Epileptic Spike Detection by Recurrent Neural Networks with Self-Attention Mechanism

Kosuke Fukumori[1], Noboru Yoshida[2], and Toshihisa Tanaka[1,3,4,5,6]

***Abstract*— Automated identification of epileptiform discharges for the diagnosis of epilepsy can mitigate the burden of the exhaustive manual search in electroencephalogram (EEG). Recent studies have indicated that a two-step method that consists of detection of candidate waveforms with signal processing and pattern matching followed by machine learning-based classification is effective. However, the overall performance depends on the detector of candidates. This paper thus considers a scenario without candidate waveforms, that is, we propose a recurrent neural network (RNN)-based self-attention model that can be fitted from the EEG segments generated without detecting spike candidates. In comparison with the state-of-the-art machine learning models which can be applied for EEG classification (LightGBM and EEGNet), the proposed model achieved higher performance (average accuracy: 90.2%). This result strongly suggests that the self-attention mechanism is suitable to an automated identification of the epileptiform discharge in the EEG.**

I. Introduction

Epilepsy is a chronic brain disease that causes seizures of disorientation and convulsions associated with excessive electrical excitation of neurons in the brain. It is estimated that there are about 50 million people with epilepsy in the world, and the chronic lack of epilepsy specialists (epileptologists) is a severe problem. For example, in Japan, there are only 700 epileptologists, although there are one million patients [1]. Besides, since the diagnosis of epilepsy requires specialized knowledge or skill, a great deal of epileptologist's work is necessary to care for each patient. This has motivated the development of an automated diagnostic aid to support epileptologists.

One of the diagnosis process steps is the measurement and analysis of the electroencephalogram (EEG). This is because the symptoms of epilepsy are determined based on a vital bio-marker called epileptiform discharges (epileptic spikes), which are often observed in the EEG [2]. Since the discovery of these epileptic spikes is very complex, several automated detection methods have been greatly advanced to support this process [3]–[6].

One of the most effective methods to automatically detect epileptic spikes is machine learning, especially supervised learning. In recent studies, there has been a tendency to create its dataset in the following steps [7]–[9].

1) Detect the temporal locations of epileptiform discharge candidate in the EEG.
2) Annotate supervised labels—typically, whether it is true epileptic or non-epileptic—by an epileptic specialist to the candidate waveforms.
3) Generate small interval EEG segments including the candidate waveforms based on the detected candidates.
4) Training a machine learning model with the EEG segments and the annotated labels.

For example, in the study of [7], candidate waveforms were firstly extracted based on EEG local maxima. Then, an annotator labeled epileptic or non-epileptic labels to the EEG maxima, and the EEG including the peak point—300 ms before the peak and 700 ms after the peak—are cropped as the segments. These pairs of segments and labels were used to fit a random forest model. Similarly, in the study of [8], 1-second EEG segments containing epileptic spikes annotated by an epileptologist were generated, and these were used to fit a support vector machine (SVM). Besides, in the study of [9], a neurologist manually identified and annotated epileptic spikes and non-epileptic spikes that resembled epileptic spikes in the EEG recordings. Based on these labels, 0.2-second segments containing epileptic or non-epileptic spikes were extracted. Then, some machine learning models such as k-nearest neighbors (k-NN), naive Bayes (NB), decision tree (DT), or SVM are fitted using these segments.

However, a method that requires the candidate detection step may lead to the oversight of epileptic spikes or or over-detection of non-epileptic waveforms such as environmental noise. In addition, in such methods with fixed spike locations within the segments, since the fitted model performance is influenced by the window taken around the peak, the segmentation constraints should be faithfully adhered even when predicting the EEG of other unknown patients. That is, the fitting of the machine learning models is highly dependent on the manner in which the EEGs containing epileptic spikes are segmented.

In this paper, we hypothesize that a machine learning model which embeds a temporal self-attention mechanism [10] could extract locations and ranges of particular interest within a segment in order to detect epileptic spikes while simplifying the rules for EEG segment creation. To this end, this paper proposes a recurrent neural network (RNN) model with a self-attention mechanism that learns a large

*This work was supported in part by JST CREST (JPMJCR1784).

[1]K. Fukumori is with Tokyo University of Agriculture and Technology, 2–14–16, Nakacho, Koganei-shi, Tokyo, Japan

[2]N. Yoshida is with the Department of Pediatrics, Juntendo University Nerima Hospital, 3–1–10, Takanodai, Nerima-ku, Tokyo, Japan

[1,3,4,5,6]T. Tanaka is with Tokyo University of Agriculture and Technology, 2–14–16, Nakacho, Koganei-shi, Tokyo, Japan, and the Department of Neurosurgery, Juntendo University, Tokyo, Japan, and RIKEN Center for Brain Science, Saitama, Japan, and RIKEN Center for Advanced Intelligent Project, Tokyo, Japan, and Hangzhou Dianzi University, Hangzhou, China. tanakat@cc.tuat.ac.jp

number of EEG segments including epileptic spikes, where the peak location is not aligned in segments. The results showed that without candidate spike detection, the proposed model achieved a higher classification performance than conventional classifiers used in other studies.

II. Related works

Many studies of epileptic spike detection have trained machine learning models with segments generated with respect to the peak point of the spike [7]. In the study of [7], firstly common epileptiform discharge, namely candidate epileptic spikes, are detected based on a threshold value. Each candidate spike is annotated as either epileptic or not and extracted from the EEG as a 1-second segment containing 300 ms before the considered peak point and 700 ms after. Then, they classified these segments using a random forest model. In the study of [8], similar to [7], candidate epileptic spikes detected using a threshold in the EEG were extracted as 1-second segments. These segments are classified using an SVM or a feed-forward neural network model. In addition, in the study of [9], 56-point segments around the locations of spikes were extracted from the EEG recorded at 256 Hz. Hence, a 0.2-second EEG recording around the location of the peak was segmented for each spike. Then, several machine learning models such as k-NN, NB, DT, or SVM are used to classify these segments. In the study of [11] as well as [9], 0.2-second segments were extracted to classify them using k-NN, DT, or SVM.

In these cases, the locations of the peaks are pre-defined in the segments, and each machine learning model is trained on them. Furthermore, the performance of the training model depends on the window taking around the peak which must be determined appropriately. Therefore, the misalignment of the peak locations has a critical effect on the prediction performance of the model.

III. Method

A. Dataset

Table I summarize the dataset. EEG recordings were collected from 50 patients (23 males 27 females) with either benign epilepsy with centro-temporal spikes (BECTS) [12] or focal epilepsy [12] at the Department of Pediatrics, Juntendo University Nerima Hospital. The age at the time of the examination was 3–12 years old. The data were recorded with the international 10–20 methods using the Nihon Koden EEG-1200 system. The sampling frequency was 500 Hz. This dataset was recorded and analyzed under approval from the Juntendo University Hospital Ethics Committee and the Tokyo University of Agriculture and Technology Ethics Committee.

First, two neurosurgeons, one pediatrician, and two clinical technologists selected a focal channel associated with the origin of the epileptic spike. Typically, one EEG dataset may contain multiple epileptic focal channels, and annotators selected the most intense channel as the focal channel. Then, epileptic spikes in the EEG were manually identified and labeled by the experts. Similarly, non-epileptic waveforms were also labeled by experts in order to generate accurate non-epileptic segments. This is because epileptic spikes may have appeared even in unlabeled locations. Second, the EEG recordings were cropped every 1-second from the head, independent of the temporal locations of the labels. Then, as shown in Fig. 1, each segment that contains either epileptic or non-epileptic labeled waveforms were extracted as a valid segment. Last, z-score normalization was applied with mean value and standard deviation for each segment.

TABLE I: Dataset summary of 50 epileptic EEG. This dataset was labeled by two neurosurgeons, two clinical technologists, and one pediatrician. The total number of labeled samples is 31,486.

#Male patients	23
#Female patients	27
Age (Ave. ± STD.)	7.9 ± 2.0
Recording duration [min] (Ave. ± STD.)	27.5 ± 3.31
#Labeled as epileptic spikes	16,008
#Labeled as non-epileptic discharges	15,478
#Total labeled samples	31,486

B. RNN-based model with self-attention mechanism

In this paper, we propose a model that can be fitted from EEG segments that are not temporally aligned with spikes. Fig. 2 illustrates the architecture of the proposed RNN-based model. As this figure shows, the model mainly has two long short-term memory (LSTM) [13] layers and one self-attention layer. This paper adopts the dot-product attention [14] as this attention layer. That is, in this layer, three hidden features $\boldsymbol{Q}^{(i)} \in \mathbb{R}^{\tau\times d}$, $\boldsymbol{K}^{(i)} \in \mathbb{R}^{\tau\times d}$, and $\boldsymbol{V}^{(i)} \in \mathbb{R}^{\tau\times d}$ are calculated from the input features $\boldsymbol{X}^{(i)} \in \mathbb{R}^{\tau\times d}$ using the three weight matrices $\boldsymbol{W}_Q \in \mathbb{R}^{d\times d}$, $\boldsymbol{W}_K \in \mathbb{R}^{d\times d}$, and $\boldsymbol{W}_V \in \mathbb{R}^{d\times d}$ as:

$$\boldsymbol{Q}^{(i)} = \boldsymbol{W}_Q \boldsymbol{X}^{(i)}, \tag{1}$$

$$\boldsymbol{K}^{(i)} = \boldsymbol{W}_K \boldsymbol{X}^{(i)}, \tag{2}$$

$$\boldsymbol{V}^{(i)} = \boldsymbol{W}_V \boldsymbol{X}^{(i)}, \tag{3}$$

where i, τ, and d are the index of EEG segments, the temporal length of the input feature, and the number of the feature channels, respectively. Then, using the softmax function and one weight matrix $\boldsymbol{W}_O \in \mathbb{R}^{d\times d}$, the output of the self-attention layer $\boldsymbol{Y}^{(i)} \in \mathbb{R}^{\tau\times d}$ is obtained as:

$$\boldsymbol{Y}^{(i)} = \boldsymbol{W}_O \ \mathrm{softmax}\left(\boldsymbol{Q}^{(i)}(\boldsymbol{K}^{(i)})^\top\right)\boldsymbol{V}^{(i)}. \tag{4}$$

Therefore, it is expected that the first block of an LSTM and an average-pooling layers will output feature vectors every 8 ms, and the self-attention layer will search for relationships between these feature vectors. features. For the generation of initial weights of this model, the Xavier initializer [15] is used.

C. Experimental implementation

In order to verify the effectiveness of the proposed model, a numerical experiment is conducted by using surface EEGs recorded from epileptic patients. In addition, EEGNet [16]

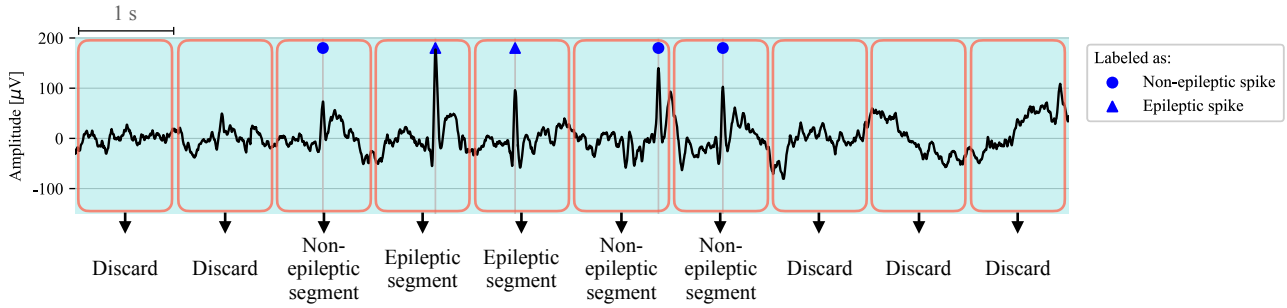


Fig. 1: Example of generating segments from the EEG recordings with supervised labels. Unlabeled segments are discarded because epileptic spikes may have appeared in the unlabeled range.

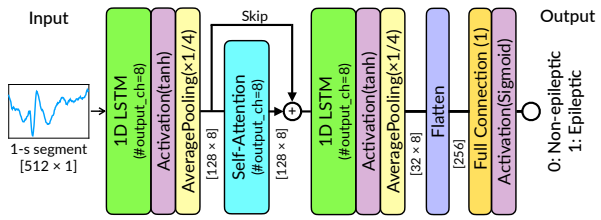
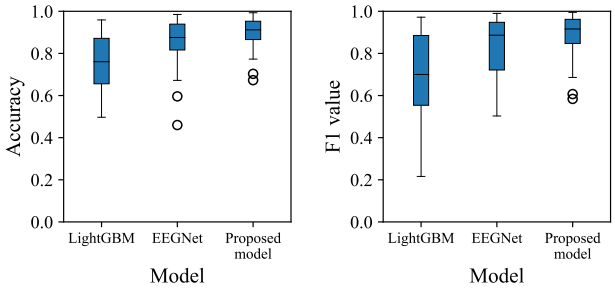


Fig. 2: Proposed 1D-RNN architecture with self-attention mechanism.



(a) Accuracies (b) F1 values

Fig. 3: Graphical results of the 50 inter-subject validations.

and LightGBM [17], which have been attracted for EEG analysis in recent years, are employed as comparison models. A 1-second raw EEG segment is input to each model to classify whether it is an epileptic spike or non-epileptic spike.

In this experiment, an inter-subject validation is conducted with segments from 49 patients as a train-data and remaining segments as a test-data in all combinations. A randomly selected 20 % segments of the train-data are used as a validation-data. EEGNet and the proposed model are fitted by the Adam optimizer [18] while suppressing overfitting using early stopping [19] with the validation-data. In the training of LightGBM, the number of estimators, which is its hyperparameter, is grid-searched from [5, 10, 20, 30, 50, 100, 300] using the validation data. To evaluate the models, the classification accuracy and the F1 value are adopted. When a dataset contains an imbalanced sample ratio—the sample ratio of a few patients is not balanced—, the F1 value is one of the useful evaluation metrics [20]. This can be calculated by the harmonic mean of precision and recall, given as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{6}$$

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{7}$$

where TP, FP, FN, and TN are the numbers of true positive, false positive, false negative, and true negative, respectively.

IV. Experimental Result

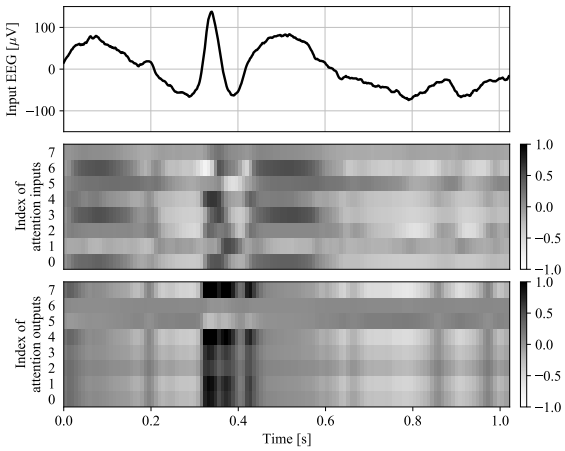
Fig. 3 graphically displays the accuracies and F1 values for each model in 50 inter-subject validations and Table II represents the average scores for each model. As shown in Table II, the average accuracy and the average F1 value of the proposed model performed higher than the other models. Fig. 4 illustrates examples of the input features to the attention layer (namely, the 8-channel output of the recurrent layer) and their output features by the attention layer. Fig. 4a shows that feature extraction by recurrent layer tends to be affected by EEG amplitude. On the other hand, the feature extraction by the attention layer responds more strongly to the temporal location of the epileptic spike. In addition, the duration of the strong response seems to be approximately 100 ms. Similarly, Fig. 4b shows the features for non-epileptic segment input. This figure shows that even though the attentional layer responds to EEG maxima at around 600 ms and 750 ms, the responses are weaker and their duration is shorter than the response to epileptic spikes.

TABLE II: Average accuracy and F1 value of the evaluation results in the inter-subject validations

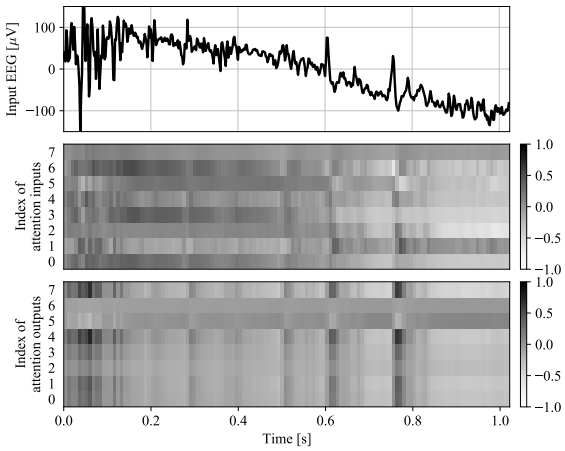
Model	Accuracy (Ave. ± STD.)	F1 value (Ave. ± STD.)
LightGBM	0.760 ± 0.127	0.702 ± 0.206
EEGNet	0.859 ± 0.102	0.821 ± 0.149
Proposed model	**0.902 ± 0.068**	**0.887 ± 0.098**

V. Discussion

This paper proposed a method to detect epileptic EEG segments that are not temporally aligned with spikes in order to eliminate the process of the candidate spike detection. The proposed RNN-based model combined with the self-attention mechanism achieved higher detection performance than the conventional models (average accuracy is 90.2



(a) When predicting an epileptic segment. The peak waveform at around 0.3 seconds is an epileptic spike.



(b) When predicting a non-epileptic segment.

Fig. 4: The top rows of each figure show the input EEG segments to the proposed model. The middle and bottom rows show the input features and output features of the self-attention layer when fed its segments (the size of the features is 128 length $\times$ 8 ch, but the temporal direction is aligned to the same length as the input segment).

%). Therefore, although previous studies have generated segments based on the location of the spike candidates [7]–[9], [11], it is fully detectable even if the spikes are not aligned within the segment. In other words, detection of candidate spikes was regarded as essential for machine learning spike detection, however, the results suggest that this detection may no longer be necessary. It is considered that the self-attention mechanism contributed to the improvement of accuracy by extracting the temporal location and range of interest accurately.

Furthermore, the features extracted by the self-attention layer suggested that the required temporal length of the EEG is around 100 ms for epileptic spike detection using machine learning techniques. In addition, the results of this paper, which aim to assist diagnosis, have the potential to indicate to medical specialists "which time EEG is suspicious".

References

[1] T. J. E. Society, "Clinical Specialists Certified by JES," http://square.umin.ac.jp/jes, 2021, (Accessed in March, 2021).

[2] A. Parmeggiani, G. Barcia, A. Posar, E. Raimondi, M. Santucci, and M. C. Scaduto, "Epilepsy and EEG paroxysmal abnormalities in autism spectrum disorders," *Brain and Development*, vol. 32, pp. 783–789, July 2010.

[3] O. Ozdamar and T. Kalayci, "Detection of spikes with artificial neural networks using raw EEG," *Computers and Biomedical Research*, vol. 31, pp. 122–124, 1998.

[4] J. Algimantas, B. Gytis, B. Dmitrijus, and S. Ruta, "EEG analysis - automtic spike detection," *Nonlinear Analysis: Modelling and Control*, vol. 16, no. 4, pp. 375–386, 2011.

[5] A. R. Johansen, J. Jin, T. Maszczyk, J. Dauwels, S. S. Cash, and M. B. Westover, "Epileptiform spike detection via convolutional neural networks," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 754–758.

[6] L. Xuyen, L. Thanh, D. Viet, T. Long, N. Trung, and N. Thuan, "Deep learning for epileptic spike detection," *VNU Journal of Science: Comp. Science & Com. Eng*, vol. 33, no. 2, pp. 1–13, 2017.

[7] J. E. L. Douget, A. Fouad, M. M. Filali, J. Pyrzowski, and M. L. V. Quyen, "Surface and intracranial EEG spike detection based on discrete wavelet decomposition and random forest classification," *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 475–478, 2017.

[8] K. Fukumori, H. T. Thu Nguyen, N. Yoshida, and T. Tanaka, "Fully data-driven convolutional filters with deep learning models for epileptic spike detection," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 2772–2776.

[9] L. T. Thanh, N. T. A. Dao, N. V. Dung, N. L. Trung, and K. Abed-Meraim, "Multi-channel EEG epileptic spike detection by a new method of tensor decomposition," *Journal of Neural Engineering*, vol. 17, no. 1, pp. 016023, Jan. 2020.

[10] R. M. Farsani and E. Pazouki, "A transformer self-attention model for time series forecasting," *Journal of Electrical and Computer Engineering Innovations (JECEI)*, vol. 9, no. 1, pp. 1–10, 2021.

[11] N. T. Dao, L. Thanh, N. Viet-Dung, N. Linh-Trung, and H. Le, "New feature selection method for multi-channel EEG epileptic spike detection system," *VNU Journal of Science: Computer Science and Communication Engineering*, vol. 35, Dec. 2019.

[12] W. D. Shields and O. C. Snead III, "Benign epilepsy with centrotemporal spikes," *Epilepsia*, vol. 50, no. s8, pp. 10–15, August 2009.

[13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[14] T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Sep. 2015, pp. 1412–1421, Association for Computational Linguistics.

[15] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

[16] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, pp. 056013, Jul. 2018.

[17] P. Detti, G. Z. M. Lara, R. Bruni, M. Pranzo, F. Sarnari, and G. Vatti, "A patient-specific approach for short-term epileptic seizures prediction through the analysis of EEG synchronization," *IEEE Transactions on Biomedical Engineering*, vol. 66, pp. 1494–1504, Oct. 2019.

[18] M. Bernico, *Deep Learning Quick Reference : Useful hacks for training and optimizing deep neural networks with TensorFlow and Keras*, Packt Publishing Limited, 2018.

[19] D. Sarkar, R. Bali, and T. Ghosh, *Hands-On Transfer Learning with Python*, Packt Publishing Limited, 2018.

[20] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.