FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Continuous EEG monitoring for the Prediction of the Outcome of Traumatic Brain Injury

Inês Cabaço da Silva Arriaga

Mestrado em Bioengenharia

Co-Supervisor: Miguel Velhote Correia (FEUP)

Co-Supervisor: Michel van Putten and
Marleen Tjepkema-Cloostermans (CNPH)

October 31, 2022

Continuous EEG monitoring for the Prediction of the Outcome of Traumatic Brain Injury

Inês Cabaço da Silva Arriaga

Mestrado em Bioengenharia

Approved in oral examination by the committee:

Chair: Prof. João Paulo Cunha
External Examiner: Prof. Maria Carla Piastra
Supervisor: Prof. Miguel Correia

October 31, 2022

Abstract

Background and Objectives. The clinical severity of Traumatic Brain Injury (TBI) ranges from mild to severe, being one of the primary causes of disability in the world and responsible for nearly 50% of deaths from trauma. There is a need for trustworthy methods to better predict the outcome of TBI and aid doctors' decision-making. However, it is difficult due to the injury's heterogeneity. Developments in this area could improve healthcare, leading to changes in treatment, and avoiding unnecessary spending. We aimed at creating a Machine Learning model capable of predicting the long-term neurological outcome of patients with a TBI, by assessing the dynamics over time of the continuous electroencephalogram (EEG) during the patient's stay in the Intensive Care Unit (ICU).

Methods. We performed continuous EEG in 111 patients with moderate to severe TBI after admission to the ICU, continued for 7 days or until discharge or death. Demographic, clinic, laboratory, radiologic and other trauma-related parameters were collected. Neurological outcome at 12 months was assessed using the Extended Glasgow Outcome Scale (GOSE) dichotomized into poor (GOSE 1-3) or good (GOSE 4-8). We extracted EEG features related to power, continuity, synchrony, symmetry, and randomness for each hour. We evaluated trends over time of the EEG features using linear and non-linear regressions in 15 possible time intervals. A Random Forest classifier with 5-fold cross-validation was trained using admission parameters and the time-dependent regressions' features for binary prediction of the outcome. Feature selection included a filter based on ranking and backward elimination. We evaluated and compared models with different inputs using sensitivity, specificity and the AUC of the ROC curve.

Results. The best two prediction models were found using EEG features from 12 to 36 hours (AUC = 0.83 [0.75-0.91], sensitivity = 0.83 [0.72-0.94], and specificity = 0.80 [0.61-0.94]) or from 48 to 72h (AUC = 0.85 [0.73-0.97], sensitivity = 0.91 [0.83-1.00], and specificity = 0.67 [0.32-1.00]) combined with predictors collected at the time of admission. The combination of admission predictors with the EEG features outperformed models trained with only admission or EEG features separately. The findings also hint at better performance using EEG trends compared to features from single time points.

Discussion. This study allows for an early prediction of the outcome of TBI resorting to a machine learning approach. EEG trends over time point to added value to the prediction, highlighting the importance of monitoring continuously the EEG and encouraging its use in severe TBI patients, as it is frequently done in other brain injury populations. We suggest that future works explore the possibility of using a multi-class prediction. Furthermore, we also suggest training other algorithms than Random Forest, which if possible allows understanding of the features' contribution to the prediction.

Keywords: Traumatic Brain Injury, Eletroencephalogram, Machine Learning, Random Forest, Outcome prediction.

Resumo

Contexto e Objectivos. A gravidade do Traumatismo Cranianoencefálico (TBI) varia de suave a severo, sendo uma das principais causas de incapacidade no mundo e responsável por quase 50% das mortes por trauma. Há uma necessidade de métodos fiáveis para prever o prognóstico do TBI e ajudar os médicos a tomar decisões. Contudo, tal é difícil devido à heterogeneidade da lesão. Desenvolvimentos na área podem melhorar os cuidados de saúde, levando a mudanças no tratamento e reduzindo os custos do mesmo. O objectivo foi criar um modelo de aprendizagem automática para prever o progóstico a longo prazo de pacientes com TBI, avaliando a dinâmica ao longo do tempo do EEG contínuo durante a estadia do paciente na Unidade de Cuidados Intensivos(ICU).

Métodos. Realizámos medições contínuas de EEG em 111 pacientes com TBI moderado a severo após a admissão na ICU, durante 7 dias ou até à alta ou morte. Os parâmetros demográficos, clínicos, laboratoriais, radiológicos e parâmetros relacionados com o trauma foram recolhidos à admissão. O resultado neurológico aos 12 meses foi avaliado com a Escala de Resultados de Glasgow Estendida (GOSE) dicotomizada em mau (GOSE 1-3) ou bom (GOSE 4-8). Extraímos características relacionadas com potência, continuidade, sincronia, simetria e aleatoriedade para cada hora. Avaliámos as tendências das características do EEG utilizando regressões lineares e não lineares em 15 intervalos de tempo. Um classificador de Floresta Aleatória com validação cruzada de 5 conjuntos foi treinado utilizando parâmetros de admissão e a informação das regressões dependentes do tempo para a previsão binária do resultado. A selecção dos preditores incluiu um filtro baseado em categorização, e na eliminação de retrocesso. Avaliámos e comparámos modelos com diferentes preditores utilizando a sensibilidade, especificidade e a área da curva ROC.

Resultados. Os dois melhores modelos de previsão foram encontrados utilizando características EEG de 12 a 36 horas (AUC = 0,83 [0,75-0,91], sensibilidade = 0,83 [0,72-0,94], e especificidade = 0,80 [0,61-0,99]) ou de 48 a 72h (AUC = 0,85 [0,73-0,97], sensibilidade = 0,91 [0,83-1,00], e especificidade = 0,67 [0,32-1,00]) combinados com preditores do momento da admissão. A combinação de preditores de admissão com do EEG superou os modelos treinados apenas com admissão ou características do EEG separadamente. Os resultados também sugerem um melhor desempenho utilizando as tendências do EEG em comparação com as de pontos de tempo únicos.

Discussão. Este estudo permite uma previsão antecipada do prognóstico de TBI com uma abordagem de aprendizagem da máquina. As tendências do EEG ao longo do tempo apontam para um valor acrescido à previsão, salientando a importância de monitorizar continuamente o EEG e encorajar o uso em doentes com TBI grave, como é feito em outras populações de lesões cerebrais. Sugerimos que trabalhos futuros explorem fazer uma previsão multiclasse. Além disso, sugerimos ainda o uso de outros algoritmos para além do de Floresta Aleatória, e, se possível, que permitam a compreensão da contribuição das características para a previsão.

Keywords: Traumatismo Craninano, Eletroencefalografia, Aprendizagem Computacional, Floresta Aleatória, Prognóstico

Agradecimentos

Em primeiro lugar, dirijo o meu agradecimento ao professor Michel van Putten pela oportunidade de poder regressar a Enschede para embarcar neste projeto e por toda a ajuda ao longo dos meses. Também aos meus co-orientadores Marleen, Preejas e ao professor Miguel Velhote pelo apoio durante esta investigação. A todo o grupo do CNPH, gostava de dizer que foi um privilégio trabalhar convosco, em especial aos parceiros da sala de Masters e de PhD. Gostava também de deixar um especial agradecimento à Tanja, por ter sido sempre atenciosa.

Numa nota mais pessoal, gostaria de agradecer aos meus pais que desde criança me incentivaram a dar o melhor de mim, sempre, sem exceção. Serviram como exemplo de pessoas trabalhadoras e esforçadas, e que me mostraram que tudo é possível quando queremos e trabalhamos para isso. Reconheço-os também por me terem dado a oportunidade de estudar, de ir para fora da Madeira e de Portugal. Agradeço ainda todo o apoio, o amor incondicional e as palavras de força que me deram, não só nesta etapa de conclusão da tese mas também ao longo dos 5 anos do curso. Não poderia faltar uma palavra de apreço à minha avó, por se lembrar sempre de mim e por depositar tanta esperança no meu futuro. Um grande obrigada ao meu irmão, que me manteve entretida com os seus acontecimentos diários dignos de filme, que me incentiva a ler e a estar informada, e que me apoiou nesta fase. Sem a minha família não teria sido possível.

Esta tese simboliza o terminar de uma grande etapa, por isso não posso deixar de fazer um agradecimento muito especial a todos os que me acompanharam durante este percurso. Às Plenas, às Chloes e ao SAX, que fizeram estes anos valerem realmente a pena. A todos os Feupcaffes, às galerias e até à errónea decisão de ir ao boite (repetidamente). Às serenatas e às tardes no Kosovo, aos finos na AE, aos arraiais e às queimas (mesmo as que não tivemos). Sei que esta etapa vai deixar muitas saudades mas os momentos que levo comigo são memórias para a vida. Ao Verdomd, por proporcionar uma primeira experiência fora de Portugal melhor do que podia imaginar. Aos Rutbeer, aos Bored Neighbours e aos Tugas em Enschede que me ensinaram que vale sempre a pena voltar onde fomos felizes. Agradeço às meninas da rua Maria Feliciana por ouvirem os meus dias, por criarem um ambiente tão bom em casa e por estarem presentes nos momentos tão simples que se tornam tão especiais. Agradeço também às housemates da Lipperkerkstraat que conseguiram tornar todos os dias mais alegres. Gostava também de deixar um obrigada às amigas de [preencher a gosto], porque me mostraram que independentemente da distância, as verdadeiras amizades não cedem. Um curso não se faz sozinho e muito menos uma tese. Estas pessoas permitiram-me ter o melhor balanço possível, e tornaram esta jornada mais meiga, porque foi partilhada com elas.

Por fim, agradeço a todas as pessoas que ao longo da minha vida me influenciaram para que este percurso terminasse assim. Acredito que todas as pessoas que passam pela nossa vida, mesmo que não saibam, têm um propósito: ajudar-nos a crescer, relembrar-nos dos aspetos importantes, influenciar-nos e tornar-nos as pessoas que somos hoje. Por isso agradeço-te a ti, que estás a ler isto, e a vocês, que me ajudaram a ser quem sou hoje. E agradeço a mim, por nunca ter desistido ou duvidado, e sem a qual esta tese não teria acontecido.

"The beginning of all things is small. "

Cícero, Roman statesman

Contents

1 Introduction **1**
- 1.1 Motivation and Objectives . . . 1
- 1.2 Structure of the Dissertation . . . 2

2 Fundamental Concepts **3**
- 2.1 Traumatic Brain Injury . . . 3
- 2.2 Scoring Systems . . . 5
- 2.3 Electroencephalography . . . 6
- 2.4 Machine Learning . . . 9
  - 2.4.1 Data and Features . . . 9
  - 2.4.2 Selection and Evaluation of models . . . 10
- 2.5 Conclusion . . . 13

3 State of the Art **15**
- 3.1 Demographic, Clinical, Imaging, and Laboratory Data . . . 15
- 3.2 Machine Learning . . . 17
- 3.3 EEG . . . 18
- 3.4 Vital Signals: Dynamics over Time . . . 19
- 3.5 Conclusion . . . 20

4 Methodology **21**
- 4.1 Database . . . 21
  - 4.1.1 Clinical Parameters and Outcome Assessment . . . 21
  - 4.1.2 EEG Data . . . 22
- 4.2 Preprocessing . . . 22
  - 4.2.1 Clean data and Artefacts removal . . . 22
  - 4.2.2 EEG Feature Extraction . . . 23
- 4.3 Temporal Evolution . . . 26
- 4.4 Machine Learning . . . 26
  - 4.4.1 Random Forest Classifier . . . 27

5 Results **31**
- 5.1 Database . . . 31
- 5.2 Models' Performance . . . 32
  - 5.2.1 Predictors . . . 33
- 5.3 Feature Selection and Importance . . . 35

6 Discussion **37**
6.1 Models' performance 37
6.1.1 EEG Contribuition 39
6.1.2 Trends in EEG *vs* Fixed points 39
6.1.3 Time Window 40
6.2 Feature Importance 40

7 Conclusion and Future Work **43**

A Database Description **45**

B Parameter Tuning **47**

List of Figures

2.1 Aspects of responsiveness of the Glasgow Coma Scale. 5
2.2 The international 10-20 system for electrode placement. 7
2.3 Confusion matrix and evaluation metrics. 12
2.4 Summary of how to distinguish classifiers comparing AUC ROC. 13

4.1 Representation of the aperiodic component of the power spectrum. 24
4.2 Time intervals defined. A total of 15 intervals was used, including durations of 24, 36, 48 and over 48 hours intervals. 27
4.3 Schema of the feature selection methodology. 28

5.1 ROC curve for 48h-72h interval and linear fit. 33
5.2 ROC curve for 12h-36h interval and linear fit. 34
5.3 AUC as a function of the number of features used. 35
5.4 Feature Importance for the two selected models, including standard deviation of the cross-validation iterations. 36

B.1 OOB error as function of the number of trees. 47

List of Tables

5.1 Patient's characteristics grouped in a poor and a good outcome. 31
5.2 AUC for train and test set. 32
5.3 Models' performance using EEG predictors, admission predictors and both EEG and admission combined. 34
5.4 Comparison between time-points and temporal information 34

A.1 Patient's characteristics, including medical history, trauma related factors, administrated drugs and occurence of secondary insults. 45
A.2 Frequency table of the neurological tests at admission. Marshall, Rotterdam score and Basal Cisterns values are assessed through CT imaging. Pupil response is assessed at admission. 46

Abbreviations

AC	Alternating Current
ADR	Alpha Delta Ratio
AED	Anti-Epileptic Drugs
AI	Artificial Intelligence
ANN	Artificial Neural Network
APACHE	Acute Physiology and Chronic Health Evaluation
AUC	Area Under the Receiver Operating Characteristic curve
BSI	Brain Symmetry Index
cEEG	continuous EEG
CRASH	Corticosteroid Randomization After Significant Head Injury
CT	Computed Tomography
DFA	Detrended Fluctuation Analysis
ECG	Electrocardiogram
EEG	Eletroencephalogram
FN	False Negative
FP	False Positive
FPR	False Positive Rate
FOOOF	Fitting Oscillations and One Over F
GCS	Glasgow Coma Scale
GCS-M	Motor Glasgow Coma Scale
GOS	Glasgow Outcome Scale
GOSE	Extended Glasgow Outcome Scale
h	Hours
HB	Haemoglobine
HIE	Hypoxic-Ischemic Encephalopathy
ICA	Independent Component Analysis
ICP	Intracranial Pressure
ICU	Intensive Care Unit
IMPACT	International Mission for Prognosis and Analysis of Clinical Trials in Traumatic Brain Injury

MAP	Mean Arterial Pressure
MC	Motorcycle
MIII	metabolic inflammatory infectious origin of consciousness impairment
ML	Machine Learning
MRI	Magnetic Ressonance Imaging
MV	Motor Vehicle
OOB	Out-of-bag
PPV	Positive Predictive Value
PSD	Power Spectrum (or spectral) Density
qEEG	quantitative EEG
RBF	Radial Basis Function
ROC	Receiver Operating Characteristic curve
SAH	Subarachnoid Hematoma
SEF90	Spectral Edge Frequency of 90%
SVM	Support Vector Machines
TBI	Traumatic Brain Injury
TN	True Negative
TP	True Positive
TPR	True Positive Rate
WMO	Medical Research Involving Human Subjects Act

Chapter 1

Introduction

Traumatic Brain Injury (TBI) is an acute event that often triggers long-term developing injuries related to behaviour, emotions, cognition, and psychiatric disorders. The clinical severity of TBI ranges from mild to fatal, being a primary cause of disability worldwide and responsible for nearly 50% of deaths from trauma [1]. The impact of TBI results in great suffering for victims and their families, with enormous costs to society [2].

As injuries are never identical, clinical research into TBI is particularly challenging. In addition, the primary injury is altered by possible secondary insults and patient-related factors, premorbid and physiologic states, influencing the response to trauma. Finally, it is very difficult to standardise and categorise endpoints and distinguish between a "good" and a "poor" outcome in an objective manner. The Glasgow Coma Scale (GCS) is a popular way to categorise the consequences of sustaining a TBI. The outcome of a patient with TBI may vary, for instance, as a result of hospital approaches and fluctuations in available resources from place to place. This imbalance in treatment and patient prognoses is significant for the outcome of the injury [3, 4].

1.1 Motivation and Objectives

Early support for medical staff decisions can have an impact on the treatment decisions. Aid clinicians in the classification of the severity of TBI, decision making, decrease in labour intensity, and subjectivity is desired [5]. Prognostic models arise as a way to achieve that goal. By combining patients' data to predict clinical outcomes, physicians would be able to adequate the treatment strategy for each patient and diminishing the overall expenses associated with the treatment and rehabilitation. Besides that, knowledge based on prompt and dependable outcome predictions can empower family members, as it prevents medical practitioners from providing false hope to patients and relatives, and avoid unrealistic expectations. The mentioned points translate into a better allocation of resources, as unnecessary expenses are mitigated. In short, new developments in this area could benefit healthcare and promote the treatment of patients with TBI [6].

The project comprised in this master thesis is part of a long ongoing investigation on TBI by researchers from the Clinical Neurophysiology group at the University of Twente and the Medisch Spectrum Twente. This dissertation aims at creating a Machine Learning (ML) model capable of predicting the neurological outcome of patients following a TBI. The present study considers the inclusion of electroencephalogram (EEG) features and clinical predictors at the time of admission, in an attempt to predict the outcomes. Our primary goal is to investigate the dynamics over time of the EEG extracted features and how those changes combined with clinical predictors can ameliorate the prognostication of the disease.

1.2 Structure of the Dissertation

The document is structured in 7 chapters. This first chapter introduces the topic and provides a rationale for the conducted research. A broad overview of the theoretical concepts required for fully understanding the work herein is provided in chapter 2. The State of the Art of current alternatives in outcome prediction of TBI is presented in chapter 3. Chapter 4 will cover the methodological steps employed throughout the dissertation. In chapter 5, we shall present and assess the results by using the aforementioned methods. The discussion of the results is carried out in chapter 6. At last, chapter 7 will present the conclusions to be drawn from the research conducted and suggestions for future work.

Chapter 2

Fundamental Concepts

This chapter is reserved for the theoretical conceptualisation of TBI, EEG, ML, and related topics.

2.1 Traumatic Brain Injury

TBI is a condition caused by an external or internal mechanical force, such as a blow, an impact, or a concussion which leads to a change in brain function [7–9]. It comprises either the loss of at least one of the subsequent events: consciousness, memory of events, motor control, vision, speech, concentration, and orientation or other neurological deficits; and a change in mental status. Normally the effects occur in temporal proximity to the injury but manifestations can occasionally be slightly delayed [9].

Symptoms of a TBI range from mild and moderate to severe, defined by the extension of the damage to the brain. It varies from a brief change in mental status or consciousness, for mild, to prolonged unconsciousness, coma and death, for more severe cases. Injuries associated with TBI include contusions, intracerebral haemorrhage, subarachnoid haemorrhage, diffuse injury, diffuse axonal injury, ischemia, and skull fractures [7,8]. The survivors not only experience a considerable weight of physical, psychiatric, emotional, and cognitive disabilities during the in-hospital time, as they also experience it for the rest of their lives [10].

Brain injury is confirmed by neuroradiologic examination or laboratory testing. Tests at the time of admission include the assessment of pupillary dilatation and reflexes, measurements of Intracranial Pressure (ICP), imaging techniques e.g. computed tomography (CT) or magnetic resonance imaging (MRI), assessment of cardiac and pulmonary function, a full-body examination, and a neurological examination often using the GCS. Surgery and medication are the gold standard treatment after sustaining a TBI [7, 9].

Several possible events can lead to a TBI, the most common being: road-traffic accidents, military blast injuries, hitting blunt objects, falls, and sports-related concussions [8]. According to a Dutch study [11], the most common causes of TBI are either injury sustained at home and

during leisure activities (47.9%) or as a result of traffic accidents (33.5%). Regarding traffic-related injuries, 56.9% of accidents concern bicyclists [2]. In Portugal, the major cause of TBI is falls, followed by traffic accidents [12].

Approximately 50 million TBIs occur in the world annually, of which 2.5 million are reported in Europe every year [10]. In the Netherlands the incidence of TBI is about 213.6 per 100,000 persons per year [2], whilst in Portugal, the incidence is lower, being around 65 per 100,000 persons per year [13]. Hospitalisation rates were numbered at 1.5 million, with 57,000 TBI resulting in death [10] every year in Europe. There are also considerable differences between countries. For instance, admission rates in Germany and Austria are eight times higher than in Spain and Portugal [10].

In general, males are twice as likely to suffer TBIs than females. The age groups where the incidence is the highest comprise children between 0 and 4 years old and teenagers between the ages of 15 to 19. Concerning hospitalisations, people over 75 years of age are the most likely to be affected by TBI along with people incarcerated in prisons, military personnel, rescue workers and victims of terrorist attacks [14].

Costs generally depend on the length of stay, surgical procedures, and the severity of injury, with more severe cases being associated with higher costs, higher mortality, and more unfavourable outcome rates [11]. The average cost per case was found to increase with age. On the one hand, even though younger people are associated with a higher incidence and disease burden, treatment costs are lower. On the other hand, people aged 25 to 64 years old have a considerable low incidence but high economic costs [11]. Most of the time it is hard to quantify the costs associated with TBI. Not only due to the disease's heterogeneity but also because of the differences between treatment centres all over the world. Furthermore, in addition to the direct and in-hospital or medical costs, there are also costs associated with rehabilitation and lost productivity. Additionally, indirect effects on friends, families, caregivers, and the community also have to be taken into consideration [10, 14, 15].

People who have recovered from a TBI suffer from its repercussions, resulting in long-lasting effects that impact the patient's life and productivity [14, 16]. Since TBI is one of the most debilitating injuries, it reduces the patient's ability to regain independence in performing personal, social, and labour activities. In addition to disability, TBI is often seen as a risk factor for other health conditions. Among the list of long-lasting mental disabilities following TBI, the most common were complications related to memory, problem-solving, stress, and the management of emotional upsets and temper control. People who have sustained a TBI were more likely to binge drink and develop depression, epilepsy, and Alzheimer's disease [14]. The aforementioned factors contributed to disturbed relationships: divorce rates were higher, with social isolation being also frequent in these cases [15], and a reduced life expectancy was observed [10].

Hereby, it is possible to infer that the implications and costs of TBI are wide. The estimation of treatment cost-effectiveness is crucial to avoid unnecessary spending and elevate treatment management. This becomes particularly important when dealing with patients in severe critical states, in which a poor outcome can lead to very high costs, as it happens with patients who suffered

severe TBI [11].

2.2 Scoring Systems

In order to evaluate the condition of a patient at admission to the hospital or at any time point following a neurological injury, it is necessary to resort to an objective measure. The most commonly used for trauma include the Glasgow Coma Scale (GCS) and the Glasgow Outcome Scale (GOS) [17]. Both scales allow for the grouping of patients according to the severity of sustained injuries and have been proven to be correlated with neurological outcomes [4, 18].

The GCS is used to quantitatively describe impairment in consciousness in acute trauma patients. It was first described in 1974 and is used in many medical units to this day. The scale rates patients according to three aspects of responsiveness: eye-opening, motor (GCS-M), and verbal responses, as exemplified in figure 2.1. Reporting each of these parameters separately allows us to infer the state of a patient clearly and easily. The use of the responsiveness evaluation allows a more effective management of patients during acute care, a period during which many decisions have to be made. These vital decisions range from securing the airway and triage to determining patient transfer, in more severe patients, or the need for neuroimaging, admission for observation or discharge, for less severely impaired patients [4]. GCS can be employed as a series of measurements to monitor the clinical course of a patient and guide changes in management. Deterioration of a patient's condition should always be carefully assessed [17].

Glasgow Coma Scale

EYE OPENING RESPONSE		VERBAL RESPONSE		MOTOR RESPONSE	
SCALE	SCORE	SCALE	SCORE	SCALE	SCORE
EYES OPENING SPONTANEOUSLY	4	ORIENTATED	5	OBEYS COMMANDS FOR MOVEMENT	6
EYES OPEN TO VERBAL COMMAND OR SPEECH	3	CONFUSED CONVERSATION BUT ABLE TO ANSWER QUESTIONS	4	PURPOSEFUL MOVEMENT	5
EYES OPEN TO PAIN	2	INNAPPROPRIATE RESPONSES	3	WITHDRAWS FROM PAIN	4
NO EYE OPENING	1	INCOMPREHENSIBLE SOUND OR SPEECH	2	ABNORMAL FLEXION OR DECORTICATE POSTURE	3
		NO VERBAL RESPONSE	1	EXTENSOR RESPONSE OR DECEREBRATE POSTURE	2
				NO MOTOR RESPONSE	1

Figure 2.1: Aspects of responsiveness of the Glasgow Coma Scale. Adapted from [19].

The GOS was also created in 1974. However, this assessment was oriented towards the evaluation of the outcome after a brain injury. It has five categories namely: 1) Dead; 2) Vegetative State, when a patient is unable to interact with its surroundings; 3) Severe Disability, when a patient can follow commands but cannot live independently; 4) Moderate Disability, when a patient is capable of living independently but is unable to return to work or school; 5) Good recovery, when the patient can fully return to daily activities [20]. The scale was redefined in 1981 to extend its 5 categories to 8, originating the Extended Glasgow Outcome Scale (GOSE). It has arisen from the limitations of the original one, e.g. the insensitivity to change, use of broad categories, and difficulties with reliability due to the lack of a structured interview format. Therefore this addition has improved the reliability of the rating and is also more sensitive to variations in mild to moderate TBI. The categories are divided into 1) Dead, 2) Vegetative State, 3) Lower Severe Disability, 4) Upper Severe Disability, 5) Lower Moderate Disability, 6) Upper Moderate Disability, 7) Lower Good Recovery, 8) Upper Good Recovery [21].

The scales can be used for patients' stratification of TBI into mild, moderate or severe and access favourable or unfavourable outcomes. Mild TBI is characterised by a GCS between 13 and 15. Usually, these patients experience headaches, dizziness and irritability but the outcome is favourable, translating to a GOS higher than 3. GCS of 9 to 12 is attributed to moderate TBI and patients might either recover, experience moderate disability, become in a vegetative state, or die. Severe patients are characterised by a GCS of 8 or fewer. These patients are usually less likely to have positive outcomes. Oftentimes the outcome is either death (GOS = 1) or vegetative state (GOS 1-3) [20, 22].

Nowadays, scoring models based on CT imaging are an option to consider for prognostic, especially in patients that are under sedation, influence of alcohol or psychoactive drugs or are intubated, as it occurs in severe TBI patients [23]. Specific grading scales based on CT can provide useful information on a patient's health condition. As a further matter, the Marshall scale, Rotterdam scale, and Helsinki scale are radiological imaging scales that aid prognoses and predict the risk for increased ICP and outcome in adults. They differ in the factors used for evaluation. Particularly, Helsinki includes the presence of an epidural hematoma, while Rotterdam takes into account the presence of subarachnoid haemorrhage but does not distinguish mass lesion in terms of type and size, and Marshall takes into account the presence of a mass lesion, the midline shift and the status perimesencephalic cisterns but relies on subjective assessments [24].

2.3 Electroencephalography

Electroencephalography provides an electro-biological measurement of the brain, by evaluating the electrical activity of a group of neurons [25,26]. It is a way to assess the brain's electrical function and allows for a time display of the difference in voltages in two sites of the brain [27]. Some challenges experienced when using this technique are that the human cortex is under the scalp

surface, which is a barrier to this measurement, and that the transformation of a three-dimensional source into a two-dimensional projection is required [27].

Traditionally, EEG recording is performed as a scalp EEG, with the placement of the electrodes on the surface of the skull [25]. By doing so, it is possible to measure the electrical activity of large, synchronously firing, populations of neurons [28]. The advantage of the extra-cranial EEG comprises its non-invasive, non-expensive (cost of hardware is low [25]), and painless assessment of the neuro-physiological function [26,28]. The standard electrode placement is the international 10-20 system, shown in figure 2.2, in which the name derives from the 10% and 20% spaced interval division of the sensors locations between the ears and nose [26,27]. At least 21 electrodes are recommended for scalp EEG [29].

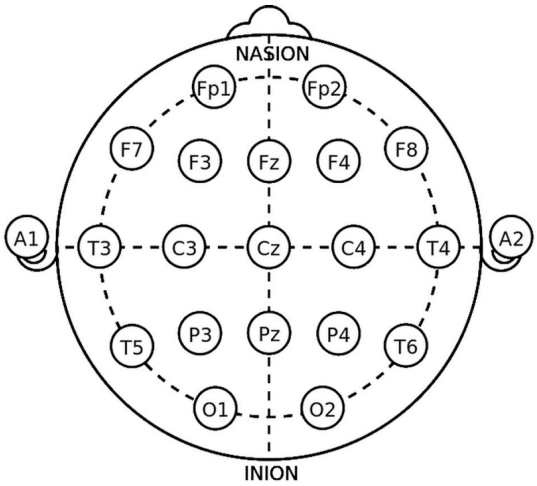


Figure 2.2: The international 10-20 system for electrode placement.
Reproduced from [30].

The EEG potentials are displayed in channels, which are the difference in potential between two electrodes [29]. The electrodes are designated as frontal (F), frontopolar (Fp), occipital (O), central (C), temporal (T), and parietal (P), according to their location in the brain areas, and a number is also given (odd for left and even for right) according to the distance of the electrode to the midline placement (to which the letter z is attributed instead of a number). This spatial array is shown in figure 2.2. From this, it is possible to obtain an electrical map which is named montage [26,27]. The montages are the different possible arrangements of derivations or channels that can be used to display brain activity. In a bipolar montage, the local potentials are amplified due to a difference in potential between contiguous electrodes [29]. The referential montage is more susceptible to external noise but detects local and distant potentials [29]. The combination of different montages allows for a better interpretation of the EEG by circumventing the drawback of a two-dimensional projection of a three-dimensional activity, allowing the display of the activity spatially. This ensures that there is no missing activity, two sides of the brain can be compared and it is possible to accurately locate the activity [31].

The EEG is sensitive to different states such as stress, alertness, drowsiness, resting, hypnosis, and sleep [26, 29]. The EEG detects several artefacts that can be originated from patient-related reasons (electrical sources), eye blink, eye movements, muscle activity, electrocardiogram (ECG), or any other body movements considered unwanted physiological signals that cause perturbation in the EEG [26, 27, 29]. In addition, artefacts can result from technical sources (electrical environment), such as alternating current (AC) power line noise associated with the impedance or electrode wires [26, 27]. Occasionally, single electrode artefacts or faulty channels may happen. Artefacts might also be caused by unequal electrode impedance and other specific events. In the Intensive Care Unit (ICU) it is also possible to often observe "noise" from mechanical and instrumental, or environmental source. The mentioned artefacts can be detected during preprocessing stages [27], and can be removed using filtering and noise reduction methods [25]. The most frequently employed methods for filtering are bandpass, wavelet, finite impulse response and adaptive filters [25].

EEG allows us to identify normal and abnormal electrical activity in the brain. Therefore it is an extremely important tool in neurology as well as in clinical neurophysiology [26]. EEG findings can support diagnosis and have many clinical applications. Visual interpretation of EEG allows for the identification of seizures, epileptiform activities, and other patterns with important information, including status epilepticus [27]. It is also possible to identify alertness, coma, and brain death and locate areas of damage [26]. Thus, using EEG allows practitioners to detect if an intervention can prevent irreversible damage, enhancing its use in the ICU.

Quantitative EEG (qEEG) is the digitised EEG that can be analysed, extending the visual EEG interpretation and allowing for the extraction of more information. qEEG can be processed using the “Fourier” or, although less common, the “Wavelet” algorithms. This enables the visualisation of the brain’s cognitive processing tasks [32]. qEEG allows us to determine spatial structures, location of brain activity and abnormalities [26]. This way, EEG interpretation becomes easier since the waveforms are presented in an easy-to-read format [33, 34]. Normally, visual EEG analysis can take up to 1-2 hours per day, which can be reduced by computer algorithms that additionally make it possible to analyse more data [5].

The mentioned qEEG techniques enhance the use of continuous EEG (cEEG), which is a non-invasive procedure to monitor the consciousness of patients with spatial and temporal resolution [33, 35]. cEEG is a paraclinical examination tool in patients with hypoxic-ischemic encephalopathy (HIE) but it is infrequently used in patients with TBI [36]. Although it is not yet a common practice, in the last years its use has attracted much attention. Some locations are now adopting the standard practice of cEEG monitoring for TBI patients in the ICU [37] since it allows for the detection of non-convulsive status epilepticus, assesses the sedation, detects cerebral ischemia, and estimates the outcome of patients with neurological disorders [33]. Furthermore, cEEG allows for the detection of changes over time, allowing for earlier medical interventions [33]. For those reasons, it should be initiated as early as possible after admission to the hospital [37]. Nonetheless, the use of cEEG can be limited by technical, patient-related or system resources [38].

To decode the activity in the brain by using the EEG, it is required to either extract features or use spectral information using the Fourier transform [25]. From the EEG it is possible to extract univariate features when they are taken from each EEG channel separately or multivariate when taken from several channels [25]. EEG features can be separated into time, frequency, and time-frequency domain features [39].

In conclusion, the EEG technique has the advantage of not requiring external radiation or injected substances in order to produce a direct measurement of brain activity. Not only the use of visual EEG but the extraction of qEEG features has also been mentioned to be relevant for diagnostic and prognostic purposes [40]. However, when working with EEG signals it is always required to take into consideration the preprocessing, and artefacts removal which will be highly time-consuming.

2.4 Machine Learning

Artificial Intelligence (AI) dates back to the 1950s when the term started to be used as a description of the capability of machines to mimic human intelligence [41, 42]. The term AI is used when a machine can perform tasks related to "problem-solving" or "learning" [42]. It allows for the computer to copy human decisions through a process of learning, reasoning, and self-correction but in a faster and easier way [43].

ML, a subset of AI, makes use of large data sets as inputs and identifies its patterns. It allows users to train the machine to make recommendations or decisions autonomously [41]. After a certain number of iterations, the machine can take an input and predict the output. The output can be compared to the known outputs for each input to judge the accuracy of the algorithm and adjust, when necessary, to improve the prediction of future inputs [41]. Nowadays, the use of extremely large data sets has led to ML becoming popular and integrated into everyday life [41]. ML is applied to the health field in many different ways either in diagnosis, treatment, or personalised care, among others. ML allows us to learn about the structure and the functional anatomy of the brain [25], holding great potential for neuroscience and neurocritical care. The use of accurate ML options allows us to directly improve the patient's outcome [41] and to obtain an individual and tailored therapy [44]. AI allows for the transformation of medical practice by aiding the physician's interpretation of the data and improving the performance of diagnosis, prognosis, and the management of decisions [43].

2.4.1 Data and Features

ML makes use of data sets, comprising multiple data points and the measured or calculated features, which describe each data point [45]. It is possible to classify the features as categorical (predefined values without a specific order), ordinal (predefined values with a specific order) or numerical (real values). Each of those features is a one-dimensional representation of the feature space. The actual value of a feature for one data point allows the definition of the point in a place in the space dimension. All the values of all the features of one data point constitute the feature

vector. If we have more features for each data set, the dimensionality is higher. The higher the dimensionality, the harder it becomes for a human to visualise the information and detect patterns in the data. However, computers can handle this efficiently to a certain degree [45]. Furthermore, it is still considered a challenge to include time as a continuous variable in ML algorithms. Time-dependent data in ML data sets can be included: i) without changes, having each time point representing a feature ii) applying Fourier transform, using the coefficients as features [45].

Regarding terminology, features are also mentioned as predictors, variables input or attributes. The target is also known as an outcome, output, response variable, dependent variable or label.

Raw data does not always fit the models being used, thus preprocessing it is an important step in ML. Very often, the features have to be extracted from the data [46]. Most ML algorithms can handle high-dimensional data sets but feature engineering, which is the data transformation into advanced combinations, such as log-transformed data, or simpler ones, as products and ratios, is important [45]. Additionally, feature transformation is also essential. It usually includes either the normal transformation or the standardisation of the data, for all features to be in the same range [46], and dimensionality reduction, especially in the presence of highly correlated or irrelevant features [46]. Furthermore, it is also important to pay attention to data quality which evaluates outliers, missing values (identifying, imputation or eliminating [46]), bias in data, unbalance in data set, and definition of measures of similarity [45]. Note that when working with categorical data, encoding might be necessary [46]. Feature selection, normally used to reduce noise, remove redundant or irrelevant attributes, and reduce dimensionality, might also be necessary. It allows us to decrease the chances of overfitting [47], which occurs when the models fit perfectly the training data but cannot generalise for new data [48]. Feature selection approaches are usually divided into filters and wrappers, whereas the first ones usually make use of the data's general characteristics, and the latter usually depend on the predictor. Wrappers are considered to have better results while filters are less computationally expensive. The use of these approaches is not mutually exclusive [47].

Data can be separated into training, validation, and testing set. The training set is the one employed for building various models while the validation set is for the selection of the algorithm and its respective hyper-parameters [45].

2.4.2 Selection and Evaluation of models

There are three possible types of ML: i) supervised, when data is labelled, the algorithm extracts a relation between the given labels and the data, hence there is feedback which allows the model to predict the output for unseen data; ii) unsupervised, when there are no given labels to the data and no feedback, but the model finds a hidden structure in data; iii) semi-supervised (less common), which is the situation when there is a decision process, a reward system and the model can learn a series of actions [25, 45, 46].

Classification tasks, a subcategory of supervised learning, try to predict the labels of new instances, using the previous observations. The class classification can be binary, having a positive

and negative class, and the goal is to find the separation boundary between the classes. Multi-class classification is also possible when the inputs are also multiclass [46]. For the prediction of continuous outputs, a different subcategory is used, the regression analysis. In this case, many predictor variables and a continuous response variable are given, and a relation between the two is searched for [46]. Regarding unsupervised learning techniques, no prior information is given about the labels and the purpose of the methods is to explore the structure of the data and try to extract significant knowledge, for example using clustering and dimensionality reduction approaches [45, 46].

Learning algorithms aim to learn from the given data set, create a model that describes it, and predict the output for a new data point. Different ML algorithms have been developed to solve different prediction tasks. However, no solution can be universally applied to tackle other problems [45,49]. Performance metrics lead users to improve the training of the models since they allow both the evaluation of the performance of the model itself, and for the selection of optimal hyper-parameters [45]. Moreover, performance metrics enable the comparison of different models and algorithms [46], since the selection of the most adequate for a specific problem is possible.

When the model cannot provide predicted values close to the observed values of the training set it is considered underfitting, usually resulting from the use of models too simplistic or selection of irrelevant features. When the model cannot generalise beyond the given training set it is considered overfitting, which often occurs because of high complexity or a too large number of features for a small set of training examples. The model should avoid both underfitting and overfitting of the training data. This can also be explained as a bias/variance trade-off. Bias is the difference between the average of the predicted values and the true mean that we are trying to predict. The variance is a measure of the sensitivity of the model to the training set. Ideally, bias and variance should be minimum. However, if the complexity of a model is increased, the variance also increases whilst the bias decreases. A balance between these values should be found, allowing for a good performance of the model in a new data set [45]. The models are selected according to their performance on the validation set. The generalisation error is evaluated with the test set, being for that reason called test error. Cross-validation can also be put to use. In this case, there is only a separation into the train and the test set, in which the train set is divided into k subsets. k-1 subsets are used for training and one is used to evaluate the performance of the model. The process is repeated k times for each one of the k subsets in validation, being the performance score the average of each set of hyper-parameters to test [45]. This allows for the fine-tuning of the parameters of each of the models [46].

The most commonly used performance measures for model assessment are derived from the confusion matrix [45]. Figure 2.3 represents a confusion matrix, where it is possible to define four different metrics: when the data is correctly predicted as positive - True Positive (TP), or negative - True Negative (TN); or when wrongly predicted as negative - False Negative (FN), or as positive - False Positive (FP) [25]. Some metricts to take into consideration are [45, 51]:

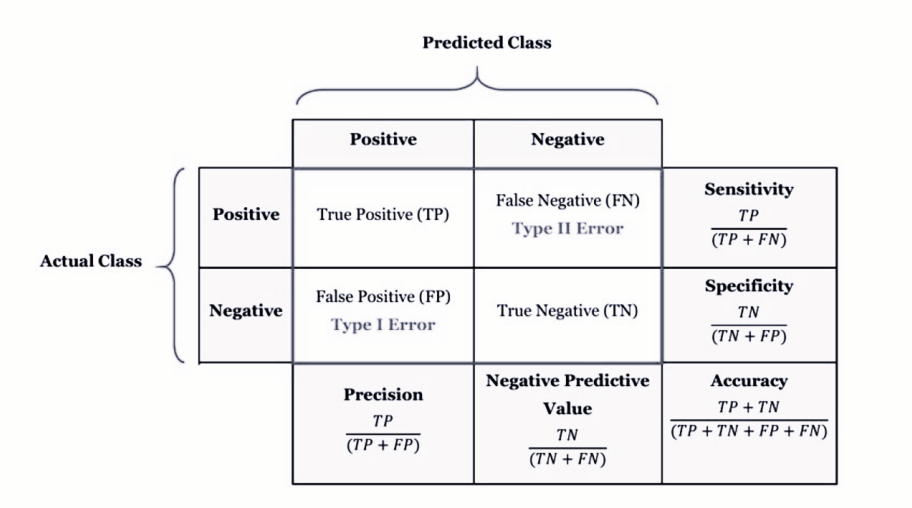


Figure 2.3: Confusion matrix and evaluation metrics.
Adapted from [50].

- Precision: ratio of correctly predicted positive values with the total number of predicted positive values. In clinical context it is also known as positive predictive value (PPV).

$$Precision = PPV = \frac{TP}{TP+FP} \tag{2.1}$$

- Sensitivity: ratio of correctly predicted positive values with the total number of positive values in the data set. Also known as recall.

$$Sensitivity = Recall = \frac{TP}{TP+FN} \tag{2.2}$$

- Specificity: ratio of correctly predicted negative values with the total number of true negatives and false positives.

$$Specificity = \frac{TN}{TN+FP} \tag{2.3}$$

- F1-score is a harmonic mean of precision and recall, allowing for a combined idea of the two metrics. The value is maximum when precision and recall are balanced as equal.

$$F1-score = \frac{2}{\frac{1}{Recall}+\frac{1}{Precision}} \tag{2.4}$$

- Accuracy: number of correct predictions, divided by the total number of predictions made.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{2.5}$$

- Area Under the ROC curve (AUC): the Receiver operating characteristic curve (ROC) is a plot of the sensitivity, also known as True Positive Rate (TPR), against False Positive Rate

(FPR), which is the ratio of negative values predicted incorrectly. The points on the ROC curve are obtained by selecting different thresholds for the classification of a point in the positive or the negative classes. Ideally, the best result of AUC is obtained when there is perfect separation between the positive and the negative classes i.e. TPR=1 and FPR=0, as represented in figure 2.4. Therefore, the goal is to maximise TPR and minimise FPR, which results in a higher AUC and better the performance.

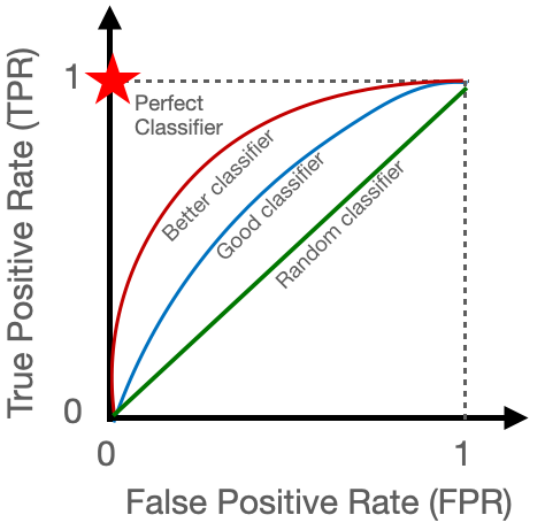


Figure 2.4: Summary of how to distinguish classifiers comparing AUC ROC. Reproduced from [52].

2.5 Conclusion

TBI is a high incidence disease worldwide. Advances in informatics and prognostic tools have the potential to promote neurocritical care by developing an outcome prediction model to greatly assist medical decision-making, upgrade patient care, reduce the incidence of secondary insults, assist in cost reallocation, provide more robust support for family members, and improve patient care. The use of ML allows the analysis of large quantities of data, making it possible to analyse information from cEEG signals in a more effective way.

Chapter 3

State of the Art

Neurocritical care relies on the continuous and real-time measurement of physiological parameters. The main focus of neurocritical care and TBI management is on the prevention, prediction, and detection of secondary brain insults, that tend to often appear after an injury. The use of multimodal monitoring and informatic tools in ICU concedes an improved follow-up of the patients since the transformation of raw data into useful information is possible [53].

In this chapter, we will focus on the description of the currently available and under-study options for the management of TBI and outcome prediction. It comprises a description of the models available in clinics and studies, including their advantages and limitations.

3.1 Demographic, Clinical, Imaging, and Laboratory Data

Prognostic models have been developed and improved over the years. Some include clinical data, laboratory or CT information. The commonly established prediction models, that are currently used in some centres are the Acute Physiology and Chronic Health Evaluation (APACHE) [54,55], the International Mission for Prognosis and Analysis of Clinical Trials in TBI (IMPACT) [56], and the Corticosteroid Randomization After Significant Head Injury (CRASH) [57]. Other than the available models, numerous research efforts are being focused on this matter, using different predictors, ML models, and timing of outcome prediction. They aim at tackling some of the existing flaws encountered in the existing models such as the use of old data, lack of validation or calibration or the restricted scenarios to which the models can be applied [58]. This section describes the traditional and some non-trivial approaches for the prediction of outcomes after brain injury using either demographic, clinical, imaging, and laboratory data individually or combined.

APACHE [54, 55] is a mortality and functional outcome prediction model that was firstly developed in the ‘90s and has undergone several updates. It includes 12 physiological variables (including GCS, age, diagnosis, surgical information, and history prior to admission). The model stratifies patients according to mortality risk. Although for the first day the prediction is not optimal, it still allows for the testing of therapeutic efficacy and is a good indicator of later mortality

and functional recovery [54, 55]. Nonetheless, APACHE is not specific to TBI but both the IMPACT [56] and CRASH [57] studies are currently the most used prognostic models for TBI.

IMPACT comprises 3 models of increasing complexity: 1) the core model, considering age, motor score, and pupillary light; 2) the extended model, which includes CT, and secondary insult variables; 3) the lab model, with levels of glucose and haemoglobin (Hb). One of IMPACT's advantages is the use of admission data making it independent of hospital care. However, it does not take into consideration the patient's changes in condition during the hospital stay. Furthermore, it is also robust due to cross-validation with the original data set and external validation with CRASH [3]. However, according to some authors, re-calibration of IMPACT is also needed [58]. Furthermore, Raj *el al.* [58] pretended to improve the IMPACT's performance by combining it with the APACHE. A combination of the two models resulted in better mortality prediction but not of the neurological outcome [58].

Likewise, the CRASH [57] study, which includes a large dataset (over 10,000 patients) also includes predictors such as age, GCS, pupillary reactivity, and presence of severe extracranial injury and, for the CT model, the presence of petechial haemorrhage, obliteration of the third ventricle or basal cisterns, subarachnoid haemorrhage, midline shift, and non-evacuated hematoma were considered. They predict mortality within 14 days, and both mortality and unfavourable outcome (using 5 classes) within 6 months. There are two variants, one for low and middle-income countries, and one for high-income countries. In 2008, Steyberg *et al.* conducted a similar study [59] using demographic predictors at admission (age, gender, race, education) in combination with clinical severity (cause of injury, GCS components, pupillary reactivity), secondary insults (hypoxia, hypotension, hypothermia), blood pressure (systolic, diastolic), CT characteristics, and biochemical variables using a Logistic Regression model to help clinicians assess the severity and prognosis of a patient with TBI, using mortality and unfavourable outcome prediction at 6-month. They concluded that the most relevant information for prognosis was contained in age, motor score, and pupillary reactivity on admission [59].

More recent studies, using patient data from Nijmegen, the Netherlands [60] aimed at developing a validated model to predict outcome in moderate and severe TBI using demographic, clinical, and radiological parameters, including two prognostic models: 1) demographic and clinical models: using age, pupillary responses, GCS, the occurrence of hypotensive episodes after injury; 2) CT models. They used a multivariate binary Logistic Regression between predictors of three possible outcomes and was externally validated in a Dutch TBI cohort. The outcome predictors for both death and unfavourable outcome after multivariate analysis were age, GCS at admission, the occurrence of a hypotensive episode, and pupillary reactivity. Despite IMPACT and CRASH's use of bigger data sets, this model relies on more recent data [61].

A comparison between IMPACT, CRASH, and Nijmegen models concluded that all three show good discriminatory ability for 6-months outcome in TBI patients [61], with small AUC differences when applied to a moderate-to-severe TBI population. In general, mortality was better predicted than the unfavourable outcome. Age, GCS, and pupillary reactivity were common

predictors for all three models but the models including CT information showed better performance [61].

Most of the patient data used for these studies were derived from highly developed countries and, for that reason, not applicable to global intensive care. Of the three mentioned, only CRASH includes a model that takes into consideration the differences between middle-income countries and high-income countries. With this motivation, Amorim *et al.* [62] focused in researching prediction of outcome in low-to-middle-income countries. Predictors included were gender, age, pupillary reactivity at admission, GCS, presence of hypoxia and hypotension, CT findings, trauma severity score, and laboratory results. Different algorithms were tested, the best ones were Random Forest for in-hospital mortality and conditional inference tree model for length of stay in the ICU [62].

Most of the previously mentioned methods included the use of CT in prognostic models. Actually, neuroimaging techniques, such as CT and MRI, are methods used in hospitals for the assessment of neurological damage and therefore are very commonly included in prognostic models [40]. CT findings such as the status of the basal cistern, midline shift, associated traumatic subarachnoid haemorrhage and intraventricular haemorrhage are useful predictors of outcome and valid options for prognostication of patients with TBI according to the literature [23]. However, the use of imaging portrays disadvantages arising from the CT findings not being always objective [63]. Relying on CT information is limited in countries where the use of radiologic information is uncommon, or for prognostic of young children or pregnant women, where radiation exposure is not recommended [64].

Despite numerous attempts to formulate new and better models, most do not reach clinic implementation either for being developed on small samples, having a poor methodology, not having external validation, and not being able to generalise [57, 61, 65].

3.2 Machine Learning

As previously mentioned in chapter 2, the decision on the best ML algorithm is not always unyielding. Authors compare the use of different algorithms and obtain different results. Comparisons might also be dependent on the evaluation metric employed. A study for prediction of mortality in severe TBI patients [66] found that the most significant variables were different for each model but that among the tested algorithms, Artificial Neural Networks (ANN) had the highest AUC in the train set (0.968) and sensitivity (80.59%). In the test set, ANN had the highest sensitivity (84.38%). Cubic Support Vector Machines (SVM), quadratic SVM, and linear SVM were reported to be the best algorithms in a study conducted to predict survival [67]. In contrast to another study [68], that concluded that for the prediction of poor outcome (GOS 1-3) Random Forest was the model that showed the highest sensitivity (97.2%), Gaussian Naïve Bayes had the highest specificity (82.8%) but the highest accuracy was attained with Gradient Boosting (87.5%), and highest AUC was in SVM with Radial Basis Function (RBF) Kernel (0.894). On the bootstrapped test data Random Forest outperformed the other models in terms of accuracy and AUC. For the

prediction of death, the best sensitivity was found for ridge regression (85.1%), the best specificity for Random Forest (99.3%), the highest accuracy for linear SVM (89.8%), and the best AUC for Random Forest (0.960). Ridge regression surpassed the other models in sensitivity and AUC in the validation set. Furthermore, for in-hospital prediction of poor outcome, Random Forest was the best whilst for in-hospital mortality, Ridge Regression was the leading one. These differences among values of the different metrics for each algorithm indicate that the choice of algorithm to use may portray a big challenge.

ML, by revealing patterns and relations in the data, portrays as a great advantage to optimise the outcome in neurological patients since it allows for a better understanding of the relevant factors, leading to a more tailored therapy for each patient [44]. Nonetheless, the use of ML also accounts for disadvantages derived from the lack of interpretability, explainability, and difficulty assessing the contribution of each variable to prediction [64] due to the use of black-box techniques. It can be discouraging for doctors since it comes with the uncertainty of responsibility on wrongly classified cases [44] and mistakes committed based on such. To tackle this problem, some authors opt for models that allow us to understand the decisions behind the algorithm. One example was a study in which the decision was based on a tree model to predict the long-term outcome of severe TBI patients [69]. In this case, variables used included the Abbreviated Injury Scale, Marshall score, and pupil reactivity as potential predictors. They obtained results of AUC of 0.67. The advantages of such models are that they are simple, easy to remember, explainable, and easy to incorporate into the clinic.

One further disadvantage of the use of ML if that the shift from research to clinical settings is not always adequate [44]. One way to avoid this is to ensure the models are largely validated in different cohorts. Finally, the authors also clarify that ML models are only meant to reinforce and aid medical decision-making, not to replace human opinion [44].

3.3 EEG

Regarding the features which ought to be used for the outcome prediction of TBI, qEEG parameters related to absolute band power and variability have been considered [70]. For absolute band power, median alpha power showed a very strong relation with outcome [70]. Regarding variability, relative fast theta power variability and relative alpha power variability were pointed [70, 71]. Furthermore, both variability of mean frequency and spectral entropy, and the total power in beta, fast theta, and alpha frequency bands, have been noted to be correlated to the outcome of TBI [71]. Literature reviews concluded that the most used qEEG features are spectral analysis [8], absolute and relative amplitude [8, 72], power within a frequency rate or on each channel [8, 72], total EEG power [72], coherence [8, 72], and symmetry between homologous electrodes pairs [8].

Regarding modelling approaches using qEEG, the models most commonly used in the literature are Logistic Regression and Random Forest [72]. Often others such as Generalised Linear model, Linear Discriminant analysis, Multivariate Autoregressive model, SVM, and Least Absolute Shrinkage and Selection Operator are used [22].

A pilot study sought to use qEEG measures and clinically relevant parameters to predict the outcome of patients with TBI [5]. Their work combines EEG monitoring with the IMPACT predictors. The best result was obtained by using 19 features (8 qEEG, Mean Arterial Pressure (MAP), age, and 9 IMPACT parameters). In all models, the mean amplitude of EEG, age and MAP were shown to be important for prediction capability. Features showed different significance when the time interval was changed. Total power (mean amplitude), spectral edge frequency of 90%, and relative alpha power at 72h were found to be significant. Similar to this study, other authors used Random Forest classifiers with 8 visual EEG features and clinical characteristics to predict survival at 6 months for different groups (stroke, TBI, the metabolic inflammatory infectious origin of consciousness impairment (MIII) and HIE [36]. Results showed that the most important clinical characteristics were GCS (closely followed by age) for predicting survival, and age for predicting unfavourable outcome. EEG background reactivity was the most important EEG feature for both outcomes. In this study, the use of clinical variables with EEG did not change classification performance, but increased specificity and decreased sensitivity [36].

3.4 Vital Signals: Dynamics over Time

The consideration of temporal information might be one factor worth taking into account, especially when using physiological signals and continuous measures.

Early in 2012, Feng *et al* [73] developed a model which considered information regarding historic ICU readings for ischemic brain injury. The model performed better compared to those that did not consider temporal information. Therefore a temporal parameter may hold promise for the prediction in TBI patients. Temporal models are more complex and require more training time but can outperform non-temporal models in terms of accuracy, AUC, and F-measure [73].

Dynamic algorithms might also be an option to improve prediction. Raj *et al*. [74] used ICP, MAP, Cerebral Perfusion Pressure, and GCS values for motor and eye responses, to predict mortality 30 days after a TBI. The first prediction was at 24h after trauma, which was then updated every 8 to 120 hours. The performance was shown to be best at 48h after injury. Medical staff benefits from real-time predictions based on dynamics since it allows the physician to become aware of the worsening of the patient's condition. It also allows the identification of the need for intervention [74].

In addition, substituting defined time points for the use of continuous measures may be helpful in predicting TBI's outcome. Lee *et al* [75] attempted to correlate the use of cEEG features with functional outcome assessed with the GOSE. The authors found that the unfavourable outcome was not associated with age or pupillary reactivity in this cohort, but rather with both GCS after resuscitation and injury severity score. Ictal-interictal continuum patterns correlated significantly with the IMPACT score. Furthermore, the cEEG background features: the absence of a posterior dominant rhythm, the absence of N2 sleep transients, a predominant delta activity, a discontinuous background at each time point of recording, and the presence of a moderate to severe background at each time point of recording were found to be related with functional outcome. In summary, the

data suggest that cEEG background can provide additional prognostic information related to the established clinical variables as early as 24h after recording onset [75].

3.5 Conclusion

Predicting long-term outcomes for individual patients is difficult, especially in the first hours or days after injury and it is currently limited by available scoring systems or clinical models. Outcome prediction models with clinical data are used including the APACHE [54,55], IMPACT [56], and CRASH [57]. Besides clinical data, it is also possible to include laboratory and CT options.

Variables commonly used are age, admission GCS (or one specific component of the GCS), and pupillary reactivity [59–62]. Combining admission features with others collected later during ICU stay can improve the accuracy of the results. Some authors have also mentioned correlations with other variables and patient's outcome such as hypotension [60,62], gender [62], hypoxia [62]. The use of CT information is also frequent [23, 40, 62, 63]. Several authors demonstrated that the models may improve their accuracy and performance when physiological signals are included in this evaluation [53, 76]. It is also expected that the addition of EEG features to the clinical data will improve the performance of the model, and provide medical staff with a reliable prognostic tool [5, 36, 70, 75].

Furthermore, the use of ML for such models has also been discussed since it presents both advantages, such as allowing for a more tailored therapy to the individual patient and for the understanding of patterns that are not easily interpreted by humans, and disadvantages associated with black-box models lacking explainability [44]. The use of regression models instead of ML is also often considered in such models but does only achieve superior performance in very specific conditions [77, 78]. Different ML algorithms should be considered [22, 68, 69]. Most studies regarding outcome prediction of TBI either use Logistic Regression [58, 60], Random Forest [62, 68], ANN [63, 64, 66], SVM [67] or tree-based models [69].

Several points for outcome assessment can also be considered such as mortality (survival *vs* death) in-hospital, after 14 days or in the long-term, and the outcome assessment (favourable *vs* unfavourable) at 3, 6, 9, and 12 months after injury.

Chapter 4

Methodology

4.1 Database

The study population for this work were patients with moderate to severe TBI admitted between 2013 and 2021 to the ICU in the Medisch Spectrum Twente. Inclusion parameters were: 1) GCS of 3-12 at the trauma site or emergency department; 2) admission to the ICU and expectancy of stay of at least 24 hours; 3) age older than 18. Patients were excluded if trauma was combined or following cardiac arrest, had a medical history of TBI, stroke without full recovery or brain illness (such as brain tumour or neuro-degenerative disease) or limited life expectancy before the TBI, high risk for development of iodine contrast induce nephropathy or contrast allergy, potentially childbearing or pregnancy. The study was conducted according to the principles of the Declaration of Helsinki 2013 and in accordance with the Medical Research Involving Human Subjects Act (WMO). Verbal and written consent is provided as soon as possible to the legal representative. When the patient regains consciousness, verbal and written study information is provided to the patient. The consent can be withdrawn at any time during the study. Data was handled confidentially and coded. Data collection protocol can be consulted in [79].

4.1.1 Clinical Parameters and Outcome Assessment

The noted demographic, clinical, and trauma-related parameters were age, gender, trauma cause, impact seizure, history of prior seizures or TBI, alcohol ingestion and pupil response.

During the stay in the ICU, the administration of sedatives (midazolam and propofol) and anti-epileptic drugs (AED) were indicated. The occurrence of secondary insults such as hypotension (characterized as mean arterial pressure below 90 mmHg) and hypoxia (arterial oxygen saturation below 90%) was also reported. Laboratory parameters included were the glucose and Hb levels at the time of admission.

Concerning CT characteristics both Marshall and Rotterdam scales scores were noted. Information on the existence of basal cisterns, midline shift bigger than 5mm, epidural mass lesion, intraventricular air blood or traumatic subarachnoid hematoma (SAH) and occurrence or not of neurosurgical interventions (such as craniotomy and ICP monitoring) were available.

Regarding the GCS different measurements were obtained: GCS motor score, GCS in the site and GCS in hospital (at admission). Neurological outcome was evaluated with GOSE, 12 months after the admission using telephonic follow-up. GOSE was dichotomized into poor outcome (GOSE 1–3) and good outcome (GOSE 4–8).

Some patients (n=27) also presented recordings of ICP.

4.1.2 EEG Data

cEEG recordings started as soon as possible after admission to the ICU and continued for 7 days after injury, until discharge from the ICU or death. A total of 21 silver-silver chloride cup electrodes were placed on the scalp of the patients following the 10–20 International System. A Neuro center EEG recording system was employed. EEG was stored in the database of the clinical neurophysiology department. Files are saved in an edf format. The file's header has indications on the starting time and date of recording, number of data records, duration of the data records, number of recorded channels and corresponding labels. 19 channels were used in a G19 montage. In this montage, the names of the corresponding signals for each channel are Fp2, Fp1, F8, F7, F4, F3, T4, T3, C4, C3, T6, T5, P4, P3, O2, O1, Fz, Cz, Pz. Furthermore, the file also has indications on the transducer, units, pre-filtering and sampling rate of the signal.

4.2 Preprocessing

Files belonging to the same patient were ordered by date and hour of acquisition and divided into 1h intervals. For each hour, a 5 minutes window was used for processing and feature extraction. The algorithm starts by using the 5 minutes in the middle of the duration of the interval, otherwise, it would look for the nearest quality approved 5 minutes window inside the one-hour interval. If no proper window would be found, the hour was excluded from the analysis.

4.2.1 Clean data and Artefacts removal

We firstly filtered EEG data using a zero-phase sixth-order Butterworth bandpass filter of 0.1-40 Hz and corrected EEG for the mean. A semi-automated algorithm to detect and remove artefacts within windows of 10 seconds in the common average was employed. We identified unrealistic high amplitudes, muscle artefacts, and flat/empty channels as artefacts. If the number of artefacts was excessive or the channel was flat, the channel was classified as a bad channel. If only present in small segments, those segments were removed from the signal. We also searched for ECG artefacts using temporal Independent component analysis (ICA) [80] (Matlab code available in: [81]) .

If the remaining signal was smaller than 2 minutes or a threshold of half the number of channels plus 3 as a minimum number of channels was not fulfilled, the signal was not considered for the analysis and another interval was searched for.

4.2.2 EEG Feature Extraction

In the frequency domain, it is possible to calculate the power spectral density (PSD, also mentioned as power spectrum) either using parametric or non-parametric methods [25, 39] . We obtained the PSD, using Welch's method. It is also possible to calculate features for different frequency bands. The bands differ slightly between authors [25–27, 39, 70, 82] but in this work are considered as following: Delta: 0.5 and 4 Hz; Theta: 4 to 8 Hz; Alpha: 8 to 13 Hz; Beta: 13 to 30 Hz; Gamma: > 30Hz;

The features extracted from the EEG signals were the following:

- **Mean Amplitude**: corresponds to the standard deviation of the signal [5] and was calculated for each channel.

- **Coherence:** is a non-directed spatio-temporal feature. It is considered the correspondent to a cross-correlation in the frequency domain and reflects how the changes of frequency components between channels are synchronized [39]. It can be calculated as the mean of all magnitude squared coherences (function available in Matlab [83]) between all possible combinations of channels [5]. We calculated it in each channel for all frequency bands and averaged using a Hann window of 4 seconds and an overlap of 2 seconds. Results in a value between 0 (not synchronized) and 1 (synchronized).

- **Regularity:** allows the evaluation of the continuity of a signal and indicates if the amplitude is within normal values [34]. Higher values indicate a more regular amplitude of the signal. We calculated it for each channel and averaged it, according to equation 4.1 [34], where N is the length of the signal q.

$$Regularity = \sqrt{\frac{\sum_{i=1}^{N} i^2\, q(i)}{\frac{1}{3} N^2 \sum_{i=1}^{N} q(i)}} \tag{4.1}$$

- **Spatial Brain Symmetry Index:** The Brain Symmetry Index (BSI) calculates the symmetry between every two pairs of electrodes from left and right hemispheres [84–86] i.e. allows the quantification of spatial and temporal spectrum characteristics. This feature is useful to monitor changes in brain function [86]. The index allows the quantification of ischemic damage [84] and takes values between 0 (symmetric) and 1 (not symmetric). Formulas for BSI can be found in [85]. If any flat or bad channels were present they were not included in the analysis but the order was preserved.

- **Shannon Entropy:** this concept allows to quantify uncertainty or randomness and is, for this reason, a measure of complexity and predictability in the time-series observed [5,39,87]. Higher entropy corresponds to a less predictable system [39]. It was calculated for each channel (following the equation 4.2 [34, 88], where where x_i the amplitude of the signal and

$p(x_i)$ the probability of its occurrence in the signal, estimated using the histogram) and then averaged.

$$Shannon\ Entropy = -\sum_{i=1}^{N} p(x_i)\ \log_2 p(x_i) \tag{4.2}$$

- **Total Power:** calculated as the sum of the PSD of all frequency bands [5].
- **Alpha-delta ratio:** several ratios can also be calculated between frequency bands and used as features [39]. The alpha delta ratio (ADR), as the name indicates, is the division of the alpha frequency band by the delta frequency band [5].
- **SEF90:** The spectral edge frequency 90% (SEF90) is the frequency at which 90% of the total power lies below this defined cut-off [5].
- **Fitting oscillations and one over f (FOOOF)**: The signals can show both periodic and aperiodic components that have been associated with a diversity of states (physiological, cognitive, behavioural) and diseases. Therefore it is possible to characterize the shape of the PSD in one aperiodic component ($1/f^{a}$) distribution with an exponential decrease across frequency bands, while the oscillations, considered the periodic component, can be seen as peaks of power above the aperiodic component in the frequency domain. The aperiodic exponent, which can be calculated as the negative slope of the PSD in log-log space, represents the pattern of aperiodic power across frequencies. The 'offset' allows to reflect the uniform shift of power across frequencies [82, 89]. Figure 4.1 presents a representation of these components. The FOOF analysis described in [82] including the 'FOOOF' toolbox for Python and Matlab wrapper were used to extract the aperiodic offset and exponent. Code available in [90].

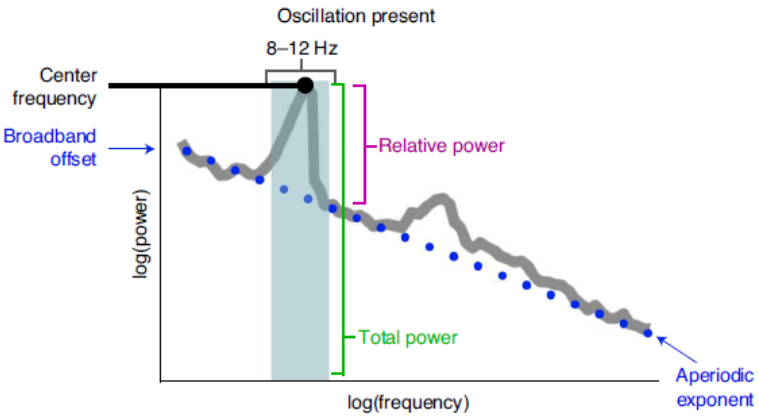


Figure 4.1: Representation of the aperiodic component of the power spectrum, described by an aperiodic exponent and an offset. Reproduced from [82].

The following features were extracted for the delta, theta, alpha and beta frequency bands:

- **Absolute Power:** is the integration of the PSD within each frequency band.
- **Relative Power:** calculated as the PSD in the frequency band divided by the total power [5, 70].
- **Variability per frequency band**: the ratio between the median absolute deviation and the median power [5, 70].
- **Center of gravity:** The center of gravity in the left-right and anterior-posterior brain direction are two time-and-frequency dependent functions which allow both to quantify the distribution of power over the head and to represent topographically where the power is maximum for each frequency band [5, 91]. Center of Gravity in the left to right and anterior to posterior direction was calculated for each frequency band separately and all, according to the equation 4.3 from [91]. The Fourier coefficients ($A_{i,j}$) are weighted in the direction d, which can be left to right or anterior to posterior. If there was any bad channel, that channel and its symmetric in the brain were excluded from the analysis.

$$Center\ of\ gravity = \frac{\sum_{j=1}^{M} A_{i,j}\, d_j}{\sum_{j=1}^{M} A_{i,j}} \tag{4.3}$$

- **Detrended Fluctuation Analysis (DFA):** Physiological processes might manifest fluctuations without characteristic scales. Often, the use of mean or median might be considered incorrect in systems that are considered "scale-free" [92]. For that reason, DFA, which allows quantifying power-law scaling, has received attention since it is considered a method to analyse the scaling behaviour of a time series and to determine power-law behaviour [92,93]. Power laws adopt the same form in any time scale, depending only on a scaling factor with exponent H - Hurst exponent. The scaling characteristics of amplitude and phase fluctuations may help understand brain operation in a critical state and comprehending cortical functioning. Therefore, the DFA allows to analyse the long-range temporal auto-correlation structure of EEG activity and global amplitude and phase synchronization [94]. It aims to remove trends in the signal and determine the mean squared fluctuations in consecutive intervals. The result of the DFA is the estimation of the Hurst exponent. An exponent H= 0.5 represents an uncorrelated process with no memory, while H > 0.5 represents the presence of long-range temporal correlations and memory. The DFA analysis which was obtained using the implementation of [94, 95] (Matlab code available: [96]) allowed us to determine power-law behaviour. The Hilbert transform was employed to extract the analytical signal of bandpass filtered data. DFA allowed us to estimate the phase and amplitude envelope dynamics across each frequency band. As input to the DFA analysis, either the sum of the amplitude envelopes (which we mention as DFA amplitude) or the global phase (mentioned as DFA phase) dynamics were used. Depending on the frequency band the window size

ranged between one oscillation to one-fourth of the recording time for the auto-correlation function computation.

- **Broken Detail:** Detailed balance is a concept that insinuates that the transitions between two states are pairwise (i.e. without any specific preference between states). The brain breaks detailed balance at large scales when it executes physical movements, processes information and performs cognitive functions. Thus, it is also expected that violations of detailed balance in neural dynamics increase with physical and cognitive effort. States of the brain can be estimated from the first two principal components of the bandpass filtered EEG data based on their current and future magnitudes to estimate the transition probability matrix. The asymmetry of this transition probability matrix or the entropy production and the curl of the corresponding probability flux can be used to define the broken detailed balance. Broken detail approach was followed as described in [97] (Matlab code available: [98]). The value of the sum of the probabilities of the flux vector (mentioned as S) and curl of flux for each frequency band was obtained.

It is important to note that the center of gravity and brain symmetry are spatial features and therefore the order of the channels was taken into consideration.

4.3 Temporal Evolution

In order to evaluate the temporal evolution of the features, a regression was fit for each feature in different periods after the first 12h after admission, when available, in 24, 36, 48 and over 48 hours possible intervals. A total of 15 intervals were defined according to figure 4.2.

Different regressions were applied, linear (4.4) and non-linear (4.5):

$$F(x) = p_1 x + p_2 \tag{4.4}$$

$$F(x) = p_1 x^2 + p_2 x + p_3 \tag{4.5}$$

where the variables p_1, p_2 or p_1, p_2 and p_3 were saved as inputs for the model. We also obtained the mean and standard deviation of the signal in each interval.

4.4 Machine Learning

Data were normalized using z-score normalization. GOSE outcome was dichotomized into poor (GOSE of 1,2 or 3) and good (GOSE >3). Categorical data was encoded. EEG predictors with the most correlations above 0.9 with other predictors were removed from the data.

We employed a feature selection method - relief-F [47, 99] to reduce the number of EEG and clinical predictors. Relief-F is an iterative, randomised and supervised approach. The goal of the

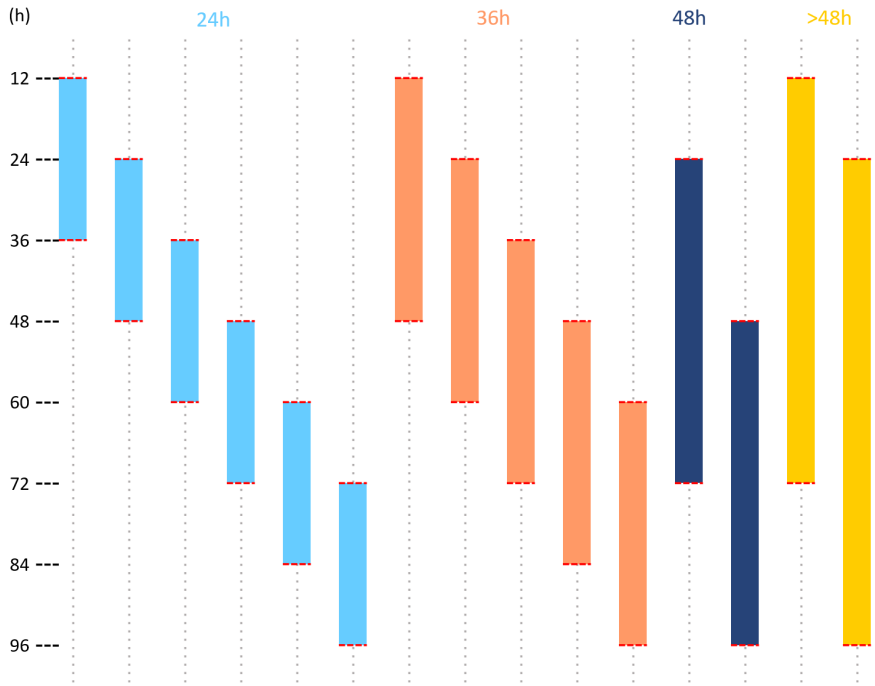


Figure 4.2: Time intervals defined. A total of 15 intervals was used, including durations of 24, 36, 48 and over 48 hours intervals.

filter is to estimate the ranking of the features based on their capability to differentiate data samples close to each other [47, 99]. The filter was followed by a backward elimination approach using Random Forest to find the best set of predictors for the model. Here, all features were used to train the model and the feature with the lowest importance was eliminated (feature importance ranking), and the model was retrained. AUC values as a function of the number of features were saved and the set of features corresponding to the maximum AUC value (after smoothing of the curve) was used as the optimal number of features. Figure 4.3 provides an overview of the methodology employed for feature selection. The optimal set of features was then applied to a Random Forest classifier to predict clinical outcome.

4.4.1 Random Forest Classifier

Random Forest is a supervised ML algorithm that allows both classification and regression tasks. The principle behind this algorithm is to build several decision trees [100–102] .

To understand Random Forest it is necessary to understand how decision trees and generated. They are composed of three components: decision nodes, leaf nodes and a root node. The training data is divided into branches which further divide into other branches. This iteration continues until a leaf node, which can no longer be segregated, is reached. The nodes represent the attributes used for the prediction of the output [101].

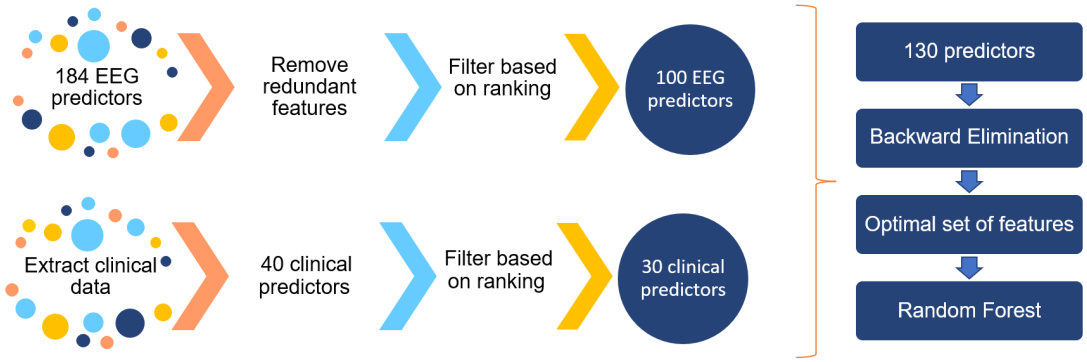


Figure 4.3: Schema of the feature selection methodology.

Random Forest is an ensemble of decision trees, thus having more than one individual model contributing to the prediction. In this case, the ensemble is based on bagging techniques (or bootstrap aggregation) in which random samples from the data set are chosen to generate the model - bootstrap step. Each model is trained individually and the output is generated considering predictions from all models - aggregation step [100, 102]. In classification tasks, the majority of the votes is used while in regression the average of the decisions is taken into consideration [100]. Results are usually better for classification [100].

Some important aspects of Random Forest make its use advantageous. The use of subsets of data allows not only to obtain a big diversity, since not all samples are used in each tree, creating always different trees but also avoids overfitting [100–103]. Additionally, it is possible to view the features' relative importance in the prediction [102, 103]. Since all trees are created in parallel and independently, it also allows full use of the Central Processing Unit for generation [100] and allows the use of large data sets [101]. Furthermore, it is immune to the curse of dimensionality [100]. The averaging/voting allows us to obtain stability in the results [100]. Random Forests handles continuous (regression) and categorical (classification) variables in the data sets and performs well even if the data contains missing or null values [100, 101]. However, Random Forests can be considered complex when compared to decision trees since they required more computational resources and more time [100, 101].

Hyperparameters to take into consideration related to the predictive power are the number of built trees, the maximum number of features considering in a splitting node, and the minimum number of examples in each leaf for internal splitting [100, 102]. In this work parameter tunning was performed using the out-of-bag (OOB) error to determine the optimal number of trees. This method considers the error on the OOB samples, which are the samples that are not selected in each bootstrap done by the Random Forest [104]. The optimal number of trees was selected as the index in which the error converged [103] (consult Appendix B for more information). A quantile error and bayesian optimization (MatLab code available in [105]) were employed for the minimum leaf size, the number of predictors to sample and in-bag fraction. In the Bayesian Optimization

method for hyper-parameter tuning, the model is run several times with different sets of hyper-parameter values [106]. However, contrary to other approaches to parameter tuning, this method evaluates the information from the previous model in order to select the hyper-parameter for the most recent one. This approach is considered to find the highest accuracy model faster [106].

The predictive value was evaluated in a 5-fold-cross validation, using 80% of data to train and 20% to test. We evaluated and compared the results using sensitivity, specificity and the AUC of the ROC curve values. The threshold was defined as the optimal operating point of the ROC curve using FPR and TPR. This optimal operating point is obtained with the function *perfcurve* [107] - which moves the straight line with slope S from the upper left corner of the ROC plot (FPR=0, TPR=1) down and to the right until it intersects the ROC curve. Values were averaged for 5 cross-validations and displayed with a 95% confidence interval.

Values from the different time intervals were fed to the model, which for each selected a different set of features. Models using only clinical, only EEG and both EEG and clinical predictors combined were compared. The inclusion of ICP as a predictor in the prediction model was also assessed. Differences between model predictions (with a statistical significance of $p < 0.05$) were assessed using McNemar's test. The McNemar's test is a non-parametric test for paired data and allows the evaluation of the changes in proportion for the paired data [108]. At last, a comparison with actual time point results and temporal evolution (trends over time) was also performed.

Chapter 5

Results

5.1 Database

The present database contains a total of 111 patients. From this set, patients were excluded for lack of EEG data before 96h after trauma (n=3), when EEG total recordings after cleaning were 2h or less (n=1), if concurrent with post-anoxic coma (n=3) and due to missing GOSE outcome scores (n=2). After exclusion, 102 patients were left for inclusion in the analysis, the majority male (n=75). Table 5.1 presents their baseline characteristics. The dataset is balanced in terms of the number of patients with poor (GOSE =1-3) and good (GOSE > 4) outcome: 54 of the sustained TBI resulted in a poor outcome, from which 43 patients died (GOSE = 1), contrasting to 48 patients that had a good admission outcome (GOSE 3-8). Patients with poor outcomes were significantly older than patients with good outcomes, being a mean age of 57 and 41 years old, respectively. GCS at admission were similar between both groups, since all patients included were admitted at the hospital with a moderate to severe TBI, but outcomes were very different. All values can be consulted in the table below. More details on the CT assessment, cause of trauma, and pupil evaluation can be found in Appendix A.

Table 5.1: Patient's characteristics grouped in a poor and a good outcome.

	Poor Outcome GOSE= 1-3 (n=54)	Good Outcome GOSE = 4-8 (n=48)
Gender | females (%)	13 (24,1%)	14 (29,2%)
Age in years | median(IQR)	59 (28)	42 (35)
GCS in hospital | median(IQR)	3 (5)	3 (6)
GOSE | median(IQR)	1 (0)	6 (2)

5.2 Models' Performance

The values for the AUC in different time intervals with EEG and admission predictors varied considerably and the results of AUC for train and test set are presented in table 5.2.

Table 5.2: AUC for train and test set. The AUC values for 30 tested models using admission and EEG predictors vary from 0.87 and 0.97 in the training set and between 0.65 and 0.85 in the test set. Values presented include the confidence interval. Models were identified using the fit function type (poly1 for linear and poly2 for polynomial) and the hours of start and end of the time interval used. The two models considered as the best models are in bold.

Interval (h)	FIT	AUC train	AUC test
12-36	poly1	0.91 [0.86-0.96]	**0.83 [0.75-0.91]**
	poly2	0.92 [0.85-0.98]	0.79 [0.69-0.88]
12-48	poly1	0.97 [0.94-1.00]	0.79 [0.73-0.85]
	poly2	0.96 [0.94-0.98]	0.75 [0.57-0.93]
12-72	poly1	0.88 [0.87-0.89]	0.80 [0.65-0.95]
	poly2	0.87 [0.83-0.91]	0.76 [0.60-0.92]
24-48	poly1	0.94 [0.89-0.99]	0.80 [0.65-0.94]
	poly2	0.87 [0.83-0.91]	0.76 [0.65-0.87]
24-60	poly1	0.92 [0.88-0.96]	0.76 [0.65-0.88]
	poly2	0.93 [0.86-1.00]	0.74 [0.63-0.85]
24-72	poly1	0.88 [0.79-0.97]	0.71 [0.66-0.76]
	poly2	0.91 [0.79-1.02]	0.72 [0.59-0.85]
24-96	poly1	0.88 [0.75-1.00]	0.70 [0.47-0.93]
	poly2	0.89 [0.79-0.99]	0.75 [0.63-0.86]
36-60	poly1	0.91 [0.86-0.97]	0.79 [0.76-0.82]
	poly2	0.93 [0.85-1.02]	0.65 [0.42-0.87]
36-72	poly1	0.89 [0.77-1.00]	0.78 [0.72-0.84]
	poly2	0.96 [0.92-1.00]	0.76 [0.62-0.91]
48-72	poly1	0.97 [0.95-0.99]	**0.85 [0.73-0.97]**
	poly2	0.95 [0.90-0.99]	0.83 [0.77-0.89]
48-84	poly1	0.95 [0.91-0.98]	0.81 [0.64-0.97]
	poly2	0.97 [0.94-1.00]	0.81 [0.63-0.98]
48-96	poly1	0.89 [0.86-0.92]	0.78 [0.68-0.89]
	poly2	0.95 [0.89-1.01]	0.81 [0.66-0.96]
60-84	poly1	0.95 [0.89-1.01]	0.79 [0.66-0.92]
	poly2	0.94 [0.89-0.99]	0.79 [0.67-0.92]
60-96	poly1	0.94 [0.87-1.01]	0.80 [0.65-0.95]
	poly2	0.96 [0.93-0.99]	0.83 [0.74-0.93]
72-96	poly1	0.95 [0.91-0.99]	0.77 [0.70-0.85]
	poly2	0.93 [0.87-0.99]	0.77 [0.60-0.93]

From the 30 available models, we selected the best two. The model with the highest AUC in the test set was obtained using data between 48h to 72h with both admission and EEG predictors

(figure 5.1). It predicted outcome yielding an AUC of 0.85 [0.73-0.97], sensitivity of 0.91 [0.83-1.00] and specificity of 0.67 [0.32-1.00] in the test set. Moreover, in an earlier time interval, using data from 12h to 36h, the following values were obtained: AUC = 0.83 [0.75-0.91], sensitivity of 0.83 [0.72-0.94] and specificity of 0.80 [0.61-0.94] (figure 5.2) in the test set. These two models were the target for further analysis. The correspondent ROC curves and evaluation metrics can be found in figures 5.1 and 5.2 for both the train and test sets. Values of AUC, sensitivity and specificity are presented with a 95% confidence interval.

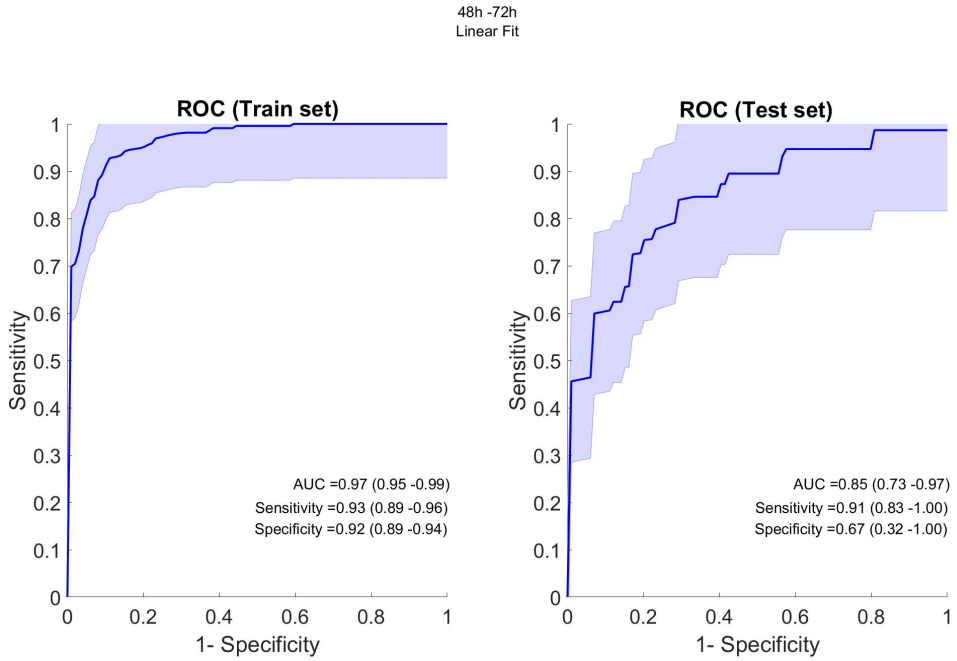


Figure 5.1: ROC curve. Train set (left) and test set (right) for the 48h-72h interval and linear fit.

5.2.1 Predictors

The results for the previously mentioned time intervals, using EEG and admission predictors combined and independently can be observed in table 5.3. Models combining EEG and admission predictors obtained better AUC than models using them separately. The models were proved to be different by Mcnemar's test at a 5% significance level. The inclusion of ICP as a predictor of the outcome of TBI did not improve the performance. ICP was not selected after the feature selection stage and therefore had no contribution to the prediction.

Furthermore, the AUC results obtained using features from the exact time points instead of the temporal information (slopes, offsets, mean and std) ranged from 0.73 to 0.80. Results for 36h

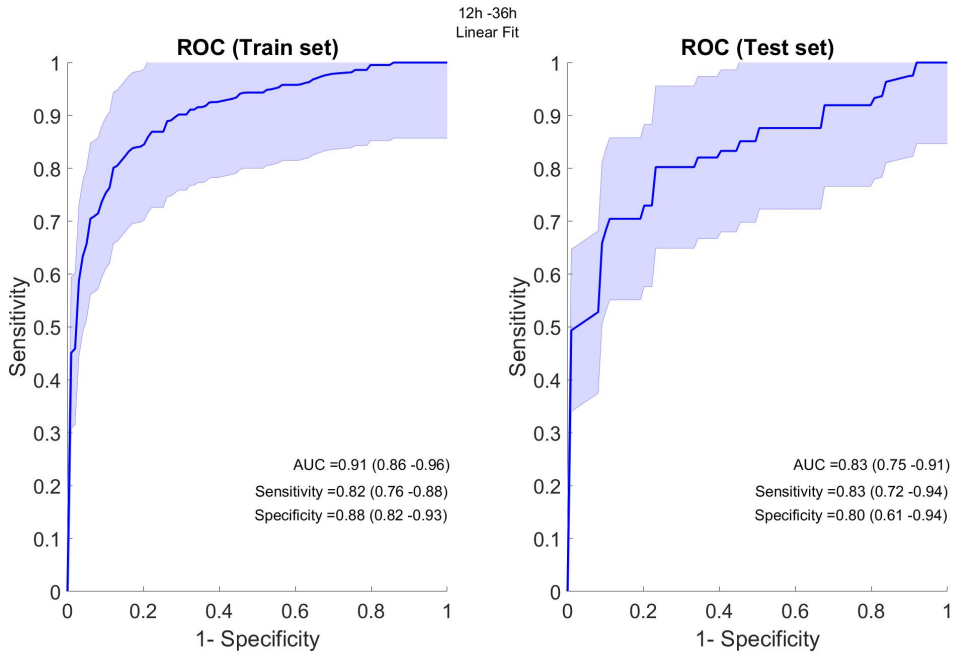


Figure 5.2: ROC curve. Train set (left) and test set (right) for the 12h-36h interval and linear fit.

Table 5.3: Model performance using EEG predictors, admission predictors and both EEG and admission combined. Values of AUC, sensitivity and specificity are presented with a 95% confidence interval. Values of AUC are higher when EEG predictors are combined with the admission ones.

Time Interval	12h-36h Linear			48h-72h Linear		
Predictors	EEG	Admission	EEG + Admission	EEG	Admission	EEG + Admission
AUC	0.72 [0.61-0.83]	0.71 [0.58-0.84]	**0.83 [0.75-0.91]**	0.75 [0.68-0.83]	0.77 [0.67-0.86]	**0.85 [0.73-0.97]**
Sensitivity	0.83 [0.52-1.00]	0.86 [0.69-1.01]	0.83 [0.72-0.94]	0.81 [0.57-1.00]	0.80 [0.60-1.01]	0.91 [0.83-1.00]
Specificity	0.57 [0.12-1.00]	0.58 [0.44-0.73]	0.80 [0.61-0.94]	0.65 [0.19-1.00]	0.74 [0.55-0.93]	0.67 [0.32-1.00]

and 72h, which are comparable with the time intervals of 12h-36h and 48h-72h, are indicated in table 5.4.

Table 5.4: Model performance using only EEG features from the 36 and 72 hours after injury compared with time intervals between 12-36h and 48-72h. AUC, sensitivity and specificity results in the test set are presented with 95% confidence interval. AUC is greater in the 12-36h model than 36h as well as in 48h-72h compared to 72h.

Time	36h	12-36h	72h	48-72h
AUC	0.78 [0.72-0.84]	**0.83 [0.75-0.91]**	0.80 [0.68-0.92]	**0.85 [0.73-0.97]**
Sensitivity	0.81 [0.69-0.83]	0.83 [0.72-0.94]	0.74 [0.51-0.98]	0.91 [0.83-1.00]
Specificity	0.81 [0.64-0.97]	0.80 [0.61-0.94]	0.79 [0.57-1.00]	0.67 [0.32-1.00]

5.3 Feature Selection and Importance

The feature selection with backward elimination resulted in the AUC as a function of the number of features (Figure 5.3). The AUC increases while the number of features decreases until the optimal data set is found. The ideal number of features used to make predictions changed for each model.

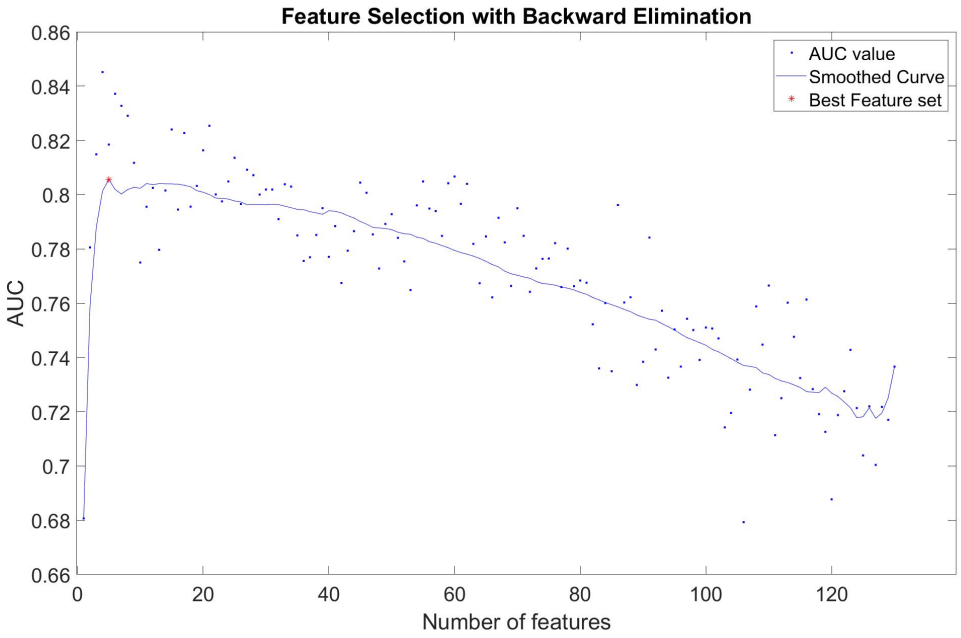
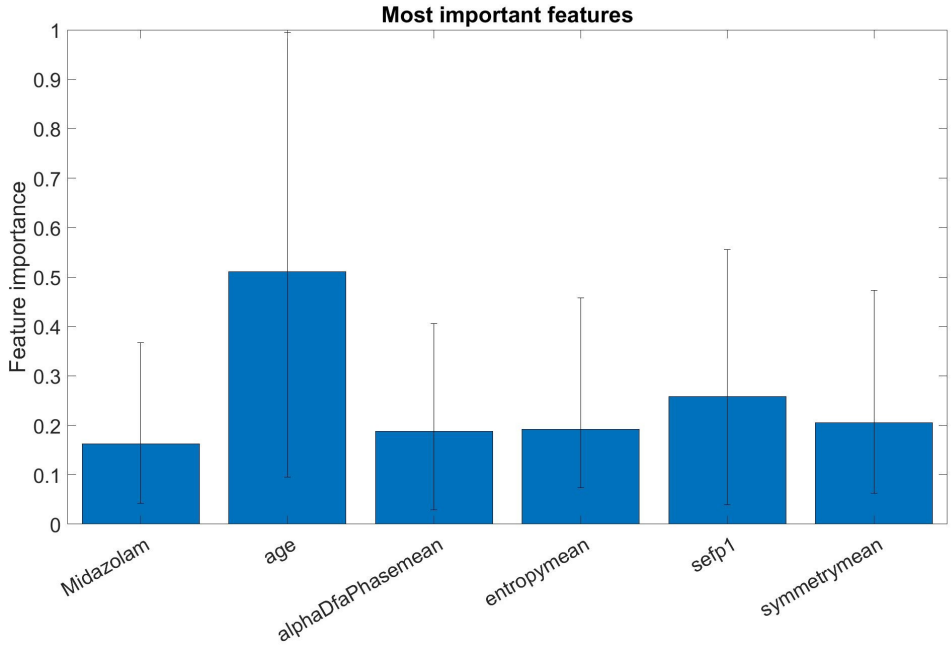
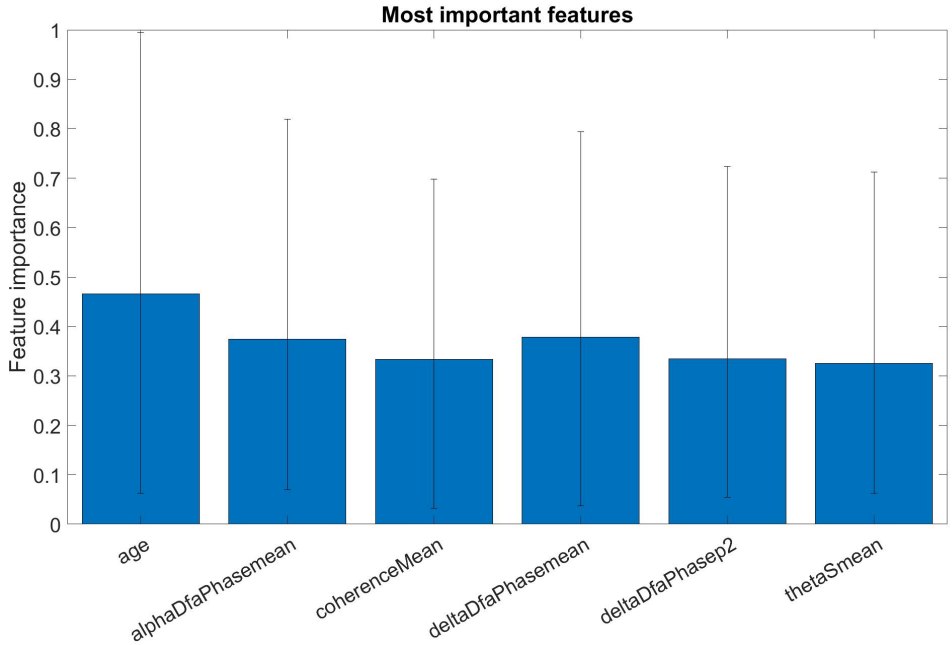


Figure 5.3: Example of the AUC as a function of the number of features used. The optimal set of data that resulted from the backward elimination is marked on the plot. In most models the ideal set of features for prediction was small (<20).

For the time interval 48h-72h, the most important features were age, the mean value of the Hurst exponent using the phase dynamics in the delta band (deltaDfaPhaseMean), the mean value of Hurst exponent using the phase dynamics in the alpha band (alphaDfaPhaseMean), the p_2 of Hurst exponent using the phase dynamics in the delta band (deltaDfaPhasep_2), the mean of the coherence, and the mean value the sum of the probabilities of the curl flux vector in the theta band (thetaSmean) (Figure 5.4b), in respective order of importance. For the time interval 12h-36h, the most important features were age, the slope of SEF90%, mean value of BSI (symmetry), mean of entropy, the mean Hurst phase in the alpha band (alphaDfaPhaseMean), mean of entropy and administration of midazolam (Figure 5.4a).



(a) 12h-36h Linear fit



(b) 48h-72h Linear Fit

Figure 5.4: Feature Importance for the two selected models, including standard deviation of the cross-validation iterations. alphaDfaPhasemean and deltaDfaPhasemean stand for the mean of the Hurst exponent in phase dynamics in the alpha and delta band respectively (results from the detrended fluctuation analysis - DFA). thetaSmean stands for the mean of the sum of the curl of the probability flux in the theta band (results from the Broken Detail analysis).

Chapter 6

Discussion

The purpose of this study was to obtain a ML algorithm capable of predicting the outcome after TBI using EEG and admission data. This study was spawned from the lack of research on the contribution of EEG trends over time to the prediction of outcome in TBI patients. The main aim herein was to investigate how changes in EEG features over time could improve the prognostication of the disease when combined with admission parameters.

The model to predict the outcome of TBI with the highest AUC was obtained for the interval between 48h to 72h (AUC = 0.85 [0.73-0.97], sensitivity = 0.91 [0.83-1.00], and specificity = 0.67 [0.32-1.00]) (figure 5.1). The model using data between 12h to 36h hours (AUC = 0.83 [0.75-0.91]; sensitivity = 0.83 [0.72-0.94]; specificity = 0.80 [0.61-0.94]) (figure 5.2), also obtained good performance and is balanced, both in sensitivity and specificity, as well as in the values obtained for the confidence interval. Both models incorporated admission and EEG predictors, using a Random Forest algorithm with a 5-fold cross-validation. Overall, the combination of admission predictors with the EEG features outperformed models trained with only admission or EEG features separately. The findings also hint at a better performance using EEG trends (regression features) compared to features from single time points.

This study allows for an early prediction of neurological outcome following a TBI, by using multiple EEG features and a ML approach. We believe these points stress the importance of continuously monitoring the EEG and encouraging its adoption in moderate to severe TBI patients in the ICU, as it is frequently done in other brain injury populations.

6.1 Models' performance

AUC values between linear and polynomial models for the same interval were different. The findings suggest that one function fits the data better than the other, depending on the time span selected.

Furthermore, differences between time intervals can be caused by the absence of data in specific time points within each interval, for example, due to the patient's dismissal from ICU or death, or due to surgical interventions, which cannot be prevented in patients with a moderate to

severe condition, as the ones in the data set used. When this occurs, it results in a regression with a bigger error and less adjustment to the available data. Thus, it was expected that the results obtained were both different between fit-type functions and between time intervals.

Numerous studies have attempted at predicting the outcome of a TBI. Comparison between studies is challenging due to disparities amongst populations of different cohorts. There are models available for other brain injuries, or in data sets with other characteristics (size, age, causes of trauma, severity, etc), different times of outcome assessment, and the actual scale for assessment varies significantly. Additionally, occasionally different evaluation metrics are employed, which does not allow to have a direct comparison between different studies. Using only admission variables, values obtained are scattered. Markedly, the CRASH studies [57] use as a performance metric the C statistic. They obtained a C value of 0.81 in the prediction of mortality in high-income countries and a C value of 0.86 for mortality after 14 days. It is not possible to compare these values with our study. Previous studies [69] have also reported values for sensitivity and specificity of tree models for the prediction of outcome. The values were 72.3% [66.4–77.6%] and 62.5% [54.9–69.6%]. These values are lower than the ones obtained for our two best models, which have sensitivity of 0.83 [0.72-0.94] and 0.91 [0.83-1.00] and specificity of 0.80 [0.61-0.94] and 0.67 [0.32-1.00], as it is evident from table 5.3. Our models using only EEG or only admission variables also obtained a sensitivity higher than 72.3%. However, the specificity values were only higher for the model from 48h-72h. Furthermore, studies [60] have reported the prediction of the outcome of TBI focusing on CT characteristics. Results yield an AUC of 0.83 for predicting unfavourable outcome but only 0.69 in surviving patients. Therefore the results were shown to be insufficient for predicting disability in survival patients. An outcome prediction model developed with admission characteristics using cross-validation [59] obtained an AUC of 0.80 [0.66 - 0.84]. This value was one of the closest to the work carried by us, amongst the ones found in the literature for the prediction of the neurological outcome of TBI at 6 months. The prognostic model was developed using a data set with 8,509 patients that suffered from moderate to severe TBI. Employing a data set 77 times the size of the one used herein, permits the model to obtain better generalisation, possibly leading to better performance in the test set. As our results were obtained in a smaller set of data, we conclude that a comparison between the works is not accurate. Furthermore, the authors employed an external validation which was not possible for us with the data available. Regarding studies employing the use of EEG, a study [5] obtained values of AUC in the validation set of 0.81 combining EEG and IMPACT predictors with admission and EEG information at 72h and 96h together. Our model obtained a higher AUC, at an earlier time point, but with an overlapping confidence interval.

As far as we know, no one has attempted to predict the outcome of TBI by considering the temporal evolution of the EEG features as we do in this study. Furthermore, the use of cross-validation is not always employed in all studies. The use of cross-validation, verification of the 95% confidence interval and evaluation in a test set constitute important steps to provide trustworthy results. To our knowledge, there are also no records of prediction of TBI by using EEG features as early as 36h with similar performance to the one obtained in this study.

All in all, our models obtained a good performance in determining the outcome of patients who suffered a TBI compared to the available literature. We resorted to the evaluation of EEG trends over time and combined them with admission predictors. We obtained predictions with good performance as early as 36h and 72h after trauma. The models presented might provide an additional tool for clinicians during the treatment of TBI, helping decision-making and proper allocation of resources. However, as we will discuss in chapter 7, before implementing it in clinical practice we still believe some future work to be required.

The use of ML to obtain the described results was critical. Random Forest was selected as the preferred algorithm due to its ability to deal with missing values, which occurs often since the EEG recording may stop due to medical interventions. Furthermore, it is also robust to outliers and noise [103], which is extremely important since it is impossible to remove all artefacts in the signals. It is also fast and simple compared to other algorithms [103]. Finally, one of the most important factors for selecting Random Forest is that provides the user with internal estimates of error and variable importance [103], which is to be taken into consideration when working with medical data. This algorithm allows us to understand the features being used, providing more confidence and comfort for the medical staff.

6.1.1 EEG Contribuition

The performance of the model with only EEG features was lower when compared with the models combining EEG and admission features. The number of patients did not allow us to use only EEG features: cEEG among patients do not have the same duration of recordings as hospitalisation time varies, and some time intervals have fewer data available.

Moreover, the protocol for data collection used for this work was aimed at a study focused on EEG. Therefore, some clinical variables presented in the literature were not noted, which did not allow to obtain a good performance considering only admission data. To sum up, the combination of EEG features with the admission variables is essential to obtain good discriminant performance and generalisation.

The recording of cEEG after TBI is widely used to detect seizures [75] but it is not a standard practice to predict the outcome [5]. The results presented in this work support the implementation of this procedure for the prediction of the outcome of TBI in the clinical setting. Nonetheless, as previously mentioned, the use of cEEG in daily clinical practice can be hampered by technical, patient-related or system resources [33], thus making its practice not always possible.

6.1.2 Trends in EEG *vs* Fixed points

Our current work suggests more fidelity in models using temporal evolution rather than defined points. One disadvantage of using continuous EEG is that recording and preprocessing the signals is not always easy [109]. In our data set, often there are medical interventions that can take hours in the ICU that are detrimental to the recordings. Furthermore, external and environmental artefacts

are also present in EEG [109]. The unwanted components are not always removed, affecting the value of the features for single data-points. When such happens, either a suitable window for feature extraction is not found or the feature is miss-calculated. Therefore, if that would occur at the moment of the pre-defined time point (24h, 36h, 48h, 72h) the data would not be available. . The use of the regression functions to approximate the data allows to redraw the information from the data, dealing properly with missing data or outliers that caused by miss-calculation of the features. This could explain temporal evolution models outperforming the fixed time point.

The changes in the values of the features over time may have some significance in the outcome of a patient with TBI. This is the case when p_1, p_2 or p_3 values (the slope of the feature over time) are selected as important features for the prediction. However, in our models, it was frequent for the mean values in the specified time interval to be selected. In those cases, the changes over time might not be as significant. Basically, to summarize, the use of the mean value of the feature is diminishing the impact of outliers in the data.

6.1.3 Time Window

The different intervals used resulted in different performance although the algorithm employed was the same. Most models, independently of the interval, achieved good performance. There was no particular pattern observed in the data through time since a good performance was obtained for both earlier and later time intervals. There is also no relevant conclusion to redraw from the duration of the time intervals. The two models selected as the best ones were both 24 hours of duration intervals. However, AUC values above 0.80 were found in 36 hours intervals (48h-84h and 60h-96h), 48 hours intervals (48h-96h), and over 48 (12h-72h). We can only conclude that by using information from 24 time points it is already possible to obtain a good prediction.

6.2 Feature Importance

As previously mentioned, both variation in time (slope), as well as the mean values in the time interval, were frequently selected as important features. The feature selection was different from one interval to another. Such findings hint at changes in features' relevance throughout different periods of a patient's stay in the ICU. However, some intervals have less available data, due to the late start of recordings of the signals, dismissal from the ICU (patients passing or awakening) or due to interruptions for medical interventions. These factors do not make it possible to determine if the change of the features between intervals is due to the features' importance over time or due to a lack of data.

Our best TBI prediction models showed strong discriminating ability in regards to age. Age has been used several times in outcome prediction of TBI [5,36,57,59–62,68,110–113] and proven to be highly relevant, supporting the theory that younger brains are more likely to recover or that age could be influencing the doctor's treatment decisions.

Our work also a showed strong discriminating ability concerning EEG features. The use of EEG features has been reported by other authors with successful results for both the outcome

prediction of TBI and the correlation of the EEG features with the outcome score [5, 70–72]. Brain symmetry, which evaluates damage in the brain [84] and helps in monitoring changes in brain function [86], entropy, which quantifies the complexity of the signals and allows for the analysis of their dynamic behaviour [114], SEF90, that has been described as an parameter in the prediction of neurological outcome [5], and coherence, a measure of "space synchrony" that was linked with a reduced cognitive function [115], were selected by our models and have also been previously described in literature [5, 8, 71, 72]. Therefore, their appearance as relevant parameters was not surprising. However, our models also showed a strong discriminant ability for features that are not as trivial such as the Hurst exponent of the Detrended Fluctuation Analysis. Nonetheless, the Hurst exponent was shown to be a helpful measurement for the interpretation of epileptic and interictal EEGs and their neurodynamics [116]. For that reason, one would expect that it contributed to the prediction of the TBI. At last, there have also been studies relating the DFA and neurological functioning that concluded that values from DFA were associated with neurological outcomes after TBI [117], supporting our results. Finally, the broken detailed analysis also showed to be relevant since it was found that the sum of the probabilities of the flux vector were relevant for the prediction. Although this feature has not been reported to be related to the outcome of TBI previously, it has been mentioned that evaluating the violations of the detailed balance is a way of quantifying entropy production [97]. Furthermore, big-violations also indicate asymmetries in the system [97]. Therefore it is reasonable that this measure can quantify the extent of damage of the brain after a TBI.

Some EEG features mentioned in the literature and also included in this work did not show a high predicting value. Examples are the variability of the relative power [70, 71], variability of the mean frequency [71], total power [71], absolute power [70]. Furthermore, admission features also mentioned in literature but not found in the most relevant features in our work include the occurrence of multiple injuries [110] or extracranial injury [57], GCS [57, 60, 68], motor GCS [59, 110], pupil reactivity [57, 59, 60], mechanical ventilation [110], presence of hypotension [60] and the degree of midline shift in CT scan [23] or presence of SAH [59]. ICP also did not appear to contribute to the prediction. However, the latest might be due to the reduced number of patients with this measurement available.

The EEG measurements and the calculation of the features can be a more time-intensive process than the sole use of admission variables. However, the results obtained allow us to conclude that the incorporation of EEG in a predictive model is relevant.

Chapter 7

Conclusion and Future Work

We were able to predict the outcome of TBI based on EEG parameters and the patient's admission information. One strength of this work was that employing cEEG dynamics over time allowed us to take into consideration the course of treatment and progression of the disease, which would not be possible when solely using assessments at the time of admission. We also demonstrated the potential of the Random Forest models for the purpose mentioned. We showed that the trends over time of the EEG features contain interesting information for the prediction of outcomes. Their potential is increased when combined with admission parameters. The model obtained is simple and clear and allows the support of medical staff. It is important to point out that the goal of such prediction models is to aid decision-making and improve healthcare treatments. It should never replace the doctor's decision and assessment.

One disadvantage of our study is that TBI is a very heterogeneous disease [118]. Our measurements for the outcome are solely based on the GOSE scale that is being divided into a good and a poor outcome. Since the boundary between a good and a poor outcome is not always as clear-cut, this measurement can be insufficient due to its variability in interpretation [119]. One suggestion would be to employ more than one measurement to distinguish between a poor and a bad outcome. Additionally, we suggest future work to explore the possibility of utilising a multi-class prediction. However, for this to be possible, a bigger data set will be required to ensure a reliable number of patients in each class.

In this work, we proved to obtain a reliable prediction model using ML. Authors have compared the performance of some ML algorithms. It was concluded that, among the tested algorithms, for the prediction of survival of TBI patients, cubic SVM, quadratic SVM, and Linear SVM performed better than Logistic Regression [67]. Other authors [68] compared several algorithms for the prediction of in-hospital outcome and mortality of TBI. They concluded that Random Forest had the best discriminant ability for the in-hospital outcome while Ridge Regression for the in-hospital mortality. Likewise, for predicting mortality, [66] 5 different ML algorithms were tested, concluding that ANN obtained the best performance, compared to Logistic Regression, SVM, Naive Bayes, and Decision Tree in moderate to severe TBI patients. On the same line, another study for prediction of outcome after TBI [111] concluded that among ANN, Naive Bayes,

Decision Tree and Logistic Regression, the highest AUC was found for ANN. Furthermore, when deciding on which algorithms to employ, Random Forest has several advantages such as allowing to understand features' importance [103], the ability to deal with outliers and noise [103] and avoiding overfitting [100–103]. However, in the future, it would also be interesting to evaluate the performance of other algorithms, giving preference to those that allow the understanding of the features' contribution to the prediction. As previously mentioned, the models should be simple and explainable in order to be possible to be implemented in the clinical setting. That being the case, the aforementioned parameters should be taken into consideration for any future work developed with a different algorithm.

To reach the clinic, models need to be largely validated in different cohorts. External validation allows us to update or tailor the approach [120]. The model obtained in our work is capable to generalise since a train and test set were employed with a 5-fold cross-validation. However, we suggest that external validation is employed, preferably with information from different centres, similar to what other authors did to validate CRASH, IMPACT and Nijmegen models [61]. However, it is important to note that such can be challenging since not many centres have the monitoring of cEEG for TBI patients as a standard practice yet. Furthermore, the use of a big data set also adds challenges in methodology, as previously described [120].

Finally, the employment of temporal evolution and the addition of the changes of features over time is a new idea. In this work, we showed the potential of applying linear or polynomial regressions to the data over time and using the slope and offset to describe its behaviour. However, we suggest future works adopt several functions to be used and selected individually (automatically, if possible) for each feature instead of using the same function type for all features. We also considered the development of an automated algorithm to identify outliers and exclude them from the fit function but it was not possible to determine when an outlier resulted from an artefact or an important sign of behaviour of the progression of the injury. Therefore future research on this topic remains necessary.

To conclude, the main goal of the project was to create a model capable of predicting the neurological outcome of patients following TBI. This goal was full-filled using EEG features and admission data in a Random Forest model. We concluded that EEG trends have an added value to the prediction. The model developed aims at aiding in-hospital decisions, and improving treatment management, which will ultimately lead to better healthcare.

Appendix A

Database Description

Very few patients had alcohol ingestion (n=3) and impact seizures (n=6). Drugs administrated (midazolam, propofol and AED) were also registered. Associated hypotension and hypoxia were higher in patients with poor outcome (20.4% and 9.3% versus 10.4% and 8.3%, respectively). 22 patients were submitted to craniectomy post TBI, 13 of them were patients with a poor outcome. Values are presented in table A.1.

Table A.1: Patient's characteristics, including medical history, trauma related factors, administrated drugs and occurence of secondary insults.

	Poor Outcome GOSE= 1-3 (n=54)	Good Outcome GOSE = 4-8 (n=48)
Alcohol Ingestion (n/%)	3(5.6%)	0(0%)
Impact Seizure (n/%)	4(7.4%)	2(4.2%)
Midazolam (n/%)	43(79.6%)	28(58.3%)
Propofol (n/%)	52(96.3%)	47(97.9%)
AED (n/%)	1(1.9%)	1(2.1%)
Hypotension (n/%)	11(20.4%)	5(10.4%)
Hypoxia (n/%)	5(9.3%)	4(8.3%)
Cranietomy (n/%)	13(24.1%)	9(18.8%)
Prior seizures (n/%)	2(3.7%)	1(2.1%)
Prior TBI (n/%)	2(3.7%)	2(4.2%)
Midline shift >5mm (n/%)	20(37%)	8(16.7%)
Epidural mass lesion (n/%)	11(20.4%)	9(18.8%)
SAH (n/%)	54(100%)	44(91.7%)

Table A.2 presents the frequency of the variables assessed at admission. Marshall's score ranges from 1 to 6 being the most common score 2 both in patients with good (64.6%) and poor outcome (46.4%). Similarly, the Rotterdam score was also assessed. The values range from 0 to 5, scoring

1 the most frequent (48.1% in poor outcome and 58.3% in good outcome). For most patients, there were no basal cisterns present. Of the possible causes of trauma falls (stairs or bicycles) were the most common cause of both good and poor outcome. From the pupil evaluation, most patients did not have pupil response (61.1% in poor outcome and 79.2% in good outcome).

Table A.2: Frequency table of the neurological tests at admission. Marshall, Rotterdam score and Basal Cisterns values are assessed through CT imaging. Pupil response is assessed at admission.

		Poor Outcome GOSE=1-3 (n=54)	Good Outcome GOSE = 4 - 8 (n=48)
Marshall score	1	0(0%)	4(8.3%)
	2	25(46.3%)	31(64.6%)
	3	5(9.3%)	3(6.3%)
	4	7(13%)	1(2.1%)
	5	16(29.6%)	9(18.8%)
	6	1(1.9%)	0(0%)
Rotterdam score	0	0(0%)	3(6.3%)
	1	26(48.1%)	28(58.3%)
	2	12(22.2%)	11(22.9%)
	3	9(16.7%)	5(10.4%)
	4	5(9.3%)	1(2.1%)
	5	2(3.7%)	0(0%)
Basal cisterns	0	40(74.1%)	41(85.4%)
	1	6(11.1%)	6(12.5%)
	2	8(14.8%)	1(2.1%)
Cause of trauma	Bicycle vs MV	12(22.2%)	8(16.7%)
	MC crash	2(3.7%)	2(4.2%)
	MV crash	4(7.4%)	9(18.8%)
	MV *vs* MV	4(7.4%)	3(6.3%)
	Person *vs* MV	4(7.4%)	5(10.4%)
	Assault	0(0%)	1(2.1%)
	Others	1(1.9%)	2(4.2%)
	Fall from stairs or bike	27(50%)	18(37.5%)
Pupil Response	0	33(61.1%)	38(79.2%)
	1	10(18.5%)	6(12.5%)
	2	11(20.4%)	4(8.3%)

Note: MV - motorvehicle; MC - motorcycle.

Appendix B

Parameter Tuning

There is a trade-off between the number of trees used and the time to train the models [121]. To surpass that challenge, the model was tested with a high number of trees and the out-of-bag error was observed. The OOB error becomes stable (converges) after a certain number of trees [103] (B.1), which can be used for further development of the models. This allows to decrease the time and space required for training each of the 30 models.

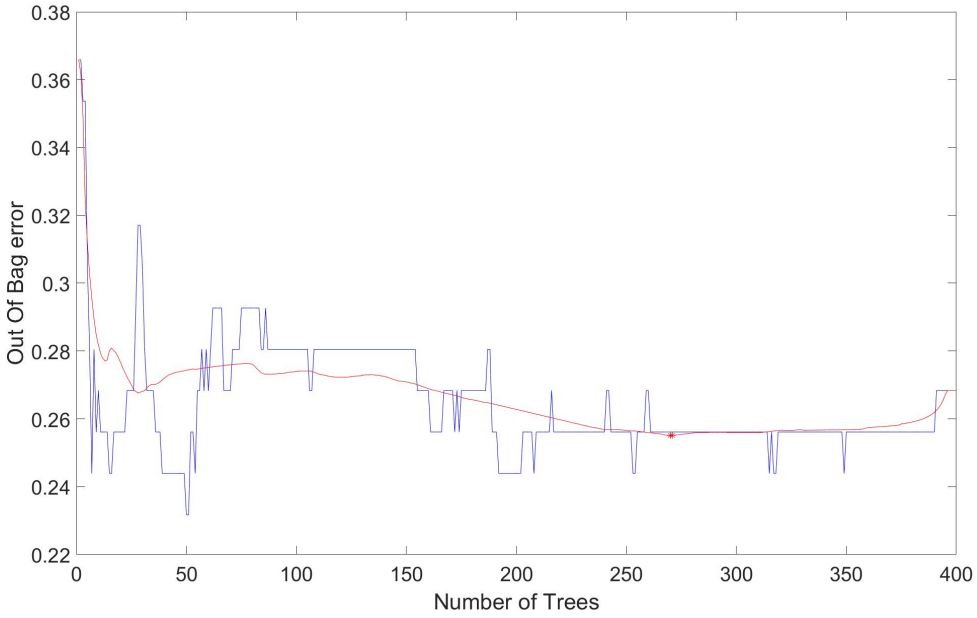


Figure B.1: OOB error as function of the number of trees.

Bibliography

[1] T. Fu, R. Jing, S. R. McFaull, and M. D. Cusimano, "Health and economic burden of traumatic brain injury in emergency department," *The Canadian Journal of Neurological Sciences Inc.*, vol. 43, pp. 238–247, 2016. doi: 10.1017/cjn.2015.320.

[2] A. C. Scholten, J. A. Haagsma, M. J. Panneman, E. F. V. Beeck, and S. Polinder, "Traumatic brain injury in the netherlands: Incidence, costs and disability-adjusted life years," *PLoS ONE*, vol. 9, 10 2014. doi: 10.1371/JOURNAL.PONE.0110905.

[3] S. Reynolds, "Predicting outcome after traumatic brain injury," *Journal of the Intensive Care Society*, vol. 12, pp. 196–200, 2011. doi: 10.1177/175114371101200306.

[4] S. Jain and L. M. Iverson, "Glasgow coma scale," 2022. In: StatPearls [Internet],Treasure Island (FL): StatPearls Publishing [Online.] Available: https://www.ncbi.nlm.nih.gov/books/NBK513298/ [Accessed Mar. 22, 2022].

[5] M. E. Haveman, M. J. V. Putten, H. W. Hom, C. J. Eertman-Meyer, A. Beishuizen, and M. C. Tjepkema-Cloostermans, "Predicting outcome in patients with moderate to severe traumatic brain injury using electroencephalography," *Critical Care*, vol. 23, 12 2019. doi: 10.1186/s13054-019-2656-6.

[6] A. I. Maas, D. K. Menon, E. W. Steyerberg, G. Citerio, F. Lecky, G. T. Manley, S. Hill, V. Legrand, A. Sorgner, N. Andelic, L. Andreassen, P. Andrews, G. Audibert, A. Audny, P. Azouv, P. Barzó, R. Beer, B. M. Bellander, A. Belli, H. Benali, M. Berardino, L. Beretta, A. Bražinová, H. Binder, F. Brehar, A. Buki, M. Bullinger, E. Cakmak, I. Callebaut, P. Cameron, G. C. Lozano, K. L. Carpenter, A. Chieregato, M. Coburn, J. P. Coles, J. Cooper, M. Cnossen, N. Curry, E. Czeiter, M. Czosnyka, C. Dahyot-Fitzelier, F. Damas, H. Dawes, V. D. Keyser, A. D. Luca, G. C. D. Ruiter, O. D. Witte, F. D. Corte, B. Demeter, B. Depreitere, D. W. Dippel, K. Dizdarevic, J. P. Dreier, G. Eapen, A. Ercole, P. Esser, M. Fabricius, L. Feremans, V. L. Feigin, F. Fossi, F. Forsyth, S. Florian, S. K. Frisvold, C. Frosini, A. Furmanov, J. Frantzén, D. Gadda, P. Gagliardo, D. Galanaud, G. Gao, A. Ghuysen, A. Godbolt, V. Gonšorová, Z. Grigore, R. Gruen, J. A. Haagsma, G. Hallaert, E. Hadzic, I. Haitsma, J. A. Hartings, R. Helbok, E. Helseth, S. Hoefer, M. Holling, M. Hunfeld, P. J. Hutchinson, R. Illéš, K. Janssens, T. J. Bovend'Eerdt, J. Y. Jiang, K. M. Jones, J. P. Kalala, F. Kalovits, G. Kasprian, A. Katila, N. Ketharanathan, A. G. Kolias,

F. Kolibay, D. Kondziella, L. O. Koskinen, A. Lagares, L. Lanyon, S. Laureys, R. Lefering, L. Levi, R. Lightfoot, H. F. Lingsma, D. Loeckx, L. N. Lohkamp, J. Lötjönen, C. Lumenta, M. Lyttle, M. Maegele, M. Majdan, A. Manara, H. Maréchal, C. Martino, L. Mascia, W. Mauritz, C. McMahon, T. Menovsky, P. Mitchell, N. Mladenov, C. Morganti-Kossmann, D. Nelson, E. Neugebauer, V. F. Newcombe, M. Oddo, M. Oresic, V. Orzalesi, J. G. Outtrim, A. Palotie, P. Parizel, J. F. Payen, V. Perlbarg, W. Peul, N. Pichon, A. Piippo, S. P. Floury, H. Ples, S. Polinder, A. Preiksaitis, M. Psota, P. Pullens, L. Puybasset, A. Ragauskas, R. Raj, M. Reiner, J. K. Rhodes, S. Richardson, S. Ripatti, S. Rocka, J. Roosenfeld, J. Rosand, C. Rosenlund, G. Rosenthal, R. Rossaint, S. Rossi, D. Rueckert, M. Rusnák, M. A. Rynkowski, J. Sahuquillo, O. Sakowitz, J. Sandor, S. Schmidt, H. Schoechl, R. Schou, T. Skandsen, M. Sonne, E. Schwendenwein, D. Smeets, P. Smieleweski, J. Söderberg, E. Stamatakis, S. Stanworth, R. Stefini, R. Stevens, W. Stewart, N. Stocchetti, W. Stummer, J. Szabó, A. Tascu, O. Tenovuo, A. Theadom, D. Tibboel, C. M. Tolias, A. Unterberg, P. Vajkoczy, A. Vargiolu, J. V. D. Naalt, T. V. Essen, W. V. Hecke, D. V. Praag, D. V. Roost, T. Vandenbulcke, L. V. Hauwe, M. V. D. Jagt, E. Vega, J. Verheyden, V. Verma, P. M. Vespa, A. Vik, R. Vilcinis, N. V. Steinbüchel, P. Vulekovic, K. K. Wang, E. Wildschut, G. Williams, M. Wilson, L. Wilson, S. Wolf, P. Ylén, M. Zaaroor, P. Zolfaghari, D. Martin, V. D'Orio, P. Damas, and J. F. L. Tshibanda, "Collaborative european neurotrauma effectiveness research in traumatic brain injury (center-tbi): A prospective longitudinal observational study," *Neurosurgery*, vol. 76, pp. 67–80, 2015. doi: 10.1227/NEU.0000000000000575.

[7] AANS, "American association of neurological surgeons," 2021. [Online.] Available: https://www.aans.org/ [Accessed Nov. 22, 2021].

[8] J. N. Ianof and R. Anghinah, "Lesão cerebral traumática: um ponto de vista eeg," *Dementia e Neuropsychologia*, vol. 11, pp. 3–5, 2017. doi: 10.1590/1980-57642016dn11-010002.

[9] D. K. Menon, K. Schwab, D. W. Wright, and A. I. Maas, "Position statement: Definition of traumatic brain injury," *Archives of Physical Medicine and Rehabilitation*, vol. 91, pp. 1637–1640, 2010. doi: 10.1016/j.apmr.2010.05.017.

[10] E. B. I. Consortium and S. Europa, "Traumatic brain injury fact sheets and policy brief can affect anyone, anywhere: Causing death or changing lives forever." [Online.] Available: www.center-tbi.eu [Accessed Dec. 18, 2021].

[11] J. T. van Dijck, M. D. Dijkman, R. H. Ophuis, G. C. de Ruiter, W. C. Peul, and S. Polinder, "In-hospital costs after severe traumatic brain injury a systematic review and quality assessment enhanced reader," *Plos One*, vol. 14, 2019. doi: 10.1371/journal.pone.0216743.

[12] C. Dias, J. Rocha, E. Pereira, and A. Cerejo, "Traumatic brain injury in portugal: Trends in hospital admissions from 2000 to 2010," *Acta Médica Portuguesa*, vol. 27, p. 349, 6 2014. doi: 10.20344/amp.4892.

[13] M. E. Santos and N. Agrela, "Traumatic brain injury in portugal: progress in incidence and mortality," *Brain Injury*, vol. 33, pp. 1552–1555, 10 2019. doi: 10.1080/02699052.2019.1658227.

[14] J. A. Langlois, W. Rutland-Brown, and M. M. Wald, "The epidemiology and impact of traumatic brain injury. a brief overview.," *Journal of Head and Trauma Rehabilitation*, vol. 21, pp. 375–378, 2006. doi: 10.1097/00001199-200609000-00001.

[15] I. Humphreys, R. L. Wood, C. L. Phillips, and S. Macey, "The costs of traumatic brain injury: a literature review," *ClinicoEconomics and Outcomes Research*, vol. 5, pp. 281–287, 2013. doi: 10.2147/CEOR.S44625.

[16] A. S. Alali, K. Burton, R. A. Fowler, D. M. Naimark, D. C. Scales, T. G. Mainprize, and A. B. Nathens, "Economic evaluations in the diagnosis and management of traumatic brain injury: A systematic review and analysis of quality," *Value in Health*, vol. 18, pp. 721–734, 2015. doi: 10.1016/j.val.2015.04.012.

[17] G. Teasdale, B. Jennett, P. Brennan, E. McElhinney, D. Allan, and F. Reith, "Recording the glasgow coma scale," 2022. [Online.] Available: https://www.glasgowcomascale.org/recording-gcs/ [Accessed Nov. 25, 2021].

[18] D. Teasdale and D. Jennett, "Glasgow coma scale score (gcs) calculator," 2022. [Online.] Available: https://www.mdcalc.com/glasgow-coma-scale-score-gcs [Accessed Jan. 03, 2022].

[19] R. N. RN, "Glasgow coma scale (gcs) nclex questions quiz nursing." [Online.] Available: https://www.registerednursern.com/glasgow-coma-scale-gcs-nclex-questions-quiz-nursing/ [Accessed Sep. 18, 2022].

[20] J. Wright, "The glasgow outcome scale. the center for outcome measurement in brain injury," 2000. [Online.] Available: http://www.tbims.org/combi/gos [Accessed Nov. 25, 2021].

[21] A. Sander, "Extended glasgow outcome scale," 2002. [Online.] Available: http://www.tbims.org/combi/gose [Accessed Nov. 25, 2021].

[22] N. S. E. M. Noor, H. Ibrahim, M. H. C. Lah, and J. M. Abdullah, "Improving outcome prediction for traumatic brain injury from imbalanced datasets using rusboosted trees on electroencephalography spectral power," *IEEE Access*, vol. 9, pp. 121608–121631, 2021. doi: 10.1109/ACCESS.2021.3109780.

[23] S. G. Palekar, M. Jaiswal, M. Patil, and V. Malpathak, "Outcome prediction in patients of traumatic brain injury based on midline shift on ct scan of brain," *Indian Journal of Neurosurgery*, vol. 10, pp. 210–215, 1 2021. doi: 10.1055/s-0040-1716990.

[24] S. N. Lueckel, A. H. Stephen, S. F. Monaghan, W. Binder, and C. A. Adams, "Predicting outcomes in acute traumatic brain injury (tbi)," *Rhode Island Medical Journal*, vol. 102, pp. 34–36, 7 2019. doi: 10.3171/2017.12.JNS172780.

[25] K. Rasheed, A. Qayyum, J. Qadir, S. Sivathamboo, P. Kwan, L. Kuhlmann, T. O'Brien, and A. Razi, "Machine learning for predicting epileptic seizures using eeg signals: A review," *IEEE Reviews in Biomedical Engineering*, vol. 14, pp. 139–155, 2021. doi: 10.1109/RBME.2020.3008792.

[26] M. Teplan, "Fundamental of eeg measurement," *MEASUREMENT SCIENCE REVIEW*, vol. 2, 01 2002.

[27] W. O. Tatum, A. M. Husain, S. R. Benbadis, and P. W. Kaplan, *Hand book of EEG INTERPRETATION*. Demos, 2008.

[28] G. A. Light, L. E. Williams, F. Minow, J. Sprock, A. Rissling, R. Sharp, N. R. Swerdlow, and D. L. Braff, "Electroencephalography (eeg) and event-related potentials (erps) with human participants," *Current Protocols in Neuroscience*, vol. 6, 7 2010. doi: 10.1002/0471142301.NS0625S52.

[29] M. Z. Koubeissi and N. J. A. Editors, *Epilepsy Board Review A Comprehensive Guide*. Springer, 2017. doi:10.1007/978-1-4939-6774-2.

[30] G. M. Rojas, C. Alvarez, C. E. Montoya, M. de la Iglesia-Vayá, J. E. Cisternas, and M. Gálvez, "Study of resting-state functional connectivity networks using eeg electrodes position as seed," *Frontiers in Neuroscience*, vol. 12, 4 2018. doi: 10.3389/fnins.2018.00235.

[31] J. N. Acharya and V. J. Acharya, "Overview of eeg montages and principles of localization," *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society*, vol. 36, pp. 325–329, 9 2019. doi: 10.1097/WNP.0000000000000538.

[32] qeegsupport, "What is qeeg/ brain mapping?." [Online.] Available: https://qeegsupport.com/what-is-qeeg-or-brain-mapping/ [Accessed Jan. 03, 2022].

[33] Y. Kubota, H. Nakamoto, S. Egawa, and T. Kawamata, "Continuous eeg monitoring in icu," *Journal of Intensive Care*, vol. 6, 7 2018. doi: 10.1186/s40560-018-0310-z.

[34] M. C. Tjepkema-Cloostermans, F. B. V. Meulen, G. Meinsma, and M. J. V. Putten, "A cerebral recovery index (cri) for early prognosis in patients after cardiac arrest," *Critical Care*, vol. 17, p. 1, 2013. doi: 10.1186/cc13078.

[35] M. L. Scheuer, "Continuous eeg monitoring in the intensive care unit," *Epilepsia*, vol. 43, pp. 114–127, 2002. doi: 10.1161/01.STR.0000144649.49861.1d.

[36] M. Müller, A. O. Rossetti, R. Zimmermann, V. Alvarez, S. Rüegg, M. Haenggi, W. J. Z'Graggen, K. Schindler, and F. Zubler, "Standardized visual eeg features predict outcome in patients with acute consciousness impairment of various etiologies," *Critical Care*, vol. 24, 12 2020. doi: 10.1186/s13054-020-03407-2.

[37] A. Caricato, I. Melchionda, and M. Antonelli, "Continuous electroencephalography monitoring in adults in the intensive care unit," *Critical Care*, vol. 22, 2018. doi: 10.1186/s13054-018-1997-x.

[38] P. L. Roux, "Physiological monitoring of the severe traumatic brain injury patient in the intensive care unit," *Current Neurology and Neuroscience Reports*, vol. 13, 2 2013. doi: 10.1007/s11910-012-0331-2.

[39] I. Stancin, M. Cifrek, and A. Jovic, "A review of eeg signal features and their application in driver drowsiness detection systems," *Sensors*, vol. 21, 2021. doi: 10.3390/s21113786.

[40] B. Albert, J. Zhang, A. Noyvirt, R. Setchi, H. Sjaaheim, S. Velikova, and F. Strisland, "Automatic eeg processing for the early diagnosis of traumatic brain injury," *Procedia Computer Science*, vol. 96, pp. 703–712, 2016. doi: 10.1016/j.procs.2016.08.253.

[41] J. Helm, A. M. Swiergosz, H. S. Haeberle, J. M. Karnuta, J. L. Schaffer, V. E. Krebs, A. I. Spitzer, and P. N. Ramkumar, "Machine learning and artificial intelligence: Definitions, applications, and future directions," *Current Reviews in Musculoskeletal Medicine*, vol. 13, pp. 69–76, 1 2020. doi: 10.1007/s12178-020-09600-8.

[42] P. P. Shinde and D. Shah, "A review of machine learning and deep learning applications," *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pp. 1–6, 2018. doi: 10.1109/ICCUBEA.2018.8697857.

[43] D. Ben-Israel, W. B. Jacobs, S. Casha, S. Lang, W. H. A. Ryu, M. de Lotbiniere-Bassett, and D. W. Cadotte, "The impact of machine learning on patient care: A systematic review," *Artificial Intelligence in Medicine*, vol. 103, 3 2020. doi: 10.1016/j.artmed.2019.101785.

[44] J. T. Senders, P. C. Staples, A. V. Karhade, M. M. Zaki, W. B. Gormley, M. L. Broekman, T. R. Smoth, and O. Arnaout, "Machine learning and neurosurgical outcome prediction systematic review," *World Neurosurgery*, vol. 109, pp. 476–486, 2018. doi: 10.1016/j.wneu.2017.09.149.

[45] S. Badillo, B. Banfai, F. Birzele, I. I. Davydov, L. Hutchinson, T. Kam-Thong, J. Siebourg-Polster, B. Steiert, and J. D. Zhang, "An introduction to machine learning," *Clinical Pharmacology and Therapeutics*, vol. 107, pp. 871–885, 4 2020. doi: 10.1002/CPT.1796.

[46] S. Raschka and V. Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, sikit-learn and TensorFlow 2*. Packt Publishing Ltd., 3rd ed., 2019.

[47] G. Roffo, "Feature selection library (matlab toolbox)," 2016. [Online.] Available: http://arxiv.org/abs/1607.01327 [Accessed May. 01, 2022].

[48] X. Ying, "An overview of overfitting and its solutions," *Journal of Physics: Conference Series*, vol. 1168, 3 2019. doi: 10.1088/1742-6596/1168/2/022022.

[49] A. Singh, "A review of supervised machine learning algorithms; a review of supervised machine learning algorithms," *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 1310–1315, 2016.

[50] C. Debaditya, "Confusion matrix," 2021. [Online] Available: https://www.debadityachakravorty.com/ai-ml/cmatrix/[Accessed Mar. 29, 2022].

[51] H. Dalianis, *Clinical text mining: Secondary use of electronic patient records*. Springer International Publishing, 5 2018. doi: 10.1007/978-3-319-78503-5.

[52] R. D. Sipio, "Towards data science," 5 2021. [Online.] Available: https://towardsdatascience.com/a-quick-guide-to-auc-roc-in-machine-learning-models-f0aedb78fbad[Accessed Sep. 17, 2022].

[53] A. Sivaganesan, G. T. Manley, and M. C. Huang, "Informatics for neurocritical care: Challenges and opportunities," *Neurocritical Care*, vol. 20, pp. 132–141, 2014. doi: 10.1007/s12028-013-9872-8.

[54] D.-Y. Cho and Y.-C. Wang, "Comparison of the apache iii, apache ii and glasgow coma scale in acute head injury for prediction of mortality and functional outcome," *Intensive Care Med*, vol. 23, pp. 77–84, 1997.

[55] W. A. Knaus, E. A. Draper, D. P. Wagner, and J. E. Zimmerman, "Apache ii: A severity of disease classification system," *Critical Care Medicine*, vol. 13, pp. 818–829, 1985. doi: 10.1097/00003246-198510000-00009.

[56] A. I. Maas, E. Steyerberg, G. Murray, and A. Maas, "International mission for prognosis and analysis of clinical trials in tbi." [Online.] Available: http://www.tbi-impact.org [Accessed Jan. 19, 2022].

[57] M. C. T. Collaborators, "Predicting outcome after traumatic brain injury: pratical prognostic models based on large cohort of international patients," *BMJ Research*, pp. 336 – 425, 2008. doi: 10.1136/bmj.39461.643438.25.

[58] R. Raj, J. Siironen, R. Kivisaari, J. Hernesniemi, and M. B. Skrifvars, "Predicting outcome after traumatic brain injury: Development of prognostic scores based on the impact and the apache ii," *Journal of Neurotrauma*, vol. 31, pp. 1721–1732, 10 2014. doi: 10.1089/neu.2014.3361.

[59] E. W. Steyerberg, N. Mushkudiani, P. Perel, I. Butcher, J. Lu, G. S. McHugh, G. D. Murray, A. Marmarou, I. Roberts, J. D. F. Habbema, and A. I. Maas, "Predicting outcome after traumatic brain injury: Development and international validation of prognostic scores based on admission characteristics," *PLoS Medicine*, vol. 5, pp. 1251–1261, 8 2008. doi: 10.1371/journal.pmed.0050165.

[60] B. Jacobs, T. Beems, T. M. V. D. Vliet, A. B. V. Vugt, C. Hoedemaekers, J. Horn, G. Franschman, I. Haitsma, J. V. D. Naalt, T. M. Andriessen, G. F. Borm, and P. E. Vos, "Outcome prediction in moderate and severe traumatic brain injury: A focus on computed tomography variables," *Neurocritical Care*, vol. 19, pp. 79–89, 8 2013. doi: 10.1007/s12028-012-9795-9.

[61] M. Majdan, H. F. Lingsma, D. Nieboer, W. Mauritz, M. Rusnak, and E. W. Steyerberg, "Performance of impact, crash and nijmegen models in predicting six month outcome of patients with severe or moderate tbi: An external validation study," *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 22, 2014. doi: 10.1186/s13049-014-0068-9.

[62] R. L. Amorim, L. M. Oliveira, L. M. Malbouisson, M. M. Nagumo, M. Simoes, L. Miranda, E. Bor-Seng-Shu, A. Beer-Furlan, A. F. D. Andrade, A. M. Rubiano, M. J. Teixeira, A. G. Kolias, and W. S. Paiva, "Prediction of early tbi mortality using a machine learning approach in a lmic population," *Frontiers in Neurology*, vol. 10, 1 2020. doi: 10.3389/fneur.2019.01366.

[63] A. T. Hale, D. P. Stonko, A. Brown, J. Lim, D. J. Voce, S. R. Gannon, T. M. Le, and C. N. Shannon, "Machine learning analysis outperforms conventional statistical models and ct classification systems in predicting 6-month outcomes in pediatric patients sustaining traumatic brain injury," *Neurosurgical Focus*, vol. 45, 11 2018. doi: 10.3171/2018.8.FOCUS17773.

[64] S. L. Chong, N. Liu, S. Barbier, and M. E. H. Ong, "Predictive modeling in pediatric traumatic brain injury using machine learning," *BMC Medical Research Methodology*, vol. 15, 2015. doi:10.1186/s12874-015-0015-0.

[65] P. Perel, P. Edwards, R. Wentz, and I. Roberts, "Systematic review of prognostic models in traumatic brain injury," *BMC Medical Informatics and Decision Making*, vol. 6, 11 2006. doi: 10.1186/1472-6947-6-38.

[66] C. S. Rau, P. J. Kuo, P. C. Chien, C. Y. Huang, H. Y. Hsieh, and C. H. Hsieh, "Mortality prediction in patients with isolated moderate and severe traumatic brain injury using machine learning models," *PLoS ONE*, vol. 13, 11 2018. doi: 10.1371/journal.pone.0207192.

[67] J. zhou Feng, Y. Wang, J. Peng, M. wei Sun, J. Zeng, and H. Jiang, "Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries," *Journal of Critical Care*, vol. 54, pp. 110–116, 12 2019. doi: 10.1016/j.jcrc.2019.08.010.

[68] K. Matsuo, H. Aihara, T. Nakai, A. Morishita, Y. Tohma, and E. Kohmura, "Machine learning to predict in-hospital morbidity and mortality after traumatic brain injury," *Journal of Neurotrauma*, vol. 37, pp. 202–210, 1 2020. doi:10.1089/neu.2018.6276.

[69] S. Rizoli, A. Petersen, E. Bulger, R. Coimbra, J. D. Kerby, J. Minei, L. Morrison, A. Nathens, M. Schreiber, and A. L. de Oliveira Manoel, "Early prediction of outcome after severe traumatic brain injury: A simple and practical model," *BMC Emergency Medicine*, vol. 16, 8 2016. doi: 10.1186/s12873-016-0098-x.

[70] A. Tolonen, M. O. K. Särkelä, R. S. K. Takala, A. Katila, J. Frantzén, J. P. Posti, M. Müller, M. van Gils, and O. Tenovuo, "Quantitative eeg parameters for prediction of outcome in severe traumatic brain injury: Development study," *Clinical EEG and Neuroscience*, vol. 49, pp. 248–257, 7 2018. doi: 10.1177/1550059417742232.

[71] A. Mikola, I. Ratsep, M. Sarkela, and T. Lipping, "Prediction of outcome in traumatic brain injury patients using long-term qeeg features," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2015-November, pp. 1532–1535, 11 2015. doi: 10.1109/EMBC.2015.7318663.

[72] N. S. E. M. Noor and H. Ibrahim, "Machine learning algorithms and quantitative electroencephalography predictors for outcome prediction in traumatic brain injury: A systematic review," *IEEE Access*, vol. 8, pp. 102075–102092, 2020. doi: 10.1109/ACCESS.2020.2998934.

[73] M. Feng, Z. Zhang, C. Guan, D. R. Hardoon, N. K. K. King, B. C. Pang, and B. T. Ang, "Utilization of temporal information for intracranial pressure development trend forecasting in traumatic brain injury," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 3930–3934, 2012. doi: 10.1109/EMBC.2012.6346826.

[74] R. Raj, T. Luostarinen, E. Pursiainen, J. P. Posti, R. S. Takala, S. Bendel, T. Konttila, and M. Korja, "Machine learning-based dynamic mortality prediction after traumatic brain injury," *Scientific Reports*, vol. 9, 12 2019. doi: 10.1038/s41598-019-53889-6.

[75] H. Lee, M. A. Mizrahi, J. A. Hartings, S. Sharma, L. Pahren, L. B. Ngwenya, B. D. Moseley, M. Privitera, F. C. Tortella, and B. Foreman, "Continuous electroencephalography after moderate to severe traumatic brain injury," *Critical Care Medicine*, vol. 47, pp. 574–582, 2019. doi: 10.1097/CCM.0000000000003639.

[76] C. Melinosky, S. Yang, P. Hu, H. C. Li, C. H. Miller, I. Khan, C. Mackenzie, W. T. Chang, G. Parikh, D. Stein, and N. Badjatia, "Continuous vital sign analysis to predict secondary neurological decline after traumatic brain injury," *Frontiers in Neurology*, vol. 9, 9 2018. doi: 10.3389/fneur.2018.00761.

[77] B. Y. Gravesteijn, D. Nieboer, A. Ercole, H. F. Lingsma, D. Nelson, B. van Calster, E. W. Steyerberg, and C.-T. collaborators, "Machine learning algorithms performed no better than regression models for prognostication in traumatic brain injury," *Journal of Clinical Epidemiology*, vol. 122, pp. 95–107, 2020. doi: 10.1016/j.jclinepi.2020.03.005.

[78] K. Kazazian, L. Norton, G. Laforge, A. Abdalmalak, T. E. Gofton, D. Debicki, M. Slessarev, S. Hollywood, K. S. Lawrence, and A. M. Owen, "Improving diagnosis and prognosis in acute severe brain injury: A multimodal imaging protocol," *Frontiers in Neurology*, vol. 12, 12 2021. doi: 10.3389/fneur.2021.757219.

[79] M. Tjepkema-Cloostermans, H. Hom, A. Beishuizen, and R. Geelkerken, "Multimodal brain monitoring in traumatic brain injury: A prospective observational cohort study," *Medisch Spectrum Twente, Enschede*, 2017.

[80] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE TRANSACTIONS ON NEURAL NETWORKS*, vol. 10, 1999. doi: 10.1109/72.761722.

[81] E. Oja, J. Karhunen, Z. Yang, A. Ilin, J. Luttinen, T. Hao, R. Vigario, and J. Ylipaavalniemi, "The fastica software package," 2011. [Online.] Available: http://research.ics.aalto.fi/ica/members.shtml [Accessed Apr. 03, 2022].

[82] T. Donoghue, M. Haller, E. J. Peterson, P. Varma, P. Sebastian, R. Gao, T. Noto, A. H. Lara, J. D. Wallis, R. T. Knight, A. Shestyuk, and B. Voytek, "Parameterizing neural power spectra into periodic and aperiodic components," *Nature Neuroscience*, vol. 23, pp. 1655–1665, 12 2020. doi: 10.1038/s41593-020-00744-x.

[83] MathWorks, "mscohere magnitude squared coherence." [Online.] Available: https://www.mathworks.com/help/signal/ref/mscohere.html [Accessed Sep. 18, 2022].

[84] R. V. Sheorajpanday, G. Nagels, A. J. Weeren, M. J. van Putten, and P. P. D. Deyn, "Reproducibility and clinical relevance of quantitative eeg parameters in cerebral ischemia: A basic approach," *Clinical Neurophysiology*, vol. 120, pp. 845–855, 5 2009. doi: 10.1016/j.clinph.2009.02.171.

[85] M. J. van Putten, "The revised brain symmetry index," *Clinical Neurophysiology*, vol. 118, pp. 2362–2367, 11 2007. doi: 10.1016/j.clinph.2007.07.019.

[86] M. J. A. M. V. Putten, . Dénes, and L. J. Tavy, "Continuous quantitative eeg monitoring in hemispheric stroke patients using the brain symmetry index," *Stroke*, vol. 35, pp. 2489–2492, 2004. doi: 10.1161/01.STR.0000144649.49861.1d.

[87] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–656, 1948.

[88] I. Djordjevic, "Quantum circuits and quantum information processing fundamentals," *Quantum Information Processing and Quantum Error Correction*, pp. 91–117, 2012. doi: 10.1016/B978-0-12-385491-9.00003-4.

[89] VoytekLab, "Fooof fitting oscillations and one over f fooof 1.0.0 documentation," 2018. [Online.] Available: https://fooof-tools.github.io/fooof/index.html [Accessed Apr. 01, 2022].

[90] T. Donoghue, "Foof- fooof tools," 2020. [Online.] Available: https://github.com/fooof-tools/fooof [Accessed Apr. 17, 2022].

[91] M. van Putten, "The colorful brain: Visualization of eeg background patterns," *Journal of Clinical Neurophysiology*, vol. 25, 4 2008. doi: 10.1097/WNP.0b013e31816bdf85.

[92] R. Hardstone, S. S. Poil, G. Schiavone, R. Jansen, V. V. Nikulin, H. D. Mansvelder, and K. Linkenkaer-Hansen, "Detrended fluctuation analysis: A scale-free view on neuronal oscillations," *Frontiers in Physiology*, vol. 3 NOV, 2012. doi: 10.3389/fphys.2012.00450.

[93] R. Ton and A. Daffertshofer, "Model selection for identifying power-law scaling," *NeuroImage*, vol. 136, pp. 215–226, 8 2016. doi: 10.1016/j.neuroimage.2016.01.008.

[94] A. Daffertshofer, R. Ton, M. L. Kringelbach, M. Woolrich, and G. Deco, "Distinct criticality of phase and amplitude dynamics in the resting brain," *NeuroImage*, vol. 180, pp. 442–447, 10 2018. doi: 10.1016/j.neuroimage.2018.03.002.

[95] A. Daffertshofer and R. Ton, "Model selection for identifying power-law scaling," *NeuroImage*, vol. 136, pp. 215–226, 2016. doi:10.1016/j.neuroimage.2016.01.008.

[96] Ton and Daffertshofer, "Fluctuation analysis," 2016. [Online.] Available: https://github.com/fooof-tools/fooof [Accessed Apr. 11, 2022].

[97] C. W. Lynn, E. J. Cornblath, L. Papadopoulos, M. A. Bertolero, and D. S. Bassett, "Broken detailed balance and entropy production in the human brain," *PNAS*, 2021. doi: 10.1073/pnas.2109889118/-/DCSupplemental.

[98] C. W. Lynn, E. J. Cornblath, L. Papadopoulos, M. A. Bertolero, and D. S, "Broken detailed balance and entropy production in the human brain," 2021. [Online.] Available: https://github.com/ChrisWLynn/Broken_detailed_balance [Accessed Apr. 08, 2022].

[99] H. Liu and H. Motoda, *Computational methods of feature selection*. Chapman and Hall/CRC, 2008.

[100] E. R. Sruthi, "Understanding random forest," 2021. [Online.] Available: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest [Accessed Apr. 06, 2022].

[101] O. Mbaabu, "Introduction to random forest in machine learning," 2020. [Online.] Available: https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/ [Accessed Apr. 06, 2022].

[102] N. Donges, "Random forest algorithm: A complete guide," 2021. [Online.] Available: https://builtin.com/data-science/random-forest-algorithm [Accessed Apr. 06, 2022].

[103] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5–32, 2001. doi: 10.1023/A:1010933404324.

[104] E. Zvornicanin, "Out-of-bag error in random forests," 10 2021. [Online.] Available: https://www.baeldung.com/cs/random-forests-out-of-bag-error [Accessed Jun. 16, 2022].

[105] MathWorks, "Statistics and machine learning toolbox." [Online.] Available: https://nl.mathworks.com/help/stats/tune-random-forest-using- [Accessed Jun. 14, 2022].

[106] R. Kurnia, "Bayesian optimization: bayes opt or hyperopt," 10 2021. [Online.] Available: https://www.analyticsvidhya.com/blog/2021/05/bayesian\protect\discretionary{\char\hyphenchar\font}{}{}optimization\protect\discretionary{\char\hyphenchar\font}{}{}bayes_opt\protect\discretionary{\char\hyphenchar\font}{}{}or\protect\discretionary{\char\hyphenchar\font}{}{}hyperopt/ [Accessed Jun. 16, 2022].

[107] MathWorks, "perfcurve receiver operating characteristic (roc) curve or other performance curve for classifier output." [Online.] Available: https://www.mathworks.com/help/stats/perfcurve.html [Accessed Sep. 18, 2022].

[108] S. Glen, "Mcnemar test definition, examples, calculation from statisticshowto.com: Elementary statistics for the rest of us!." [Online.] Available: https://www.statisticshowto.com/mcnemar-test/ [Accessed Jun. 10, 2022].

[109] A. C. Knoll, J. Jin, M. Rashid, P. P. A. M. A, A. A. Nasir, N. Sulaiman, A. P. P. A. Majeed, R. M. Musa, A. F. A. Nasir, B. S. Bari, and S. Khatun, "Current status, challenges,

and possible solutions of eeg-based brain-computer interface: A comprehensive review," *Article 25 Comprehensive Review. Front. Neurorobot*, vol. 1, p. 25, 2020. doi: 10.3389/fnbot.2020.00025.

[110] S. Lenell, L. Nyholm, A. Lewén, and P. Enblad, "Clinical outcome and prognostic factors in elderly traumatic brain injury patients receiving neurointensive care," *Acta Neurochirurgica*, vol. 161, pp. 1243–1254, 6 2019. doi: 10.1007/s00701-019-03893-6.

[111] H. Y. Lu, T. C. Li, Y. K. Tu, J. C. Tsai, H. S. Lai, and L. T. Kuo, "Predicting long-term outcome after traumatic brain injury using repeated measurements of glasgow coma scale and data mining methods," *Journal of Medical Systems*, vol. 39, 1 2015. doi: 10.1007/s10916-014-0187-x.

[112] E. P. Thelin, E. Jeppsson, A. Frostell, M. Svensson, S. Mondello, B. M. Bellander, and D. W. Nelson, "Utility of neuron-specific enolase in traumatic brain injury; relations to s100b levels, outcome, and extracranial injury severity," *Critical Care*, vol. 20, 9 2016. doi: 10.1186/s13054-016-1450-y.

[113] L. Yuan, X. Wei, Y. Jin, G. Wang, H. Tian, and S. Chen, "Use of multisequence 3.0-t mri to detect severe traumatic brain injury and predict the outcome," *J Radiol*, vol. 88, 2015. doi: 10.1259/bjr.20150129.

[114] K. Kalpakis, S. Yang, P. F. Hu, C. F. Mackenzie, L. G. Stansbury, D. M. Stein, and T. M. Scalea, "Permutation entropy analysis of vital signs data for outcome prediction of patients with severe traumatic brain injury," *Computers in Biology and Medicine*, vol. 56, pp. 167–174, 1 2015. doi: 10.1016/j.compbiomed.2014.11.007.

[115] R. W. Thatcher, C. Biver, R. Mcalaster, and A. Salazar, "Biophysical linkage between mri and eeg coherence in closed head injury," 1998.

[116] S. Geng, W. Zhou, Q. Yuan, D. Cai, and Y. Zeng, "Eeg non-linear feature extraction using correlation dimension and hurst exponent," *Neurological Research*, vol. 33, pp. 908–912, 11 2011. doi: 10.1179/1743132811Y.0000000041.

[117] R. Burr, C. Kirkness, and P. Mitchell, "Detrended fluctuation analysis of intracranial pressure predicts outcome following traumatic brain injury," *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 2509–2518, 11 2008. doi: 10.1109/TBME.2008.2001286.

[118] A. Maas, "Traumatic brain injury: Changing concepts and approaches," *Elsevier Enhanced Reader*, vol. 19, pp. 3–6, 2016. doi: 10.1016/j.cjtee.2016.01.001.

[119] J. Ranson, B. E. Magnus, N. Temkin, S. Dikmen, J. T. Giacino, D. O. Okonkwo, A. B. Valadka, G. T. Manley, and L. D. Nelson, "Diagnosing the gose: Structural and psychometric properties using item response theory, a track-tbi pilot study," *Journal of Neurotrauma*, vol. 36, pp. 2493–2505, 2018. doi: 10.1089/neu.2018.5998.

[120] R. D. Riley, J. Ensor, K. I. E. Snell, T. P. A. Debray, D. G. Altman, K. G. M. Moons, and G. S. Collins, "External validation of clinical prediction models using big datasets from e-health records or ipd meta-analysis: opportunities and challenges," *BMJ*, vol. 353, 2016. doi: 10.1136/bmj.i3140.

[121] Y. LeCun, "Rebellion research: Ai asset management," 4 2022. [Online.] Available: https://www.rebellionresearch.com/ [Accessed Mar. 06, 2022].