2024 20th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT) | 979-8-3503-6944-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/DCOSS-IoT61029.2024.00097

Robustness of ML-Based Seizure Prediction Using Noisy EEG Data From Limited Channels

Umair Mohammad and Fahad Saeed
Knight Foundation School of Computing and Information Sciences
Florida International University, Miami, FL, USA
{umohamma, fsaeed}@fiu.edu

***Abstract*—Seizures pose a significant health hazard for over 50 million individuals with epilepsy worldwide, with approximately 56% experiencing uncontrollable seizures according to the CDC. Predicting seizures is challenging even with the availability of various sensors (gyroscopes, pulse rate sensors, heart rate monitors, etc). Electroencephalography (EEG) data can directly measure the activity of the brain and has been the choice of leveraging deep learning (DL) models for seizure prediction. Despite DL models achieving over 95% accuracy on retroactive clinical-grade EEG data, this performance fails to translate in real-world settings where the accuracy goes down to 66% - which warrants further investigation. Moreover, consumer-grade wearable EEG headsets, characterized by lower data quality and a varying number of channels across brands, present additional challenges. In this paper, we estimate the robustness of DL models which are trained on clinical-grade EEG data but tested on the type of data expected from consumer-grade wearable EEG headsets. We select the previously published model SPERTL to estimate its robustness when: (1) predicting with data from less leads/channels, (2) predicting when faced with streaming data, (3) evaluating performance on imbalanced data with more interictal segments. Our results are compared against baseline results from the SPERTL model which we have re-configured to operate independently of the number of channels with an average baseline area under the curve (AUC) score of 98.56%. Our results demonstrate that though the model is surprisingly resilient to streaming and noisy data, reducing the number of channels and a higher class imbalance have a more severe degradation. The AUC across all cross-validation sets degrades only by 2% and 3% on average for noisy and streaming data, respectively. However, a performance reduction, on average, is observed by 32% when imbalance is increased with higher percentage of interictal samples, and up to 16% when using lower number of channels.**

***Index Terms*—Electroencephalography (EEG), Wearable, Seizure Prediction, Robustness, Lower Quality Data**

I. Introduction

Epilepsy is a neurological disorder characterized by repeat unpredictable seizures which affect more than 3 million people in the USA and up-to 50 million people worldwide. Though treatable with medications, the centers for disease (CDC) estimates that 56% of the epileptic population do not achieve seizure freedom. These uncontrolled seizures pose significant health, and quality of life challenges. Electroencephalography (EEG), a technique to measure the brain's electrical activity, is used clinically by neurologists and epileptologists (neurologists specializing in epilepsy) for tracking, diagnosing and managing seizure activity.

Increasingly, machine learning (ML)-based approaches are playing a key role in seizure detection/prediction, spike detection and EEG transcription due to the availability of long-term EEG datasets annotated with different events including seizures. One such dataset is the Children's Hospital Boston - Massachusetts Institute of Technology (CHB-MIT) dataset, which comprises long-term EEG recordings annotated with the start and end times of seizures from 24 subjects. CHB-MIT has become the standard dataset for ML and advanced deep learning (DL)-based seizure prediction. Several models in literature have shown success in predicting preictal segments (indicating an imminent seizure) against interictal periods.

While developing DL models is an active area of research, several challenges remain to make these models functional on real-world wearable EEG headsets. Firstly, datasets such as the CHB-MIT are obtained in controlled-settings with clinical-grade equipment - which may not be possible in a real-world scenario. Further, most models train using balanced preictal and interictal classes which again does not represent real-world conditions; since the number of seizure events are much smaller than "normal" periods. In contrast, wearable EEG-based seizure prediction for ambulatory or home-care use is likely to be implemented on EEG headsets that have a good form factor, fixed-position dry electrodes and a capability for wireless communication. Hence, such EEG systems are expected to have fewer than the 23 channels common to datasets such as CHB-MIT. Data collected in real-time will be a stream of mostly interictal samples followed by successive preictal segments on rare occasions whenever a seizure occurs. Lastly, wearable wireless EEG has a lower signal-to-noise ratio (SNR) compared to clinical leading to reduced data quality.

The goal of this work is to investigate the robustness of a ML model (SPERTL [1] for this study) trained on the CHB-MIT dataset, and its ability to translate to a wearable EEG with less and lower quality data. This critical objective will bridge the gap between models that work well in literature but their robustness to consumer-grade wearable EEG for long-term seizure prediction usage is not well characterized. This will enable better understanding of the performance gaps for a given model, and what steps the ML practitioners can take to ensure model robustness for real-world settings.

This work was supported by the National Science Foundation (NSF) grants NSF TI-2213951 and OAC-2312599. The content is solely the responsibility of the authors and does not represent the official views of the NSF.

2325-2944/24/$31.00 ©2024 IEEE
DOI 10.1109/DCOSS-IoT61029.2024.00097

II. Related Work

The trend in healthcare is to move towards continuous monitoring or real-time detection systems especially for seizure care in ambulatory settings or even at home. This is aided by the development of high accuracy ML prediction and diagnosis models along with the explosion in the internet of medical things (IoMT) or health IoT (H-IoT) and wearable technology. Specifically, attention has been given to designing ML models that are implementable on system-on-chip (SoC) for use on wearables in an energy-efficient manner.

For example, the works of [2]–[4] use a combination of feature extraction and lightweight model design for lower complexity energy-efficient execution on SoC hardware. Further, the works of [5] improve the AUC/ROC of the seizure prediction model in continuous EEG (cEEG) whereas [6] proposes a dynamic architecture based on data collection from various sensors (no EEG) with preprocessing on an SOC or on a smartphone followed by classification using GAs and k-NNs. A reasonable accuracy is achieved with a latency of 0.5s and 2.25 days hours battery life. In contrast, the work of [7] makes ES detection patient-specific by offering personalized sensor selection and positioning on body. The works of [8]–[10] explore different IoMT based continuous ambulatory monitoring systems including a micro controller - TCP/IP network - user application based design [8], a RaspBerry Pi and an Android app-based solution with cloud-based ML inference [9], and distributing the ML model among the edge, fog nodes, and the cloud which results in significant energy savings when tested on the EPILEPSIAE dataset [10].

It can be observed that most of the works discussed focus on the architecture of the IoT framework rather than the performance of the ML models. The paradigm of edge computing with associated end IoT devices further enables the implementation of SoC systems mounted on wearable devices. The models can either be executed directly on chip and an alarm can be sent to the caregivers/physicians via the edge/cloud or the execution can be done at the edge or cloud after data collection from the wearable. Some of the relevant works in literature have attempted to quantify the ML model performance but mostly for detection, not prediction. For example, the work of [11] develops and tests lightweight models using feature dimensionality reduction by correlations that give an accuracy of 97.5% and lightweight detection with Kriging methods that result in a 100% accuracy and a 0.85s delay when tested on a RaspBerry Pi [12]. EZCap [13] is another wearable solution based on Kringing methods that is designed and tested on the Bonn dataset where results indicate an accuracy of 97.5% with a delay of 0.81s.

Because complete caps with scalp EEG may result in some social stigma, wearables with lower and more conspicuous placement of electrodes have been designed such as the 4-electrode systems mounted behind-the-ear [14] that gives a 94.5% accuracy for an FAR of 0.52 and on a smart glass [15] that can give a sensitivity of 93.8% on a 12 patient dataset with an average 2.71 day battery life. A simple real-time ES detection system is designed in [16] based on DWT based feature extraction and DNN which can provide an accuracy of 98.6% on the Bonn dataset.

Although some of these models provide a sensitivity of greater than 95%, either these models are trained on limited data collected from wearables (e.g. [11]) or tested on the Bonn dataset for seizure detection, *not prediction* on a dataset with long-term recordings. Overall, though making models energy-efficient and capable of IoMT deployment for real-time ES detection either on SoC or via the edge/cloud has been studied, few efforts have been expanded on quantifying the robustness of ML models trained for *prediction* and their ability to translate to IoT systems.

III. Problem Description

A major obstacle in developing end-to-end early seizure prediction solutions based on wearable EEG is the *inability* of ML-based models to translate to hardware. As described prior, most predictive models are trained and tested on the CHB-MIT where the reported sensitivities are in excess of 98% with a nearly zero false positive rate (FPR). However, as reported by [17], some of these models fail when tested with the leave-one-seizure-out (LOSO) validation even with a balanced class approach providing sensitivities as low as 67%. It is expected that when we consider less number of leads, streaming and imbalanced data in the real-world and lower EEG data quality, this sensitivity will be reduced even further with a higher FPR. The implications are that the model's ability to predict seizures will be severely reduced while the false alarm rate (FAR) will be higher due to a high FPR. To frame this problem scientifically, it is crucial to consider the inherent differences between the CHB-MIT dataset and wearable EEG data, as well as the specific challenges that need to be addressed.

A. Varying numbers and locations of EEG Channels:

The CHB-MIT dataset contains long-term recordings collected at a sampling rate of 256 Hz from a scalp EEG of up-to 22 channels with the electrodes placed in the standard 10-20 configuration. This dataset has become the benchmark for testing ML/AI models for seizure detection and prediction. On the contrary, consumer-grade wearable EEG's that have proliferated the market for applications including mental wellness and well-being, sleep monitoring and sports-related performance tracking, may have as few as four EEG channels. Others may have up-to 16 channels (and though there a few research-grade EEG headsets with up-to 32 and 64 channels, they are rarely used for commercial applications). Further, the location of electrodes on each of these brands will be different. Consequently, deploying pre-trained ML models on CHB-MIT dataset must demonstrate a degree of channel independence to accommodate the variable number of EEG channels available in commonly accessible wearable headsets. Such models would allow us to quantify the performance of the seizure prediction system across multiple types of wearable EEG with different form factors and numbers/locations of the lead.

B. Streaming/Imbalanced Validation

Traditionally, for seizure prediction tasks, the long-term EEG recording is divided into smaller segments. The decision-making is done at the level of each segment which can be used to make a universal decision at the level of the seizure. Further, though the start and end of seizures are known and annotated/verified based on clinical observations, the start and end of preictal duration is not know or clinically well-defined. That means researchers usually assume a set preictal period duration and define the seizure prediction horizon (SPH) and seizure occurrence period (SOP) based on this assumption. The preictal samples are usually extracted from the SPH and there is a gap equivalent to the SOP between the end of the preictal duration and start of the ictal onset. To build up the dataset, multiple segments are extracted from this assumed preictal duration.

All other labels that do not fall into the category of preictal or interictal (with the exception of additional gap times in some works [18]) are assumed to be interictal. Because of a significantly higher proportion of interictal samples, most works simply sample an equal amount of interictal and preictal samples in a balanced way. Further, training algorithms typically employ stochastic gradient descent where the preictal and interictal labeled samples are randomly shuffled before the cross-validation results are presented. However, this controlled division is not similar to real-world streaming EEG data.

In practice, validation should be performed with datasets that emulate the true temporal dynamics of epilepsy, where a series of interictal (non-seizure) samples are followed by a series of preictal (seizure-indicating) samples. Further, an unequal distribution of preictal and interictal samples should be tested with increasing ratios of interictal to preictal segments all the way up-to the maximum possible interictal segments.

C. Signal-to-Noise Ratio (SNR)

Wearable EEG headsets are usually designed for better form factor and ease-of-use which leads to a lower data quality compared to scalp EEG. This quality is further degraded in wireless headsets due to physical affects such as wireless fidelity and packet loss. Overall, datasets such as the CHB-MIT which collect EEG recordings from clinical-grade headsets can expect to have data characterized by a higher signal-to-noise ratio (SNR). Although ML models especially those based on DL have a degree of robustness to noise, the lower SNR observed in data from consumer-grade EEG can still negatively impact the ML model's predictive performance. Despite other factors such as packet loss from wireless communication, sampling rates and bit resolution, we focus on the data quality itself and characterize it with SNR as is common among clinical practices and EEG manufacturers.

The overall goal is to test the high performance of predictive models in controlled CHB-MIT dataset type settings in constrasting real-world situations characterized by variable numbers/locations EEG electrodes, streaming and imbalanced data, and SNR disparities. Our work is a progress towards solving this vital problem for the successful deployment of real-time seizure prediction models on wearables, ultimately improving the lives of individuals living with epilepsy and enhancing their safety and well-being.

TABLE I
CHANNELS USED FROM THE CHB-MIT FOR VALIDATION COMPARED TO AVAILABLE BRANDS.

CHB-MIT	Brand 1	Brand 2	Brand 3a	Brand 3b
Fp1	Fp1	-	-	AF3
FP2	FP2	-	-	AF4
F7	F7	-	-	F7
F3	F3	-	F3	F3
F4	F4	-	F4	F4
F8	F8		-	F8
T7	T3	T3	T7	T7
C3	C3	-	-	C5
C4	C4	-	-	C6
T8	T4	T4	T8	-
P7	T5	-	-	P7
P3	P3	-	-	-
P4	P4	-	-	T8
P8	T6	-	-	P8
O1	O1	O1	-	O1
O2	O2	O2	-	O2
FZ	FZ	-	-	-
CZ	CZ	-	-	-
PZ	PZ	-	Pz	-

IV. METHODS

In this study, we propose an approach to assess and quantify the robustness of an ML model trained and validated on the CHB-MIT dataset. Firstly, we draw inspiration from Wang et al.'s work [18], where they employed LOSO cross-validation and only 7/24 patients that were able to provide at least three seizures fitting the selection criteria are chosen. For our work, we will focus on each of the three cross-validation (CV) folds of patient 1 to quantify the different metrics. The ML model under consideration is a modified version of the residual neural network (ResNet) that has been previously trained and tested on all 7 patients and provided up-to 88% sensitivity [1].

To make the model independent of the number of channels and enhance the model versatility, we create a 1D ResNet equivalent of SPERTL and train each channel of each segment individually. During validation, majority voting is employed to decide if each segment belongs to the preictal or interictal class. This approach allows us to validate the model's performance effectively, even when confronted with EEG data generated by wearable headsets that may feature a significantly reduced number of channels compared to the CHB-MIT dataset. Table I shows the equivalent electrode placement on commonly available consumer grade wearable EEG compared to CHB-MIT. Only channels using these electrodes will be used to validate the performance of each brand.

We introduce a two-stage process to address the temporal complexities of seizure prediction in the real world. In the first stage, we let the model observe the validation set as a continuous stream of EEG data rather than randomly dividing the preictal and interictal samples among the CV folds. This mirrors the actual progression of an epileptic seizure, where

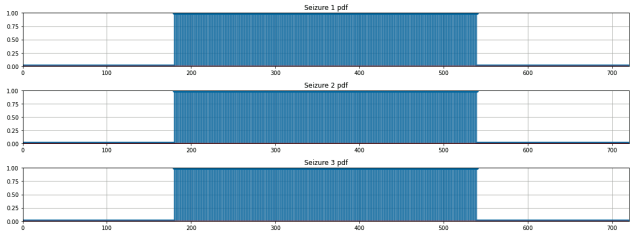


Fig. 1. Example of validation dataset left as a stream of interictal samples followed by preictal.

a sequence of interictal segments are followed by preictal segments. In the next stage, we test for the ability of the model to predict seizures with increasing interictal to preictal ratios. Fig. 1 represents the data labels collated in a streaming manner. Each row represents the preictal and interictal labels associated for a particular seizure. For example, the images shown are for seizures 1, 3 and 7 for patient CHB01 of the CHB-MIT dataset.

Consumer-grade wearable EEG inherently has a lower SNR and hence, we we validate the model's resilience by introducing white Gaussian noise to the EEG data, emulating scenarios where SNR disparities exist. In general, let us say the difference in SNR of a medical grade EEG is given by SNR_M and a wearable by SNR_W. Then, the difference in SNR (SNR_{diff}) can be given by:

$$SNR_{diff} = SNR_M - SNR_W \tag{1}$$

Let us say that an EEG segment of length L is represented by $\mathbf{E}_W$. Then, to test with CHB-MIT, we will bring down the quality by the addition of the following noise:

$$\mathbf{E}_W = \mathbf{E}_W - SNR_{diff} N_0 / L \tag{2}$$

where $N_0 \sim \mathcal{N}(0,\ 1)$; where N_0 is drawn from a zero-mean unit-variance Gaussian random variable.

V. Experimental Setup

We focus on EEG data from patient 01 of the CHB-MIT dataset with a total recording time of 40 hours with seven seizures annotated after verification by 2 independent neurologists. For our preictal/interictal sampling strategy, we adopted a SPH of 30 minutes, and a SOP of 5 minutes meaning a preictal duration of 30 minutes is assumed and that seizures are predicted at least 5 minutes ahead of time. These durations demonstrated to be sufficient to capture seizure activity from prior work in the literature [1], [18]. Further, a 2-hour gap is assumed between 2 consecutive seizures to clearly distinguish preictal and interictal periods and account for potential post-ictal effects. After applying this stringent inclusion-exclusion-inclusion criteria, we retained seizures 1, 3, and 7 and generate *three CV sets* based on the LOSO technique.

To ensure uniformity, all preictal and interictal data is segmented into non-overlapping 5-second frames, corresponding to a frame size of 1280 samples based on a sampling rate of 256 Hz. The major task in seizure prediction is solving the binary classification problem of discriminating between preictal and interictal segments. The data comes from 22 EEG channels, each segment contains 1280 samples and for every seizure, we have 360 preictal segments (and 360 interictal segments in the balanced case), leading to a total size of $720 \times 22 \times 1,280$. To test for channel-independence, each channel was treated as an independent subject of interest which equates to 7,920 positive labels in total per CV set. This meticulous approach to data selection and labeling laid the groundwork for our subsequent experiments in which we assessed the translatability of machine learning models to real-world wearable EEG data.

TABLE II
Comparison of the baseline ROC-AUC score against the results from brands 2, 3a and 3b.

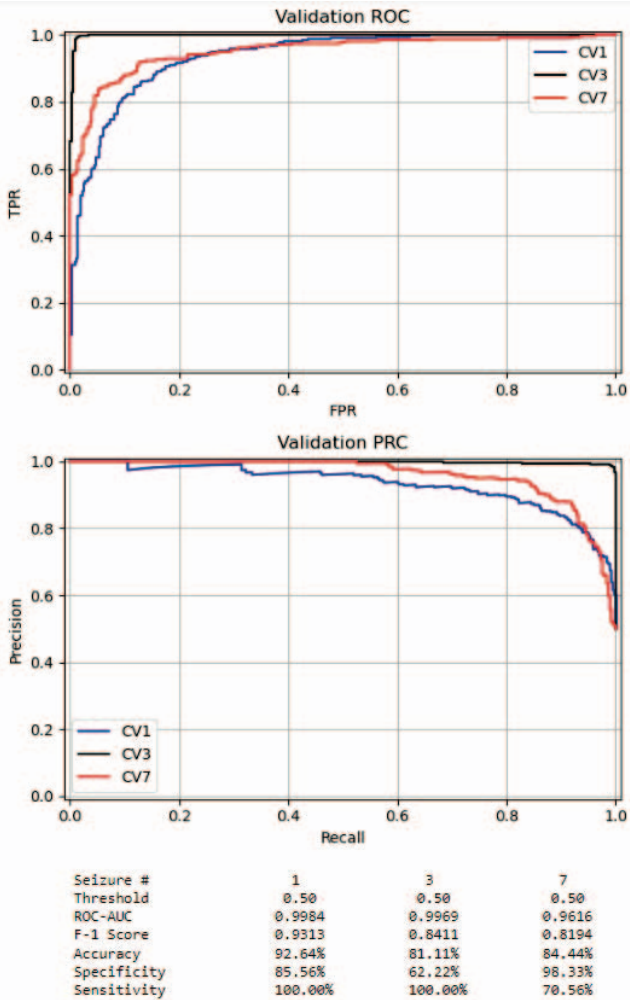
Seizure	Baseline	Brand 2	Brand 3a	Brand 3b
1	0.9984	0.6570	0.5781	0.9508
3	0.9969	0.8765	0.5208	0.9808
7	0.9616	0.9554	0.8922	0.9216
Average	0.9856	0.8296	0.6637	0.9511

A. Simulation Scenarios:

In our baseline training and validation process with the LOSO method, we designate two seizures for training purpose while reserving the third for validation. Notably, the training validation split is a very stringent 66.7%-33.3%. This ratio differs significantly from the more common 80-20 or 75-25 splits seen in traditional ML-based classification. To illustrate, for CV set 1, we use the preictal and interictal samples associated with seizures 3 and 7 for training, whereas we validate the performance using the preictal and interictal samples from seizure 1, and so forth for CV sets 3 and 7.

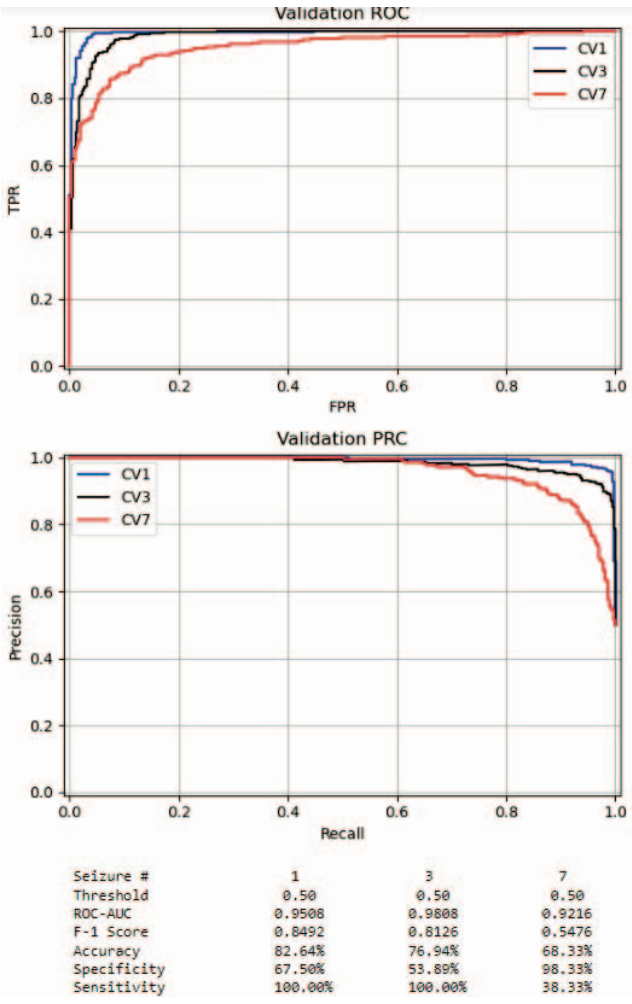
The baseline model is validated balanced classes, including 30 minutes of randomly sampled interictal segments out of the contiguous interictal segments collected from the long-term records based on the above defined exclusion criteria. Table I shows the receiver operating characteristic - area under the curve (ROC-AUC) scores for the baseline model. In our exploration of streaming data scenarios, we re-validated the model for each cross-validation fold by re-arranging the data so that a series of interictal labels are followed by preictal labels. The contiguous interictal samples that preceded preictal samples for each seizure, without random shuffling of the validation set. The streaming approach is illustrated in Figure 1, highlighting the importance of adapting models to streaming data dynamics.

To assess the impact of channel scarcity and class imbalance on model performance, we conducted a series of experiments. First, we validated the streaming data with extreme levels of imbalances by adding additional interictal samples to the validation set. This allowed us to quantify the effects of class imbalance on model performance systematically. Additionally, to simulate scenarios with fewer EEG leads, we simply excluded data from the leads described in Table 1, thereby mimicking the limitations of wearable EEG headsets. Our



Seizure #	1	3	7
Threshold	0.50	0.50	0.50
ROC-AUC	0.9984	0.9969	0.9616
F-1 Score	0.9313	0.8411	0.8194
Accuracy	92.64%	81.11%	84.44%
Specificity	85.56%	62.22%	98.33%
Sensitivity	100.00%	100.00%	70.56%

Fig. 2. Baseline results for SPERTL re-trained and cross-validated on patient 1 of the CHB-MIT dataset.



Seizure #	1	3	7
Threshold	0.50	0.50	0.50
ROC-AUC	0.9508	0.9808	0.9216
F-1 Score	0.8492	0.8126	0.5476
Accuracy	82.64%	76.94%	68.33%
Specificity	67.50%	53.89%	98.33%
Sensitivity	100.00%	100.00%	38.33%

Fig. 3. Validation on Streaming type data but with balanced preictal and interictal ratios.

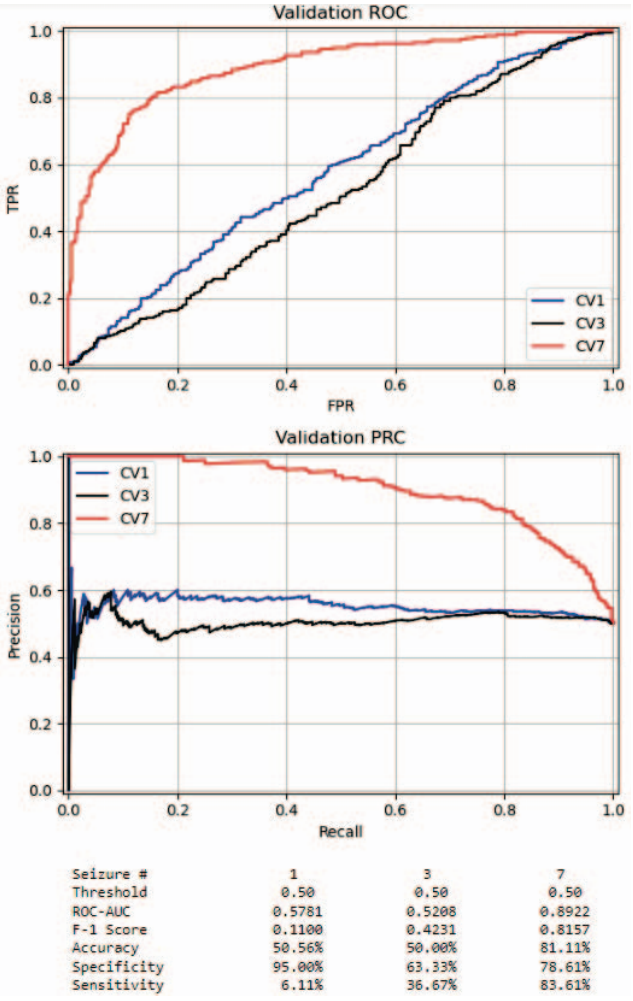
decision-making process remained consistent, classifying a segment as preictal if more than half of the leads predicted the positive class. To assess the influence of lower data quality characterized by noise, we intentionally introduced White Gaussian noise to the validation set, as detailed in (2). These experiments provided valuable insights into how the model adapted to noisy data and the implications for its predictive accuracy.

To demonstrate the results for all scenarios, we plot the receiver operating characteristic (ROC) and precision recall curves (PRC). We also compare the performance along several metrics including the ROC-AUC score, F-1 score, accuracy, sensitivity and specificity which are shared below each figure as a snapshot. On each plot, CV set 1 (CV1) is represented by the blue line, CV3 with the black line and CV7 with the red line.

VI. Results and Discussion

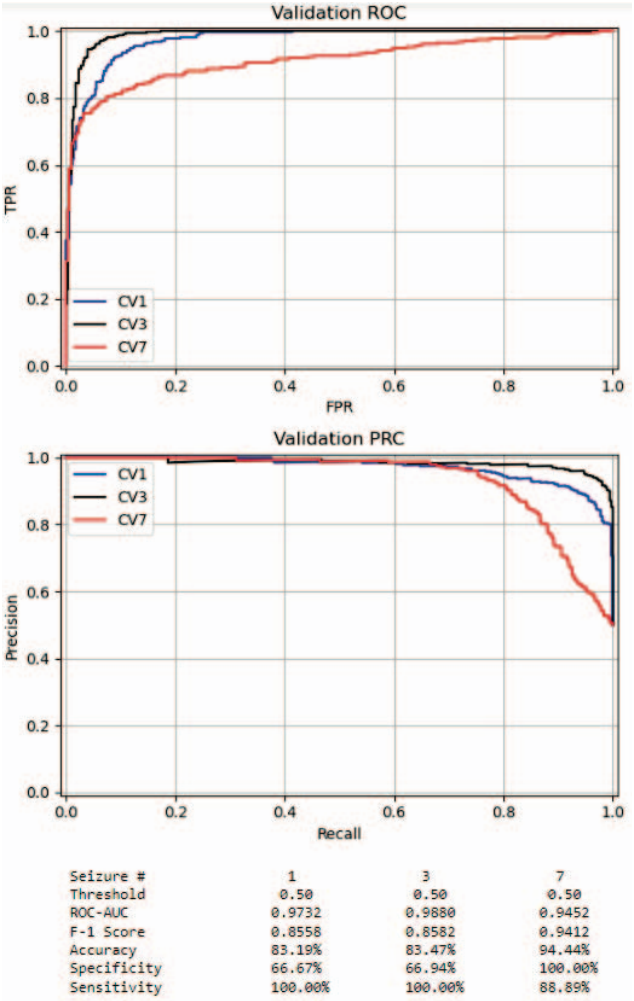
In this work, we begin by demonstrating the baseline results for each CV fold from the version of SPERTL re-configured for channel independent training and validation. Fig. 2 illustrates that the model provides excellent results for all CV sets with an ROC-AUC scores in excess of 0.99 for CV1 and CV3 and a perfect sensitivity whereas a ROC-AUC score of 0.9610 is achieved for CV7. The experiments are then extended to the case with less channels corresponding to the configurations in Table I. As observable, brands 2 and 3a which have the lowest number of channels (4 and 5, respectively), result in the lowest levels of performance. However, there is an interesting anomaly where the performance on CV7 does not degrade significantly with a reduction in channels. The reason for this appears to be that the included channels are sufficient to capture the preictal activity of the seizure type in CV7 whereas they are not adequate for CV1 and CV3.

Fig. 3 illustrates the model's performance when the valida-



Seizure #	1	3	7
Threshold	0.50	0.50	0.50
ROC-AUC	0.5781	0.5208	0.8922
F-1 Score	0.1100	0.4231	0.8157
Accuracy	50.56%	50.00%	81.11%
Specificity	95.00%	63.33%	78.61%
Sensitivity	6.11%	36.67%	83.61%

Fig. 4. Validation on Streaming type data but with severely imbalanced interictal to preictal ratios for each cross-validation fold.



Seizure #	1	3	7
Threshold	0.50	0.50	0.50
ROC-AUC	0.9732	0.9880	0.9452
F-1 Score	0.8558	0.8582	0.9412
Accuracy	83.19%	83.47%	94.44%
Specificity	66.67%	66.94%	100.00%
Sensitivity	100.00%	100.00%	88.89%

Fig. 5. Validation on from all channels corresponding to CHB-MIT but on noisy data with an average degradation of 10 dB.

tion set is observed as streaming data whereas Fig. 4 shows the results with an additional imbalance component added on by increasing the number of interictal samples. As observable, when we simply present the validation data in a streaming manner, the model is quite robust and there is only an average drop of 0.02 in the ROC-AUC score. However, once we overload the validation sets with interictal data (on average there is a 16:1 interictal to preictal ratio), the average ROC-AUC score across all sets falls by about 32%. The impact may be so severe because we directly test the imbalance on streaming data and instead of a gradual increase, opt for the more extreme 16:1 ratio.

Lastly, we examine the impact of lower quality data. Overall, the average AUC-ROC score is only reduced by about 0.02. However, some of the cross-validation folds appear to be more robust such as CV1 and CV3. In contrast, CV7 has a sharper fall in performance compared to the baseline. One reason for the overall smooth performance may be the role of normalization after the first input layer and before the convolutional layer in SPERTL. The comparison of all approaches including streaming data, imbalanced classes and lower quality data against the baseline in terms of the ROC-AUC are presented in Table VI. It is clearly observable that the performance drops in each case, but the most severe drop occurs in the case of imbalanced classes. It will be interesting to study different levels of class-imbalance to quantify the performance drop with increasing imbalance.

TABLE III
COMPARISON OF THE ROC-AUC SCORE FOR ALL CROSS-VALIDATION SETS UNDER DIFFERENT TEST CONDITIONS.

Seizure	Baseline	Streaming	Imbalanced	Noisy
1	0.9984	0.9508	0.5781	0.9732
3	0.9969	0.9808	0.5208	0.9880
7	0.9616	0.9216	0.8922	0.9452

VII. Conclusion

In this paper, we evaluated the robustness and ability of a previously developed DL-based seizure prediction model to translate to differences in quality of data from consumer-grade wearable EEG headsets. First, we set up a baseline and then quantified the performance variations by selectively ignoring channels, considering the distinctive characteristics of four common wearable EEG brands. Second, we validated the model robustness by introducing noise, reducing the signal-to-noise ratio in the validation data, and simulating lower-quality consumer-grade EEG recordings. Lastly, we explore the impact of various levels of class imbalance, reflecting the real-life rarity of seizure events and the prevalence of a higher proportion of interictal samples. In this last step, we also make additional performance characterizations based on the model's performance when presented with data in a streaming format, reflecting real-time collection of data from EEG sensors. Overall, we demonstrated that imbalance and a reduction in the number of channels has the most significant impact as compared to streaming or noisy/low-quality data. Further experiments are needed to validate these results by testing different levels of imbalance, and more comprehensive characterization across all of the patients of the CHB-MIT dataset.

References

[1] U. Mohammad and F. Saeed, "SPERTL: Epileptic Seizure Prediction using EEG with ResNets and Transfer Learning," in *BHI-BSN 2022 - IEEE-EMBS International Conference on Biomedical and Health Informatics and IEEE-EMBS International Conference on Wearable and Implantable Body Sensor Networks, Symposium Proceedings*, pp. 1–5, Institute of Electrical and Electronics Engineers Inc., 2022.

[2] J. Yoo, "On-chip epilepsy detection: Where machine learning meets patient-specific healthcare," in *Proceedings - International SoC Design Conference 2017, ISOCC 2017*, pp. 146–147, Institute of Electrical and Electronics Engineers Inc., may 2018.

[3] M. Shoaran, B. A. Haghi, M. Taghavi, M. Farivar, and A. Emami-Neyestanak, "Energy-efficient classification for resource-constrained biomedical applications," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 8, pp. 693–707, dec 2018.

[4] S. Iranmanesh, G. Raikos, S. A. Imtiaz, and E. Rodriguez-Villegas, "A Seizure-Based Power Reduction SoC for Wearable EEG in Epilepsy," *IEEE Access*, vol. 7, pp. 151682–151691, 2019.

[5] A. Hanin, S. Demeret, V. H. Nguyen-Michel, V. Lambrecq, and V. Navarro, "Continuous EEG monitoring in the follow-up of convulsive status epilepticus patients: A proposal and preliminary validation of an EEG-based seizure build-up score (EaSiBUSSEs)," *Neurophysiologie Clinique*, vol. 51, pp. 101–110, mar 2021.

[6] P. M. Vergara, E. De La Cal, J. R. Villar, V. M. González, and J. Sedano, "An IoT Platform for Epilepsy Monitoring and Supervising," *Journal of Sensors*, vol. 2017, pp. 1–18, 2017.

[7] S. A. McHale, E. Pereira, U. Weishmann, M. Hall, and H. Fang, "An IoT approach to personalised remote monitoring and management of epilepsy," in *Proceedings - 14th International Symposium on Pervasive Systems, Algorithms and Networks, I-SPAN 2017, 11th International Conference on Frontier of Computer Science and Technology, FCST 2017 and 3rd International Symposium of Creative Computing, ISCC 2017*, pp. 414–418, Institute of Electrical and Electronics Engineers Inc., nov 2017.

[8] F. Pinho, J. H. Correia, N. J. Sousat, J. J. Cerqueira, and N. S. Dias, "Wireless and wearable EEG acquisition platform for ambulatory monitoring," in *SeGAH 2014 - IEEE 3rd International Conference on Serious Games and Applications for Health, Books of Proceedings*, pp. 1–7, Institute of Electrical and Electronics Engineers Inc., 2014.

[9] M. Manimegalai, P. Nirmala Devi, V. Mekala, S. Haripriya, B. Jawahar, and M. Mageshkumar, "A Machine learning approach to monitor Brain Health and epilepsy detection," in *IOP Conference Series: Materials Science and Engineering*, vol. 1055, pp. 1–14, IOP Publishing, feb 2021.

[10] F. Forooghifar, A. Aminifar, and D. Atienza, "Resource-Aware Distributed Epilepsy Monitoring Using Self-Awareness From Edge to Cloud," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 6, pp. 1338–1350, 2019.

[11] Y. Guo, Y. Zhang, M. Mursalin, W. Xu, and B. Lo, "Automated epileptic seizure detection by analyzing wearable EEG signals using extended correlation-based feature selection," in *2018 IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks, BSN 2018*, pp. 66–69, Institute of Electrical and Electronics Engineers Inc., apr 2018.

[12] I. L. Olokodana, S. P. Mohanty, and E. Kougianos, "Ordinary-kriging based real-time seizure detection in an edge computing paradigm," in *Digest of Technical Papers - IEEE International Conference on Consumer Electronics*, pp. 1–6, Institute of Electrical and Electronics Engineers Inc., jan 2020.

[13] I. L. Olokodana, S. P. Mohanty, E. Kougianos, and R. S. Sherratt, "EZcap: A Novel Wearable for Real-Time Automated Seizure Detection from EEG Signals," *IEEE Transactions on Consumer Electronics*, vol. 67, pp. 166–175, may 2021.

[14] Y. Gu, E. Cleeren, J. Dan, K. Claes, W. Van Paesschen, S. Van Huffel, and B. Hunyadi, "Comparison between Scalp EEG and Behind-the-Ear EEG for Development of a Wearable Seizure Detection System for Patients with Focal Epilepsy," *Sensors*, vol. 18, p. 29, dec 2017.

[15] D. Sopic, A. Aminifar, and D. Atienza, "E-Glass: A Wearable System for Real-Time Detection of Epileptic Seizures," in *Proceedings - IEEE International Symposium on Circuits and Systems*, pp. 1–5, Institute of Electrical and Electronics Engineers Inc., apr 2018.

[16] M. A. Sayeed, S. P. Mohanty, E. Kougianos, and H. P. Zaveri, "Neuro-Detect: A Machine Learning-Based Fast and Accurate Seizure Detection System in the IoMT," *IEEE Transactions on Consumer Electronics*, vol. 65, pp. 359–368, aug 2019.

[17] Z. Wang, J. Yang, and M. Sawan, "A Novel Multi-scale Dilated 3D CNN for Epileptic Seizure Prediction," in *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems, AICAS 2021*, (Washington, DC), pp. 1–4, Institute of Electrical and Electronics Engineers Inc., jun 2021.

[18] Z. Wang, J. Yang, H. Wu, J. Zhu, and M. Sawan, "Power efficient refined seizure prediction algorithm based on an enhanced benchmarking," *Scientific Reports*, vol. 11, dec 2021.