Convolutional Transformer Networks For Epileptic Seizure Detection

Nan Ke[1], Tong Lin[1] ✉, Zhouchen Lin[1,2,5], Xiao-Hua Zhou[3,4,5], Taoyun Ji [6]
[1] Key Lab. of Machine Perception (MoE), School of AI, Peking University, Beijing, China
[2] Institute for Artificial Intelligence, Peking University, Beijing, China
[3] Beijing International Center for Mathematical Research, Peking University, Beijing, China
[4]Department of Biostatistics, Peking University, Beijing, China
[5] Pazhou Lab, Guangzhou, China
[6] Department of Pediatrics, Peking University First Hospital, Beijing, China

ABSTRACT

Epilepsy is a chronic neurological disease that affects many people in the world. Automatic epileptic seizure detection based on electroencephalogram (EEG) signals is of great significance and has been widely studied. The current deep learning epilepsy detection algorithms are often designed to be relatively simple and seldom consider the characteristics of EEG signals. In this paper, we propose a promising epilepsy detection model based on convolutional transformer networks. We demonstrate that integrating convolution and transformer modules can achieve higher detection performance. Our convolutional transformer model is composed of two branches: one extracts time-domain features from multiple inputs of channel-exchanged EEG signals, and the other handle frequency-domain representations. Experiments on two EEG datasets show that our model offers state-of-the-art performance. Particularly on the CHB-MIT dataset, our model achieves 96.02% in average sensitivity and 97.94% in average specificity, outperforming other existing methods with clear margins.

CCS CONCEPTS

• **Applied computing** → **Health informatics**; • **Computing methodologies** → **Machine learning**.

KEYWORDS

Neural Networks, Convolutional Transformer Networks, Epileptic Seizure Detection

ACM Reference Format:
Nan Ke[1], Tong Lin[1] ✉, Zhouchen Lin[1,2,5], Xiao-Hua Zhou[3,4,5], Taoyun Ji [6]. 2022. Convolutional Transformer Networks For Epileptic Seizure Detection. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (CIKM2022).* ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

CIKM2022, October 17-22, 2022, Georgia, USA

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

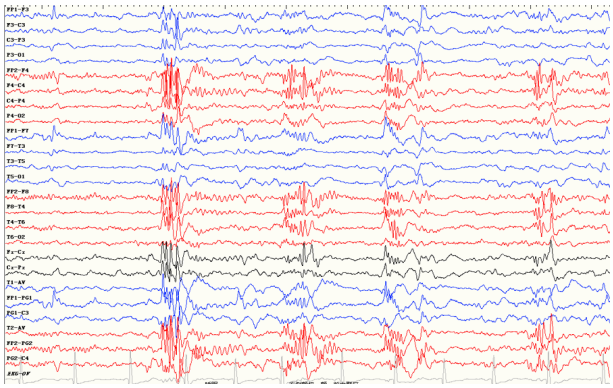

Figure 1: The waveform graph sampled from the PKU1st EEG dataset.

1 INTRODUCTION

Epilepsy, caused by abnormal discharges of brain nerve cells, is one of the most common neurological disorders [15]. According to the World Health Organization Report, there are around 50 million people worldwide suffering from epilepsy. Electroencephalography (EEG, as shown in Figure 1), which can record brain wave signals, has been widely used for epilepsy diagnosis and treatment [3]. Specialized medical knowledge is required to analyze those EEG signals. However, the amount of EEG signals is often too huge for trained neurologists to efficiently analyze [13]. Therefore, the development of automated epilepsy detection algorithms to reduce the burden on physicians has become a valuable research direction [17].

To achieve automatic detection of epilepsy, many machine learning methods have been proposed. The current epilepsy detection methods are mainly divided into two categories, one is based on manual feature engineering, and the other is based on deep learning. The former category of methods uses various signal processing methods to extract discriminant features, and then sends them to the binary classifier to determine the ictal session. Commonly used EEG signal processing methods include fourier transform and wavelet transform [7] [8], as well as other signal processing methods such as the local mean decomposition [19] (LMD) and the empirical mode decomposition (EMD) [14]. These processed features will be sent to classifiers such as SVM and random forest to obtain classification results [10].

However, these feature engineering methods are often complicated and cumbersome, and the effect is not good enough. Epilepsy detection methods based on deep learning mainly use convolutional neural networks (CNNs) [22] [20] and LSTM [1]. Hu [9] used a Bi-LSTM with some time-frequence features. Acharya [2] proposed a deep CNN consisting of 13 layers for automatic seizure detection. The current deep learning algorithms are often designed to be relatively simple and seldom take into account the characteristics of EEG signals. There is still a lot of room for exploration in this direction.

Transformer models are not only widely used in the field of natural language processing [21] [5] , but also make significant progress in the field of computer vision [6] [23] . The EEG segment data is a time series in essence and is also image-like, which is suitable for processing with transformer models. However, transformer models have not been successfully used for epilepsy detection yet in the literature. In this paper, we propose a novel epileptic seizure detection model based on the transformer networks and achieve state-of-the-art results on two datasets. The contributions of this paper are listed as follow:

- We propose to use the transformer model on epilepsy detection tasks for the first time, and show that a hybrid approach combining convolution with transformer blocks works best.
- We propose an end-to-end seizure detection model, MFCvT, based on convolutional transformer. This model, designed for levaraging the multi-channel characteristics of EEG, uses mutil-view gated inputs and Fourier transform to improve the seizure detection performance.

2 METHODOLOGY

In this section, we propose the convolutional transformer model CNViT and its extended model MFCvT.

2.1 The CNVIT Model

Directly applying the vision transformer models such as ViT [6] and CvT [23] to the epilepsy detection task does not perform well in the experiment. This is because the transformer models have a larger hypothesis space and requires more data to train. However, the amount of EEG data is often insufficient when comparing with the huge number of parameters in transformers, so that the transformer models cannot be trained well. On the other hand, the convolutional neural network models such as CW-SRNet [11] achieve good results in epilepsy detection tasks, while the alternating use of convolution and transformer block in the CvT [23] model performs poorly. We found that it would be better to directly use multi-stage convolutional layers to extract the representation information of epileptic seizures. Based on this insight, we design a network CNViT (Convolutional Vision Transformer) that first uses multi-layer convolution to extract features, and then adopts transformer blocks. The model architecture of CNViT is shown in Figure 2.

The CNViT model is divided into two stages. In the first stage, we use a eight-layer convolutional neural network to extract the features of epileptic EEG segments, and then flatten the obtained feature map to output a sequence $S = (s_1, s_2, ..., s_n)$, where s_i is the value of a pixel in the feature map. The sequence S will be used as

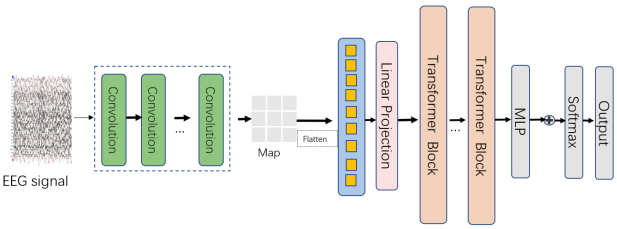

Figure 2: The Architecture of The CNViT Model

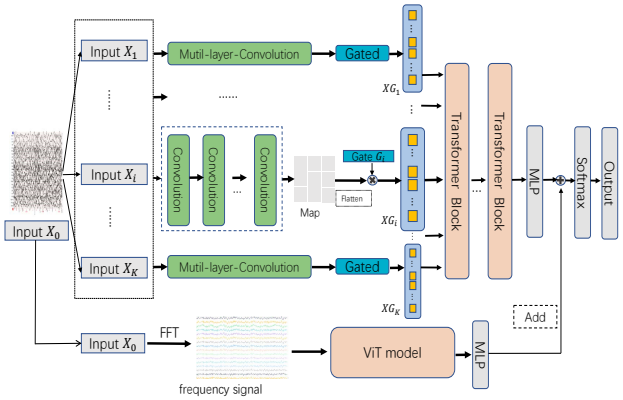

Figure 3: The Architecture of The MFCVT Model

the input of the second stage of the transformer module. For each s_i, we consider it to be a value of a certain type of feature, and map it to a new semantic space through an embedding layer. At the same time the order of the sequence is encoded by position, and the encoding of each position is represented by a one-dimensional embedding. Then the position encoding is directly added to the feature map embedding of the previous step. These embeddings will be sent to the transformer blocks in the next stage. Empiracally, we found that the detection effect of deeper and narrower structures will be better. We finally selected six consecutive transformer blocks. The output of the last transformer block is sent to an MLP classification network to output the classification prediction results.

The experimental results of this model are shown in the table 1 . It can be seen that the transformer model with this structure can do better than the original optimal convolutional network on the epilepsy detection task.

2.2 The MFCVT Model

In this subsection, we further propose two improvements based on the CNViT model. For an EEG segment, when we arbitrarily change the order between EEG channels, theoretically it should not change the result of the model to determine whether the epilepsy is onset. But when we use the convolution network, the convolution operation is related to the image position information. The electrical channel order may affect the results of the convolutional network. At the same time, when an epileptic seizure occurs, there will be

partial connections between neighbouring brain regions. Certain adjacent brain regions may have similar phenomena at the same time during an epileptic seizure. In the presence of specific signals, the convolution operation captures the information of adjacent regions, which means that the order of EEG channels has an impact on the convolutional network. We propose to take this impact into account by adopting multiple channel order inputs.

The convolutional network actually extracts more features in the time domain, but during epileptic seizures, the frequency of brain waves will change significantly. We hope that the model can explicitly consider the characteristics of this frequency change, so we introduce the discrete Fourier transform method into the model to directly extract the frequency domain features.

Based on the above two points, we propose a model MFCvT (Convolutional transformer network with multiple gated inputs and Fourier transform) with stronger seizure detection capabilities. The architecture of MFCvT is shown in the figure 3.

For a raw EEG fragment sample $X_0 \in \mathbb{R}^{C\times T}$, We get multiple inputs $(X_1, X_2, ..., X_K)$ by randomly shuffling the order of EEG channels, where K is the number of shuffles. Each X_i is modified EEG fragment sample obtained by randomly swapping the channel dimensions of X_0. For each input sample X_i, we use an eight-layer convolutional neural network to obtain its feature map. After the feature map is flattened, a one-dimensional vector XF_i $(i = 1, ..., K)$ can be obtained. After feeding the K ways into the convolutional network we can get a set $(XF_1, XF_2, ..., XF_K)$. Because different EEG channel sequences may have good or bad effects on the results and we cannot know in advance which sequence is better, we add a gating module here to dynamically learn the importance of each channel. The gating module consists of a set of multiple gating vectors $(g_1, g_2, ..., g_K)$. Each gating vector g_i is a learnable vector of the same dimension as XF_i. After passing g_i through the sigmoid function we get a vector $G_i$whose value is between 0 and 1. Then we let multiply XF_i and G_i to get the vector XG_i.

The vector set $(XG_1, XG_2, ..., XG_K)$ is obtained after the same processing as above for the K-way input, which will be the input to the following transformer blocks. Specifically, we first use an embedding layer to do a semantic space transformation, then feed the obtained multiple embeddings directly into the transformer blocks. Since there is no positional sequence relationship between multiple paths, we do not need to add position encoding here. Like the CNViT model mentioned above, here we use six consecutive transformer blocks. Before they pass the softmax of the classification layer, we can get the vector $logits_G$.

The above is the main part of the MFCvT model. In order to combine the frequency domain features to improve the model performance, we add the auxiliary branche to the network. For the above-mentioned original EEG segment input X_0, we perform one-dimensional discrete Fourier transform on each channel to obtain the frequency domain signal X_F, where X_F and X_0 have the same dimensions. Next, we feed the frequency domain signal X_F into the ViT [6] network, and the patch for ViT [6] model is set as $1 \times C$. Before passing through the softmax of the classification layer, we can get the vector $logits_F$. Since the vector $logits_G$ and the vector $logits_F$ have the same dimensions, we use the direct weighted fusion method to get the final logits

$$logits = logits_G * (1 - \alpha) + logits_F * \alpha$$

where α is the weighting factor. This logits can output the probability value of classification prediction through the sigmoid layer. In the next experiments, we can see that the MFCvT model achieves the best results in experiments on multiple datasets.

3 EXPERIMENTS

In this section, we compare our models with other state-of-the-art methods for seizure detection.

3.1 Experimental Setup

Dataset. We use the popular CHB-MIT benchmark dateset and the dataset collected from The Peking University First Hospital (referred as PKU1st dataset). The CHB-MIT EEG dataset, gathered at the Boston Children's Hospital [18], is one of the largest and most used public datasets for epilepsy. This dataset consists of long-duration multi-channel EEG recordings from 23 pediatric patients with intractable seizures. The details of the mit dataset are described in [9] [11]. We split the continuous EEG into many two second segments. By means of the dataset seizure annotation, we can easily distinguish between interictal and ictal phase. The PKU1st dataset was collected from the department of Pediatrics of Peking University First Hospital. We use data from 18 patients with a similar composition to the MIT dataset. Mixed and single-patient experiments were performed on each dataset. A single-patient experiment operates on a single person's data that is partitioned into training set and test set, while in a mixed experiment the data from all patients are randomly divided into training and test sets. In all experiments we compare averaged metrics.

3.2 Measurements

In our experiments, four statistical indicators are used for the performance evaluation of the proposed method. Some indicators are defined as follows:

$$\text{Sensitivity (Sen)} = \frac{TP}{TP + FN}, \tag{1}$$

$$\text{Specificity (Spe)} = \frac{TN}{TN + FP}, \tag{2}$$

$$\text{Accuracy (Acc)} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

where TP is true positive, FP is false positive, TN is true negative and FN is false negative. We also use AUC(Area under the ROC Curve) for performance evaluation. In clinical practice, the most concerned indicator is Sensitivity.

3.3 Results

We compare our transformer models with other epilepsy detection models. CW-SRNet [11] is currently the best performing CNN-based epilepsy detection model. Table 1 shows the comparison results of various transformer models. It can be seen that the direct use of the classic visual transformer model (ViT model [6] and CvT model [23]) is not as useful as the CNN model, and the sensitivity of the CvT model 84.72% is even far lower than the 94.17 of the CNN model. The CNViT model, which we designe first to use convolution to extract

features and then use transformer, slightly exceeds CNN with a sensitivity of 94.47%, indicating that this transformer structure is more suitable for epilepsy detection tasks. Our final MFCvT model achieves 96.71% sensitivity, 97.23% specificity, 97.15% accuracy and 99.54% AUC, significantly outperforming other models.

The results of the single-person experiment on the CHB-MIT dataset are shown in Table 2. Our MFCvT model achieves an average sensitivity of 96.02%, specificity of 97.94%, accuracy of 97.56% and AUC of 99.37%, which is the best performing model on the CHB-MIT dataset.

The results of the 18-person mixed experiment of the PKU1st dataset are shown in Table 3. The CNN model CW-SRNet [11] achieves 87.64% sensitivity, and our MFCvT model achieves 90.12% sensitivity, 96.14% specificity, 94.74% accuracy and 97.99% AUC, the best performance on this dataset 's model. The results of the single-person experiment on the PKU1st dataset are shown in Table 4. Our MFCvT model achieves an average sensitivity of 86.45%,better than CNNs and other transformer models.These experiments fully demonstrate the effectiveness of our model.

Table 1: Results on the Mixed CHB-MIT Dateset

Method	Sen (%)	Spe (%)	Acc (%)	AUC (%)
CW-SRNet [11]	94.17	96.76	96.08	98.78
ViT [6]	92.18	94.73	94.31	97.96
CvT [23]	84.72	95.28	93.53	96.54
CNViT	94.47	97.03	96.61	99.02
MFCvT	**96.71**	**97.23**	**97.15**	**99.54**

Table 2: Single Patient Results on the CHB-MIT Dataset

Method	Sen (%)	Spe (%)	Acc (%)	AUC (%)
Dyadic WT [16]	92.60	**99.90**	-	-
Discrete WT [4]	91.71	92.89	92.30	-
CNN+MIDS [22]	74.08	92.46	-	-
Bi-LSTM [9]	93.61	91.85	-	-
CE-STNet [12]	92.41	96.05	95.96	-
CW-SRNet [11]	94.48	96.94	96.45	99.03
MFCvT (Ours)	**96.02**	97.94	**97.56**	**99.37**

4 CONCLUSION

In this paper, we propose a seizure detection model based on convolutional transformer. Our MFCvT model, designed for the multi-channel characteristics of EEG, uses mutil-view gated inputs and fourier transform to improve the seizure detection performance.

Table 3: Results on the Mixed PKU1st Dateset

Method	Sen (%)	Spe (%)	Acc (%)	AUC (%)
CW-SRNet [11]	87.64	95.35	93.56	97.06
ViT [6]	84.54	92.44	90.61	94.54
CvT [23]	64.16	89.17	83.4	85.50
CNViT	87.41	92.82	96.48	96.48
MFCvT	**90.12**	**96.14**	94.74	**97.99**

Table 4: Single Patient Results on the PKU1st Dataset

Method	Sen (%)	Spe (%)	Acc (%)	AUC (%)
CW-SRNet [11]	86.45	**97.65**	**96.21**	97.53
ViT [6]	82.85	96.68	94.97	94.88
CvT [23]	74.19	96.89	93.25	93.94
CNViT	87.02	96.44	95.21	96.12
MFCvT	**88.43**	97.16	95.79	**97.63**

The proposed method achieves an an average sensitivity of 96.02% on the single patient setting of CHB-MIT dataset, which offers the state-of-the-art results. These experiments convincingly prove the effectiveness of our proposed epilepsy detection model.

ACKNOWLEDGMENTS

This work was supported by NSFC Tianyuan Fund for Mathematics (No. 12026606), National Key RD Program of China (No. 2018AAA0100300), Beijing Academy of Artificial Intelligence(BAAI), NSF China (No. 61731018), and Project 2020BD006 supported by PKU-Baidu Fund.

REFERENCES

[1] Ahmed M Abdelhameed, Hisham G Daoud, and Magdy Bayoumi. 2018. Epileptic seizure detection using deep convolutional autoencoder. In *2018 IEEE International Workshop on Signal Processing Systems (SiPS)*. IEEE, 223–228.

[2] U Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Hojjat Adeli. 2018. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in biology and medicine* 100 (2018), 270–278.

[3] MZ Ahmad, Maryam Saeed, Sajid Saleem, and Awais M Kamboh. 2016. Seizure detection using EEG: A survey of different techniques. In *2016 International Conference on Emerging Technologies (ICET)*. IEEE, 1–6.

[4] Duo Chen, Suiren Wan, Jing Xiang, and Forrest Sheng Bao. 2017. A high-performance seizure detection algorithm based on Discrete Wavelet Transform (DWT) and EEG. *PloS one,* 12, 3 (2017), e0173138.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[6] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020).

[7] Kais Gadhoumi, Jean-Marc Lina, and Jean Gotman. 2012. Discriminating preictal and interictal states in patients with temporal lobe epilepsy using wavelet analysis of intracerebral EEG. *Clinical neurophysiology* 123, 10 (2012), 1906–1916.

[8] Samanwoy Ghosh-Dastidar, Hojjat Adeli, and Nahid Dadmehr. 2007. Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection. *IEEE transactions on biomedical engineering* 54, 9 (2007), 1545–1551.

[9] Xinmei Hu, Shasha Yuan, Fangzhou Xu, Yan Leng, Kejiang Yuan, and Qi Yuan. 2020. Scalp EEG classification using deep Bi-LSTM network for seizure detection. *Computers in Biology and Medicine* 124 (2020), 103919.

[10] Jing Jin, Xingyu Wang, and Bei Wang. 2007. Classification of Direction perception EEG Based on PCA-SVM. In *Third International Conference on Natural Computation (ICNC 2007)*, Vol. 2. IEEE, 116–120.

[11] Nan Ke, Tong Lin, and Zhouchen Lin. 2021. Channel-Weighted Squeeze-and-Excitation Networks For Epileptic Seizure Detection. In *2021 IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*. 666–673.

[12] Yang Li, Yu Liu, Wei-Gang Cui, Yu-Zhu Guo, Hui Huang, and Zhong-Yi Hu. 2020. Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 28, 4 (2020), 782–794.

[13] A Liu, JS Hahn, GP Heldt, and RW Coen. 1992. Detection of neonatal seizures through computerized EEG analysis. *Electroencephalography and clinical neurophysiology* 82, 1 (1992), 30–37.

[14] Roshan Joy Martis, U Rajendra Acharya, Jen Hong Tan, Andrea Petznick, Ratna Yanti, Chua Kuang Chua, EY Kwee Ng, and Louis Tong. 2012. Application of empirical mode decomposition (EMD) for automated detection of epilepsy using EEG signals. *International journal of neural systems* 22, 06 (2012), 1250027.

[15] Florian Mormann, Ralph G Andrzejak, Christian E Elger, and Klaus Lehnertz. 2007. Seizure prediction: the long and winding road. *Brain* 130, 2 (2007), 314–333.

[16] Lorena Orosco, Agustina Garcés Correa, Pablo Diez, and Eric Laciar. 2016. Patient non-specific algorithm for seizures detection in scalp EEG. *Computers in biology and medicine* 71 (2016), 128–134.

[17] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human brain mapping* 38, 11 (2017), 5391–5420.

[18] Ali Hossam Shoeb. 2009. *Application of machine learning to epileptic seizure onset detection and treatment.* Ph. D. Dissertation. Massachusetts Institute of Technology.

[19] Jonathan S Smith. 2005. The local mean decomposition and its application to EEG perception data. *Journal of the Royal Society Interface* 2, 5 (2005), 443–454.

[20] M Asjid Tanveer, Muhammad Jawad Khan, Hasan Sajid, and Noman Naseer. 2021. Convolutional neural networks ensemble model for neonatal seizure detection. *Journal of Neuroscience Methods* 358 (2021), 109197.

[21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[22] Zuochen Wei, Junzhong Zou, Jian Zhang, and Jianqiang Xu. 2019. Automatic epileptic EEG detection using convolutional neural network with improvements in time-domain. *Biomedical Signal Processing and Control* 53 (2019), 101551.

[23] Haiping Wu, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, and Lei Zhang. 2021. Cvt: Introducing convolutions to vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 22–31.