HHS Public Access

Author manuscript

Ann Neurol. Author manuscript; available in PMC 2021 May 01.

Published in final edited form as:

Ann Neurol. 2020 September ; 88(3): 588–595. doi:10.1002/ana.25812.

Development and Validation of Forecasting Next Reported Seizure Using e-Diaries

Daniel M. Goldenholz, MD, PhD[1,2], **Shira R. Goldenholz, MD, MPH**[1], **Juan Romero, MS**[1,2], **Rob Moss, BS**[3], **Haoqi Sun, PhD**[2,4], **Brandon Westover, MD, PhD**[2,4]

[1]Department of Neurology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA

[2]Harvard Medical School, Boston, Massachusetts, USA

[3]Seizure Tracker, Springfield, Virginia, USA

[4]Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts, USA

Abstract

Objective: There are no validated methods for predicting the timing of seizures. Using machine learning, we sought to forecast 24-hour risk of self-reported seizure from e-diaries.

Methods: Data from 5,419 patients on SeizureTracker.com (including seizure count, type, and duration) were split into training (3,806 patients/1,665,215 patient-days) and testing (1,613 patients/549,588 patient-days) sets with no overlapping patients. An artificial intelligence (AI) program, consisting of recurrent networks followed by a multilayer perceptron ("deep learning" model), was trained to produce risk forecasts. Forecasts were made from a sliding window of 3-month diary history for each day of each patient's diary. After training, the model parameters were held constant and the testing set was scored. A rate-matched random (RMR) forecast was compared to the AI. Comparisons were made using the area under the receiver operating characteristic curve (AUC), a measure of binary discrimination performance, and the Brier score, a measure of forecast calibration. The Brier skill score (BSS) measured the improvement of the AI Brier score compared to the benchmark RMR Brier score. Confidence intervals (CIs) on performance statistics were obtained via bootstrapping.

Results: The AUC was 0.86 (95% CI = 0.85–0.88) for AI and 0.83 (95% CI = 0.81–0.85) for RMR, favoring AI ($p < 0.001$). Overall (all patients combined), BSS was 0.27 (95% CI = 0.23–0.31), also favoring AI ($p < 0.001$).

Interpretation: The AI produced a valid forecast superior to a chance forecaster, and provided meaningful forecasts in the majority of patients. Future studies will be needed to quantify the clinical value of these forecasts for patients.

Address correspondence to Dr Goldenholz, Department of Neurology, Beth Israel Deaconess Medical Center, 330 Brookline Ave, Baker 5, Boston, MA 02215. daniel.goldenholz@bidmc.harvard.edu.

Author Contributions

Conception and design: D.G., B.W., H.S. Acquisition and analysis of data: D.G., R.M. Drafting and preparing figures: all authors.

Potential Conflicts of Interest

Nothing to report.

Introduction

One of the most debilitating aspects of epilepsy is its unpredictability. Not knowing when the next seizure will come results in, among other things, loss of driving privileges, decreased employability, social stigma, and anxiety.[1] This lack of foreknowledge is one of the reasons most patients take daily antiseizure medications rather than only during higher risk periods.

Patients and caregivers are very interested in noninvasive forecasting tools capable of providing 24-hour warnings, even if the forecasts are imperfect.[2] Based on several recent long-term outpatient studies, there may be predictable features in seizure diaries.[3–7] Advances in machine learning techniques, particularly deep learning,[8] and the availability of large seizure diary datasets[9,10] now permit a novel approach to the challenge of risk forecasting.

We hypothesized that using a database of >1 million patient-reported seizures,[9] it is possible to accurately train an algorithm to forecast the probability of future seizures. The present study aimed to design and validate an artificial intelligence (AI) algorithm to forecast the chance of a reported seizure occurring within 24 hours based on electronic seizure diary entries.

Patients and Methods

Data

Seizure diary data were obtained from SeizureTracker.com[9,10] between December 1, 2007 and March 13, 2018 in deidentified and unlinked format. All patients included did not opt out of research on the Seizure Tracker website. The study was determined exempted by the Beth Israel Deaconess Medical Center Institutional Review Board. Seizure Tracker is a free-to-users multiplatform resource that includes web-based tools, mobile device applications, and a smart speaker “skill” that together help patients and caregivers manage seizure diary information.[10] This patient cohort has been previously characterized, including seizure rates, seizure types, and epilepsy types.[9]

All seizure records (training and testing) with reported seizure dates outside the export range (December 1, 2007–March 13, 2018) or with erroneous (negative) seizure duration values were excluded. No data imputation was used. Patients were considered eligible if they had at least 3 seizures recorded and had at least 85 days between the first and last recorded seizure.

Patient data were divided into a training set and a testing set using a cutoff date of November 30, 2015 (Fig 1). All eligible patients up to the cutoff were included in the training set, with diaries being truncated at the cutoff date. All diaries of eligible patients that began after the cutoff date were included in the test set. Accordingly, the testing set comprised only “new” patients who were not encountered by the AI in the training phase.

Forecasts were computed each day in each patient, using a sliding window (Fig 2). A 3-month “look-back” history was composed into a 3 × 84 matrix for each AI forecast. The rows were comprised of: (1) daily seizure counts, (2) average daily seizure duration in

seconds, and (3) daily generalized tonic–clonic seizure counts. These 3 covariates were abstracted from patient-reported diary entries. Each column represented one 24-hour day. There were 84 columns representing days in the 3-month look-back window. Each row was independently normalized by the maximum value per row.

Model Training

AI Forecasting Model.—The algorithm consisted of a multilayer artificial neural network (ie, deep learning model) comprised of a recurrent neural network connected to a multilayer perceptron[11] (see Fig 2). The output of the system was the estimated probability of a seizure occurring within the next 24 hours. Dropout regularization at 50% was used after each recurrent layer. Three hyperparameters were grid searched with cross-validation: number of covariates (1 or 3), number of history days (84 or 56 or 28), and recurrent network layer type. If number of covariates was 1, only daily seizure counts were used; if the number was 3, input also included a row of daily seizure durations and a row of number of generalized tonic–clonic seizures. The recurrent neural network layers were adjusted to be long short-term memory (LSTM), bidirectional LSTM, or gated recurrent unit methods.[11]

Training used K-fold (K = 11) cross-validation to optimize the above hyperparameters. The hyperparameters that minimized the binary cross-entropy in the training set were used on the testing set without further modification. The finalized flowchart is shown in Figure 2, comprising 1 covariate, 84-day history window, and LSTMs in the recurrent layers.

Rate-Matched Random Forecasting Model.—A benchmark comparison forecasting model was developed that represents an informed “random forecaster.” Like a coin flip, a random forecasting model (rate-matched random [RMR]) represents a random guess. Unlike the flip of a fair coin, RMR was biased to make guesses aligned with the individual’s historical “usual” seizure rate. As such, RMR assumed one’s future risk was equal to the average seizure count per day over the prior 3 months (see Fig 2). This rate was obtained by taking the same 84-day history of seizure counts used by the AI (described above). Because RMR is recalculated each forecasted day, it is possible for RMR to achieve a better accuracy than a stationary forecast if a patient’s average rate changes over the course of months or years.[12]

Evaluating Model Performance

The final AI model was applied to the testing set by processing the seizure diaries in the same manner as above. Forecasts were formed for each day in each patient in the testing set. Using the known seizure outcome and the forecast from each of the models, we evaluated model performance in 2 ways. First, we evaluated the model’s ability to make binary predictions (seizure vs no seizure) using the receiver operating characteristic (ROC) analysis on the test set; performance is summarized as the area under the ROC curve (AUC). Using these ROC plots, the position with the optimal balance between sensitivity and specificity was used to compute an F1 score, which is a balanced metric of false positives and false negatives. Second, we evaluated model calibration, that is, how closely the predicted probability of seizures matched the observed rate of seizures in the test set. Forecasting models are said to be “well calibrated” when the predicted risk closely matches the observed

risk. Calibration was quantified using Brier scores.[13] These were calculated for the 2 forecasting methods. A perfect Brier score is 0; an always-wrong forecaster has a score of 1.

Brier skill scores (BSSs)[14] were also calculated; these allow for a direct comparison of the AI method to the RMR method. A value of BSS > 0 suggests that AI is superior to the RMR method. BSS values were calculated based on Brier scores and modified Brier scores.

Confidence intervals (CIs) were calculated using 5,000 rounds of bootstrapping (repeated random selection of patients with replacement), with each bootstrap sample consisting of 1,613 patients randomly selected with replacement from the testing set, permitting a way to assess the reliability of the performance metrics. Brier scores and AUC metrics were compared using the permutation test.

Individual level performance was explored using a dichotomized version of the AI forecasts. To dichotomize, a per-patient threshold was selected by identifying the value that optimized the sensitivity/specificity tradeoff with the ROC. That threshold was then applied, and sensitivity, time in warning, percentage of seizures in HIGH, and percentage of seizures in LOW were computed and plotted for each patient in the testing set.

This study adheres to the principles of TRIPOD (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis).[15,16]

Results

The full database included 15,855 users of Seizure Tracker with any seizure documented. Of those, the subset of eligible patients was divided into a training set (3,806 patients, 1,665,215 patient-days total) and testing set (1,613 patients, 549,588 patient-days total). A summary of basic characteristics of the patients is shown in the Table.

A calibration curve compares forecasted risk to the true risk. An ideal forecaster would forecast a risk of X when the true risk is X. For example, consider a weather forecaster who states, “The chance of snow today is 45%.” For an ideal forecaster, when we look back at all past days when the same forecast was given, we will find that it snowed on 45% of such days. In other words, for an ideal forecaster, we find a diagonal line y = x when plotting the true risk (y) versus the forecasted risk (x). Figure 3 shows the AI forecaster to be much closer to the ideal than RMR. RMR has poor performance at the higher end of the forecast probability scale, because it rarely forecasts high values, and when it does, it is extremely inaccurate. Conversely, the AI calibration curve remains close to the y = x (ideal forecaster) line throughout the curve. In an analogous fashion, the ROC plot (see Fig 3) demonstrates that the AI surpasses the RMR across thresholds.

The Brier score was 0.10 (95% CI = 0.09–0.11) for AI and 0.14 (95% CI = 0.13–0.15) for RMR, favoring AI ($p = 0.0005$). The BSS was 0.27 (95% CI = 0.23–0.31), showing that the AI improves over RMR ($p < 0.0001$). The AUC was 0.86 (95% CI = 0.85–0.88) for AI and 0.83 (95% CI = 0.81–0.85) for RMR, again showing AI to be superior ($p = 0.0005$). The F1 score for AI was 0.56 (95% CI = 0.53–0.60) and for RMR was 0.53 (95% CI = 0.50–0.57), also showing AI superior ($p = 0.0005$).

Comparing the AUC values at the individual level, 57% of patients demonstrated $AUC_{AI} > AUC_{RMR}$.

An example patient diary is shown to illustrate the result of forecasts (Fig 4). An optimal threshold is highlighted to facilitate dichotomizing the forecast values if desired. In that case, forecasts above the threshold would be considered warning days, and below threshold would be considered no-warning days. Other patients may prefer to use the actual forecast value rather than a dichotomized forecast.

Figure 5 shows the metrics for performance on the patient level for all patients in the testing set, with seizure rates indicated by color.

Discussion

This study demonstrates that even without electroencephalography (EEG) or implanted biosensors—using only electronic seizure diaries—an AI approach is able to produce accurate forecasts of the next recorded seizure event. The AI was trained on one of the world's largest patient-reported seizure diary databases. Using a set of 1,613 patients whom the AI had never previously encountered in training, the technique made forecasts that were more accurate than a matched random forecaster according to multiple metrics. Metrics at the individual patient level indicated that the majority of patients had forecasts that were more accurate using the AI compared with RMR. Taken together, these results indicate that the AI is a valid forecasting method for seizures occurring 24 hours in the future.

Studies from intracranial EEG[4] and patient-reported diaries[17] report that circadian[14] and other multiscale cycles appear in seizure timing. Harnessing these patterns would be valuable for a successful forecaster system. In the absence of a priori knowledge of these cycles (including how many are present and their phase, amplitude, and evolution in time), a forecasting method would need to fit a large number of covariates and would require a detailed model of how multiscale cycles work. Instead, the present study employed a deep learning algorithm to essentially teach itself what patterns and models would be most appropriate by observing a very large set of patient diaries.

The primary outcome evaluation metrics used in this study differ from recommendations of some groups that focus on binary outcomes, such as "time in warning" and "sensitivity."[18–20] However, we recommend that patients should be given actual forecast probabilities, rather than binarized forecasts. Although having true probabilities may be challenging initially for some users, they have the distinct advantage of providing more flexibility in interpretation, as patients can individualize their responses to different forecast probabilities. Probabilities also offer patients the opportunity to combine forecasts from other technologies (such as wearables or implanted devices) in a Bayesian manner.[14] Moreover, it is trivial to binarize forecasts for less sophisticated users, if needed or desired (eg, see Figs 4 and 5). Using binary metrics can help to frame the simplified value of a forecasting tool at the individual level. In the case of the AI forecaster, there is significant heterogeneity of outcomes (see Fig 5), reflecting the reality that the present forecasting system can be more accurate for some patients but much less so for others.

There are several important limitations to this work. First, the primary dataset evaluated includes self-reported seizure diary information without physician curation. As a result, the underlying data are subject to quality considerations, such as over-reporting, under-reporting, and misclassification of events.[10,21–24] Nevertheless, these concerns are manifest in all aspects of the clinical care of seizures. Despite recent advances in seizure detection technology,[25,26] there are currently no approved devices for detecting all seizures in outpatients. Moreover, it has been shown in multiple studies that data collected from self-reported seizure diaries produce valuable insights about epilepsy, and a number of results originating in patient-reported databases have been corroborated with other more reliable datasets.[6,17,27,28] Nevertheless, self-reported seizure diary data can differ substantially from seizure patterns characterized objectively using continuous EEG data or external observers.[23] Consequently, the results here must be interpreted as an early step toward the ultimate goal of forecasting all clinical seizures.

It is important to note that forecasting self-reported seizures is not synonymous with forecasting confirmed seizures. In a number of studies, incomplete recording of seizures was common,[24] presumably due to impaired seizure perception. When compared to objective seizure detection, self-reported seizure diaries may account for <50%, on average, of the true seizure count.[24] The only way to mitigate this limitation effectively would be to make use of a dataset comprising long-term objective seizure recordings (for all clinical seizure subtypes) from a large number of patients. Such data do not yet exist, but there are reasons to be optimistic.[29]

In addition, the present AI forecasts were limited to 24-hour blocks in the future. Obviously, patients with high seizure rates (>1 per day) would not benefit from the present AI system. Other forecasting horizons are possible, and may have value to patients. Moreover, this study used all available seizures, meaning that even seizures that would be considered “less burdensome” to some patients were treated equally with those that might be considered more severe. Similarly, seizures that would qualify as status epilepticus according to some definitions[30,31] were treated no differently than others in this analysis. Patients who had very few or very many seizures were equally included in this study, without a minimum or maximum monthly seizure frequency requirement. Many of these design decisions were made to capture the largest possible group of patients to determine the feasibility of electronic diary-based forecasting. Future work is expected to benefit from subgroups enriched with patients most likely to benefit from more patient-specific forecasts.

The reason that RMR is able to achieve AUC > 0.50 in our data is that the cohort represents a mixture of patients, each with a different seizure rate (for additional information see https://tinyurl.com/y847tfxy). This is comparable to flipping a set of weighted coins, where each coin has a different weight to make heads more or less likely from a uniform distribution of weights between always heads and always tails. Here, a coin represents a patient, each coin flip represents 1 day, heads represents “seizure,” and tails represents “no seizure.” If one knew the exact bias of each coin and used that to make the analogous RMR forecast of every flip, the AUC would be expected to achieve the value reported here. Conversely, if all coins had the exact same bias, then the AUC would be expected to be 0.50.

Therefore, due to expected high AUC from the RMR estimate, we also incorporated other metrics to evaluate the AI forecast, such as F1 score, Brier score, and BSS.

The effect size of the AI forecaster could be conceptualized using various metrics, but each of them indicate that the system is not perfect. In contrast to what we have shown here, others have demonstrated a subset of about 20 to 30% of patients who have some powers of self-prediction.[32,33] Here, we find that a much larger set of patients (57% of patients with $AUC_{AI} > AUC_{RMR}$) may benefit from the AI. What has been demonstrated here is not a perfect deployable system. Rather, the system outperforms a random system. We have shown that a majority of patients from a community sample could find benefit from the AI system. Other complementary systems may be required (eg, wearable or implantable devices) to improve the accuracy and utility of a forecaster based on diaries alone.

Left untested in the present work is the issue of clinical utility. Many questions arise based on the potential availability of a seizure forecasting tool. For instance, if a patient receives a forecast that has higher accuracy than a rate-matched chance forecast, will he or she actually benefit in their daily life? Will patients make different safety and medication choices based on the forecast results? As importantly, will patients truly want such forecasts? Moreover, is it safe to provide forecasts to patients when the forecasts are imperfect? These and related questions reveal fundamental limitations of retrospective review of seizure diaries for testing a forecasting system. Future studies should address these questions using a prospective cohort study design.

We look forward to the day when all people with epilepsy can use a reliable and accurate seizure forecasting system. As technology advances, the quality and quantity of objective data available for seizure diary collection and accuracy of training sets can be improved. Meanwhile, the present study offers hope that diary-based forecasting systems may be helpful.

Acknowledgment

This project was made possible by the International Seizure Diary Consortium (https://sites.google.com/site/isdchome/) and the ongoing collaboration between SeizureTracker.com and our laboratory. This work was supported in part by NIH National Institute of Neurological Disorders and Stroke grants T32NS048005, K23NS090900, R01NS102190, R01NS107291 and a BIDMC departmental grant. Portions of this research were conducted on the O2 High Performance Compute Cluster (http://rc.hms.harvard.edu), supported by the Research Computing Group at Harvard Medical School.

References

1. Institute of Medicine (US) Committee on the Public Health Dimensions of the Epilepsies In: England MJ, Liverman CT, Schultz AM, Strawbridge LM, eds. Epilepsy across the spectrum: promoting health and understanding. Washington, DC: National Academies Press, 2012.
2. Janse SA, Dumanis SB, Huwig T, et al. Patient and caregiver preferences for the potential benefits and risks of a seizure forecasting device: a best-worst scaling. Epilepsy Behav 2019;96:183–191. [PubMed: 31150998]
3. Herzog AG, Fowler KM, Sperling MR, Massaro JM. Distribution of seizures across the menstrual cycle in women with epilepsy. Epilepsia 2015;56:e58–e62. [PubMed: 25823700]
4. Baud MO, Kleen JK, Mirro EA, et al. Multi-day rhythms modulate seizure risk in epilepsy. Nat Commun 2018;9:1–10. [PubMed: 29317637]

5. 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. [PubMed: 30219655]
6. Goldenholz DM, Goldenholz SR, Moss R, et al. Is seizure frequency variance a predictable quantity? Ann Clin Transl Neurol 2018;5: 201–207. [PubMed: 29468180]
7. Chiang S, Vannucci M, Goldenholz DM, et al. Epilepsy as a dynamic disease: a Bayesian model for differentiating seizure risk from natural variability. Epilepsia Open 2018;3:236–246. [PubMed: 29881802]
8. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521: 436–444. [PubMed: 26017442]
9. Ferastraoaru V, Goldenholz DM, Chiang S, et al. Characteristics of large patient-reported outcomes: where can one million seizures get us? Epilepsia Open 2018;3:364–373. [PubMed: 30187007]
10. Casassa C, Rathbun Levit E, Goldenholz DM. Opinion and special articles: self-management in epilepsy: web-based seizure tracking applications. Neurology 2018;91:e2027–e2030. [PubMed: 30455263]
11. Chollet F. Deep learning with Python. 1st ed Shelter Island, NY: Manning Publications, 2017.
12. Schelter B, Feldwisch-Drentrup H, Schulze-Bonhage A, Timmer J. Seizure prediction: an approach using probabilistic forecasting In: Osorio I, Zaveri H, Frei M, Arthurs S, eds. Epilepsy: the intersection of neurosciences, biology, mathematics, engineering and physics. Boca Raton, FL: CRC Press, 2011:249–256.
13. Brier GW. Verification of forecasts expressed in terms of probability. Mon Weather Rev 1950;78:1–3.
14. Karoly PJ, Ung H, Grayden DB, et al. The circadian profile of epilepsy improves seizure forecasting. Brain 2017;140:2169–2182. [PubMed: 28899023]
15. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Eur Urol 2015;67: 1142–1151. [PubMed: 25572824]
16. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015; 162:W1–W73. [PubMed: 25560730]
17. Karoly K, Goldenholz D, Moss R, et al. Human epilepsy is characterised by circadian and circaseptan rhythms. Lancet Neurol 2018;17:977–985. [PubMed: 30219655]
18. Snyder DE, Echauz J, Grimes DB, Litt B. The statistics of a practical seizure warning system. J Neural Eng 2008;5:392–401. [PubMed: 18827312]
19. Schelter B, Winterhalder M, Maiwald T, et al. Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. Chaos 2006;16:013108.
20. Winterhalder M, Maiwald T, Voss HU, et al. The seizure prediction characteristics: a general framework to assess and compare seizure prediction methods. Epilepsy Behav 2003;4:318–325. [PubMed: 12791335]
21. Fisher RS, Blum DE, DiVentura B, et al. Seizure diaries for clinical research and practice: limitations and future prospects. Epilepsy Behav 2012;24:304–310. [PubMed: 22652423]
22. Le S, Shafer PO, Bartfeld E, Fisher RS. An online diary for tracking epilepsy. Epilepsy Behav 2011;22:705–709. [PubMed: 21975298]
23. 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. [PubMed: 23642342]
24. Elger CE, Hoppe C. Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection. Lancet Neurol 2018;17: 279–288. [PubMed: 29452687]
25. Onorati F, Regalia G, Caborni C, et al. Multicenter clinical assessment of improved wearable multimodal convulsive seizure detectors. Epilepsia 2017;58:1870–1879. [PubMed: 28980315]
26. Halford JJ, Sperling MR, Nair DR, et al. Detection of generalized tonic-clonic seizures using surface electromyographic monitoring. Epilepsia 2017;58:1861–1869. [PubMed: 28980702]
27. Goldenholz DM, Goldenholz SR, Moss R, et al. Does accounting for seizure frequency variability increase clinical trial power? Epilepsy Res 2017;137:145–151. [PubMed: 28781216]

28. Goldenholz DM, Strashny A, Cook M, et al. A multi-dataset time-reversal approach to clinical trial placebo response and the relationship to natural variability in epilepsy. Seizure 2017;53:31–36. [PubMed: 29102709]

29. 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: ENEURO.0349–17.2017. 10.1523/eneuro.0349-17.2017.

30. Rossetti AO, Lowenstein DH. Management of refractory status epilepticus in adults: still more questions than answers. Lancet Neurol 2011;10:922–930. [PubMed: 21939901]

31. Trinka E, Cock H, Hesdorffer D, et al. A definition and classification of status epilepticus—report of the ILAE task force on classification of status epilepticus. Epilepsia 2015;56:1515–1523. [PubMed: 26336950]

32. Privitera M, Haut SR, Lipton RB, et al. Seizure self-prediction in a randomized controlled trial of stress management. Neurology 2019;93: e2021–e2031. [PubMed: 31645468]

33. Haut SR, Hall CB, LeValley AJ, Lipton RB. Can patients with epilepsy predict their seizures? Neurology 2007;68:262–266. [PubMed: 17242331]

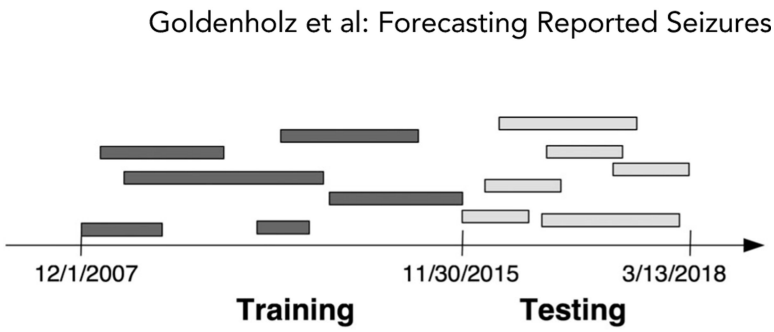

FIGURE 1:

Illustration of training/testing data split. The entire seizure diary database was split based on the cutoff date of November 30, 2015. All diaries that began before that date that met eligibility criteria were included in the training set, but truncated at that date. All diaries that began after that date were included in the testing set. This scheme allowed for testing to occur on patients that the artificial intelligence was not exposed to during training.

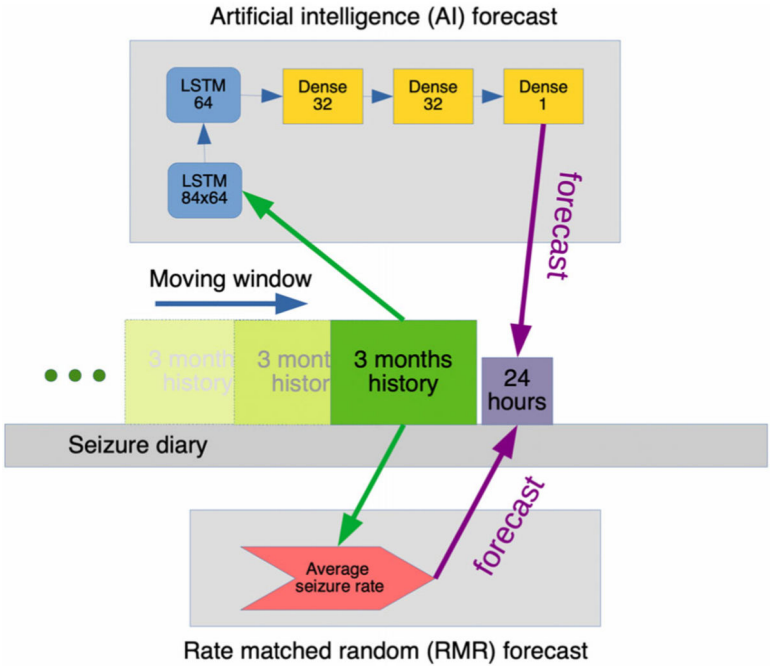

FIGURE 2:

How forecasts were made. A moving window of time, comprising 3 months of history and a 24-hour period, was slid across each patient's seizure diary. For each position, the artificial intelligence (AI) and the rate-matched random (RMR) methods were used to produce one forecast. As shown, the AI method employed a preprocessing stage to generate an 84 × 3 matrix for input, and the RMR method took the first row from that matrix as input. The AI then processed the input using 2 long short-term memory (LSTM) layers, and then 3 densely connected layers to produce a single output forecast. The RMR calculated the average daily seizure rate from the 3-month history to produce an estimate.

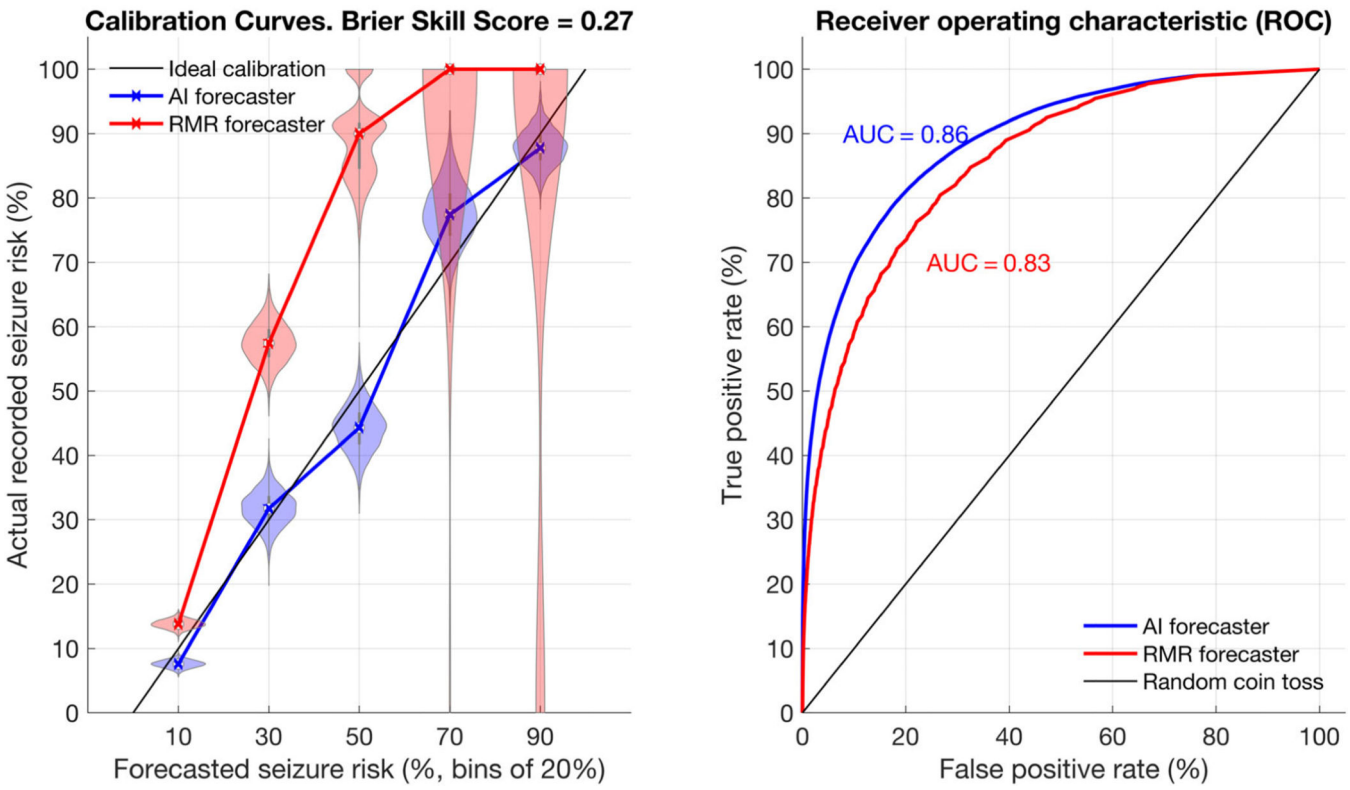

FIGURE 3:

Calibration and receiver operator characteristic (ROC). Calibration shows the relationship between actual recorded seizure probability and seizure forecast values. The rate-matched random (RMR) and artificial intelligence (AI) forecasters are shown in the left calibration curve with violin plots. On the right, the ROC plot compares both forecasters directly. For calibration plots, 5 bins of size 20% were used, meaning forecast values from 0–20, 20–40, 40–60, 60–80, and 80–100 are summarized in the figure along the x-axis. The ideal calibration for a hypothetical perfect forecaster is shown as a dotted black line. The violin plots represent a histogram of values, allowing a clear representation of the spread of values observed. The wider the plot becomes, the more likely a given value was. For instance, when RMR forecasted 20–40%, the true risk often was 60–75%. The calibration curve intersects the median values from each violin plot. RMR shows very poor calibration, particularly at higher forecast values, whereas the AI shows good calibration (ie, close to the idealized calibration curve). Of note, RMR rarely forecasted high-valued (>80%) forecasts, but when it did the true risk was always very low. The ROC plot shows the AI consistently outperforms RMR for any given threshold. AUC = area under the ROC curve.

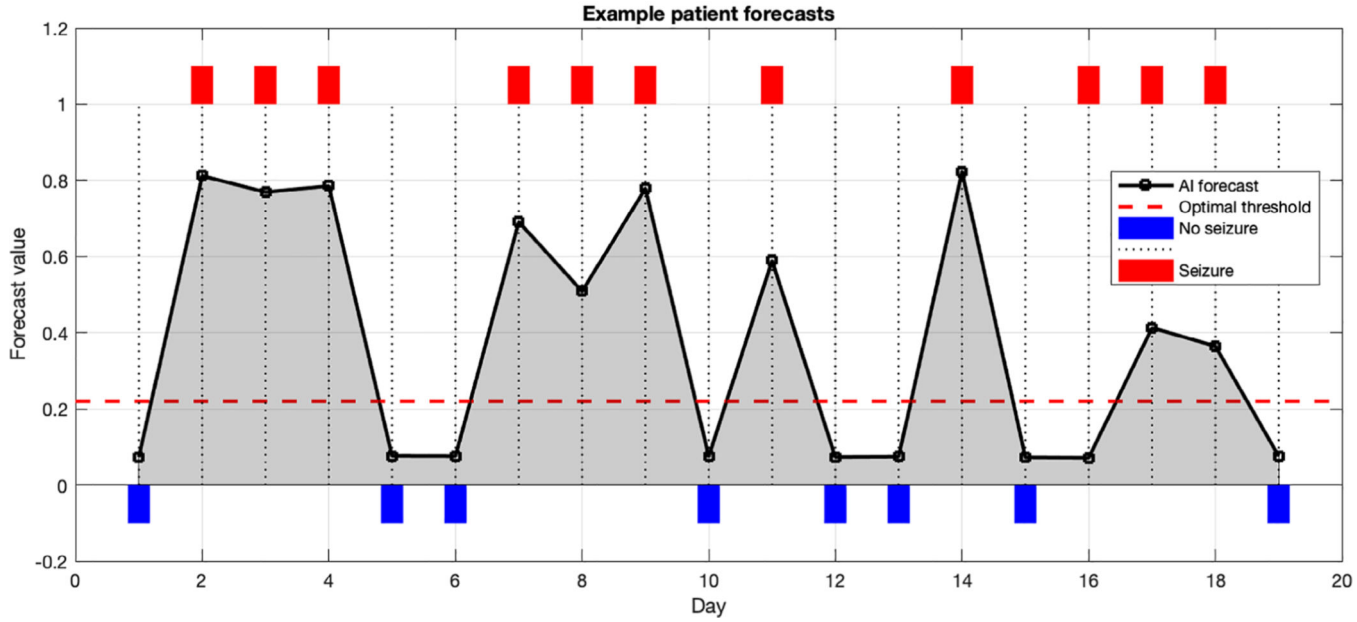

FIGURE 4:

Example patient diary. Shown here is a 19-day diary excerpt from one patient. The black line indicates the forecasts from each day. Red rectangles indicate seizure days, whereas blue rectangles indicate nonseizure days. The red dashed line shows the threshold that this particular patient can optimally use as a cutoff (based on the receiver operating characteristic over this patient's entire 96-day diary). Optimal threshold was selected by finding the threshold that optimally trades off sensitivity and specificity. As seen here, most seizure days have forecasts above the optimal threshold, and most nonseizure days have forecasts below. AI = artificial intelligence.

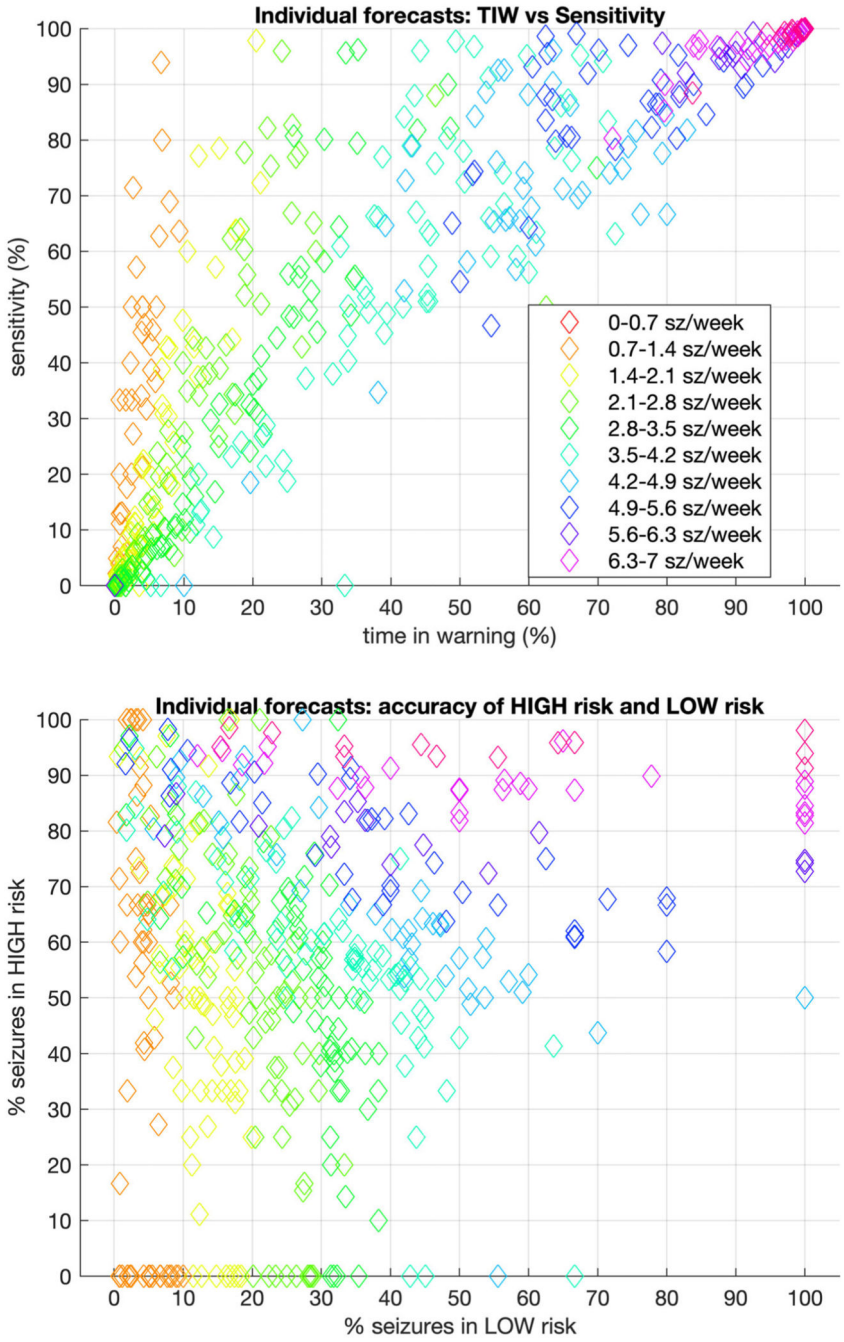

FIGURE 5:

Individual level forecasting metrics for dichotomized forecasts. Taking the optimal threshold cutoff for each patient, forecasts were recasts as “high risk” versus “low risk.” In that context, the upper graph compares time in warning (TIW) to sensitivity, whereas the lower graph compares the accuracy of high-risk to low-risk forecasts. In both, the color of each marker indicates the seizure rate for that patient. The ideal for the upper figure would be very low TIW (which would depend on seizure rate) and extremely high sensitivity. The

ideal for the lower figure would be 0% seizures in low warning, and 100% seizures in high warning. sz = seizures.

TABLE.

Characteristics of the Patient Diaries Included in the Training and Testing Sets

Characteristic	Training	Testing
Patients, n	3,806	1,613
Seizure frequency, per mo: median (min—max)[a]	2.56 (0.04–503.63)	3.67 (0.09–882.63)
Diary duration, days: median (min—max)	364 (85–2,362)	329 (85–1,366)
Total number of seizures: median (min—max)	31 (3–23,838)	43 (3–19,873)
Gender, %, M/F/ unknown	47.3/50.6/2.1	49.0/48.2/2.8
Age, yr, median (min—max)[b]	17.3 (0.6–84.5)	16.4 (0.9–87.6)
VNS usage, n (%)	643 (16.9%)	312 (19.3%)
Brain surgery, n (%)	112 (2.9%)	39 (2.4%)
Epilepsy type, %, focal/nonfocal/unknown[c]	16.8/9.7/73.5	15.6/10.0/74.5

Numbers summarizing across patients are always given as median, and a range between the minimum (min) and maximum (max) value.

[a]Twelve patients in the training set and 0 in the testing set had seizure frequencies < 1/yr, with the minimum rate being 1 seizure every other year.

[b]Age on the last date of the exported dataset (training or testing).

[c]Epilepsy type was defined based on user profiles in the Seizure Tracker database. Focal epilepsy = users who checked any of the list A items AND did not check any of the list B items. Nonfocal epilepsy = users who did not check any of the list A items AND did check any of the list B items. List A: brain tumors, brain trauma, brain hematoma, stroke, brain surgery, brain malformations, tuberous sclerosis. List B: Alzheimer disease, metabolic disorder, genetic abnormalities, electrolyte abnormalities, alcohol or drug abuse, Dravet syndrome, Angelman syndrome, neurofibromatosis, Down syndrome, Aicardi syndrome, Sturge–Weber syndrome, Rett syndrome, hypothalamic hamartoma.

F = female; M = male; VNS = vagal nerve stimulation.