Methodological Issues in Evaluating Machine Learning Models for EEG Seizure Prediction: Good Cross-Validation Accuracy Does Not Guarantee Generalization to New Patients

Sina Shafiezadeh [1,*], Gian Marco Duma [2], Giovanni Mento [1,3], Alberto Danieli [2], Lisa Antoniazzi [2], Fiorella Del Popolo Cristaldi [1], Paolo Bonanni [2] and Alberto Testolin [1,4,*]

1   Department of General Psychology, University of Padova, 35131 Padova, Italy
2   Epilepsy and Clinical Neurophysiology Unit, Scientific Institute, IRCCS E. Medea, 31015 Conegliano, Italy
3   Padova Neuroscience Center, University of Padova, 35131 Padova, Italy
4   Department of Mathematics, University of Padova, 35131 Padova, Italy
*   Correspondence: sina.shafiezadeh@studenti.unipd.it (S.S.); alberto.testolin@unipd.it (A.T.)

Abstract: There is an increasing interest in applying artificial intelligence techniques to forecast epileptic seizures. In particular, machine learning algorithms could extract nonlinear statistical regularities from electroencephalographic (EEG) time series that can anticipate abnormal brain activity. The recent literature reports promising results in seizure detection and prediction tasks using machine and deep learning methods. However, performance evaluation is often based on questionable randomized cross-validation schemes, which can introduce correlated signals (e.g., EEG data recorded from the same patient during nearby periods of the day) into the partitioning of training and test sets. The present study demonstrates that the use of more stringent evaluation strategies, such as those based on leave-one-patient-out partitioning, leads to a drop in accuracy from about 80% to 50% for a standard eXtreme Gradient Boosting (XGBoost) classifier on two different data sets. Our findings suggest that the definition of rigorous evaluation protocols is crucial to ensure the generalizability of predictive models before proceeding to clinical trials.

Keywords: seizure prediction; epilepsy; electroencephalography; feature extraction; machine learning; signal processing; artificial intelligence; model validation

1. Introduction

Epilepsy, a severe neurological disease that leads to recurrent seizures, affects more than 65 million people worldwide, with an incidence rate of 61.44 per 100,000 person-years [1,2]. Although antiepileptic drugs can reduce clinical complications and mortality rates, 30% of patients are refractory to such drugs [3], thus urging the development of alternative treatments. The unpredictable nature of seizures increases the risk of injury and psychosocial disability, significantly affecting the patient's quality of life [4]. However, evidence suggests that specific alterations in brain dynamics can be observed before epileptic attacks [5]. This discovery spurred the interest of academic centers and medical companies in building devices to anticipate seizures, primarily by analyzing the electroencephalogram (EEG) [6,7]. Monitoring devices would allow patients to avoid dangerous situations and plan the administration of preventive treatments, such as electrical stimulation or targeted drug delivery, with much greater precision.

Seizure *prediction* aims to anticipate an upcoming seizure before it clinically manifests. This task differs significantly from seizure *detection*, a simpler binary classification problem that requires discriminating between normal and seizure brain activity. However, predicting seizures using EEG analysis is challenging, as EEG manifestations vary widely between patients and even within the same patient. Ten years ago, the prototype of the first implanted seizure advisory system was tested in human patients [8]. The system could

identify periods of low, moderate and high probability of seizures and was developed with precise target performance criteria, such as reaching a sensitivity of high probability warnings greater than 65% and an accuracy at least higher than the probability prediction of random events. Less invasive devices that rely on scalp EEG recordings in the channel or source space have also been proposed. One hypothesis is that functional connectivity patterns can reveal information about the dynamics of the epileptic brain, which can be used to predict the onset and locations of seizures [9].

In recent years, the increasing success of artificial intelligence techniques in clinical diagnosis [10] and disease forecasting [11] has revived the interest in leveraging machine learning for the challenging task of seizure prediction (for reviews, see [12–14]). One common approach involves extracting various descriptive features from EEG recordings and using them to train machine learning algorithms to identify time blocks proximal (e.g., 2 h, 1 h, or 30 min) to an upcoming seizure. These features include time- and frequency-based indexes, information theory measures, and sophisticated metrics derived from dynamical systems theory [15,16]. For example, some studies extracted 22 linear univariate features from 6 EEG channels and achieved an average sensitivity of approximately 73% using classifiers such as support vector machines (SVMs) and artificial neural networks [17,18]. A similar approach trained SVMs with a reduced set of bivariate features and achieved slightly better accuracy [19]. An alternative method based on the extraction of histogram bins combined with Gaussian mixture models reported an average sensitivity of 88% [20].

The application of machine learning techniques to predict seizures has thus shown great potential. However, concerns have been raised about the possibility of robustly detecting preictal states with respect to the reproducibility and statistical validation of these techniques [13]. For example, despite achieving impressive performance in various epilepsy research applications, follow-up validation studies have often shown that the reported findings may not accurately reflect the robustness of the methods [21]. This issue is especially evident when models are trained using small data sets, which is common in epilepsy research. In such cases, machine learning algorithms are prone to overfitting [22], which occurs when the model learns irrelevant patterns originating from noise in the data. Although performance is high in the training set, an overfitted model will fail to predict future observations; that is, it will not be able to generalize to unseen data. For example, in medical applications, it could happen that a model will learn to perform a classification task by relying on patient-specific characteristics that are not representative of the clinical population or by detecting spurious features related to the measurement tools [23]. It is therefore important to evaluate machine learning models on left-out test samples, which should be as independent as possible from those used during the training phase.

In this work, we evaluate the performance of seizure prediction models based on standard machine learning algorithms by systematically comparing two cross-validation methods. To allow comparison with existing approaches, various supervised classifiers were trained with a commonly used set of features extracted from scalp EEG recordings. Two data sets were evaluated: the classic benchmark of the seizure prediction literature (CHB-MIT) and a new data set collected by the Epilepsy and Clinical Neurophysiology Unit of the Eugenio Medea IRCCS Hospital in Conegliano (Italy). The latter will be called "our data set" throughout the paper.

The first part of the Materials and Methods section illustrates the details of the two EEG data sets, including the data annotation procedure. Then, we describe the EEG signal processing pipeline and the feature extraction stage. The machine learning models considered and the characteristics of the two evaluation procedures are described at the end of the Materials and Methods section.

Our best model achieved an average precision of 79% on the CHB-MIT data set and 82% on our data set. However, these numbers dropped to approximately 50% (chance level) when the models were evaluated using a more challenging leave-one-patient-out validation scheme, where the entire set of recordings from single patients was iteratively excluded from the training phase. We conclude our article with a critical discussion of our

findings and propose research directions to build more robust predictive models of seizure occurrence.

2. Materials and Methods

2.1. EEG Data Sets

In this study, we used two EEG data sets containing long-term continuous multichannel recordings, which were obtained using the international standard 10–20 EEG scalp electrode positioning system, with a sampling rate of 256 Hz. We randomly selected 8 patients from the CHB-MIT data set and 10 patients from our data set to further demonstrate that high accuracy can be achieved even with a small subset of patients, as reported in similar studies [24,25]. A data cleaning phase was carried out to ensure that at least one recording containing at least 4 hours of monitoring was available before a seizure for all patients, and seizures separated by intervals of less than 3 h were removed [26]. The first data set contained recordings of eight patients (two males, five females, and one unknown) selected from the CHB-MIT data set [27]. It includes 28 seizures, recorded using the 22 common EEG channels listed in Table A1. Our data set instead contained recordings of 10 patients (4 males and 6 females) totaling 40 seizures. These recordings were obtained using 20 common EEG channels, and their distribution in the scalp is represented in Figure 1.
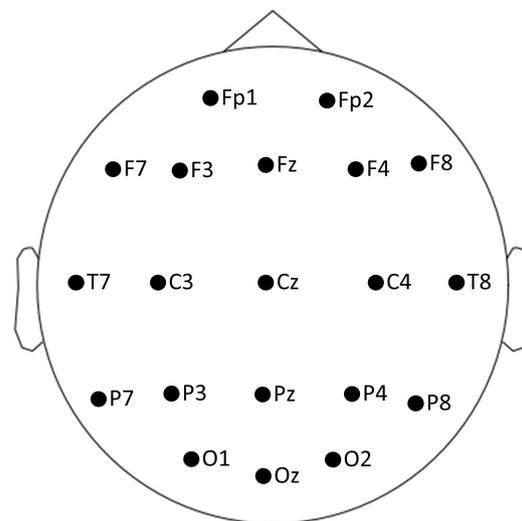

Figure 1. Scalp positioning of the 20 common EEG channels used in our data set.

2.2. Data Labeling

The EEG signal recorded in epileptic patients can be categorized into four main stages: (1) the interictal state, which is a period of regular brain activity between two consecutive seizures; (2) the preictal state, which refers to the period included between approximately 60 and 90 min before seizure onset; (3) the ictal state, which is when the seizure occurs; and (4) the postictal state, which is the period immediately following a seizure for a few minutes [28–30]. The beginning and end of the ictal state were manually marked by the clinicians A.D. and P.B. based on the electroclinical and video-recorded information derived from video EEG monitoring. Our prediction task was designed to discriminate between the portion of the preictal state immediately preceding the seizure and a normal signal recorded during the interictal state (see Figure 2). To this end, we defined two binary categories based on the distance from the upcoming seizure: class 1 contained signals sampled from the time window between 0 and 30 min before the seizure, while class 2 had signals sampled from 30 min randomly selected from the interictal state activity.
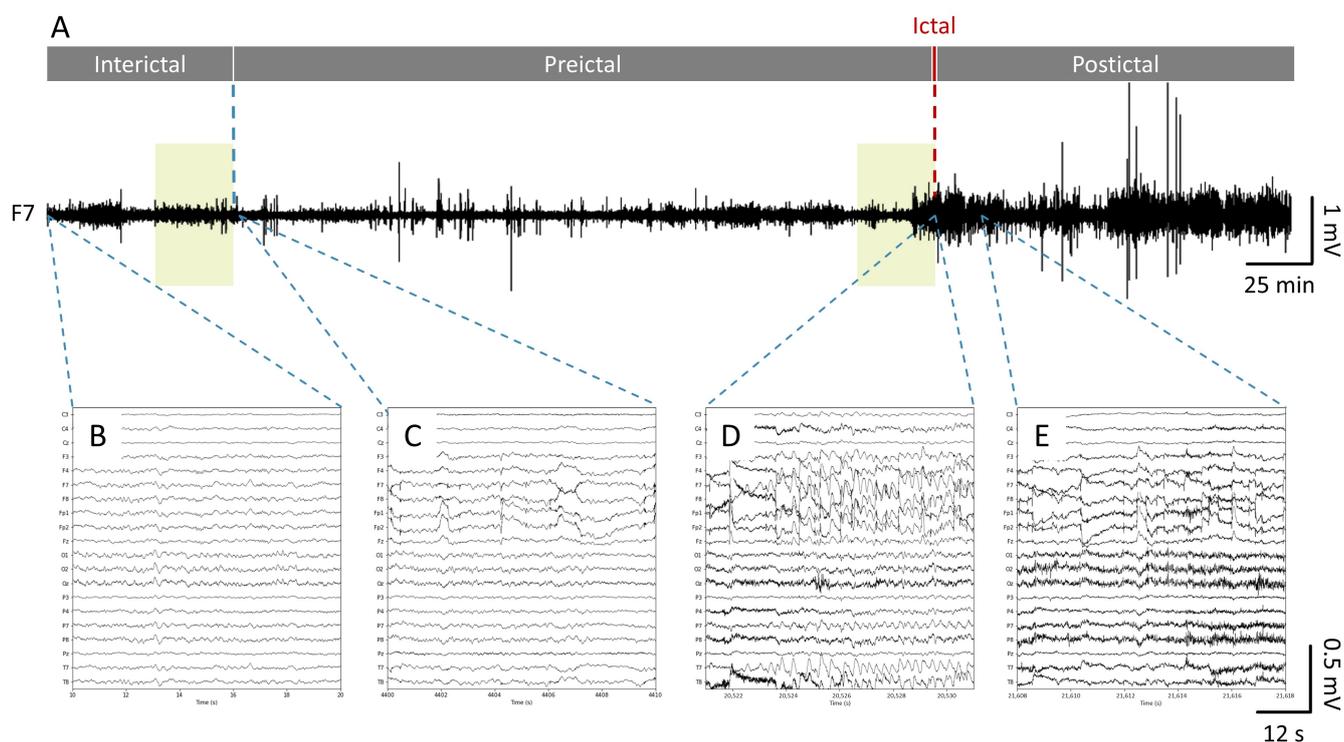
Figure 2. Standard segmentation for the EEG recording of an epileptic patient. (**A**) The trace depicts 480 min of recording from the F7 channel during a seizure divided into four main stages: interictal, preictal, ictal, and postictal. The green areas represent the 30 min from the preictal and interictal states used for the binary prediction task. Panels (**B–E**) illustrate magnifications of 20 s of recordings from 20 channels at the beginning of the interictal, preictal, ictal, and postictal states, respectively.

2.3. Data Preprocessing and Feature Extraction

The signal was initially processed by applying Notch filters at 50 and 100 Hz to eliminate power line interference, and a high-pass filter at 1 Hz was implemented to remove the DC offset and baseline fluctuations on all signals using the MNE package Python version 3.8.5 [31,32]. Next, a low-pass filter was implemented at 125 Hz to maintain higher frequencies that could characterize abnormal brain activity [33,34]. After preprocessing, the EEG signal was divided into 5 s non-overlapping time windows, and a set of 53 commonly used features [16] was extracted using the MNE-Features subpackage (see Table A2). These features contained time domain features such as the mean, variance, standard deviation, skewness, and kurtosis, as well as essential frequency domain features such as the power spectral density, spectral entropy, and Hjorth parameters (mobility and complexity).

2.4. Machine Learning Models

The primary purpose of this study was to predict seizures through a binary classification task, which would allow raising warning alarms before the appearance of a seizure. Different supervised machine learning models, including support vector machines (SVMs), decision trees, k-nearest neighbors, logistic regression, naive Bayes, random forests, and gradient boosting, were applied to discriminate between the preictal and interictal states. The best-performing model was selected, considering the highest average accuracy in all patient data. The best model was XGBoost, which is well suited to classifying large-scale data thanks to its scalability and parallelization [35].

The Optuna framework [36] was used to tune the hyperparameters and improve the performance of the default settings. The search space included the booster type (gbtree, gblinear, or dart), lambda (1×10^{-8} to 1.0), and alpha (1×10^{-8} to 1.0) parameters. Additionally, depending on whether the booster was gbtree or dart, the max_depth (1 to 9),

eta (from 1×10^{-8} to 1.0), gamma (from 1×10^{-8} to 1.0), and grow_policy (depthwise and lossguide) were also tuned. When the dart booster was selected, the sample_type (uniform and weighted), normalize_type (tree and forest), rate_drop (from 1×10^{-8} to 1.0), and skip_drop (from 1×10^{-8} to 1.0) were finally searched to find more effective values. This dynamic space search was iterated 30 times to find optimized hyperparameters.

2.5. Model Evaluation

The performance was measured considering the accuracy (ACC), sensitivity (SEN), and specificity (SPE), which vary between 0 and 1:

$$Accuracy = ((tp + tn)/(tp + tn + fn + fp)), \tag{1}$$

$$Sensitivity = (tp/(tp + fn)), \tag{2}$$

$$Specificity = (tn/(tn + fp)), \tag{3}$$

where *tp* indicates true positives, *tn* indicates true negatives, *fn* indicates false negatives, and *fp* indicates false positives.

The randomized cross-validation (RCV) and leave-one-patient-out (LOO) validation methods were compared to evaluate the performance of the model. In RCV, the data samples were randomly split into training and test sets using fivefold cross-validation. The performance metrics were computed separately for each patient in the test set, and the process was repeated five times. The average performance metrics were then reported [24]. On the other hand, in LOO validation, each patient's recordings were iteratively left out of the training set. The performance of the model was evaluated separately for each patient, and the results were compared for both data sets to assess the differences between the two validation strategies.

3. Results

3.1. RCV

The RCV results for the CHB-MIT data set are presented in Table 1. The findings show that chb01 had the highest ACC (91.33%) and SEN (100%), indicating that the model could accurately identify the correct preictal and interictal states of this patient. Furthermore, three patients had 100% SPE, reflecting the ability of the model to correctly identify interictal states. On average, the model achieved an ACC of 78.75%, SEN of 64.48%, and SPE of 78.20% in all patients. In contrast, chb04 had the lowest ACC (68.85%), and chb05 had the lowest SEN (33.28%), indicating that the model's performance was relatively lower for these patients. For chb01, the model failed to identify the interictal states, probably due to the very short interictal time interval, leading to an SPE of 0%. It is interesting to notice that the prediction performance can significantly differ among patients. One possible explanation is that the considered cohort might contain patients with heterogeneous types of seizures, which can manifest with a variety of EEG signatures that might be more or less challenging to detect.

The RCV results in our data set were similar, as shown in Table 2. The best accuracy of 86.23% was achieved for p8, whereas p4 had the highest sensitivity of 83.40%. Eight patients obtained a specificity of 100%. The mean values for ACC, SEN, and SPE in all patients were 81.68%, 64.66%, and 96.12%, respectively. Among the patients, p10 had the smallest ACC and SEN with 73.24% and 49.80%, respectively, while p7 had the minimum SPE with 72.39%. These results demonstrate that the RCV validation method can achieve high prediction performance in both data sets.

Table 1. RCV results for the CHB-MIT data set. Bold format highlights maximum values.

Patient Number	Gender	Total Seizures	ACC (%)	SEN (%)	SPE (%)
chb01	f	3	**91.33**	**100.00**	0.00
chb04	m	3	68.85	66.59	71.29
chb05	f	3	72.18	33.28	89.38
chb06	f	5	81.81	60.05	97.37
chb07	f	2	75.95	49.90	**100.00**
chb12	f	4	86.05	75.08	**100.00**
chb15	m	6	76.53	80.73	67.57
chb24	-	2	77.32	50.20	**100.00**
		Average	**78.75**	**64.48**	**78.20**

Table 2. RCV results for our data set. Bold format highlights maximum values.

Patient Number	Gender	Total Seizures	ACC (%)	SEN (%)	SPE (%)
p1	m	10	82.90	68.99	**100.00**
p2	f	3	85.75	66.61	**100.00**
p3	m	2	79.49	49.91	**100.00**
p4	f	6	85.92	**83.40**	88.82
p5	f	3	84.20	62.08	**100.00**
p6	f	4	80.36	69.19	**100.00**
p7	f	5	76.95	79.99	72.39
p8	m	3	**86.23**	66.68	**100.00**
p9	m	2	81.77	49.97	**100.00**
p10	f	2	73.24	49.80	**100.00**
		Average	**81.68**	**64.66**	**96.12**

Figure 3 compares the RCV performance metrics between the two data sets. The results indicate that the model achieved similar performance in both data sets, with a noticeable difference of almost 18% only for the SPE metric.
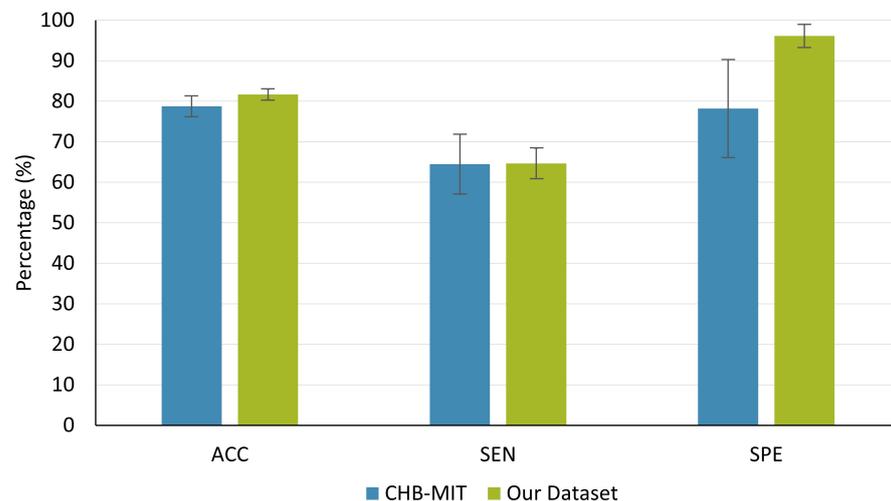
Figure 3. Average performance computed using the RCV validation method on the two data sets. The bar chart displays the average performance metrics and the standard error of the mean.

3.2. LOO

Table 3 reports the performance measured using the LOO validation method on the CHB-MIT data set. The average values of ACC, SEN, and SPE in all patients were 49.55%, 55.56%, and 43.54%, respectively, representing a dramatic decrease compared with the RCV validation method. The results indicate that chb07 achieved the highest ACC (57.64%) and SEN (79.31%) among all patients, while chb12 had the highest SPE (51.53%).

The performance measured using LOO validation in our data set is shown in Table 4. In addition, in this case, we observed a significant drop in performance, with average metric values of 50.93%, 48.62%, and 54.58% for ACC, SEN, and SPE, respectively. The highest values were obtained in the same patient p5, with 79.20% for ACC, 77.79% for SEN, and 80.09% for SPE.

Table 3. LOO validation method's results on the CHB-MIT data set. Bold format highlights maximum values.

Patient Number	Gender	Total Seizure	ACC (%)	SEN (%)	SPE (%)
chb01	f	3	43.23	39.88	46.58
chb04	m	3	48.19	49.54	46.85
chb05	f	3	55.28	62.31	48.24
chb06	f	5	46.86	57.11	36.61
chb07	f	2	**57.64**	**79.31**	35.97
chb12	f	4	49.38	47.22	**51.53**
chb15	m	6	49.15	51.34	46.96
chb24	-	2	46.67	57.78	35.56
		Average	**49.55**	**55.56**	**43.54**

Table 4. LOO validation method's results on our data set. Bold format highlights maximum values.

Patient Number	Gender	Total Seizure	ACC (%)	SEN (%)	SPE (%)
p1	m	10	44.93	52.80	36.26
p2	f	3	43.09	59.81	32.00
p3	m	2	56.02	52.78	58.04
p4	f	6	37.64	51.30	24.08
p5	f	3	**79.20**	**77.79**	**80.09**
p6	f	4	59.57	48.21	77.61
p7	f	5	54.27	44.22	67.34
p8	m	3	50.41	20.56	68.93
p9	m	2	35.48	59.31	23.32
p10	f	2	48.68	19.44	78.07
		Average	**50.93**	**48.62**	**54.58**

The comparison of performance metrics between the two data sets using the LOO validation method is reported in Figure 4. Similar to the RCV set-up, in this case, we observed only minor differences between the data sets, with slight variations mainly in SEN (6.94%) and SPE (11.04%).
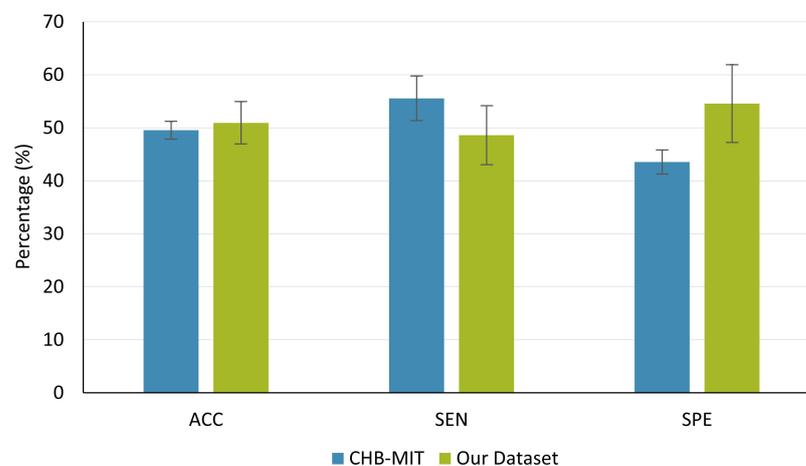

Figure 4. The results of the LOO validation method on the CHB-MIT data set and our data set to classify interictal and preictal phases are indicated. The bar chart displays the average performance metrics and the standard error of the mean.

3.3. Comparison between RCV and LOO

Table 5 and Figure 5 illustrate the difference in performance metrics resulting from the two validation methodologies. Overall, we can observe a striking drop in performance according to all metrics when adopting the more stringent LOO validation procedure. In particular, the accuracy decreased by almost 30% for both data sets, and the sensitivity decreased by 8.92% for CHB-MIT and 16.04% for our data set. The specificity decreased by 35% and 42% for CHB-MIT and our data set, respectively.

The performance metrics' averages, standard deviations, and results of statistical significance testing (two-tailed student's t-test) are reported in Table 5. The estimated p-values indicate that the differences were always statistically significant; only the sensitivity metric measured on the CHB-MIT data set did not seem to become significantly worse when the LOO validation procedure was adopted.

Table 5. Results of the t-test comparing RCV and LOO validation methods on the CHB-MIT data set and our data set. All performance metrics are illustrated by mean (%) $\pm$ standard deviation.

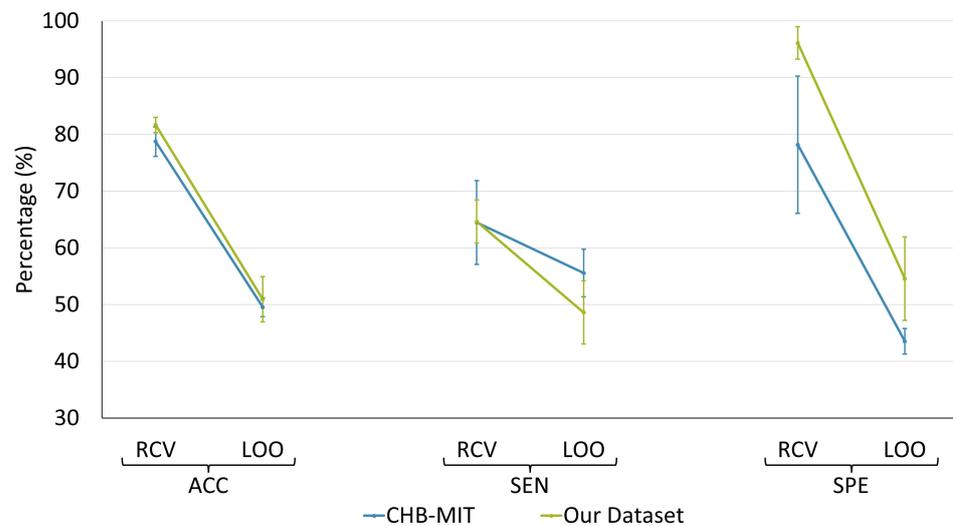| Data Sets	Metrics	RCV	LOO	*p*-Value
CHB-MIT	ACC	78.75 ± 7.34	49.55 ± 4.72	<0.001
	SEN	64.48 ± 20.88	55.56 ± 11.87	0.31
	SPE	78.2 ± 34.21	43.54 ± 6.40	<0.05
Our data set	ACC	81.68 ± 3.45	50.93 ± 12.80	<0.001
	SEN	64.66 ± 10.34	48.62 ± 15.96	<0.05
	SPE	96.12 ± 9.99	54.58 ± 21.81	<0.001

Figure 5. Comparison of the average performance metrics for classifying interictal and preictal phases on the CHB-MIT data set and our data set using RCV or LOO validation methods. Error bars represent the standard error of the mean.

4. Discussion

In this study, we used two data sets containing multichannel EEG recordings to evaluate the adequacy of randomized (RCV) and leave-one-patient-out (LOO) cross-validation strategies to measure machine learning algorithms' performance in a seizure prediction task. To this end, 53 features were first extracted from preprocessed EEG data, and various standard machine learning classifiers were trained to predict whether a signal window belonged to a preictal *vs.* interictal state. The best-performing model (XGBoost) was optimized using a Bayesian hyperparameter tuning procedure based on the Optuna framework and was finally deployed to compare two cross-validation schemes using standard performance metrics: accuracy, specificity, and sensitivity.

The results obtained using the RCV validation scheme suggest that machine learning algorithms can achieve remarkable performance in seizure prediction, with an average accuracy of 79% for the CHB-MIT data set and 82% for our data set. The accuracy was even more impressive for individual patients, reaching 91% in CHB-MIT and 86% in our data set. These findings align with previous results reported in the literature, which used similar machine learning models for seizure prediction. For example, a study reported an average sensitivity of 75.8% in discriminating interictal and preictal states using SVMs [19]. Another reported an average sensitivity of 80% [37], and yet another study obtained an average accuracy of 81.17% using time-frequency feature extraction combined with classification techniques [24].

However, our results clearly highlight that the RCV validation method could lead to overly optimistic conclusions. Indeed, when using the more robust LOO validation procedure, all performance metrics dramatically dropped, often by more than 20%. This suggests that a random splitting of EEG windowed signals might consistently increase the risk of overfitting the training data, making it easier for the model to learn spurious statistical features that are not representative of the clinical condition. For example, the classifier could learn to predict an upcoming seizure based on a systematic but uninformative alteration of the EEG recording, such as a temporary increase in skin conductivity caused by a patient's sweating. When tested on a completely different patient, such a classifier would misinterpret sweating as an alerting signal.

5. Conclusions

The main objective of the present study was to establish a solid validation methodology, which could be used in future studies to more robustly assess the performance of machine learning models in epilepsy research. Building a system that could work accurately "out of the box" with new patients is one of the greatest challenges in seizure prediction, and we argue that the leave-one-patient-out validation strategy explored in our study is closer to real-life operating scenarios compared with randomized cross-validation procedures. Such a conclusion is aligned well with recent proposals that call for the adoption of more stringent evaluation criteria in seizure prediction [38], as well as with more general guidelines for the application of artificial intelligence tools in medicine [39]. Indeed, measuring the performance of a detection model on a completely left-out set of patients, rather than on a randomly selected split of the data, would allow avoiding introducing spurious correlations in the training/testing split and thus better assess the robustness and generalization across clinical settings and patient populations.

Once a proper evaluation methodology has been established, we aim to explore more advanced machine learning techniques, such as deep neural networks [40]. This would require collecting and annotating a much higher volume of EEG recordings, but at the same time, it could significantly improve the prediction accuracy, even in the leave-one-patient-out setting. We believe that success in this challenging task would finally pave the way for the clinical testing of supporting technologies based on machine learning, which holds great potential to improve the lives of epileptic patients.

Institutional Review Board Statement: This study was conducted in accordance with the Declaration of Helsinki and approved by the local ethics committee (n.350/CE-Medea).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data used in the present study are not publicly available due to privacy issues related to the involvement of clinical populations.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EEG	Electroencephalography
XGBoost	eXtreme Gradient Boosting
SVM	Support vector machine
RCV	Randomized cross-validation
LOO	Leave-one-patient-out
ACC	Accuracy
SEN	Sensitivity
SPE	Specificity

Appendix A

Table A1. List of 22 channels used in the CHB-MIT data set.

Data Set	Channel Names
CHB-MIT	FP1-F7, F7-T7, T7-P7, P7-O1, FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2, FP2-F8, F8-T8, T8-P8, P8-O2, FZ-CZ, CZ-PZ, P7-T7, T7-FT9, FT9-FT10, FT10-T8

Table A2. List of 27 APIs used to extract the signal features.

Feature Type	API Names
Univariate features	compute_mean, compute_variance, compute_std, compute_ptp_amp, compute_skewness, compute_kurtosis, compute_rms, compute_quantile, compute_decorr_time, compute_pow_freq_bands, compute_hjorth_mobility_spect, compute_hjorth_complexity_spect, compute_hjorth_mobility, compute_hjorth_complexity, compute_higuchi_fd, compute_katz_fd, compute_zero_crossings, compute_line_length, compute_spect_slope, compute_spect_entropy, compute_energy_freq_bands, compute_spect_edge_freq, compute_wavelet_coef_energy, compute_teager_kaiser_energy
Bivariate features	compute_max_cross_corr, compute_phase_lock_val, compute_nonlin_interdep

References

1. Beghi, E. The epidemiology of epilepsy. *Neuroepidemiology* **2020**, *54*, 185–191. [CrossRef] [PubMed]
2. Fisher, R.S.; Boas, W.V.E.; Blume, W.; Elger, C.; Genton, P.; Lee, P.; Engel, J., Jr. Epileptic seizures and epilepsy: Definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia* **2005**, *46*, 470–472. [CrossRef] [PubMed]
3. Kwan, P.; Schachter, S.C.; Brodie, M.J. Drug-resistant epilepsy. *N. Engl. J. Med.* **2011**, *365*, 919–926. [CrossRef] [PubMed]

4. Fisher, R.S.; Vickrey, B.G.; Gibson, P.; Hermann, B.; Penovich, P.; Scherer, A.; Walker, S. The impact of epilepsy from the patient's perspective I. Descriptions and subjective perceptions. *Epilepsy Res.* **2000**, *41*, 39–51. [CrossRef]

5. Andrzejak, R.G.; Lehnertz, K.; Mormann, F.; Rieke, C.; David, P.; Elger, C.E. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E* **2001**, *64*, 061907. [CrossRef]

6. Lehnertz, K.; Mormann, F.; Kreuz, T.; Andrzejak, R.G.; Rieke, C.; David, P.; Elger, C.E. Seizure prediction by nonlinear EEG analysis. *IEEE Eng. Med. Biol. Mag.* **2003**, *22*, 57–63. [CrossRef]

7. Iasemidis, L.D. Epileptic seizure prediction and control. *IEEE Trans. Biomed. Eng.* **2003**, *50*, 549–558. [CrossRef]

8. Cook, M.J.; O'Brien, T.J.; Berkovic, S.F.; Murphy, M.; Morokoff, A.; Fabinyi, G.; D'Souza, W.; Yerra, R.; Archer, J.; Litewka, L.; 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. [CrossRef]

9. Van Mierlo, P.; Papadopoulou, M.; Carrette, E.; Boon, P.; Vandenberghe, S.; Vonck, K.; Marinazzo, D. Functional brain connectivity from EEG in epilepsy: Seizure prediction and epileptogenic focus localization. *Prog. Neurobiol.* **2014**, *121*, 19–35. [CrossRef]

10. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [CrossRef]

11. Calesella, F.; Testolin, A.; De Filippo De Grazia, M.; Zorzi, M. A comparison of feature extraction methods for prediction of neuropsychological scores from functional connectivity data of stroke patients. *Brain Inform.* **2021**, *8*, 1–13. [CrossRef]

12. Abbasi, B.; Goldenholz, D.M. Machine learning applications in epilepsy. *Epilepsia* **2019**, *60*, 2037–2047. [CrossRef] [PubMed]

13. Assi, E.B.; Nguyen, D.K.; Rihana, S.; Sawan, M. Towards accurate prediction of epileptic seizures: A review. *Biomed. Signal Process. Control* **2017**, *34*, 144–157. [CrossRef]

14. Gadhoumi, K.; Lina, J.M.; Mormann, F.; Gotman, J. Seizure prediction for therapeutic devices: A review. *J. Neurosci. Methods* **2016**, *260*, 270–282. [CrossRef] [PubMed]

15. Greene, B.R.; Faul, S.; Marnane, W.; Lightbody, G.; Korotchikova, I.; Boylan, G.B. A comparison of quantitative EEG features for neonatal seizure detection. *Clin. Neurophysiol.* **2008**, *119*, 1248–1261. [CrossRef] [PubMed]

16. Temko, A.; Thomas, E.; Marnane, W.; Lightbody, G.; Boylan, G. EEG-based neonatal seizure detection with support vector machines. *Clin. Neurophysiol.* **2011**, *122*, 464–473. [CrossRef]

17. Rasekhi, J.; Mollaei, M.R.K.; Bandarabadi, M.; Teixeira, C.A.; Dourado, A. Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods. *J. Neurosci. Methods* **2013**, *217*, 9–16. [CrossRef]

18. Teixeira, C.A.; Direito, B.; Bandarabadi, M.; Le Van Quyen, M.; Valderrama, M.; Schelter, B.; Schulze-Bonhage, A.; Navarro, V.; Sales, F.; Dourado, A. Epileptic seizure predictors based on computational intelligence techniques: A comparative study with 278 patients. *Comput. Methods Programs Biomed.* **2014**, *114*, 324–336. [CrossRef]

19. Bandarabadi, M.; Teixeira, C.A.; Rasekhi, J.; Dourado, A. Epileptic seizure prediction using relative spectral power features. *Clin. Neurophysiol.* **2015**, *126*, 237–248. [CrossRef]

20. Zandi, A.S.; Tafreshi, R.; Javidan, M.; Dumont, G.A. Predicting epileptic seizures in scalp EEG based on a variational Bayesian Gaussian mixture model of zero-crossing intervals. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 1401–1413. [CrossRef]

21. Shazadi, K.; Petrovski, S.; Roten, A.; Miller, H.; Huggins, R.M.; Brodie, M.J.; Pirmohamed, M.; Johnson, M.R.; Marson, A.G.; O'Brien, T.J.; et al. Validation of a multigenic model to predict seizure control in newly treated epilepsy. *Epilepsy Res.* **2014**, *108*, 1797–1805. [CrossRef] [PubMed]

22. Dietterich, T. Overfitting and undercomputing in machine learning. *ACM Comput. Surv. (CSUR)* **1995**, *27*, 326–327. [CrossRef]

23. Mutasa, S.; Sun, S.; Ha, R. Understanding artificial intelligence based radiology studies: What is overfitting? *Clin. Imaging* **2020**, *65*, 96–99. [CrossRef] [PubMed]

24. Tamanna, T.; Rahman, M.A.; Sultana, S.; Haque, M.H.; Parvez, M.Z. Predicting seizure onset based on time-frequency analysis of EEG signals. *Chaos Solitons Fractals* **2021**, *145*, 110796. [CrossRef]

25. Kitano, L.A.S.; Sousa, M.A.A.; Santos, S.D.; Pires, R.; Thome-Souza, S.; Campo, A.B. Epileptic seizure prediction from EEG signals using unsupervised learning and a polling-based decision process. In Proceedings of the Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; Proceedings, Part II 27; Springer: Berlin/Heidelberg, Germany, 2018; pp. 117–126.

26. Abdelhameed, A.M.; Bayoumi, M. An Efficient Deep Learning System for Epileptic Seizure Prediction. In Proceedings of the 2021 IEEE International Symposium on Circuits and Systems (ISCAS), Daegu, Republic of Korea, 22–28 May 2021; pp. 1–5.

27. Shoeb, A.H. Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.

28. Selim, S.; Elhinamy, E.; Othman, H.; Abouelsaadat, W.; Salem, M.A.M. A review of machine learning approaches for epileptic seizure prediction. In Proceedings of the 2019 14th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 17–18 December 2019; pp. 239–244.

29. Usman, S.M.; Khalid, S.; Bashir, Z. Epileptic seizure prediction using scalp electroencephalogram signals. *Biocybern. Biomed. Eng.* **2021**, *41*, 211–220. [CrossRef]

30. Patel, V.; Buch, S.; Ganatra, A. A review on EEG based epileptic seizure prediction using machine learning techniques. In *Proceedings of the International Conference on Intelligent Computing, Information and Control Systems*; Springer: Cham, Switzerland, 2019; pp. 384–391.

31. Niknazar, H.; Maghooli, K.; Nasrabadi, A.M. Epileptic seizure prediction using statistical behavior of local extrema and fuzzy logic system. *Int. J. Comput. Appl.* **2015**, *113*. [CrossRef]

32. Thangavel, P.; Thomas, J.; Peh, W.Y.; Jing, J.; Yuvaraj, R.; Cash, S.S.; Chaudhari, R.; Karia, S.; Rathakrishnan, R.; Saini, V.; et al. Time–frequency decomposition of scalp electroencephalograms improves deep learning-based epilepsy diagnosis. *Int. J. Neural Syst.* **2021**, *31*, 2150032. [CrossRef]

33. Allen, P.; Fish, D.; Smith, S. Very high-frequency rhythmic activity during SEEG suppression in frontal lobe epilepsy. *Electroencephalogr. Clin. Neurophysiol.* **1992**, *82*, 155–159. [CrossRef]

34. Arroyo, S.; Uematsu, S. High-frequency EEG activity at the start of seizures. *J. Clin. Neurophysiol.* **1992**, *9*, 441–448. [CrossRef]

35. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.

36. Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; Koyama, M. Optuna: A Next-generation Hyperparameter Optimization Framework. In Proceedings of the Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, 4–8 August 2019.

37. Gadhoumi, K.; Lina, J.M.; Gotman, J. Discriminating preictal and interictal states in patients with temporal lobe epilepsy using wavelet analysis of intracerebral EEG. *Clin. Neurophysiol.* **2012**, *123*, 1906–1916. [CrossRef]

38. Peng, P.; Song, Y.; Yang, L.; Wei, H. Seizure prediction in EEG signals using STFT and domain adaptation. *Front. Neurosci.* **2022**, *15*, 1880. [CrossRef] [PubMed]

39. Rajpurkar, P.; Chen, E.; Banerjee, O.; Topol, E.J. AI in health and medicine. *Nat. Med.* **2022**, *28*, 31–38. [CrossRef] [PubMed]

40. Shoeibi, A.; Khodatars, M.; Ghassemi, N.; Jafari, M.; Moridian, P.; Alizadehsani, R.; Panahiazar, M.; Khozeimeh, F.; Zare, A.; Hosseini-Nejad, H.; et al. Epileptic seizures detection using deep learning techniques: A review. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5780. [CrossRef] [PubMed]