2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Seizure Prediction using Convolutional Neural Networks and Sequence Transformer Networks

Ryan Chen and Keshab K. Parhi, *Fellow*, *IEEE*

Abstract— **Accurate seizure prediction is important for design of wearable and implantable devices that can improve the lives of subjects with epilepsy. Such implantable devices can be used for closed-loop neuromodulation. However, there are many challenges that inhibit the performance of prediction models. One challenge in accurately predicting seizures is the nonstationarity of the EEG signals. This paper presents a patient-specific deep learning approach to improve predictive performance by transforming EEG data before extracting features for seizure prediction. In the proposed approach, a Sequence Transformer Network (STN) is first used to learn temporal and magnitude invariances in EEG data. The proposed method further computes the short-time Fourier transform (STFT) of the transformed EEG signals as input features to a convolutional neural network (CNN). A k-out-of-n post-processing method is used to reduce the significance of isolated false positives. The approach is tested using intracranial EEG from the American Epilepsy Society Seizure Prediction Challenge dataset. Leave-one-out cross validation is used to evaluate the model. The proposed model achieves an overall sensitivity of 82%, false prediction rate of 0.38/hour, and average AUC of 0.746.**

I. Introduction

Epilepsy is one of the most common neurological diseases, affecting approximately 70 million people worldwide [1]. It is a disorder in the brain that is diagnosed when an individual has had two or more unprovoked seizures. Epileptic seizures are accompanied by abnormal electrical activity in the brain, as opposed to other types of seizures that may be caused by external factors such as head injury or reaction to medication. Because epileptic seizures occur with no apparent cause, the ability to predict the onset of an epileptic seizure could considerably increase the quality of life in epileptic patients. Providing a warning or alarm prior to a seizure would allow epileptic patients to utilize treatment such as anti-epileptic drugs or brain stimulation more effectively. Seizure prediction can be accomplished by analyzing electroencephalogram (EEG) or intra-cranial EEG (iEEG) recordings of the brain.

The EEG pattern of an epileptic patient can be split into four phases: preictal phase, occurring before the actual seizure; ictal phase, which includes the onset of the seizure and the time when an individual is actively seizing; postictal phase, occurring after the seizure; and the interictal phase, accounting for all other time periods. Specifically, the seizure prediction problem can be defined as a binary classification of EEG data between the interictal and preictal phases. This is possible because studies have demonstrated that EEG patterns during the preictal phase are different than patterns during the interictal phase [2]. However, the underlying cause behind an epileptic seizure remains unclear. This has led to extensive research on determining the features of EEG signals that best differentiate the interictal and preictal phases of epileptic seizures. Previous features used for seizure prediction include power spectral density, short-time Fourier transform, and univariate linear measures [3]–[6] of the EEG signal.

Past seizure prediction algorithms utilizing traditional machine learning methods such as support vector machines (SVM) have achieved high sensitivity and low false positive rates [4], [7]. However, algorithms that achieve outstanding results often rely on hand-crafted, patient-specific feature engineering which can be time consuming. Additionally, EEG signals are nonstationary [8]. Thus, the features that are optimal for predicting seizures could change over time and render a prediction model obsolete. This paper attempts to use a sequence transformer network (STN) to first transform EEG data, in an effort to overcome some of the challenges associated with the nonstationary nature of EEG signals. The STN learns parameters to apply temporal and magnitude transformations to EEG data. In the proposed approach, EEG data is first transformed with the STN. The short-time Fourier transform is then applied to the transformed EEG data and used as input feature maps to a convolutional neural network (CNN) for further feature extraction and classification.

II. Materials and Methods

A. Data Description

The American Epilepsy Society Prediction Challenge dataset consists of intracranial EEG from five canine subjects and two human subjects [9]. The Institutions Ethical Review Board approved all experimental procedures involving human patients. The experimental procedures involving animal models were approved by the Institutional Animal Care and Ethics Committee.

Intracranial EEG was recorded from dogs with naturally occurring epilepsy using an ambulatory monitoring system. For four of the canine subjects, intracranial EEG was sampled from 16 electrodes at 400 Hz. For the last canine subject, EEG was sampled from 15 electrodes at 400 Hz. The canine data supplied for the American Epilepsy Society Prediction Challenge is a subset of long duration recordings, spanning multiple months up to a year. To avoid potential contamination between interictal, preictal, and postictal EEG signals, interictal data is taken from at least one week before or after any seizure.

The human patients' intracranial EEG are sampled at 5000 Hz from a varying number of electrodes: 15 electrodes for

This research was supported in part by the National Science Foundation under grant number CCF-1954749.

R. Chen and K. K. Parhi are with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455, USA (email: chen5296@umn.edu; parhi@umn.edu).

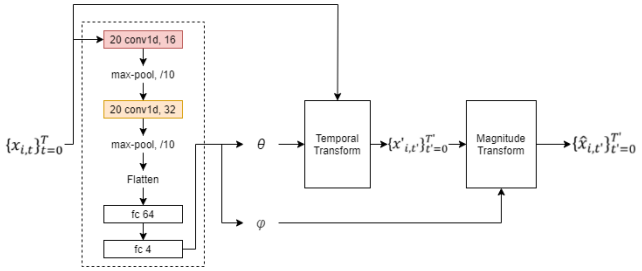

Fig. 1: The architecture for the Sequence Transformer Network used. Input EEG signal denoted as $\{x_{i,t}\}_{t=0}^{T}$. Convolution layers are annotated with the filter size, convolution type, and number of filters. Convolution layers have stride 1 and ReLU activation. Max-pool layers have the same size and stride of 10. Temporal Transform is applied with a discrete mapping, and the Magnitude Transform applies a linear transformation.

Patient 1, and 24 electrodes for Patient 2. The human data is taken from shorter monitoring sessions, which may last less than one week. To avoid potential contamination between the different phases of the EEG signal, interictal data is taken from at least four hours before or after any seizure.

The data is supplied as a collection of ten minute EEG clips, labeled as either preictal or interictal. The preictal data covers one hour prior to each seizure with a five minute seizure horizon (i.e., from 1:05 to 0:05 before seizure onset). This paper uses only 30 minutes of preictal data (i.e., from 0:35 to 0:05 before seizure onset). Additionally, the EEG data is downsampled to reduce storage requirements and computational complexity: 200 Hz for dog data, and 1000 Hz for human data.

B. Sequence Transformer Network

One of the biggest challenges in predicting seizures is the nonstationary nature of EEG signals. This introduces complexity when attempting to extract features for classification because metrics such as amplitude, offset, and time scale of the EEG signal could vary over time within a class. Therefore, applying a transform to align EEG signals in both the time and magnitude axes may help overcome these complexities. Inspired by [10], we use an STN to learn and apply such transformations to exploit invariances in EEG signals.

The STN (shown in Fig. 1) uses a CNN to learn the parameters θ and φ. Specifically, the CNN consists of two convolution layers, two max-pool layers, and two fully-connected layers. The first convolution layer has 16 filters with size 20 and stride 1, using rectified linear unit (ReLU) activation function. The second convolution layer has 32 filters with size 20 and stride 1, and ReLU activation. Both of the max-pooling layers pool over a region of size 10 with stride 10. Prior to each fully-connected layer is a dropout layer with dropout rate 0.5. The first fully-connected layer has 64 nodes followed by a ReLU activation function. The second fully-connected layer outputs four values: θ_1 and θ_0 for the θ parameter, and φ_1 and φ_0 for the φ parameter. That is,

$$\begin{bmatrix}\theta \\ \varphi\end{bmatrix} = \begin{bmatrix}\theta_1 & \theta_0 \\ \varphi_1 & \varphi_0\end{bmatrix}$$

The parameters θ and φ are used to apply temporal and magnitude transformations, respectively, to the EEG signal. This network is typically incorporated into an end-to-end trainable architecture. As shown in Fig. 3, the proposed seizure prediction architecture is end-to-end trainable and uses the STN prior to feature extraction and classification by a CNN.

The temporal transformation parameter θ maps each sample in the input signal to a new temporal location (i.e., from time t to t' for the time steps $t = 0, ..., T$). This operation is linear and is shown by the expression,

$$t' = \theta \begin{bmatrix}t \\ 1\end{bmatrix} = \begin{bmatrix}\theta_1 & \theta_0\end{bmatrix} \begin{bmatrix}t \\ 1\end{bmatrix} = \theta_1 t + \theta_0$$

However, $t' = \theta_1 t + \theta_0$ is not guaranteed to be an integer, so the value of $x_{t'}$ must be inferred in this case. This is completed by linearly interpolating between the two nearest points. For example, $x_{2.5}$ does not exist in the original data signal; so we calculate $x'_{2.5}$ as the average of x_2 and x_3. The temporal transform maps each sample x_t to $x'_{t'} = x'_{\theta_1 t + \theta_0}$. Although the EEG signal could be stretched, compressed, or shifted along the temporal axis, the order of the samples is preserved because the transform is linear.

The magnitude transform aims to exploit amplitude or offset invariances that could exist in the EEG signal. The parameter φ is used to apply a linear transformation to the value of each sample (after temporal transformation) and is shown by the expression,

$$\hat{x}_{t'} = \varphi \begin{bmatrix}x'_{t'} \\ 1\end{bmatrix} = \begin{bmatrix}\varphi_1 & \varphi_0\end{bmatrix} \begin{bmatrix}x'_{t'} \\ 1\end{bmatrix} = \varphi_1 x'_{t'} + \varphi_0$$

This transformation allows the signal to be stretched, compressed, or shifted along the magnitude axis. Overall, the temporal and magnitude transformations in the STN attempt to normalize the nonstationary EEG signal. This could improve prediction performance for long-term EEG recordings.

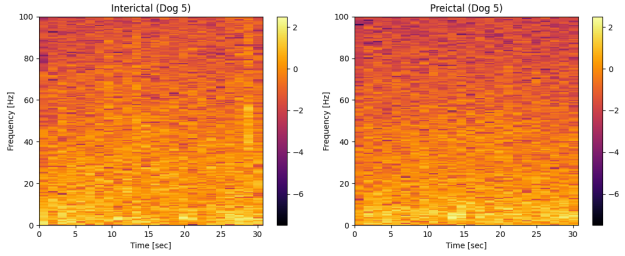

Fig. 2: Logarithm of the absolute value of the short-time Fourier transform is shown for Dog 5 in the American Epilepsy Society Prediction Challenge dataset. The same electrode is shown for the interictal (*left*) and preictal (*right*) data clips.

TABLE I: Summary of the dataset used to evaluate the proposed seizure prediction model. Clips are 30 seconds long

Subject	Sampling Frequency (Hz)	Total Interictal Hours	Total # of Seizures	Oversampling		Total # of Preictal Clips	Total # of Interictal Clips
				Time Step (sec)	Preictal Clips Generated		
Dog 1	200	80	4	0.730	9348	9588	9600
Dog 2	200	83.3	7	1.250	9555	9975	9996
Dog 3	200	240	12	0.730	28044	28764	28800
Dog 4	200	134	17	1.935	14994	16014	16080
Dog 5	200	75	5	0.980	8685	8985	9000
Patient 1	1000	8.3	3	6.333	810	990	1000
Patient 2	1000	7	3	7.916	648	828	840

C. Data Preprocessing

In order to use a two-dimensional CNN for further feature extraction and classification, the short-time Fourier transform (STFT) is used to convert the EEG signal into an image-like matrix with time and frequency axes. First, the EEG signal is split into 30 second clips. For each clip, the STFT is computed from 512-point FFTs using a Hann window with 50% overlap. The logarithm of the absolute value of the STFT is taken as the input feature map to the CNN. This results in a $n \times 22 \times 257$ feature map for the dog EEG at 200 Hz and a $n \times 116 \times 257$ feature map for the human EEG at 1000 Hz, where the dimensions are $channels \times time \times frequency$. An example of the transformed data for a single electrode of Dog 5 is shown in Fig. 2.

Additionally, the dataset supplied for the American Epilepsy Society Prediction Challenge contains much more interictal data than preictal data for all of the subjects. This is undesirable as highly imbalanced datasets often lead to poor performance in classification with a CNN [11]. To overcome this challenge, the preictal data is oversampled using a sliding window. The window is set to 30 seconds to match the length of the original data clips and is shifted across the time axis of the raw EEG signal. The time step is chosen such that the total number of preictal data clips after oversampling is similar to the number of interictal data clips. Table I shows the sliding window step size and the number of preictal clips generated for each subject. If the total number of clips is not equal in the preictal and interictal classes after oversampling, then clips are randomly removed from the majority class until they are equal. As seen in Table I, a number of interictal clips must be removed to balance the classes for all the subjects.

D. Proposed Prediction Model

The architecture of the proposed prediction model is shown in Fig. 3 (*right*). The EEG data is transformed by the STN as described previously. Then the STFT is computed for the transformed data as input to the CNN. The CNN consists of three convolution layers and two fully-connected layers. Each convolution layer contains a convolutional layer with ReLU activation, a batch normalization layer, and a max-pooling layer. The number of filters for the three convolutional layers is 16, 32, and 64 respectively. The first convolutional layer has filters of size 5×5 with stride 2×2. The larger stride size acts as an additional downsampling operation. For the remaining two convolutional layers, filters of size 3×3 with stride 1×1 are used. All of the max-pooling layers have size 2×2. After the third max-pooling layer, the features are flattened. The first fully-connected layer has 128 nodes and is followed by a sigmoid activation function. The second fully-connected layer has 2 nodes and is followed by a softmax activation function for classification. There are two dropout layers, each one is placed prior to each fully-connected layer with 0.5 dropout rate.

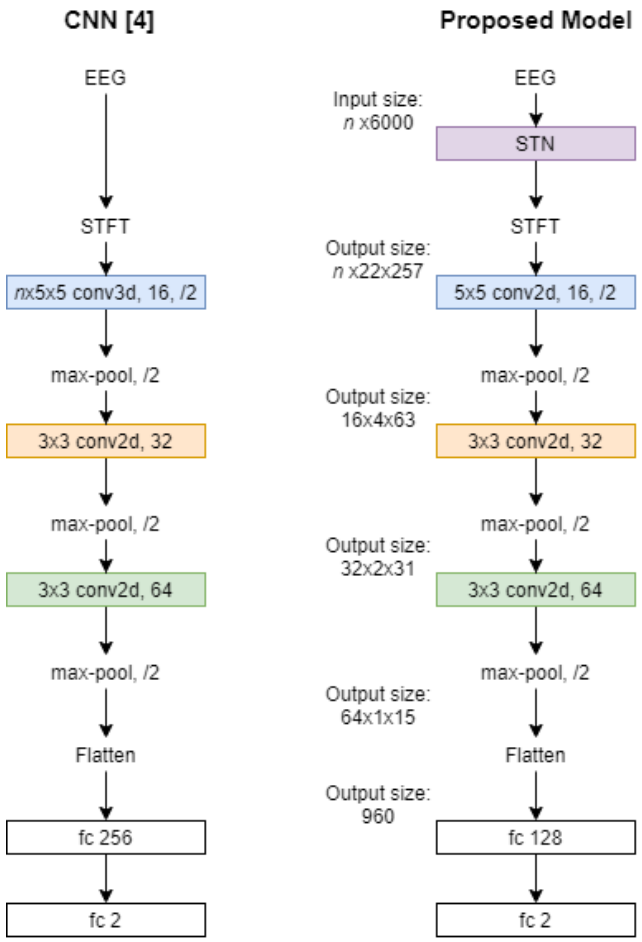

Fig. 3: The architecture for the convolutional neural networks used for seizure prediction. Convolution layers are annotated with the filter size, convolution type, and number of filters. The first convolution layer has stride 1x2x2 in the *left* model and stride 2x2 in the *right* model. Max-pool layers have size 2x2. *n* refers to the number of channels in the EEG data. **Left:** the CNN model from [5] as reference. Output sizes of each layer are not shown. **Right:** the proposed model incorporates a STN in addition to the CNN. The output sizes shown are for the Dog EEG downsampled to 200 Hz using short-time Fourier transform with 30s window, 512pt FFT, and 50% overlap.

Adam optimization is used with learning rate $1e{-}5$, β_1 value 0.9, and β_2 value 0.999 [12]. Binary cross-entropy is used as the loss-function. The model is trained for 50 epochs with early stopping regularization based on the calculated loss on the validation set. The model is implemented in Python 3.8.5 with Tensorflow 2.3.0 backend.

In practice, a user should not be notified of an impending

seizure if a single 30 second input results in a preictal classification. This is because isolated false positives are common during interictal periods. Thus, in order to mitigate the effects of isolated false positives, a seizure is predicted only if *k*-out-of-*n* inputs result in a preictal classifictaion. The proposed prediction model uses *k*=6 and *n*=8, meaning that a seizure is predicted when 180 seconds out of the last 240 seconds result in preictal classification.

After a seizure is predicted, an alarm period starts and consists of the seizure prediction horizon (SPH) and seizure occurrence period (SOP). The SPH allows for intervention before the seizure occurs and is set to 5 minutes as provided by the American Epilepsy Society Prediction Challenge dataset. The SOP is the time frame where the seizure is predicted to occur and is set to 30 minutes. Hence, only 30 minutes of preictal data is used per seizure. Further predictions are disabled for the 35 minute duration of the alarm period.

III. Experimental Results

The proposed model was trained and tested on each subject individually using leave-one-out cross validation. This means that for a subject with N seizures, we create N folds where each fold contains preictal data for one seizure. The interictal data is also partitioned evenly into the N folds. Each fold is used once as the test set and the remaining $N-1$ folds are used in the training process. The data in the $N-1$ folds are split into 75% for training and 25% for validation. The loss on the validation data is monitored for early stopping. Preictal clips that were generated from oversampling are removed from the test set. The test results are averaged over three independent runs where the interictal data is randomly shuffled.

TABLE II: Prediction Performance of the Proposed System

Subject	Sensitivity		False Predictions (/hour)		AUC	
	CNN [5]	STN + CNN	CNN [5]	STN + CNN	CNN [5]	STN + CNN
Dog 1	50%	50%	0.19	0.66	-	0.594
Dog 2	100%	100%	0.04	0.20	-	0.956
Dog 3	58.3%	78%	0.14	0.30	-	0.860
Dog 4	78.6%	92%	0.48	0.56	-	0.795
Dog 5	80%	80%	0.08	0.14	-	0.933
Patient 1	100%	78%	0.42	0.24	-	0.710
Patient 2	66.7%	44%	0.86	1.33	-	0.373
Total	75%	82%	0.21	0.38	-	0.746

Table II shows the results of the proposed model in comparison to the results of a CNN without STN from [5]. The structure of the compared CNN is shown in Fig. 3 (*left*). The sensitivity of the proposed model is higher for Dog 3 and Dog 4 at 78% and 92% in comparison to the results of the CNN at 58.3% and 78.6%, respectively. However, in the proposed model the sensitivity for Patient 1 and Patient 2 is lower at 78% and 44% compared to the results of the CNN at 100% and 66.7%, respectively. This suggests that the use of an STN may not work well on short-term EEG recordings with few training examples. Overall, the proposed model achieves an average sensitivity of 82%, an average false prediction rate of 0.38/hour, an average prediction lead time of 29 minutes, and an average AUC of 0.746.

IV. Conclusion

In this paper, a patient-specific model for seizure prediction using an STN to normalize EEG data before feature extraction and classification by a CNN has been proposed. The proposed model achieves an average sensitivity of 82%, an average false prediction rate of 0.38/hour, an average prediction lead time of 29 minutes, and an average AUC of 0.746. Prediction performance for subjects with short-term EEG recordings is poor, suggesting that a STN may not be effective at exploiting temporal and magnitude invariances without longer term recordings. More research should be conducted to determine methods to overcome the challenges of the nonstationary EEG signal.

References

[1] "Epilepsy." [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/epilepsy

[2] K. Gadhoumi, J.-M. Lina, and J. Gotman, "Discriminating preictal and interictal states in patients with temporal lobe epilepsy using wavelet analysis of intracerebral EEG," *Clinical Neurophysiology*, vol. 123, no. 10, pp. 1906–1916, Oct. 2012. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1388245712002179

[3] Y. Park, L. Luo, K. K. Parhi, and T. Netoff, "Seizure prediction with spectral power of EEG using cost-sensitive support vector machines," *Epilepsia*, vol. 52, no. 10, pp. 1761–1770, Jun. 2011. [Online]. Available: https://doi.org/10.1111/j.1528-1167.2011.03138.x

[4] Z. Zhang and K. K. Parhi, "Seizure prediction using polynomial SVM classification," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Milan: IEEE, Aug. 2015, pp. 5748–5751. [Online]. Available: https://ieeexplore.ieee.org/document/7319698/

[5] N. D. Truong, A. D. Nguyen, L. Kuhlmann, M. R. Bonyadi, J. Yang, S. Ippolito, and O. Kavehei, "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Networks*, vol. 105, pp. 104–111, Sep. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0893608018301485

[6] B. Abbaszadeh, T. Haddad, and M. C. E. Yagoub, "Probabilistic prediction of Epileptic Seizures using SVM," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Berlin, Germany: IEEE, Jul. 2019, pp. 3442–3445. [Online]. Available: https://ieeexplore.ieee.org/document/8856286/

[7] Z. Zhang and K. K. Parhi, "Low-Complexity Seizure Prediction From iEEG/sEEG Using Spectral Power and Ratios of Spectral Power," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 10, no. 3, pp. 693–706, Jun. 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7307237/

[8] W. Klonowski, "Everything you wanted to ask about EEG but were afraid to get the right answer," *Nonlinear Biomedical Physics*, vol. 3, no. 1, p. 2, Dec. 2009. [Online]. Available: https://nonlinearbiomedphys.biomedcentral.com/articles/10.1186/1753-4631-3-2

[9] "American Epilepsy Society Seizure Prediction Challenge." [Online]. Available: https://kaggle.com/c/seizure-prediction

[10] J. Oh, J. Wang, and J. Wiens, "Learning to Exploit Invariances in Clinical Time-Series Data using Sequence Transformer Networks," *arXiv:1808.06725 [cs, stat]*, Aug. 2018, arXiv: 1808.06725. [Online]. Available: http://arxiv.org/abs/1808.06725

[11] M. Buda, A. Maki, and M. A. Mazurowski, "A systematic study of the class imbalance problem in convolutional neural networks," *Neural Networks*, vol. 106, pp. 249–259, Oct. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0893608018302107

[12] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv:1412.6980 [cs]*, Jan. 2017, arXiv: 1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980