13th International Conference on Electrical, Electronics and Computer Engineering (IcETRAN), Silver Lake, 8–11 June 2026

Automated Detection of Epileptic Seizures from EEG: From Model Predictions to Real-World Applications

Aneta Kartali
University of Ljubljana
Faculty of Computer and Information Science
Ljubljana, Slovenia
aneta.kartali@fri.uni-lj.si
0009-0007-0134-0989

Octavian Mihai Machidon
University of Ljubljana
Faculty of Computer and Information Science
Ljubljana, Slovenia
octavian.machidon@fri.uni-lj.si
0000-0003-3133-1008

Alina Luminita Machidon
University of Ljubljana
Faculty of Computer and Information Science
Ljubljana, Slovenia
alina.machidon@fri.uni-lj.si
0000-0002-9330-3865

Veljko Pejović
University of Ljubljana
Faculty of Computer and Information Science
Ljubljana, Slovenia
veljko.pejovic@fri.uni-lj.si
0000-0002-9009-0024

***Abstract*—Reliable automated seizure detection from EEG recordings is essential for continuous, out-of-clinic monitoring and patient care. Despite the availability of standardized guidelines for evaluation, many developed approaches do not follow the recommended practices, making direct comparison difficult. Reproducibility presents another major concern due to the lack of complete documentation of preprocessing and training pipelines. Furthermore, subject-independent models frequently exhibit significant performance drops when applied to unseen subjects, implying the lack of robustness and generalizability in existing approaches. All these shortcomings hinder the development of seizure detection systems suitable for real-world applications. In this work, we analyze reported results and evaluate multiple seizure detection techniques under standardized, subject-independent settings. We introduce a subject-independent lightweight temporal convolutional model trained on continuous multichannel EEG recordings from the CHB-MIT dataset. The proposed approach yields 78.00% sensitivity and a false alarm rate of 1.59 per hour for subject-wise 10-fold cross-validation analysis. Our results further emphasize the impact of evaluation methodology on reported performance as well as the need for transparent reporting to support reproducibility and the development of reliable seizure detection systems for real-world deployment.**

***Index Terms*—seizure detection, electroencephalogram (EEG), deep learning, temporal convolution**

I. Introduction

Reliable out-of-clinic seizure detection is becoming increasingly important with the growing use of ambulatory and long-term EEG monitoring [1]. In this context, there exists a growing need for automated systems that can generalize across patients and perform reliably with minimal subject-specific adaptation [2]. Such subject-independent approaches are crucial for real-world deployment, where acquiring sufficient labeled data for each individual is impractical and often impossible. A robust generalized seizure detection system can reduce reliance on patient-specific adaptation, support large-scale use while enabling continuous monitoring in diverse populations and settings.

Seizures manifest as characteristic temporal and spectral patterns in electroencephalography (EEG) signals, which can be identified by analyzing changes in amplitude, frequency content, and spatial distribution across channels. Deep learning methods have enabled end-to-end learning directly from raw or minimally processed EEG signals, with convolutional and recurrent architectures demonstrating strong performance in capturing both spatial and temporal dependencies [3], [4]. Most existing approaches formulate seizure detection as a binary classification problem on short EEG segments, typically ranging from a few seconds to tens of seconds, where each segment is labeled as seizure or non-seizure.

Subject-independent seizure detection aims to train models that generalize across individuals by utilizing data from multiple subjects [5]. In this setting, the model is expected to output predictions for unseen subjects without requiring subject-specific adaptation. However, improper data partitioning can lead to overoptimistic performance estimates. In particular, data leakage occurs when recordings from the same subject appear in both training and evaluation sets, allowing the model to make use of subject-specific patterns rather than learning generalizable seizure characteristics. To mitigate this, strict subject-wise separation between training, validation, and test sets is required. When implemented correctly, subject-independent evaluation provides a more realistic estimate of real-world performance, as the model is tested on entirely unseen individuals.

Despite being more realistic, subject-independent approaches typically yield lower performance compared to subject-specific models due to substantial inter-subject variability in EEG signals. Differences in brain anatomy, electrode placement, seizure morphology, and recording conditions pose significant challenges for generalization. In addition, seizure detection is inherently affected by severe class imbalance, as seizure events occupy only a small fraction of long-term EEG recordings. This imbalance can bias performance

evaluation if not properly addressed. Metrics such as accuracy may be misleading in this context, while metrics that better capture detection quality under imbalance, such as F1-score, sensitivity, precision, and false alarm rate (FAR), provide a more informative assessment. However, inconsistent reporting practices and selective use of metrics can conceal true model performance, which emphasizes the need for standardized and transparent evaluation protocols [1].

Since most models operate on short EEG segments, their outputs must be aggregated to produce clinically meaningful seizure detections. This transition from segment-based predictions to event-based evaluation is non-trivial and has a significant impact on reported performance. Recent recommendations, such as those proposed in the SzCORE framework [1], advocate for standardized evaluation procedures that include clear definitions of detection latency, false alarm rate, and event matching criteria. These guidelines aim to improve comparability across studies and ensure that reported results reflect practical utility.

In this work, we propose a lightweight convolutional neural network for seizure detection from multichannel continuous EEG recordings. We evaluate the model on the widely used CHB-MIT dataset [6] using a subject-independent training protocol with strict subject-wise data separation. Unlike many prior studies, we include all available subjects in the analysis and systematically examine the impact of different evaluation and reporting practices. Through our results, we highlight the importance of transparent methodology, appropriate metric selection, and standardized evaluation in developing reliable and reproducible seizure detection algorithms suitable for real-world deployment.

II. Materials and Methods

A. Dataset

In this study, we utilize a widely used Children's Hospital Boston-Massachusetts Institute of Technology (CHB-MIT) Scalp EEG Database. The dataset consists of 982 hours of EEG data collected from 22 pediatric subjects (5 males, aged 3–22, 17 females, aged 1.5–19) with intractable seizures [6]. The recordings are grouped into 23 cases, from chb01 to chb23 (recordings chb01 and chb21 belong to the same participant, with chb21 obtained 1.5 years later). In total, the dataset contains 686 recordings, ranging from 60 minutes to 4 hours. The EEG signals were recorded using a bipolar montage, with electrodes placed according to the international 10-20 system. Signals were sampled at 256 Hz with a 16-bit resolution.

We utilized the BIDS-compatible version of the dataset publicly available on Zenodo [7]. The repository contains EEG data in European Data Format (EDF) files, modified to standardize the channel configurations across all participants. All processed recordings have 18 channels (Fp1-F3, F3-C3, C3-P3, P3-O1, Fp2-F4, F4-C4, C4-P4, P4-O2, Fp1-F7, F7-T7, T7-P7, P7-O1, Fp2-F8, F8-T8, T8-P8, P8-O2, Fz-Cz, and Cz-Pz) from a double banana bipolar montage. The dataset contains seizure annotations as start and stop times. There are 664 EDF recordings, 129 of which contain one or more seizures. In total, the records include a total of 198 seizures.

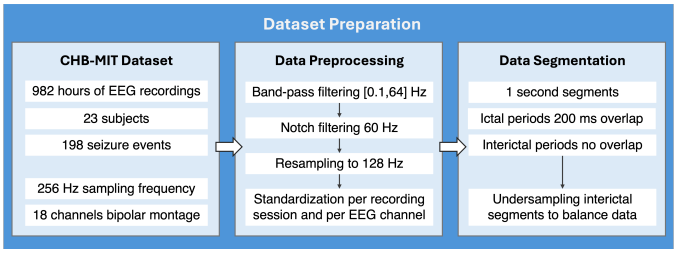

Fig. 1. Dataset preparation procedure utilized in the study, including EEG signal preprocessing and segmentation pipeline.

B. Preprocessing

All EEG data were band-pass filtered between 0.1 Hz and 64 Hz. A notch filter at 60 Hz was further applied to remove the power line noise. The data was resampled to 128 Hz to reduce the computational complexity and accelerate training. Before segmentation, all data were standardized per channel and per recording session (mean and standard deviation were calculated per recording session, for each EEG channel individually). We utilize all 18 EEG channels, thus not lowering the spatial complexity of the data seen by the model.

To train the model to detect seizures, we utilize both ictal and inter-ictal waveforms and segment the dataset into 1-second-long segments. To address the problem of class imbalance, we apply two data augmentation techniques: oversampling the minority class and sub-sampling the majority class. Data segments that belong to the seizure class are sampled with an overlap of 80% within the seizure duration. Non-seizure data were not overlapped. Furthermore, the majority class segments were randomly subsampled to match the number of seizure segments. As a result, the CHB-MIT dataset was transformed into 113000 1-second-long EEG segments used for training the proposed deep learning model to perform seizure detection. The dataset preparation procedure is depicted in Fig. 1.

C. Model Architecture

Model architecture (Fig. 2) consists of a temporal convolutional network and a linear decoder. The input sequence is a $C = 18$-channel vector of length $T = 128$ samples (1 second) and is denoted by $\mathbf{X}^{(T)} = [\mathbf{x}_1, ..., \mathbf{x}_C] \in \mathbb{R}^{C \times T}$. We opted for a simple model design, aiming for a less computationally demanding alternative that could be integrated into wearable, resource-constrained devices. Unlike recurrent and attention-based modules, temporal convolutions improve training stability and reduce processing time while maintaining large receptive fields with the use of dilations [8]. Recently, they have been successfully adopted to detect seizures from EEG data [3].

In this work, the temporal convolutional module extracts features and maps the input sequence $\mathbf{X}^{(T)}$ into an embedding vector $\mathbf{Z} = f(\mathbf{X}^{(T)})$ of size $d_f = 18$. It consists of three causal

TCN blocks, each containing two convolutional layers. Model parameters were selected through a hyperparameter search. All convolutional layers have a kernel size of 3, a stride of 1, and increasing dilations starting from the second block (1, 1, 2, 4, 8, 16). The temporal dimension across all layers is conserved with the use of zero-padding. All convolutional layers are followed by a Rectified Linear Unit (ReLU) activation function and Batch Normalization.

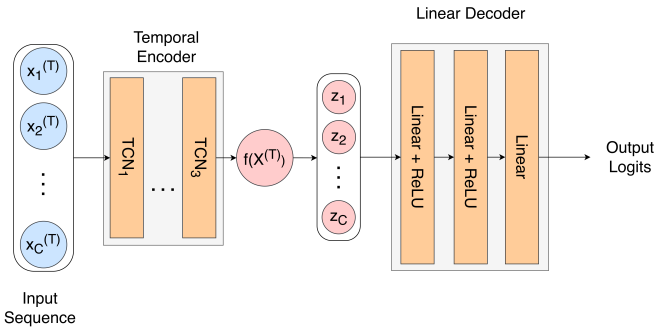

Fig. 2. Proposed model architecture for seizure detection based on raw EEG waveforms. The model consists of a 3-layer temporal convolutional module and a linear decoder.

D. Training Algorithm

Each segmented EEG interval is treated as a data example and labeled as normal or seizure. The dataset amounts to around 113000 1-second-long EEG segments. We perform a binary classification to differentiate between the seizure and normal waveforms. For all training, we employ a leave-k-participants-out cross-validation strategy or a cross-patient data partitioning approach [5]. In each fold, all 1-second segments from the same participant belong to either the training or test set, thus avoiding data leakage.

We trained the proposed model architecture employing 10-fold cross-validation. The dataset is first randomly shuffled, then divided into 10 folds using stratified folds with non-overlapping groups of subjects. This means that each subject will appear exactly once in the test set. For reproducibility purposes, we present the specific per-fold subject split, depicted in Table I.

TABLE I
SUBJECT SPLIT PER CROSS-VALIDATION FOLD. ONLY TEST SUBJECTS ARE LISTED, WHILE THE REMAINING SUBJECTS' DATA WERE UTILIZED FOR TRAINING.

CV fold	Test subjects
Fold 1	chb03, chb11, chb15
Fold 2	chb04, chb14
Fold 3	chb07, chb16, chb22
Fold 4	chb08, chb19
Fold 5	chb09, chb18
Fold 6	chb02, chb12
Fold 7	chb07, chb17
Fold 8	chb01, chb20, chb24
Fold 9	chb10, chb13, chb23
Fold 10	chb06

We optimize the models using Adam with a learning rate of 0.001 for 100 epochs. We use an objective function consisting of binary cross-entropy loss and L_2 regularization term with $\lambda_{L2} = 0.05$.

$$Loss(\mathbf{w}, \mathcal{D}) = \sum_i CE(f(\mathbf{x}_i), y_i) + \lambda_{L2} \sum_j w_j^2 \quad (1)$$

In 1, $\mathbf{w}$ represents the set of all trainable model parameters and includes encoder and decoder weights. The set of training sequences is given as $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$ where N is the total number of training sequences.

E. Evaluation Pipeline

In practice, seizure detection is performed on continuous EEG data, which, as a consequence, results in a highly imbalanced evaluation dataset. To perform seizure detection on a continuous EEG signal using the model previously trained on segmented data, we utilize an approach that relies on obtaining segment-based predictions. We first obtain segment-based probabilities from the trained model, post-process the output signal, and then retrieve segment-based and event-based seizure detection metrics. This pipeline is further illustrated in Fig. 3.

The post-processing procedure is based on filtering, thresholding, and binarization of output probabilities, similarly to [4] and [9]. First, we apply a moving average filter to smooth out the obtained probabilities, using a sliding window width of 10, averaging ten one-second predictions. Next, a threshold value H is used to generate a binary output, indicating seizure or non-seizure outcome. The post-processing procedure was adopted after observing significantly improved model performance compared to seizure detection based solely on model outputs.

Binary output of the thresholded smoothed probability signal is used to obtain segment-based predictions. On the other hand, to obtain event-level predictions, we further apply the processing rules as suggested in [1]. First, we convert sample-based predictions into events by merging consecutive positive predictions, and then we retain only events that are at least 10 seconds long. This significantly reduces the number of false positives. We further merge neighboring events if they are less than 90 seconds apart. Finally, we consider a predicted event a true positive if it overlaps with the true seizure annotation, allowing for 30-s preictal tolerance and 60-s postictal tolerance. Such a strategy is suggested in [1] to enhance sensitivity.

III. RESULTS

A. Evaluation of Model Performance

In Fig. 4, we show cross-validation performance results. These results represent the raw performance of the trained model, evaluated on the test set of each training fold, without any additional post-processing of the model predictions. Binary classification performance of our subject-independent

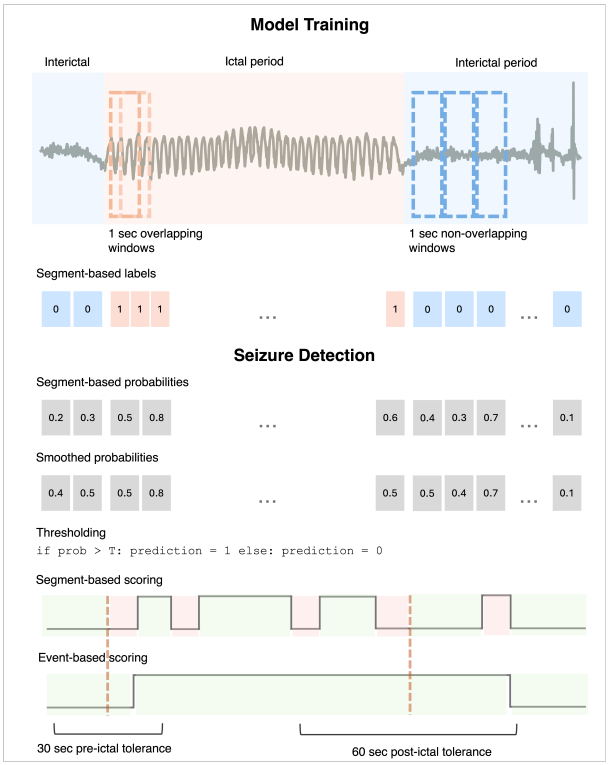

Fig. 3. Model training and evaluation pipeline.

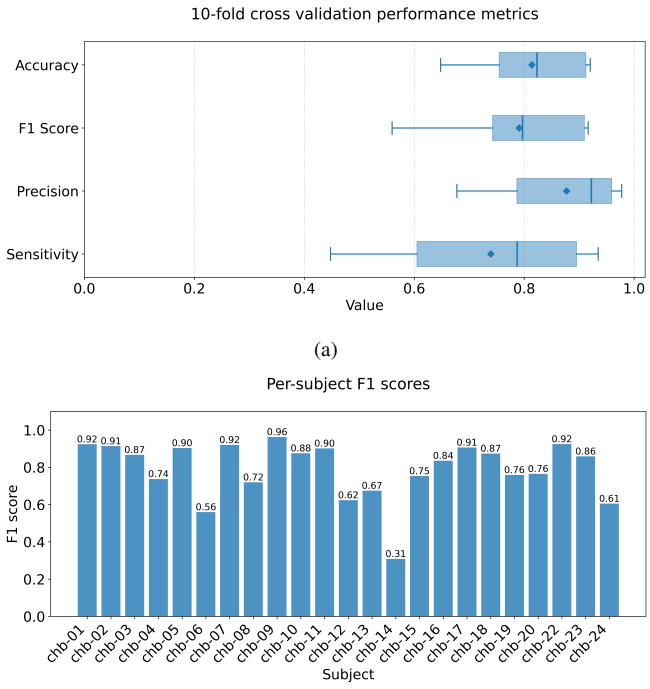

Fig. 4. (a) Cross-validation performance of the causal TCN model trained to perform the EEG-based seizure detection task. The reported results are based on raw model predictions and are aggregated over 10 folds. The model from each training fold is evaluated on subjects not included in the training set. (b) Per-subject F1-scores demonstrate subject-dependent differences and represent the source of high variability present in cross-validation results.

seizure detection model, in terms of accuracy, F1-score, precision, and sensitivity, is 81.40±9.83% (mean ± standard deviation), 79.08±12.41%, 87.71±10.50%, and 73.94±17.31%, respectively, averaged over the 10 folds. Per-subject results were obtained by evaluating each model on subjects that were not seen in the training set of that specific fold.

All performance metrics exhibit significant variation, which is a consequence of large variation in per-subject results. In particular, subjects chb06, chb12, chb14, and chb24 (Fig. 4b) exhibit the worst performance over all metrics, all of which were excluded from studies such as [9] and [10]. However, excluding the worst-performing subjects from the evaluation significantly impacts the overall reported model performance in the published works. As a consequence, this often leads to misleading results and overoptimistic solutions.

Furthermore, we compare our model's performance with other published works, focusing on studies that apply deep learning to segmented EEG data from the CHB-MIT dataset, particularly those implementing subject-independent training without data leakage. We compare segment-based seizure detection performance, and summarize the results in Table II. Additionally, we reproduce the methodology from [11] using their publicly available code. While our approach does not achieve the highest overall performance, it demonstrates significantly improved F1-score and precision on held-out subjects.

To further evaluate the performance of the proposed model architecture, we provide results for segment-based and event-based seizure detection, as recommended in [1]. For each subject, the model used for evaluation was taken from the fold in which the particular subject was not present in the training set. Furthermore, to obtain segment-based and event-based predictions, we post-process model predictions as described in Section II-E. Per-subject detection results are presented in Tables III and IV. The utilized performance metrics include segment-based: accuracy (ACC), F1 score (F1), precision (PREC), and sensitivity (SENS), as well as event-based: F1 score (F1), precision (PREC), sensitivity (SENS), false alarm rate (FAR) per hour, and latency in seconds.

Due to varying model performance on different subjects' data, we have tuned the threshold parameter individually for each subject as in [4]. We have done so to minimize false positive and false negative rates. chb06 was excluded from

TABLE II
OVERALL MODEL PERFORMANCE COMPARISON WITH PUBLISHED SUBJECT-INDEPENDENT DEEP-LEARNING MODELS ON SEGMENT-BASED SEIZURE DETECTION. ALL METRICS ARE GIVEN IN PERCENTAGES. *FOR [11], WE REPORT RESULTS OBTAINED AFTER REPEATING THE AUTHORS' APPROACH BASED ON THE AVAILABLE CODE.

Approach	ACC	F1	PREC	SENS
Jana et al. (2023) [12]	75.32	-	-	74.45
Ali et al. (2024) [2]	85.06	5.46	2.97	64.24
Lim et al. (2025) [3]	82.66	68.02	68.09	67.95
Kashefi et al. (2025) [11]*	79.76	38.13	98.68	23.83
Our approach	81.40	79.08	87.71	73.94

the result aggregation, since all but one seizure event from this subject is less than 10 seconds long and thus could not be correctly detected with our approach.

To improve transparency and support reproducibility, we provide information about each threshold applied in Table III. The same thresholds are further applied to obtain event-level predictions reported in Table IV.

TABLE III
PER SUBJECT SEGMENT-BASED PERFORMANCE EVALUATION. ALL METRICS, INCLUDING ACCURACY (ACC), F1-SCORE (F1), PRECISION (PREC), AND SENSITIVITY (SENS), ARE REPORTED IN PERCENTAGES.

Subject	**Thr.**	**Segment-Based Evaluation**			
ID	**H**	***ACC***	***F1***	***PREC***	***SENS***
chb01	0.7	88.73	63.92	54.33	77.62
chb02	0.8	85.50	77.92	86.33	71.01
chb03	0.7	79.45	42.97	33.72	59.24
chb04	0.8	82.35	13.64	7.62	65.24
chb05	0.9	88.66	8.01	4.20	84.99
chb06	0.4	62.84	0.43	0.21	34.96
chb07	0.8	86.28	58.21	48.55	72.67
chb08	0.5	84.45	42.55	30.38	71.01
chb09	0.9	82.71	68.86	72.65	65.44
chb10	0.9	82.61	77.99	96.96	65.22
chb11	0.8	87.00	73.75	73.39	74.12
chb12	0.4	62.01	3.18	1.67	38.47
chb13	0.9	67.21	45.14	65.34	34.47
chb14	0.8	72.48	9.65	5.40	45.61
chb15	0.8	87.29	27.24	16.52	77.60
chb16	0.7	86.82	32.38	20.72	73.99
chb17	0.8	82.65	3.48	1.78	83.40
chb18	0.9	79.66	65.22	72.38	59.35
chb19	0.8	63.71	40.70	78.55	27.46
chb20	0.6	88.97	22.61	13.18	79.26
chb22	0.7	89.80	40.61	27.23	79.90
chb23	0.8	77.52	37.62	28.45	55.51
chb24	0.6	85.92	26.16	15.93	73.07
Average	0.75	81.44	40.08	38.89	65.21

When individual subjects are investigated (Table III and Table IV), it can be observed that chb06, chb12, and chb14 exhibit the worst overall performance. This is also the case in [2], and the mentioned subjects have been excluded from other studies [9], [10]. On the other hand, the best overall performance is achieved for subjects chb02, chb09, chb10, and chb11.

Seizure detection for the majority of subjects is performed by adopting a threshold of 0.7 or higher. Having in mind that the model outputs probabilities in the range from 0 to 1, and that the classification threshold is most commonly set to 0.5, the distinction between positive and negative cases is not trivial when the model performs classification from raw EEG waveforms. This per-subject threshold adaptation is also reported in [4] and [9].

On the seizure event detection task (Table IV), compared to the performance reported in [2], our model achieves only 0.65% lower sensitivity, while maintaining a significantly lower average false alarm rate of 0.65 per hour, compared to the 5.32 reported in [2]. On the other hand, when selecting a subject-specific threshold to maximize sensitivity, the model achieves an accuracy of 78.00% and a false alarm rate of 1.59 per hour. In this case, the model achieves perfect sensitivity

TABLE IV
PER SUBJECT EVENT-BASED PERFORMANCE EVALUATION BASED ON THE TRAINED CAUSAL TCN MODEL. F1-SCORE (F1), PRECISION (PREC), AND SENSITIVITY (SENS) ARE GIVEN IN PERCENTAGES. FALSE ALARM RATE (FAR) IS CALCULATED PER HOUR, AND THE DETECTION LATENCY (LAT) IS REPORTED IN SECONDS.

Subject	**Thr.**	**Event-Based Evaluation**				
ID	**H**	***F1***	***PREC***	***SENS***	***FAR***	***LAT [s]***
chb01	0.7	51.61	40.00	72.72	0.16	7.00
chb02	0.8	80.00	100.00	66.67	0.00	8.00
chb03	0.7	52.63	41.66	71.43	0.18	9.0
chb04	0.8	25.00	15.00	75.00	0.11	18.67
chb05	0.9	7.94	4.13	100.00	2.97	8.00
chb06	0.4	0.72	0.37	10.00	3.86	4.00
chb07	0.8	40.00	25.00	100.00	0.13	11.00
chb08	0.5	29.41	17.24	100.00	1.20	3.2
chb09	0.9	88.89	80.00	100.00	0.01	14.00
chb10	0.9	85.71	85.71	85.71	0.02	6.83
chb11	0.8	70.00	70.00	70.00	0.04	5.42
chb12	0.4	10.92	6.40	37.21	3.31	6.75
chb13	0.9	51.61	66.67	42.10	0.04	9.62
chb14	0.8	9.52	5.88	25.00	0.26	18.67
chb15	0.8	21.74	12.99	66.67	1.16	6.25
chb16	0.7	24.39	17.86	38.46	0.26	5.6
chb17	0.8	8.63	4.58	75.00	2.08	9.0
chb18	0.9	85.71	81.82	90.00	0.02	12.88
chb19	0.8	87.50	87.50	87.50	0.02	14.29
chb20	0.6	35.42	22.10	89.47	0.59	7.65
chb22	0.7	32.65	24.24	50.00	0.21	6.25
chb23	0.8	40.00	30.61	57.69	0.29	5.33
chb24	0.6	40.28	26.61	82.86	0.65	6.10
Average	0.75	44.52	39.36	71.98	0.62	9.07

for half of the subjects, while chb06, chb12, and chb14 remain the worst-performing cases.

IV. DISCUSSSION

Throughout our experiments, we observed that the choice of the threshold H used during post-processing has a significant impact on the results. Lowering the threshold can significantly improve segment-based accuracy and sensitivity. Specifically, a lower threshold reduces missed detections, increasing the true positive rate (correct detections) and decreasing the false negative rate, thereby improving sensitivity. Although this also increases the number of false positives, the highly imbalanced nature of continuous EEG data leads to an overall improvement in accuracy. Finally, due to the high class imbalance, specificity remains high despite the increased number of false positives.

On the contrary, from the perspective of event-based seizure detection, the choice of the threshold H is highly subject-dependent. In other words, if high sensitivity is a priority, the threshold must be tuned for each subject. We should, however, keep in mind that higher threshold values reduce the false alarm rate but at a cost of increased detection latency.

Increasing the threshold H improves the precision of both segment-based and event-based detection by reducing the number of false positives, which in turn positively impacts the F1-score. Given these trade-offs, studies that do not report both F1-score and precision for segment-based and event-based evaluation may present a misleading picture of model performance.

The dataset utilized in this study [7] is highly imbalanced. Seizure events are rare, representing only 0.34% of the total data, and their distribution is uneven across subjects. As a consequence, some subjects contribute with fewer seizure events. When these subjects' data is included in training, the model is exposed to fewer examples and reduced variability of seizure patterns, which can increase the misclassification rate [2].

This imbalance also limits the model's exposure to non-seizure patterns. During evaluation, the model encounters long continuous segments of non-seizure EEG, many of which differ from those seen during training due to dataset balancing procedures, such as the one applied in this study. This introduces additional uncertainty and can further increase misclassification. Performance degradation associated with subject-independent training can be observed in prior studies [2], [13], [14].

Finally, the CHB-MIT dataset consists of recordings acquired within a single clinical setting under equally controlled conditions. Even when a model performs well on unseen subjects from this dataset, such performance does not necessarily generalize to other datasets. This limitation is particularly relevant for EEG data recorded in out-of-clinic or home environments, where variability can be induced from a multitude of factors. Thus, we highlight the necessity to evaluate models not only on unseen subjects but also across datasets not utilized during model training. Such an evaluation will yield a more realistic assessment of the model's robustness and applicability in real-world settings.

V. Conclusion

In this paper, we present a lightweight CNN-based approach for seizure detection, with a focus on analyzing threshold-based post-processing techniques and the evaluation metrics used to report model performance. We adopt a subject-independent training strategy to promote generalization across individuals. All subjects from the CHB-MIT dataset are included in both training and evaluation, thus ensuring an overall assessment of model performance.

The proposed causal temporal convolutional model requires only 1 second of multichannel EEG data to produce a prediction, enabling a lower latency for seizure onset detection in potential real-time applications. The model achieves an average sensitivity of up to 78.00% on unseen subjects, with a false alarm rate of 0.62 per hour. While these results exceed those reported in related work, our findings confirm that inter-subject variability presents a significant challenge for subject-independent seizure detection. Therefore, increasing data diversity, incorporating additional datasets, and improving seizure pattern representation present essential steps for developing robust and reliable seizure detection systems.

We plan to extend our research investigating different techniques to further differentiate between seizure and non-seizure segments in the latent space. Furthermore, our future work will explore the precision of seizure onset detection and its feasibility for integration in wearable, real-world seizure-monitoring systems. Finally, as our goal is to arrive at a system that would be feasible to use in real-world conditions, we will investigate approaches for making the model more robust to input data variability, testing our developed approaches on datasets not used for model training and validation.

Our future work will investigate techniques for better separating seizure and non-seizure representations in the latent space, as well as improving the precision of seizure onset detection. To explore the feasibility of deploying seizure detection models in wearable, real-world monitoring systems, we will further focus on improving robustness to input variability by evaluating the model on datasets not used during training or validation.

Acknowledgments

Research presented in this publication was partly funded by the Slovenian Research Agency project "approXimation for adaptable distributed artificial intelligence" (N2-0393), as well as core funding No. P2-0098.

References

[1] J. Dan, U. Pale, A. Amirshahi, W. Cappelletti, T. M. Ingolfsson, X. Wang, A. Cossettini, A. Bernini, L. Benini, S. Beniczky *et al.*, "Szcore: seizure community open-source research evaluation framework for the validation of electroencephalography-based automated seizure detection algorithms," *Epilepsia*, vol. 66, pp. 14–24, 2025.

[2] E. Ali, M. Angelova, and C. Karmakar, "Epileptic seizure detection using chb-mit dataset: The overlooked perspectives," *Royal Society open science*, vol. 11, no. 5, 2024.

[3] Z. Y. Lim, Y. H. Pang, S. Y. Ooi, W. H. Khoh, and F. S. Hiew, "Mltcn-eeg: metric learning-based temporal convolutional network for seizure eeg classification," *Neural Computing and Applications*, vol. 37, no. 4, pp. 2849–2875, 2025.

[4] H. Li, X. Dong, X. Zhong, C. Li, H. Cui, and W. Zhou, "End-to-end model for automatic seizure detection using supervised contrastive learning," *Engineering Applications of Artificial Intelligence*, vol. 133, p. 108665, 2024.

[5] X. Zhang, X. Zhang, Q. Huang, and F. Chen, "A review of epilepsy detection and prediction methods based on eeg signal processing and deep learning," *Frontiers in Neuroscience*, vol. 18, p. 1468967, 2024.

[6] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, T. Treves, and J. Guttag, "Patient-specific seizure onset detection," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*. IEEE, 2004, pp. 419–422.

[7] D. Jonathan and A. Shoeb, "BIDS CHB-MIT Scalp EEG Database," 2023.

[8] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," August 2018. [Online]. Available: http://arxiv.org/abs/1803.01271

[9] Y. G. Chung, A. Cho, H. Kim, and K. J. Kim, "Single-channel seizure detection with clinical confirmation of seizure locations using chb-mit dataset," *Frontiers in Neurology*, vol. 15, p. 1389731, 2024.

[10] M. Shen, P. Wen, B. Song, and Y. Li, "An eeg based real-time epilepsy seizure detection approach using discrete wavelet transform and machine learning methods," *Biomedical Signal Processing and Control*, vol. 77, p. 103820, 2022.

[11] H. Kashefi Amiri, M. Zarei, and M. R. Daliri, "Epileptic seizure detection from electroencephalogram signals based on 1d cnn-lstm deep learning model using discrete wavelet transform," *Scientific Reports*, vol. 15, no. 1, p. 32820, 2025.

[12] G. C. Jana, K. Swami, and A. Agrawal, "Capsule neural network based approach for subject specific and cross-subjects seizure detection from eeg signals," *Multimedia Tools and Applications*, vol. 82, no. 23, pp. 35 221–35 252, 2023.

[13] Z. Wei, J. Zou, J. Zhang, and J. Xu, "Automatic epileptic eeg detection using convolutional neural network with improvements in time-domain," *Biomedical Signal Processing and Control*, vol. 53, p. 101551, 2019.

[14] Y. Zhao, G. Zhang, Y. Zhang, T. Xiao, Z. Wang, F. Xu, and Y. Zheng, "Multi-view cross-subject seizure detection with information bottleneck attribution," *Journal of Neural Engineering*, vol. 19, no. 4, p. 046011, 2022.