Faculdade de Engenharia da Universidade do Porto

Machine learning in the diagnosis of disorders of consciousness using SMART

Manuel Curral

Mestrado Integrado em Engenharia Informática e Computação

Supervisor: Professor Doutor Jaime Cardoso

Co-Supervisor: Professora Liliana Teixeira

Co-Supervisor: Professor Doutor Nuno Rocha

October 31, 2021

© Manuel José Pereira Curral, 2021

Machine learning in the diagnosis of disorders of consciousness using SMART

Manuel Curral

Mestrado Integrado em Engenharia Informática e Computação

October 31, 2021

Abstract

People with disorders of consciousness (DOC) are vulnerable to misdiagnosis which can negatively affect their rehabilitation process. The incorrect diagnosis of people with DOC is common, reaching 43 %. This has possible implications for decisions regarding the provision of health care to this group. Diagnosis of this population depends on the assessment of their behavioral responses to stimuli. The intentionality and types of behavior exhibited by people in a vegetative state (VS) and a minimally conscious state (MCS) can be difficult to distinguish and subtle signs of consciousness can go unnoticed. It is widely recognized that the use of standardized and sensitive behavioral assessment scales, such as the Sensory Modality Assessment and Rehabilitation Technique (SMART), can help healthcare professionals to identify subtle signs of awareness. SMART is an assessment tool that combines communication, motor and sensory modalities to diagnose the condition of patients who have suffered severe brain injuries. This method is quite credible and accepted by the healthcare community that deals with this clinical population. It requires and consumes many resources in making the diagnosis.

However, less experienced SMART assessors can be misled by some types of patient responses and even in the analysis of session data, namely in different diagnostic limit zones. Hence a second opinion based on Machine Learning can prove to be very useful. In addition, diagnosis is made session by session and the cumulative diagnostic certainty as the sessions progress. This tool can be detected as well as the expert assessors, in the future it can be very useful to detect in advance (in a smaller number of sessions) the state of consciousness of the patient to be analyzed.

SMART evaluation has already been explored with statistical software such as statistical package for the social sciences (SPSS) combining analysis methods and techniques such as analysis of variance (ANOVA), etc. So far, no machine learning methods have been found in partnership with this technique and diagnostic tools (SMART).

The best diagnosis, through a second opinion performed by the machine, is expected to increase the confidence level in decision-making by SMART assessors. More protected and less subject to criticism of negligence, data that possible errors if detected, can be bridged and safeguarded or at least become noticeable therefore, it leads to higher hit rates. Minimizing the allocation of time to human resources for this specific task, can be beneficial for these professionals due to the useful / effective time to perform tasks (elimination of extra hours to do a task that was previously done). Institutions: speeds up the prognosis and diagnosis process, making it financially convenient to make professionals more available, resulting in greater performance and efficiency, with the possibility of performing other essential tasks.

Keywords: Machine Learning, disorders of consciousness, diagnosis, SMART, minimally conscious state, vegetative state, brain injury

Resumo

Pessoas com perturbações de consciência (PdC) são vulneráveis a diagnósticos errados que podem afetar negativamente o seu processo de reabilitação. O diagnóstico incorreto de pessoas com PdC é comum, podendo atingir os 43%. Isto acarreta possíveis implicações nas decisões relacionadas com a prestação de cuidados de saúde desta população. o diagnóstico desta população depende da avaliação das suas respostas comportamentais a estimulação. A intencionalidade e os tipos de comportamentos exibidos por pessoas em estado vegetativo e estado de consciência mínima, podem ser difíceis de distinguir e sinais subtis de consciência podem passar despercebidos. É amplamente reconhecido que o uso de escalas de avaliação comportamental padronizadas e sensíveis, tal como o Sensory Modality Assessment and Rehabilitation Technique (SMART), pode ajudar os profissionais de saúde a identificar sinais subtis de consciência. O SMART é um instrumento de avaliação que combina funções comunicacionais, motoras e de aferição de sentidos para diagnosticar o estado de pacientes que sofreram lesões cerebrais graves. Este método é bastante credível e aceite pela comunidade da área da saúde que lida com esta população clínica. Ele requer e consome muitos recursos na elaboração do diagnóstico.

Contudo, avaliadores SMART menos experientes podem ser induzidos em erro por algum tipos de respostas dos pacientes e mesmo na análise dos dados das sessões nomeadamente em zonas limite de diagnósticos distintos. Daí uma segunda opinião com base em Machine Learning pode vir a ser muito útil. Além disso, o diagnóstico é feito sessão a sessão sendo a certeza de diagnóstico cumulativa no progredir das sessões.

Esta ferramenta se detetar tão bem como os avaliadores expert, pode no futuro ser muito útil a detetar antecipadamente (num menor número de sessões) o estado de consciência do paciente analisado. - A avaliação SMART já foi explorada através de software estatístico como SPSS usando métodos e técnicas de análise ANOVA, etc. Até ao momento não foram encontrados métodos de machine learning em parceria desta técnica e ferramenta de diagnóstico (SMART).

Melhor diagnóstico, através de uma segunda opinião realizada pela máquina é expectável que aumente o índice de confiança na sua tomada de decisão por parte dos avaliadores SMART . Mais resguardados e menos sujeitos a críticas de negligência, dados que possíveis erros se detectados, possam ser colmatados e salvaguardados ou pelos menos se tornem perceptíveis portanto, conduz a taxas de acerto mais elevadas. Redução dos tempos de alocação a recursos humanos destinados a esta tarefa específica, pode ser benéfico para estes profissionais pelo tempo útil de realização de tarefas (eliminação das horas a mais a fazer uma tarefa já realizada outrora). Instituições: agiliza o processo de prognóstico e diagnóstico, financeiramente conveniente tornar profissionais mais libertos consequente maior rendimento e eficácia com possibilidade de realização de outras tarefas primordiais.

Keywords: Aprendizado Máquina, Perturbações de consciência PdC,diagnóstico de consciência, SMART, estado de consciência mínimo, estado vegetativo, lesão cerebral

Acknowledgements

Hello world!

I want to thank FEUP for the fantastic conditions of labour, the people who work here and people (students like me) by promoting activities that enable the development of areas of our interest. Dedicated to the people who want to learn and people with a desire to teach in general. It moves the world [1].

Above all, I would like to thank the fantastic synergies of the Group Coordinator (VCMI) and supervisor Prof. Jaime Cardoso as well the Institutional Coordinator of the Superior School of Health of Porto in name of Prof. Nuno Rocha and Prof. Liliana Teixeira of Polytechnic of Leria as the promoters of this interesting idea.

I want to thank humanity or friendship, the exchange and debate of ideas that exist in the world, which somehow inspires us, gives us encouragement in the most difficult moments and decisions and, on the other hand, gives greater meaning to life, shares happiness and other important feelings with us and vice versa. The power that positive influence can have in one another's lives. The intrinsic willpower that makes us think, research, understand, solve problems and also make us ask for help and listen. My thanks to family, affective, academic/professional, community, financial, sports and health support in general, without you it was not the same thing.

Manuel Curral

“You can’t stop the waves, but you can learn to surf. ”

Jon Kabat-Zinn

“O Universo criou um cérebro para permitir ver-se a si mesmo, para estar consciente de si mesmo. ”

Henry Markram

Contents

List of Figures

List of Tables

Abbreviations

ANOVA	Analysis of Variance
DOC	Disorders of Consciousness
DRS	Disability Rating Scale
GCS	Glasgow Coma Scale
ICC	Intra Class Correlation
fMRI	functional Magnetic Resonance Imaging
MCS-	Minimally Conscious State Minus
MCS+	Minimally Conscious State Plus
MMN	Multifocal Motor Neuropathy
NN	Neural Network
SMART	Sensory Modality Assessment and Rehabilitation Technique
WNSSP	Western Neuro Sensory Stimulation Profile
TBI	Traumatic brain injury
VS	Vegetative State
WWW	*World Wide Web*

Chapter 1

Introduction

Context and Motivation

Consciousness is the individual's ability to be aware of the knowledge of self and the environment. Furthermore, it is the ability to respond to various voluntary internal and external stimuli [2]. In basic neurological terms, it is composed of awareness and wakefulness [3]. The different states of consciousness are represented in the table 1.1 .

.

Condition	**Wakefulness**	**Awareness**
Coma	-	-
Vegetative State	+ to ++	-
Minimally Conscious State	+ to ++	+
Emerged from Minimally Conscious State	++	++

Table 1.1: Disorders of consciousness categorization

Brain Injury (BI) is a head injury that damages the brain and its complex connections [4].

This causes problems with how a person can think and interact with the world around them. Following a BI, there are specific cognitive skills that are no longer functioning in the same capacity [5]. Perception, observation, and recognition of information are deeply affected.

There are events that damage areas of the brain that control parts of the human body. And the patient's faculties are conditioned. The origin of BI can be:

- acute: as in a virus or haemorrhage
- traumatic: for example road traffic accidents, impacts where there are head injuries
- non-traumatic: such as drowning, organ's infarction, drug overdose, etc.

All of the above cause brain damage that leads to consequences in terms of disturbing the person's consciousness [6].

After the coma, the rates of diagnostic errors, namely in the distinction between vegetative state (VS) and the minimally conscious state (MCS), are high $\approx 40\%$ [7] [8] [9].

The diagnosis of the patient who suffers brain damage is made using scales assessing the behavioural response to stimuli: CRS-r, SMART, WHIM, WNSSP, Rancho levels physicians, etc. [10]. But also with technological advancement such as neuronal imaging technologies such including: functional magnetic resonance imaging (fMRI), electroencephalogram (EEG) positron emission tomography (PET) [11] [12] [13].

It has been possible to leverage the expertise in the area of neuroscience and the identification, consequent diagnosis, of the cerebral behaviour of patients [5]. This helps to identify patients with DOC better because it demonstrates the level of responsiveness that could not be obtained by the methods previously described 1.

VS is defined as the absence of self-awareness and the environment 1.1. Behaviours are limited to reflexive responses indicating no purposeful movement, neither experience of suffering or evidence of comprehension [14].

MCS is serious but does not represent a complete lack of awareness resulting from widespread damage to the telencephalon (the part of the brain that controls thinking and behaviour). The intentionally and types of behaviour exhibited by people in a VS and a MCS, can be challenging to distinguish, and subtle signs of consciousness can go unnoticed. It is widely recognized that the use of standardized and sensitive behavioural assessment scales, such as the Sensory Modality Assessment and Rehabilitation Technique (SMART), can help healthcare professionals identify subtle signs of awareness.

With the data gathered from various patient assessments, we think it is possible to train an artificial intelligence model to recognize hidden patterns or minute details resulting from the multiple data that were fed to the model. The new data will generalise the upcoming data and produce a diagnosis enough to be of help to SMART assessors. [15].

1.1 SMART

SMART is an assessment tool that combines communication, motor and sensory modalities to diagnose patients who have suffered severe brain injuries [16].

Five levels: no response, reflex response, withdrawal response, localized response and differentiated response.

The main advantage of the method is entirely credible and accepted by the healthcare community that deals with this clinical population.

The disadvantage is that it requires and consumes many resources in making the diagnosis.

Research continues on clinical tools such as (fMRI) with improved diagnostic certainty and prognostic applications [5]. There are three main factors that influence the prognosis of patients in VS and MCS:

- Time (the longer in the state, the more complicated functional recovery becomes)
- Age (young people have a higher recovery rate, linked to physiological recovery processes and brain plasticity)

- Type (if non-traumatic, there is a shorter potential recovery window)
- Note: The more severe the degree of injury, the rarer the recovery.

Note: The diagnostic evaluation is done over a period of 3 weeks over 10 sessions, distributed in equal numbers in the morning and afternoon.[10]

1.2 Objectives

The goal of this dissertation is to give greater certainty in the diagnosis, or for instance reduce levels of misdiagnosis of DOC. The following steps materialize the aim:

1. Classification in 2 possible stages MCS and VS
2. Reduce time procedures diagnosis: after the first complete goal check, reduce the number of sessions to less than 10
3. Try to find correlations between origin and possible stages and gauge the accuracy of results with basis on number of sessions available

1.3 Structure of the document

In total, the dissertation has five chapters.

Chapter 1 contextualizes DOC and SMART methodologies and the advantages of combining them both with artificial intelligence 1.

The chapter 2 contains state of the art, description of the most common diagnostic tools and their impact on the development of medical diagnosis 2.

Chapter 3 contains machine learning theory, a guide on how to create a model and how to validate the results 3.

Chapter 4 talks about experimental studies related to SMART and the database 4.

Chapter 5 is the conclusions that describe what was done, the results obtained and what there is more to explore 5.

Areas:
CCS →Computing methodologies →Machine learning
CCS →Applied computing →Life and medical sciences

Chapter 2

Diagnostic tools using Machine Learning

2.1 Diagnostic tools

Neuroscientific technologies use experimental methods to study the patients' brain processes. Those methods include neuroimaging, psychophysical techniques or psychological tests which are used to study processes such as learning, attention, memory, or emotion [17].

The following methods stand out:

- Electroencephalogram (EEG)

  Electroencephalography is a simple, non-invasive technique based on recording and evaluating brain activity using electrodes placed on the skull surface [18] [12].The (EEG) is the record that results from the measurement of the electrical potentials of the brain [5] [19]. EEG shows the electrical fluctuation in the different locations of the cortex [20]. However, it has the disadvantage of having an insufficient resolution to register the neuronal activity in deeper brain structures, such as the *nucleus accumbens* related to the processing of emotions[21].

  Study using Mismatch Negativity (MMN) tomography technique resulting in classification of patients with DOC in figure 2.1

- (Functional) Magnetic Resonance

  It is a method of diagnosis and fundamental research in market research analysing emotions planning surgery brain mapping neurosciences [23]. It allows the analysis of the subjects' response to different activities or stimuli. Magnetic resonance imaging is a non-invasive technique used to obtain information about the subjects' response to different stimuli (very common in research and analysis of neurological diseases such as Alzheimer's Disease and

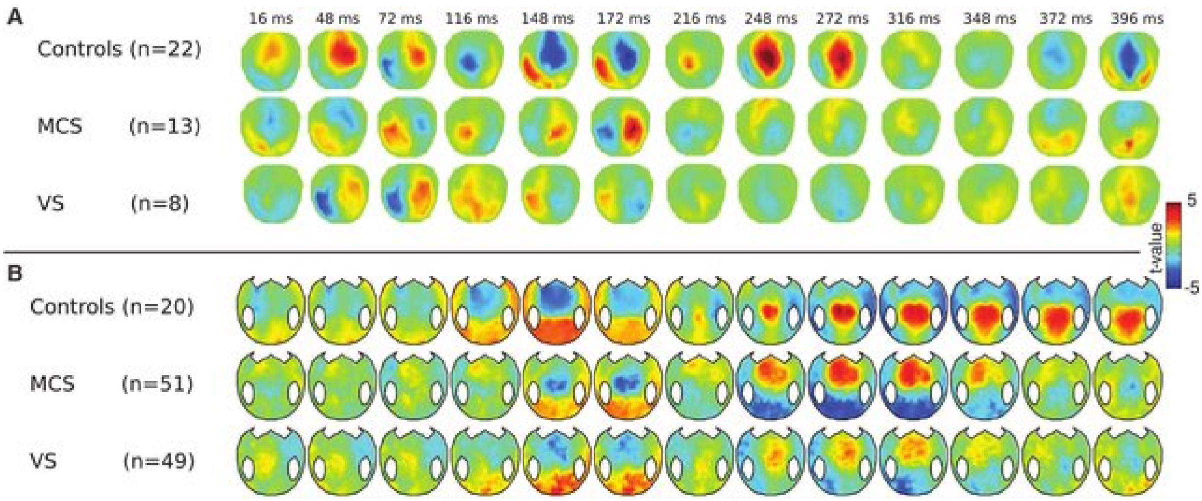


Figure 2.1: MMN topography in healthy controls patients versus DOC's patients. Observation of t test maps (deviation and standard trials) in A comparing with similar maps obtained in France between 2008 and 2011[11] [22]

also for soft tissue injuries and inflammation). This technique evaluates the activation and the emotional state of the subject when exposed to certain stimuli [24].

- GSR (galvanic skin response)

  Measurement of the galvanic response is done by placing electrodes on the fingers. Studies measure resistance skin and its conductance.

  A GSR amplifier applies constant tension to the skin using low voltage that the individual cannot perceive it through electrodes. The current generated in the skin by tension can be detected and recorded. The output of the GSR amplifier determines the conductance. The conductance of the skin gives feedback on the body's response to the stimulus; it is widely used in post-coma and coma phases [25, 26].

- Eye-tracking

  Eye-tracking refers to recording the movements of an individual's eye while examining a visual stimulus. In broad, it is responsible for measuring eye movements using a camera that quantifies them. Modern eye trackers record eye position and movement using contrast to locate the central point of the pupil and create a reflection of the cornea using infrared light. It is possible to analyse of the position of the gaze and the movement of the eyes in three-dimensional environments [27]. Eye-tracking techniques apply to domains where interaction and interests matter, looking to sell or immerse to improve the customer or user experience. In medicine, it is essential to recognize behaviours and patterns to investigate further and characterize[28].

- Face reading

Facial expressions are one of the most robust visual methods for conveying emotions. The face plays a crucial role in the cognitive processes of individuals, since the signs that show facial expressions denote internal states or emotions. The analysis of facial expressions provides valuable information when combined with other tools that allow sensory information collection, such as eye-tracking or EEG.

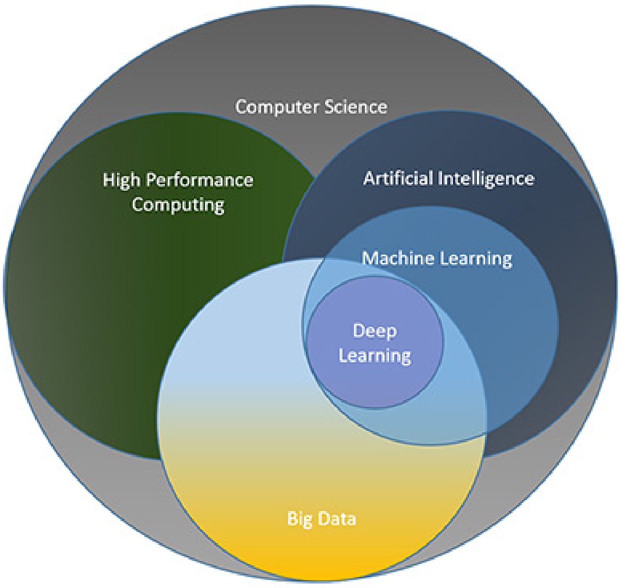
2.2 General idea of application

It is common in medical science to have pre-determined datasets for training and a new dataset(s) for testing. A typical use case is, for instance, a model based on the knowledge of n patients, trying to classify the status of other patients who will arrive [29]. The raise of high-performance computing (which includes GPU's, CPU's, high-speed interconnect, memory, libraries, storage, compilers, etc.) increased the speed and productivity of developing the algorithms that would generally take months to be completed, even at a distance, as in the case of cloud computing A [30].

Chapter 3

Machine Learning, a field of AI

Machine Learning is a subfield of Artificial Intelligence (AI), a branch of computer science whose fundamental aim is to develop computational systems (smart machines) that show operational behaviour and are capable of performing tasks that typically require human intelligence [31]. Those algorithms often exceed the capabilities of human at doing these tasks.



[30]

Figure 3.1: Machine learning and directly related areas

3.1 Historical Perspective

The field of AI began in the 1940s with Alan Turing, an English mathematician and computer scientist. Who is widely know as the father of Artificial Intelligence and Computer Science [32]. His outstanding achievements include:

- Cracking the Enigma code which, according to Keegan, was allegedly decisive in the course of the Second World War. As a result the war was significantly shortened and cracking the Enigma code contributed to the known outcome [33].

- Inventing of the first model for the general-purpose computers: Turing machine [34]

- Creating the Turing test, currently used, to find out whether an intelligent agent is capable of thinking like a human being. Through a conversation in natural language between a human and a computer, a human evaluator has to try to distinguish the computer from the human. When this happens, the test is negative. Else, it means that a computational system achieves a human-like level of artificial intelligence [35].

Up to this moment applications have been developed in the Narrow AI branch, this means that AI tries to perform well on one task, rather than strive for human intelligence, named Artificial General Intelligence (AGI) or beyond the best human brains of that, named Artificial Super Intelligence (ASI).

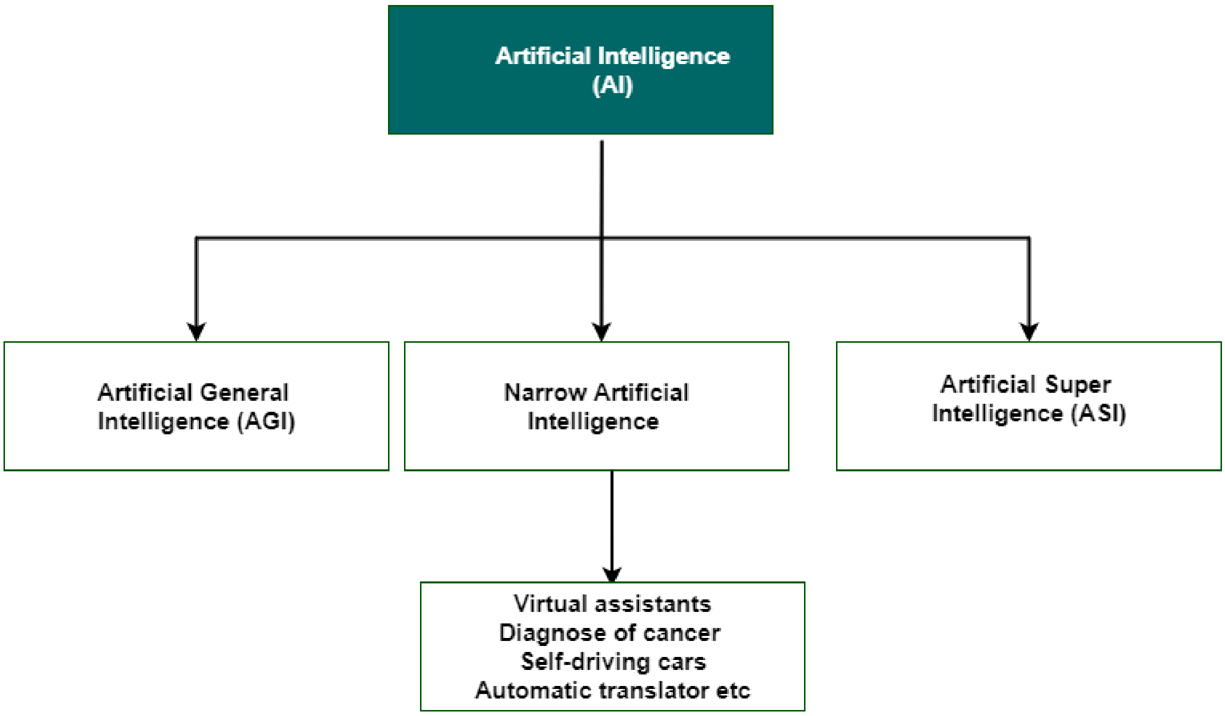


Figure 3.2: Types of AI

To reach the same objective that makes people seem intelligent using study of those processes/principles that make computers execute tasks more effectively and in a manner similar to humans [36]. To achieve that purpose we can use various programming techniques including non-deterministic search based partially on declarative languages, including logic-based, functional or object-oriented programming.

3.1.1 The evolution of AI in board games

The theory of neural network (NN) began in 1943 using electrical circuits as a nerve cell with a logic gate with binary outputs. In the 1950s, computer scientists simulated the idea in their work, a hypothetical NN.

Over time, we have seen successive application of artificial intelligence in board games, because they are both understandable and complex.

Followed events chronologically:

- In the 50's, the investigation on the game of checkers stands out. So learn by playing and giving a set of rules game as well a sense of direction and some parameters without including the weight of your importance. Arthur Samuel developed software that helped the IBM computer get better at checkers, the more it played. With the premise of the possibility of scaling and transposing these machine learning techniques to other areas [37].

- Already in 1997, IBM's Deep Blue (chess-playing computer), capable of evaluating 200 million positions per second (twice as fast as the 1996 version), beat grandmaster Garry Kasparov.

- And more recently in 2016, human intelligence head to head with machines, AlphaGo — AI program powered by Google. In the game Go, considered as the oldest in the world and probably the most complex. Google DeepMinds AlphaGo combines reinforcement learning techniques with simulation so is powered with all types of Go matches, and then study, learn from them, and discover your own moves. It knows the rules, and discovers highly efficient creative moves never before thought of by humans [38] .

Due to the evolution of the processing power of computers, it is now possible to run computationally demanding algorithms in terms of calculations and/or the amount of data used [39]. In this way, there was a path that made possible a new era, a revolution of Artificial intelligence called deep learning, that mimics neural networks of the human brain.

3.1.2 Perceptron

The Perceptron is like an atom of artificial intelligence that performs computations to extract information or knowledge from input data [?].

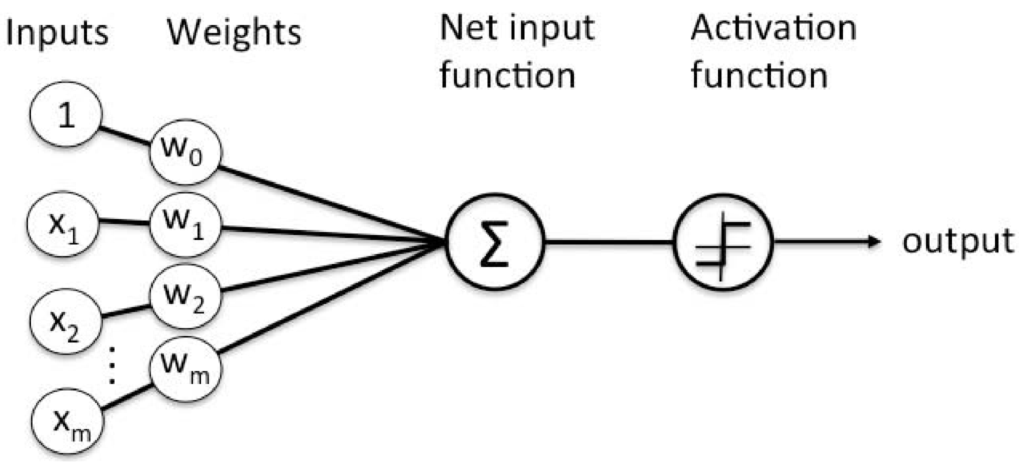


Figure 3.3: Perceptron model representation
[40]

3.2 Machine Learning types

Machine learning algorithms are categorized into three main categories:

- supervised learning 3.2.0.1
- unsupervised learning
- reinforcement learning

3.2.0.1 Supervised learning

Supervised learning is the subset of machine learning algorithms. The aim of supervised algorithms is to create a new model based on the features and labels to classify any given new point of data. The dataset consists of labeled examples $(x_i, y_i)_{i=1}^N$ where x_i is called a feature vector. A feature vector is a multidimensional vector that describes the example. The label y_i can be a real number, a vector or a matrix. It can also be a finite set of classes [41].

The supervised learning algorithms are broadly categorized as:

- Classification which predicts a category of a new data point.
- Regression which predicts a continuous variable.

Examples:

Linear Regression, assumption: the class is expected to be a linear combination 3.2.0.1 of the features. It could be adapted for higher dimensions.

$f(X) = \beta_0 + \beta_1 X_1 + \beta 2 X_2 + ... + \beta_p X_p$

There are two types of algorithms:

- **Lazy** like:
  - K-Nearest Neighbours or KNN (distance based technique)

    When a new point appears, we calculate the similarity (or distances, as in clustering) of the K-nearest Neighbours. Where k is a positive number, to obtain optimal value should use an error plot or accuracy plot. The class of our new input is assigned to the class most common, like voting system in democracy [29].
  - Rule Based The rule-based technique was widely used in healthcare in the 1970s. However, it has its limitations in comparison with the methods used nowadays. Those limitations include high monetary cost, dependence on specialized human resources, and complexity of updating the rules. The table below illustrates Practical implementation constraints.

Approaches	Model comprehensibility	Performance	Reproducibility	Dependency on prior knowledge	Development and training costs[a]	Running costs	Around-the-clock availability	Update costs
Human evaluation	High	Moderate or high	Moderate	High	High	High	Low	High
Rule-based algorithms	High	Moderate or high	High	High	Moderate or high	Low	High	High
Feature-based machine-learning methods	Moderate or high	Moderate or high	High	Moderate[b]	Moderate	Low	High	Moderate[c]
Deep artificial neural networks	Low or moderate	High	High	Low	Moderate	Low	High	Low

[a]The estimated cost of training professionals that carry out the clinical tasks (human evaluation) or of developing the automated system (rule-based, feature-based or deep-artificial-neural-network-based) that performs the tasks. [b]For feature-based machine-learning methods, prior knowledge may facilitate the derivation of useful features from the raw data. [c]When the update requires encoding new features, the update cost of feature-based machine-learning methods includes feature engineering and model retraining.

Figure 3.4: Human evaluation vs different types of AI paradigms
[15]

  - Naive Bayes Rules comes from the Bayes theorem, applies and calculates conditional probabilities and probabilities. Suitable for large and unknown datasets. Assumes that features are independent or unrelated to each other. Then the algorithm learns and collects information from each feature class individually. In Scikit-learn, there are three types of Naive Bayes classifiers: Multinomial, BernoulliNB and GaussianNB. GuassianNB is present in the project.

- **Eager** like:
  - Neural Networks behaviour similar to that of the human brain, as it connects small units, the neurons, in an organized way. When a neuron is sending information, we say it is "activated". Through these highly complex information exchanges, pattern

recognition is possible. When neural networks have many hidden layers, it is deep learning.

- SVMs Intends to maximize the decision boundary margin. The orange dots are the support vector, the points closer to the boundary of decision Figure 3.5.

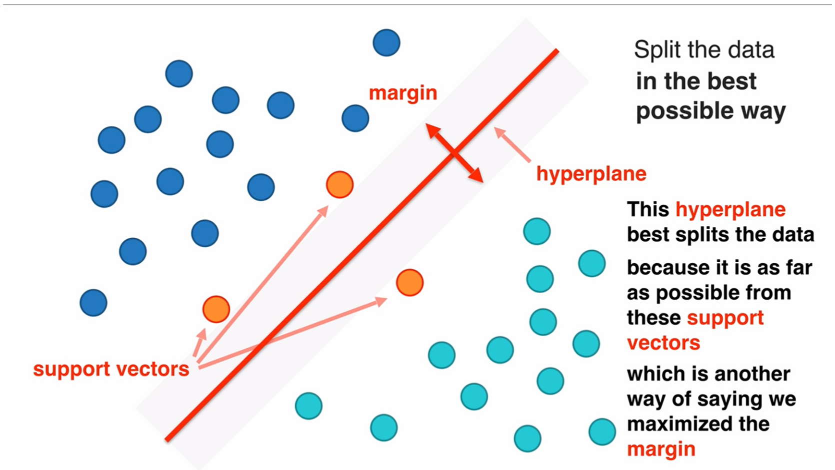


Figure 3.5: Visual Explanation of SVM
[42]

When the problem is not linearly separable use **the kernel trick**:

* Transition to a higher dimensional space — feature space
* Get an optimal boundary
* Return to input space

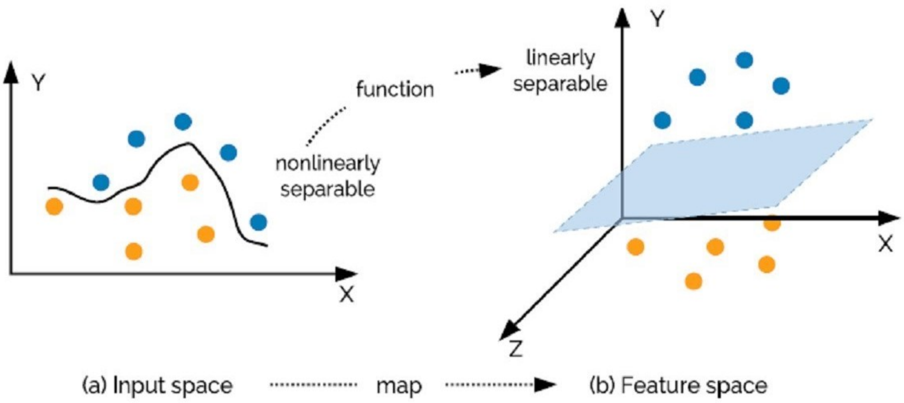


Figure 3.6: Visual Explanation of the kernel trick

- Decision trees

Hierarchical learning (decomposes in subtrees), to build one:

1. Select feature for root node and create a branch for each possible feature value

2. : split instances into subsets: one for each branch extending from the node

3. Repeat for each branch, using only the instances that reach the branch

A selected feature corresponds a question or "test" to the database in order to subdivide data until obtaining a branch of the same class called pure leaves. The goal is to pick the most significant features to subdivide in the faster way possible (minimum number of tests), avoiding overfitting. Note that only 1 feature is picked by the tree if there are more than encodes the same information. The benefits compared with other algorithms is ease of being understandable by the people, the ability to visualize beyond the algorithms doesn't variant to scaling of the data.

– Random Forests
  Multiple decision trees (ensemble method) , hierarchical learning, divide and conquer method. Bootstrapping of samples for each decision tree, and all decision trees, makes a prediction. In regression, it gives an average of results. For the other side in classification, makes a soft voting that consist that: the most probable class, the result of the average of the probabilities of all the probabilities of the predictions of the decision trees, is chosen with the representative label [43].

– Gradient boosted regression trees:

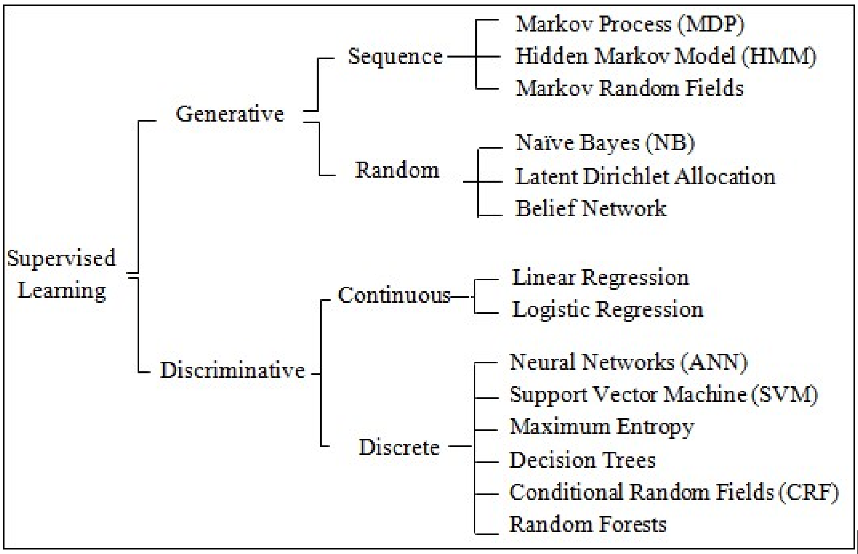
  * used for regression and classification

  * considered one of the most powerful boosting algorithms and widely used in supervised learning, nevertheless, it is required to tune the parameters carefully.

  * have a strong pre-pruning which results in shallow trees, typically no more than 5 levels deep consequently is smaller in memory and faster to predict. With more trees and n_estimators allow more chances to correct mistakes on the training set.

  * make a serialization of trees with the aim of predicting the error value of the previous tree. And with the learning rate parameter, manage the intensity of correctness in previous trees to the definition in its final result.

In this project we focus, use and explore Naive Bayes more supervised algorithms of the discriminative and discrete type according to the input data.



[44]

Figure 3.7: Taxonomy of supervised learning algorithms

In this image Figure 3.8, we summarize the main features of the algorithms that stand out.

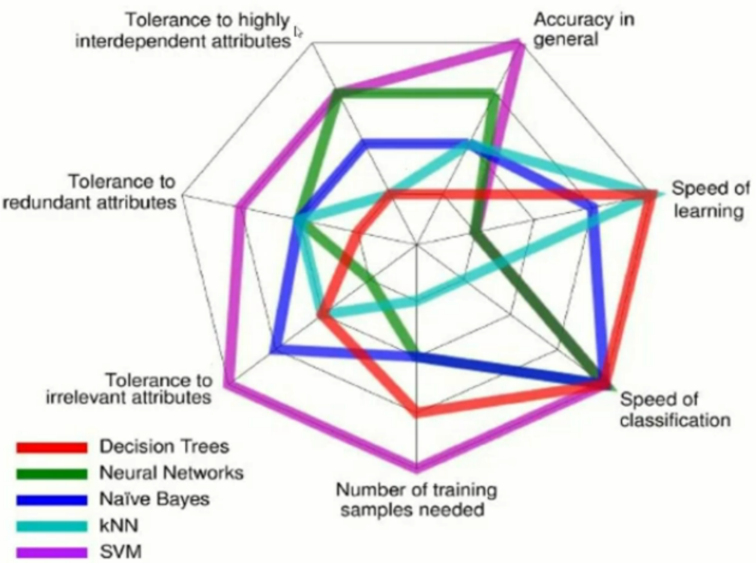


Figure 3.8: Main features of the algorithm

3.3 Model Evaluation

The common practice in the field is to evaluate the model to see if it has the desired performance.

3.4 Pipeline steps of a project

- Collecting the data

- Get some features
  - Extract, Analyse and Select
- Get a model
- Train a model
- Test the model
- Evaluate the model, adjust and repeat if necessary

It is worth pointing out that in case of data limitation, it is important to test with the new data. In the evaluation phase, it is expected that:

- A model is able to generalize from seen data to unseen data.
- Is “Able” according to some metrics

Furthermore, there are statistical methods that allow us to assess the model in terms of predictions and get more information.

3.5 Evaluation process

- We split the dataset into a training dataset and a test dataset. Under no circumstances, we should use the "test" dataset to train and perform evaluation.

Otherwise, the data will be skewed. That may lead to the misdiagnosis of performance at levels higher than the real ones. Testing the model will take an example from the dataset, evaluate it without adjusting the model. The evaluation result or prediction of the model will be compared with the ground truth labels.

- The ground truth labels are typically not given to the model to make the prediction.
- They are used to confirm if the prediction was correct or not.

Let us examine a simple example with two categorical classes: "yes" or "no" If a dataset example has a label "yes" and the model predicted as a "yes", the model has correctly classified the example. Otherwise, it is an erroneous prediction.

For each input x with ground truth class y we have: Correctly discriminated/classified examples Incorrectly discriminated/classified examples

From the data in the other table, we can deduce a set of metrics [46].

It is good practice to save this data (confusion matrix) because from it, we can calculate everything from these metrics [48].

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	True Positives (TP)	False Negatives (FN)
	Negative	False Positives (FP)	True Negatives (TN)

Figure 3.9: Confusion Matrix
[45]

		Predicted condition			
	Total population $= P + N$	**Positive (PP)**	**Negative (PN)**	Informedness, bookmaker informedness (BM) $= TPR + TNR - 1$	Prevalence threshold (PT) $= \frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$
Actual condition	**Positive (P)**	**True positive (TP),** **hit**	**False negative (FN),** **type II error, miss, underestimation**	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{TP}{P} = 1 - FNR$	False negative rate (FNR), miss rate $= \frac{FN}{P} = 1 - TPR$
	Negative (N)	**False positive (FP),** **type I error, false alarm, overestimation**	**True negative (TN),** **correct rejection**	False positive rate (FPR), probability of false alarm, fall-out $= \frac{FP}{N} = 1 - TNR$	True negative rate (TNR), specificity (SPC), selectivity $= \frac{TN}{N} = 1 - FPR$
	Prevalence $= \frac{P}{P + N}$	Positive predictive value (PPV), precision $= \frac{TP}{PP} = 1 - FDR$	False omission rate (FOR) $= \frac{FN}{PN} = 1 - NPV$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Negative likelihood ratio (LR−) $= \frac{FNR}{TNR}$
	Accuracy (ACC) $= \frac{TP + TN}{P + N}$	False discovery rate (FDR) $= \frac{FP}{PP} = 1 - PPV$	Negative predictive value (NPV) $= \frac{TN}{PN} = 1 - FOR$	Markedness (MK), deltaP (Δp) $= PPV + NPV - 1$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
	Balanced accuracy (BA) $= \frac{TPR + TNR}{2}$	F_1 score $= \frac{2PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$	Fowlkes–Mallows index (FM) $= \sqrt{PPV \times TPR}$	Matthews correlation coefficient (MCC) $= \sqrt{TPR \times TNR \times PPV \times NPV} - \sqrt{FNR \times FPR \times FOR \times FDR}$	Threat score (TS), critical success index (CSI), Jaccard index $= \frac{TP}{TP + FN + FP}$

Figure 3.10: Full matrix confusion table with metrics included
[47]

Chapter 4

Experimental Study

4.1 SMART database considerations

This project consists of taking up the technique of diagnosing patients with DOC. For that, I had access to a dataset with 35 records of the assessment of this technique that took place in a hospital in London, where Professor Liliana Teixeira worked. Professor Liliana allowed me to access this dataset and was kind enough to explain how this technique works and clarify all the doubts that followed from reading the data and its particularities. The dataset has 147 columns of feature data. Many columns are repeated from 10 sessions that the technique uses for the evaluation of 8 modalities. There are 5 sensory modalities (gustatory, olfactory, auditory, visual and tactile), the rest are functional communication, wakefulness/arousal and motor function [49]. In this way, 8 modalities are evaluated in 10 times, to make up 80 features that enlarge the data set. The multiple assessments of the same modality are, often, synonymous with feature redundancy. The individual assessment of each session is done at the end of each session and this technique covers 2 evaluators, in 10 sessions and each of these is associated with the degree of certainty. In total we have over 40 columns/features. In addition, there is a subsequent evaluation where the assessors discuss and reach a consensus on the joint result for each session and the associated degree of certainty. So we have 20 more features. Of the other 7 features that remain, 2 are for patient identification and their evaluation order. These two features introduce noise to the model creation and are not taken into consideration while creating supervised learning models. Another 2 columns correspond to the y or diagnostic labels (in case of 2 and 3 possible states). The labelling mode is shown in the following tables:

Table 4.1: Classification with 2 states

Label	Condition	Quantity
1	Vegetative State	10
2	Minimally Conscious State	25

Table 4.2: Classification with 3 states

Label	Condition	Quantity
1	VS	10
2	MCS-	11
3	MCS+	14

The **time** variable is the time interval, in months, between brain injury and SMART assessment. Age refers to the patient's age. These features are continuous, so, in the following table, you can check information about these:

Table 4.3: Describe Age and Time

Features	Mean	Minimum	Maximum	Median
Time (in months)	8,2	3	70,0	52,0
Age (in years)	49,7	19,0	77,0	6,0

The other two features are the patient's gender and the etiology/origin of the disease, they are binary with the following meaning:

Table 4.4: Gender code definition

Label	Gender	Quantity
1	Female	7
2	Male	28

Table 4.5: Etiology code definition

Label	Etiology	Quantity
1	Traumatic	11
2	Not traumatic	24

4.2 Results

The project repository is available at: https://github.com/ManJ-PC/ML-DoC . The chosen platform is GoogleColab.

In this section we distinguish machine learning algorithms and deep learning algorithms (in the case of neural networks).

In the confusion matrix, the positive corresponds to the VS classes and the negative to the MCS classes.

The choice of two states is motivated to preserve the collected data. The algorithms classify the data into three states (MCS-, MCS+ and VS) not present in the older collections. The binary classification was used to ensure more records in the database. In the AUC graphs, we can evaluate the prediction of the model at the correctness level. The best results have been achieved with the deep learning approach (using neural networks) and algorithms such as SVM and Random forest.

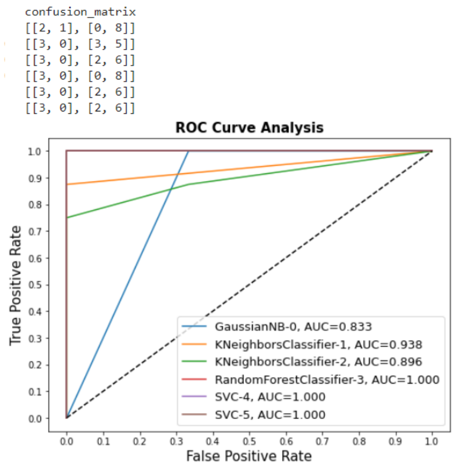


Figure 4.1: Confusion Matrix and ROC curve of a Bunch of Algorithms

4.3 Data exploration and analysis

Upon close examination of the data summary, we can surmise that the classification for the two states is unbalanced. 4.1.

And it turns out that there are fields with null values, 111 in total. This happens when the patient is blind or the gustatory sensory system is not possible to assess. It happens in the situations like patients are fed through a gastric tube or other alternative feeding mechanisms such as transcheostemia, etc. All 'NA' are changed and assigned a non-significant value: -1. The reason for doing so is that data is precious and scarce, and removing records would further shorten our dataset. If we do not have sufficient. Without data, you cannot make the algorithm make a perfect model. The ID and order insignificant columns are removed. And the column of the diagnosis made by human evaluators, our label, is separated from X for an isolated panda's series y.

4.4 Machine Learning

X and y are split into training and testing subsets. Taking care to stratify the data in relation to the label (stratify=y). This is to ensure that all classes are proportional in relation to the training side to the testing side, avoiding the risk of poorly trained classes.

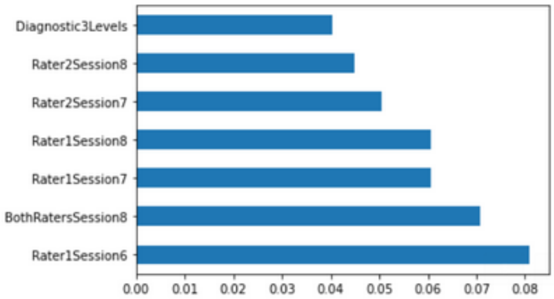
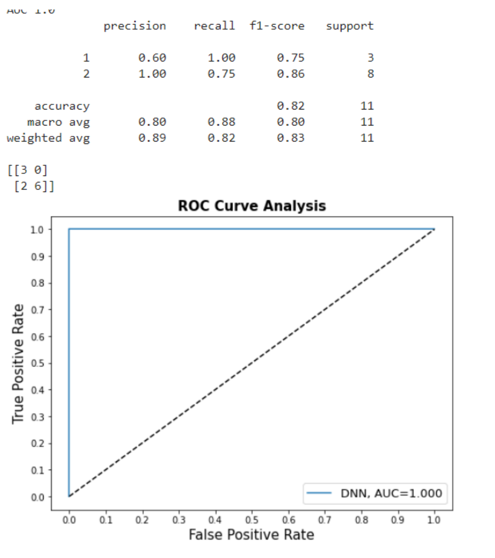


Figure 4.2: Features importances in the Random Forest model



	precision	recall	f1-score	support
1	0.60	1.00	0.75	3
2	1.00	0.75	0.86	8
accuracy			0.82	11
macro avg	0.80	0.88	0.80	11
weighted avg	0.89	0.82	0.83	11

```
[[3 0]
 [2 6]]
```


Figure 4.3: Neural Network, Confusion Matrix and ROC curve resulting from the iterative method with 100 epochs, 25 neurons and ADAM optimizer

Chapter 5

Conclusions and Future Work

In this chapter, we introspect the project. We compare the proposed objectives with those achieved, and we will see the differences and why they exist. The main objective of the dissertation is to carry out the steps of the machine learning process such as: data pre-processing (analysis, cleaning), mining and application of techniques such as selection, reduction and sub-setting, in the first part. Through the generated models, we reach the DOC's distinction states with a high rate of precision. However, a lot of work can still be done, as the aim is that the model can shorten or optimize the work performed by health professionals.

5.1 Further Work

The project can be further developed to find correlation of features in algorithms. Divide in two subgroups based on etiology. We can also see if it is possible to do fewer sessions to minimize resources of this technique (human resources, time spend). Since it may no longer be adding value to the data.

- A broader view of classification in 3 possible stages (VS, MCS- and MCS+)
- Iterate over the various machine learning algorithms
- Unification of similar features [50]
- Predict the correlation of selected features and limit the number of sessions needed
- Use of diagnostic tools to sens ory data acquisition
- Test the model with new and actual data where it does not contain as rather sessions
- Automate the entire process, acquire technological tools that allow you to track patients regularly

Appendix A

Glossary

Coma: from the Greek kôma = deep sleep, can be defined as a state of total or partial loss of consciousness, voluntary motor skills and sensitivity, generally due to brain damage, intoxications, metabolic and endocrine problems, without which, regardless of severity, vital functions are maintained to a greater or lesser degree [51]. When physiological, the state of coma can be measured using the Glasgow Coma Scale (GCSl) and when pharmacological using the Ramsay Sedation Scale (RSS).

Cloud Computing: technology that allows remote use of computing resources through Internet connectivity

Bibliography

[1] M. T. SERAFINI and G. MONIZ, “Saber Estudar e Aprender,” p. 249, 1991.

[2] A. M. Owen, “Disorders of consciousness,” *Annals of the New York Academy of Sciences*, vol. 1124, no. 1, pp. 225–238, 2008.

[3] Z. S. Zheng, N. Reggente, E. Lutkenhoff, A. M. Owen, and M. M. Monti, “Disentangling disorders of consciousness: Insights from diffusion tensor imaging and machine learning,” *Human brain mapping*, vol. 38, no. 1, pp. 431–443, 2017.

[4] Hong Song, Qiang Li, Shixiong Li, Wei Kang, Jian Yang, Yi Yang, and Jianghong He, “Classification of patients with disorder of consciousness based on DTI sequence analysis,” *2017 IEEE International Conference on Smart Cloud (SmartCloud). Proceedings*, 2017.

[5] A. Bender, R. J. Jox, E. Grill, A. Straube, and D. Lulé, “Wachkoma und minimaler Bewusstseinszustand: Systematisches Review und Metaanalyse zu diagnostischen Verfahren,” *Deutsches Arzteblatt International*, vol. 112, pp. 235–242, apr 2015.

[6] L. Teixeira, D. Blacker, and N. Rocha, “Disorders of consciousness,” *Multidisciplinary Interventions for People with Diverse Needs-A Training Guide for Teachers, Students, and Professionals*, p. 27, 2020.

[7] K. Andrews, L. Murphy, R. Munday, and C. Littlewood, “Misdiagnosis of the vegetative state: retrospective study in a rehabilitation unit,” *Bmj*, vol. 313, no. 7048, pp. 13–16, 1996.

[8] H. Gill-Thwaites and R. Munday, “The sensory modality assessment and rehabilitation technique (smart): a valid and reliable assessment for vegetative state and minimally conscious state patients,” *Brain injury*, vol. 18, no. 12, pp. 1255–1269, 2004.

[9] C. Schnakers, A. Vanhaudenhuyse, J. Giacino, M. Ventura, M. Boly, S. Majerus, G. Moonen, and S. Laureys, “Diagnostic accuracy of the vegetative and minimally conscious state: clinical consensus versus standardized neurobehavioral assessment,” *BMC neurology*, vol. 9, no. 1, pp. 1–5, 2009.

[10] H. Gill-Thwaites and R. Munday, “The sensory modality assessment and rehabilitation technique (SMART): A valid and reliable assessment for vegetative state and minimally conscious state patients,” *Brain Injury*, vol. 18, no. 12, pp. 1255–1269, 2004.

[11] J. R. King, T. Bekinschtein, and S. Dehaene, “Comment on "preserved feedforward but impaired top-down processes in the vegetative state",” dec 2011.

[12] C. Da and S. Lourenço, “Deep Learning for EEG Analysis in Epilepsy,” 2019.

[13] d. C. T. L, G.-T. H, R. F, and D. S, “Can behavioural observations made during the SMART assessment detect the potential for later emergence from vegetative state?,” *Neuropsychological rehabilitation*, vol. 28, pp. 1340–1349, nov 2018.

[14] M.-S. T. F. on PVS, “Medical aspects of the persistent vegetative state,” *New England Journal of Medicine*, vol. 330, no. 21, pp. 1499–1508, 1994.

[15] K.-H. Yu, A. L. Beam, and I. S. Kohane, “Artificial intelligence in healthcare,” *Nature Biomedical Engineering 2018 2:10*, vol. 2, pp. 719–731, oct 2018.

[16] L. da Conceição Teixeira, H. Gill-Thwaites, F. Reynolds, and S. Duport, “Can behavioural observations made during the SMART assessment detect the potential for later emergence from vegetative state?,” *Neuropsychological Rehabilitation*, vol. 28, pp. 1340–1349, nov 2018.

[17] S. Corchs, G. Chioma, R. Dondi, F. Gasparini, S. Manzoni, U. Markowska-Kacznar, G. Mauri, I. Zoppis, and A. Morreale, “Computational methods for resting-state EEG of patients with disorders of consciousness,” 2019.

[18] O. Avilov, S. Rimbert, A. Popov, and L. Bougrain, “Deep Learning Techniques to Improve Intraoperative Awareness Detection from Electroencephalographic Signals,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2020-July, pp. 142–145, Institute of Electrical and Electronics Engineers Inc., jul 2020.

[19] Z. Xu, J. Wang, R. Wang, Z. Zhang, and S. Yang, “Deep Convolutional Neural Network for Detection of Disorders of Consciousness,” in *Chinese Control Conference, CCC*, vol. 2020-July, pp. 7084–7089, IEEE Computer Society, jul 2020.

[20] Z. S. Zheng, N. Reggente, E. Lutkenhoff, A. M. Owen, M. M. Monti, and Z. Sheng Zheng, “Medical Research Council,” 2016.

[21] P. J. Lang and M. M. Bradley, “Emotion and the motivational brain,” jul 2010.

[22] M. Boly, M. I. Garrido, O. Gosseries, M. A. Bruno, P. Boveroux, C. Schnakers, M. Massimini, V. Litvak, S. Laureys, and K. Friston, “Response to comment on "preserved feedforward but impaired top-down processes in the vegetative state",” dec 2011.

[23] “fMRI Imaging: How Is an fMRI Done? - fMRI Imaging | HowStuffWorks.”

[24] “FMRI Functional Magnetic Resonance Imaging Lab.”

[25] Ç. G. Altıntop, F. Latifoğlu, A. K. Akın, R. İleri, and M. A. Yazar, "Analysis of Consciousness Level Using Galvanic Skin Response during Therapeutic Effect," *Journal of Medical Systems*, vol. 45, jan 2021.

[26] J. Luauté, A. Dubois, L. Heine, C. Guironnet, A. Juliat, V. Gaveau, B. Tillmann, and F. Perrin, "Electrodermal reactivity to emotional stimuli in healthy subjects and patients with disorders of consciousness," *Annals of Physical and Rehabilitation Medicine*, vol. 61, pp. 401–406, nov 2018.

[27] R. Taherkhani and M. Kia, "Designing a high accuracy 3D auto stereoscopic eye tracking display, using a common LCD monitor," 2012.

[28] W. K. C. Ting, J. L. Perez Velazquez, and M. D. Cusimano, "Eye movement measurement in diagnostic assessment of: Disorders of consciousness," 2014.

[29] E. Muller, J. P. Shock, A. Bender, J. Kleeberger, T. Högen, M. Rosenfelder, B. Bah, and A. Lopez-Rolon, "Outcome prediction with serial neuron-specific enolase and machine learning in anoxic-ischaemic disorders of consciousness," *Computers in Biology and Medicine*, vol. 107, pp. 145–152, 2019.

[30] "HPC Advisory Council - Subgroups - HPC | GPU."

[31] D. B. Spalding, "A general purpose computer program for multi-dimensional one- and two-phase flow," *Mathematics and Computers in Simulation*, vol. 23, pp. 267–276, sep 1981.

[32] A. Beavers, "Alan turing: Mathematical mechanist," *Cooper, S. Barry; van Leeuwen, Jan. Alan Turing: His Work and Impact. Waltham: Elsevier*, pp. 481–485, 2013.

[33] S. J. Keegan, *Intelligence in War: Knowledge of the Enemy from Napoleon to Al-Qaeda*. New York: Alfred A. Knopf, 2003.

[34] A. R. Booker, "Turing and the riemann hypothesis."

[35] J. P. Bowen, "Alan Turing: Founder of Computer Science," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10215 LNCS, pp. 1–15, 2016.

[36] S. J. Russell and P. Norvig, *Artificial Intelligence: a modern approach.* Pearson, 3 ed., 2009.

[37] A. L. Samuel, "Some Studies in Machine Learning Using the Game of Checkers,"

[38] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, "Mastering the game of Go without human knowledge," *Nature 2017 550:7676*, vol. 550, pp. 354–359, oct 2017.

[39] "Explore the history of machine learning."

[40] F. Rosenblatt, "THE PERCEPTRON: A PROBABILISTIC MODEL FOR INFORMATION STORAGE AND ORGANIZATION IN THE BRAIN 1," *Psychological Review*, vol. 65, no. 6, pp. 19–27, 1958.

[41] A. Burkov, "The Hundred-Page Machine Learning Book, Quebec City, Canada, 2019, ISBN 978-1999579517," 2020.

[42] "(3) Support Vector Machines: A Visual Explanation with Sample Python Code - YouTube."

[43] A. MUller, *Introduction to machine learning with Python : a guide for data scientists*. 2017.

[44] "Taxonomy of machine learning algorithms | Scala for Machine Learning - Second Edition."

[45] B. Sheng, O. M. Moosman, B. Del Pozo-Cruz, J. Del Pozo-Cruz, R. M. Alfonso-Rosa, and Y. Zhang, "A comparison of different machine learning algorithms, types and placements of activity monitors for physical activity classification," *Measurement: Journal of the International Measurement Confederation*, vol. 154, mar 2020.

[46] "What is a Confusion Matrix? - Unite.AI."

[47] "Confusion matrix - Wikipedia."

[48] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011.

[49] H. Gill-Thwaites and R. Munday, "The Sensory Modality Assessment and Rehabilitation Technique (SMART): A Comprehensive and Integrated Assessment and Treatment Protocol for the Vegetative State and Minimally Responsive Patient," *http://dx.doi.org/10.1080/096020199389392*, vol. 9, pp. 305–320, jul 2010.

[50] J. Patel, "Unification of machine learning features," in *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, pp. 1201–1205, 2020.

[51] "The language and imagery of coma and brain injury: Representations in ... - matthew colbeck - google livros."