Seizure prediction and forecasting: a scoping review

Joshua C. Cheng and Daniel M. Goldenholz

Purpose of review
This scoping review summarizes key developments in the field of seizure forecasting.

Recent findings
Developments have been made along several modalities of seizure forecasting, including long term intracranial and subcutaneous encephalogram, wearable physiologic monitoring, and seizure diaries. However, clinical translation of these tools is limited by various factors. One is the lack of validation of these tools on an external dataset. Moreover, the widespread practice of comparing models to a chance forecaster may be inadequate. Instead, the model should be able to at least surpass a moving average forecaster, which serves as a 'napkin test' (i.e., can be computed on the back of a napkin). The impact of seizure frequency on model performance should also be accounted for when comparing performance across studies. Surprisingly, despite the potential for poor quality forecasts, some individuals with epilepsy still want access to imprecise forecasts and some even alter their behavior based upon them.

Summary
Promising advances have been made in the development of tools for seizure forecasting, but current tools have not yet overcome clinical translation hurdles. Future studies will need to address potentially dangerous patient behaviors as well as account for external validation, the napkin test, seizure frequency dependent metrics.

Keywords
encephalogram, machine learning, seizure diaries, seizure forecasting

INTRODUCTION

One of the most debilitating components of living with epilepsy is the uncertainty regarding the next upcoming seizure. During a survey conducted by the Fourth International Workshop on Seizure prediction, a majority of respondents endorsed that 'fear' was the most difficult part of living with epilepsy [1]. An accurate seizure forecast could both help reduce uncertainty as well as offer time to increase safety.

In 2013, a pilot study of patients with drug-resistant epilepsy participating in the NeuroVista trial showed that intracranial encephalogram (EEG) recordings may have the potential to forecast upcoming seizures [2]. In parallel, other early studies showed that seizure diaries [3], or physiologic signals from wearable wristbands [4] also have potential for seizure forecasting. Yet, despite the promise of these early studies and subsequent developments, the widespread integration of a reliable tool for seizure forecasting for clinical use has not yet been achieved.

Here, we will discuss some potential concerns and barriers which limit current clinical translation of tools for seizure forecasting (see Table 1). We will also discuss several recent developments across various seizure forecasting tools, highlighting their novel advancements as well as limitations (see Table 2).

CONCERNS AND LIMITATIONS

The application of machine learning algorithms is integral to most studies on seizure forecasting. Ideally, the dataset for developing the algorithm contains a training set (from which the algorithm learns patterns/associations), a validation set (often the left-out data in cross validation on which the model is validated/optimized), as well as a hold-out test set as a final assessment of its performance.

Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

Correspondence to Daniel M. Goldenholz, MD, PhD, 330 Brookline Ave, Baker 5, Boston, MA 02215, USA. Tel: +1 617 632 8930; e-mail: daniel.goldenholz@bidmc.harvard.edu

Curr Opin Neurol 2025, 38:135–139

DOI:10.1097/WCO.0000000000001344

KEY POINTS

- Advances have been made in seizure forecasting using intracranial and subcutaneous encephalogram data, wearables, and seizure diaries.
- Clinical translation of seizure forecasting tools developed may be limited by factors such as lack of external validation, inability to pass the napkin test (risk = average seizure rate calculated on the back of a napkin), and not accounting for seizure frequency.
- Remarkably, despite the potential for imprecise forecasts, individuals with epilepsy continue to want access to seizure forecasts and some will even alter their behavior based upon these forecasts.

Importantly, no model changes occur after the hold-out test set is assessed, to prevent data leakage and overfitting. Currently, many studies on seizure forecasting lack an external held-out test set for assessing its performance and generalizability [5▪▪]. Rather, performance is often derived from the validation set during cross-validation. Often, how model hyperparameters are chosen is also not well specified. This is potentially concerning as the model performance may be overstated due to unreported data leakage and overfitting, which could negatively impact generalizability.

Table 1. List of key pitfalls in seizure forecasting along with various methods to avoid them

Key pitfalls	How to avoid
Lack of external validation dataset.	Ensure that the seizure forecasting model is evaluated once only on a separate external validation dataset as an assessment of its performance
Statistically significant but not clinically meaningful model	Compare the performance of the seizure forecasting model against a moving average forecaster (the 'napkin test'), rather than comparing to a chance forecaster
Not accounting for seizure frequency	Ensure that seizure frequencies of the participants are reported, and how model performance may scale with seizure frequency
Highlighting a single performance metric	Show performance in calibration (Brier Score, calibration curve) and discrimination (AUC-ROC, AUC-PR), to demonstrated robust results

AUC, area under curve.

Another significant concern is the current determination of 'significance' of a machine learning model in seizure forecasting. One common approach across many studies is to compare its performance against a 'chance' forecaster (i.e., such as permuted model forecasts). If the improvement in performance of the seizure forecaster over the chance model is statistically significant, then this is often considered an indicator of a well performing model. However, the threshold to be better than a chance forecaster is low, and as such the performance of the model can be statistically significant without being clinically meaningful. For example, one commonly examined metric is area under curve (AUC) of the model, which accounts for both sensitivity and specificity. Although there is variability on what is a 'good' AUC score, a value of 0.5 is essentially chance discrimination, with values up to 0.6 still being within the random or poor range [6]. It would not be unexpected to see seizure forecasting models within that range still be statistically significant when compared to a chance forecaster. Such significant models would still be unlikely to translate well clinically. Moreover, a rate-matched random forecast was found to have an AUC of 0.83, highlighting the importance of contextualizing AUC values [3]; the rate-matched random forecast was very poorly calibrated and not well suited for clinical use. Additionally, the AUC reported in that study reflected all forecasts considered together (rather than per patient AUC), blending results from patients with low and high seizure frequencies, thereby obscuring poor forecasting performance.

In an effort to improve benchmarking, one validation study applied a prior machine learning model based on seizure diaries on a new prospective dataset [7▪]. Notably, beyond comparing the performance of the machine learning model against a chance forecaster, they also compared it to a moving average forecaster. The moving average (MA) forecaster for each participant simply took their total number of seizure days in the prior 84 days divided by 84 to predict future daily risk. Although the machine learning algorithm based on seizure diaries was able to outperform the chance forecaster, performance was not significantly better than the simplistic MA. The MA benchmark might be thought of as the 'napkin test' because any clinically relevant forecast must be better than a method that can be calculated by hand on a napkin. This is so for two reasons. First, patients and clinicians likely already have an intuitive feel for the MA level of risk which only changes over long time-scales. Second, any model which is less accurate than MA should not

Table 2. Summary of key studies across various modalities of seizure forecasting and their limitations

Modality	Study	External hold-out validation?	Compared against a moving average?	Accounts for seizure frequency?
Intracranial EEG seizure cycles	Leguia *et al.*, 2023 Khambati *et al.*, 2024	No No	No No	No No
Subcutaneous EEG	Viana *et al.*, 2023	No	No	No
Scalp EEG	Wong *et al.*, 2023	Sometimes[a]	No	No
Wearables	Xiong *et al.*, 2023	Yes	No	No
Diaries	Gleichgerrcht *et al.*, 2022 Goldenholz *et al.*, 2024	No Yes	No Yes	No Yes

EEG, encephalogram.
[a]The referenced study Wong *et al.* (2023) is a review of many studies.

be preferred to the more intuitive MA. Newer analysis demonstrates that comparing any candidate forecasting tool against a MA is much more challenging than the permutation test (i.e. chance), because the MA forecast is already surprisingly accurate [8▪▪]. Patients and caregivers can already access the MA calculation without the need for any technical devices or software, and likely implicitly account for the MA risk. As such, the development of a seizure forecaster which cannot pass the napkin test is unlikely to be clinically meaningful.

Another important consideration is the influence of seizure frequency on model performance. In one study, when various metrics of model performance were stratified by seizure frequency, it was found that performance generally varied as a function of seizure frequency [8▪▪]. Of note, one performance metric (AUC receiver-operating characteristics) was found to be independent from seizure frequency. Whereas AUC is good for testing discrimination, it is insufficient to test the calibration of a forecasting tool. As different patients have different seizure frequencies, this work suggests the importance of taking seizure frequency into account when comparing performance across datasets, and including multiple metrics for both discrimination and calibration.

NOVEL DEVELOPMENTS IN SEIZURE FORECASTING

In 2018, several studies including intracranial EEG and seizure diaries revealed seizure risk exhibits cycles which span multiple timescales, including circadian (24 h) and multidien (multiday) cycles [9–11]. Using multidien seizure cycles based on intracranial and subcutaneous EEG data, one study attempted to develop a model which could make seizure forecasts in new participants [12]. Being able to do so would be advantageous as it would provide the potential for seizure forecasts to be available for a given patient from the start, rather than warranting the time and need for a model to be trained from that individual's data first. To do this, they characterized multidien seizure cycles based on interictal epileptiform activity (IEA) for each participant derived from their EEG. Seizures were assessed by self-reported diaries. They then extracted the instantaneous phase of the multidien seizure cycles and used this data in either a point process general linear model or recurrent neural network to forecast seizures above chance in 79% and 81% respectively of unseen participants. However, this work was limited by their use of a noncausal wavelet transform, which requires the use of information from the future. In addition, their model was only compared to a chance forecaster.

A follow-up study used functional connectivity (instead of IEA) to compute seizure risk cycles for forecasting [13▪]. Specifically, they first showed that functional connectivity of the hippocampus fluctuates with IEA multidien seizure cycles, demonstrating that hippocampal functional connectivity exhibits multidien cycles. Using a leave-one-out cross validation approach, they then showed that models based on hippocampal functional connectivity can perform statistically significant seizure forecasting compared to chance. However, no separate hold-out test set was used to independently assess model performance (see Table 2). As the model was compared to chance rather than a moving average forecaster, even relatively low individual AUCs (range: 0.53–0.91) were still able to achieve statistical significance.

Independent of the cycle approach, one study used long-term subcutaneous two-channel EEG for seizure forecasting [14]. In brief, the initial one third of an individual's data was used as the training data, with the remaining data used as the testing data in a pseudoprospective approach. Across the various algorithms applied, at least half of the 6 subjects had statistically significant seizure forecasting when compared to chance, with AUC ranging from 0.65 to 0.74.

One limitation of the study was benchmarking against chance (see Table 2). However, another was the limited spatial coverage of the device given that only two channels were available. Towards the latter, a novel full-head subscalp EEG for long-term monitoring has been developed for individuals with drug resistant epilepsy [15]. Over 9 days, this was shown to be non-inferior to scalp EEG for measuring stage-specific brain oscillations and detecting interictal spikes/ictal signals. Although early in development, one remains hopeful that these developments may lead to ambulatory tools that could aid seizure forecasting.

SEIZURE FORECASTING IN SCALP ENCEPHALOGRAM

There are multiple publicly available scalp EEG datasets from individuals with epilepsy, with numerous studies making use of them for seizure forecasting [16]. For example, accuracies $> 90\%$ for seizure prediction have been reported [17,18]. However, no prospective validation has been performed. This may be limited by the lack of a form factor for scalp EEG to be worn long-term that would allow for real-time forecasting. Other issues include lack of uniformity across datasets such as number of channels/channel placements or sampling frequency [16]. One insightful study found that typical EEG forecasting techniques run the risk of confusing the difference between true seizure forecasting with detecting the temporal proximity to similar appearing EEG noise. This was confirmed by using standard forecasting practices with the goal of detecting randomly selected points in time, rather than true seizures [5▪▪]. Thus, very careful controls need to be used with EEG and other physiologic signals to avoid these pitfalls.

SEIZURE FORECASTS BASED ON PHYSIOLOGIC DATA

Through the Empatica wearable wristband, a wide variety of physiologic signals including temperature, heart rate, heart rate variability (HRV), electrodermal activity, and 3-axis accelerometry can be obtained. Work has shown that these physiologic signals may have cycles as well that appear synchronized with seizure cycles [19]. However, HRV-based algorithms for seizure forecasting should be reserved for those who have prominent ictal autonomic changes, especially as nonconvulsive seizures in adults may be associated with minimal autonomic changes [20].

For more natural settings, one study in 13 participants with epilepsy and uncontrolled seizures collected continuous long-term heart rate data via a commercial smartwatch (Fitbit) with self-reported seizures through a smartphone app [21]. They used the data leading up to and including the 10^{th} seizure as the training data, from which they extracted heart rate and seizure cycles. They then fit the 'Prophet' model (Facebook, USA) on this data to generate future seizure likelihoods at an hourly resolution. By using this approach, better than chance seizure forecasts were obtained during the testing period in the majority of patients (9/13 for hourly resolution, 11/13 for daily resolution forecasts). Notably, they then validated this model by applying it in a prospective manner to show that it can forecast seizures above chance in 4/6 or 5/6 individuals based on hourly or daily resolutions respectively (see Table 2).

SEIZURE FORECASTING BASED ON DIARIES

Using data gathered from the Human Epilepsy Project on individuals with focal epilepsy, one study applied machine learning on seizure diaries to predict seizure timing in novel participants pseudo-prospectively [22]. Using their approach, they were able to predict the day for 87% of seizures in unseen individuals within a margin of error of 4 days. The margin of error did not account for baseline seizure frequency (see Table 2).

Another study, based on initial work [3] showing a promising deep learning approach to seizure forecasting based on diaries, tested the same technique on a prospective dataset of 25 individuals with epilepsy. Although group level results were encouraging with an AUC of 0.82, individual level results were poor ($\text{AUC} = 0.43 \pm 0.21$). Notably, the seizure forecaster failed the napkin test (see Table 2). The conclusion of that study is the initially promising approach fell short of being clinically useful.

PERCEPTION AND RESPONSE TO SEIZURE FORECASTS

Remarkably, despite the potential for poor quality seizure forecasts, one surprising finding is that when individuals with epilepsy were surveyed, 91% of respondents did not mind poor quality forecasts and still wanted access to seizure forecasts irrespective of accuracy [7▪]. Similarly, one study showed that more than half of individuals altered their behavior based on seizure forecasts, such as how they scheduled their travel or social activities despite inaccurate forecasts [23▪▪]. One rationale for the continued desire for seizure forecasts was that it improved individuals' perception of control in their lives and reduced uncertainty, regardless of whether these perceptions were backed by accurate models [23▪▪]. Additional safeguards may be needed

when forecasts are less than 100% accurate, but even perfect probabilistic forecasts may be challenging for some patients to use safely. More study is needed to determine safer ways to present probabilistic forecasts in order to minimize patient harm.

CONCLUSION

There remains an unmet need for tools that can accurately and reliably forecast seizures. Advancements have been made along several modalities, including long-term EEG, wearables, and diaries. Unfortunately, several pitfalls limit widespread clinical translation of currently developed tools: including lack of external validation, benchmarking against chance rather than the napkin test, and not accounting for the impact of seizure frequencies on performance metrics.

Despite the fact that some patients are ready to embrace imperfect forecasts now, it is incumbent of the field to continue advancing the field. We look forward to a day when our patients can experience less uncertainty and more safety.

Acknowledgements

None.

Financial support and sponsorship

Funding: DMG was supported by NIH K23NS124656.

Conflicts of interest

J.C.C. has no disclosures. DMG is an unpaid advisor for Epilepsy AI and Eysz. He has been a paid advisor for Magic Leap. He has been provided speaker fees from AAN, AES, ACNS and AI in Epilepsy and Neurology. He also previously has been a paid consultant for Neuro Event Labs, IDR, LivaNova and Health Advances. He has received grants from NIH and BIDMC. None of the above relationships pose a financial conflict of interest.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- ■ of special interest
- ■■ of outstanding interest

1. Arthurs S, Zaveri HP, Frei MG, Osorio I. Patient and caregiver perspectives on seizure prediction. Epilepsy Behav 2010; 19:474–477.
2. Cook MJ, O'Brien TJ, Berkovic SF, *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; 12:563–571.
3. Goldenholz DM, Goldenholz SR, Romero J, *et al.* Development and validation of forecasting next reported seizure using e-diaries. Ann Neurol 2020; 88:588–595.
4. Nasseri M, Pal Attia T, Joseph B, *et al.* Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning. Sci Rep 2021; 11:21935.
5. ■■ West J, Dasht Bozorgi Z, Herron J, *et al.* Machine learning seizure prediction: one problematic but accepted practice. J Neural Eng 2023; 20:.

This paper provides a detailed exploration into how standard practices in EEG forecasting can result in a nongeneralizable model. The ideas here may also have implications for other wearables in the forecasting arena.

6. Hond AAH de, Steyerberg EW, Calster B van. Interpreting area under the receiver operating characteristic curve. Lancet Digital Health 2022; 4: e853–e855.
7. ■ Goldenholz DM, Eccleston C, Moss R, Westover MB. Prospective validation of a seizure diary forecasting falls short. Epilepsia 2024; 65:1730–1736.

This study includes several useful findings. After finding a previously reported model has poor validation using rigorous methods, it explores the subjective responses of patients to having inaccurate forecasts.

8. ■■ Chang CY, Zhang B, Moss R, *et al.* Necessary for seizure forecasting outcome metrics: seizure frequency and benchmark model. Epilepsy Res 2024; 208:107474.

This paper helps to clarify improved methods for characterizing forecasting tools, but also shows the advantage of using moving average as a benchmark to compare other models rather than chance forecasting.

9. Leguia MG, Andrzejak RG, Rummel C, *et al.* Seizure cycles in focal epilepsy. JAMA Neurol 2021; 78:454–463.
10. Karoly PJ, Goldenholz DM, Freestone DR, *et al.* Circadian and circaseptan rhythms in human epilepsy: a retrospective cohort study. Lancet Neurol 2018; 17:977–985.
11. Baud MO, Kleen JK, Mirro EA, *et al.* Multiday rhythms modulate seizure risk in epilepsy. Nat Commun 2018; 9:88.
12. Leguia MG, Rao VR, Tcheng TK, *et al.* Learning to generalize seizure forecasts. Epilepsia 2023; 64(Suppl 4):S99–S113.
13. ■ Khambhati AN, Chang EF, Baud MO, Rao VR. Hippocampal network activity forecasts epileptic seizures. Nat Med 2024; 30:2787–2790.

This paper reveals a potential new avenue for exploration that may be useful for implanted devices to achieve meaningful seizure forecasting.

14. Viana PF, Pal Attia T, Nasseri M, *et al.* Seizure forecasting using minimally invasive, ultra-long-term subcutaneous electroencephalography: Individualized intrapatient models. Epilepsia 2023; 64(S4):S124–S133.
15. van Maren E, Alnes SL, Ramos da Cruz J, *et al.* Feasibility, safety, and performance of full-head subscalp EEG using minimally invasive electrode implantation. Neurology 2024; 102:e209428.
16. Wong S, Simmons A, Rivera-Villicana J, *et al.* EEG datasets for seizure detection and prediction—a review. Epilepsia Open 2023; 8:252–267.
17. Wang Y, Shi Y, Cheng Y, *et al.* A Spatiotemporal graph attention network based on synchronization for epileptic seizure prediction. IEEE J Biomed Health Inform 2023; 27:900–911.
18. Zhang Y, Li X, Wang S, *et al.* A robust seizure detection and prediction method with feature selection and spatio-temporal casual neural network model. J Neural Eng 2023; 20:056036.
19. Gregg NM, Pal Attia T, Nasseri M, *et al.* Seizure occurrence is linked to multiday cycles in diverse physiological signals. Epilepsia 2023; 64: 1627–1639.
20. Mason F, Scarabello A, Taruffi L, *et al.* Heart rate variability as a tool for seizure prediction: a scoping review. J Clin Med 2024; 13:747.
21. Xiong W, Stirling RE, Payne DE, *et al.* Forecasting seizure likelihood from cycles of self-reported events and heart rate: a prospective pilot study. EBioMedicine 2023; 93:104656.
22. Gleichgerrcht E, Dumitru M, Hartmann DA, *et al.* Seizure forecasting using machine learning models trained by seizure diaries. Physiol Meas 2022; 43. doi: 10.1088/1361-6579/aca6ca.
23. ■■ Stirling RE, Nurse ES, Payne D, *et al.* User experience of a seizure risk forecasting app: a mixed methods investigation. Epilepsy Behav 2024; 157:109876.

This study is extremely important because it reveals the self-reported behavior changes from patients who had real-time access to seizure forecasts that were less than 100% accurate. The psychological factors of forecasting is under-explored, and in that regard, this study is a landmark looking at what kinds of behaviors one might expect.