Heuristic Active Learning for the Prediction of Epileptic Seizures Using Single EEG Channel

Joao M. C. Marques[1,2], Hilda A. Cerdeira[1,3], Edgar Tanaka[1], Conrado de Vitor[1] and Paula Gomez[1]

***Abstract*— Predicting epileptic seizure occurrence has long been a goal of the community surrounding it. Accurate prediction, however, is still elusive. This work presents a modified pipeline for the training of seizure prediction systems which aims to attenuate the effects of current data labeling strategies - and consequent data mislabeling of samples that heavily affect classifiers that are trained on it. This paper also presents a seizure prediction system trained following the proposed pipeline, which improved our system's performance by reducing its time-in-warning (TiW) by over 14%, while improving its prediction sensitivity to 72.4%, bringing its performance closer to the state-of-the-art performance (83.1% prediction sensitivity) for systems with similar TiW (41%) [1], while only requiring input from two scalp EEG electrodes - without making use of any variables external to the single EEG channels.**

I. INTRODUCTION

The capacity to predict epileptic seizures and give warnings of their occurrence in advance could give patients more independence from their caregivers. Patients could take appropriate precautions or protection measures, substantially improving their quality of life. Epistemic™ has developed a prediction software based on the idea of the existence of a preictal period with detectable characteristics in an EEG signal. The software is proprietary an is based on non-linear dynamical systems theory [2]. This article presents a heuristic active learning process that improves the performance of Epistemic™'s previously developed epileptic seizure prediction software.

Seizures can be clearly identified from EEG graphs in time by neurologists and other trained professionals. The problem has always been automatizing this identification. Citing R. P. Lesser, "Seizure detection algorithms have been used for several decades... Despite these, seizure detection often seems easy for an EEGer, but hard for a machine" [3].

Likewise, Epistemic™'s prediction software presents results that can be identified by a trained professional to define whether a seizure will happen in the near future. This article proposes a new machine learning algorithm that substitutes the trained professional with an automated methodology and defines whether an EEG point in time presents the anomalies expected in what we define as a pre-seizure. A pre-seizure therefore is an anomaly that happens in a patient's EEG time series that is indicative of an incoming seizure. Combining the active learning framework with Epistemic™'s prediction software results in an automated system for seizure prediction.

While the specifics of Epistemic™'s algorithm are kept as a trade secret, the team will gladly provide the feature sets used in this publication that were generated by the algorithm for the sake of the reproducibility of the results upon request.

[1] Epistemic Gomez & Gomez Ltda. ME - Av. Professor Lineu Prestes 2242, Cietec, sala 244, 05508-000, Cidade Universitria, Sao Paulo, Brazil contato@epistemic.com.br
[2] Escola Politecnica - USP, Universidade de Sao Paulo, Rua Professor Luciano Gualberto travessa 3, no 380, 05508-010 Sao Paulo, Brazil joao.marques@usp.br
[3] Instituto de Fisica Teorica - UNESP, Universidade Estadual Paulista, Rua Dr. Bento Teobaldo Ferraz 271, Bloco II, Barra Funda, 01140-070 Sao Paulo, Brazil cerdeira@ift.unesp.br

II. EPISTEMIC™'S PREVIOUS WORK

Many studies have found that epileptic seizures are not sudden episodes, but they evolve over an interval of time called preictal period [4]. The definition of a preictal period per-se assumes that there are different behaviors in that period when compared to the interictal period. Epistemic™'s software is based on the assumption that anomalies in the preictal stage can be discriminated from EEGs signal behavior in preictal state.

A. *Previous epistemic detection algorithm*

Epistemic™'s previously developed epileptic seizure prediction software performs two different transforms on the EEG time series on a single EEG Channel. The input of the software is sampled and digitized data from patients' electroencephalogram (EEG) time series signals defined by the difference in potential from a unique pair of electrodes. The output are Transforms 1 and 2 as shown in Fig. 1.

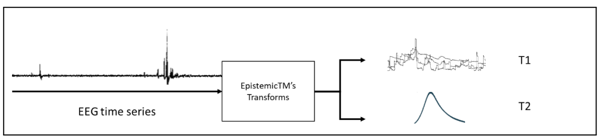

Fig. 1. Epistemic™'s feature generation procedure

Both transforms have the purpose of indicating whether an instability is present at a determined time in the EEG. Epistemic™'s software gives an alert whenever both transforms are indicating an instability simultaneously. For T1 that happens when its three curves display a synchronous increase in amplitude as shown in Fig 2. For T2 that happens when its visual representation, usually similar to a single Gaussian curve, changes format shown in Fig. 3.

This method, however, had a high Time-in-Warning, which inspired the authors to pursue the framework described below.

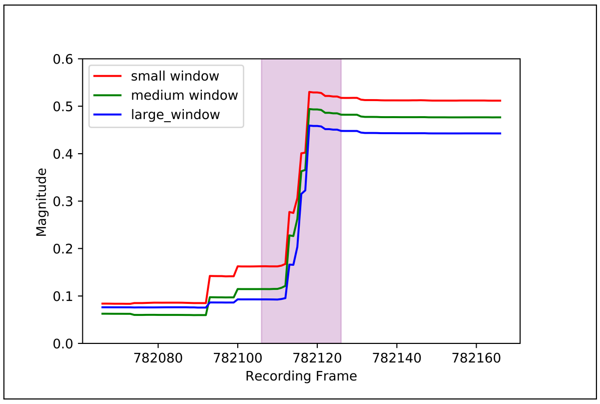

Fig. 2. Synchronous uptick in T1, indicative of a seizure warning

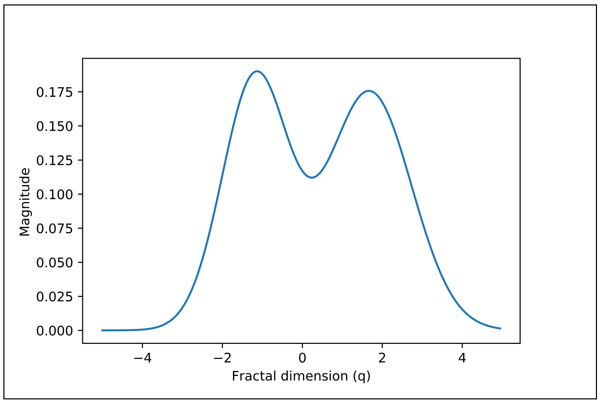

Fig. 3. Altered curve T2, also indicative of a seizure warning

III. DEFINITIONS

A. Formal Statement of the Problem

A peculiarity of the seizure prediction problem based on EEG signals is that while there is literature backing for the presence of discrete alteration of EEG patterns in the period up to several hours before a seizure - also known as pre-ictal -[5] , the precise characterization of such activity is still elusive. As such, most machine learning approaches to seizure prediction have relied on blindly labeling EEG chunks according to their relative position to an observed episode - *i.e* - labeling fixed-length snippets of EEG data as pre-ictal if they corresponded to an hour [6] or fifteen minutes [1] before the clinical onset of a seizure. This procedure, however, leads to massive mislabeling of data, since the alterations for which these studies search are often described as brief [5] and surrounded by seemingly normal brain activity. Hence, when labeling the data according to this procedure, one would be labeling normal snippets as characteristic of abnormal pre-ictal activity, which jeopardizes the performance of most machine learning algorithms, causing increased false positive rates.

This problem of data labeling under uncertain conditions can be formalized in the following manner: Let there $x_i \in \mathbb{X}$ denote the m-dimensional feature vector of the i_{th} sample of the data and X be the $n \times m$ feature matrix of a dataset containing n samples with m features. Let $y_i \in \mathbb{L}$ denote the label attributed to sample i, unknown, belonging to the label space of labels, $\mathbb{L}$. Let $O(x \in \mathbb{X}, k \in \mathbb{N}) \mapsto \mathbb{L}$ denote an oracle that, when given a set of k samples x_i from the feature space $\mathbb{X}$, provides them with k labels from the label space $\mathbb{L}$, which are stored in the label vector $\hat{Y}$. This oracle's probability of error, *i.e.*, the probability that it will mislabel any given sample i is given by an increasing function in the number of labeled examples by the oracle, $E(k \in \mathbb{N}) \mapsto [0,1] \in \mathbb{R}$. This accuracy function models the falling accuracy of the oracle due to the heuristic process of labeling data points based solely on their time distance from any given seizure. This function, however, may also be used in different contexts to model other sorts of errors, which may happen when labeling datasets, like increased error rates which are incurred when labeling large datasets by hand or the increasing prevalence of false positives caused by testing healthy people for a rare disease by using a test that has a non-zero false positive rate. We must then train a classifier on the dataset $(X, \hat{Y})$, labeled by the oracle. The goal of the entire procedure is to maximize performance, measured by some metric, of the classifier $\mathbb{C}(x \in X) \mapsto \mathbb{L}$ when applied to a real unseen dataset (X, Y), whose performance will be verified (in this specific case) using the metrics defined by Snyder et al. [7]. This pipeline can be summarized by figure 4, which depicts the current procedure adopted by machine learning practitioners in the research community - that is - the procedure of blindly labeling all time-samples (be them frame-by-frame) or in finite recordings with respect to to their position to the clinical manifestation of the crisis.

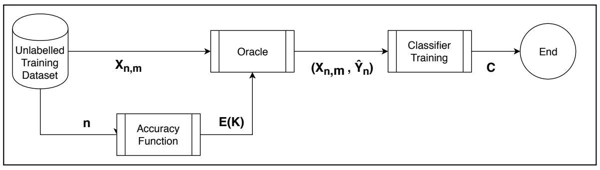

Fig. 4. Current procedure for training seizure prediction algorithms

B. Proposed Heuristic

While this problem cannot possibly have an analytic solution until an oracle Accuracy function $E(k)$ has been defined, one way to approach this problem is by considering it as a *minmax* problem, in which we wish to minimize the number of examples necessary to maximize the performance of the classifier. This *minmax* problem, when associated with the knowledge that pre-ictal signals are surrounded by areas of relative normal brain activity [5], allows this problem to be tackled heuristically. Our proposed heuristic separates this problem into two distinct phases: the data selection and classifier training.

In order to tackle the problem of decreasing oracle accuracy, an alteration is proposed to this pipeline, in which we introduce an entity called the example selector. This entity is responsible for selecting the least possible number of examples that would suffice for the training of a well-performing classifier. One way of doing so is treating the example selector as an outlier detector, which selects only outliers to be used for both training the classifier and classifying after training. This detector must be calibrated, however, in order to avoid being too strict (and consequently miss relevant seizure-indicating outliers) or too permissive, increasing the number of fed into the Oracle, increasing the number of Oracle mislabelings.

This outlier detection calibration problem is then defined as follows: Let there s_{tot} denote the total number of seizures s_i that have been identified in the training files. Let us also define as potentially pre-ictal, ppi_{s_i}, all data points located up to 15 minutes before the clinical onset of a seizure s_i; as potentially interictal, pii, all the points which are outside this interval; and as certainly ictal, ci_{s_i}, all the points contained within the reported clinical onset of seizure s_i and up to ten minutes after its end, to allow for the normalization of brain activity after an event. These labels constitute the label set $\mathbb{L}$. Finally, let Ω be the set of all outliers detected by an outlier detector $\mathbb{D}$. Let us define the representation, r_{s_i}, of a given seizure s_i by equation 1. The coverage $\Gamma(\mathbb{D})$ of an outlier detector $\mathbb{D}$ is defined according to equation 2. Finally, let us define the pseudo-accuracy of the detector Ψ according to equation 3, where $|\cdot|$ is the cardinality of the set.

$$r_{s_i} = \begin{cases} 1, & \text{if } \exists y \in \Omega \text{ such that } y \in [ci_{s_i}, ppi_{s_i}] \\ 0, & \text{else} \end{cases} \tag{1}$$

$$\Gamma(\mathbb{D}) = \left(\frac{\sum_{i=1}^{s_{tot}} r_{s_i}}{s_{tot}} \right) \tag{2}$$

$$\Psi(\mathbb{D}) = \frac{|y \in \Omega \text{ such that } y \neq pii|}{|\Omega|} \tag{3}$$

The problem of finding the optimal detector $\mathbb{D}_{best}$ for this case was then heuristically defined according to equation 4. In practice, the problem was thus defined in order to ensure that the selected detector had the best possible pseudo-accuracy while guaranteeing it would not be overly restrictive to the point of jeopardizing its coverage.

$$\mathbb{D}_{best} = \arg\max_{\mathbb{D}} \left(\Psi(\mathbb{D})\Gamma(\mathbb{D})^2 \right) \tag{4}$$

Once the outlier detector is defined, the labels of each outlier in Ω are then considered as the labels provided by the oracle and these points are used to constitute the labeled training dataset $(X_{train}, \hat{Y}_{train})$ which is then used to train the classifier $\mathbb{C}$. Once tuned and trained, the entire system (detector + classifier) is exposed to a test set - and its warnings are evaluated according to the same criterion as before, *i.e.*, if they occur up to 15 minutes before a seizure or during a seizure they are considered as a true positive and as a false positive otherwise. The resulting pipeline of the proposed solution can be seen in figure 5

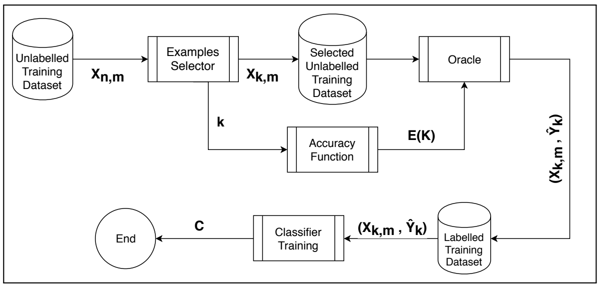

Fig. 5. Proposed procedure for training seizure prediction algorithms

IV. METHODS

A. Data

In this paper, we used the CHB-MIT Scalp EEG Database [8], [9]. This dataset is originally comprised of 23 cases collected from 22 pediatric patients with intractable seizures (5 males aged between 3 and 22 and 17 females aged between 1.5 and 19). Each of the cases, labeled chb01-chb23, contains between 9 and 42 continuous .edf files from the same subject. Each recording was made at 256Hz with 16-bit resolution. The complete dataset is comprised of 664 .edf files, 129 of which contain seizures. Besides running the files through Epistemic™'s proprietary software, a few considerations were done when pre-processing the data to be used in this paper. Firstly, data from patient CHB24 was discarded, as its files do not contain the times when each of the files were collected - thus making it impossible to establish if any given point in them occurred before a seizure. For the same reason, files that were isolated in time - that is, with relevant gaps in time before and after them, like file chb08_29, were also disconsidered in this analysis, as the large gaps between themselves and the following files, and the lack of annotations regarding the time in-between them, make it impossible to accurately label their points. Other than that, no additional preprocessing was done to the files. The files were then split by patient, then split again in two categories, that of files that contained seizures and that of files that did not. Each category was then chronologically split in a proportion of 80-20, assigning the first 80% of the files as training files and the remaining 20% as test files. The training files were then again split chronologically in the same 80-20 proportion, with the initial 80% of the training files being assigned to training and the remaining 20% to validation. This chronological split was done in order to avoid data leakage and ensure the robustness of the validation process - *i.e.* avoiding validating or testing using data from the past on a model that was trained on future data.

B. Outlier Detector

As mentioned previously,Epistemic™'s prediction software creates visual representation that indicates anomalies in EEG time-signals. These are curves for every time interval whose shapes are related to the possibility of a pre-seizure state of the EEG. For instance, figure 6 indicates a curve highly associated with pre-seizures, figure 7 indicates one that is a strong indication of no upcoming seizures and figure 8 represents a curve whose warning status is uncertain. In addition, we have come to believe that the evolution of these curves in time is also valuable for detection, as the evolution of the topography of the curve is also correlated with the occurrence of a seizure. It is, therefore, also essential to find the optimal number of frames of past conditioning curves (T2) to fully characterize a given data point in the EEG. These curves are numerically calculated in real-time and they are not easily parametrized to substitute visual inspection. However, they can be satisfactorily described by a set of few points, namely the local extrema. The coordinates of these points in the graph were then used to construct the feature vector v_i of each of the frames in the EEG, namely $v_i = [x_{peak1}, y_{peak1}, x_{peak2}, y_{peak2}, x_{valley1}, y_{valley1}]$.

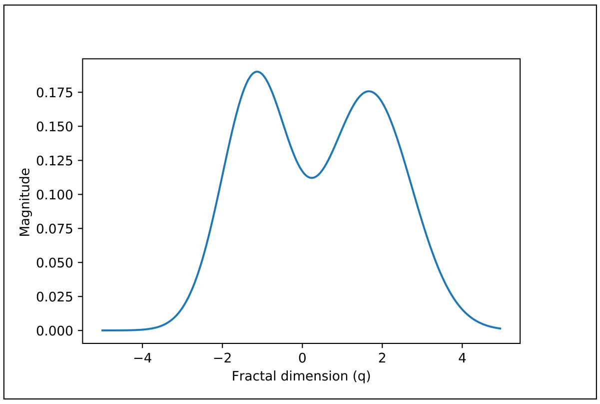

Fig. 6. Conditioning Curve T2 highly related to seizure warnings

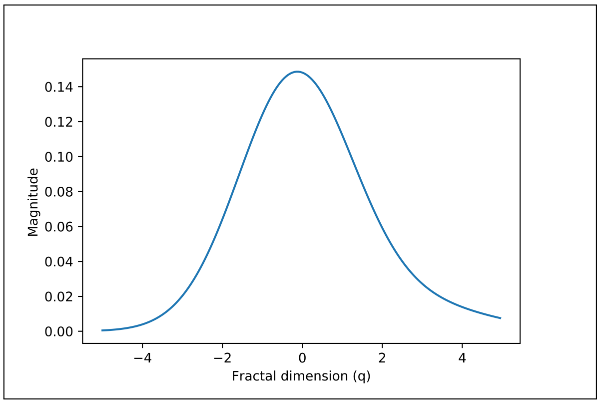

Fig. 7. Conditioning Curve T2 not related to seizure warnings

While there exist several viable strategies to perform outlier detection, such as Isolation Forests [10], Long Short Term Memory Networks [11], Hierarchical Temporal Memory based algorithms [12], we have decided to implement a simpler strategy inspired by two-tailed statistical tests which are easier to interpret. Let us first define the set of parameters used in this detector. Let there TW denote the number of frames of the conditioning curves that are considered at any given moment when evaluating the anomaly status of a given step in the EEG. Let $skip$ denote the number of frames skipped between consecutive frames when constructing the following observation window. That is, for every frame of the EEG i, a feature vector x_i is build according to (5), defining the window set of each sample i, $\mathbb{W} = [i, i-skip, \cdots, i-TW \cdot skip]$.

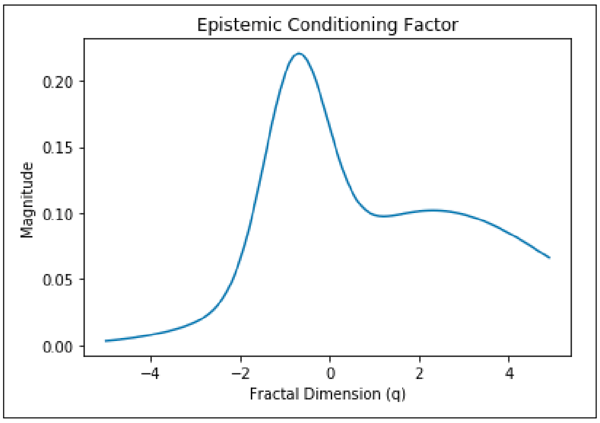

Fig. 8. Conditioning curve T2 with unkown relation to seizure warnings

$$x_i = \left[v_i, v_{i-skip}, v_{i-2\cdot skip}, \cdots, v_{i-TW\cdot skip}\right] \quad (5)$$

Let us define as well μ_{d_i}, σ_{d_i} as the rolling mean and rolling standard deviation of the distance between peaks d_i within each frame, *i.e.*, $d_i = x_{peak2_i} - x_{peak1_i}$, defined over a window of stability of 20,000 frames. The outlier detector was then defined according to (6), where DF is the minimum fraction of samples in a window that have to be outliers for the total feature vector to be considered an outlier, out_i is defined in (7), where DFM is the minimum number of standard deviations from the mean necessary to consider the differences between peaks an outlier and MR is the minimum ratio between the y coordinates of the peaks so that this outlier is considered relevant.

$$\mathbb{D}(x_i) = \begin{cases} 1, & \text{if } \frac{\sum_{i\in\mathbb{W}}(out_i)}{TW+1} > DF \\ 0, & \text{else} \end{cases} \quad (6)$$

$$out_i = \begin{cases} 1, & \text{if } (d_i - \mu_{d_i} > DFM \cdot \sigma_{d_i}) \,\&\, \left(\frac{y_{peak2_i} - y_{valley_i}}{y_{peak1_i} - y_{valley_i}} > MR\right) \\ 0, & \text{else} \end{cases} \quad (7)$$

Together, TW, $skip$, DF, DFM, MR are the set of parameters to be adjusted for this outlier detector. As mentioned previously, this setup is highly interpretable, since DFM can be associated intuitively with a p-value for the observed distance between peaks and TW and $skip$ provide us with an intuitive sense of a relevant duration window to be observed. The adjustment of these hyperparameters was done through a process of Bayesian optimization performed exclusively on the training files, using the Python implementation of the BayesianOptimization library [13], inspired by [14]. The results of this process are reported in the following section.

C. Classifier

For the seizure prediction task a standard gradient boosting classifier was used, CatBoost [15], namely its python implementation. One of the advantages of CatBoost is that its performance is robust to hyperparameter choice, unlike other boosting libraries such as XGBoost [16], [15]. In order to ease the evaluation of the model during training, the classification problem was turned into a binary classification problem, with frames labeled as ppi_{s_i} or ci_{s_i} being labeled as True, while the remaning frames were labeled as False. This was done so that the validation metric used, Area Under the Receiver Operating Characteristic curve (AUROC) could be used unequivocally, which has several benefits as reported by [17], particularly when dealing with a classification problem with severe class imbalance. We note, however, that during the evaluation procedure, only ppi_{s_i} points are considered as a True prediction.

D. Evaluation

Once the system has been trained to optimize the AUROC, we then evaluate both the validation and testing performance following the procedure defined by [7]. In their paper, Snyder et al. [7] describe a set of 4 performance metrics for seizure prediction systems, namely: Prediction Sensitivity, Time-in-Warning (TiW), Improvement-Over-Chance (IoC), as well as the p-value associated with that IoC. The prediction sensitivity is the ratio between the number of seizures that were predicted by your system within the specified prediction window (in our case, of 15 minutes). Considering that every time a system emits an alarm indicating an upcoming seizure the detector remains in a warning state for the duration of its prediction window, Snyder et al. define the TiW as being the fraction of total experiment time that a system spends in a state of warning. For instance, should a system remain in warning for 2 hours during a 10-hour experiment, that system's TiW would be of 0.2. Considering the TiW metric of a given system, Snyder et al. define a random system with an identical expected TiW to that system. This random system has an expected sensitivity associated to it. The difference between the sensitivities of system being evaluated and the random system with identical TiW is defined as being the IoC of the system.

V. RESULTS

A. Outlier Detection Results

The optimization of the outlier detector was done using the reinforcement learning library for Gaussian optimization BayesianOptimization [13], inspired by [14]. The optimal parameters for the outlier detector can be found in tableI, while its performance in both the validation and test sets can be found in Fig. 9, where it can be seen that coverage above 90% was obtained by our proposed heuristic, while keeping the pseudo-accuracy above 3% even in the files within the testing dataset, which were not used in the optimization, following standard machine learning procedures to avoid data leakage. The outliers identified by this detector were then used in the following step of training the seizure prediction algorithm.

B. Supervised Classification Results

We then trained the classifier on the labeled outlier database created in the previous step. For this step we used a Catboost Classifier[15] with default parameters. This classifier's performance can be seen in Fig. 12. The validation

TABLE I
OPTIMAL OUTLIER DETECTOR PARAMETERS

Skip	TW	DFM	MR	DF
0	10	3.1	0.06	0.45

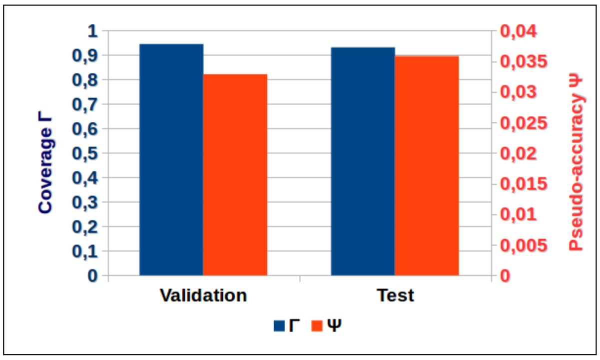

Fig. 9. Outlier detector performance metrics

performance is, as expected, higher than test performance. That was expected, since the validation dataset takes place closer in time to the training dataset, and it is known that the circadian rhythm has As a final step in the calibration step we then calibrated a leaky accumulator to the end of our classifier, to help reduce the time-in-warning of the system, as proposed by [1]. In order to do so we looked at cumulative histograms of our validation data for both the probability assigned by the classifier as well as the number of consecutive warnings, which can be seen, respectively, in figures 10 and 11, from which we inferred that the optimal parameters for this accumulator were a probability higher than 0.5 and a number of at least 10 consecutive frames of warning in order to consider a warning was indeed triggered. Once these parameters had been defined, we finally evaluated our complete model- Outlier Detector + Classifier + Leaky Accumulator according to the metrics defined in [7].

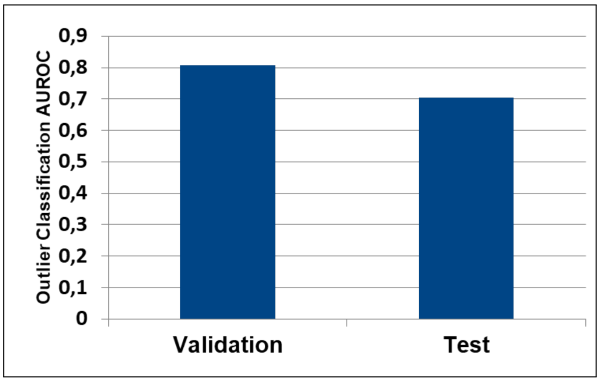

Fig. 10. Trained Classifier AUROC performance for outlier classification task

C. Seizure Prediction Results

The final results obtained by our combined system are reported in table II as an aggregate of the performance for all patients.

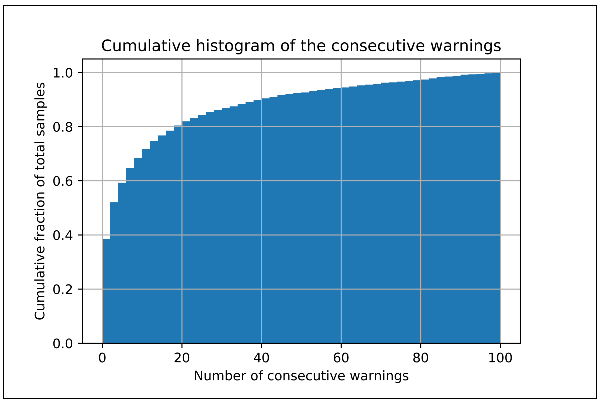

Fig. 11. Cumulative normalized histogram plot of consecutive warnings

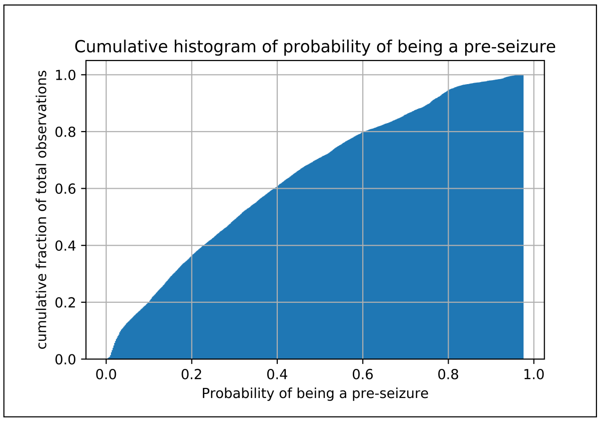

Fig. 12. Cumulative normalized histogram of classifier-assigned probabilities

TABLE II
FINAL PREDICTOR PERFORMANCE METRICS

Database (Algorithm)	s_{tot}	IoC	Sensitivity	TiW	p-value (IoC)
Validation(new)	18	0.45	0.88	0.46	< 0.0002
Test(new)	29	0.34	0.74	0.41	< 0.0003
Full(old)	169	0.15	0.67	0.55	< 0.0005

VI. DISCUSSION

As can be seen in table II, our proposed system achieved (marked as "new") a test prediction sensitivity of 74% with a TiW of 41%. This test performance more than doubles the IoC of our previous algorithm (denoted in the table as "old", in the last row), increasing sensitivity by 10% and reducing the system's TiW by 25%. When compared to the state-of-the-art for systems with similar TiW, such as the one reported in [1], which presented 83.1% prediction sensitivity with 41% TiW , our system presents a comparable, though slightly inferior performance. Our proposed system also has the added benefit of relying solely on two EEG electrodes to perform its prediction task, as well as working with scalp EEGs, rather than implanted ones, which is considerably less invasive for patients and less likely to result in medical complications. This proposed system, is therefore, an important step in developing a practical portable seizure prediction system that is viable and minimally invasive. It is also notable that only a single model was trained using the data from all patients combined. It is known, however, that many of the features governing seizure prediction tend to be patient specific [1]. It is, therefore, expected that if the models were to be trained on a patient-specific basis, our system's performance would be further improved.

VII. FUTURE RESEARCH

The following steps in this research will be to apply this method to bigger databases of EEG signals which has a wider range of patients and a database which has longer-term data for a smaller number of patients. Larger databases will allow us to verify the effects of tuning the outlier detector on a patient-specific manner on the overall performance of the outlier detector system, as well as provide us with a larger set of epilepsy types, to ensure the robustness of our system.

REFERENCES

[1] I. Kiral-Kornek, S. Roy, E. Nurse, B. Mashford, P. Karoly, T. Carroll, D. Payne, S. Saha, S. Baldassano, T. O'Brien, D. Grayden, M. Cook, D. Freestone, and S. Harrer, "Epileptic Seizure Prediction Using Big Data and Deep Learning: Toward a Mobile System," *EBioMedicine*, vol. 27, pp. 103–111, 1 2018.

[2] H. A. Cerdeira and P. Gomez, "Early Detection of Epileptic Seizures," in *IWSP7:Epilepsy Mechanisms, Models, Prediction and Control*, (Melbourne), 2015.

[3] R. P. Lesser and W. R. S. Webber, "Seizure detection: reaching through the looking glass.," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 119, pp. 2667–8, 12 2008.

[4] Shufang Li, Weidong Zhou, Qi Yuan, and Yinxia Liu, "Seizure Prediction Using Spike Rate of Intracranial EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, pp. 880–886, 11 2013.

[5] B. Litt, R. Esteller, J. Echauz, M. D'Alessandro, R. Shor, T. Henry, P. Pennell, C. Epstein, R. Bakay, M. Dichter, and G. Vachtsevanos, "Epileptic Seizures May Begin Hours in Advance of Clinical Onset: A Report of Five Patients*," in *Applications of Intelligent Control to Engineering Systems*, pp. 225–245, Dordrecht: Springer Netherlands, 2009.

[6] S. Jukic and A. Subasi, "A MapReduce-based rotation forest classifier for epileptic seizure prediction," *CoRR*, vol. abs/1712.0, 2017.

[7] D. E. Snyder, J. Echauz, D. B. Grimes, and B. Litt, "The Statistics of a Practical Seizure Warning System," *Journal of neural engineering*, vol. 5, pp. 392–401, 12 2008.

[8] A. H. Shoeb, *Application of machine learning to epileptic seizure onset detection and treatment.* PhD thesis, Massachusetts Institute of Technology, 2009.

[9] A. L. Goldberger, L. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Circulation," *Discovery*, vol. 101, no. 23, p. 1, 1997.

[10] K. M. Ting, F. T. Liu, and Z. Zhou, "Isolation Forest," in *2008 Eighth IEEE International Conference on Data Mining(ICDM)*, vol. 00, pp. 413–422, 2008.

[11] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal, "Long short term memory networks for anomaly detection in time series," in *Proceedings*, p. 89, Presses universitaires de Louvain, 2015.

[12] A. Lavin and S. Ahmad, "Evaluating Real-Time Anomaly Detection Algorithms – The Numenta Anomaly Benchmark," in *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pp. 38–44, IEEE, 12 2015.

[13] F. Nogueira, "Bayesian Optimization: Open source constrained global optimization tool for Python," 0.

[14] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," in *Advances in neural information processing systems*, pp. 2951–2959, 2012.

[15] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," *arXiv preprint arXiv:1706.09516*, 2017.

[16] T. Chen and C. Guestrin, "{XGBoost}: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, (New York, NY, USA), pp. 785–794, ACM, 2016.

[17] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms," *Pattern Recognition*, vol. 30, pp. 1145–1159, 7 1997.