


CRITICAL REVIEW

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

Mauro F. Pinto¹  | Adriana Leal¹ | Fábio Lopes^{1,2} | José Pais³ |
António Dourado¹ | Francisco Sales⁴ | Pedro Martins¹ | César A. Teixeira¹

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

²Department Neurosurgery, Epilepsy Center, Medical Center - University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany

³Hospital CUF Tejo, Lisbon, Portugal

⁴Refractory Epilepsy Reference Centre, Centro Hospitalar e Universitário de Coimbra, EPE, Coimbra, Portugal

Correspondence

Mauro F. Pinto, Department of Informatics Engineering, CISUC, University of Coimbra, Coimbra, Portugal.
Email: mauropinto@dei.uc.pt

Funding information

Fundação para a Ciência e a Tecnologia, Grant/Award Number: 2020.04537. BD, SFRH/BD/139757/2018, SFRH/BD/147862/2019, UID/CEC/00326/2020 and PTDC/EEI-EEE/5788/2020

Abstract

Seizure prediction may be the solution for epileptic patients whose drugs and surgery do not control seizures. Despite 46 years of research, few devices/systems underwent clinical trials and/or are commercialized, where the most recent state-of-the-art approaches, as neural networks models, are not used to their full potential. The latter demonstrates the existence of social barriers to new methodologies due to data bias, patient safety, and legislation compliance. In the form of literature review, we performed a qualitative study to analyze the seizure prediction ecosystem to find these social barriers. With the Grounded Theory, we draw hypotheses from data, while with the Actor-Network Theory we considered that technology shapes social configurations and interests, being fundamental in healthcare. We obtained a social network that describes the ecosystem and propose research guidelines aiming at clinical acceptance. Our most relevant conclusion is the need for model explainability, but not necessarily intrinsically interpretable models, for the case of seizure prediction. 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, if they can deliver human-comprehensible explanations. Due to skepticism and patient safety reasons, many authors advocate the use of transparent models which may limit their performance and potential. Our study highlights a possible path, by using model explainability, on how to overcome these barriers while allowing the use of more computationally robust models.

KEYWORDS

actor network theory, grounded theory, interpretability/explainability, machine learning, seizure prediction

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Epilepsia Open* published by Wiley Periodicals LLC on behalf of International League Against Epilepsy.

1 | INTRODUCTION

Despite antiepileptic 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,¹⁻⁴ as its timely anticipation opens the way to several seizure control strategies, such as: (a) closed-loop systems that trigger drug delivery or electrical brain stimulation; (b) warning devices that inform the patient to prevent accidents (eg, falling from stairs) and/or to self-administer rescue medication.⁵⁻⁷

Although seizure prediction research started in the 1970s through electroencephalogram (EEG) analysis,⁸⁻¹⁰ few predictive devices¹¹ and closed-loop systems¹² 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} 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.¹⁴⁻¹⁷

In the form of review, we inspected the seizure prediction literature to understand the social difficulties, based on the Grounded Theory (GT)¹⁸ and the 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 preexisting 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

We present here a social network²¹ 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

Key points

- This paper aims at providing solutions for researchers to develop new prediction methodologies with higher rates of clinical acceptance
- We built a social network of the seizure prediction ecosystem to obtain an overall view while grasping the existing social barriers
- Our greatest finding is a possible answer to the clinical use of deep learning approaches
- It is possible to develop clinically accepted deep learning approaches, to some extent, if authors deliver human-comprehensible explanations

when they fail, leads researchers to question and mistrust its use, and thus rising skepticism. This is the reason why some authors argue the use of only interpretable models.²²

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.²³ 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 toward 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.

2 | MATERIALS AND METHODS

We can divide the used methodology into five stages (see Figure 1). Firstly, we choose studies from the literature that we considered significantly relevant and that

1. Choosing initial literature

We chose studies from the literature that addressed prediction models, patients, legislation, and algorithm explainability.



3. Network Refinement

Iteratively, we refined relations with our knowledge derived from seizure prediction experience, and encapsulated actors and relations until we reached saturation.



5. Guidelines Development

We developed four major guidelines to help authors guiding their research towards clinically approved studies.



2. Network Creation

Iteratively, we analyzed our initial literature and constructed a social network until actors and relations saturated. When emerging actors and relations were cited in other articles, we also analyzed those.



4. Network Study

Based on the obtained network, we discussed the future of seizure prediction research and patient device application in the real world, where we raised questions about it.



FIGURE 1 The five-stage methodology followed in this work. Icons obtained from Refs [24–28]

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. Other topics and studies also arose in this discussion with all authors and reviewers, which led us to select more papers (please see in File S1 the “Paper Route” section to understand how all papers were selected). Finally, we chose four guidelines we found crucial for the faster progress toward new clinically accepted studies.

2.1 | 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 survey¹⁶ on seizure detection and prediction devices, and one article¹⁷ presenting the Drug-Resistant Epilepsy (DRE) patients' view on seizure intervention devices. Finally, as we have a biomedical engineering background and machine learning background, we chose a book on interpretable machine learning,²⁹ 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 analyzed 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.

2.2 | 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, who 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.²¹ 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 File S1.

Grounded theory 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.³⁰ 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.^{18,31} 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

Actor-Network theory 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 nonhuman objects.³² It is a framework based on the following concepts^{19,33}: (a) actors, the participants in the network which are human and nonhuman objects; (b) heterogeneity, each actor importance is given by the web of relations; (c) quasi-objects, the successful outcomes which pass from actor to actor within the network; (d) punctualization, a similar concept to abstraction in object-oriented programming, referred here as encapsulation; (e) obligatory passage point, situations that have to occur for all actors to satisfy the interests of the network; and (f) irreversibility, wherein healthcare is not likely to occur due to the importance of developing robust and effective studies to maintain patient safety.

At its heart, ANT tackles the notion of an organizational identity.³³ 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.³⁴ Some criticisms¹⁹

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.

2.3 | 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 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 colors 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).

2.4 | 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 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. To be more precise, one researcher performed the initial codification and created the network. Then, the network was presented to all team members separately. Each member discussed the network with the researcher that performed the initial codification. These discussions had as many iterations as necessary until all disagreements were solved. Finally, based on the social network, we questioned ourselves on probable paths for seizure prediction future where several questions arose.

2.5 | 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.

2.6 | Interactive presentation

In the end, we developed an interactive presentation provided in File S2 Interactive Presentation and in File S3 Interactive Light Presentation. 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 presurgical monitoring acquisition until prospective application development. Also, the reader is allowed to interactively explore the whole ecosystem.

3 | RESULTS

We present here a summarized version of the seizure prediction ecosystem, which is shown chronologically, and our proposed guidelines. In File S1, 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.

3.1 | Seizure prediction ecosystem

Figure 2 depicts the obtained social network, which describes the relations between actors. Actors (x) and relations (x - y) are named with numbers and grouped in colors 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 DRE patient (1). Years after diagnosed with DRE, a patient is referred to an epilepsy center to undergo presurgical 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.^{8,35} These data will be stored and constitute retrospective data (7). Most of the databases available to perform academic studies (8) concerns presurgical 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 File S1 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 sections (we include here preprocessing, 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 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, if authors provide efforts on avoiding data bias, ensuring patient safety, and explaining their models' decisions. Furthermore, explanations not only increase trust and mitigate skepticism on artificial intelligence

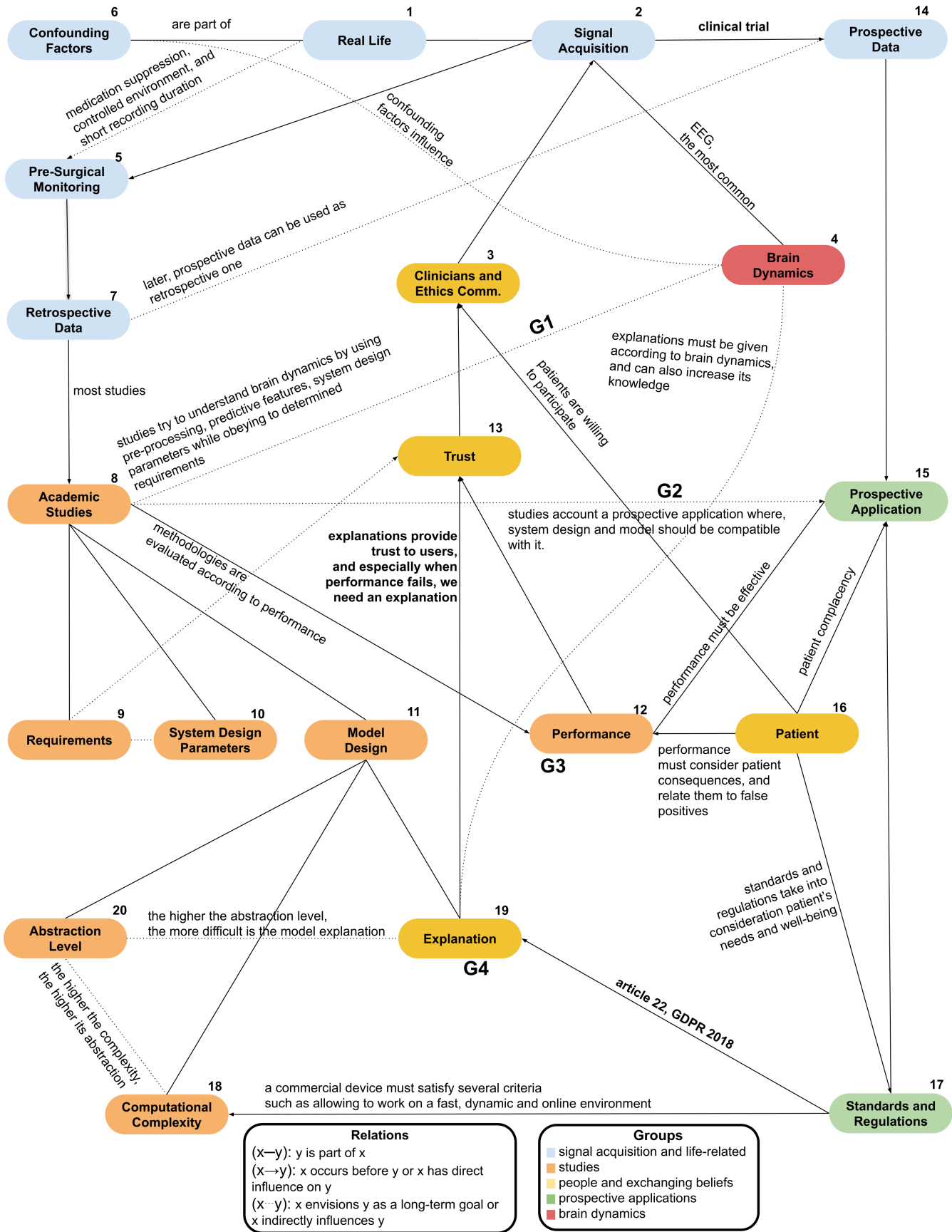


FIGURE 2 The relations between the major actors of epilepsy seizure prediction ecosystem. All actors are numbered to provide an intuition regarding the first steps to consider when developing seizure prediction systems. G1, G2, G3, and G3 are the proposed guidelines

algorithms, but they can also deliver new knowledge on brain dynamics (19→4).

Concerning legislation (17), the 2018 General Data Protection Regulation (GDPR)¹⁵ and the more recent 2021 European Union Medical Devices Regulation (EU MDR)³⁶ also promote the delivery of model explanations (not necessarily intrinsically interpretable models). Article 22 and rule 11 from GDPR and EMD, respectively, are clear examples. 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.

3.2 | Studies guidelines

By describing and discussing all relations, we inferred four guidelines that may help authors in guiding their research on seizure prediction. Figure 3 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 presenting 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

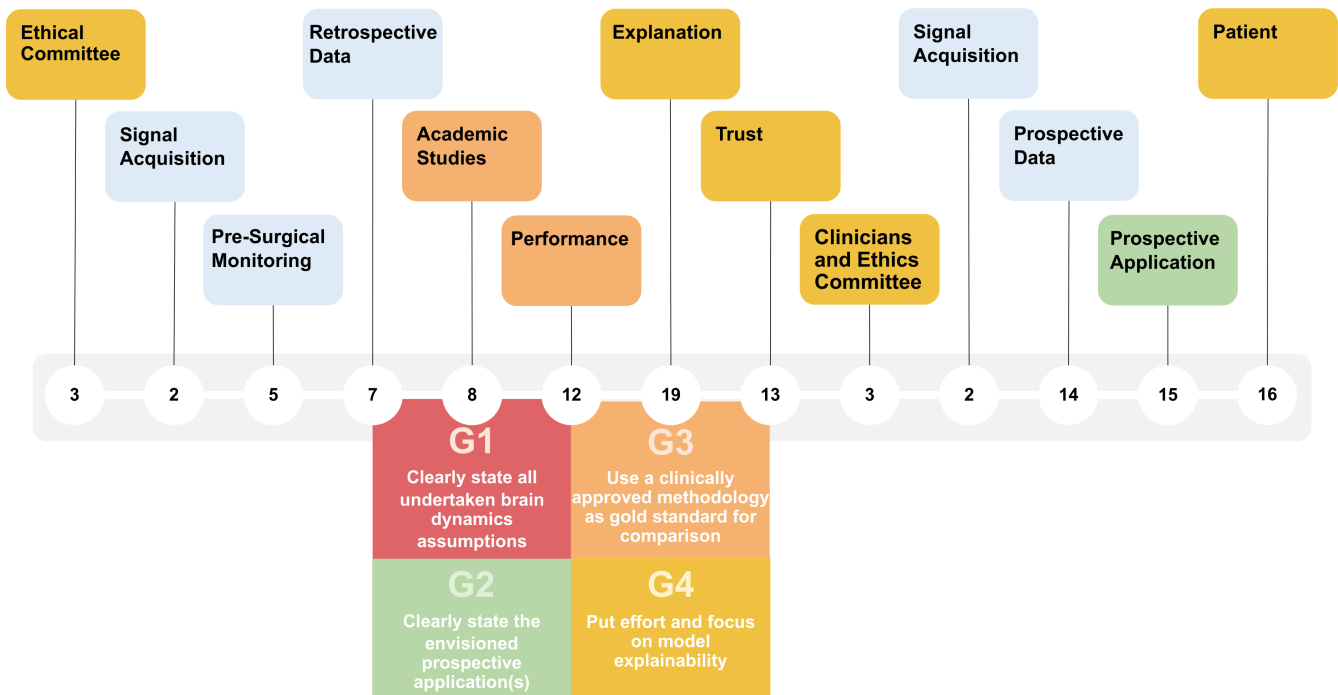


FIGURE 3 A product process for a seizure prediction prospective application, while showing our guidelines concerning designing academic studies

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, such 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 Seizure Occurrence Periods (SOP) periods for 30-60 minutes. Nevertheless, the Responsive Neurostimulation® (RNS®) System is programmed to make electrical discharges up to 5000 ms¹². Possibly, for closed-loop systems, these SOP intervals are too long to deliver an effective intervention. Additionally, many authors use short Seizure Prediction Horizon (SPH) intervals in scalp EEG studies.³⁶ In these cases, an SPH of 10 seconds or even 1 minute is not enough to intervene after an alarm, such as reaching a secure place or taking rescue medication. For example, diazepam rectal gel (the only Food and Drug Administration (FDA) drug approved for seizure cluster, and which might be tested as prevention) takes 5–10 minutes to work,³⁷ while oral diazepam or lorazepam takes 15 minutes.³⁸ 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).

3.3 | 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 analyzing data from longer recordings and/or a higher number of patients, trust increases as the testing data are 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,³⁹ 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 skepticism concerning machine learning and high-level abstraction models may be due to the difficulty in delivering explanations about models' decisions.²⁹ 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® system¹² (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 System¹¹ (NCT01043406) is more complex, using a preprocessing 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 (also known as proictal). This model uses as input the best 16 features (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 neighbors (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 preacquired 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, if they produce human-comprehensible explanations while ensuring patient safety, handling data bias, and achieving model robustness.

4 | 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 ecosystems was difficult due to our data science/clinical background. Hence the natural bias/emphasis on academic studies. Although we previously mentioned our limitations toward qualitative research tools, we stress its importance in the discussion as it constitutes a study limitation.

We analyzed 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 subguidelines from the obtained ones. In the "Questions about the seizure prediction future" section in File S1, 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 patients might be able to track their data, analyze it, and, therefore, better control their closed-loop

systems.^{40,41} Dana Lewis created the "Do-It-Yourself Pancreas System" (#DIYPS), founded the open-source artificial pancreas system movement (#OpenAPS), and advocates patient-centered, -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 electro stimulates 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 relationship 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.⁴² 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 preictal concept as a point of no return, the two clinically approved studies use seizure susceptibility instead, which shows potential for seizure forecasting. Forecasting is different from prediction, as it shifts away from whether a seizure will occur or not and focuses instead on identifying periods of a high probability of seizure occurrence.⁴³ Despite this study's particular emphasis on seizure prediction, we firmly believe that these guidelines and conclusions can be adapted and, thus, hold for seizure forecasting (see "Forecasting Extrapolation" section from File S1).

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.^{9,36} 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.⁴⁴⁻⁴⁶ 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 toward 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. For the warning device case, the literature has pointed to a maximum of 0.15 in FPR/h⁴⁸ and a minimum of 90% sensitivity.¹⁷ For more details and to better understand what an acceptable performance for a clinical setting could be, see the “An acceptable performance for a clinical setting” subsection in File S1.

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 al⁸ declared that algorithms were still too limited in performance to justify enrolling in clinical trials using responsive stimulation. Despite this paper being 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. With this, we claim the following: despite some authors advocating performance limitations to justify clinical trials, these were performed in the past and continue to be. Thus, the idea of a limited performance to justify a clinical trial may be misleading. Clinical trials continue to be performed (as in the case of SeizeIT2, which ended in 2021) because researchers and ethical committees find them necessary, existing a favorable benefit/risk ratio. There is an ongoing necessity to perform clinical trials, especially to avoid publication bias. In the literature, it is easy to find prediction performances that are overestimated as authors, in some cases, only report the best results. When a methodology appears promising, there is the need to test it in different datasets and contexts.

Moreover, 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 Long-term Treatment (LTT) study, NCT00572195) that started in 2006. 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 al⁴⁷ 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 noninterpretable models, such as support vector machines, multilayer perceptron and radial basis function neural networks. Based on this, we concluded that an increased performance is not the single criterion 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 Toward this, more studies, such as the one by Sisterson et al,⁴⁸ 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.^{13,49} As these approaches, along with rigorous preprocessing⁵⁰ 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^{29,51} and may also be context dependent, as ethical issues may arise (eg, 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.⁵¹⁻⁵³

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 situations. 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.⁵⁴ In the case of seizure prediction, obtaining interpretability can become even harder because (a) there is no clinical annotation on the preictal period⁸ and (b) 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 preictal 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,⁵⁵ 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.⁵⁶ 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 favorable. 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, if researchers put efforts in ensuring patient safety in each stage of each study and clinical trials. If 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.

ACKNOWLEDGMENTS

This work is funded by FCT- Foundation for Science and Technology, I.P., within the scope of the projects: CISUC - UID/CEC/00326/2020 with funds from the European Social Fund, through the Regional Operational Program Centro 2020; and project RECoD - PTDC/EEI-EEE/5788/2020 financed with national funds (PIDDAC) via the Portuguese State Budget. 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. Adriana Leal gratefully acknowledges the Portuguese funding institution FCT (Foundation for Science and Technology), the Human Capital Operational Program (POCH), and the European Union (EU) for supporting this research work under PhD grant SFRH/BD/147862/2019. Fábio Lopes gratefully acknowledges the Portuguese funding institution FCT (Foundation for Science and Technology), the Human Capital Operational Program (POCH) and the European Union (EU) for supporting this research work under PhD grant 2020.04537.BD. This research is supported (not financially) by the European Reference Network for rare and complex epilepsies (ERN EPICARE)—Project ID No 769051. ERN EPICARE is partly co-funded by the European Union within the framework of the Third Health Programme "ERN-2016—Framework Partnership Agreement 2017–202. This study has been previously reported in a preprint server. It can be accessed in the following link: <https://www.researchsquare.com/article/rs-886717/v1>.

CONFLICT OF INTEREST

The authors declare no competing financial interests. None of the authors has any conflict of interest to disclose. We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

AUTHOR CONTRIBUTIONS

MFP, CT, and PM designed the experiment. MFP built the network and developed the guidelines. AL, MFP, FL, and AD interpreted and discussed the results concerning a machine learning prediction context. JP and FS interpreted and discussed the results concerning the clinical context. AD reviewed substantially the manuscript. MFP wrote the manuscript. All authors reviewed the manuscript.

ORCID

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

REFERENCES

- Laxer KD, Trinkka E, Hirsch LJ, Cendes F, Langfitt J, Delanty N, et al. The consequences of refractory epilepsy and its treatment. *Epilepsy Behav.* 2014;37:59–70.
- Fiest KM, Sauro KM, Wiebe S, Patten SB, Kwon C-S, Dykeman J, et al. Prevalence and incidence of epilepsy: a systematic review and meta-analysis of international studies. *Neurology.* 2017;88(3):296–303.
- Ihle M, Feldwisch-Drentrup H, Teixeira CA, Witon A, Schelter B, Timmer J, et al. EPILEPSIAE - a European epilepsy database. *Comput Methods Programs Biomed.* 2012;106(3):127–38. <https://doi.org/10.1016/j.cmpb.2010.08.011>
- Alvarado-Rojas C, Valderrama M, Fouad-Ahmed A, Feldwisch-Drentrup H, Ihle M, Teixeira CA, et al. Slow modulations of high-frequency activity (40–140 Hz) discriminate pre-ictal changes in human focal epilepsy. *Sci Rep.* 2014;4:4545.
- Klatt J, Feldwisch-Drentrup H, Ihle M, Navarro V, Neufang M, Teixeira C, et al. The EPILEPSIAE database: an extensive electroencephalography database of epilepsy patients. *Epilepsia.* 2012;53(9):1669–76.
- Jette N, Engel J. Refractory epilepsy is a life-threatening disease: lest we forget. *AAN Enterprises.* 2016;86(21):1932–3.
- Cloppenborg T, May TW, Blümcke I, Grewe P, Hopf LJ, Kalbhenn T, et al. Trends in epilepsy surgery: stable surgical numbers despite increasing presurgical volumes. *J Neurol Neurosurg Psychiatry.* 2016;87(12):1322–9.
- Mormann F, Andrzejak RG, Elger CE, Lehnertz K. Seizure prediction: the long and winding road. *Brain.* 2007;130(2):314–33.
- Gadhoumi K, Lina JM, Mormann F, Gotman J. Seizure prediction for therapeutic devices: a review. *J Neurosci Methods.* 2016;260:270–82. <https://doi.org/10.1016/j.jneumeth.2015.06.010>
- Iasemidis LD. Epileptic seizure prediction and control. *IEEE Trans Biomed Eng.* 2003;50(5):549–58.
- Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, 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. *Lancet Neurol.* 2013;12(6):563–71.
- Sun FT, Morrell MJ. The RNS System: responsive cortical stimulation for the treatment of refractory partial epilepsy. *Expert Rev Med Devices.* 2014;11(6):563–72.
- Freestone DR, Karoly PJ, Cook MJ. A forward-looking review of seizure prediction. *Curr Opin Neurol.* 2017;30(2):167–73.
- Kuhlmann L, Lehnertz K, Richardson MP, Schelter B, Zaveri HP. Seizure prediction — ready for a new era. *Nat Rev Neurol.* 2018;14(10):618–30.
- Goodman B, Flaxman S. European Union regulations on algorithmic decision-making and a “right to explanation”. *AI Magazine.* 2017;38(3):50–7.
- Ramgopal S, Thome-Souza S, Jackson M, Kadish NE, Sánchez Fernández I, Klehm J, et al. Seizure detection, seizure prediction, and closed-loop warning systems in Epilepsy. *Epilepsy Behav.* 2014;37:291–307.
- Schulze-Bonhage A, Sales F, Wagner K, Teotonio R, Carius A, Schelle A, et al. Views of patients with epilepsy on seizure prediction devices. *Epilepsy Behav.* 2010;18(4):388–96.
- Chapman AL, Hadfield M, Chapman CJ. Qualitative research in healthcare: an introduction to grounded theory using thematic analysis. *J R Coll Physicians Edinb.* 2015;45(3):201–5.
- Cresswell KM, Worth A, Sheikh A. Actor-Network Theory and its role in understanding the implementation of information technology developments in healthcare. *BMC Med Inform Decis Mak.* 2010;10(1):67.
- Boyatzis RE. Transforming qualitative information: Thematic analysis and code development. Thousand Oaks, CA: Sage; 1998.
- Scott J. Social network analysis. *Sociology.* 1988;22(1):109–27.
- Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell.* 2019;1(5):206–15.
- Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining explanations: An overview of interpretability of machine learning. In: 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA); 2018. p. 80–9.
- Icon F, Icon F. Flat Icon made by Flat Icon. 2021. Available from: <https://www.flaticon.com>. Accessed in 10 March 2021.
- Becris. Becris Icon made by Flat Icon. 2021. Available from: <https://www.flaticon.com>. Accessed in 10 March 2021.
- Becris. Neural Icon made by Becris. 2021. Available from: <https://www.flaticon.com>. Accessed in 10 March 2021.
- Freepik. Brainstorming Icon made by Freepik. 2021. Available from: <https://www.flaticon.com>. Accessed in 10 March 2021.
- Freepik. Book Icon made by Freepik perfect. 2021. Available from: <https://www.flaticon.com>. Accessed in 10 March 2021.
- Molnar C. Interpretable machine learning. 2019. Available from: <https://christophm.github.io/interpretable-ml-book/>
- Charmaz K, Belgrave LL. Grounded theory. In: Ritzer G, editor. *The blackwell encyclopedia of sociology.* New York, NY: Blackwell Publishing Ltd; 2007. p. 2023–7.
- Davey B, Adamopoulos A. Grounded theory and actor-network theory: a case study. *IJANTTI.* 2016;8(1):27–33.
- Troshani I, Wickramasinghe N. Tackling complexity in e-health with actor-network theory. In: 2014 47th Hawaii International Conference on System Sciences. 2014. p. 2994–3003.
- Wickramasinghe N, Bali RK, Tatnall A. Using actor network theory to understand network centric healthcare operations. *Int J Electron Healthc.* 2007;3(3):317–28.
- Iyamu T, Mguclwa S. Transformation of healthcare big data through the lens of actor network theory. *Int J Healthc Manag.* 2018;11(3):182–92.
- Engel J. What can we do for people with drug-resistant epilepsy?: the 2016 Wartenberg lecture. *Neurology.* 2016;87(23):2483–9.
- Beckers R, Kwade Z, Zanca F. The EU medical device regulation: implications for artificial intelligence-based medical device software in medical physics. *Physica Med.* 2021;83:1–8.
- Bou Assi E, Nguyen DK, Rihana S, Sawan M. Towards accurate prediction of epileptic seizures: a review. *Biomed Signal Process Control.* 2017;34:144–57. <https://doi.org/10.1016/j.bspc.2017.02.001>
- Dreifuss FE, Rosman NP, Cloyd JC, Pellock JM, Kuzniecky RI, Lo WD, et al. A comparison of rectal diazepam gel

- and placebo for acute repetitive seizures. *N Engl J Med.* 1998;338(26):1869–75.
39. Epilepsy Foundation. Accessed in 10 March 2021.
 40. Lombrozo T. The structure and function of explanations. *Trends Cogn Sci.* 2006;10(10):464–70.
 41. Lewis D. History and perspective on DIY closed looping. *J Diabetes Sci Technol.* 2019;13(4):790–3.
 42. Chu LF, Utengen A, Kadry B, Kucharski SE, Campos H, Crockett J, et al. “Nothing about us without us”—patient partnership in medical conferences. *BMJ.* 2016;354:i3883.
 43. Okun MS, Foote KD. Parkinson’s disease DBS: what, when, who and why? The time has come to tailor DBS targets. *Expert Rev Neurother.* 2010;10(12):1847–57.
 44. Dumanis SB, French JA, Bernard C, Worrell GA, Fureman BE. Seizure forecasting from idea to reality. Outcomes of the my seizure gauge epilepsy innovation institute workshop. *eNeuro.* 2017;4(6):ENEURO.0349-17.2017. <https://doi.org/10.1523/ENEURO.0349-17.2017>
 45. Tasker RC. Emergency treatment of acute seizures and status epilepticus. *Arch Dis Child.* 1998;79(1):78–83.
 46. Gaínza-Lein M, Benjamin R, Stredny C, McGurl M, Kapur K, Loddenkemper T. Rescue medications in epilepsy patients: a family perspective. *Seizure.* 2017;52:188–94.
 47. Scheepers M, Scheepers B, Clarke M, Comish S, Ibitoye M. Is intranasal midazolam an effective rescue medication in adolescents and adults with severe epilepsy? *Seizure.* 2000;9(6):417–21.
 48. Winterhalder M, Maiwald T, Voss HU, Aschenbrenner-Scheibe R, Timmer J, Schulze-Bonhage A. The seizure prediction characteristics: a general framework to assess and compare seizure prediction methods. *Epilepsy Behav.* 2003;4(3):318–25.
 49. Teixeira C, Favaro G, Direito B, Bandarabadi M, Feldwisch-Drentrup H, Ihle M, et al. Brainatic: a system for real-time epileptic seizure prediction. In: Guger C, Allison B, Leuthardt EC, editors *Brain-computer interface research*. Berlin: Springer; 2014. p. 7–17.
 50. Sisterson ND, Wozny TA, Kokkinos V, Bagic A, Urban AP, Richardson RM. A rational approach to understanding and evaluating responsive neurostimulation. *Neuroinformatics.* 2020;18(3):365–75.
 51. Nurse E, Mashford BS, Yepes AJ, Kiral-Kornek I, Harrer S, Freestone DR. Decoding EEG and LFP signals using deep learning: heading TrueNorth. In: *Proceedings of the ACM international conference on computing frontiers*. 2016. p. 259–66.
 52. Islam MS, El-Hajj AM, Alawieh H, Dawy Z, Abbas N, El-IJ. EEG mobility artifact removal for ambulatory epileptic seizure prediction applications. *Biomed. Signal Process. Control.* 2020;55:101638.
 53. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. *arXiv.* 2017; <https://doi.org/10.48550/arXiv.1702.08608> [preprint]
 54. Schirrmester RT, Springenberg JT, Fiederer LDJ, Glasstetter M, Eggenberger K, Tangermann M, et al. Deep learning with convolutional neural networks for EEG decoding and visualization. *Hum Brain Mapp.* 2017;38(11):5391–420.
 55. Schirrmester R, Gemein L, Eggenberger K, Hutter F, Ball T. Deep learning with convolutional neural networks for decoding and visualization of EEG pathology. In: *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. 2017. p. 1–7.
 56. Granger CB, Goldberg RJ, Dabbous O, Pieper KS, Eagle KA, Cannon CP, et al. Predictors of hospital mortality in the global registry of acute coronary events. *Arch Intern Med.* 2003;163(19):2345–53.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher’s website.

How to cite this article: Pinto MF, Leal A, Lopes F, et al. On the clinical acceptance of black-box systems for EEG seizure prediction. *Epilepsia Open.* 2022;7:247–259. <https://doi.org/10.1002/epi4.12597>