were removed and a unique set of the remaining features was taken, the resultant feature sets had a length of 52 and 61 features for the proximal and distal sets respectively. With the resulting features, RFECV was applied to the feature subsets to determine the most predictive features.

Figure 7.2 illustrates the results of the RFECV for both proximal and distal datasets. The x-axis represents the number of features selected, while the y-axis shows the neg_RMSE. The blue lines indicate the mean neg_RMSE, calculated by averaging the neg_RMSE val-

ues obtained across all validation folds (held out patients) for each number of features. Error bars around the blue lines represent the variability in model performance, showing one standard deviation above and below the mean neg_RMSE. The red dashed lines denote the optimal number of features, determined by the maximum mean neg_RMSE value.

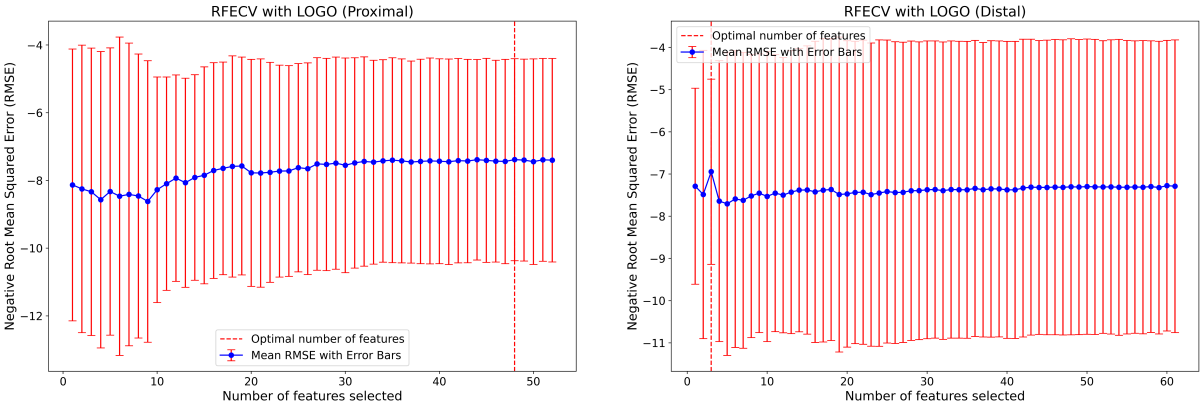

Figure 7.2: A figure with two subplots illustrating recursive feature elimination plots applied to a subset of features. The plots demonstrate the variation in negative Root Mean Squared Error as the number of features changes. A dashed red line represents the number of features with support returned by recursive feature elimination. The left subplot corresponds to the proximal dataset, while the right subplot corresponds to the distal dataset.

The RFECV approach identified 48 features and 3 features yielded the best performance on the wrapped, random forest regression models across validation folds for the proximal and distal datasets respectively. The identified 3 features for the distal dataset led to underfitting. To counteract this, the number of features which returned the second highest neg_RMSE for the distal dataset was selected (n=18). Table 7.1 contains the selected features for both the proximal and distal datasets. Within the table, feature names are followed by a number if they have an association with the pulse prominence such as "pulse width 100" or "diastolic width 50 (deriv 1)". These numbers indicate the specific percentage of the pulse prominence at which the feature was extracted. For instance, "pulse width 100" means that the feature was measured at 100% of the pulse prominence, representing the full width of the pulse. Similarly, "diastolic width 50 (deriv 1)" means that the feature was extracted at 50% of the pulse prominence.

Proximal	Distal
onset-end slope (deriv 1)	pulse width 100
ds ratio 25	kurt (deriv 2)
std ibi (deriv 2)	pulse width 10
rise time - decay time ratio (deriv 1)	diastolic width 50 (deriv 1)
systolic width 100	systolic width 75
upslope	rise time - decay time ratio
kurt	rise time - decay time ratio (deriv 1)
rise time - decay time ratio	max end datum difference
pulse length - height ratio	ibi (deriv 2)
std ibi	skew
ds ratio 25 (deriv 1)	ds ratio 50 (deriv 2)
kurt (deriv 1)	start datum difference
systolic width 10	kurt (deriv 1)
systolic width 75 (deriv 2)	zcr
max start datum difference (deriv 2)	skew (deriv 1)
zcr	ds ratio 50 (deriv 1)
ds ratio 75 (deriv 2)	skew (deriv 2)
skew	systolic width 25 (deriv 2)
end datum difference (deriv 1)	
downslope length (deriv 2)	
onset-end slope (deriv 2)	
pulse length - height ratio (deriv 2)	
end datum difference (deriv 2)	
max start datum difference	
start datum area (deriv 1)	
end datum area (deriv 1)	
systolic width 100 (deriv 1)	
max end datum difference	
ds ratio 10	
downslope	
downslope (deriv 1)	
std ibi (deriv 1)	
systolic width 25 (deriv 2)	
end datum difference	
onset-end slope	
zcr (deriv 2)	
ds ratio 25 (deriv 2)	
start datum difference (deriv 2)	
skew (deriv 1)	
ds ratio 75	
max end datum difference (deriv 1)	
start datum difference (deriv 1)	
diastolic width 10 (deriv 1)	
end datum area (deriv 2)	
diastolic width 75 (deriv 2)	
pulse width 10 (deriv 2)	
systolic width 10 (deriv 1)	
diastolic width 10	

Table 7.1: Table containing the selected features for both the the proximal and distal datasets.

7.3 Bayesian optimisation

The hyperparameters of the five regression models were optimised using the optimisation dataset exclusively.

Hyperparameter optimisation was conducted using the Optuna library, a Bayesian optimisation framework, using the Tree-structured Parzen Estimator (TPESampler) algorithm [151]. Bayesian optimisation operates by constructing a probabilistic model that maps hyperparameter values to a target objective function. This probabilistic model is then used to estimate the performance of hyperparameters, helping to steer hyperparameter selection to evaluate in the actual objective function. In contrast to methods such as grid or random search, which search the hyperparameter space either systematically or randomly, Bayesian optimisation uses prior knowledge from previous evaluations to inform its search strategy. This approach enables a more efficient search of hyperparameter spaces, often resulting in superior hyperparameter solutions.

The hyperparameters were optimised using the RMSE across leave-one-patient-out cross-validation (LOPOCV). In LOPOCV, the model is trained on the data from all but one patient and validated on the data from the left-out patient.The reason for adopting LOPOCV for both optimisation and future evaluation of the models was to optimise and evaluate the model's for robustness to inter-patient variability. By mirroring a real-world scenario where the model must make predictions on new patient data, LOPOCV serves as a robust test of the model's generalisability whilst minimising the possible positive bias of including data from the same patient in the training and testing dataset. In scenarios where the ultimate goal is to apply the model across different individuals, the ability to generalise well to unseen patients' data is crucial.

For each model 50 hyperparameter optimisation "trials" were performed. To mitigate the risk of overfitting to the training dataset during model optimisation, the hyperparameter search spaces were tuned. Initially, the hyperparameter search spaces encompassed broader ranges, but as the optimisation progressed if any instances of overfitting to the training data were seen the search space was refined. This involved narrowing the range of specific hyperparameters, implementing stronger regularisation penalties, and adjusting learning rates to encourage improved generalisation. The primary objective was to strike a balance, achieving good performance on the training data while ensuring the model's effectiveness in generalising to the validation data. The hyperparameter search spaces for

each model are outlined in Table 7.2. The hyperparameter optimisation was carried out twice, once for both the proximal and distal optimisation datasets.

Model	Hyperparameter Search Space	Optimised Hyperparameters (Proximal)	Optimised Hyperparameters (Distal)
RF	max depth = 15 to 25 number of estimators = 100 to 500 min samples split = 10 to 20 min samples leaf = 10 to 15	max depth = 18 number of estimators = 101 min samples split = 13 min samples leaf = 11	max depth = 17 number of estimators = 196 min samples split = 10 min samples leaf = 11
KNN	number of neighbours = 3 to 10 algorithm = [auto, ball tree, kd tree, brute] metric = [euclidean, manhattan, chebyshev, minkowski]	number of neighbours = 10 algorithm = auto metric = chebyshev	number of neighbours = 10 algorithm = auto metric = chebyshev
LGBM	max depth = 5 to 10 number of iterations = 100 to 250 learning rate = 0.1 to 0.5 number of leaves = 25 to 35 extra tress = True or False	max depth = 10 number of iterations = 102 learning rate = 0.179 number of leaves = 28 extra tress = False	max depth = 8 number of iterations = 241 learning rate = 0.160 number of leaves = 35 extra tress = True
XGB	number of estimators = 10 to 20 max depth = 10 to 15 learning rate = 0.3 to 0.7 min child weight = 45 to 55	number of estimators = 19 max depth = 10 learning rate = 0.351 min child weight = 53	number of estimators = 18 max depth = 14 learning rate = 0.382 min child weight = 55
SVR	C = 10 to 50 gamma (DISTAL) = 5 to 10 gamma (PROXIMAL) = 1 to 6 epsilon = 0.0001 to 0.01	C = 10.855 gamma = 5.986 epsilon = 0.000164	C = 10 gamma = 9.972 epsilon = 0.002

Table 7.2: Table containing the defined search spaces for each regression model and the resultant optimised hyperparameters.

7.4 Model Evaluation

The mean squared error (MSE), RMSE, mean absolute error (MAE), coefficient of determination R^2, the upper and lower Bland-Altman limits of agreement as well as the correlation coefficient were calculated. The training and testing results are presented in Tables 7.3 and 7.4.

Distal Results (Evaluation)								
Model	**RMSE Test**	**MSE Test**	**MAE Test**	**R2 Test**	**Bland-Altman Upper Test**	**Bland-Altman Lower Test**	**Bland-Altman Bias Test**	**Test Corr Coef**
KNN	5.529	32.735	4.445	-1.496	10.057	-9.202	0.427	0.021
LGBM	5.253	29.902	4.233	-1.217	9.220	-8.862	0.179	-0.016
RF	5.030	27.364	4.067	-1.046	8.782	-8.487	0.147	-0.007
SVR	5.473	32.148	4.404	-1.370	9.266	-9.348	-0.041	0.027
XGB	5.379	30.986	4.338	-1.354	9.372	-9.154	0.109	0.001
Proximal Results (Evaluation)								
Model	**RMSE Test**	**MSE Test**	**MAE Test**	**R2 Test**	**Bland-Altman Upper Test**	**Bland-Altman Lower Test**	**Bland-Altman Bias Test**	**Test Corr Coef**
KNN	5.734	35.212	4.650	-1.766	10.055	-8.995	0.530	-0.052
LGBM	5.627	33.680	4.634	-1.677	9.414	-8.661	0.376	-0.022
RF	5.538	32.969	4.538	-1.555	9.423	-8.598	0.413	-0.026
SVR	5.640	33.268	4.608	-1.701	9.289	-9.159	0.065	-0.020
XGB	5.689	34.444	4.663	-1.729	9.625	-9.025	0.300	-0.013

Table 7.3: Table containing the results (test) for the 5 regression models, estimating intracranial pressure. The table contains the mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), the upper and lower Bland-Altman limits of agreement as well as the correlation coefficient (Corr Coef) values for each model.

Distal Results (Evaluation)								
Model	**RMSE Train**	**MSE Train**	**MAE Train**	**R2 Train**	**Bland-Altman Upper Train**	**Bland-Altman Lower Train**	**Bland-Altman Bias Train**	**Train Corr Coef**
KNN	3.063	9.387	2.011	0.625	6.021	-5.986	0.017	0.792
LGBM	2.993	8.961	2.121	0.642	5.866	-5.866	0.000	0.807
RF	2.709	7.341	1.854	0.707	5.316	-5.303	0.006	0.861
SVR	3.203	10.267	1.986	0.590	5.965	-6.542	-0.288	0.772
XGB	2.573	6.625	1.816	0.735	5.041	-5.045	-0.002	0.862
Proximal Results (Evaluation)								
Model	**RMSE Train**	**MSE Train**	**MAE Train**	**R2 Train**	**Bland-Altman Upper Train**	**Bland-Altman Lower Train**	**Bland-Altman Bias Train**	**Train Corr Coef**
KNN	3.613	13.058	2.586	0.478	7.128	-7.034	0.047	0.696
LGBM	3.052	9.316	2.198	0.628	5.981	-5.981	0.000	0.798
RF	2.556	6.540	1.738	0.739	5.019	-5.002	0.008	0.877
SVR	2.427	5.892	1.199	0.765	4.474	-4.985	-0.256	0.880
XGB	2.643	6.991	1.876	0.721	5.177	-5.184	-0.003	0.855

Table 7.4: Table containing the results (train) for the 5 regression models, estimating intracranial pressure. The table contains the mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), coefficient of determination (R^2), the upper and lower Bland-Altman limits of agreement as well as the correlation coefficient (Corr Coef) values for each model.

Extrapolating from the results presented in Table 7.3, an examination of the predictive performance is conducted, focusing on four key metrics: RMSE, MAE, correlation coefficient and Bland-Altman limits of agreement. MAE assesses the average absolute errors between predicted and actual values. Unlike MSE, MAE does not square errors meaning it does not amplify the influence of outliers or large errors. By not penalising large errors, MAE offers a balanced view of model performance.

On the other hand, RMSE also gauges the discrepancy between predicted and actual values. In contrast to MAE, RMSE calculates the square root of the average of squared errors. Squaring the errors gives more weight to larger errors, meaning they have a proportionally greater impact on the RMSE. This makes RMSE sensitive to outliers and large errors, emphasising their importance. It is suggested that in the context of estimating ICP, the clinical implications of errors can vary significantly. Larger errors in ICP prediction can have critical consequences for patient care and safety. For instance, underestimating ICP might lead to inadequate intervention, risking patient health. Therefore, prioritising the reduction of larger errors is crucial to enhance the clinical utility and safety of the model. RMSE aligns well with this objective. Given this, the feature selection process using RFECV and model hyperparameter optimisation were centred around minimising squared errors as the primary cost function. RMSE, being the square root of MSE, allows for a more intuitive interpretation of the model's performance as it brings the evaluation metric back to the original units of the target variable. Both MAE and RMSE are presented in the same units as the target variable (mmHg), facilitating easier comprehension and comparison.

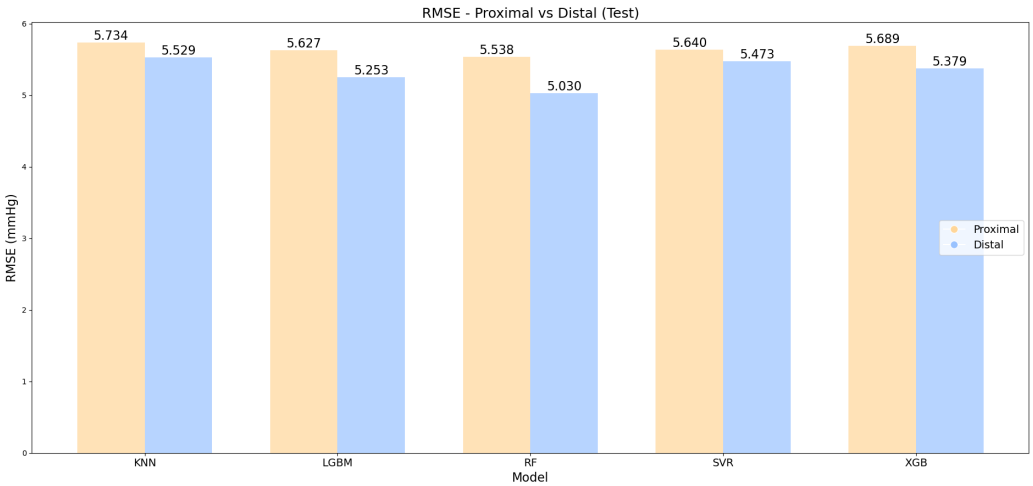

Figure 7.3: A bar chart presenting the root mean squared error of each model for both the short and long distance near-infrared spectroscopy data.

Although it is appreciated and understood that a comparison between the proximal and distal results can be challenging due to the variance in feature and hyperparameter selection. Overall the models developed using distal features performed better than model developed using proximal features. The mean MAE and RMSE across models developed using distal features were 4.297 and 5.333 mmHg, approximately 5.7% and 7.2% lower than the mean MAE and RMSE across models developed using proximal features. Of the models developed using distal data, the best performing model with reference to RMSE and MAE was a RF model which returned a mean RMSE and MAE across the testing folds of LOPOCV of 5.030 and 4.067 mmHg respectively. The worst performing distal model with reference to RMSE and MAE was a KNN model which produced an RMSE and MAE of 5.529 and 4.445 mmHg.

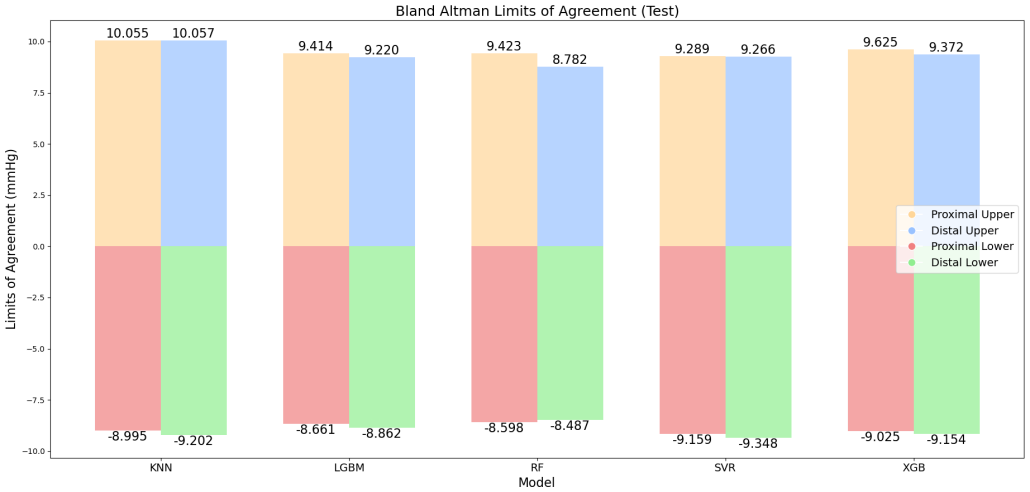

Figure 7.4: A bar chart representing the Bland-Altman upper and lower limits of agreement for each model, calculated at the 95% confidence level. The bars in the chart indicate the upper and lower limits of agreement, representing the range within which most of the differences between predicted and actual values lie with 95% confidence.

To complement the RMSE and MAE the correlation coefficient and Bland-Altman limits of agreement of the developed models are also discussed. Figure 7.4 depicts the Bland-Altman limits of agreement for each developed model. Bland-Altman is a statistical technique used to assess the agreement between two quantitative measurements. The Bland-Altman plot provides insights into the extent of agreement and any potential biases between the predicted and actual values. The limits of agreement, marked in the plot, illustrate the range within which most of the differences between predicted and actual values lie. Specifically, they represent the mean difference between the predicted and actual values, with 95% confidence intervals meaning that 95% of the differences fall

within this range, offering a valuable indication of the agreement's reliability.

The mean upper and lower limits of agreement of the distal and proximal derived models are 9.334 and -9.011 and 9.561 and -8.888 mmHg accordingly. The limits of agreement between the models trained using distal and proximal data are similar. Both groups of models returned negligible mean biases of 0.337 and 0.164 mmHg suggests a minimal systematic difference between the predicted and actual ICP values. This implies that, on average, the ICP predictions from the models are within approximately 8 mmHg of the actual values.

In the case of ICP prediction a preferable correlation coefficient would be a strong positive correlation with label. The mean correlation coefficient returned by both the distal and proximal derived models are very low. The distal derived models returned a mean correlation coefficient of approximately 0.005, 118.5% greater than the mean correlation coefficient returned by the proximal derived models. The higher and positive mean correlation coefficient of the distal derived models apposed to the mean negative correlation coefficient of the proximal derived models suggests that distal features may be more useful for the prediction of ICP. Despite this, the correlation coefficients remain low.

7.5 Discussion

This study has investigated the effectiveness of non-invasive ICP monitoring using NIR-PPG derived features. Five classical machine learning models were optimised and evaluated on both short distance and long distance NIR-PPG data.

Despite appreciating the challenges of comparison due to the variance in feature and hyperparameter selection. From the models developed, the distal models returned lower RMSE and MAE errors on average than the models developed using proximal derived features. This supports our hypothesis that long distance NIR-PPG data is associated with changes in ICP. Additionally although both groups of models returned low correlation coefficients the distal models returned a higher mean correlation coefficient in comparison to the proximal models which returned a mean negative correlation coefficient.

The low correlation coefficients and wide limits of agreement observed in our analysis suggest that the models struggled to generalise to new, unseen data. It is hypothesised that this challenge is primarily attributable to noise within the data and inter-patient

data heterogeneity. Specifically, variances in the distribution of ICP among patients may impact model performance.

Moreover, the presence of inter-patient feature data heterogeneity and the consequential challenge of creating a shared distribution exacerbates the difficulty in achieving optimal model generalisation. This heterogeneity can stem from diverse patient characteristics, disease severity, and treatment responses among others. This coupled with varying distributions of ICP levels among the patients, can impair the model's ability to learn underlying patterns effectively thereby resulting in wider limits of agreement.

Addressing these challenges could involve increasing the dataset, particularly with more data collected at the high and low ends of the ICP distribution as well as increasing the the diversity of the patient population which comprise the dataset. A larger and more diverse dataset could help in reducing the observed wide limits of agreement by enabling the model to better learn and predict the outcomes for patients with less common characteristics.

Animal studies have investigated the estimation of ICP non-invasively using NIRS, features extracted from DCS pulsations and NIR-PPG cardiac waveforms [93]–[95]. The approaches within these studies have demonstrated significant agreement with ICP returning R^2 of up to 0.937 and MSE lower than < 3 mmHg. Despite this, these studies use MAP as an additional feature for model development and a random sampling cross-validation evaluation approach. MAP was found to be the strongest predictive feature across both studies [94], [95]. Although the small subject size of these studies is understood, it is suggested that the inclusion of MAP and the evaluation methodologies of these studies may positively influence the results of these studies. This study aimed to investigate the efficacy of predicting ICP non-invasively using NIR-PPG morphological and time-series features exclusively.

The evaluation and optimisation methodology could be optimised further. The feature selection and model optimisation were performed using the optimisation dataset exclusively to minimise potential data leakage. Model optimisation utilised the aggregate RMSE across LOPOCV. Following hyperparameter optimisation, model evaluation occurred through LOPOCV on the evaluation dataset. This methodology was chosen over nested cross-validation to strike a balance between performance and computational efficiency. Despite this choice, there remains a possibility of bias in the models towards the optimisation set's data. The 30% of patients' data constituting the optimisation dataset

was deemed a reasonably representative sample. Even if the models exhibit bias towards the optimisation dataset, they do not positively bias the evaluation results; if anything, they could negatively impact them. This approach was considered reasonable and serves to prevent data leakage and provide a robust evaluation of model performance.

The evaluation and optimisation methodology could be optimised further. The feature selection and model optimisation were performed using the optimisation dataset exclusively to minimise potential data leakage. Model optimisation utilised the aggregate RMSE across LOPOCV. Following hyperparameter optimisation, model evaluation occurred through LOPOCV on the evaluation dataset. This methodology was chosen over nested cross-validation to strike a balance between performance and computational efficiency. However, this approach does have some limitations.

There remains a possibility of bias in the models towards the optimisation set's data. Although the 30% of patients' data constituting the optimisation dataset was deemed a reasonably representative sample, it may not capture the full diversity of the entire dataset. This could lead to models that perform well on the optimisation dataset but less so on the new, unseen data.

The use of the optimisation dataset exclusively for feature selection and model optimisation might result in an overfitting risk. The selected features and hyperparameters are tailored to the specific characteristics of the optimisation dataset, which could limit the generalisability of the model when applied to the evaluation dataset. Even if the models exhibit bias towards the optimisation dataset, they do not positively bias the evaluation results; if anything, they could negatively impact them.

Nested cross-validation would have allowed for the simultaneous optimisation of model parameters and feature selection while evaluating model performance in an unbiased manner across different folds. However, nested cross-validation could result in each fold having different hyperparameters and selected features. In contrast, the chosen approach ensures that the same hyperparameters and features are used across all patients during evaluation reflecting a real-world scenario where a single model is deployed universally across different patients.

The chosen approach ensures consistent model parameters and selected features across all evaluated patients, balancing computational efficiency with a robust evaluation of model performance.

7.6 Summary

This study aimed to assess the potential of non-invasive ICP monitoring using NIR-PPG derived features. Five classical machine learning models were optimised and evaluated on both short distance and long distance NIR-PPG data. The results indicated that distal models returned better results than models developed using proximal features returning a higher mean RMSE, MAE and correlation coefficient. The best peforming model was a RF model which returned a mean RMSE and MAE across the testing folds of LOPOCV of 5.030 and 4.067 mmHg respectively with Bland-Altman limits of agreement of approximately 8.5 mmHg and a low correlation coefficient of -0.007.

This research provides insights to non-invasive ICP monitoring through the investigation of the largest clinically collected NIR-PPG ICP labelled dataset from patients with severe TBI. The outcomes of this study provide solid foundation for future research within this domain especially the investigation of the causes of possible inter-patient heterogenity.

The search for non-invasive alternatives to ICP monitoring is significant. The potential of NIR-PPG technology to change TBI diagnosis and ICP monitoring arises as a promising prospect. NIR-PPG based monitoring offers the potential for an inexpensive, easy-to-use, safe, and real-time alternative. Moreover, a NIR-PPG device as a consequence of its non-invasive nature and simplicity of use could make an ideal point-of-care monitor in various settings, from emergency rooms and ambulances to sports events and remote locations. The cost-effectiveness of NIR-PPG further improves its appeal, as it offers a more affordable approach to TBI diagnosis and ICP monitoring.

Chapter 8

Discussion and Conclusion

8.1 Discussion

8.1.1 Summary of thesis and findings

The aim of this research was to explore the dynamic relationship between PPG derived features and variations in ICP levels. This work contributes to the nascent research field of non-invasive ICP monitoring. It provides insights derived from the largest, clinically collected PPG dataset labelled with invasive ICP data, to-date. The research was predicated upon the data collected by the nICP sensor. The nICP sensor's design was based upon the results of a Monte-Carlo simulation of light-tissue interaction which indicated certain photodiode-photodetector separations in which NIR light penetrated deeper into the head tissues allowing for the collection of data from the white matter of the brain [101]. The subsequent sensor shown in Figure 2.8 consisted of two photodiodes "proximal" and "distal" positioned at 10 mm and 35 mm from the light source representing short-distance and long-distance data respectively. It was hypothesised that the data collected from proximal photodiode contained extracerebral information and the distal contained both extracerebral and cerebral information. This research focused upon the 810nm wavelength data as it represents the isosbestic point allowing for an optical signal independent of blood oxygenation, eliminating it as a confounding factor [103].

Using the nICP sensor a data collection was performed over a 78-week period from 40 patients from the ITU of the Royal London Hospital who had received a severe TBI diagnosis. Data from the nICP sensor and from a gold standard, invasive ICP monitor was collected synchronously for an average of approximately 42 hours for each patient.

There were significant challenges associated with the data collected. Firstly, although it was assumed that the invasive and non-invasive data began collection synchronously in many cases one data source was of a different size to the other or the data of patients were made up of multiple different recordings which could have been caused by the need to remove sensors for routine scans or interventions. When loading and structuring the raw data files, assumption had to be made in order to deal with these discrepancies.

Once the raw NIR-PPG and ICP data had been loaded it became apparent that there was a significant amount of noise and hypothesised measurement errors present within both signals. There were a number of patients which had either physiologically implausible ICP values or clear measurement errors, as depicted by Figures 3.4 and 3.3. In order to handle this, data from patient's with hypothesised measurement error were excluded. The remaining data was handle through the definition of physiologically plausible bounds of ICP. Any data indexes which fell below or above these bounds were identified and removed from both the reference ICP data and the NIR-PPG data. On investigation of the NIR-PPG there was a significant presence of noise within the data as shown in Figure 3.6. This noise served as a major challenge which underpinned this research. The noise within the data was primarily characterised by high amplitudinal variance which is often associated with motion artifact and flat lines which are associated with photodetector saturation. In order to minimise the presence of noise within the data without onerously having to label large quantities of the dataset in order to develop machine learning models a novel, statistical PPG denoising algorithm was developed and applied to the dataset named the *Envelope PPG Denoising Algorithm*. The EPDA was evaluated and statistically shown to reduce the noise present within the signal across four metrics.

Following the denoising and the removal of patient's with missing or erroneous data the dataset was comprised of approximately 38,544.56 minutes or 642.41 hours of data from 27 patients. Feature extraction was then performed using this data. The data of each patient was segmented into 1 minute windows of data. From each one minute window of data 141 time-series and morphological features were extracted from the original signal as well as the first and second derivatives. Considering that the derived NIR-PPG features underpin this research and given the significant presence of noise within the signal three decisions were made in an effort to maximise the reliability of the extracted features: Firstly a z-score outlier detection algorithm was developed using three pulse characteristics. This was done in an effort to identify and ignore possible residual anomalous pulses which may

be present within a window of data. Secondly, all extracted features were predicated upon three fiducial points, the pulse onset, peak and end. This decision was made due to the lack of a reliable diacrotic notch across patients and the presence of noise in the data. It was assumed focusing the detection and subsequent feature extraction around three "easier" to detect fiducial points would yield more reliable features. Finally, the first and second derivatives were calculated using a Savitzky-Golay filter, this again was done due to the noise within the data. The filter's polynomial smoothing affect was hypothesised to result in more efficacious fiducial point identification and subsequently more robust feature extraction, the Figures 4.3 and 4.4 depict the pulse detection on the calculated first and second derivatives of a window of data.

Once the features had been extracted from both the proximal and distal NIR-PPG data, the three central hypotheses of the research could be tested. These included: (i) if alterations in PPG features correlate with changes in ICP levels, (ii) if the association between PPG features and variations in ICP levels is stronger in long-distance NIR-PPG data compared to short-distance NIR-PPG data, and (iii) if PPG-derived features can accurately estimate ICP non-invasively.

In order to test the first two hypotheses a statistical analysis was conducted. In order to increase the robustness of the statistical analysis, a data selection protocol was defined. A graphical user interface was developed as shown in Figure 5.1 which enabled the classification of "good" and "bad" windows of data for each patient. For each patient, 6 hours of data was classified as good or the end of a specific patient's data was reached before stopping. This meant that the data of which the statistical analysis was based upon was of "good" quality meaning the subsequent extracted features should also be more reliable. Given the slow changing nature of ICP it was assumed that 6 hours of data represented a reasonable representative sample for each patient. The findings of the correlation and group analysis conducted using Spearman correlation and Kruskal-Wallis affirmed the first two hypotheses. The results demonstrated significant correlations between the features and ICP levels. Specifically, 77.30% and 79.43% of features significantly correlated ($p < 0.05$) with the label in distal and proximal datasets, respectively. The Kruskal-Wallis analysis revealed that 81.56% and 75.89% of features changed significantly ($p < 0.05$) across ICP groups. Although there is a non-significant difference in the number of significantly correlated features (2.75%) and significant differences across groups 7.21% between the proximal and distal datasets, the correlation analysis shows that the absolute

average correlation coefficient of all features and significantly correlated features in the distal dataset is 25.76% and 24.24% higher, respectively, than in the proximal dataset.

Furthermore the findings indicate that many of the most significant changes in features affected the pulse width. It is hypothesised that the recorded changes in pulse width may reflect alterations in blood flow and vessel elasticity. As ICP increases, reducing CPP below the critical threshold and consequently diminishing CBF, less blood volume per cardiac cycle reaches the brain. A narrower pulse width may signify a shorter duration of blood volume change, potentially associated with decreased compliance or lower CBF. The majority of the most significant feature changes were observed in the first derivative, particularly within the distal dataset, which is believed to encompass "cerebral" information. The first derivative, representing flow velocity. An increase in ICP can compress cerebral vessels, thereby diminishing CBF. This constriction not only reduces the volume of blood that can flow through these vessels but may also affect the velocity of blood flow, likely captured by changes in the first derivative of the PPG signal indicating that understanding these changes in features related to blood flow velocity may be significant in describing ICP dynamics via PPG data.

In order to test the final hypothesis if NIR-PPG derived features can be used to estimate ICP non-invasively, five classical machine learning models were developed. To perform feature selection, model optimisation and evaluation, both the proximal and distal NIR-PPG datasets were partitioned by patient in a ratio 70:30, as shown in Figure 7.1 creating an *evaluation* and *optimisation* dataset for both the proximal and distal data. Feature selection was conducted for both the proximal and distal optimisation datasets using RFECV employing LOPOCV and 10 fold cross validation. The feature selection, hyperparameter optimisation and evaluation were conducted using LOPOCV to optimise and evaluate the model's for robustness to inter-patient variability. By mirroring a real-world scenario where the model must make predictions on new patient data. Following the feature selection the models hyperparameters' were optimised using Bayesian optimisation. Following this each model was evaluated separately on both the proximal and distal data. The lowest MAE and RMSE were obtained using features derived from long-distance data. A RF model achieved the lowest RMSE and MAE of 5.030 and 4.067 mmHg respectively. The RF exhibited wide limits of agreement with the reference method. This was reflected in the 95% Bland-Altman limits of agreement, ranging from 8.782 to -8.487 mmHg. It is hypothesised that the ICP estimation is primarily challenged by the noise

within the data, inter-patient data heterogeneity and ICP distribution. Despite this the findings present novel insights. This research area is nascent and evolving and future work is required. The outcomes of this research provide a promising foundation for future research within this domain.

8.1.2 Comparison with other related works

The field of ICP monitoring using non-invasive physiological data is in its early stages and is developing. The initial pilot study in 1997 demonstrated the potential of NIRS data as a viable method for assessing elevated ICP. Subsequent studies exploring the link between PPG derived features and ICP have been conducted (with the exception of a small and opaque study) with the use of non-human primates [93]–[95], [152]. Notably, this body of research, to the best of the author's knowledge has focused almost exclusively on a limited set of morphological and time-series features derived directly from the original signal, without the analysis or incorporation of features from the derivative of the signal.

Additionally, these studies have relied on continuous, invasively collected ABP derived features, which has consistently been shown to be the strongest predictive features across all the studies which used it [94], [95]. The majority of predictive models developed in this domain have utilised the RF algorithm [94], [95], achieving promising outcomes, such as a R^2 of 0.937 and a RMSE of 2.703 mmHg [95]. Despite this and whilst appreciating the small sample size of these studies the methodological approach of using random sampling for cross-validation raise concerns about the generalisability of these findings. Random cross-validation can introduce a positive bias by including data from the same subjects in both training and testing datasets, potentially inflating the performance metrics.

Direct comparisons with related literature are challenging due to differences in data sources and evaluation methodologies. This research distinguishes itself by leveraging the largest clinically collected dataset and the most extensive set of extracted PPG features to date for this application. It also marks a significant difference from previous studies by exclusively utilising NIR-PPG derived features and implementing a more rigorous LOPOCV approach for model evaluation. It is the first to perform a comprehensive statistical analysis explicitly examining the relationship between PPG features and ICP and evaluating a broad group of machine learning models for this predictive task.

8.2 Strengths and limitations

8.2.1 Strengths

- The dataset used in this study is the largest clinically collected dataset of its kind to-date. NIR-PPG signals labelled with invasive ICP data were collected from 40 patients. When denoised and refined the dataset consisted of approximately 38,000.00 minutes or 640 hours of data from 27 patients.
- A novel, effective and efficient PPG denoising algorithm was developed. When evaluated on a subset of the NIR-PPG data it demonstrated its efficacy by yielding significant differences ($p < 0.05$) for all evaluation metrics before and after denoising.
- A significant number of features (141) were extracted from the original signal and its first and second derivatives. The feature extraction was conducted from both short-distance (proximal) and long-distance (distal) datasets.
- Using both short and long distance NIR-PPG derived features a robust investigation of the dynamic relationship between the derived features and ICP was conducted.
- A robust feature selection methodology was carried out and the largest group of machine learning models to-date were implemented and evaluated, for this predictive task.

8.2.2 Limitations

- A significant challenge of this research was the noise and measurement errors present within the NIR-PPG and ICP data. The presence of noise within the NIR-PPG led to challenges in maximising efficacious detection of fiducial points from which the subsequent features were based upon and which drove this research. Although a substantial number of features were extracted, a higher quality signal could enable the detection and extraction of more features such as feature derived from the diacrotic notch. This was not done within this research due to data integrity and inter-patient variability and in an attempt to make feature extraction more reliable. Additionally, although reasonable bounds were defined to remove erroneous ICP data, more reliable ICP data and/or narrower bounds for ICP removal could lead to better ICP estimation results.

- The developed and implemented denoising algorithm was effective at removing noise present across the signal which significantly aided the preprocessing of the data. Despite this, the algorithm was not focused on pulse level quality assessment. A pulse level anomaly detection was implemented at the time of feature extraction, which although effective was predicated upon the characteristics of the pulses within each window. This anomalous pulse detection approach therefore is limited in moments where poor quality pulses make up the entire window.

- The developed pulse detection algorithm, tailored specifically for this dataset, was qualitatively assessed through visual inspection of the detection on NIR-PPG data during its development rather than through quantitative evaluation. We acknowledge that the absence of a quantitative assessment may lead to potential errors in pulse detection and subsequent feature extraction. However, we are reasonably assured of the accuracy of the extracted features for the included data. This assurance stems from the fact that the pulse detection algorithm's results were visually inspected and verified during the data selection process outlined within the methodology of the statistical analysis discussed in Chapter 8. Thus, despite the potential limitations, we believe the qualitative assessment provided a satisfactory verification of the algorithm's performance for our purposes.

- The implemented classical machine learning models do not inherently consider the temporal structure of the data. They generally treat each instance of data as independent, without accounting for potential temporal dependencies between observations. As a result, they might fail to capture trends, seasonality, and other time-related dynamics which may prove important for the estimation of ICP.

  Using a group holdout approach for feature selection and model optimisation with a subset of the entire dataset presents limitations. The assumption that the optimisation subset is representative of the entire dataset may not hold true, increasing the risk of overfitting. This method can result in the selection of features and hyperparameters that are overly specific to the characteristics of the subset, potentially limiting the model's ability to generalise to new, unseen data.

8.3 Conclusion and future work

An important area for future research involves the collection of a new and more comprehensive dataset. The current dataset, as described in Section 3.1, was limited by being a single-center collection and included data from 27 patients after data exclusion. This relatively small sample size and the single-arm nature of the study poses possible limitations. A small sample size and single-center study may limit the generalisability and reliability of the findings. Factors such as patient demographics and treatment protocols can vary significantly across different regions and centers. Data collected from a smaller sample from a single center might not capture the variability seen in a more diverse, multi-center study.

Additionally, the dataset suffered from considerable noise, both in the non-invasive NIR-PPG and the invasive ICP data. An attempt to mitigate the noise present within the data was taken through the implementation of a novel PPG denoising algorithm, the "EPDA" described in Section 3.2.1 and by identifying and removing invasive ICP values based on physiologically feasible bounds.

Moreover, the dataset lacked stratified demographic and medical information, such as medication usage and skull thickness which could potentially serve as important features for non-invasive ICP estimation. Due to effective clinician intervention another limitation of the dataset was the scarcity of elevated and valuable ICP data above the clinically significant threshold of circa 20 mmHg.

To address these challenges, a larger and multi-center data collection would be useful to increase the sample size and diversity. A focus should be placed on collecting more data above the 20 mmHg threshold to possibly improve the robustness of the ICP estimation models and statistical analysis. A revised data collection protocol, which may involve shorter but more supervised collection periods, could help ensure higher data quality. Furthermore, potential modifications to the non-invasive NIR-PPG sensor could enhance the quality of the NIR-PPG signals collected.

Another important direction for future research involves the quantitative evaluation of the PPG peak detection algorithm. The developed pulse detection algorithm, specifically tailored for this dataset, was assessed qualitatively through visual inspection of the detection on NIR-PPG data during its development, rather than through a quantitative evaluation. This qualitative approach, while useful, may not fully capture the algorithm's

accuracy and robustness.

The absence of a quantitative assessment introduces potential limitations, as it may lead to errors in pulse detection and subsequent feature extraction. Such errors could affect the reliability of the findings and the overall performance of the non-invasive ICP estimation model. While reasonably assured of the accuracy of the extracted features owing to the visual inspection and verification of the pulse detection algorithm during the data selection process, as outlined in the methodology of the statistical analysis discussed in Chapter 8, this assurance is limited by the subjective nature of visual inspection.

To enhance the reliability and validity of the pulse detection algorithm, future work should focus on developing and implementing a quantitative evaluation framework. This framework could use a labelled version or subset of the dataset and assess how many pulses are correctly identified. Additionally, validation with other datasets and comparison with existing pulse detection algorithms could provide further insights into its accuracy and generalisability.

Enhancing the machine learning and feature selection process represents another significant area for future work. The classical machine learning models implemented and evaluated in this research do not inherently consider the temporal structure of the data, treating each data instance as independent without accounting for potential temporal dependencies between observations. This limitation may result in the failure to capture trends, seasonality, and other time-related dynamics, which are potentially important for accurate ICP estimation. To address these limitations, future research could explore the implementation and evaluation of time series-specific models, such as recurrent neural networks (RNNs) or long short-term memory networks (LSTMs) designed to handle temporal dependencies and can more effectively model the sequential nature of the data.

The exclusive use of the optimisation dataset for feature selection and model hyperparameter optimisation using a group holdout approach has its limitations. Although the data constituting the optimisation dataset was deemed a reasonably representative sample, it may not capture the full diversity of the entire dataset. Subsequently the selected features and hyperparameters might overfit to the optimisation dataset, potentially limiting the generalisability of the model when applied to new, unseen data.

Nested cross-validation offers a more rigorous and unbiased method for model evaluation and optimisation. This approach involves an inner loop for hyperparameter tuning and feature selection and an outer loop for model evaluation. Each fold in the outer loop

uses the inner loop to find the best model configuration, ensuring that the evaluation process does not influence the model optimisation. By performing feature selection and hyperparameter tuning within the inner loop, nested cross-validation reduces the risk of overfitting to a specific subset of the data perhaps enhancing the model's generalisability.

Despite its computational expense implementing nested cross-validation in future research may lead to the development of more robust, generalisable models for ICP estimation.

TBI represents a significant global health concern with an estimated annual incidence ranging between 50 to 69 million cases worldwide. The burden of TBI disproportionately affects LMICs, where incidences are reported to be significantly higher than in HICs. As regions with the lowest resources bear the highest disease burden its seems reasonable to suggest that the incidence of TBI is also under-reported. The main monitoring modality for severe TBI is ICP monitoring. The current gold standard is invasive, high-risk, requires a high level of expertise, is expensive and only accessible within a hospital setting. In addition to this the overwhelming majority of TBI cases are mild, the implications of which can be serious and long-term. Current gold standard monitoring is only used in severe cases, the diagnosis of mild TBI remains challenging due to its subjective nature and the reliance on self-reported symptoms. Currently there is no existing, non-invasive monitoring approach which is accurate and continuous.

There is a substantial and pressing need for the development of a continuous noninvasive, easy-to-use, inexpensive and efficacious monitoring device. Such a device would reduce the barrier to entry for ICP monitoring across all severities, providing immediate diagnosis of elevated ICP whilst effectively serving the currently undeserved majority of TBI cases. PPG signals are non-invasive and cost-effective. The wider body of research investigating the use of PPG signals for ICP monitoring is emerging and expanding. This work provides a credible and encouraging foundation for further research in this domain.

Appendix

Features
AUC
Systolic AUC
Diastolic AUC
Ratio between the systolic and diastolic AUC
Rise time (samples between the pulse onset and peak)
Decay time (samples between the pulse peak and end)
Ratio between the rise time and decay time
Number of beats
Inter-beat interval
Standard deviation of the inter-beat interval
Prominence
Upslope (slope between pulse onset to peak)
Downslope (slope between peak to pulse end)
Onset-end slope (slope between pulse onset and end)
Ratio between the upslope and downslope
Ratio between the pulse length and height
Start datum area (area between a straight line between the pulse onset and peak and pulse data between those points)
End datum area (area between a straight line between the pulse peak and end and pulse data between those points)
Ratio between the start datum area and end datum area
Max start datum difference (maximum element-wise difference between a straight line between the pulse onset and peak and pulse data between those points)
Max end datum difference (maximum element-wise difference between a straight line between the pulse peak and end and pulse data between those points)
The median of the element-wise difference between a straight line between the pulse onset and peak and pulse data between those points
The median of the element-wise difference between a straight line between the pulse peak and end and pulse data between those points
Pulse width at 10% of the pulse prominence (in samples)
Pulse width at 25% of the pulse prominence (in samples)
Pulse width at 50% of the pulse prominence (in samples)
Pulse width at 75% of the pulse prominence (in samples)
Pulse width at 100% of the pulse prominence (in samples)
Systolic width at 10% of the pulse prominence (in samples)
Systolic width at 25% of the pulse prominence (in samples)
Systolic width at 50% of the pulse prominence (in samples)
Systolic width at 75% of the pulse prominence (in samples)
Systolic width at 100% of the pulse prominence (in samples)
Diastolic width at 10% of the pulse prominence (in samples)
Diastolic width at 25% of the pulse prominence (in samples)
Diastolic width at 50% of the pulse prominence (in samples)
Diastolic width at 75% of the pulse prominence (in samples)
Diastolic width at 100% of the pulse prominence (in samples)
Ratio between the diastolic and systolic pulse width at 10% of the pulse prominence
Ratio between the diastolic and systolic pulse width at 25% of the pulse prominence
Ratio between the diastolic and systolic pulse width at 50% of the pulse prominence
Ratio between the diastolic and systolic pulse width at 75% of the pulse prominence
Ratio between the diastolic and systolic pulse width at 100% of the pulse prominence
Variance of the pulse data
Skew of the pulse data
Kurtosis of the pulse data
Zero-crossing rate of the pulse data

Table A1: Table containing the list of features extracted from the original, first derivative and second derivative of the signal.

References

[1] A. Brazinova, V. Rehorcikova, M. S. Taylor, *et al.*, "Epidemiology of traumatic brain injury in europe: A living systematic review," en, *J Neurotrauma*, vol. 38, no. 10, pp. 1411–1440, Dec. 2018.

[2] M. C. Dewan, A. Rattani, S. Gupta, *et al.*, "Estimating the global incidence of traumatic brain injury," en, *J Neurosurg*, pp. 1–18, Apr. 2018.

[3] J. T. J. M. van Dijck, M. D. Dijkman, R. H. Ophuis, G. C. W. de Ruiter, W. C. Peul, and S. Polinder, "Correction: In-hospital costs after severe traumatic brain injury: A systematic review and quality assessment," en, *PLoS One*, vol. 14, no. 7, e0219529, Jul. 2019.

[4] S. Badri, J. Chen, J. Barber, *et al.*, "Mortality and long-term functional outcome associated with intracranial pressure after traumatic brain injury," en, *Intensive Care Med*, vol. 38, no. 11, pp. 1800–1809, Aug. 2012.

[5] S. W. Bothwell, D. Janigro, and A. Patabendige, "Cerebrospinal fluid dynamics and intracranial pressure elevation in neurological diseases," en, *Fluids Barriers CNS*, vol. 16, no. 1, p. 9, Apr. 2019.

[6] J. Donnelly, M. Czosnyka, H. Adams, *et al.*, "Twenty-Five years of intracranial pressure monitoring after severe traumatic brain injury: A retrospective, Single-Center analysis," en, *Neurosurgery*, vol. 85, no. 1, E75–E82, Jul. 2019.

[7] www.ninds.nih.gov, Ed., *Traumatic brain injury (tbi)*, Nov. 2023. [Online]. Available: `https://www.ninds.nih.gov/health-information/disorders/traumatic-brain-injury-tbi`.

[8] A. I. R. Maas, D. K. Menon, G. T. Manley, *et al.*, "Traumatic brain injury: Progress and challenges in prevention, clinical care, and research," en, *Lancet Neurol*, Sep. 2022.

[9] D. I. Katz, S. I. Cohen, and M. P. Alexander, "Chapter 9 - mild traumatic brain injury," in *Traumatic Brain Injury, Part I*, ser. Handbook of Clinical Neurology, J. Grafman and A. M. Salazar, Eds., vol. 127, Elsevier, 2015, pp. 131–156. DOI: https://doi.org/10.1016/B978-0-444-52892-6.00009-X. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B978044452892600009X.

[10] N. D. Silverberg, G. L. Iverson, ACRM Brain Injury Special Interest Group Mild TBI Task Force members: *et al.*, "The american congress of rehabilitation medicine diagnostic criteria for mild traumatic brain injury," en, *Arch Phys Med Rehabil*, vol. 104, no. 8, pp. 1343–1355, May 2023.

[11] M. McCrea, T. Hammeke, G. Olsen, P. Leo, and K. Guskiewicz, "Unreported concussion in high school football players: Implications for prevention," en, *Clin J Sport Med*, vol. 14, no. 1, pp. 13–17, Jan. 2004.

[12] W. P. Meehan 3rd, R. C. Mannix, M. J. O'Brien, and M. W. Collins, "The prevalence of undiagnosed concussions in athletes," en, *Clin J Sport Med*, vol. 23, no. 5, pp. 339–342, Sep. 2013.

[13] T. Tanielian, L. H. Jaycox, T. L. Schell, *et al.*, "Invisible wounds: Mental health and cognitive care needs of america's returning veterans," 2008.

[14] B. M. Asken, M. A. McCrea, J. R. Clugston, A. R. Snyder, Z. M. Houck, and R. M. Bauer, ""playing through it": Delayed reporting and removal from athletic activity after concussion predicts prolonged recovery," en, *J Athl Train*, vol. 51, no. 4, pp. 329–335, Apr. 2016.

[15] H. F. Lingsma, J. K. Yue, A. I. R. Maas, E. W. Steyerberg, G. T. Manley, and TRACK-TBI Investigators, "Outcome prediction after mild and complicated mild traumatic brain injury: External validation of existing models and identification of new predictors using the TRACK-TBI pilot study," en, *J Neurotrauma*, vol. 32, no. 2, pp. 83–94, Nov. 2014.

[16] D. C. Voormolen, M. Zeldovich, J. A. Haagsma, *et al.*, "Outcomes after complicated and uncomplicated mild traumatic brain injury at three-and six-months post-injury: Results from the center-tbi study," *Journal of Clinical Medicine*, vol. 9,

no. 5, 2020, ISSN: 2077-0383. DOI: 10.3390/jcm9051525. [Online]. Available: https://www.mdpi.com/2077-0383/9/5/1525.

[17] M. Brasure, G. J. Lamberty, N. A. Sayer, *et al.*, *Multidisciplinary Postacute Rehabilitation for Moderate to Severe Traumatic Brain Injury in Adults*, en. Rockville (MD): Agency for Healthcare Research and Quality (US), Jun. 2012.

[18] L. Rangel-Castilla, S. Gopinath, and C. S. Robertson, "Management of intracranial hypertension," en, *Neurol Clin*, vol. 26, no. 2, pp. 521–41, x, May 2008.

[19] A. Naik, M. M. Bederson, D. Detchou, *et al.*, "Traumatic brain injury mortality and correlates in low- and Middle-Income countries: A Meta-Epidemiological study," en, *Neurosurgery*, vol. 93, no. 4, pp. 736–744, Apr. 2023.

[20] V. G. Coronado, L. C. McGuire, K. Sarmiento, *et al.*, "Trends in traumatic brain injury in the U.S. and the public health response: 1995-2009," en, *J Safety Res*, vol. 43, no. 4, pp. 299–307, Aug. 2012.

[21] M. Faul, M. M. Wald, L. Xu, and V. G. Coronado, "Traumatic brain injury in the united states: Emergency department visits, hospitalizations, and deaths, 2002-2006," 2010.

[22] N. Kureshi, M. Erdogan, G. Thibault-Halman, L. Fenerty, R. S. Green, and D. B. Clarke, "Long-Term trends in the epidemiology of major traumatic brain injury," en, *J Community Health*, vol. 46, no. 6, pp. 1197–1203, Jun. 2021.

[23] R. Y. Hsia, A. J. Markowitz, F. Lin, J. Guo, D. Y. Madhok, and G. T. Manley, "Ten-year trends in traumatic brain injury: A retrospective cohort study of california emergency department and hospital revisits and readmissions," en, *BMJ Open*, vol. 8, no. 12, e022297, Dec. 2018.

[24] GBD 2016 Traumatic Brain Injury and Spinal Cord Injury Collaborators, "Global, regional, and national burden of traumatic brain injury and spinal cord injury, 1990-2016: A systematic analysis for the global burden of disease study 2016," en, *Lancet Neurol*, vol. 18, no. 1, pp. 56–87, Nov. 2018.

[25] F. E. Lecky, O. Otesile, C. Marincowitz, *et al.*, "The burden of traumatic brain injury from low-energy falls among patients from 18 countries in the CENTER-TBI registry: A comparative cohort study," en, *PLoS Med*, vol. 18, no. 9, e1003761, Sep. 2021.

[26] D. Waltzman, J. Haarbauer-Krupa, and L. S. Womack, "Traumatic brain injury in older Adults-A public health perspective," en, *JAMA Neurol*, vol. 79, no. 5, pp. 437–438, May 2022.

[27] C. Yang, L. Lang, Z. He, *et al.*, "Epidemiological characteristics of older patients with traumatic brain injury in china," en, *J Neurotrauma*, vol. 39, no. 11-12, pp. 850–859, Mar. 2022.

[28] M. Majdan, J. Melichova, D. Plancikova, *et al.*, "Burden of traumatic brain injuries in children and adolescents in europe: Hospital discharges, deaths and years of life lost," en, *Children (Basel)*, vol. 9, no. 1, Jan. 2022.

[29] J. van Ierssel, M. Osmond, J. Hamid, M. Sampson, and R. Zemek, "What is the risk of recurrent concussion in children and adolescents aged 5-18 years? a systematic review and meta-analysis," en, *Br J Sports Med*, vol. 55, no. 12, pp. 663–669, Oct. 2020.

[30] G. Gao, X. Wu, J. Feng, *et al.*, "Clinical characteristics and outcomes in patients with traumatic brain injury in china: A prospective, multicentre, longitudinal, observational study," en, *Lancet Neurol*, vol. 19, no. 8, pp. 670–677, Aug. 2020.

[31] D. F. Mackay, E. R. Russell, K. Stewart, J. A. MacLean, J. P. Pell, and W. Stewart, "Neurodegenerative disease mortality among former professional soccer players," en, *N Engl J Med*, vol. 381, no. 19, pp. 1801–1808, Oct. 2019.

[32] A. Theadom, S. Mahon, P. Hume, *et al.*, "Incidence of Sports-Related traumatic brain injury of all severities: A systematic review," en, *Neuroepidemiology*, vol. 54, no. 2, pp. 192–199, Jan. 2020.

[33] J. R. Marin, M. D. Weaver, and R. C. Mannix, "Burden of USA hospital charges for traumatic brain injury," en, *Brain Inj*, vol. 31, no. 1, pp. 24–31, Nov. 2016.

[34] B. Te Ao, P. Brown, M. Tobias, *et al.*, "Cost of traumatic brain injury in new zealand: Evidence from a population-based study," en, *Neurology*, vol. 83, no. 18, pp. 1645–1652, Sep. 2014.

[35] S. C. Stein, P. Georgoff, S. Meghan, K. Mizra, and S. S. Sonnad, "150 years of treating severe traumatic brain injury: A systematic review of progress in mortality," en, *J Neurotrauma*, vol. 27, no. 7, pp. 1343–1353, Jul. 2010.

[36] S. Rajesh, D. Wonderling, I. Bernstein, C. Balson, and F. Lecky, "Head injury: Assessment and early management—summary of updated nice guidance," *BMJ*, vol. 381, A. Helmy, *et al.*, Eds., 2023. DOI: 10.1136/bmj.p1130. eprint: https://www.bmj.com/content/381/bmj.p1130.full.pdf. [Online]. Available: https://www.bmj.com/content/381/bmj.p1130.

[37] G. Teasdale and B. Jennett, "Assessment of coma and impaired consciousness. a practical scale," en, *Lancet*, vol. 2, no. 7872, pp. 81–84, Jul. 1974.

[38] T. A. Gennarelli, H. R. Champion, W. S. Copes, and W. J. Sacco, "Comparison of mortality, morbidity, and severity of 59,713 head injured patients with 114,447 patients with extracranial injuries," en, *J Trauma*, vol. 37, no. 6, pp. 962–968, Dec. 1994.

[39] A. B. Newberg and A. Alavi, "Neuroimaging in patients with head injury," en, *Semin Nucl Med*, vol. 33, no. 2, pp. 136–147, Apr. 2003.

[40] S. C. Wei, S. Ulmer, M. H. Lev, S. R. Pomerantz, R. G. González, and J. W. Henson, "Value of coronal reformations in the CT evaluation of acute head trauma," en, *AJNR Am J Neuroradiol*, vol. 31, no. 2, pp. 334–339, Oct. 2009.

[41] T. T. Zacharia and D. T. D. Nguyen, "Subtle pathology detection with multidetector row coronal and sagittal CT reformations in acute head trauma," en, *Emerg Radiol*, vol. 17, no. 2, pp. 97–102, Oct. 2009.

[42] P. M. Parizel and C. D. Philips, "Traumatic neuroemergency: Imaging patients with traumatic brain injury—an introduction," in *Diseases of the Brain, Head and Neck, Spine 2020–2023: Diagnostic Imaging*, J. Hodler, R. A. Kubik-Huch, and G. K. von Schulthess, Eds. Cham: Springer International Publishing, 2020, pp. 77–92, ISBN: 978-3-030-38490-6. DOI: 10.1007/978-3-030-38490-6_7. [Online]. Available: https://doi.org/10.1007/978-3-030-38490-6_7.

[43] B. Mokri, "The Monro-Kellie hypothesis: Applications in CSF volume depletion," en, *Neurology*, vol. 56, no. 12, pp. 1746–1748, Jun. 2001.

[44] O. B. Paulson, S. Strandgaard, and L. Edvinsson, "Cerebral autoregulation," en, *Cerebrovasc Brain Metab Rev*, vol. 2, no. 2, pp. 161–192, 1990.

[45] D. DeMers and D. Wachs, "Physiology, mean arterial pressure," en, in *StatPearls*, Treasure Island (FL): StatPearls Publishing, Jan. 2024.

[46] O. B. Paulson, S. Strandgaard, and L. Edvinsson, "Cerebral autoregulation," en, *Cerebrovasc Brain Metab Rev*, vol. 2, no. 2, pp. 161–192, 1990.

[47] J. Ghajar, "Traumatic brain injury," en, *Lancet*, vol. 356, no. 9233, pp. 923–929, Sep. 2000.

[48] R. D. Stevens, M. Shoykhet, and R. Cadena, "Emergency neurological life support: Intracranial hypertension and herniation," en, *Neurocrit Care*, vol. 23 Suppl 2, no. Suppl 2, S76–82, Dec. 2015.

[49] C. A. Adams, D. M. Stein, J. J. Morrison, and T. M. Scalea, "Does intracranial pressure management hurt more than it helps in traumatic brain injury?" en, *Trauma Surg Acute Care Open*, vol. 3, no. 1, e000142, Jan. 2018.

[50] D. C. Engel, A. Mikocka-Walus, P. A. Cameron, and M. Maegele, "Pre-hospital and in-hospital parameters and outcomes in patients with traumatic brain injury: A comparison between german and australian trauma registries," en, *Injury*, vol. 41, no. 9, pp. 901–906, Jan. 2010.

[51] N. Carney, A. M. Totten, C. O'Reilly, *et al.*, "Guidelines for the management of severe traumatic brain injury, fourth edition," en, *Neurosurgery*, vol. 80, no. 1, pp. 6–15, Jan. 2017.

[52] J. Guillaume and P. Janny, "[continuous intracranial manometry; importance of the method and first results]," *Rev Neurol (Paris)*, vol. 84, no. 2, pp. 131–142, Feb. 1951.

[53] N. Lundberg, "Continuous recording and control of ventricular fluid pressure in neurosurgical practice," en, *Acta Psychiatr Scand Suppl*, vol. 36, no. 149, pp. 1–193, 1960.

[54] R. Muralidharan, "External ventricular drains: Management and complications," en, *Surg Neurol Int*, vol. 6, no. Suppl 6, S271–4, May 2015.

[55] D. R. Huyette, B. J. Turnbow, C. Kaufman, D. F. Vaslow, B. B. Whiting, and M. Y. Oh, "Accuracy of the freehand pass technique for ventriculostomy catheter placement: Retrospective assessment using computed tomography scans," en, *J Neurosurg*, vol. 108, no. 1, pp. 88–91, Jan. 2008.

[56] A. Quiñones-Hinojosa, *Schmidek and Sweet Operative Neurosurgical Techniques: Indications, Methods, and Results: Sixth Edition* (Schmidek and Sweet Operative Neurosurgical Techniques: Indications, Methods, and Results: Sixth Edition). 2012, vol. 1-2.

[57] C. Chau, C. Craven, A. Rubiano Escobar, *et al.*, "The evolution of the role of external ventricular drainage in traumatic brain injury," *Journal of Clinical Medicine*, vol. 8, p. 1422, Sep. 2019.

[58] *Neurology update*, https://mmcneuro.wordpress.com/2013/03/24/abraham-lincolns-ventriculostomy/, Accessed: 2024-04-2024.

[59] C. Miller and R. P. Tummala, "Risk factors for hemorrhage associated with external ventricular drain placement and removal," en, *J Neurosurg*, vol. 126, no. 1, pp. 289–297, Apr. 2016.

[60] A. Saladino, J. White, E. Wijdicks, and G. Lanzino, "Malplacement of ventricular catheters by neurosurgeons: A single institution experience," *Neurocritical care*, vol. 10, pp. 248–52, Nov. 2008. DOI: 10.1007/s12028-008-9154-z.

[61] S. Hagel, T. Bruns, M. W. Pletz, C. Engel, R. Kalff, and C. Ewald, "External ventricular drain infections: Risk factors and outcome," en, *Interdiscip Perspect Infect Dis*, vol. 2014, p. 708 531, Nov. 2014.

[62] K. Wright, P. Young, C. Brickman, *et al.*, "Rates and determinants of ventriculostomy-related infections during a hospital transition to use of antibiotic-coated external ventricular drains," en, *Neurosurg Focus*, vol. 34, no. 5, E12, May 2013.

[63] Q. Aten, J. Killeffer, C. Seaver, and L. Reier, "Causes, complications, and costs associated with external ventricular drainage catheter obstruction," *World Neurosurgery*, vol. 134, pp. 501–506, 2020, ISSN: 1878-8750. DOI: https://doi.org/10.1016/j.wneu.2019.10.105. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1878875019327251.

[64] P. Le Roux, "Intracranial pressure monitoring and management," en, in *Translational Research in Traumatic Brain Injury*, Boca Raton (FL): CRC Press/Taylor and Francis Group, 2016.

[65] U. Kawoos, R. M. McCarron, C. R. Auker, and M. Chavko, "Advances in intracranial pressure monitoring and its significance in managing traumatic brain injury," en, *Int J Mol Sci*, vol. 16, no. 12, pp. 28 979–28 997, Dec. 2015.

[66] E. Münch, R. Weigel, P. Schmiedek, and L. Schürer, "The camino intracranial pressure device in clinical practice: Reliability, handling characteristics and complications," en, *Acta Neurochir (Wien)*, vol. 140, no. 11, 1113–9, discussion 1119–20, 1998.

[67] A. Bekar, S. Doğan, F. Abaş, *et al.*, "Risk factors and complications of intracranial pressure monitoring with a fiberoptic device," en, *J Clin Neurosci*, vol. 16, no. 2, pp. 236–240, Dec. 2008.

[68] W.-C. Hong, Y.-K. Tu, Y.-S. Chen, L.-M. Lien, and S.-J. Huang, "Subdural intracranial pressure monitoring in severe head injury: Clinical experience with the codman MicroSensor," en, *Surg Neurol*, vol. 66 Suppl 2, S8–S13, 2006.

[69] L.-O. D. Koskinen and M. Olivecrona, "Clinical experience with the intraparenchymal intracranial pressure monitoring codman MicroSensor system," en, *Neurosurgery*, vol. 56, no. 4, 693–8, discussion 693–8, Apr. 2005.

[70] J.-M. Lang, J. Beck, M. Zimmermann, V. Seifert, and A. Raabe, "Clinical evaluation of intraparenchymal spiegelberg pressure sensor," en, *Neurosurgery*, vol. 52, no. 6, 1455–9, discussion 1459, Jun. 2003.

[71] Y. H. Yau, I. R. Piper, R. E. Clutton, and I. R. Whittle, "Experimental evaluation of the spiegelberg intracranial pressure and intracranial compliance monitor. technical note," en, *J Neurosurg*, vol. 93, no. 6, pp. 1072–1077, Dec. 2000.

[72] S. L. Bratton, R. M. Chestnut, J. Ghajar, *et al.*, "Vii. intracranial pressure monitoring technology," *Journal of Neurotrauma*, vol. 24, no. supplement 1, S-45-S–54, 2007. DOI: 10.1089/neu.2007.9989. eprint: https://doi.org/10.1089/neu.2007.9989. [Online]. Available: https://doi.org/10.1089/neu.2007.9989.

[73] L. D. DeWitt and L. R. Wechsler, "Transcranial doppler," en, *Stroke*, vol. 19, no. 7, pp. 915–921, Jul. 1988.

[74] S. Purkayastha and F. Sorond, "Transcranial doppler ultrasound: Technique and application," en, *Semin Neurol*, vol. 32, no. 4, pp. 411–420, Jan. 2013.

[75] S.-J. Lee, J. Ahn, and J. M. Hong, "Neurosonology for unconscious or neurocritically ill patients," *Journal of Neurosonology and Neuroimaging*, vol. 11, pp. 46–61, Jun. 2019. DOI: 10.31728/jnn.2019.00048.

[76] S. G. Voulgaris, M. Partheni, H. Kaliora, N. Haftouras, I. S. Pessach, and K. S. Polyzoidis, "Early cerebral monitoring using the transcranial doppler pulsatility index in patients with severe brain trauma," en, *Med Sci Monit*, vol. 11, no. 2, CR49–52, Feb. 2005.

[77] J. Bellner, B. Romner, P. Reinstrup, K.-A. Kristiansson, E. Ryding, and L. Brandt, "Transcranial doppler sonography pulsatility index (PI) reflects intracranial pressure (ICP)," en, *Surg Neurol*, vol. 62, no. 1, 45–51, discussion 51, Jul. 2004.

[78] H. Maeda, H. Etani, N. Handa, *et al.*, "A validation study on the reproducibility of transcranial doppler velocimetry," en, *Ultrasound Med Biol*, vol. 16, no. 1, pp. 9–14, 1990.

[79] R. W. Baumgartner, J. Mathis, M. Sturzenegger, and H. P. Mattle, "A validation study on the intraobserver reproducibility of transcranial color-coded duplex sonography velocity measurements," en, *Ultrasound Med Biol*, vol. 20, no. 3, pp. 233–237, 1994.

[80] C. J. McMahon, P. McDermott, D. Horsfall, J. R. Selvarajah, A. T. King, and A. Vail, "The reproducibility of transcranial doppler middle cerebral artery velocity measurements: Implications for clinical practice," en, *Br J Neurosurg*, vol. 21, no. 1, pp. 21–27, Feb. 2007.

[81] Q. Shen, J. Stuart, B. Venkatesh, J. Wallace, and J. Lipman, "Inter observer variability of the transcranial doppler ultrasound technique: Impact of lack of practice on the accuracy of measurement," en, *J Clin Monit Comput*, vol. 15, no. 3-4, pp. 179–184, May 1999.

[82] G. Tsivgoulis, A. V. Alexandrov, and M. A. Sloan, "Advances in transcranial doppler ultrasonography," en, *Curr Neurol Neurosci Rep*, vol. 9, no. 1, pp. 46–54, Jan. 2009.

[83] K. Helmke and H. C. Hansen, "Fundamentals of transorbital sonographic evaluation of optic nerve sheath expansion under intracranial hypertension. i. experimental study," en, *Pediatr Radiol*, vol. 26, no. 10, pp. 701–705, Oct. 1996.

[84] H. C. Hansen and K. Helmke, "Validation of the optic nerve sheath response to changing cerebrospinal fluid pressure: Ultrasound findings during intrathecal infusion tests," en, *J Neurosurg*, vol. 87, no. 1, pp. 34–40, Jul. 1997.

[85] J. P. Jeon, S. U. Lee, S.-E. Kim, *et al.*, "Correlation of optic nerve sheath diameter with directly measured intracranial pressure in korean adults using bedside ultrasonography," en, *PLoS One*, vol. 12, no. 9, e0183170, Sep. 2017.

[86] V. Rajajee, M. Vanaman, J. J. Fletcher, and T. L. Jacobs, "Optic nerve ultrasound for the detection of raised intracranial pressure," en, *Neurocrit Care*, vol. 15, no. 3, pp. 506–515, Dec. 2011.

[87] J. Dubourg, E. Javouhey, T. Geeraerts, M. Messerer, and B. Kassai, "Ultrasonography of optic nerve sheath diameter for detection of raised intracranial pressure: A systematic review and meta-analysis," en, *Intensive Care Med*, vol. 37, no. 7, pp. 1059–1068, Apr. 2011.

[88] T. Cimilli Ozturk, H. Demir, R. Yorulmaz, S. Ozdemir, G. Isat, and O. Ecmel Onur, "Assessment of intra-interobserver reliability of the sonographic optic nerve sheath diameter measurement," en, *Kaohsiung J Med Sci*, vol. 31, no. 8, pp. 432–436, Jul. 2015.

[89] O. S. Aduayi, C. M. Asaleye, V. A. Adetiloye, E. O. Komolafe, and V. A. Aduayi, "Optic nerve sonography: A noninvasive means of detecting raised intracranial pressure in a resource-limited setting," en, *J Neurosci Rural Pract*, vol. 6, no. 4, pp. 563–567, Oct. 2015.

[90] S. Ballantyne, G. O'Neill, R. Hamilton, and A. Hollman, "Observer variation in the sonographic measurement of optic nerve sheath diameter in normal adults," *European Journal of Ultrasound*, vol. 15, no. 3, pp. 145–149, 2002, ISSN: 0929-8266. DOI: https://doi.org/10.1016/S0929-8266(02)00036-8. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0929826602000368.

[91] P. H. Raboel, J. Bartek Jr, M. Andresen, B. M. Bellander, and B. Romner, "Intracranial pressure monitoring: Invasive versus Non-Invasive Methods-A review," en, *Crit Care Res Pract*, vol. 2012, p. 950 393, Jun. 2012.

[92] A. Kampfl, B. Pfausler, D. Denchev, H. P. Jaring, and E. Schmutzhard, "Near infrared spectroscopy (nirs) in patients with severe brain injury and elevated intracranial pressure," in *Brain Edema X*, H. E. James, L. F. Marshall, H. J. Raulen, *et al.*, Eds., Vienna: Springer Vienna, 1997, pp. 112–114, ISBN: 978-3-7091-6837-0.

[93] A. Ruesch, S. Schmitt, J. Yang, M. A. Smith, and J. M. Kainerstorfer, "Fluctuations in intracranial pressure can be estimated non-invasively using near-infrared spectroscopy in non-human primates," *Journal of Cerebral Blood Flow & Metabolism*, vol. 40, no. 11, pp. 2304–2314, 2020, PMID: 31775565. DOI: 10.1177/0271678X19891359.

[94] A. Ruesch, J. Yang, S. Schmitt, D. Acharya, M. A. Smith, and J. M. Kainerstorfer, "Estimating intracranial pressure using pulsatile cerebral blood flow measured with diffuse correlation spectroscopy," en, *Biomed Opt Express*, vol. 11, no. 3, pp. 1462–1476, Feb. 2020.

[95] F. A. J. Relander, A. Ruesch, J. Yang, *et al.*, "Using near-infrared spectroscopy and a random forest regressor to estimate intracranial pressure," *Neurophotonics*, vol. 9, no. 4, p. 045 001, Oct. 2022.

[96] B. Dixon, J. M. Sharkey, E. J. Teo, *et al.*, "Assessment of a Non-Invasive brain pulse monitor to measure Intra-Cranial pressure following acute brain injury," en, *Med Devices (Auckl)*, vol. 16, pp. 15–26, Jan. 2023.

[97] A. V. Creaser, S. A. Clemes, S. Costa, *et al.*, "The acceptability, feasibility, and effectiveness of wearable activity trackers for increasing physical activity in children and adolescents: A systematic review," en, *Int J Environ Res Public Health*, vol. 18, no. 12, Jun. 2021.

[98] D. Castaneda, A. Esparza, M. Ghamari, C. Soltanpur, and H. Nazeran, "A review on wearable photoplethysmography sensors and their potential future applications in health care," en, *Int J Biosens Bioelectron*, vol. 4, no. 4, pp. 195–202, Aug. 2018.

[99] P. Kyriacou and J. Allen, *Photoplethysmography: Technology, Signal Analysis and Applications*. Elsevier Science, 2021, ISBN: 9780128235256. [Online]. Available: https://books.google.co.uk/books?id=d8wnEAAAQBAJ.

[100] J. Moyle, "Pulse Oximetry, 2nd Edn," *BJA: British Journal of Anaesthesia*, vol. 89, no. 5, pp. 802–803, Nov. 2002, ISSN: 0007-0912. DOI: 10.1093/bja/89.5.802.

eprint: https://academic.oup.com/bja/article-pdf/89/5/802/696342/aef568.pdf. [Online]. Available: https://doi.org/10.1093/bja/89.5.802.

[101] M. Roldan, S. Chatterjee, and P. A. Kyriacou, "Brain Light-Tissue interaction modelling: Towards a non-invasive sensor for traumatic brain injury," en, *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2021, pp. 1292–1296, Nov. 2021.

[102] M. Roldan and P. A. Kyriacou, "A non-invasive optical multimodal photoplethysmography-near infrared spectroscopy sensor for measuring intracranial pressure and cerebral oxygenation in traumatic brain injury," *Applied Sciences*, vol. 13, no. 8, 2023, ISSN: 2076-3417. DOI: 10.3390/app13085211. [Online]. Available: https://www.mdpi.com/2076-3417/13/8/5211.

[103] J. M. Murkin and M. Arango, "Near-infrared spectroscopy as an index of brain and tissue oxygenation," en, *Br J Anaesth*, vol. 103 Suppl 1, pp. i3–13, Dec. 2009.

[104] G. R. Bradley and P. A. Kyriacou, "Opening the envelope: Efficient envelope-based ppg denoising algorithm," *Biomedical Signal Processing and Control*, vol. 88, p. 105 693, 2024, ISSN: 1746-8094. DOI: https://doi.org/10.1016/j.bspc.2023.105693. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809423011266.

[105] E. Mejía-Mejía, J. Allen, K. Budidha, C. El-Hajj, P. A. Kyriacou, and P. H. Charlton, "4 - photoplethysmography signal processing and synthesis," in *Photoplethysmography*, J. Allen and P. Kyriacou, Eds., Academic Press, 2022, pp. 69–146, ISBN: 978-0-12-823374-0. DOI: https://doi.org/10.1016/B978-0-12-823374-0.00015-3. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128233740000153.

[106] J. Lee, S. Sun, S. M. Yang, *et al.*, "Bidirectional recurrent Auto-Encoder for photoplethysmogram denoising," en, *IEEE J Biomed Health Inform*, vol. 23, no. 6, pp. 2375–2385, Dec. 2018.

[107] J. H. Kwon, S. E. Kim, N. H. Kim, E. C. Lee, and J. H. Lee, "Preeminently robust neural PPG denoiser," en, *Sensors (Basel)*, vol. 22, no. 6, Mar. 2022.

[108] F. Mohagheghian, D. Han, O. Ghetia, *et al.*, "Noise reduction in photoplethysmography signals using a convolutional denoising autoencoder with unconventional training scheme," en, *IEEE Trans Biomed Eng*, vol. PP, Sep. 2023.

[109] K. Xu, X. Jiang, and W. Chen, "Photoplethysmography motion artifacts removal based on signal-noise interaction modeling utilizing envelope filtering and time-delay neural network," *IEEE Sensors Journal*, vol. 20, no. 7, pp. 3732–3744, 2020. DOI: 10.1109/JSEN.2019.2960370.

[110] W.-H. Lin, N. Ji, L. Wang, and G. Li, "A characteristic filtering method for pulse wave signal quality assessment," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 603–606. DOI: 10.1109/EMBC.2019.8856811.

[111] D. Dao, S. M. A. Salehizadeh, Y. Noh, *et al.*, "A robust motion artifact detection algorithm for accurate detection of heart rates from photoplethysmographic signals using Time-Frequency spectral features," en, *IEEE J Biomed Health Inform*, vol. 21, no. 5, pp. 1242–1253, Oct. 2016.

[112] M. Raghu Ram, K. Venu Madhav, E. Hari Krishna, K. Nagarjuna Reddy, and K. Ashoka Reddy, "Computation of spo2 using non-parametric spectral estimation methods from wavelet based motion artifact reduced ppg signals," in *2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies*, 2011, pp. 776–780. DOI: 10.1109/ICSCCN.2011.6024656.

[113] M. R. Ram, K. V. Madhav, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "A novel approach for motion artifact reduction in ppg signals based on as-lms adaptive filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 5, pp. 1445–1457, 2012. DOI: 10.1109/TIM.2011.2175832.

[114] M. Merino-Monge, J. A. Castro-García, C. Lebrato-Vázquez, I. M. Gómez-González, and A. J. Molina-Cantero, "Heartbeat detector from ECG and PPG signals based on wavelet transform and upper envelopes," *Physical and Engineering Sciences in Medicine*, vol. 46, no. 2, pp. 597–608, Jun. 2023.

[115] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020. DOI: 10.1038/s41592-019-0686-2.

[116] M. Rinkevičius, P. H. Charlton, R. Bailón, and V. Marozas, "Influence of photoplethysmogram signal quality on pulse arrival time during polysomnography," en, *Sensors (Basel)*, vol. 23, no. 4, Feb. 2023.

[117] M. Elgendi, "Optimal signal quality index for photoplethysmogram signals," en, *Bioengineering (Basel)*, vol. 3, no. 4, Sep. 2016.

[118] G. D. Clifford, J. Behar, Q. Li, and I. Rezek, "Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms," en, *Physiol Meas*, vol. 33, no. 9, pp. 1419–1433, Aug. 2012.

[119] T. Y. Abay and P. A. Kyriacou, "5 - photoplethysmography in oxygenation and blood volume measurements," in *Photoplethysmography*, J. Allen and P. Kyriacou, Eds., Academic Press, 2022, pp. 147–188, ISBN: 978-0-12-823374-0. DOI: https://doi.org/10.1016/B978-0-12-823374-0.00003-7. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128233740000037.

[120] J. de Trafford and K. Lafferty, "What does photoplethysmography measure?" en, *Med Biol Eng Comput*, vol. 22, no. 5, pp. 479–480, Sep. 1984.

[121] A. A. Kamal, J. B. Harness, G. Irving, and A. J. Mearns, "Skin photoplethysmography–a review," en, *Comput Methods Programs Biomed*, vol. 28, no. 4, pp. 257–269, Apr. 1989.

[122] L.-G. Lindberg and P. A. Oberg, "Optical properties of blood in motion," *Optical Engineering*, vol. 32, pp. 253–257, Feb. 1993. DOI: 10.1117/12.60688.

[123] W. B. Murray and P. A. Foster, "The peripheral pulse wave: Information overlooked," en, *J Clin Monit*, vol. 12, no. 5, pp. 365–377, Sep. 1996.

[124] A. A. Awad, A. S. Haddadin, H. Tantawy, *et al.*, "The relationship between the photoplethysmographic waveform and systemic vascular resistance," en, *J Clin Monit Comput*, vol. 21, no. 6, pp. 365–372, Oct. 2007.

[125] L. Wang, E. Pickwell-MacPherson, Y. P. Liang, and Y. T. Zhang, "Noninvasive cardiac output estimation using a novel photoplethysmogram index," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2009, pp. 1746–1749. DOI: 10.1109/IEMBS.2009.5333091.

[126] R. Ferizoli, P. Karimpour, J. M. May, and P. A. Kyriacou, "Arterial stiffness assessment using PPG feature extraction and significance testing in an in vitro cardiovascular system," *Scientific Reports*, vol. 14, no. 1, p. 2024, Jan. 2024.

[127] J. M. Ahn, "New aging index using signal features of both photoplethysmograms and acceleration plethysmograms," en, *Healthc Inform Res*, vol. 23, no. 1, pp. 53–59, Jan. 2017.

[128] N. Selvaraj, A. K. Jaryal, J. Santhosh, S. Anand, and K. Deepak, "Monitoring of reactive hyperemia using photoplethysmographic pulse amplitude and transit time," *Journal of Clinical Monitoring and Computing*, vol. 23, pp. 315–22, Sep. 2009. DOI: 10.1007/s10877-009-9199-3.

[129] K. Takazawa, N. Tanaka, M. Fujita, *et al.*, "Assessment of vasoactive agents and vascular aging by the second derivative of photoplethysmogram waveform," en, *Hypertension*, vol. 32, no. 2, pp. 365–370, Aug. 1998.

[130] H. J. Baek, J. S. Kim, Y. S. Kim, H. B. Lee, and K. S. Park, "Second derivative of photoplethysmography for estimating vascular aging," in *2007 6th International Special Topic Conference on Information Technology Applications in Biomedicine*, 2007, pp. 70–72. DOI: 10.1109/ITAB.2007.4407346.

[131] M. Kachuee, M. M. Kiani, H. Mohammadzade, and M. Shabany, "Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time," in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2015, pp. 1006–1009. DOI: 10.1109/ISCAS.2015.7168806.

[132] M. H. Chowdhury, M. N. I. Shuzan, M. E. H. Chowdhury, *et al.*, "Estimating blood pressure from the photoplethysmogram signal and demographic features using machine learning techniques," en, *Sensors (Basel)*, vol. 20, no. 11, Jun. 2020.

[133] J. Yi Kim, B. Hwan Cho, S. Mi Im, M. Ju Jeon, I. Young Kim, and S. Kim, "Comparative study on artificial neural network with multiple regressions for continuous estimation of blood pressure," en, *Conf Proc IEEE Eng Med Biol Soc*, vol. 2005, pp. 6942–6945, 2005.

[134] F. S. Cattivelli and H. Garudadri, "Noninvasive cuffless estimation of blood pressure from pulse arrival time and heart rate with adaptive calibration," in *2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks*, 2009, pp. 114–119. DOI: 10.1109/BSN.2009.35.

[135] Y. Zhang and Z. Feng, "A svm method for continuous blood pressure estimation from a ppg signal," in *Proceedings of the 9th International Conference on Machine Learning and Computing*, ser. ICMLC '17, Singapore, Singapore: Association for Computing Machinery, 2017, pp. 128–132, ISBN: 9781450348171. DOI: 10.1145/3055635.3056634. [Online]. Available: https://doi.org/10.1145/3055635.3056634.

[136] C. El Hajj and P. A. Kyriacou, "Cuffless and continuous blood pressure estimation from PPG signals using recurrent neural networks," en, *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2020, pp. 4269–4272, Jul. 2020.

[137] Y. Kurylyak, F. Lamonaca, and D. Grimaldi, "A neural network-based method for continuous blood pressure estimation from a ppg signal," in *2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, 2013, pp. 280–283. DOI: 10.1109/I2MTC.2013.6555424.

[138] P. S. Addison, "Slope transit time (stt): A pulse transit time proxy requiring only a single signal fiducial point," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 11, pp. 2441–2444, 2016. DOI: 10.1109/TBME.2016.2528507.

[139] Y.-Y. Chen, C.-L. Lin, Y.-C. Lin, and C. Zhao, "Non-invasive detection of alcohol concentration based on photoplethysmogram signals," *IET Image Processing*, vol. 12, no. 2, pp. 188–193, 2018. DOI: https://doi.org/10.1049/iet-ipr.2017.0625. eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/iet-ipr.2017.0625. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-ipr.2017.0625.

[140] H. Adams, J. Donnelly, M. Czosnyka, *et al.*, "Temporal profile of intracranial pressure and cerebrovascular reactivity in severe traumatic brain injury and association with fatal outcome: An observational study," *PLOS Medicine*, vol. 14, pp. 1–21, Jul. 2017. DOI: 10.1371/journal.pmed.1002353. [Online]. Available: https://doi.org/10.1371/journal.pmed.1002353.

[141] J. Ghajar, "Traumatic brain injury," en, *Lancet*, vol. 356, no. 9233, pp. 923–929, Sep. 2000.

[142] E. R. Cardoso, J. O. Rowan, and S. Galbraith, "Analysis of the cerebrospinal fluid pulse wave in intracranial pressure," en, *J Neurosurg*, vol. 59, no. 5, pp. 817–821, Nov. 1983.

[143] T. Mitchell, *Machine Learning* (McGraw-Hill International Editions). McGraw-Hill, 1997, ISBN: 9780071154673. [Online]. Available: https://books.google.co.uk/books?id=EoYBngEACAAJ.

[144] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American Statistical Association*, vol. 53, no. 282, pp. 457–481, 1958, ISSN: 01621459. [Online]. Available: http://www.jstor.org/stable/2281868 (visited on 04/15/2024).

[145] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[146] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[147] J. H. Friedman, "Greedy function approximation: A gradient boosting machine.," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001. DOI: 10.1214/aos/1013203451. [Online]. Available: https://doi.org/10.1214/aos/1013203451.

[148] G. Ke, Q. Meng, T. Finley, *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17, Long Beach, California, USA: Curran Associates Inc., 2017, pp. 3149–3157, ISBN: 9781510860964.

[149] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16, San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 785–794, ISBN: 9781450342322. DOI: 10.1145/2939672.2939785. [Online]. Available: https://doi.org/10.1145/2939672.2939785.

[150] G. R. Bradley and P. A. Kyriacou, "Evaluating the effectiveness of non-invasive intracranial pressure monitoring via near-infrared photoplethysmography using classical machine learning methods," *Biomedical Signal Processing and Control*, vol. 96, p. 106 517, 2024, ISSN: 1746-8094. DOI: https://doi.org/10.1016/j.bspc.

2024.106517. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809424005755.

[151] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 2623–2631.

[152] K. J. Dixon, "Pathophysiology of traumatic brain injury," en, *Phys Med Rehabil Clin N Am*, vol. 28, no. 2, pp. 215–225, Mar. 2017.