

On the clinical acceptance of black-box systems for EEG seizure prediction

Mauro Pinto (✉ mauro_p@live.com.pt)

Univ Coimbra, CISUC, Department of Informatics Engineering, Coimbra, Portugal

<https://orcid.org/0000-0002-9359-4324>

Adriana Leal

Univ Coimbra, CISUC, Department of Informatics Engineering, Coimbra, Portugal

Fábio Lopes

Univ Coimbra, CISUC, Department of Informatics Engineering, Coimbra, Portugal

José Pais

Hospital CUF Tejo

António Dourado

Univ Coimbra, CISUC, Department of Informatics Engineering, Coimbra, Portugal

Francisco Sales

Refractory Epilepsy Reference Centre, Centro Hospitalar e Universitário de Coimbra, EPE,
Coimbra, Portugal

Pedro Martins

Univ Coimbra, CISUC, Department of Informatics Engineering, Coimbra, Portugal

César Teixeira

Univ Coimbra, CISUC, Department of Informatics Engineering, Coimbra, Portugal

Article

Keywords: seizures, grounded theory, actor-network theory, black-box systems

Posted Date: September 21st, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-886717/v2>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Epilepsia Open on April 4th, 2022. See the published version at <https://doi.org/10.1002/epi4.12597>.

On the clinical acceptance of black-box systems for EEG seizure prediction

Mauro. F. Pinto^{1,*}, Adriana Leal¹, Fábio Lopes¹, José Pais²,
António Dourado¹, Francisco Sales³, Pedro Martins¹, and César A.
Teixeira¹

¹Univ Coimbra, CISUC, Department of Informatics Engineering,
Coimbra, Portugal

²Hospital CUF Tejo, Lisbon, Portugal

³Refractory Epilepsy Reference Centre, Centro Hospitalar e
Universitário de Coimbra, EPE, Coimbra, Portugal

*mauropinto@dei.uc.pt

Abstract

Despite 46 years of seizure prediction research, few devices/systems underwent clinical trials and/or are commercialised, where the most recent state-of-the-art approaches are not used to their full potential. This demonstrates the existence of social barriers to new methodologies. Based on the literature, we performed a qualitative study to analyse the seizure prediction ecosystem to find these barriers. With Grounded Theory and Actor-Network Theory, we draw hypothesis from data and considered that technology shapes social configurations and interests. For seizure prediction, as long as an algorithm proves to be useful to the patient, we conclude that we may only need to explain the model's decisions, and not to necessarily obtain intrinsically interpretable models. Accordingly, we argue that it is possible to develop robust prediction models, including black-box systems to some extent, while avoiding data bias, ensuring patient safety, and still complying with legislation, as long as they can deliver human-comprehensible explanations.

1 Introduction

Despite anti-epileptic drugs and surgical treatments, more than 20 million people suffer from uncontrolled epileptic seizures, bringing social and economic impact. Patients may suffer from discrimination and stigma along with significant health care needs, loss of productivity, and death. A solution for uncontrolled seizures might come from prediction [1, 2, 3, 4], as its timely anticipation opens

the way to several seizure control strategies, such as: i) closed-loop systems that trigger drug delivery or electrical brain stimulation; ii) warning devices that inform the patient to prevent accidents (e.g., falling from stairs) and/or to self-administer rescue-medication [5, 6, 7].

Although seizure prediction research started in the 1970s through electroencephalogram (EEG) analysis [8, 9, 10], few predictive devices [11] and closed-loop systems [12] have been clinically approved for trial. Additionally, these were based on "detection features alone" (line-length, bandpass, and energy-related) [13], which may be less robust than current state-of-the-art approaches [9]. In fact, an overview of current research uncovers the existence of major multidisciplinary barriers [9, 14, 14]. For instance, to develop a trustful, robust, and commercial solution one needs to handle expectations and beliefs from all actors of this ecosystem: technology and data scientists, clinicians, industry, legislation, ethics, and patients [14, 15, 16, 17].

We inspected the seizure prediction literature to understand the social difficulties, based on Grounded Theory (GT) [18] and Actor-Network Theory (ANT) [19]. GT is a standard methodology applied in qualitative research where researchers draw hypotheses from data: unlike most quantitative methods, data collection is not part of a process to test a pre-existing hypothesis. In short, it is the identification, and iterative refinement of relevant subjects from data [18, 20]. ANT main characteristics are its focus on inanimate entities and subsequent effects on social processes. Technology emerges from social interests and configures social interactions instead of handling technology as an external force. Thus, ANT can be useful for studying information technology implementations in healthcare settings [19, 21].

We present here a social network [22] that describes the relations between all actors. By using encapsulation, we can deliver a more general overview while deepening technical aspects that can be accessed individually. Furthermore, we explored how and why this ecosystem operates like it does, which helped to unravel paths that may lead to the successful development of new seizure prediction devices. Our main conclusion is that trust plays a fundamental role in increasing the number of clinically approved studies and subsequent commercial devices. The absence of an explanation for black-box decision models, especially when they fail, leads researchers to question and mistrust its use, and thus rising scepticism. This is the reason why some authors argue the use of only interpretable models [23].

However, for the specific case of seizure prediction, we believe that efforts should focus on explainability (and not necessarily on intrinsically interpretable models) as it is sufficient to reinforce trust, patient safety, ethics, and compliance with applicable law and industry standards. Explainability may be the key aspect that allows the entrance of promising deep learning approaches in clinical practice, as these hold great potential. Note that interpretability and explainability are different concepts [24]. While the former regards the extent to which a system output can be predicted by a given input, which is clear by using intrinsically interpretable models with a reduced set of features, explainability concerns how to explain the decisions that were made.

By providing a social understanding and guidelines for effective communication between actors, we hope this work contributes towards new clinically trusted methodologies, particularly for the work of those who develop software seizure prediction approaches, so that they have a higher chance of clinical acceptance. Conversely, it may also help clinicians to understand this software research area. Although these guidelines may have been implicitly used by the academic community for several years, we believe that their formalization is interesting and particularly useful. However many concepts developed here may also be applied to other healthcare areas where built devices implement algorithms developed from clinical data.

2 Results

We present here a summarized version of the seizure prediction ecosystem, which is shown chronologically, and our proposed guidelines. In Supplementary Material, we provide the social network in an interactive presentation, where encapsulation aspects, other details, and a step-by-step product design explanation are more intuitive. Thus, the reader is allowed to interactively explore the whole ecosystem. In addition, we also focus here on the findings that relate to clinical trials, explainability, and interpretability.

Seizure prediction ecosystem

Figure 1 depicts the obtained social network, which describes the relations between actors. Actors (x) and relations ($x-y$) are named with numbers and grouped in colours to provide a better understanding. We will explain these relations throughout this section while deepening parts that require more detail. In the end, we provide guidelines to help authors design their research.

We begin with a Drug-Resistant Epilepsy (DRE) patient (1). Years after diagnosed with Drug-Resistant Epilepsy (DRE), a patient is referred to an epilepsy centre to undergo pre-surgical monitoring (5). The EEG signal (4) is acquired to inspect brain activity to localize the epileptic focus. If easily localized, removing the epileptic region is a possible solution [8, 25]. This data will be stored and constitute retrospective data (7). The majority of databases available to perform academic studies (8) concerns pre-surgical monitoring conditions.

Studies try to capture and understand brain dynamics with the goal of predicting seizures (8- -4). Inevitably, we make several assumptions (see "Assumptions" section in Supplementary Material for more information) when we design a new study. These may result from the used mathematical models, available data and other limitations, or even reflect the researcher knowledge concerning brain dynamics (8- -4). These studies must also envision a real application scenario by simulating a prospective scenario (8- -15). To do this, studies must then comply with some requirements (9), have appropriate design parameters (10) concerning the real application, propose a discriminative model (11), and discuss its performance (12). Model design (19) is one of the most explored sec-

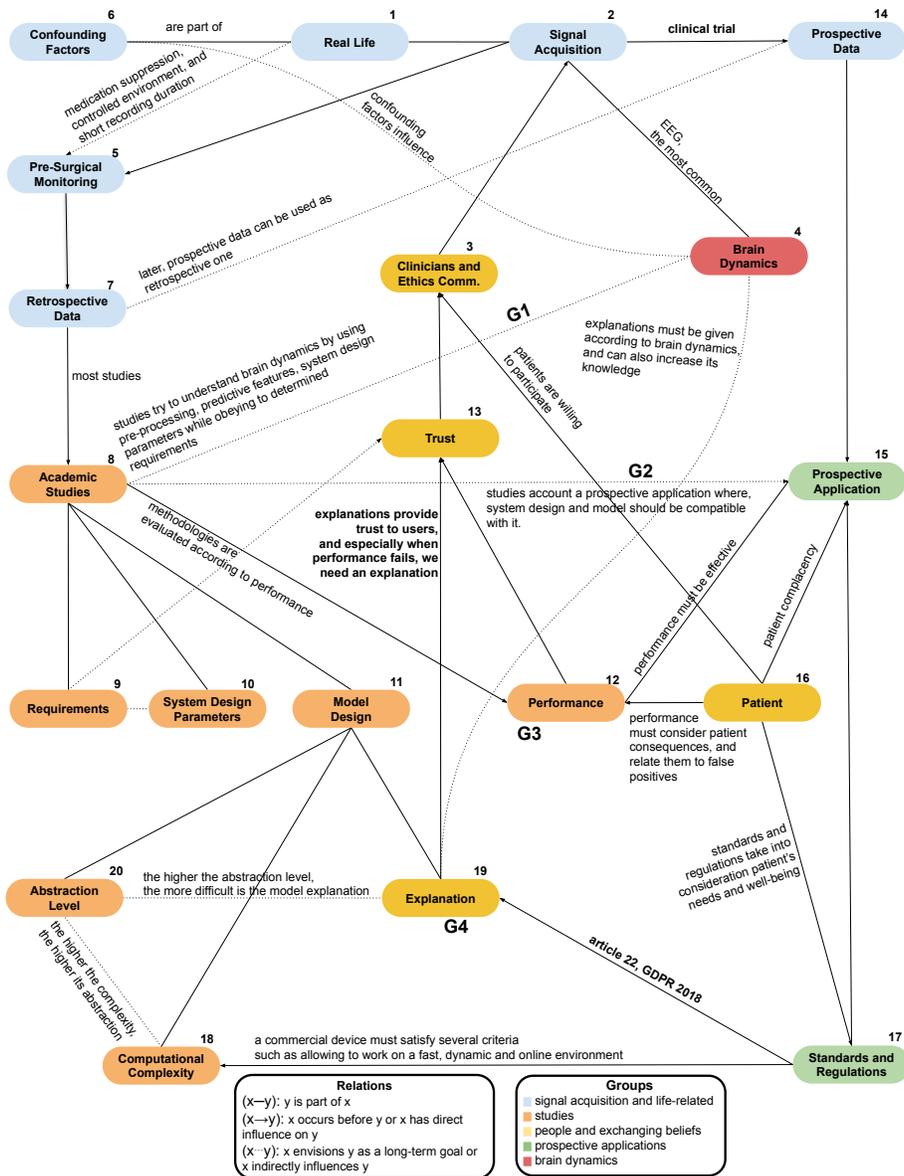


Figure 1: The obtained network: the relations between the major actors of epilepsy seizure prediction ecosystem. All actors are numbered to provide an intuition regarding the first steps to consider when developing seizure prediction systems. G1, G2, G3 and G3 are the proposed guidelines.

tions (we include here pre-processing, feature extraction, and model training).

A model can be characterized according to computational complexity (18) and abstraction level (20).

To start a clinical trial, we also need trust (13). Data scientists and clinicians need to find a given methodology trustworthy. We need to ensure patient safety, model robustness, and avoid bias. Thus, high performance is a necessary condition (12→13), but it is not enough. We also need to explain our model's decisions (19→13), to ensure safety and model effectiveness. Note that, for the particular case of seizure prediction, we need to know how to explain the model's decision but we may not necessarily need intrinsically interpretable models, as seen in the next sections with the Neurovista Advisory System [11].

For clinical trials, we argue the possibility of using complex prediction models, including black-box systems to some extent, as long as authors provide efforts on avoiding data bias, ensuring patient safety, and explaining their models' decisions. Furthermore, explanations not only increase trust and mitigate scepticism on artificial intelligence algorithms, but they can also deliver new knowledge on brain dynamics (19- -4).

Concerning legislation (17), the 2018 GDPR also promotes the delivery of model explanations (not necessarily intrinsically interpretable models). Current legislation should be seen as a reinforcement of safe methodologies, that considers patient's needs and well-being (13→17). When data scientists and clinicians trust the proposed methodology, the ethics committee can accept a clinical trial (13→3). In this case, patients are invited to participate in clinical trials (16→3).

After the ethics committee approval and patients' agreement to volunteer, a clinical trial starts. The prospective data (14) later becomes retrospective (7) and is used in an indefinite number of studies. With the prospective data, we can use intervention in real-time. By timely anticipating a seizure, we can trigger an intervention. To do this, we need to guarantee that the false-positive interventions are not harmful to the patient (16→15) and community. The intervention must also comply with all the industry standards and safety measures (17). It must have fast processing, do not have hardware problems, and be of easy placement and removal.

Studies Guidelines

By describing and discussing all relations, we inferred four guidelines that may help authors in guiding their research on seizure prediction. Figure 2 depicts a production process of a hypothetical device. Firstly, authors perform studies with retrospective data, in which they evaluate performance and the quality of given explanations. Clinicians and data scientists trust models' decisions when these are human-comprehensible, also increasing the confidence of the volunteering patients. In this case, an ethics committee may have strong reasons to approve a prospective study with an intervention system. Finally, the built device reaches its goal: improve the life of DRE patients.

The first guideline (*G1*) concerns undertaken assumptions on brain dynamics, which differ between studies due to available data and used methodology. Authors should state their assumptions regarding brain dynamics before pre-

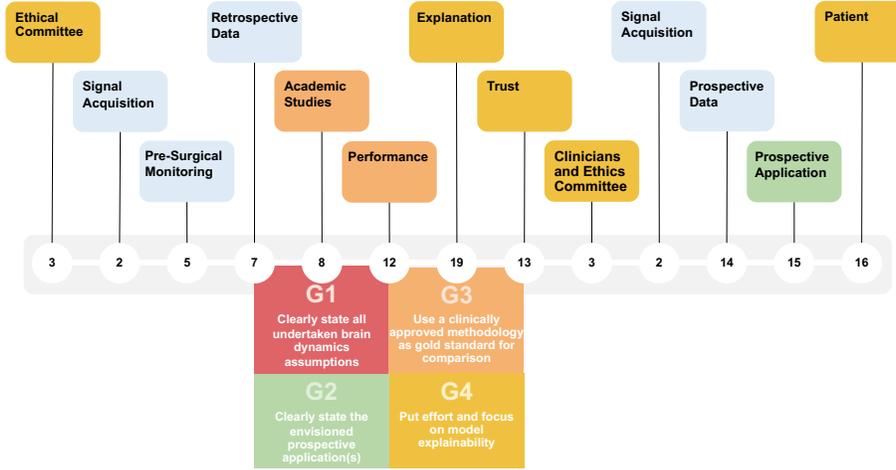


Figure 2: A product process for a seizure prediction prospective application, while showing our guidelines concerning designing academic studies.

senting the mathematical tools used in data analysis. Experienced researchers may understand what is at stake. However, others may benefit from the assumption statement by gaining faster insight, enabling easier comparison among studies, and understanding limitations. For instance, authors claim that tackling confounding factors increases performance, but believing in a direct causal relation may be naive. Reducing confounding factors does not increase performance *per se* but rather improve the experimental design and study requirements by improving assumed brain dynamics (8-4), namely in model design and problem definition. Similarly to confounding factors, aspects as problem definition and system design parameters encounter the same problem.

The second guideline (*G2*) concerns stating the prospective applications envisioned with the designed experiment (8-15). It helps readers and authors understanding what is at stake concerning system parameters, the type of data and envisioned intervention. For instance, most seizure prediction studies report optimal SOP periods for 30-60 minutes. Nevertheless, the RNS[®] system is programmed to make electrical discharges up to 5000 ms [12]. Possibly, for closed-loop systems, these SOP intervals are too long to deliver an effective intervention. Additionally, many authors use short SPH intervals in scalp EEG studies [26]. In these cases, an SPH of 10 seconds or even 1 minute is not enough for taking action after an alarm, such as reaching a secure place or taking rescue medication. For example, diazepam rectal gel (the only FDA drug approved for seizure cluster, and which might be tested as prevention) takes 5-10 minutes to work [27], while oral diazepam or lorazepam takes 15 minutes [28]. This guideline would stimulate discussion regarding study limitations, as well.

The third guideline (*G3*) is related to the use of methodologies that have been clinically approved as a gold standard for comparison. Reporting only

sensitivity, specificity, and prediction above chance-level might be limited, as these metrics strongly depend on data and may not explicitly show progress. Thus, authors should compare their approaches with the ones already clinically approved. This comparison should not only be based on performance but also explainability. The latter leads us to our most important guideline, (G4): researchers should focus on explainability (19) to promote trust among experts. It would be interesting to, at least, present a concrete example of model decisions throughout time. This way, it would demonstrate how a model could explain its predictions to an expert as a data scientist/clinician (application level), and a patient (human-level).

The importance of how explaining decisions

After proper studies comparison, one can ask what a good performance is, or even inquire about the minimum performance that justifies the design of a clinical trial. We believe that a proper methodology is the one which we trust. In literature, trust seems to be represented by literature convergence and reproducibility where studies report high performance (12→13) and comply with consensual study requirements (9- -13). By analysing data from longer recordings and/or a higher number of patient, trust increases as the testing data is more likely to represent real-life conditions [14].

High-level abstraction models may have the potential to handle complex dynamics but require strong efforts towards providing explanations (19- -20). Current clinical knowledge on physiology should be the source of explanations as well as the basis for new findings (19- -4). As an explanation is an exchange of beliefs [29], its acceptance may differ among patients, clinicians, and data scientists.

Although a given methodology, eventually, makes incorrect decisions, we can still trust it if one can explain its decisions (19→13). A great scepticism concerning machine learning and high-level abstraction models may be due to the difficulty in delivering explanations about models' decisions [30]. Although authors and/or clinicians are more willing to trust black-box models when they make correct decisions, wrong ones lead to mistrust because there is no human-comprehensible explanation [13].

The phase IV Neuropace RNS[®] system [12] (NCT00572195) can use up to two independent detections, which are highly configurable and adjusted by the physician, which ensures patient safety. Each detection performs a threshold decision, based on a given extracted feature (line-length, bandpass, and area), by comparing the current window of analysis with another considered to have interictal activity. We believe this is the most simple and explainable strategy we can obtain. One can fully understand the underlying mechanisms behind each decision. The phase I NeuroVista Seizure Advisory System [11] (NCT01043406) is more complex, using a pre-processing step, extracting similar and intuitive features (line-length, Teager-Kaiser energy, and average energy), and training a machine learning model that produced a measure of seizure-risk which concerns a seizure-susceptibility state. This model uses as input the best 16 fea-

tures (from a set of 16 channels X 6 filter/normalization options X 3 analysis methods), and it involved 10 layers (creating different decision surfaces), being inspired in k-nearest neighbours (k-NN) and decision tree classifiers, where each layer considers a different seizure-risk related to its proximity to a seizure event. This algorithm is more complex and not fully transparent. In other words, we do not understand its underlying mechanisms, despite using k-NN and decision tree classifiers (which may be intrinsically interpretable when using a reduced set of features). Calculating seizure risk in a 16-dimension feature space that is furthered divided into 2^{10} partitions (decision surfaces) is not human-comprehensible. Nevertheless, the extracted features are clinically intuitive and the model decision can produce a very human-intuitive output explanation on the obtained seizure risk. It simultaneously compares the current window of analysis with several data distributions whose time proximity to a seizure (and therefore, seizure risk) is considered. By performing multiple data-distribution classifications, it may be more robust to data bias and noise. The authors also ensured patient safety: firstly, they accessed model performance on pre-acquired patient-specific data and secondly, only patients with satisfactory performance received the advisory system.

These two clinical trials demonstrate that, despite all the scientific community efforts to develop complex models and consequent increase in performance, it may be necessary for a fully explainable model to provide trust. Additionally, the Seizure Advisory System clinical trial demonstrates the possibility of using models that are not necessarily intrinsically interpretable, as long as they produce human-comprehensible explanations while ensuring patient safety, handling data bias, and achieving model robustness.

3 Discussion

Despite being useful for clinicians and patients to understand this ecosystem, this study is directed to researchers that develop prediction approaches, so that they have a higher chance of clinical acceptance. Providing a comprehensible overview of all the ecosystem was difficult due to our data science/clinical background. Hence the natural bias/emphasis on academic studies.

We analysed literature regarding seizure prediction that has been published over the last 46 years. In the future, we plan to undergo interviews to provide possible paths and sub-guidelines from the obtained. In the "Questions about the seizure prediction future" section in Supplementary Material, we present a series of questions that arose from describing this ecosystem which we would like to tackle and that deserve our attention.

Our greatest limitation was the patient role, as we did not properly include his/her agency. We strongly believe that we (the academic community) are still far from understanding what is it like to be a patient: the patients' expectations are largely different than the ones from clinicians and data scientists. In the future, we need to be more aware of the active role that a patient can have. The case of Dana Lewis and Hugo Campos are clear examples, where the pa-

tients might be able to track their data, analyze it and therefore, better control their closed-loop systems [31, 32]. Dana Lewis created the “Do-It-Yourself Pancreas System” (#DIYPS), founded the open-source artificial pancreas system movement (#OpenAPS), and advocates patient-centred, -driven, and -designed research. She created #DIYPS to make her continuous glucose monitor (CGM) alarms louder, and developed predictive algorithms to timely forecast necessary actions in the future (<https://diyyps.org/about/dana-lewis/>). Hugo Campos was diagnosed with hypertrophic cardiomyopathy: a disease in which the heart muscle becomes abnormally thick and that can be fatal. He received an implantable defibrillator, which is a device that electrostimulates the heart in case of dangerous arrhythmias. Simply put, after losing his health insurance, he bought a pacemaker programmer on eBay and learned how to use it with a two-week course. Hugo Campos is now a data liberation advocate and leader in the e-patient movement (<https://medicinex.stanford.edu/citizen-campos/>). In fact, article 22 of GDPR 2018 not only provides patients with the right to have an explanation for any algorithm decision but also gives them the right to question those decisions. Please note that we are aware of the complexity of these issue, as we present here an oversimplification of it. We believe that patient accountability and its relation with clinical accountability will be largely discussed in the future.

Despite oriented to seizure prediction, obtained guidelines and relations may be easily translated to different healthcare problems. Other conditions may benefit from a real-life intervention, such as the case of deep brain stimulation in Parkinson’s disease [33]. Computer-aided diagnosis/prognosis software tools face similar problems on ethics, explainability, and trust given the high risk associated with model decisions in healthcare.

About guidelines, *G1* allows improving methodology comparison while delivering a deeper understanding of study limitations to clinicians (regarding assumptions on the underlying physiological mechanisms). For instance, it is interesting to note that, despite most authors with retrospective data use the pre-ictal concept as a point of no return, the two clinically approved studies deal use seizure susceptibility instead.

G2 increases author comprehension on the limitations of signal acquisition methods and patient consequences associated with the obtained specificity. Furthermore, increases in model performance at the cost of developing systems with unreal parameters may be questionable [9, 26]. Although large seizure occurrence windows may translate in higher performance, the interval to accept true alarms is larger. For the case of a warning system, we need to consider the levels of stress and anxiety-induced on patients or the consequences of frequent intake of rescue medication [34, 35, 36]. We also need to understand how/if closed-loops intervention systems can be used with significantly long occurrence periods [12]. We believe that, by considering an increase in performance as one of the primary goals of research, authors develop methodologies that may lack practical application. Although some studies may have a primary goal to increase knowledge on brain dynamics, researchers should clearly state limitations towards real application. Based on this, we encourage authors to study

the consequences for the patients stemming from the development of a given seizure intervention system, through the definition of a maximum number of false alarms.

Concerning legislation and industry standards, we understand these as keepers of best practices on patient safety and trust among all actors. Holistic understanding of trust, explainability, and performance when developing a seizure prediction methodology may be the crucial aspect of this ecosystem. In 2007, Mormann et al. [8] declared that algorithms were still too limited in performance to justify enrolling in clinical trials using responsive stimulation. Despite this paper is one of the most influential in seizure prediction, the first clinical trial (a warning system) [11] started only three years later, in March 2010 and was published in 2013. The first clinical trial using responsive stimulation (phase III RNS[®]System Pivotal Study, NCT00264810) started in 2005, which also led to the phase IV clinical trial (RNS[®]System LTT study, NCT00572195) that started in 2006. Additionally, all current-generation of clinically approved studies and intervention devices use the detection of features alone [13], which demonstrates the importance of explainability. Other examples are present in the literature that arose during discussion, as in 2014, Teixeira et al. [37] tested the Brainatic, which is a real-time scalp EEG-based seizure prediction system, approved by the Clinical Ethical Committee at the Centro Hospitalar e Universitário de Coimbra. It computed 22 univariate features per electrode, and it used non-interpretable models, such as support vector machines, multilayer perceptron and radial basis functions neural networks. Based on this, we concluded that an increased performance cannot be the single criteria for a positive ethics committee decision. This shows that there is room for improvement, possibly by exploring more complex but still explainable systems. For instance, the RNS[®]system might benefit from a more robust approach to capture dynamics before a point of no return [8]. Towards this, more studies, such as the one by Sisterson et al. [38], need to be performed to assess the algorithm effectiveness of responsive neurostimulation. Conclusively, as these methods have been clinically accepted and since a gold-standard comparison method is missing, they should be used as such, both for performance comparison and decision explanation.

Computational power has increased in the past years, which allowed deep learning approaches in several areas. Seizure prediction is no exception [13, 39]. As these approaches, [13] along with rigorous preprocessing [40] have a higher potential to handle brain dynamics, and as intrinsically interpretable models may not be a requirement to undergo a clinical trial, we believe there is an urgent demand for developing explainability methods that work on top of black-box models.

There might be a tendency to argue that by requiring an explanation, the model will be limited in terms of performance (hypothetically 12→19). However, we strongly believe that explanations may enhance the model’s functioning, by tackling the incompleteness of problem formalization. In medical contexts, for example, a correct decision only solves our problem partially [30, 41] and may also be context-dependent, as ethical issues may arise (e.g., choosing between

to save a life and prolong the suffering of a patient). We want to simultaneously deepen brain dynamics understanding, detect data bias, and improve model robustness. It is therefore important to understand possible trade-offs between potentially related aspects, that might not be easily recognized. All of these, when considered in an explanation, improve our understanding, which represents a way to promote patient safety and increases the chance of social acceptance concerning machine learning use [41, 42, 43].

These guidelines and used methodology can be applied to other healthcare settings using computer-assisted diagnosis/prognosis. However, we are aware that guideline *G4* may differ among situation. When predicting hospital mortality after acute coronary events, for example, there are established score models and therefore, using intrinsically interpretable models might be required to better integrate existing clinical knowledge [44]. In the case of seizure prediction, obtaining interpretability can become even harder because i) there is no clinical annotation on the pre-ictal period [8] and ii) the EEG is still far from being fully understood [13, 14]. Therefore, it might be hard to replicate a methodology as there is no standardized protocol to manually identify the pre-ictal period. When discussing case studies with clinicians on the EEG signal, we have observed that they often tend to point to/annotate spikes-and-wave discharges, activity increase, and rapid changes in the signal morphology and associate these to seizure events or seizure susceptibility. We suggest that a possible way "to engage in the clinical discussion", would be by using complex models such as Convolutional Neural Networks to capture complex dynamics, and then by delivering (pointing) to the EEG detected events that were considered for a given decision. This type of explanation could be performed by using, for example, Local Interpretable Model Agnostic Explanations[45], and should be, beforehand, evaluated at the application level of explainability, by discussing these detected events with clinicians. This way, we might try to emulate the process of analysis of the EEG of an epileptic patient typically conducted by a clinician. Additionally, the use of such models may also unravel new patterns (EEG morphologies) that have not yet been associated with epileptic manifestations.

Indeed, we can see our body as a black-box system. In the case of antidepressants, for example, there is still no explanation for the delayed effect of antidepressant drugs and what neurochemical changes reverse the many different symptoms of depression and anxiety [46]. Simply put, we know the inputs (medication) and the outputs (the change in the patients) but we do not fully understand the underlying mechanisms. Nevertheless, these drugs are widely used because they are effective and their risk-benefit balance is favourable. Thus, we believe that the application of Machine Learning and the consequent requirements on interpretability/explainability will depend on the context and the available medical knowledge. For the specific case of seizure prediction, we argue the clinical use of deep learning approaches, as long as researchers put efforts in ensuring patient safety in each stage of each study and clinical trials. As long as researchers can ensure a good risk-benefit balance for the patient (for instance, by providing human-comprehensible explanations) and patients

are willing to volunteer, it may even be unethical to forbid the use of these new methodologies.

As future work, we pretend to tackle the most relevant questions that arose during the previous stage by undergoing interviews with clinicians, data scientists, lawyers, and patients.

4 Methods and Materials

We can divide the used methodology into five stages (see Figure 3). Firstly, we choose studies from the literature that we considered significantly relevant and that addressed seizure prediction models, patients' point of view, legislation, and algorithm explainability.

Secondly, based on the latter, we developed a social network model until reaching saturation or more specifically, until we did not find more actors or relations. Additionally, when actors and relations emerged from referenced studies in the selected papers, we also inspected those to certify that saturation occurred. Thirdly, we refined the social network with our knowledge derived from seizure prediction, where we redefined relations and encapsulated actors. Fourthly, we studied the obtained network and discussed the future of seizure prediction and possible devices for patients. In this discussion, we attempted to list several questions that we found relevant. Finally, we chose four guidelines we found crucial for the faster progress towards new clinically accepted studies.

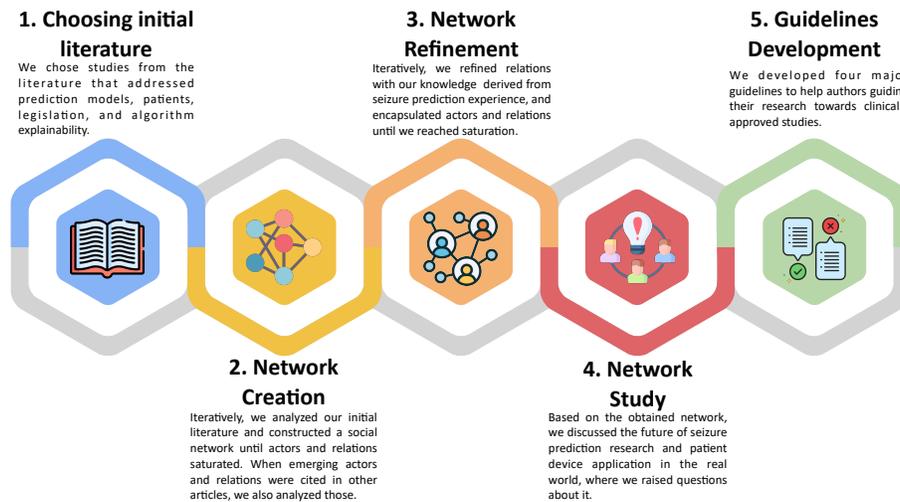


Figure 3: The five-stage methodology followed in this work. Icons obtained from [47, 48, 49, 50, 51].

Choosing initial literature

Our starting materials were the published literature on seizure prediction, as this research field has almost 46 years of existence. We chose three surveys [8, 13, 14] that provided an overall vision of past, present, and future of seizure prediction. These present a critic view of the area. Additionally, we chose a survey [16] on seizure detection and prediction devices, and that is the only article presenting DRE patient’s view on seizure intervention devices [17]. Finally, as we have a biomedical engineering background and machine learning background, we chose a book on interpretable machine learning [30], available online, as we were previously aware of some of the importance of interpretability/explainability. Despite our awareness to the importance of interpretability/explainability importance, note that we did not know, beforehand, which would be the one required (interpretability or explainability) for this specific case. These materials were analysed in the order they are referenced in this paragraph. We consider that this stage (choosing the initial literature) may be the one leading to greater discussion among seizure prediction experts.

Network Creation

We created a social network because it provides a power-model for social structure. The concept of a network here is a set of points (actors, which can be individuals or collective) connected by lines (relations). Our goal was to describe these relations and explain the patterns found. Constructing a network is not a theoretically-neutral task, as it depends on the intellectual judgement of the researcher [22]. To help us structure the network development, we based our literature analysis in GT and ANT. We provide some of the iterations of the developed network, in "Social network iteration and refinement details" section from Supplementary Material.

GT is an inductive process that has contributed to a large acceptance of qualitative methods in several social sciences [18]. Its fundamental premise is that researchers must develop a theory from empirical data. Thus, its overall process consists in the codification of gathered data and identification of emerging themes, and consequent development throughout further data collection [52]. Coded data are commonly short statements or words that capture the meaning of phrases and are used to index data and group ideas.

Additionally, we used concepts of ANT simultaneously with GT. Thus, we did not use GT to search traditional themes, but rather to search for socio-technical actors and their relations. With these, we built a social network. The GT analysis was iterative and performed until reaching saturation. More particularly, it stopped when new actors or relations were not found [18].

Although GT develops theories from rigorous data gathering, the research process requires a certain sensitivity [18, 53]. We stress that the researcher experience heavily influences the data codification and the emergence of themes and ideas. Therefore, main criticism on this theory is the possible introduction of bias, given that truly inductive analysis may not be achievable. We are limited

by prior knowledge and or applications. Due to this, we stress our background experience in developing machine learning pipelines for healthcare, particularly in seizure prediction [18].

ANT is a sociological approach to understand humans and their interaction with technology in specific settings. Its main characteristic is symmetry, which treats equally human and non-human objects [54]. It is a framework based on the following concepts [19, 55]: i) actors, the participants in the network which are human and non-human objects; ii) heterogeneity, each actor importance is given by the web of relations; iii) quasi-objects, the successful outcomes which pass from actor to actor within the network; iv) punctualization, a similar concept to abstraction in object-oriented programming, referred here as encapsulation; v) obligatory passage point, situations that have to occur for all actors to satisfy the interests of the network; and vi) irreversibility, wherein healthcare is not likely to occur due do the importance of developing robust and effective studies to maintain patient safety.

At its heart, ANT tackles the notion of an organisational identity [55]. Thus, we used it to guide our analysis to investigate, understand, and explain the processes that influence and lead to the development of clinically approved studies for seizure prediction [56]. Some criticisms [19] on ANT are that it may be too descriptive. Moreover, it fails in delivering any definitive explanation or approach that best handles the studied actors and relations. Due to this, we applied the social network concept to make it more intuitive. Other limitations of ANT is that it fails to handle human intentions, morals, backgrounds, and previous experiences of human actors. This was one of the reasons why we highlight the importance of explainability. Nevertheless, we are aware that a given explanation will depend on these. Although we did not tackle these directly, rigorous explainability evaluation on the application and human levels might account for them.

Network Refinement

After the social network reached saturation, we encountered a complex structure with many actors and relations. The network could not be delivered in that form, as it was not intuitive. Thus, we decided to refine the network based on our prior seizure prediction experience. This process was also motivated by the mentioned dependence on researcher sensitivity, and punctualization (encapsulation) concept. We believe that our inexperience in social sciences could have derived some of these problems. These could have been overcome differently by experienced researchers in social sciences, as they have a higher understanding of ANT stages [55], such as inscription, translation, and framing.

As previously stated in this paper, we redefined certain relations, such as the ones that concern brain assumptions, confounding factors, performance, and trust. We performed these until reaching saturation. We also grouped the actors in colours concerning themes we found intuitive: signal acquisition and life-related (blue), studies (orange), people and exchanging beliefs (yellow), prospective applications (green), and brain dynamics that trigger seizures and

how to capture its data (red). The final network is depicted in Figure 1.

Network Study

Then, we discussed the network to make it robust and detect possible conflicts, irregularities, and missing actors/relations. Note that ANT investigates the description of the relations, how a network comes to being, and how it temporarily holds. The addition or removal of an actor significantly affects the network. Thus, it may fail when dealing with changes by focusing on a stability situation.

As the seizure prediction experience from authors contributes to this work, we stress that the final outcome might differ among researchers. Others may include different initial articles and or perform differently on data codification, network refinement, and encapsulation. Additionally, it is relevant to remember that the network is permanently evolving as our social reality is always changing and is complex [19].

Due to the particular importance of assumptions statement on brain dynamics, we also discussed them until reaching consensus. Finally, based on the social network, we questioned ourselves on probable paths for seizure prediction future where several questions arose.

Guidelines Development

At last, we agreed on four guidelines that may lead to progress in this area. These were based on the obtained network, its development, and seizure prediction future discussion.

Interactive presentation

In the end, we developed an interactive presentation provided in Supplementary Material. It allows the reader to explore the ecosystem and to better understand the encapsulation of the network. We also present there a simplified version of a seizure prediction product process, from pre-surgical monitoring acquisition until prospective application development. In any part of that process, the product might suffer many iterations, as it is developed for an healthcare setting. Thus, ensuring patient safety is a priority.

Acknowledgement

This work is funded by national funds through the FCT - Foundation for Science and Technology, I.P., within the scope of the project CISUC - UID/CEC/00326/2020 and by European Social Fund, through the Regional Operational Program Centro 2020.

Mauro Pinto gratefully acknowledges the Portuguese funding institution FCT (Foundation for Science and Technology), Human Capital Operational

Program (POCH) and the European Union (EU) for supporting this research work under the PhD grant SFRH/BD/139757/2018.

Author contributions

M.F.P., C.T. and P.M. designed the experiment. M.F.P. built the network and developed the guidelines. A.L., M.F.P., F.L., and A.D. interpreted and discussed the results concerning a machine learning prediction context. J.P. and F.S. interpreted and discussed the results concerning the clinical context. A.D. reviewed substantially the manuscript. M.F.P. wrote the manuscript. All authors reviewed the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

References

- [1] Kenneth D Laxer, Eugen Trinka, Lawrence J Hirsch, Fernando Cendes, John Langfitt, Norman Delanty, Trevor Resnick, and Selim R Benbadis. The Consequences of Refractory Epilepsy and its Treatment. *Epilepsy & Behavior*, 37:59–70, 2014.
- [2] Kirsten M Fiest, Khara M Sauro, Samuel Wiebe, Scott B Patten, Churl-Su Kwon, Jonathan Dykeman, Tamara Pringsheim, Diane L Lorenzetti, and Nathalie Jetté. Prevalence and Incidence of Epilepsy: a Systematic Review and Meta-Analysis of International Studies. *Neurology*, 88(3):296–303, 2017.
- [3] Matthias Ihle, Hinnerk Feldwisch-Drentrup, César A. Teixeira, Adrien Witon, Björn Schelter, Jens Timmer, and Andreas Schulze-Bonhage. EPILEPSIAE - A European Epilepsy Database. *Comput. Methods Programs Biomed.*, 106(3):127–138, 2012.
- [4] Catalina Alvarado-Rojas, Mario Valderrama, A Fouad-Ahmed, H Feldwisch-Drentrup, M Ihle, CA Teixeira, F Sales, A Schulze-Bonhage, C Adam, Antonio Dourado, et al. Slow Modulations of High-Frequency Activity (40-140 Hz) Discriminate Pre-Ictal Changes in Human Focal Epilepsy. *Scientific reports*, 4:4545, 2014.
- [5] Juliane Klatt, Hinnerk Feldwisch-Drentrup, Matthias Ihle, Vincent Navarro, Markus Neufang, Cesar Teixeira, Claude Adam, Mario Valderrama, Catalina Alvarado-Rojas, Adrien Witon, Michel Le Van Quyen,

- Francisco Sales, Antonio Dourado, Jens Timmer, Andreas Schulze-Bonhage, and Bjoern Schelter. The EPILEPSIAE Database: An Extensive Electroencephalography Database of Epilepsy Patients. *Epilepsia*, 53(9):1669–1676, 2012.
- [6] Nathalie Jette and Jerome Engel. Refractory Epilepsy is a Life-Threatening Disease: Lest we Forget, 2016.
- [7] Thomas Cloppenborg, Theodor W May, Ingmar Blümcke, Philip Grewe, Lena J Hopf, Thilo Kalbhenn, Margarete Pfäfflin, Tilman Polster, Reinhard Schulz, Friedrich G Woermann, et al. Trends in Epilepsy Surgery: Stable Surgical Numbers Despite Increasing Presurgical Volumes. *Journal of Neurology, Neurosurgery & Psychiatry*, 87(12):1322–1329, 2016.
- [8] Florian Mormann, Ralph G. Andrzejak, Christian E. Elger, and Klaus Lehnertz. Seizure Prediction: The Long and Winding Road. *Brain*, 130(2):314–333, 2007.
- [9] Kais Gadhomi, Jean Marc Lina, Florian Mormann, and Jean Gotman. Seizure Prediction for Therapeutic Devices: A Review. *Journal of Neuroscience Methods*, 260(029):270–282, 2016.
- [10] Leon D. Iasemidis. Epileptic Seizure Prediction and Control. *IEEE Trans. Biomed. Eng.*, 50(5):549–558, 2003.
- [11] Mark J Cook, Terence J O’Brien, Samuel F Berkovic, Michael Murphy, Andrew Morokoff, Gavin Fabinyi, Wendyl D’Souza, Raju Yerra, John Archer, Lucas Litewka, et al. Prediction of Seizure Likelihood with a Long-Term, Implanted Seizure Advisory System in Patients with Drug-Resistant Epilepsy: a First-in-Man Study. *The Lancet Neurology*, 12(6):563–571, 2013.
- [12] Felice T Sun and Martha J Morrell. The rms system: responsive cortical stimulation for the treatment of refractory partial epilepsy. *Expert review of medical devices*, 11(6):563–572, 2014.
- [13] Dean R Freestone, Philippa J Karoly, and Mark J Cook. A Forward-Looking Review of Seizure Prediction. *Current Opinion in Neurology*, 30(2):167–173, 2017.
- [14] Levin Kuhlmann, Klaus Lehnertz, Mark P Richardson, Björn Schelter, and Hitten P Zaveri. Seizure Prediction — Ready for a New Era. *Nature Reviews Neurology*, 14(10):618–630, 2018.
- [15] Bryce Goodman and Seth Flaxman. European union regulations on algorithmic decision-making and a “right to explanation”. *AI magazine*, 38(3):50–57, 2017.

- [16] Sriram Ramgopal, Sigride Thome-Souza, Michele Jackson, Navah Ester Kadish, Iván Sánchez Fernández, Jacquelyn Klehm, William Bosl, Claus Reinsberger, Steven Schachter, and Tobias Loddenkemper. Seizure Detection, Seizure Prediction, and Closed-Loop Warning Systems in Epilepsy. *Epilepsy & Behavior*, 37:291–307, 2014.
- [17] Andreas Schulze-Bonhage, Francisco Sales, Kathrin Wagner, Rute Teotónio, Astrid Carius, Annette Schelle, and Matthias Ihle. Views of patients with epilepsy on seizure prediction devices. *Epilepsy & behavior*, 18(4):388–396, 2010.
- [18] AL Chapman, M Hadfield, and CJ Chapman. Qualitative research in healthcare: an introduction to grounded theory using thematic analysis. *Journal of the Royal College of Physicians of Edinburgh*, 45(3):201–205, 2015.
- [19] Kathrin M Cresswell, Allison Worth, and Aziz Sheikh. Actor-network theory and its role in understanding the implementation of information technology developments in healthcare. *BMC medical informatics and decision making*, 10(1):67, 2010.
- [20] Richard E Boyatzis. *Transforming qualitative information: Thematic analysis and code development*. sage, 1998.
- [21] Cristina Margarida Chuva Costa. *BIZ2BIS: a socio-technical approach to design inter-organizational business models and their underlying information systems*. PhD thesis, 00500:: Universidade de Coimbra, 2013.
- [22] John Scott. Social network analysis. *Sociology*, 22(1):109–127, 1988.
- [23] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215, 2019.
- [24] Leilani H Gilpin, David Bau, Ben Z Yuan, Ayesha Bajwa, Michael Specter, and Lalana Kagal. Explaining explanations: An overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)*, pages 80–89. IEEE, 2018.
- [25] Jerome Engel. What can we do for people with drug-resistant epilepsy?: the 2016 wartenberg lecture. *Neurology*, 87(23):2483–2489, 2016.
- [26] Elie Bou Assi, Dang K. Nguyen, Sandy Rihana, and Mohamad Sawan. Towards Accurate Prediction of Epileptic Seizures: A Review. *Biomedical Signal Processing and Control*, 34:144–157, 2017.
- [27] Fritz E Dreifuss, N Paul Rosman, James C Cloyd, John M Pellock, Ruben I Kuzniecky, Warren D Lo, Fumisuke Matsuo, Gregory B Sharp, Joan A Conry, Donna C Bergen, et al. A comparison of rectal diazepam gel and placebo for acute repetitive seizures. *New England Journal of Medicine*, 338(26):1869–1875, 1998.

- [28] Epilepsy Foundation. Seizure Rescue Therapies: Oral Rescue Therapies. Epilepsy Foundation Website in <https://www.epilepsy.com/learn/treating-seizures-and-epilepsy/seizure-rescue-therapies>, 2020.
- [29] Tania Lombrozo. The structure and function of explanations. *Trends in cognitive sciences*, 10(10):464–470, 2006.
- [30] Christoph Molnar. *Interpretable Machine Learning*. <https://christophm.github.io/interpretable-ml-book/>, 2019.
- [31] Dana Lewis. History and perspective on diy closed looping. *Journal of diabetes science and technology*, 13(4):790–793, 2019.
- [32] Larry F Chu, Audun Utengen, Bassam Kadry, Sarah E Kucharski, Hugo Campos, Jamia Crockett, Nick Dawson, and Kevin A Clauson. “nothing about us without us”—patient partnership in medical conferences. *BMJ*, 354, 2016.
- [33] Michael S Okun and Kelly D Foote. Parkinson’s disease dbs: what, when, who and why? the time has come to tailor dbs targets. *Expert review of neurotherapeutics*, 10(12):1847–1857, 2010.
- [34] Robert C Tasker. Emergency treatment of acute seizures and status epilepticus. *Archives of disease in childhood*, 79(1):78–83, 1998.
- [35] Marina Gaínza-Lein, Robert Benjamin, Coral Stredny, Marlee McGurl, Kush Kapur, and Tobias Loddenkemper. Rescue medications in epilepsy patients: a family perspective. *Seizure*, 52:188–194, 2017.
- [36] Mark Scheepers, Bruce Scheepers, Michael Clarke, Susan Comish, and Matthew Ibitoye. Is intranasal midazolam an effective rescue medication in adolescents and adults with severe epilepsy? *Seizure*, 9(6):417–421, 2000.
- [37] César Teixeira, Gianpietro Favaro, Bruno Direito, Mojtaba Bandarabadi, Hinnerk Feldwisch-Drentrup, Matthias Ihle, Catalina Alvarado, Michel Le Van Quyen, Bjorn Schelter, Andreas Schulze-Bonhage, et al. Brainatic: A system for real-time epileptic seizure prediction. In *Brain-Computer Interface Research*, pages 7–17. Springer, 2014.
- [38] Nathaniel D Sisterson, Thomas A Wozny, Vasileios Kokkinos, Anto Bagic, Alexandra P Urban, and R Mark Richardson. A rational approach to understanding and evaluating responsive neurostimulation. *Neuroinformatics*, pages 1–11, 2020.
- [39] Ewan Nurse, Benjamin S Mashford, Antonio Jimeno Yepes, Isabell Kiral-Kornek, Stefan Harrer, and Dean R Freestone. Decoding eeg and lfp signals using deep learning: heading truenorth. In *Proceedings of the ACM international conference on computing frontiers*, pages 259–266, 2016.

- [40] Md Shafiqul Islam, Ahmad M El-Hajj, Hussein Alawieh, Zaher Dawy, Nabil Abbas, and Jamil El-Imad. Eeg mobility artifact removal for ambulatory epileptic seizure prediction applications. *Biomedical Signal Processing and Control*, 55:101638, 2020.
- [41] Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.
- [42] Robin Tibor Schirrmester, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggenberger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11):5391–5420, 2017.
- [43] R. Schirrmester, L. Gemein, K. Eggenberger, F. Hutter, and T. Ball. Deep learning with convolutional neural networks for decoding and visualization of eeg pathology. In *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, pages 1–7, 2017.
- [44] Christopher B Granger, Robert J Goldberg, Omar Dabbous, Karen S Pieper, Kim A Eagle, Christopher P Cannon, Frans Van de Werf, Alvaro Avezum, Shaun G Goodman, Marcus D Flather, et al. Predictors of hospital mortality in the global registry of acute coronary events. *Archives of internal medicine*, 163(19):2345–2353, 2003.
- [45] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. ” why should i trust you?” explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.
- [46] Catherine J Harmer, Ronald S Duman, and Philip J Cowen. How do antidepressants work? new perspectives for refining future treatment approaches. *The Lancet Psychiatry*, 4(5):409–418, 2017.
- [47] Freepik. Book Icon made by Freepik perfect from www.flaticon.com. FlaIcon Website in [https://https://www.flaticon.com](https://www.flaticon.com), 2021.
- [48] Freepik. Brainstorming Icon made by Freepik from www.flaticon.com. FlaIcon Website in [https://https://www.flaticon.com](https://www.flaticon.com), 2021.
- [49] Flat Icon. Guidelines Icon made by Flat Icon from www.flaticon.com. FlaIcon Website in [https://https://www.flaticon.com](https://www.flaticon.com), 2021.
- [50] Becris. Becris Icon made by Flat Icon from www.flaticon.com. FlaIcon Website in [https://https://www.flaticon.com](https://www.flaticon.com), 2021.
- [51] Becris. Neural Icon made by Becris from www.flaticon.com. FlaIcon Website in [https://https://www.flaticon.com](https://www.flaticon.com), 2021.

- [52] Kathy Charmaz and Linda Liska Belgrave. Grounded theory. *The Blackwell encyclopedia of sociology*, 2007.
- [53] Bill Davey and Arthur Adamopoulos. Grounded theory and actor-network theory: a case study. *International Journal of Actor-Network Theory and Technological Innovation (IJANTTI)*, 8(1):27–33, 2016.
- [54] Indrit Troshani and Nilmini Wickramasinghe. Tackling complexity in e-health with actor-network theory. In *2014 47th Hawaii International Conference on System Sciences*, pages 2994–3003. IEEE, 2014.
- [55] Nilmini Wickramasinghe, Rajeev K Bali, and Arthur Tatnall. Using actor network theory to understand network centric healthcare operations. *International Journal of Electronic Healthcare*, 3(3):317–328, 2007.
- [56] Tiko Iyamu and Sibulela Mgudlwa. Transformation of healthcare big data through the lens of actor network theory. *International Journal of Healthcare Management*, 11(3):182–192, 2018.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplementarymaterial.pdf](#)
- [interactivepresentation.pptx](#)
- [interactivepresentationpdfversion.pdf](#)