Proceedings of CONF-MLA 2026 Symposium: Intelligent Systems and Automation: AI Models, IoT, and Robotic Algorithms
DOI: 10.54254/2755-2721/2026.LD37868

DARF-P1DNet: A Dual-View Attention-Guided Residual Fusion Pyramidal 1D Network for EEG-Based Epileptic Seizure Detection

Yuexi Xiang

International Education College, Beijing University of Chemical Technology, Beijing, China
xiang71869886@qq.com

Abstract. Seizures cause noticeable abnormal changes in EEG signals. Relying on manual reading of long-term EEG recordings not only costs plenty of time, but it is also easily affected by the doctor's experience and fatigue. To improve the automation and stability of EEG seizure detection, this paper proposes a Dual-view Attention Residual Fusion Pyramidal 1D Network (DARF-P1DNet). First, this method normalizes the single-channel signals from the Bonn EEG dataset. It then constructs the original view and an augmented view for the training phase. Next, two P-1D-CNN encoders are used. They have the same structure, but their parameters are independent. These encoders extract complementary features from the two views. Then, Feature-wise Attention and Residual Fusion are used to filter and keep important information from the dual views. Finally, the model uses a lightweight classification head to perform binary classification between Non-seizure and Seizure. It is trained using weighted cross-entropy loss. In the experiments, we evaluate the model using common classification metrics, ablation studies, and comparison experiments. The overall method has a clear structure. It can fully use the complementary information from the original and augmented signals while keeping the model relatively lightweight. This provides a workable solution for seizure detection in EEG data with small sample sizes.

Keywords: Electroencephalogram (EEG), Epilepsy, Seizure detection, Dual-view learning, Attention mechanism, Deep learning

1. Introduction

Epilepsy is a common type of neurological disorder. During a seizure, the brain neurons show abnormal discharge, which creates noticeable changes in rhythm and waveform in the EEG signal. EEG is a non-invasive technique with high temporal resolution, making it an important tool for epilepsy screening and auxiliary diagnosis. For long-term EEG recordings, if we rely completely on manual reading, the workload is heavy. Moreover, the results are easily affected by the reader's experience and fatigue. Therefore, using algorithms for automatic detection has practical value.

Early EEG seizure detection mostly used traditional machine learning methods. These methods usually first extract time-domain, frequency-domain, or nonlinear features, and then use classifiers

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.LD37868

such as Support Vector Machines and K-Nearest Neighbors to complete the recognition. The computational process of these methods is clear, but the feature design relies heavily on manual experience. With the development of deep learning, one-dimensional convolutional networks can directly learn features from raw time-series signals, which reduces the need for complex manual feature engineering. P-1D-CNN forms a pyramidal structure by gradually reducing the number of channels across layers. This makes the model relatively lightweight and suitable for use as a base network for EEG classification with small sample sizes.

However, a single input view is easily affected by changes in amplitude, changes in time position, and noise. When the sample size is limited, the model may also rely too much on certain local patterns. Common data augmentation can increase the variety of training samples. However, if we only train the model on the augmented data in a normal way, the complementary relationship between the original information and the augmented information is not directly used. Therefore, this study uses the original view and the training-time augmented view for simultaneous learning. It also performs fusion at the feature level. This allows the model to keep the basic patterns of the raw EEG while also learning features that are more stable to small changes.

To address the above problems, this paper proposes DARF-P1DNet. Its main idea includes three aspects. First, we normalize the central EEG segment of length 512, and create the augmented view during training using gain, cyclic shift, and Gaussian noise. Second, we use two P-1D-CNN encoders with the same structure but independent parameters to extract features from the dual views. We then use Feature-wise Attention and Residual Fusion to complete the feature fusion. Third, we use a simple fully connected classification head to perform binary classification, and we use weighted cross-entropy loss to reduce the effect of class imbalance on training. In the later experiments, we will verify the role of each module through common classification metrics, ablation studies, and method comparisons.

2. Related work

Automatic EEG seizure detection methods can be broadly divided into two categories: traditional machine learning and deep learning. Traditional methods usually first filter or transform the signal, then extract hand-crafted features such as statistical measures, spectral features, wavelet features, or entropy features, and finally use a classifier to determine the EEG state [1]. These methods typically have smaller model sizes and faster training speeds. They also make it easier to explain the roles of different features. However, the detection results are affected by the way features are selected. When the signal contains a lot of noise or when there are obvious differences between samples, fixed hand-crafted features may not maintain stable performance.

Traditional machine learning methods rely heavily on manually designed EEG features. The quality of feature design directly determines the final recognition performance of the model. WANG Yuanfa and colleagues completed EEG signal decomposition using wavelet transform, extracted multi-dimensional hand-crafted features such as approximate entropy and Teager energy, and used multi-class Support Vector Machines to recognize epileptic EEG signals. They achieved a classification accuracy of 97.21% on the Bonn dataset [2]. Kang and colleagues used dual-density dual-tree complex wavelet transform to extract wavelet entropy and signal energy features. They used Least Squares Support Vector Machines to distinguish between seizure and non-seizure EEG signals. For the binary classification task, their detection accuracy reached 98.06% [3]. YANG Shuhan and colleagues extracted time-domain statistical features and nonlinear features, combined them with the XGBoost classifier, and carried out epilepsy detection experiments. On public EEG datasets, they achieved a recognition accuracy of 96.43% [4].

DOI: 10.54254/2755-2721/2026.LD37868

Deep learning methods can learn features directly from EEG sequences. Among them, one-dimensional convolutional networks are well suited for time-series signals [5]. The convolutional layers can identify local waveform patterns, and the multi-layer structure can gradually form higher-level feature representations. Compared with larger networks, P-1D-CNN reduces model complexity through pyramidal channel settings, making it easier to train when the sample size is small. However, a single encoder still mainly learns from one input form. When the amplitude, position, or noise of the EEG signal changes, the extracted features may fluctuate. Previous studies have confirmed this limitation. Singh and colleagues found in their seizure detection experiments based on one-dimensional convolution that shifts in signal amplitude could cause shifts in the waveform features extracted by convolution, directly reducing the stability of cross-sample recognition [6]. Kumar and colleagues pointed out in their comparative experiments on scalp EEG that a single convolutional encoder is sensitive to electromyography and power-line interference. Noise disturbance can cause feature space drift, leading to a clear decline in model generalization performance [7]. The ablation comparison experiments by Omar and colleagues also showed that for networks relying only on a convolutional encoder, the discriminative ability of features decays when the time-domain position of the waveform shifts. This makes it difficult to adapt to signal distortions caused by individual differences in EEG signals [8].

To improve the model's ability to adapt to input disturbances, data augmentation and multi-view learning have been widely used in time-series signal classification tasks [9]. The attention mechanism can assign different weights to features, and residual connections can preserve the original basic features [10]. Based on the above ideas, this paper builds a dual-view encoding, feature attention, and residual fusion pipeline on top of P-1D-CNN. We do not introduce complex recurrent networks or ensemble modules. This simplifies the network structure and also makes it easy to carry out ablation experiments to verify the effectiveness of each module.

3. DARF-P1DNet model architecture

The DARF-P1DNet (Dual-view Attention Residual Fusion Pyramidal 1D Network) proposed in this paper is built on the lightweight P-1D-CNN. It is mainly designed to address the problems of large input variations and limited information from a single view in small-sample EEG seizure detection. The model does not introduce complex recurrent networks or multi-model voting structures. Instead, it gradually enhances the representation capability through three stages: input view construction, dual-path feature extraction, and feature fusion. The overall process is shown in Figure 1. First, the input EEG segment is normalized. Then, the original view and the training-time augmented view are created. These two views are sent separately into two P-1D-CNN encoders. The encoders have the same structure but independent parameters. After encoding, the features from the two paths are concatenated at the feature level. Then, Feature-wise Attention is applied to redistribute the feature weights, and Residual Fusion is used to preserve the unweighted basic information. Finally, the fused features are fed into the classification head to perform binary classification between seizure and non-seizure states. The key point of this design is not simply increasing the network depth. Instead, it explicitly uses the complementary information between the original signal and the augmented signal. This makes the role of each module clearer, and it also makes it easier to verify each module through ablation experiments later.

DOI: 10.54254/2755-2721/2026.LD37868

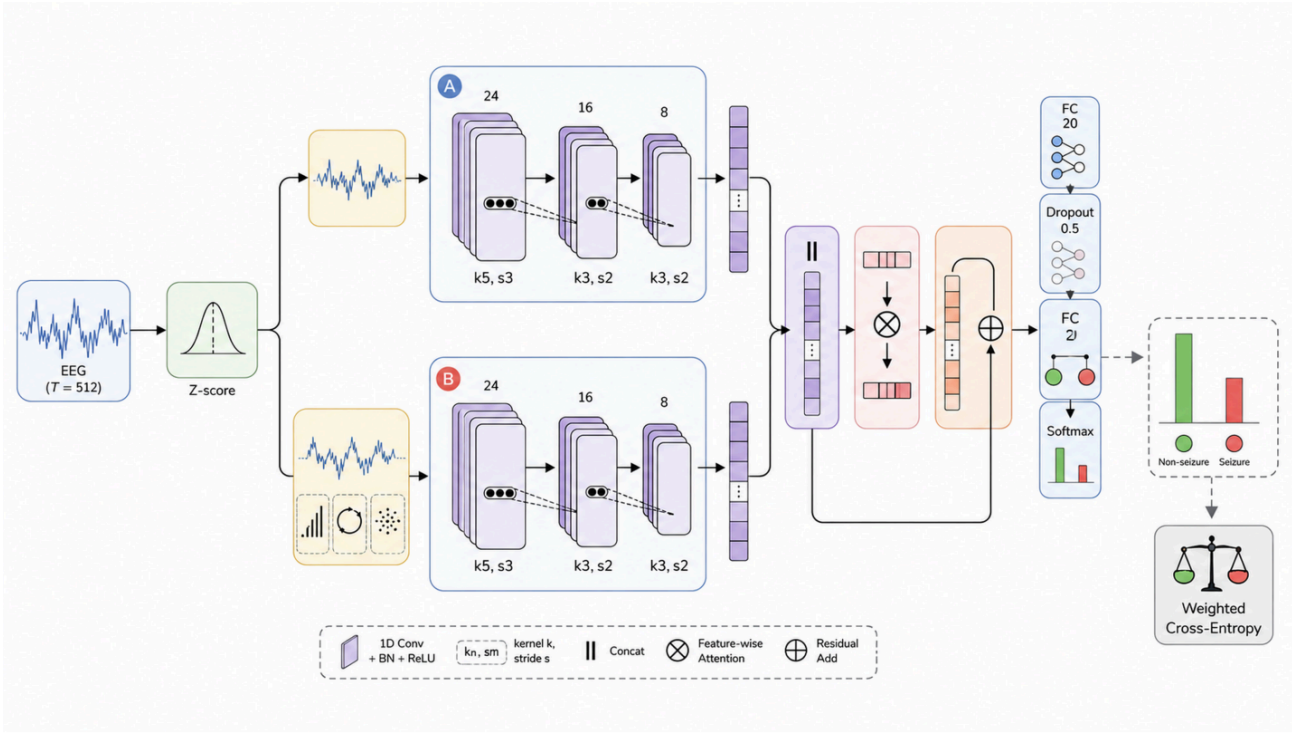

Figure 1. Architecture of the proposed DARF-P1DNet

3.1. Input preprocessing and dual-view construction

The model input is a fixed-length single-channel segment from the Bonn EEG dataset. Different EEG segments may differ in amplitude range and baseline position. If these segments are fed directly into the network, they may increase the fluctuations during model training. Therefore, we first apply Z-score normalization to each input segment. This makes different samples have more similar numerical scales. Let x_t denote the value of sample t, and let the mean and standard deviation of the segment be μ and σ, respectively. The normalized value is denoted as z_t. The calculation process is shown in Equation (1).

$$z_t = \frac{x_t - \mu}{\sigma + \varepsilon} \tag{1}$$

In Equation (1), ε is a small stability term to prevent numerical difficulties when the standard deviation is very small. Z-score normalization changes the entire distribution of the signal. It does not alter the original time points of the waveform. Thus, it can maintain the main features of the EEG waveform while decreasing the variation in scales among different samples. The normalized signal is then used as the common input for building the dual views.

To enhance the acquisition of stable features for the model with little change in the input, the original normalized form is maintained during the training period and a new form is produced. The augmentation techniques include three basic transformations: random gain, cyclic shift and the addition of Gaussian noise. They are employed to simulate amplitude variations, small shifts along the time axis of the waveforms and slight noises which frequently appear during signal acquisition respectively. These transformations only modify the appearance of the signal and do not influence the label of the sample. The augmentation procedure can be concisely represented by Equation (2).

DOI: 10.54254/2755-2721/2026.LD37868

$$x_{aug} = a \cdot Roll(x, \Delta) + n, \quad n \sim N(0, \sigma_n^2) \tag{2}$$

In Equation (2), a denotes the gain coefficient, $Roll(\cdot)$ stands for the signal shift by Δ sampling points, and n represents the added zero-mean Gaussian noise. The Original View and Training-time Augmented View are both derived from the same normalized EEG segment. Therefore, the two encoders can extract features from the original information and the perturbed information separately. Comparing with directly mixing augmented samples into the training set, the dual-view structure can maintain the relationship between the two kinds of information within the model. This gives a clearer input for the following feature fusion.

3.2. Dual P-1D-CNN encoding and feature fusion

The original view and the augmented view are inputted into two P-1D-CNN encoders separately. These two encoders have the same network architecture, but their parameters are updated individually. This enables the two paths to respond differently to their corresponding inputs. Each encoder includes three series-connected one-dimensional convolutional blocks. Each block comprises a Conv1D, Batch Normalization and a ReLU. The number of channels decreases from 24 to 16 to 8 in these three blocks. The kernel sizes are 5, 3 and 3, and the strides are 3, 2 and 2, respectively. No additional pooling layers are added. Instead, the temporal length is compressed gradually by using convolutional strides which helps to reduce the feature dimensions without losing the continuous temporal information. The general computation of the l-th convolutional block can be expressed as Equation (3).

$$F_l = ReLU(BN(Conv1D(F_{l-1}; W_l, b_l))) \tag{3}$$

In Equation (3), F denotes the output characteristic of the present convolutional layer. The characteristics of the previous layer are used as the input for this layer. W and b stand for the convolutional parameters and the bias respectively. $BN(\cdot)$ and $ReLU(\cdot)$ indicate the batch normalization and activation procedures respectively. During the computation of the network from the initial layers to the last layers, the main convolutions mainly extract the local variations in the waveform, while the succeeding convolutions construct more concentrated features based on the simplified temporal representation. The final output characteristics of the two encoders are represented by F_0 and F_a. In order to avoid losing the information of any branch too soon before merging, the outputs of the two branches are merged in the feature direction to obtain the comprehensive feature $F_c = [F_o; F_a]$. This method firstly preserves the information of both aspects and then permits the following modules to assess the importance of different features.

After obtaining the joint feature F_c, the model uses Feature-wise Attention to perform feature recalibration. This module first aggregates the global information from the joint feature, and then generates weights for each feature position through a learnable mapping. This allows the responses that are more relevant to the classification task to be strengthened, while the weaker or redundant responses are suppressed. The attention weights and the weighted features can be expressed together as shown in Equation (4).

$$\alpha = Sigmoid(W \cdot GAP(F_c) + b), \quad F_{att} = \alpha \odot F_c \tag{4}$$

In Equation (4), $GAP(\cdot)$ represents global average pooling. W and b are learnable parameters, α is the obtained feature weight, $\odot$ denotes element-wise weighting, and F_{att} is the feature after

DOI: 10.54254/2755-2721/2026.LD37868

attention processing. The attention mechanism can highlight more useful features. However, if we rely completely on the reweighted results, we may also weaken the information in the original fused feature that is still valuable. To address this, we add Residual Fusion after the attention. We add the original concatenated feature F_c to F_{att} in an element-wise manner, and obtain the final fused feature $F_f = F_{att} + F_c$. This operation is relatively straightforward, so we do not list a separate formula for it.

Through the above processing, the model retains both the basic information in Fc and the information strengthened by attention in F_{att}. This helps to avoid the possible loss of information caused by overly strong feature selection. The whole fusion process consists of three consecutive steps: first concatenation, then weighting, and finally residual preservation. The structure is relatively clear, and the role of each step can be verified separately in the later ablation experiments.

3.3. Classification output and training objective

After obtaining the final fused feature, the model uses a lightweight fully connected classification head to perform binary classification. The classification head consists of FC-20, Dropout (0.5), and FC-2 with $\mathrm{Softmax}$. The fully connected layer is used to map the fused feature into a more compact classification representation. During training, Dropout randomly suppresses some neuron responses. This helps to reduce the risk of overfitting under small-sample conditions. Finally, $\mathrm{Softmax}$ outputs the predicted probabilities for the two classes: Non-seizure and Seizure. Let the classification feature sent to the output layer be h. The predicted probability is shown in Equation (5).

$$\mathrm{p} = \mathrm{Softmax}(\mathrm{W_c h} + \mathrm{b_c}) \tag{5}$$

In Equation (5), $\mathrm{W_c}$ and $\mathrm{b_c}$ represent the weights and bias of the classification layer, respectively. p_0 and p_1 represent the predicted probabilities of the two classes. The model uses the class with the larger probability as the final prediction result. Considering that different classes may differ in sample size or classification difficulty, we use Weighted Cross-Entropy as the optimization objective during training. This adjusts the effect of errors from different samples on the total loss through class weights. Its form is shown in Equation (6).

$$\mathrm{L} = -\Sigma\, \mathrm{w_{y_i}} \log \mathrm{p_{i,y_i}} \tag{6}$$

In Equation (6). y represents the true class of the sample, w represents the weight corresponding to that class, and p represents the predicted probability given by the model for the true class. Weighted cross-entropy only affects the training objective and does not change the network structure described above. Up to this point, DARF-P1DNet completes the full processing pipeline: from input normalization, dual-view construction, dual-path P-1D-CNN encoding, feature attention and residual fusion, to the final binary classification output. The overall design keeps the relatively lightweight nature of P-1D-CNN, while also allowing the dual-view information to be explicitly used at the feature level.

DOI: 10.54254/2755-2721/2026.LD37868

4. Experiments

4.1. Dataset description

This study uses the public Bonn EEG dataset for seizure detection experiments. This dataset contains five groups of EEG recordings: A, B, C, D, and E. Group E contains seizure signals, and the other four groups contain non-seizure signals. We constructed a binary classification task from this dataset. After preprocessing steps such as normalization, the data is used for model training and testing.

In addition, this study uses the UCI Epileptic Seizure Recognition dataset for generalization testing. This dataset consists of segmented EEG signals and their corresponding class labels. We also processed it as a binary classification task between seizure and non-seizure. This is used to evaluate the model's classification ability under different data conditions.

4.2. Evaluation metrics

This study uses accuracy, sensitivity, and specificity as the main evaluation metrics. These metrics are used to quantitatively assess the binary classification performance on the Bonn dataset and the UCI dataset. Accuracy reflects the model's overall classification ability for all samples. Sensitivity measures the model's ability to detect seizure samples. It can reflect how many seizure samples are missed. Specificity evaluates the model's ability to identify non-seizure samples. It can indicate the level of false alarms that are suppressed. These three metrics evaluate the model's performance from three aspects: overall classification ability, seizure detection ability, and non-seizure recognition ability.

4.3. Ablation experiments

The DARF-P1DNet model proposed in this paper achieves good detection performance on the Bonn dataset. The complete model reached an accuracy of 99.00%, a sensitivity of 95.00%, and a specificity of 100.00%. These results indicate that the model can effectively distinguish between epileptic seizure and non-seizure EEG segments. While maintaining a high ability to identify non-seizure samples, it also shows good detection capability for seizure samples.

To check the real contribution of each module in the model, we carried out two sets of ablation experiments. When the attention module was eliminated, the model accuracy dropped to 97.00%, with a sensitivity of 100.00% and a specificity of 96.25%. After removing the residual fusion module, the accuracy, sensitivity and specificity were all 95.00%. Comparing with the whole model, the performance of both ablation groups decreased. This suggests that both the attention mechanism and the residual fusion significantly improve the model performance. Specifically, the attention mechanism can strengthen the feature responses which have distinguishing significance. The residual fusion can maintain the basic information during the feature reweighting process, thus enhancing the overall classification ability of the model.

Table 1. Comparison study

Methodology	Year	Accuracy	Sensitivity	Specificity
CNN [11]	2020	98.67%	99.00%	98.00%
1D-CNN [12]	2021	97.52%	97.11%	97.94%

DOI: 10.54254/2755-2721/2026.LD37868

Table 1. (continued)

LSTM [13]	2022	97.85%	97.43%	98.22%
SVM + wavelet-entropy features [14]	2023	98.20%	98.60%	97.80%
P-1D-CNN [15]	2024	98.11%	97.62%	98.57%
This work	2025	99.00%	95.00%	100.00%

From the table, the complete model obtains an accuracy of 99.00% on the Bonn dataset, which is an increase of 2.00 percentage points compared with Ablation Group 1 and 4.00 percentage points compared with Ablation Group 2. It shows that both the dual-view encoding and the attention-residual fusion modules have a positive effect on the performance. As for the specificity, the complete model reaches 100%, while Ablation Group 1 only has 96.25%. This indicates that multi-view feature fusion can effectively decrease the false positives. Although Ablation Group 1 has a sensitivity of 100%, its specificity decreases greatly. It suggests that after eliminating the dual-view mechanism, the model's ability to discriminate non-seizure samples is reduced. On the UCI dataset, the model still achieves an accuracy of 97.13%, which is only 1.87 percentage points lower than that on the Bonn dataset. This provides empirical evidence for the model's cross-dataset stability.

We contrast our method with other researches on the Bonn dataset. The experimental outcomes indicate that DARF-P1DNet achieves an accuracy of 99.00% and a specificity of 100.00%. The overall performance is superior to some of the compared methods. Although its sensitivity is 95.00%, which is lower than that of some others, it demonstrates remarkable strengths in accuracy and specificity. This implies that the model has good comprehensive classification ability. In order to further assess the model's generalization capability, we test it on the UCI Epileptic Seizure Recognition dataset. It obtains an accuracy of 97.13%, a sensitivity of 95.00% and a specificity of 97.66%. These results show that the model can keep a stable classification performance under different data distributions and sample lengths. This suggests that DARF-P1DNet has certain generalization ability.

5. Conclusion

This paper proposes the DARF-P1DNet model to address the problem of insufficient feature representation from a single view in small-sample EEG seizure detection. The model is built on P-1D-CNN. It extracts complementary features through dual-path encoding of the original view and the augmented view. It merges Feature-wise Attention and Residual Fusion for feature selection and information preservation. The experimental results indicate that the whole model obtains an accuracy of 99.00%, a sensitivity of 95.00% and a specificity of 100.00% on the Bonn dataset. The ablation experiments also confirm the effectiveness of the attention mechanism and the residual fusion module in enhancing the model performance.

In the experiments using UCI Epileptic Seizure Recognition dataset, the model achieves an accuracy of 97.13%, a sensitivity of 95.00% and a specificity of 97.66%. It indicates that the model has certain generalization ability under different data conditions. Further studies may confirm the model's performance on larger clinical EEG data from various sources. Additionally, future work can evaluate the reliability and feasibility of the model in assisting seizure detection in real clinical situations, which would benefit its application in future clinical use.

DOI: 10.54254/2755-2721/2026.LD37868

References

[1] Alotaiby, T., Alshebeili, S., Alotaibi, F., et al. (2021). A review of epileptic seizure detection from EEG signals: Traditional and deep learning methods. *IEEE Access, 9*, 140522–140543.

[2] Wang, Y. F., Pang, Y., Zhou, Q. N., et al. (2023). Automatic epileptic seizure detection method based on multi-class SVM algorithm. *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 35*(3), 536–544.

[3] Kang, T. Z. (2022). Automatic epileptic seizure detection algorithm based on dual density dual tree complex wavelet transform. *West China Medical Journal, 37*(5), 721–728.

[4] Yang, S. H., Li, B., & Zhou, F. F. (2021). Cross-patient automatic epileptic seizure detection algorithm based on machine learning. *Journal of Jilin University (Science Edition), 59*(1), 101–106.

[5] Wang, X., Wang, X., Liu, W., et al. (2021). One dimensional convolutional neural networks for seizure onset detection using long-term scalp and intracranial EEG. *Neurocomputing, 463*, 372–387.

[6] Singh, A., Kaur, M., & Singh, D. (2023). Novel deep learning framework for detection of epileptic seizures using EEG signals. *Computational Biology and Medicine, 155*, 106432.

[7] Kumar, S., & Srivastava, A. (2023). Automatic epileptic seizure detection based on EEG using a moth-flame optimization of one-dimensional convolutional neural networks. *Sensors, 23*(15), 6872.

[8] Omar, M., El-Sayed, A., & Abdel-Raouf, M. (2022). Effects of data augmentation with the BNNSMOTE algorithm in seizure detection using 1D-MobileNet. *Bioengineering, 9*(12), 781.

[9] Tian, X., Deng, Z., Choi, K., et al. (2021). Deep multi-view feature learning for EEG-based epileptic seizure detection. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 29*, 2478-2489.

[10] Lv, H., Zhang, Y., Xiao, T., et al. (2023). Seizure detection based on lightweight inverted residual attention network. *International Journal of Neural Systems, 33*(7), 2350032.

[11] Sahani, M., Rout, S. K., & Dash, P. K. (2021). FPGA implementation of epileptic seizure detection using semisupervised reduced deep convolutional neural network. *Applied Soft Computing, 113*, 107922.

[12] Zhang, Y., Liu, X., Wang, Y., et al. (2022). 1D-CNN based seizure detection in EEG signals with optimal channel selection. *Biomedical Signal Processing and Control, 75*, 103565.

[13] Tuncer, E., & Doğru Bolat, E. (2022). Classification of epileptic seizures from electroencephalogram (EEG) data using bidirectional short-term memory (Bi-LSTM) network architecture. *Biomedical Signal Processing and Control, 71*, 103130.

[14] Sharma, S., & Pachori, R. B. (2022). Epileptic seizure detection based on wavelet entropy features and support vector machine. *Biomedical Signal Processing and Control, 77*, 103794.

[15] Ullah, I., Hussain, M., Qazi, E.-u.-H., et al. (2021). An automated system for epileptic seizure detection using pyramidal one-dimensional convolutional neural network from EEG signals. *IEEE Access, 9*, 39928-39941.