



Automated algorithms for seizure forecast: a systematic review and meta-analysis

Ana Sofia Carmo^{1,2} · Mariana Abreu^{1,2} · Maria Fortuna Baptista^{3,4} · Miguel de Oliveira Carvalho^{3,4} · Ana Rita Peralta^{3,4} · Ana Fred^{1,2} · Carla Bentes^{3,4} · Hugo Plácido da Silva^{1,2,5}

Received: 20 June 2024 / Revised: 16 August 2024 / Accepted: 18 August 2024 / Published online: 6 September 2024
© The Author(s) 2024

Abstract

This study aims to review the proposed methodologies and reported performances of automated algorithms for seizure forecast. A systematic review was conducted on studies reported up to May 10, 2024. Four databases and registers were searched, and studies were included when they proposed an original algorithm for automatic human epileptic seizure forecast that was patient specific, based on intraindividual cyclic distribution of events and/or surrogate measures of the preictal state and provided an evaluation of the performance. Two meta-analyses were performed, one evaluating area under the ROC curve (AUC) and another Brier Skill Score (BSS). Eighteen studies met the eligibility criteria, totaling 43 included algorithms. A total of 419 patients participated in the studies, and 19442 seizures were reported across studies. Of the analyzed algorithms, 23 were eligible for the meta-analysis with AUC and 12 with BSS. The overall mean AUC was 0.71, which was similar between the studies that relied solely on surrogate measures of the preictal state, on cyclic distributions of events, and on a combination of these. BSS was also similar for the three types of input data, with an overall mean BSS of 0.13. This study provides a characterization of the state of the art in seizure forecast algorithms along with their performances, setting a benchmark for future developments. It identified a considerable lack of standardization across study design and evaluation, leading to the proposal of guidelines for the design of seizure forecast solutions.

Keywords Automated seizure forecast · Seizure likelihood · mHealth · Seizures · Epilepsy · Systematic review

Introduction

People with epilepsy (PWE), as well as their caregivers, live with the uncertainty of when the next seizure might happen, severely constraining some patients' independence and entitlement to normalcy in day-to-day activities. Despite the seeming unpredictability of seizure occurrence,

better-than-chance patient-based prediction of impending seizures has been reported [1, 2], which suggests the existence of preictal dynamics that may be leveraged by automated forecasting algorithms. In fact, PWE and caregivers have reported several non-physiological factors that, from preliminary observation, appear to be linked to seizure occurrence [3]. These include potential seizure triggers (such as alterations in sleep patterns [4]), the environment (such as changes in atmospheric pressure [5]), and cyclicity in seizures [6], which could be used as inputs for a forecast tool [3]. There have also been reports of physiological manifestations that precede seizures [7], including in electrodermal activity (EDA) [8] and electroencephalography (EEG) [9], as well as cyclic patterns in these (and other) biosignals, which appear to be phase locked with seizure occurrence [6, 10]. As such, forecasting of seizure likelihood has been the focus of a joint effort within the epilepsy community since the Epilepsy Foundation recognized its relevance in 2016 [11].

✉ Ana Sofia Carmo
ana.sofia.carmo@tecnico.ulisboa.pt

¹ Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Lisboa, Portugal

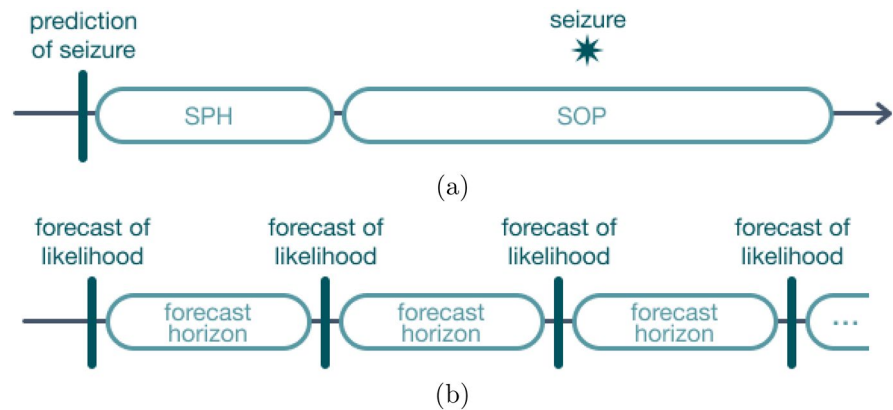
² Instituto de Telecomunicações, Lisboa, Portugal

³ Neurophysiology Monitoring Unit EEG/Sleep Laboratory, Hospital de Santa Maria, Unidade Local de Saúde Santa Maria, Lisboa, Portugal

⁴ Centro de Estudos Egas Moniz. Faculdade de Medicina da Universidade de Lisboa, Lisboa, Portugal

⁵ LUMILS The Lisbon ELLIS Unit | European Laboratory for Learning and Intelligent Systems, Lisboa, Portugal

Fig. 1 Illustration of the concepts of (a) SPH and SOP, where a correct prediction corresponds to an onset that occurs after SPH and within SOP; and (b) forecast horizon, where the forecasts are equally spaced in time in contrast with the task of seizure prediction



Some authors have implied that the forecast of seizures should be based on dynamics observed hours or days before the seizure occurs [12], including physiological alterations in signals like the EEG, EDA, and others, hereinafter called surrogate measures of the preictal state (SMPS). Others authors have gone further and restricted the basis of the forecast to intraindividual cyclic distribution of events (CDE) [7], which comprise cyclicality in the occurrence of seizures or even in certain physiological events, such as interictal epileptiform activity (IEA). Regardless, there is hardly any clear definition of what constitutes seizure forecast, and particularly what distinguishes this practice from seizure prediction. Instead, the major effort from the community toward this distinction has been on shifting the focus from a deterministic perspective (i.e., to raise, or not to raise, an alarm for an impending seizure) to assessing the body states that suggest a higher likelihood of seizure [12, 13].

In the last few years, several automated algorithms that attempt to gauge the likelihood of seizure occurrence in a highly individualized manner have been proposed in literature, and the reported performances are promising. However, no standardized review of the algorithms proposed has been attempted so far.

Moreover, the shift in perspective and problem statement (i.e., from a deterministic problem into a measure of likelihood) imposes additional challenges when it comes to evaluating and comparing forecast performances. Due to its inherent probabilistic nature, we may encounter situations in which the brain enters a state of high seizure likelihood without this actually translating into the occurrence of a seizure event [14, 15], challenging the traditional approaches to performance evaluation. Much like the task of seizure prediction lacked performance standards at the start of the 2000s [16], so does seizure forecast now.

In this paper, we review reported performance of automated algorithms that attempt to provide an individual's likelihood of having a seizure within a given time window, based on intraindividual cyclic distribution of events and/or surrogate measures of the preictal state. This includes

methods which aim to forecast seizure likelihood and that leverage seizure triggers, the environment, cyclicality in seizures, patterns in physiological data, or a combination of these. This review will address the following research questions:

RQ1: What is the current state of automated seizure forecast?

RQ2: Which data are the most relevant for the forecast of seizure likelihood?

RQ3: Which approaches and metrics are most often used to assess forecast performance?

Materials and methods

This review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [17]. Details of the protocol for this systematic review were registered on International Prospective Register of Systematic Reviews (PROSPERO), under PROSPERO ID CRD42023478920.

Concept definition

Forecast horizon

Seizure prediction and forecasting are often (incorrectly) used interchangeably. In seizure prediction, the main objective is to raise an alarm prior to a seizure occurring, providing the patient (or caregiver) the opportunity to act in accordance, either by taking fast-acting anti-seizure medication (ASM) or adopting protective measures. When developing seizure prediction algorithms, the concepts of seizure prediction horizon (SPH) and seizure occurrence period (SOP) arise. The corresponding definitions were initially proposed by Maiwald et al. [18] (as illustrated in Fig. 1a) and have since been adopted in several papers addressing seizure prediction and sometimes extended to forecasting

[19–21]. SPH is defined as the period of anticipation of a seizure event, i.e., the time interval in which a seizure should not (yet) occur after an alarm is raised, while SOP is the interval of time (after SPH) within which the seizure onset is expected to arise.

Seizure forecast, on the other hand, attempts to provide the patient (or caregiver) with an indicator of seizure likelihood for a specific time period after the forecast, called the forecast horizon (Fig. 1b). This concept denotes the time interval in the future for which the forecast is generated [22]. As such, the forecast horizon is not a direct counterpart to the concepts of SPH and SOP.

Retrospective vs pseudo-prospective approach

The train–test approach consists of the methodology used for model evaluation and can be either retrospective or (pseudo-)prospective.

In a retrospective approach, the data is randomly divided into training and test sets without considering the temporal relationship of the data points.

Conversely, in a prospective approach, the model is trained only on past data and evaluated in real time on future data. Similarly, in a pseudo-prospective approach, the prospective evaluation is emulated by using retrospective data, but dividing it into training and test sets so that the model is trained only on past data and evaluated in data posterior to that, thus also respecting time dependencies.

Deterministic vs probabilistic forecast

In seizure forecast, we encounter a view that challenges the traditional approaches to performance evaluation. Most commonly, when assessing the performance of a prediction of seizure, the concepts of SPH and SOP are leveraged, where a correct prediction corresponds to an onset that occurs after SPH and within SOP. However, when working with forecasts, instead of having an alarm raised, a likelihood is received at equally spaced intervals of time.

To handle this issue, some studies opt for converting the forecasts into binary high/low risks (using discriminative thresholds), thus allowing the use of traditional (i.e., deterministic) performance measures. A number of papers adopt metrics such as sensitivity (Sen), accuracy (Acc), and area under the ROC curve (AUC). However, these methodologies are highly dependent on the chosen SPH and SOP. In fact, by replacing the categorical notion of seizure occurrence (i.e., it either occurred when expected, or it did not) with a likelihood, we may encounter situations in which the brain enters a state of high seizure likelihood without this actually translating into the occurrence of a seizure event [15, 23]. Therefore, the traditional measures of performance are usually replaced or complemented with alternatives that allow

to properly assess the clinical utility of the forecasts, such as Brier Score (BS) and Brier Skill Score (BSS).

Systematic review

A systematic literature search was carried out on the databases IEEE Xplore, Scopus, and PubMed, as well as on the Web of Science Core Collection. The search was conducted using the query ((automated OR automatic OR algorithm OR machine learning OR deep learning OR artificial intelligence) AND (forecast OR risk OR likelihood OR prediction OR cyclic* OR rhythm*)) AND (epilepsy OR seizure)) on the title, abstract, and keywords of the studies. A first search was conducted up to November 2, 2023. Another search was conducted on May 10, 2024.

From the identified records, built-in automation tools of the search engines were used to filter the results according to language (resulting in the exclusion of all records that were not written in English, Portuguese, Spanish, or French), as well as according to document type (resulting in the exclusion of all reviews). Then, the corresponding references were uploaded into the Rayyan software,¹ in which duplicates were automatically identified. Duplicates were resolved by the Rayyan built-in automation tool for the cases where similarity was equal to or larger than 90%; the remaining duplicates were resolved manually by one of the reviewers (A.S.C.). The titles and abstracts were screened by three independent reviewers (A.S.C., M.F.B. and M.O.C.), resulting in the exclusion of the records that were not considered relevant. Conflicts were resolved in an open discussion between all reviewers. The inclusion criteria were:

- (1) Proposal of an original real-time forecast algorithm for automatic human epileptic seizure likelihood assessment, that is patient-specific, based on intraindividual cyclic distribution of events and/or cyclic distribution of events.
- (2) Any cohort dimension.
- (3) Records written in English, Portuguese, Spanish, or French.
- (4) Published as an article, proceeding paper, or abstract.

After screening, the full-text documents were retrieved and assessed for eligibility by two independent reviewers (A.S.C. and M.A.). Non-trivial cases were decided in an open discussion between the two reviewers. Studies were excluded if:

¹ Rayyan software is available at <https://www.rayyan.ai/>.

- (1) Separate sets of data were not used for training and testing of the forecast algorithm.
- (2) The proposed algorithm did not provide a measure of seizure likelihood/risk.
- (3) Failed to report at least one of the following data: methodology proposed for forecast of seizure likelihood; data considered for the algorithm; forecasting horizon; methodology and results on the evaluation of seizure forecast performance.

Finally, A.S.C extracted the data, which included:

- (1) Input data: source, type, description.
- (2) Cohort: dimension, median/total duration of recording, median/total number of seizures.
- (3) Methodology: type of algorithmic approach, forecast horizon, train–test approach.
- (4) Results: metrics used to evaluate performance, reported performance.

Meta-analysis

A meta-analysis was performed to quantify the current state of seizure forecast. Studies were stratified according to the type of input data given to the algorithms, which can be categorized as SMPS, CDE, or a combination of these, following the distinction provided in [7].

The two most highly reported deterministic and probabilistic metrics were chosen for the meta-analysis. As such, random-effects meta-analysis with the restricted maximum likelihood (REML) method was used to obtain pooled estimates of AUC and BSS. The AUC, which ranges from 0 to 1 (perfect score), can measure the forecast performance based either on true positive rate versus false positive rate or sensitivity versus portion of time in false warning. It is often used in seizure prediction literature as the main performance outcome, including in previous seizure prediction/forecast challenges [24, 25]. BSS, on the other hand, is a probability score that provides a measure of improvement over a naive forecast, ranging from $-\infty$ to 1, where 0 corresponds to a performance equivalent to the naive forecast.

A weighting scheme was also employed, where the variance of each study was weighted by its sample size. This approach ensured that studies with larger sample sizes contributed more to the overall analysis.

Finally, sources of heterogeneity were also explored through subgroup analysis for each potential moderator variable. For the eligible cases² the moderator effect of

the variables on heterogeneity was quantified by including them as covariates in the meta-analysis model, and evaluated according to changes in I^2 . Five sources of heterogeneity were considered apart from the type of input data: source of data (i.e., dataset used), input data (EEG vs heart rate vs seizure times vs other inputs), forecast horizon (<1 h, 1 h, 24 h), train/test approach (retrospective vs prospective/pseudo-prospective), and study (i.e., algorithms proposed by the same study).

Statistical analysis was performed using the `metafor` R library and, when applicable, statistical significance was set to 0.05.

Results

A total of 8249 records were identified, from which 945 were removed according to publication type and language (see Fig. 2 for the PRISMA flow diagram). Additionally, 3160 duplicates were removed. Of the remaining, 4144 records were screened for relevance to the review topic and 549 of these were sought for retrieval of the full records. Out of the 549 records, 32 could not be retrieved and a further 502 studies were excluded for not meeting at least one of the eligibility criteria. Finally, three additional reports were included, one as a result of the second search and the other two were manually added. In total, 18 studies were included in the systematic review.

Study characteristics

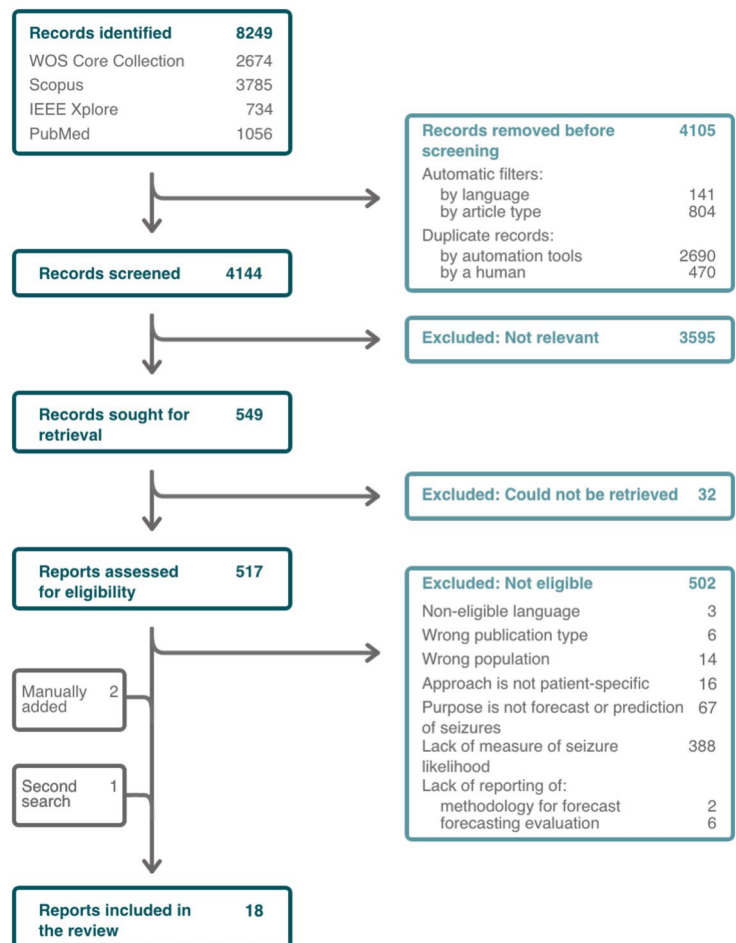
The results for the 18 studies are summarized in Tab. 1. A total of 419 patients participated in the studies (median of 12, interquartile range (IQR) of 8.25–17.25), and a total number of 19442 seizures were reported across studies³. The duration of the total/recorded data varied greatly across studies, and even within the same sample, with some studies encompassing less than 1 week [28] and others up to 9 years [22].

There was little consistency regarding the source of data used across the body of studies, with 5 studies relying on self-collected data when evaluating their proposed algorithms and the remaining 12 using previously documented datasets. The NeuroVista trial [29] was the most commonly reported (6, one of them in combination with self-collected data), consisting of 15 patients with focal seizures submitted to ambulatory intracranial EEG monitoring for more than 80 days. Two studies reported using data from the NeuroPace

² Following the recommendation of Fu et al. [26], only the cases in which all subgroups had at least four entries were considered eligible for quantification of moderator effect on heterogeneity.

³ The number of reported seizures does not account for the seizures reported in [27], since they only reported the median number of seizures (143, IQR of 13–1233).

Fig. 2 PRISMA flow diagram depicting the number of records identified, included and excluded, as well as the reasons for exclusion. Four databases and registers were searched, resulting in 15 included reports, with 3 additional reports being included at a later stage. A total of 18 reports were included in the review



trial [30], where an implanted brain-responsive neurostimulator was validated in patients with disabling partial or generalized tonic–clonic seizures, for a mean follow-up period of 5.4 years. Data from the 24/7 EEG SubQ trial [31] was used in three studies (one of them in combination with data from the NeuroPace trial), consisting of subcutaneous EEG recordings of nine participants with temporal lobe epilepsy, for up to 3 months. Two studies relied on data from the EPILEPSIAE database [32], which was created under a joint European project and is the largest collection of hospital EEG recordings, with data from more than 250 patients. Finally, a single study used data from the Minder sub-scalp system trial (ACTRN 12619001587190), where six patients were implanted with a sub-scalp EEG monitoring system. Moreover, the majority (14/18) used data collected with mobile technology, which included subcutaneous and intracranial EEG implants, smartwatches and other wrist-based devices, as well as mobile applications.

Regarding input data, seven studies used only SMPS, six reported using solely CDE, four proposed algorithms that use both types of input data, and one study proposed algorithms using both types separately. Figure 3 depicts

the use of the different data used as input in the 18 papers, as well as how the data was given to the algorithmic approaches, when applicable. EEG was the most widely used physiological signal (13/18), either as a raw input (5/13), through features derived from it (4/13), or through cyclic patterns extracted from those features (3/13). Photoplethysmography (PPG) and electrocardiography (ECG) followed as sources of heart rate (HR) (5/18). Similarly to EEG, HR was not only explored as a surrogate measure of the preictal state, but also as a source of CDE (2/5). Apart from the EEG, seizure times were the most commonly used input data, being used in 12 out of the 18 studies, and equally explored as a raw input or through cyclic profiles derived from them. Other, less commonly reported data inputs included other peripheral physiological signals, the time of day (ToD), and behavioral/external conditions.

Finally, despite the probabilistic framework of the task of seizure forecast, less than half the studies (8/18) reported probabilistic evaluation metrics, while all studies reported deterministic ones. AUC was the most reported metric (15/18), followed by Sen and time in warning

Table 1 Summary of characteristics of the studies included in the systematic review

First author, year	Data source	Sample size	Duration in days (median)	# Seizures (median)	Type of input data	Algorithmic approach	Reported metrics
Attia, 2021 [33]	24/7 EEG SubQ trial	1	230	22	SMPS	LSTM	AUC, Sen, % Sig., FPR
Chen, 2022 [34]	NeuroVista trial	15	557	151	CDE	Phase modeling	AUC, Sen, TiW, % Sig., GSS
Cook, 2013 [29]	NeuroVista trial	11	265	31	SMPS	k-NN/decision tree type classifier	Sen, TiW, Likelihood ratio
Costa, 2024 [35]	EPILEPSIAE database	40	5	5	SMPS	SVM, LR, SNN	BSS, Sen, TiW, FPR, BS
Cousyn, 2022 [36]	Self-collected	10	Mean: 10.7	2	SMPS	SVM	AUC, Acc, F1-score, BSS, BS
Cousyn, 2023 [28]	Self-collected	15	11	Mean: 25	SMPS	SVM	AUC, Acc, Sen, Spe, F1-score, BS
Karoly, 2017 [37]	NeuroVista trial	9	459	102	SMPS; CDE; both	LR; phase modeling; ensemble through LR weight updating	AUC, BSS, Sen, TiW
Karoly, 2020 [38]	Self-collected and NeuroVista trial	50	Mean: 336	Mean: 109	CDE	Phase modeling	AUC, Acc, TiW, % Sig.
Leguia, 2022 [27]	NeuroPace trial and 24/7 EEG SubQ trial	161	1722; 85	143	SMPS; CDE	GLM	AUC, BSS, % Sig.
Maturana, 2020 [39]	NeuroVista trial	14	512	151	CDE	Phase modeling	Sen, TiW
Nasseri, 2021 [40]	Self-collected	6	242	16	SMPS	LSTM	AUC, Sen, TiW, % Sig
Payne, 2020 [20]	NeuroVista trial	8	Mean: 548	157	CDE	Phase modeling; ensemble through naive Bayes	AUC, % Sig.
Proix, 2021 [22]	NeuroPace trial	18	Mean: 1484	Mean: 43	SMPS; CDE; both	GLM	AUC, BSS, % Sig.
Stirling, 2021 [41]	self-collected	11	435	94	both	LSTM + RF (ensemble through LR)	AUC, TiW, % Sig., BS
Stirling, 2021 [42]	Minder sub-scalp system trial	1	183	134	CDE	RF + LR (ensemble through sequential input)	AUC, Sen, TiW
Truong, 2021 [21]	EPILEPSIAE database	30	4	Mean: 9	SMPS; both	Bayesian CNN	AUC
Viana, 2022 [43]	24/7 EEG SubQ trial	6	80	17	SMPS	LSTM	AUC, Sen, TiW, % Sig., FPR
Xiong, 2023 [44]	self-collected	13	495	71	CDE	Phase modeling	AUC, BSS, Sen, TiW, % Sig

Duration and seizure number are given as median, unless stated otherwise. SMPS: surrogate measures of the preictal state; CDE: cyclic distribution of events. CNN: convolutional neural network; GLM: generalized linear model; k-NN: k-nearest neighbors; LR: logistic regression; LSTM: long short-term memory network; SNN: shallow neural network; RF: random forest; SVM: support vector machine. AUC: area under the ROC curve; BS: Brier Score; BSS: Brier Skill Score; FPR: false positive rate; GSS: geometric mean of sensitivity and specificity; Sen: sensitivity; Spe: specificity % Sig: percentage of patients with significant forecasts; TiW: time in warning

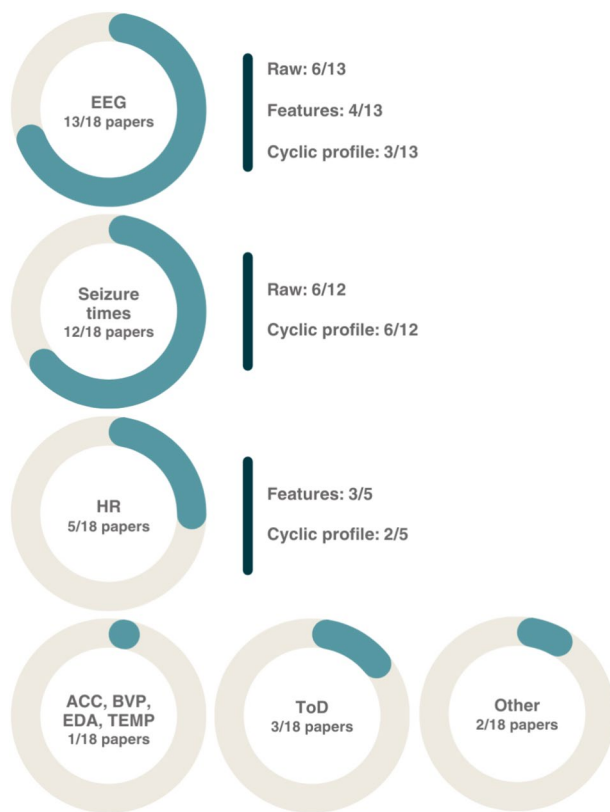


Fig. 3 Summary of reported input data across the 18 papers included in the review, along with how the data were given to the algorithms, when applicable (i.e., raw vs features vs cyclic profile). *ACC* accelerometry, *BVP* blood volume pulse, *EDA* electrodermal activity, *EEG* electroencephalography, *HR* heart rate, *TEMP* temperature, *ToD* time of day, *Other* sleep, activity, weather

(TiW) (11/18), and % of patients with significant forecasts (10/18). BS and BSS were reported in eight and six studies, respectively⁴. Moreover, when using both types of train–test approaches, i.e., retrospective vs (pseudo-) prospective, most studies (5/7) reported a different set of metrics for each.

Descriptive analysis

Table 2 summarizes the characteristics of all considered algorithms. Some studies comprise multiple table entries, corresponding to algorithms whose characteristics differ significantly with regard to the following attributes: input data, forecast horizon, and train–test approach. For each of these unique characteristics' configurations, only the best performance is reported. AUC and BSS scores are presented,

⁴ Although BS was reported in more studies than BSS, the latter was used as a performance metric in a larger number of proposed approaches.

whenever reported in the studies. An extension of this table, including all reported performance metrics, is available in Supplementary file 1.

The 18 studies proposed a total of 43 algorithms with unique configurations. The most common type of input data was SMPS (19/43), which included the time series of the physiological signals, features extracted from them, times when seizures occurred, and ToD. Seventeen algorithms relied on CDE, which were either based on times of seizure occurrence, IEA, HR, sleep, or weather features. The remaining algorithms (7/43) used a combination of the previous inputs, as well as features extracted from patient's activity. Figure 4 provides a summary of the types of input data reported across the 43 algorithms, along with the category of algorithmic approach used.

Remarkably, most algorithms relied on methodologies for phase modeling when leveraging CDE (13/17 and 1/7), which implies the identification of significant cycles or the use of histograms of the times of previous events. The remaining opted for the use of generalized linear models (GLMs) (3/17 and 1/7), deep learning (DL) models such as convolutional neural networks (CNNs) (1/7), or more classic machine Learning (ML) models such as random forest (RF) and logistic regression (LR) (1/17 and 4/7). On the other hand, algorithms using SMPS relied most commonly on classic ML algorithms (9/19 and 5/7), including support vector machines (SVMs) and LR, while the remaining opted for GLMs (4/19 and 1/7), or long short-term memory networks (LSTMs) and CNNs (6/19 and 5/7).

It was not uncommon for studies to evaluate their algorithms through a dual approach of retrospective and (pseudo-)prospective train–test schemes (7/18). In such cases, the retrospective period was used for development of the algorithm, often in a cross-validation scheme, providing an estimate of algorithm performance. Then, the (pseudo-) prospective period was used to simulate a real-world scenario, often being iteratively updated with the “new” data, in which the algorithm was validated in a setting analogous to the one it was designed to function.

Meta-analysis

Out of the 43 entries, 23 were eligible for the meta-analysis with AUC and 12 with BSS.⁵ The results of the meta-analysis are summarized in Fig. 5 and Fig. 6 for AUC and BSS, respectively. The overall mean AUC was 0.71 (95% confidence interval (CI)=0.68–0.75, $I^2=97.29%$), which was further stratified according to the type of input data given

⁵ Algorithms were considered eligible for the meta-analysis if both the mean and standard deviation (SD) across patients was reported, and if the sample size comprised more than one patient.

Table 2 Summary of characteristics of the studies included in the systematic review

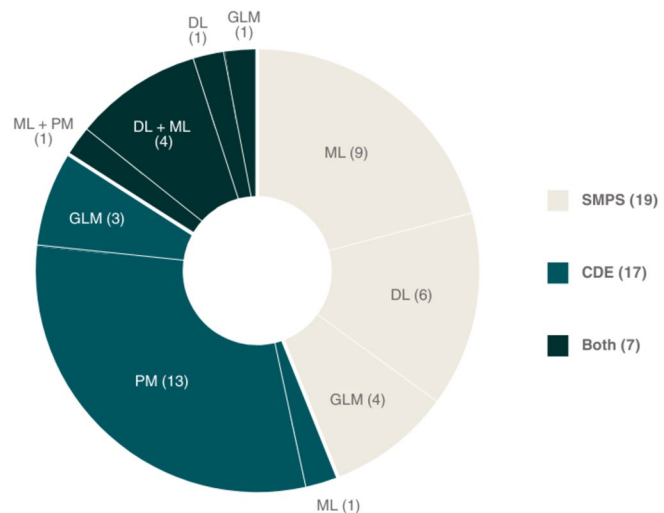
First author, year	Sample size	Total # seizures	Type of input data	Input data	Horizon	Approach	AUC	TIW	% Sig	BSS	BS
Attia, 2021 [33]	1	15	SMPS	EEG	1 h	R	0.75		100%		
Chen, 2022 [34]	15	2398	CIDE	EEG, ToD			0.81		100%		
	12	1976		EEG cyclic profile	5min	R	0.91±0.08	L:80%	100%		
Cook, 2013 [29]	11	154	SMPS	EEG features	5min	P	0.75±0.10	L:80%	92%		
	10	323				R		H:27.8% L:42.2%			
Costa, 2024 [35]	40	224	SMPS	EEG	10min	P		H:23% L:45%		0.01±0.15	0.18±0.1
Cousyn, 2022 [36]	10	38	SMPS	EEG features	24 h	R	0.79				
	15	47	SMPS	HR features	24 h	R	0.79±0.17			0.72±0.22	0.13±0.11
Karoly, 2017 [37]	14					P					
	9	1383	SMPS	EEG features	30min	R	0.79±0.08				0.3 [0.18;0.48]
Karoly, 2020 [38]	50	5450	CIDE	Seizure cyclic profile		P		H:25.3%		0.05±0.06	
	161	(*)	Both	Seizure cyclic profile	5min	P	0.85±0.05	H:24.4%		0.05±0.03	
Leguia, 2022 [27]	14		SMPS	Seizure times	24 h	R	0.63±0.08	H:24.7%		0.11±0.07	
	14	2871	CIDE	EEG features			0.65±0.05	H:14.8% L:67.1%			
Maturana, 2020 [39]	14		CIDE	EEG cyclic profile	2min	P	0.69±0.06			0.08±0.10	
	6	278	SMPS	EEG cyclic profile			H:7.75%			0.08±0.08	
Nasser, 2021 [40]	6			ACC, BVP, EDA, TEMP, HR, ToD, signal quality metrics	1 h	R	0.75±0.15		83%		
	8	1236	CIDE	Sleep features cyclic profile	10min	P	0.63±0.11				
Payne, 2020 [20]	18	767	SMPS	Weather features cyclic profile			0.55±0.10		62%		
	18		CIDE	Seizure cyclic profile			0.69±0.08		62%		
Proix, 2021 [22]	18			Seizure cyclic profile			0.68±0.11		75%		
	18	767	SMPS	EEG features, seizure times	24 h	P	0.61 [0.59,0.64]		75%	0.02 [0.01,0.04]	
Proix, 2021 [22]	18			Seizure and EEG cyclic profiles	1 h		0.60 [0.57,0.65]		11%	0.01 [0.00,0.01]	
	18		CIDE	Seizure and EEG cyclic profiles	24 h		0.73 [0.65,0.76]		67%	0.01 [0.00,0.01]	
Proix, 2021 [22]	18			Seizure and EEG cyclic profiles	1 h		0.70 [0.64,0.77]		83%	0.17 [0.11,0.26]	
	18			Seizure and EEG cyclic profiles	1 h		0.70 [0.64,0.77]		94%	0.02 [0.01,0.02]	

Table 2 (continued)

First author, year	Sample size	Total # seizures	Type of input data	Input data	Horizon	Approach	AUC	TiW	% Sig	BSS	BS
Stirling, 2021 [41]	11	1493	Both	ensemble	1 h		0.75 [0.69,0.81]		100%	0.03 [0.02,0.06]	
	8	1078	Both	HR features, sleep features, activity features, HR cyclic profile, seizure times	24 h	R	0.66±0.11	median L:18%	91%		21.91±0.72
	11	1493			1 h	R	0.59±0.16	median H:14%	50%		0.17±0.05
Stirling, 2021 [42]	8	1078			1 h	P	0.74±0.10		100%		
	1	134	CDE	Seizure and EEG cyclic profiles	1 h	P	0.65±0.18	H:26% L:63%	88%		
Truong, 2021 [21]	30	261	SMPS	EEG	30min	R	0.69				
			Both	EEG and seizure cyclic profile			0.69				
Viana, 2022 [43]	6	103	SMPS	EEG	1 h	R	0.65±0.17	H:33.3%	67%		
				EEG, ToD			0.73±0.07	H:31.47%	83%		
Xiong, 2023 [44]	13	2247	CDE	Seizure and HR cyclic profiles	24 h	R	0.70±0.15		85%	0.20±0.23	
	6	2514			1 h	P	0.74±0.17		83%	0.17±0.17	
	13	2247			1 h	R	0.71±0.12	H:27%	69%	0.04±0.03	
	6	2514			1 h	P	0.76±0.07	H:18%	67%	0.05±0.05	

Duration and seizure number are given as median, unless stated otherwise. *SMPS* Surrogate measures of the preictal state, *CDE* cyclic distribution of events, *CNN* convolutional neural network, *GLM* generalized linear model; *k-NN* k-nearest neighbors; *LR* logistic regression, *LSTM* long short-term memory network, *SNN* shallow neural network, *RF* random forest, *SVM* support vector machine, *AUC* area under the ROC curve, *BS* Brier Score, *BSS* Brier Skill Score, *FPR* false positive rate, *GSS* geometric mean of sensitivity and specificity, *Sen* sensitivity, *Spe* specificity, % *Sig* percentage of patients with significant forecasts, *TiW* time in warning

Fig. 4 Summary of types of input data reported across the 43 algorithms, along with the category of algorithmic approach used. *SMPS* Surrogate measures of the preictal state, *CDE* cyclic distribution of events, *ML* machine learning, *DL* deep learning, *GLM* Generalized linear model, *PM* phase modeling



to the algorithms. The subtotal AUC for algorithms using only SMPS was 0.71 (CI=0.65–0.77, $I^2=96.82\%$), for algorithms relying solely on CDE was 0.73 (CI=0.67–0.78, $I^2=96.97\%$), and for algorithms that used both types of input data was 0.67 (CI=0.61–0.73, $I^2=55.70\%$).

The overall mean BSS was 0.13 (CI=0.03–0.23, $I^2=99.50\%$). The subtotal BSS for algorithms using only SMPS was 0.18 (CI=–0.07–0.43, $I^2=99.82\%$) and for algorithms relying solely on CDE was 0.07 (CI=0.04–0.11, $I^2=87.37\%$). A single algorithm using both types of input data reported the BSS score at 0.11 (CI=0.07–0.15).

When considering the five potential sources of heterogeneity (source of data, input data, forecast horizon, train/test approach, and study), only forecast horizon for AUC and train/test approach for AUC and BSS were eligible for quantification of moderator effect. However, none provided a significant explanation on heterogeneity, accounting for less than 1% change in I^2 .

Regarding subgroup analysis of the remaining variables, some insights may be ascertained (all forest plots illustrating the subgroup analysis are provided as Supplementary file 2). While decreases in heterogeneity were observed in a set of subgroups in the case of BSS for source of data and forecast horizon, these subgroups were for the most part composed of results from the same studies. In fact, subgrouping by study also revealed equivalent decreases in heterogeneity for those same studies. As such, in the case of BSS, it was not possible to infer which of the two variables was the moderator.

In the case of AUC, subgrouping by study also revealed significant decreases in heterogeneity, with several subgroups presenting small to moderate I^2 values. Surprisingly, decreases in heterogeneity were also observed in one subgroup of input data and another subgroup in forecast horizon, which were not explained by the study variable.

The former corresponds to the algorithms that used seizure times in combination with other input data, and the latter to the approaches that output forecasts with a horizon of 1 h.

Discussion

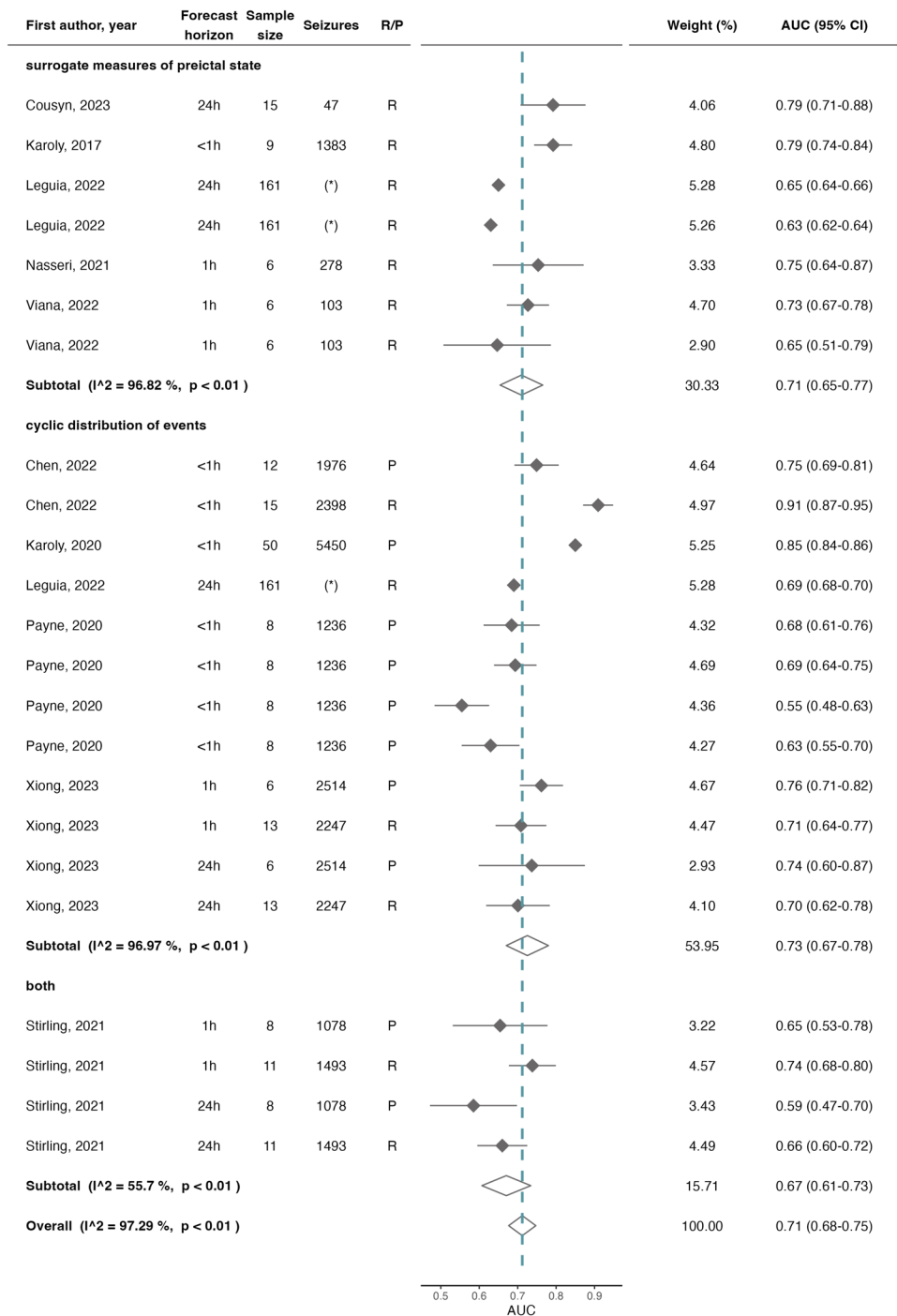
This systematic review and the meta-analyses provide a characterization of the state-of-the-art of seizure forecast algorithms along with their performances, setting a benchmark for future developments. It addresses the relevance of classic prediction performance, while also reiterating the need for human-interpretable measures of likelihood and accompanying performance metrics. Hence, two meta-analyses are provided, one on AUC (a deterministic performance metric) and another on BSS (a probabilistic performance metric).

Overview of the proposed approaches and performance

This systematic review allowed for the characterization of the current state of automated seizure forecast, effectively addressing RQ1. Firstly, it revealed a considerable diversity of the proposed approaches in regard to study design, including factors such as type of input data, algorithmic approaches, methodology for training and testing, and forecast horizons. While most proposed approaches relied on SMPS, an almost equal amount opted for exploring CDE. Moreover, ML and DL remain the most common type of algorithmic approach when SMPS is used as an input, however, phase modelling was highly favored for CDE.

The meta-analysis with the studies reporting AUC revealed an overall mean score of 0.71 (CI=0.68–0.75, $I^2=97.29\%$), with a large heterogeneity across studies. The subgroup analysis showed small to moderate heterogeneity

Fig. 5 Forest plot of AUC of seizure forecast algorithms, overall and stratified by the type of input data given to the algorithms. The dashed line illustrates the estimated overall AUC and the diamonds represent either the overall or subgroup summary. Additional information provided includes the horizon of the forecast, the sample size, total number of seizures, and the train–test approach (which was either (pseudo-)prospective, P, or retrospective, R). *AUC* area under the ROC curve. (*) Identifies the algorithm proposed by [27], since they only reported the median number of seizures (143, IQR of 13-1233)

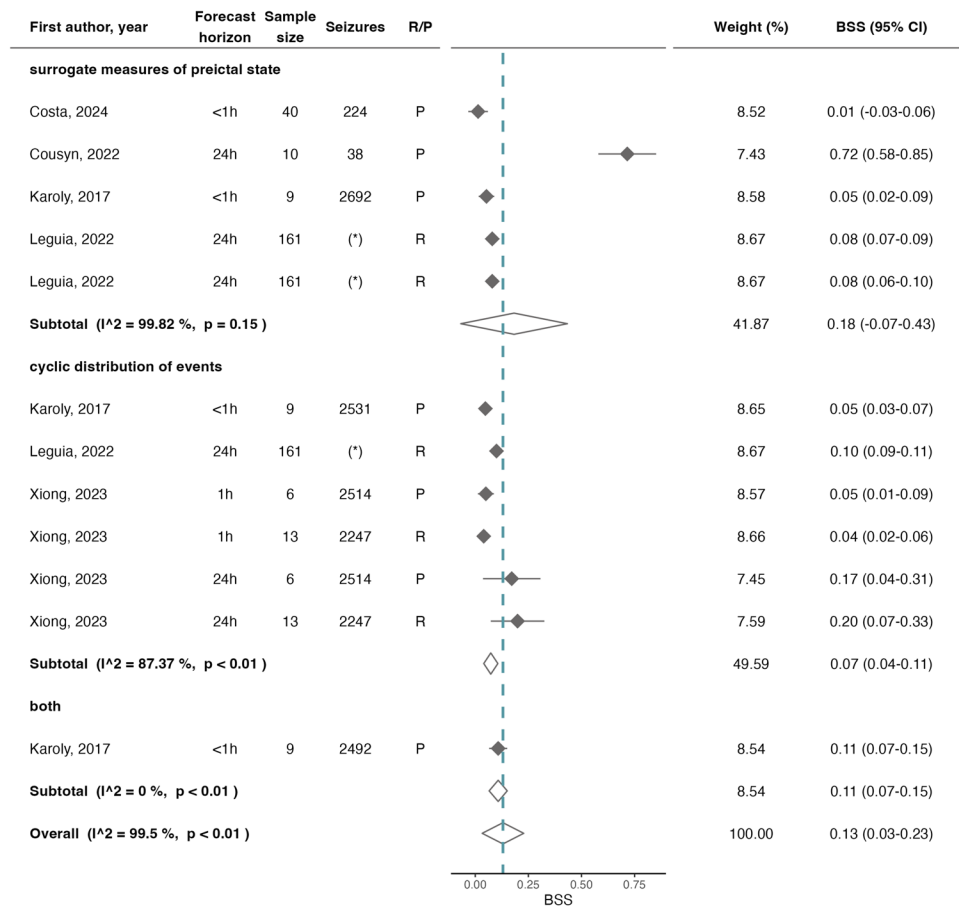


values for algorithms that used seizure times in combination with other input data and for the approaches that proposed forecasts with a horizon of 1 h. This insight potentially indicates a more reliable effect estimate for these two subgroups, with AUC of 0.70 (CI=0.67–0.73, I²=0.00%) and 0.73 (CI=0.70–0.76, I²=35.52%), respectively. Given that forecasts are often considered excellent when the AUC is above 0.9, these results indicate a

reasonable performance while also showing opportunity for improvement in regard to this evaluation metric.

Furthermore, the near-zero BSS (0.13, CI=0.03–0.23) demonstrates only a slight improvement in performance over a naive forecast. This result underscores the need for a larger focus on the development of algorithms that are optimized for accurate portrayals of the true likelihood of

Fig. 6 Forest plot of BSS of seizure forecast algorithms, overall and stratified by the type of input data given to the algorithms. The dashed line illustrates the estimated overall BSS, and the diamonds represent either the overall or subgroup summary. Additional information provided includes the horizon of the forecast, the sample size, total number of seizures, and the train–test approach (which was either (pseudo-)prospective, P, or retrospective, R). *BSS* Brier Skill Score. (*) Identifies the algorithm proposed by [27], since they only reported the median number of seizures (143, IQR of 13–1233)



a seizure occurring, instead of relying on machine-trained, threshold-based interpretations of those likelihoods.

Addressing RQ3, despite the efforts of the epilepsy community towards shifting the focus from a deterministic perspective to assessing the body states that suggest a higher likelihood of seizure, a majority of studies still relied entirely on deterministic performance metrics to evaluate their algorithms, with less than half the studies reporting probabilistic metrics, namely BS and BSS. The most common deterministic metrics were AUC, Sen, and TiW, requiring the conversion of the measures of seizure likelihood (either numerical or categorical, i.e., risk levels) into a classification of preictal/interictal segments. This reversion back to the classic notion of seizure prediction hinders the use case of seizure forecast.

Interestingly, there was not a considerable difference in AUC or BSS performance across subgroups. While these results should be interpreted with caution, due to the substantial unexplained heterogeneity and the small number of studies in each subgroup, they suggest comparable performances from both types of input data. While EEG patterns have historically been the most explored source of data due to its inherent relation to seizure onset, the pooled AUC and BSS performances along with the heterogeneity

analysis obtained in this meta-analysis would suggest that other sources of data, including the timings of seizure occurrence, peripheral physiological data, or even sleep patterns may hold relevant insights into the non-deterministic nature of seizure occurrence. However, given the observed heterogeneity, at this stage it is not possible to address RQ2 and provide a directional statement in regard to which data are the most relevant for the forecast of seizure likelihood.

Guidelines for seizure forecast

This systematic review revealed a severe lack of standardization across sources of data (i.e., datasets), train–test approaches (i.e., retrospective vs pseudo-prospective), forecast horizons (i.e., daily, hourly, or on the order of minutes), evaluation methodologies (i.e., segment based vs event based, as described in [45]), and the metrics of performance (i.e., deterministic vs probabilistic). In fact, the lack of standardization in the literature is one of the major bottlenecks identified by this review. While the objective of this work is not to establish which are the best practices or methods, future developments may propose new approaches that are comparable to the studies included in this review with respect to three domains:

1. **Data source:** It is evidently desirable to benchmark any new approach in publicly available datasets, since it allows for direct comparison of results. However, given the novelty of this area of research, publicly available datasets may not include all input data that the readers intend to explore. For this reason, self-collected datasets may offer significant added value and should, whenever possible, be made publicly available.
2. **Analysis:** While there is no directional insight regarding the preferred value for horizon of the forecasts, several studies have reported community preferences regarding this design factor [46–48] which can guide future developments. Moreover, novel approaches towards seizure forecast should explicitly state this variable, as it is crucial to not only make them comparable to the literature, but also to inform the epilepsy community how that approach would be applied as a real-world tool. Furthermore, as distinct train/test methodologies preclude quantitative comparison between two proposed approaches, we advise future developments to evaluate algorithms in a (pseudo-)prospective manner, as it provides a more comprehensive estimate on the algorithms' capacity to generalize [45] in comparison to retrospective evaluation. However, purely prospective studies may not be feasible due to resource constraints. In such cases, pseudo-prospective approaches, while not accounting for implementation constraints (such as power consumption or communication of the output to the user), aim at a more reasonable and fair estimate of performance.
3. **Reporting:** Both deterministic and probabilistic measures of performance should be reported, ideally favoring those most commonly observed in this review.

As such, beyond setting a performance benchmark for future developments in automated forecasting algorithms, we hope that the considerations found in this work drive future efforts into a convergent approach regarding study design and evaluation methodologies, in a way that reflects how the clinical application of forecasting of seizure likelihood is currently envisioned.

Limitations

This work identified some limitations in the studies included in the review. The retrospective approach during training and evaluation implies that temporal integrity is not respected, which introduces biases not only when computing profiles with cyclic nature but also when accounting for the natural evolution of patients' surrogate patterns of the preictal state across time. In this context, prospective evaluation is a valuable step towards the integration of these algorithms into the clinical setting, as it provides a more comprehensive estimate on their capacity to generalize [45]. While only one

study [29] provided a prospective period of algorithm evaluation, several have opted for a pseudo-prospective approach.

It is also important to highlight some limitations of this work. Firstly, determining the forecast horizon proved to be somewhat of a challenge. Despite efforts made by Proix et al. [22] to elucidate this concept, the forecast horizon is not consistently defined across studies (only 8 out of the 18 studies explicitly stated it): some studies propose algorithms that generate forecasts at specific clock times, naturally defining the horizon as the time between forecasts; while others solely provide the SPH and/or SOP used during event-based evaluation, without offering a clear conversion into a horizon interpretable by the user. Secondly, only a fraction of the proposed algorithms was included in each meta-analysis, despite the considerable number of algorithms found in the literature. Thirdly, there was significant unexplained heterogeneity in both meta-analyses, which may suggest the discrepancy within study designs as a potential culprit. Moreover, the significant decrease in heterogeneity when subgrouping by study may also suggest the relevance of consistency in study design for the homogeneity and robustness of future meta-analysis. The analysis on moderator effect provided very little explanation on the observed heterogeneity and the multiple subgroup analyses revealed significant imbalance in the number of approaches proposed across the different design variables.

While the results of this work provide a well-defined starting point for future developments and performance benchmarks according to study design, it is important to acknowledge that the heterogeneity here characterized may compromise pooled results and direct comparison between performances. For this reason, readers are advised to interpret the results of the meta-analyses with caution. Instead, readers are encouraged to leverage the insights from this work to identify the gaps in study design that remain underexplored and, once addressed, will enable a more reliable, quantitative summary of the state-of-the-art.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00415-024-12655-z>.

Acknowledgements This work was supported by the Fundação para a Ciência e Tecnologia (FCT), Portugal, under the grants 2021.08297.BD and 2022.12369.BD, by FCT/MCTES through national funds and when applicable co-funded EU funds under the project UIDB/50008/2020 (DOI identifier <https://doi.org/10.54499/UIDB/50008/2020>). It was done under the project “PreEpiSeizures” in a collaboration between Instituto de Telecomunicações and the Neurophysiology Monitoring Unit-EEG/Sleep Laboratory from Hospital de Santa Maria (ULSSM).

Author Contributions **Ana Sofia Carmo:** conceptualization (lead); investigation—namely following the PRISMA guidelines (equal); formal analysis (lead); visualization (lead); writing—original draft (lead); writing—review and editing (equal). **Mariana Abreu:** conceptualization (supporting); investigation—namely following the PRISMA guidelines (equal); writing—review and editing (equal). **Maria Fortuna Baptista:** investigation—namely following the PRISMA guidelines

(equal); writing—review and editing (equal). **Miguel de Oliveira Carvalho:** investigation—namely following the PRISMA guidelines (equal); writing—review and editing (equal). **Ana Rita Peralta:** conceptualization (supporting); supervision (supporting); writing—review and editing (equal). **Ana Fred:** supervision (supporting); writing—review and editing (equal). **Carla Bentes:** conceptualization (supporting); supervision (supporting); writing—review and editing (equal). **Hugo Plácido da Silva:** conceptualization (supporting); supervision (lead); writing—review and editing (equal).

Funding Open access funding provided by FCTIFCCN (b-on).

Data availability All data used for the characterization of the studies in this review, as well as for the meta-analysis are available as Supplementary Material. The details of the protocol for this systematic review are available on PROSPERO, under PROSPERO ID CRD42023478920.

Declarations

Conflicts of interest The authors have no conflict of interest to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Mackay M, Mahlaba H, Gavillet E, Whittaker RG (2017) Seizure self-prediction: myth or missed opportunity? *Seizure* 51:180–185
- Privitera M, Haut SR, Lipton RB, McGinley JS, Cornes S (2019) Seizure self-prediction in a randomized controlled trial of stress management. *Neurology* 93(22):E2021–E2031
- Grzeskowiak CL, Dumanis SB (2021) Seizure forecasting: patient and caregiver perspectives. *Front Neurol* 12:717428
- Dell KL, Payne DE, Kremen V, Maturana MI, Gerla V, Nejedly P, Worrell GA, Lenka L, Mivalt F, Boston RC, Brinkmann BH, D'Souza W, Burkitt AN, Grayden DB, Kuhlmann L, Freestone DR, Cook MJ (2021) Seizure likelihood varies with day-to-day variations in sleep duration in patients with refractory focal epilepsy: a longitudinal electroencephalography investigation. *EclinicalMedicine* 37:7
- Doherty MJ, Youn C, Gwinn RP, Haltiner AM (2007) Atmospheric pressure and seizure frequency in the epilepsy unit: preliminary observations. *Epilepsia* 48:1764–1767
- Baud MO, Kleen JK, Mirro EA, Andrechak JC, King-Stephens D, Chang EF, Rao VR (2018) Multi-day rhythms modulate seizure risk in epilepsy. *Nat Commun* 9:1–10
- Hubbard I, Beniczky S, Ryvlin P (2021) The challenging path to developing a mobile health device for epilepsy: the current landscape and where we go from here. *Front Neurol* 12:740743
- Vieluf S, Amengual-Gual M, Zhang B, El Atrache R, Ufongene C, Jackson MC, Branch S, Reinsberger C, Loddenkemper T (2021) Twenty-four-hour patterns in electrodermal activity recordings of patients with and without epileptic seizures. *Epilepsia* 62(4):960–972
- Leal A, Curty J, Lopes F, Pinto MF, Oliveira A, Sales F, Bianchi AM, Ruano MG, Dourado A, Henriques J, Teixeira CA (2023) Unsupervised EEG preictal interval identification in patients with drug-resistant epilepsy. *Sci Rep* 13(1):784
- Karoly PJ, Stirling RE, Freestone DR, Nurse ES, Maturana MI, Halliday AJ, Neal A, Gregg NM, Brinkmann BH, Richardson MP, La Gerche A, Grayden DB, D'Souza W, Cook MJ (2021) Multi-day cycles of heart rate are associated with seizure likelihood: An observational cohort study. *Ebiomedicine*. 72
- Epilepsy Foundation (2016) “2016 Community Survey,” tech. rep., Epilepsy Foundation.
- Dumanis SB, French JA, Bernard C, Worrell GA, Fureman BE (2017) “Seizure Forecasting from Idea to Reality. Outcomes of the My Seizure Gauge Epilepsy Innovation Institute Workshop. *eNeuro*. 4: 11
- Janse SA, Dumanis SB, Huwig T, Hyman S, Fureman BE, Bridges JF (2019) Patient and caregiver preferences for the potential benefits and risks of a seizure forecasting device: a best-worst scaling. *Epilepsy Behav* 96:183–191
- Badawy RA, Freestone DR, Lai A, Cook MJ (2012) Epilepsy: ever-changing states of cortical excitability. *Neuroscience* 222:89–99
- Ly JQ, Gaggioni G, Chellappa SL, Papachilleos S, Brzozowski A, Borsu C, Rosanova M, Sarasso S, Middleton B, Luxen A, Archer SN, Phillips C, Dijk DJ, Maquet P, Massimini M, Vandewalle G (2016) Circadian regulation of human cortical excitability. *Nat Commun*. 7:1–10
- Litt B, Lehnertz K (2002) Seizure prediction and the pre-seizure period. *Curr Opin Neurol* 15(2):173–177
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:3
- Maiwald T, Winterhalder M, Aschenbrenner-Scheibe R, Voss HU, Schulze-Bonhage A, Timmer J (2004) Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic. *Physica D* 194:357–368
- Truong ND, Kuhlmann L, Bonyadi MR, Querlioz D, Zhou L, Kavehei O (2019) Epileptic seizure forecasting with generative adversarial networks. *IEEE Access* 7:143999–144009
- Payne DE, Dell KL, Karoly PJ, Kremen V, Gerla V, Kuhlmann L, Worrell GA, Cook MJ, Grayden DB, Freestone DR (2021) Identifying seizure risk factors: a comparison of sleep, weather, and temporal features using a Bayesian forecast. *Epilepsia* 62(2):371–382
- Truong ND, Yang YK, Maher C, Kuhlmann L, McEwan A, Nikpour A, Kavehei O (2021) Seizure susceptibility prediction in uncontrolled epilepsy. *Front Neurol* 12:721491
- Proix T, Truccolo W, Leguia MG, Tchong TK, King-Stephens D, Rao VR, Baud MO (2021) Forecasting seizure risk in adults with focal epilepsy: a development and validation study. *Lancet Neurol* 20:127–135
- Badawy RA, Freestone DR, Lai A, Cook MJ (2012) Epilepsy: ever-changing states of cortical excitability. *Neuroscience* 222:89–99
- Kuhlmann Levin, Lopez Liz, O'Connell Meghan, rudyno5, Wang Sylvia, Cukierski Will (2016) Melbourne University AES/Math-Works/NIH Seizure Prediction.
- My Seizure Gauge Consortium (2022) My Seizure Gauge Seizure Forecasting Challenge.

26. Fu R, Gartlehner G, Grant M, Shamliyan T, Sedrakyan A, Wilt TJ, Griffith L, Oremus M, Raina P, Ismail A, Santaguida P, Lau J, Trikalinos TA (2011) Conducting quantitative synthesis when comparing medical interventions: AHRQ and the Effective Health Care Program. *J Clin Epidemiol* 64:1187–1197
27. Leguia MG, Rao VR, Tchong TK, Duun-Henriksen J, Kjær TW, Proix T, Baud MO (2022) Learning to generalize seizure forecasts. *Epilepsia*.
28. Cousyn L, Dono F, Navarro V, Chavez M (2023) Can heart rate variability identify a high-risk state of upcoming seizure? *Epilepsy Res.* 197:107232
29. Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, D'Souza W, Yerra R, Archer J, Litewka L, Hosking S, Lightfoot P, Ruedebusch V, Sheffield WD, Snyder D, Leyde K, Himes D (2013) 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* 12(6):563–571
30. Bergey GK, Morrell MJ, Mizrahi EM, Goldman A, King-Stephens D, Nair D, Srinivasan S, Jobst B, Gross RE, Shields DC, Barkley G, Salanova V, Olejniczak P, Cole A, Cash SS, Noe K, Wharen R, Worrell G, Murro AM, Edwards J, Duchowny M, Spencer D, Smith M, Geller E, Gwinn R, Skidmore C, Eisenschenk S, Berg M, Heck C, Van Ness P, Fountain N, Rutecki P, Massey A, O'Donovan C, Labar D, Duckrow RB, Hirsch LJ, Courtney T, Sun FT, Seale CG (2015) Long-term treatment with responsive brain stimulation in adults with refractory partial seizures. *Neurology* 84:810–817
31. Weisdorf S, Duun-Henriksen J, Kjeldsen MJ, Poulsen FR, Gangstad SW, Kjær TW (2019) Ultra-long-term subcutaneous home monitoring of epilepsy-490 days of EEG from nine patients. *Epilepsia* 60:2204–2214
32. Klatt J, Feldwisch-Drentrup H, Ihle M, Navarro V, Neufang M, Teixeira C, Adam C, Valderrama M, Alvarado-Rojas C, Witon A, Le Van Quyen M, Sales F, Dourado A, Timmer J, Schulze-Bonhage A, Schelter B (2012) The EPILEPSIAE database: an extensive electroencephalography database of epilepsy patients. *Epilepsia* 53:1669–1676
33. Attia TP, Viana PF, Nasser M, Richardson MP, Brinkmann BH (2021) Seizure Forecasting from Subcutaneous EEG Using Long Short Term Memory Neural Networks: Algorithm Development and Optimization. In: *Proceedings - 2021 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021*. pp. 3599–3602.
34. Chen Z, Maturana MI, Burkitt AN, Cook MJ, Grayden DB (2022) Seizure Forecasting by High-Frequency Activity (80–170 Hz) in Long-term Continuous Intracranial EEG Recordings. *Neurology*. 99: E364–E375, 7
35. Costa G, Teixeira C, Pinto MF (2024) Comparison between epileptic seizure prediction and forecasting based on machine learning. *Sci Rep* 14(1):5653
36. Cousyn L, Messaoud RB, Lehongre K, Frazzini V, Lambrecq V, Adam C, Mathon B, Navarro V, Chavez M (2022) Daily resting-state intracranial EEG connectivity for seizure risk forecasts. *Epilepsia* 64(2):E23–E29
37. Karoly PJ, Ung H, Grayden DB, Kuhlmann L, Leyde K, Cook MJ, Freestone DR (2017) The circadian profile of epilepsy improves seizure forecasting. *Brain* 140(8):2169–2182
38. Karoly PJ, Cook MJ, Maturana M, Nurse ES, Payne D, Brinkmann BH, Grayden DB, Dumanis SB, Richardson MP, Worrell GA, Schulze-Bonhage A, Kuhlmann L, Freestone DR (2020) Forecasting cycles of seizure likelihood. *Epilepsia* 61(4):776–786
39. Maturana MI, Meisel C, Dell K, Karoly PJ, D'Souza W, Grayden DB, Burkitt AN, Jiruska P, Kudlacek J, Hlinka J, Cook MJ, Kuhlmann L, Freestone DR (2020) Critical slowing down as a biomarker for seizure susceptibility. *Nat Commun.* 11:1–12, 5
40. Nasser M, Pal Attia T, Joseph B, Gregg NM, Nurse ES, Viana PF, Worrell G, Dümpelmann M, Richardson MP, Freestone DR, Brinkmann BH (2021) Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning. *Sci Rep.* 11:1–9
41. Stirling RE, Grayden DB, D'Souza W, Cook MJ, Nurse E, Freestone DR, Payne DE, Brinkmann BH, Pal Attia T, Viana PF, Richardson MP, Karoly PJ (2021) Forecasting seizure likelihood with wearable technology. *Front Neurol.* 12:704060
42. Stirling RE, Maturana MI, Karoly PJ, Nurse ES, McCutcheon K, Grayden DB, Ringo SG, Heasman JM, Hoare RJ, Lai A, D'Souza W, Seneviratne U, Seiderer L, McLean KJ, Bulluss KJ, Murphy M, Brinkmann BH, Richardson MP, Freestone DR, Cook MJ (2021) Seizure forecasting using a novel sub-scalp ultra-long term EEG monitoring system. *Front Neurol.* 12:713794
43. Viana PF, Pal Attia T, Nasser M, Duun-Henriksen J, Biondi A, Winston JS, Pavão Martins I, Nurse ES, Dümpelmann M, Schulze-Bonhage A, Freestone DR, Kjaer TW, Richardson MP, Brinkmann BH (2022) Seizure forecasting using minimally invasive, ultra-long-term subcutaneous electroencephalography: Individualized inpatient models. *Epilepsia*.
44. Xiong WJ, Stirling RE, Payne DE, Nurse ES, Kameneva T, Cook MJ, Viana PF, Richardson MP, Brinkmann BH, Freestone DR, Karoly PJ (2023) Forecasting seizure likelihood from cycles of self-reported events and heart rate: a prospective pilot study. *Ebiomedicine.* 93
45. Ren Z, Han X, Wang B (2022) The performance evaluation of the state-of-the-art EEG-based seizure prediction models. *Front Neurol* 13:11
46. Janse SA, Dumanis SB, Huwig T, Hyman S, Fureman BE, Bridges JF (2019) Patient and caregiver preferences for the potential benefits and risks of a seizure forecasting device: a best-worst scaling. *Epilepsy Behav* 96:183–191
47. Chiang S, Moss R, Black AP, Jackson M, Moss C, Bidwell J, Meisel C, Loddenkemper T (2021) Evaluation and recommendations for effective data visualization for seizure forecasting algorithms. *JAMIA Open* 4:1–14
48. Grzeskowiak CL, Dumanis SB (2021) Seizure forecasting: patient and caregiver perspectives. *Front Neurol* 12:717428–717428