

## Article

# LMA-EEGNet: A Lightweight Multi-Attention Network for Neonatal Seizure Detection Using EEG signals

Weicheng Zhou, Wei Zheng , Youbing Feng and Xiaolong Li

Ocean College, Jiangsu University of Science and Technology, Zhenjiang 212100, China;  
212241803637@stu.just.edu.cn (W.Z.); yzfyb@just.edu.cn (Y.F.); lixiaolong@just.edu.cn (X.L.)

\* Correspondence: zhengwei@just.edu.cn

**Abstract:** Neonatal epilepsy is an early postnatal brain disorder, and automatic seizure detection is crucial for timely diagnosis and treatment to reduce potential brain damage. This work proposes a novel Lightweight Multi-Attention Network, LMA-EEGNet, for diagnosing neonatal epileptic seizures from multi-channel EEG signals employing dilated depthwise separable convolution (DDS Conv) for feature extraction and using pointwise convolution followed by global average pooling for classification. The proposed approach substantially reduces the model size, number of parameters, and computational complexity, which are crucial for real-time detection and clinical diagnosis of neonatal epileptic seizures. LMA-EEGNet integrates temporal and spectral features through distinct temporal and spectral branches. The temporal branch uses DDS Conv to extract temporal features, enhanced by a channel attention mechanism. The spectral branch utilizes similar convolutions alongside a spatial attention mechanism to highlight key frequency components. Outputs from both branches are merged and processed through a pointwise convolution layer and a global average pooling layer for efficient neonatal seizure detection. Experimental results show that our model, with only 2471 parameters and a size of 23 KB, achieves an accuracy of 95.71% and an AUC of 0.9862, demonstrating its potential for practical deployment. This study provides an effective deep learning solution for the early detection of neonatal epileptic seizures, improving diagnostic accuracy and timeliness.



**Citation:** Zhou, W.; Zheng, W.; Feng, Y.; Li, X. LMA-EEGNet: A Lightweight Multi-Attention Network for Neonatal Seizure Detection Using EEG signals. *Electronics* **2024**, *13*, 2354. <https://doi.org/10.3390/electronics13122354>

Academic Editors: Shing-Hong Liu and Chiun-Li Chin

Received: 13 May 2024  
Revised: 13 June 2024  
Accepted: 14 June 2024  
Published: 16 June 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** neonatal seizure detection; EEG signals; deep learning; attention mechanism; lightweight network

## 1. Introduction

Neonatal epilepsy is a neurological disorder that occurs within the first 28 days of life, characterized by recurrent seizures. The causes are varied, including genetic factors, brain development abnormalities, infections, metabolic disorders, and hypoxic-ischemic encephalopathy. Seizure symptoms may include convulsions, apnea, and eye deviation. Timely and accurate detection of such seizures is crucial for preventing potential long-term brain damage, guiding appropriate treatment strategies, and improving health outcomes for neonates [1].

In recent years, deep learning has been widely applied across numerous fields. By constructing multi-layered neural network architectures, it can automatically extract features and recognize patterns from large datasets, thereby addressing many complex real-world problems. For instance, D Ai et al. [2] proposed an innovative deep learning approach that utilizes one-dimensional convolutional neural networks (1D CNNs) to analyze raw electromechanical impedance (EMA) data for identifying structural damage in concrete, significantly improving damage detection accuracy. Similarly, W Zhang et al. [3] introduced a deep convolutional neural network (CNN) with new training methods for bearing fault diagnosis in noisy environments and under varying working loads, enhancing the reliability and robustness of fault diagnosis. These successful applications showcase the immense potential of deep learning in tackling complex issues across different industries.

As deep learning techniques have been widely applied in the healthcare domain [4], particularly with significant advancements in medical image analysis, automated disease diagnosis methods have rapidly evolved. In recent years, various machine learning and deep learning-based methods have been proposed to improve the diagnostic accuracy of neonatal epilepsy. Y Liu et al. [5] proposed a method that utilizes ordinal pattern representation combined with a nearest neighbor classifier to detect seizures in EEG signals, demonstrating its effectiveness in epilepsy diagnosis. Utilizing convolutional neural networks (CNN), O'Shea et al. [6,7] successfully enhanced the recognition rate of neonatal epilepsy. They further developed an FCN architecture capable of learning hierarchical representations of raw EEG data, demonstrating both the efficiency and practicality of this approach in neonatal seizure detection. AM Pavel et al. [8] conducted a two-arm, parallel, controlled trial to evaluate the diagnostic accuracy of an automated machine learning algorithm known as ANSeR (Algorithm for Neonatal Seizure Recognition) in identifying neonatal epileptic seizures. The study findings indicate that the ANSeR algorithm performs well in terms of safety and accuracy, effectively detecting neonatal epilepsy. In addition, P. Striano et al. [9] explored the application of machine learning in epilepsy detection. Although still in its early stages, this approach has demonstrated potential for automatically detecting epileptic seizures from EEG signals. A Gramacki et al. [10] developed a deep learning framework for epilepsy detection and proposed an efficient automatic epilepsy detection method by analyzing selected neonatal EEG recordings. For a diagnosis of severity levels in neonatal epileptic seizures, BS Debelo et al. [11] introduced a diagnostic system based on deep convolutional neural networks, which demonstrated high efficiency and accuracy on actual medical datasets. K Visalini et al. [12] demonstrated a machine learning architecture based on Deep Belief Networks (DBN) for binary classification of epileptic and non-epileptic phases. This DBN-based approach offers a novel technological method for the automatic monitoring and diagnosis of neonatal epilepsy, possessing potential for future clinical application.

These research efforts demonstrate the applicability and potential of deep learning approaches in the detection of neonatal epilepsy. These approaches not only utilize the traditional time–frequency characteristics of EEG signals but also pioneer new directions in the in-depth analysis of EEG features through deep learning architectures. Nevertheless, existing methods still face challenges when dealing with highly complex and nonlinear EEG data [13]. These challenges include but are not limited to improving detection accuracy, reducing false-positive rates, and the computational burden of real-time monitoring.

Deep learning models typically feature extensive parameters and high computational demands, which particularly present challenges in actual medical settings. In recent years, there has been extensive research into lightweight deep learning networks due to their ability to deliver high performance with low computational costs. For instance, X Hu et al. [14] proposed a lightweight multi-scale attention-guided network for real-time semantic segmentation, significantly enhancing the efficiency and accuracy of pixel-level classification. Similarly, F Xie et al. [15] developed a multi-scale convolutional attention network designed for lightweight image super-resolution, demonstrating superior performance in enhancing image resolution while maintaining low computational costs. Moreover, Ziya Ata Yazıcı et al. [16] introduced GLIMS, an attention-guided lightweight multi-scale hybrid network for volumetric semantic segmentation, significantly improving 3D medical image analysis. Yufeng Z et al. [17] proposed a lightweight deep convolutional network with inverted residuals to effectively match optical and SAR images, enhancing the robustness and accuracy of image matching tasks. These studies highlight the potential and effectiveness of lightweight deep learning networks across various fields.

In this study, our main contributions are as follows: First, we introduced a novel lightweight multi-attention network (LMA-EEGNet) specifically designed for diagnosing neonatal epileptic seizures. Second, we integrated dilated depthwise separable convolution (DDS Conv) in the feature extraction process, which significantly reduces the model size and computational complexity, thus providing an efficient solution for resource-constrained

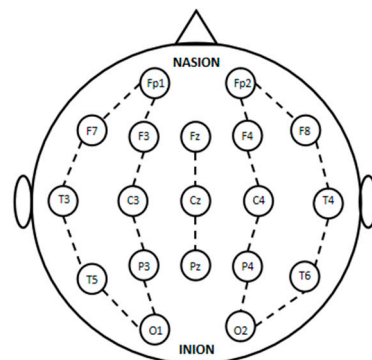
environments. Additionally, we designed temporal and spectral branches to extract the respective features of EEG signals and enhanced them using attention mechanisms, thereby improving seizure detection accuracy. Finally, unlike traditional methods that use fully connected layers for classification, we employed pointwise convolution and global average pooling layers. This approach not only ensures high accuracy but also maintains a small number of parameters and a compact model size. Through these innovations and contributions, our research provides an effective and efficient solution for detecting neonatal seizures.

## 2. Methods

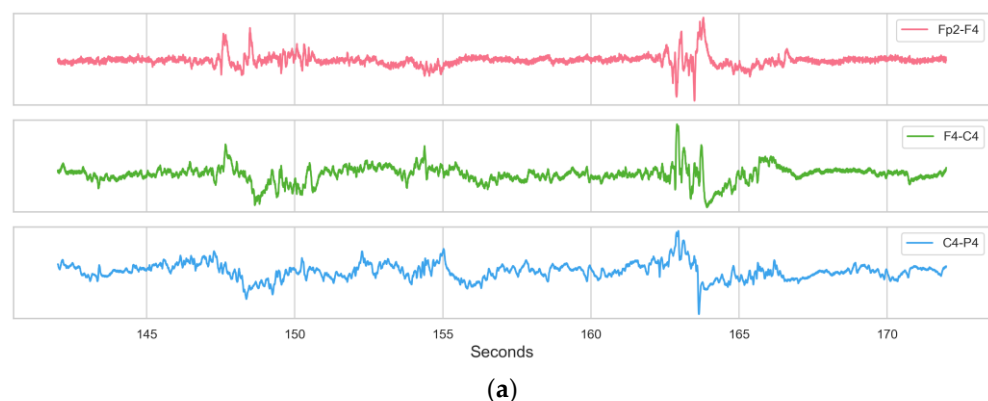
### 2.1. Dataset

The dataset applied to our algorithm contains medical data from the publicly available dataset of Helsinki University Hospital, which recorded multi-channel electroencephalograms (EEGs) from 79 full-term neonates admitted to the Neonatal Intensive Care Unit (NICU) at Helsinki University Hospital [18]. The study involving this dataset had received the necessary ethical approval. The EEG signals were sampled at 256 Hz, with a recording length of approximately 60 min.

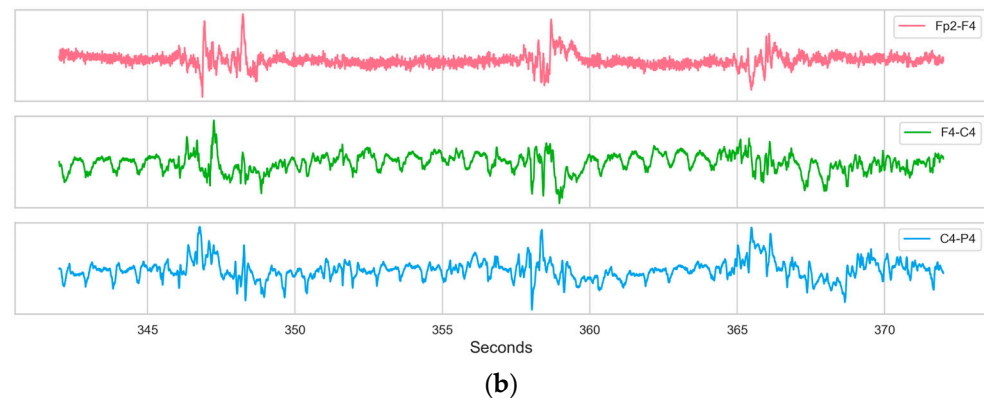
Each file includes potentials from 19 electrodes, with each electrode positioned and labeled according to the international 10–20 system. Figure 1 shows the standard 10–20 electrode placement for EEG recording. The 18 bipolar montage channels formed between the electrodes are described as follows: Fp2-F4, F4-C4, C4-P4, P4-O2, Fp1-F3, F3-C3, C3-P3, P3-O1, Fp2-F8, F8-T4, T4-T6, T6-O2, Fp1-F7, F7-T3, T3-T5, T5-O1, Fz-Cz, and Cz-Pz. Figure 2 shows the EEG activity of Sample 9, highlighting sections of non-seizure and seizure states.



**Figure 1.** Standard 10–20 electrode placement for EEG recording.



**Figure 2.** Cont.



**Figure 2.** EEG activity of Sample 9: sections of non-seizure (a) and seizure (b) states.

## 2.2. Data Preprocessing

The dataset includes annotations from three experts who independently labeled each second of each sample. A label of 0 indicates that the expert did not observe a seizure in that second, whereas a label of 1 indicates that the expert observed a seizure.

We found that there were discrepancies among the experts regarding the seizure labels. Therefore, it was necessary to unify the label processing. If two or more experts labeled a particular second as a seizure, we considered that the neonate indeed experienced a seizure during that time. Conversely, if only one or no expert labeled a particular second as a seizure, we considered that there was no seizure during that second. For each second of each infant, we calculated the seizure label frequency (frequency of 0, 1, 2, or 3). We regarded frequencies of 0 and 1 as no seizure experienced, and frequencies of 2 and 3 as a seizure experienced. Evidently, for each second, a seizure label frequency of 0 or 3 meant that the three experts had no disagreement about the state during that second, while a seizure label frequency of 1 or 2 meant that the experts had inconsistent opinions about the state during that second. We counted the total time for seizure label frequencies of 0, 1, 2, and 3, denoted as  $T_0$ ,  $T_1$ ,  $T_2$ ,  $T_3$ , respectively. In our study, the Annotation Difference Rate (ADR) is defined for each sample as a measure to quantify the degree of discrepancy in annotations among experts:

$$ADR = \frac{T_1 + T_2}{T_1 + T_2 + T_0 + T_3} \times 100\% \quad (1)$$

Among the 79 infants, 22 were considered by all three experts as not having seizures, while 39 were considered by these three experts as having experienced seizures. Among these 39 infants, there were some whose labels showed significant disagreement among the three experts, with some infants having an ADR as high as 67.8%. Table 1 shows the annotation data for partial infant EEG samples.

Considering that data with significant expert disagreement may affect the model's performance, among the samples that experts considered as having experienced seizures, we selected patients 9, 13, 14, 36, 39, 44, 47, and 62 as experimental samples, as the difference in annotations by the three experts was not too significant for these patients. At the same time, we also selected patients 3, 10, 27, 28, 30, 32, and 35 from the samples that experts considered as not having experienced seizures, to construct the dataset along with the aforementioned samples. This enriches the diversity of the dataset and improves the generalization performance of the model.

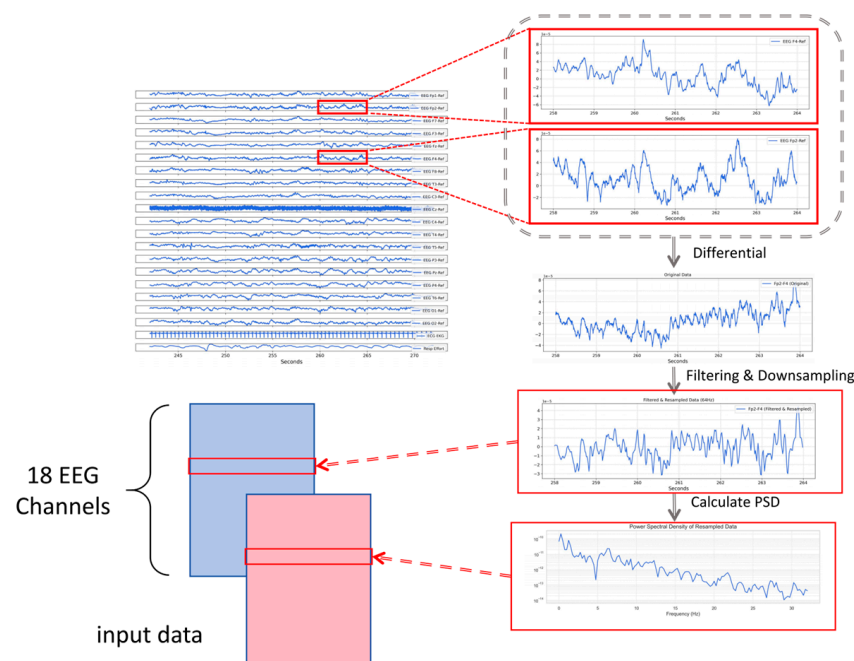
The raw EEG files in the dataset are stored in the EDF format, with each file containing data from 19 electrodes, as well as electrocardiogram (ECG) and respiration effort (Resp Effort) data. We need to subtract the data from the two electrodes at the ends of the desired channels to form the channel data, eventually forming EEG data from 18 channels. Since EEG signals can be affected by various noise such as electromyographic noise and ocular noise [19], we applied a 0.5–32 Hz bandpass filter to each channel. Additionally, the

sampling rate of 256 Hz is relatively high, requiring substantial computational resources for processing. Therefore, we chose to downsample the data from each channel to 64 Hz.

**Table 1.** Annotation data for partial infant EEG samples.

Sample ID	T <sub>0</sub> (s)	T <sub>1</sub> (s)	T <sub>2</sub> (s)	T <sub>3</sub> (s)	ADR (%)
1	3541	1909	798	745	38.7
16	1671	2560	1468	242	67.8
54	1833	1603	908	0	57.8
63	1648	1394	514	344	48.9
9	2507	163	18	862	5.1
13	13,980	69	132	1235	1.3
14	1221	225	196	2084	11.2
36	4549	42	40	451	1.6
39	2065	300	87	2177	8.3
44	2961	40	44	315	2.5
47	3291	106	9	200	3.1
62	5462	8	2	380	0.1

To obtain the input data for our deep learning network, we need to segment the EEG data into labeled windows. To ensure that the label within each window is consistent, we use a sliding window approach to segment the 18-channel EEG signals [20]. Each window contains 6 s of EEG signals, resulting in a time-domain window size of (18, 384). Next, to obtain the frequency-domain representation of each channel, we employ the Welch power spectral density estimation method [21]. This involves dividing the time signal into segments, computing the periodogram for each segment, and then averaging to obtain the power spectral density for each channel. The advantage of this method is that it can reduce the influence of noise and improve the accuracy of power spectral density estimation, particularly when the input signal is non-stationary (such as EEG signals), where the Welch method can yield more accurate results. The size of each frequency-domain window is (18, 129). Figure 3 shows the diagram of data processing.

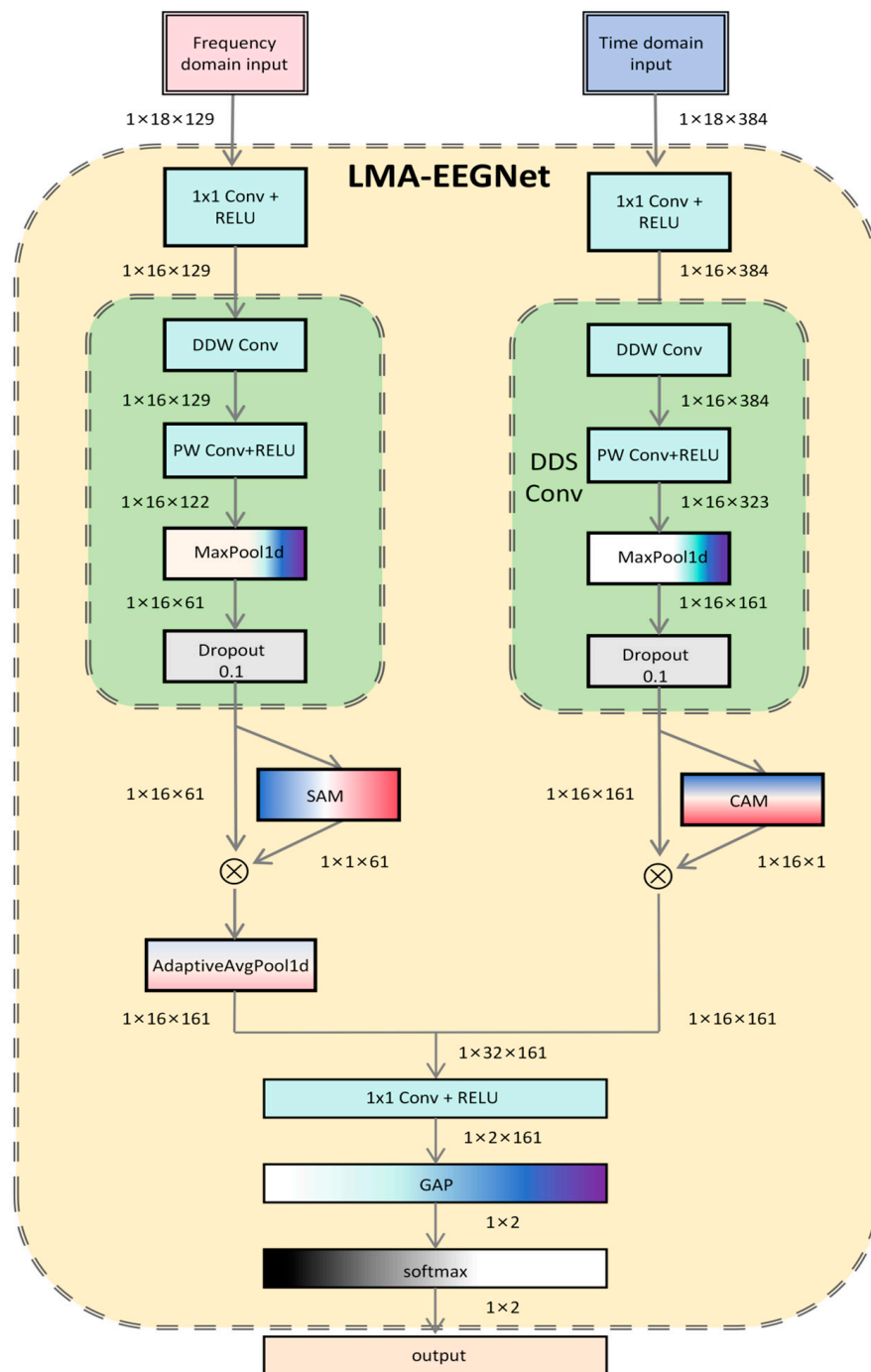


**Figure 3.** The diagram of data processing.

### 2.3. LMA-EEGNet's Entire Structure

#### 2.3.1. Model Overview

Lightweight Multi-Attention EEGNet (LMA-EEGNet) is a lightweight neural network designed for detecting neonatal epileptic seizures. The network adopts a dual-path architecture that can process time-domain and frequency-domain features in parallel while employing multiple attention mechanisms to enhance the extraction of critical information [22]. Furthermore, by replacing traditional 1D convolutional layers and fully connected layers with dilated depthwise separable convolutions and pointwise convolutions, respectively, the network reduces the number of parameters and network complexity while maintaining efficient feature extraction capabilities. Figure 4 shows the structure of LMA-EEGNet.



**Figure 4.** The structure of LMA-EEGNet.



For the time-domain path, it first uses pointwise convolution for cross-channel information fusion, then employs dilated depthwise separable convolution to extract features from the time-series data, followed by the RELU activation function to enhance non-linear representational capability. Pooling layers and dropout layers are used to reduce feature dimensions and prevent overfitting. Finally, a channel attention module (CAM) is employed to enhance the model's attention to the most informative channels.

The frequency-domain path also employs pointwise convolution and dilated depthwise separable convolution for extracting frequency-domain features. An adaptive average pooling layer (AdaptiveAvgPool1d) is used to match the time-domain sequence length, ensuring time-frequency alignment of the feature maps. After extracting the frequency-domain features, a spatial attention module (SAM) is used to generate an attention map, which is then expanded and applied to the frequency-domain feature maps, enabling the model to focus on important frequency components.

After processing the time-domain and frequency-domain features, the model fuses the two sets of feature maps. The fused feature map goes through a pointwise convolutional layer for final classification, where the number of output channels matches the number of target classes. Next, global average pooling is applied to reduce the response of each classification channel to a single scalar value, ultimately outputting a one-dimensional feature vector that represents the predicted probabilities for different classes (i.e., seizure or non-seizure).

### 2.3.2. Dilated Depthwise Separable Convolution (DDS Conv)

Depthwise separable convolutions significantly reduce the number of model parameters and computational costs by decomposing a standard convolution into two steps: a depthwise convolution and a pointwise convolution [23]. Specifically, the depthwise convolution is applied independently on each input channel, while the pointwise convolution is responsible for merging the outputs of these channels. This decomposition makes the number of parameters linearly related to the size of the convolution kernels and the number of input channels, rather than the cubic relationship in traditional convolutions.

In a standard 1D convolution, if there are  $M$  input channels and  $N$  output channels, and a convolution kernel of length  $K$  is used, then the total number of parameters  $N_{CNN-para}$  would be as follows:

$$N_{CNN-para} = K \times M \times N \quad (2)$$

In depthwise separable convolutions, the number of parameters is divided into two parts: the depthwise convolution applies a convolution kernel of length  $K$  independently to each input channel, and the pointwise convolution uses a  $1 \times 1$  convolution kernel to combine the results of the depthwise convolution across channels. The total number of parameters  $N_{DSC-para}$  is the following:

$$N_{DSC-para} = K \times M + N \times M \quad (3)$$

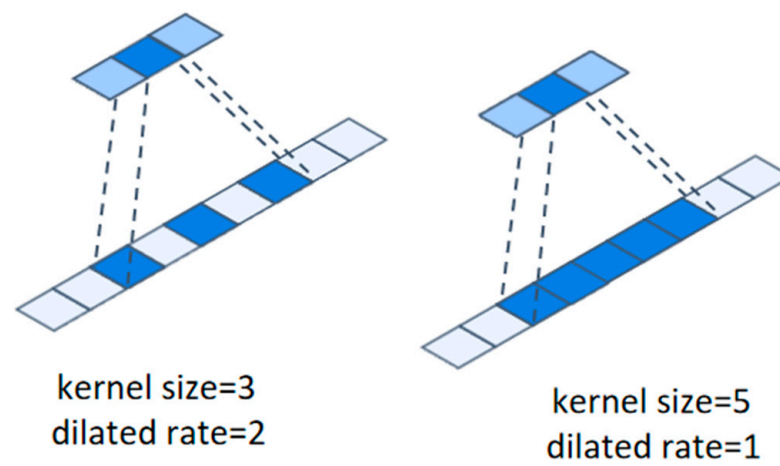
In depthwise separable convolutions, the reduction in the number of parameters relative to standard convolutions can be quantified by the following ratio:

$$\frac{N_{DSC-para}}{N_{CNN-para}} = \frac{K \times M + N \times M}{K \times M \times N} = \frac{1}{N} + \frac{1}{K} \quad (4)$$

This ratio demonstrates the direct relationship between the parameter reduction ratio and the size of the convolution kernel  $K$  and the number of output channels  $N$ . Particularly, when the size of the convolution kernel  $K$  or the number of output channels  $N$  increases, the parameter savings achieved by depthwise separable convolutions relative to standard convolutions become more significant. This parameter-saving property makes depthwise separable convolutions beneficial for constructing lightweight and computationally efficient

neural network architectures, especially when processing one-dimensional signals like EEGs with a large number of channels.

To further lighten the network, we improved the depthwise separable convolutional layer by introducing dilation in the depthwise convolutional layer, resulting in a dilated depthwise separable convolution (DDS Conv). Dilated convolutions enlarge the convolution kernel by inserting gaps between the kernel elements, effectively increasing the kernel size without increasing the number of parameters. From another perspective, with a fixed kernel size, dilated convolutions reduce the number of parameters compared to standard convolutions by removing some elements from the convolution kernel. Figure 5 shows that dilated convolution achieves the same receptive field with fewer parameters. By introducing dilation, we further reduced the number of parameters in the network, achieving the goal of a lightweight network.



**Figure 5.** Comparison between dilated convolution and regular convolution.

### 2.3.3. Channel Attention Module (CAM)

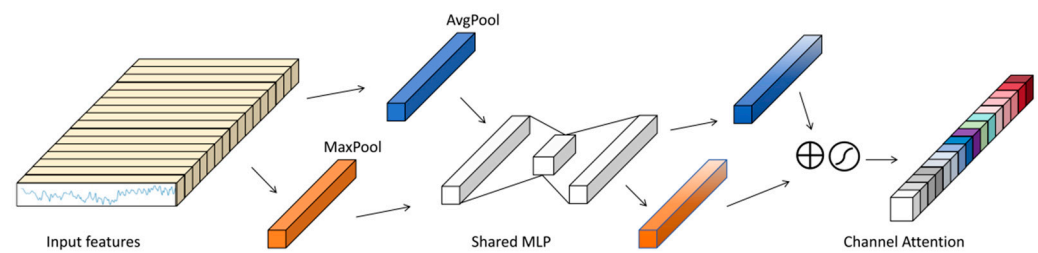
In the time-domain branch, the introduction of the channel attention module (CAM) aims to enhance the model's adaptability to the importance of different channels in representing key information in time-series data [24]. It is based on the assumption that different channels contribute differently to representing crucial information in the time series. Through global average pooling and global max pooling, CAM extracts global statistical features from the multi-channel data, which are subsequently used to train a multi-layer perceptron (MLP). The MLP learns non-linear relationships between channels by setting a bottleneck layer (i.e., reducing the number of channels via the reduction ratio), effectively reducing model complexity and mitigating the risk of overfitting. The resulting channel weight map, obtained through the sigmoid activation function, guides the model to concentrate resources on processing the most informative channels, thereby improving the representational capability of time-domain features within a limited computational budget.

The channel attention module (CAM) computes the channel attention weights as follows:

$$M_c(X) = \sigma(MLP(AvgPool(X)) + MLP(MaxPool(X))) \quad (5)$$

where  $X$  is the input feature map;  $AvgPool(X)$  and  $MaxPool(X)$  represent global average and max pooling operations, which compress the spatial dimensions to focus on channel-wise statistics. The  $MLP$  function processes these statistics to capture non-linear inter-channel relationships. The final attention weights  $M_c(X)$  are obtained by applying the sigmoid function  $\sigma$ , normalizing the outputs to emphasize more informative channels effectively. Figure 6 shows the diagram of CAM.





**Figure 6.** The diagram of CAM.

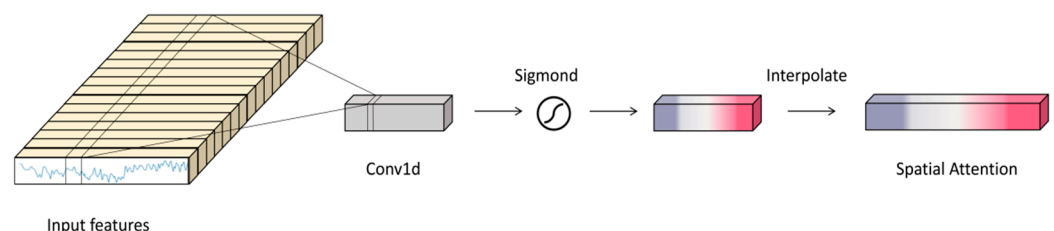
#### 2.3.4. Spatial Attention Module (SAM)

Meanwhile, in the frequency domain branch, the spatial attention module (SAM) is employed, which utilizes large-kernel convolutions along the spatial dimension to capture long-range dependencies between frequency domain features. In frequency domain analysis, different signal components (e.g., waves of different frequencies) correspond to different temporal behaviors. For instance, low-frequency components may correspond to long-term trends in the signal, while high-frequency components may relate to short-term fluctuations. By focusing on regions in the frequency domain that exhibit significant temporal structure through the spatial attention module (SAM), the model can more effectively identify and utilize these time-varying frequency domain features for decision-making or prediction tasks. SAM generates a one-dimensional attention map using single-channel convolution kernels and an expanded receptive field, which is then modulated by the sigmoid function and interpolated to match the original size of the frequency domain feature maps. This strategy allows the network to learn which frequency components are more important for the current task, enabling it to allocate more attention to processing these components. Consequently, the network can focus more on the frequency domain features that are more helpful for seizure detection rather than treating all frequency components as equally important, potentially improving model performance.

The spatial attention module (SAM) computes the spatial attention weights as follows:

$$M_s(X) = \text{Interp}(\sigma(\text{Conv1d}(X))) \quad (6)$$

where  $X$  denotes the input feature map. The operation  $\text{Conv1d}(X)$  applies a convolution to  $X$  to capture local dependencies and patterns. The sigmoid function  $\sigma$  normalizes these convolution outputs into the range  $[0, 1]$ , producing a preliminary attention map. The function  $\text{Interp}$  then adjusts the size of this attention map to match the spatial dimensions of the input  $X$ . This resizing is crucial to ensure that the attention weights can be applied element-wise to the original input feature map  $X$ . Figure 7 shows the diagram of SAM.



**Figure 7.** The diagram of SAM.

#### 2.3.5. Pointwise Convolution and Global Average Pooling

At the classification stage, to further alleviate the network burden and reduce the number of model parameters, we pioneered the use of a pointwise convolutional layer and global average pooling instead of a traditional fully connected layer for feature fusion and dimensionality reduction in the field of neonatal seizure detection. Pointwise convolution, commonly known as  $1 \times 1$  convolution, has the main advantage of enabling linear combinations across channels while preserving the spatial dimensions of the feature maps, thereby

significantly reducing the number of model parameters [25]. This design aims to achieve effective feature compression while retaining spatial information. Additionally, global average pooling is employed to further downsample the spatial dimensions, reducing the computational complexity [26].

Through the pointwise convolutional layer, we reduced the number of channels for merging features from 32 to the required number of classes. Immediately after, we applied a global average pooling layer to the output of the pointwise convolutional layer, which compressed the spatial dimensions of the entire feature map to 1. The purpose of this step was to average the global information across each feature channel, generating a tensor with dimensions (batch size, num classes). This tensor was then directly used for computing the classification loss, thereby completing the class prediction. This approach not only improved the computational efficiency of the model but also helped mitigate overfitting, making the model more robust in practical applications.

### 3. Experiment

#### 3.1. Evaluation Metrics

To evaluate the model's performance, we chose accuracy, sensitivity, specificity, and AUC as the evaluation metrics. The following equations are used to calculate these metrics:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (7)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (9)$$

At the same time, we also selected the number of model parameters and the number of floating-point operations (Flops) to assess the model's complexity, highlighting the lightweight nature of our model.

#### 3.2. Experimental Setup and Results

In our experiments, all neural networks were implemented using the PyTorch 2.1.2 framework and trained in a supervised manner on an Nvidia GPU. The Adam optimizer was used for training with a minimum batch size of 16. The learning rate was set to 0.001, and the models were trained for 150 epochs.

As mentioned above, we used the Helsinki dataset to validate the performance of our model. After segmenting the data into windows, we balanced the number of positive and negative samples in the dataset through undersampling, as the imbalance between positive and negative samples can cause the model to overfit to the majority class and perform poorly on the minority class, affecting the overall model performance [27].

The proportion of dataset division significantly impacts the final performance of the model. To determine the optimal division ratio, we experimented with three different proportions: 60% training, 20% validation, and 20% testing; 70% training, 15% validation, and 15% testing; and 80% training, 10% validation, and 10% testing. For each division ratio, we conducted five repeated experiments and compared various performance metrics (see Table 2 for detailed results). Ultimately, we selected the division ratio of 80% training, 10% validation, and 10% testing. To ensure that the training results of the model did not become biased towards one class, we maintained a balanced number of samples from both classes in each subset during the dataset division.

We trained our model on the dataset, and to achieve optimal performance and prevent overfitting, the training process was repeated multiple times, up to a maximum of 150 epochs, and stopped when the specified stopping criteria were met. We then evaluated the trained model on the test set. Our model achieved an accuracy of 95.71%, a sensitivity of 95.00%, a specificity of 96.43%, and an AUC of 0.9862. In terms of evaluating the lightweight

metrics of our model, it contains only 2467 parameters, requires only 363,248 floating-point operations, and the complete model size is merely 23.1 KB.

**Table 2.** Comparison results with different dataset division ratios.

Ratios	Mean Acc (%)	Mean Sen (%)	Mean Spe (%)	Mean AUC
60%; 20%; 20%	93	90.216	95.788	97.848
70%; 15%; 15%	93.236	91.142	95.332	97.698
80%; 10%; 10%	93.928	93.288	94.57	98.446

As research on lightweight networks in the field of neonatal epilepsy detection is scarce, in this study, we compared the performance of the LMA-EEGNet model with several other classifiers in the task of seizure detection. Table 3 shows that while maintaining comparable performance, our network significantly reduces the number of parameters compared to other studies, with the parameter count being only 0.0087% to 4.9% of other models. This significant reduction not only implies lower memory usage and computational costs but also enhances the deployability of the model on various computing devices. Furthermore, despite the drastic reduction in parameters, our network can still maintain a high level of diagnostic performance, demonstrating the effectiveness and practicality of our lightweight design.

**Table 3.** Comparison results with other epilepsy seizure detection methods.

Model	Parameters	Acc (%)	Sen (%)	Spe (%)	AUC
LMA-EEGNet	2467	95.7	95.0	96.4	0.9862
PCA+LDA [28]	-	94.7	94.8	89.1	-
DLWH [29]	-	95.1	94.3	95.4	-
2D-CNN [30]	49,560	98.2	82.7	88.2	-
TSKCNN [31]	28,459,615	98.0	96.0	99.0	-
LRCN [32]	9,695,012	99.0	84.0	99.0	-
2D-CNN [10]	424,321	96.2	-	-	-
Fd-CAE [33]	-	92.3	-	98.7	-

The symbol ‘-’ represents undisclosed model performance metrics data for which no related information has been released at present.

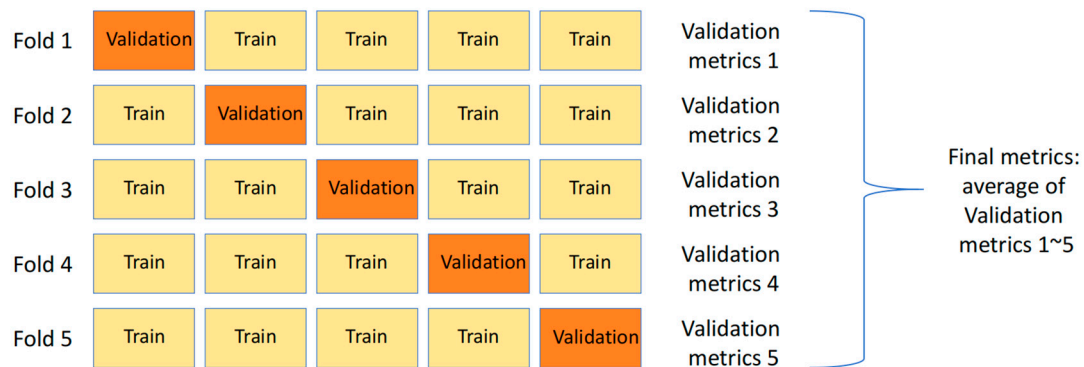
### 3.3. Exploring the Impact of Different Dilation Rates

In this study, we explored the impact of key hyperparameters on the performance of the proposed model, with the main objective of optimizing the model to enhance its overall performance. We selected 20% of the dataset as the test set and used the remaining 80% for training and validation. The training and validation sets were utilized in a 5-fold cross-validation to compare model performance. After each parameter adjustment, we retrained the model using this 5-fold cross-validation approach, took the average of the test results from each fold, and reevaluated the model using these metrics. Figure 8 Shows the diagram of five-fold cross-validation.

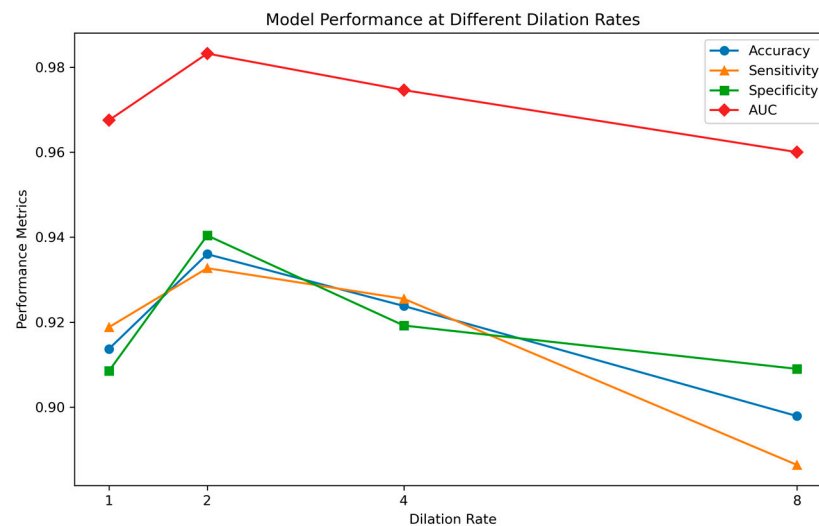
These metrics helped us understand the specific impact of different parameter configurations on the model’s predictive ability. We investigated the effect of different dilation rates on the model’s performance, and the experimental results are presented in Figure 9.

To determine if there are statistically significant differences in AUC scores among models with different dilation rates, we utilized ANOVA (see Tables 4 and 5 for detailed results). Based on Levene’s test for homogeneity of variances, all significance levels are above 0.05, indicating that the variances across groups are equal. The ANOVA results show significant differences in AUC scores among the models with different dilation rates ( $F = 15.261, p < 0.001$ ). This indicates that the dilation rates have a significant impact on model performance. To further explore these differences, we performed post hoc tests using Bonferroni’s method (see Table 6 for detailed results). The analysis revealed that the

AUC scores between Dilation Rate 1 and 2, Dilation 2 and 8, and Dilation Rate 4 and 8 differ significantly.



**Figure 8.** The diagram of five-fold cross-validation.



**Figure 9.** Performance metrics of the model under different dilation rates.

**Table 4.** Levene’s test for homogeneity of variances.

AUC	Test Statistic	df1	df2	Significance
Based on Mean	0.329	3	16	0.805
Based on Median	0.195	3	16	0.898
Based on Median and with Adjusted df	0.195	3	15.213	0.898
Based on Trimmed Mean	0.320	3	16	0.811

**Table 5.** ANOVA results for AUC scores.

AUC	Sum of Squares	df	Mean Square	F	Significance
Between Groups	10.784	3	3.595	15.261	<0.001
Within Groups	3.769	16	0.236		
Total	14.553	19			

From the experimental results, we can observe that the model achieved optimal performance when the dilation rate was set to 2. This suggests that a moderate dilation rate helps the model more effectively capture meaningful temporal features without overfitting or losing important information.

**Table 6.** Multiple comparisons of AUC for different dilation rates (Bonferroni’s method).

(I) Dilation Rate	(J) Dilation Rate	Mean Difference (I-J)	Std. Error	Significance	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	−1.572 *	0.30695	<0.001	−2.4954	−0.6486
	4	−0.700	0.30695	0.220	−1.6234	0.2234
	8	0.348	0.30695	1.000	−0.5754	1.2714
2	1	1.572 *	0.30695	<0.001	0.6486	2.4954
	4	0.872	0.30695	0.071	−0.0514	1.7954
	8	1.920 *	0.30695	<0.001	0.9966	2.8434
4	1	0.700	0.30695	0.220	−0.2234	1.6234
	2	−0.872	0.30695	0.071	−1.7954	0.0514
	8	1.048 *	0.30695	0.021	0.1246	1.9714
8	1	−0.348	0.30695	1.000	−1.2714	0.5754
	2	−1.920 *	0.30695	<0.001	−2.8434	−0.9966
	4	−1.048 *	0.30695	0.021	−1.9714	−0.1246

\* The mean difference is significant at the 0.05 level.

When the dilation rate was 1 (in which case the convolutional layer was a regular deep convolution), the model had the smallest receptive field, which might have caused the model to be overly sensitive to noise and minor variations, affecting its generalization ability. As the dilation rate increased from 2, the model’s performance showed a significant decline. In the case of a dilation rate of 8, the model’s performance dropped noticeably, which could be attributed to the excessively large dilation rate, leading to the loss of important local features despite increasing the receptive field, thus impacting the overall judgment capability of the model. For tasks like epilepsy detection, precise temporal and frequency information is crucial, and if the receptive field is too large, ignoring these details may result in performance degradation.

### 3.4. Ablation Studies

To verify the performance improvements brought by the introduction of various attention mechanisms, we conducted ablation experiments on the attention mechanism modules. Specifically, we performed experiments by separately removing the temporal branch attention module, the frequency branch attention module, and both modules simultaneously. For each configuration, we trained the model five times using a dataset split of 80% training, 10% validation, and 10% testing. We reported the average performance metrics to ensure the reliability and robustness of our results.

In this experiment, we designate the full LMA-EEGNet model as Model 1. Model 2 refers to the LMA-EEGNet model with the CAM module removed, while Model 3 denotes the LMA-EEGNet model without the SAM module. Finally, Model 4 represents the LMA-EEGNet model with all attention modules removed. This nomenclature allows for a clear distinction between the different versions of the LMA-EEGNet model throughout the discussion and analysis.

To determine if there are statistically significant differences in the AUC scores among the models presented in Table 7, we employed a one-way ANOVA (see Tables 8 and 9 for detailed results). This method allows us to rigorously assess the performance variations and ensure the reliability of our ablation study findings.

**Table 7.** Performance metrics of various models on the test set from the ablation study.

Model	Mean Acc (%)	Mean Sen (%)	Mean Spe (%)	Mean AUC
1	93.29	94.43	92.14	0.9853
2	91.00	89.86	92.09	0.9723
3	91.36	89.57	93.14	0.9744
4	89.86	89.14	90.57	0.9648

**Table 8.** Levene’s test for homogeneity of variances in the ablation Study.

AUC	Test Statistic	df1	df2	Significance
Based on Mean	1.710	3	16	0.205
Based on Median	0.852	3	16	0.486
Based on Median and with Adjusted df	0.852	3	9.447	0.498
Based on Trimmed Mean	1.668	3	16	0.214

**Table 9.** ANOVA results for AUC scores in the ablation Study.

AUC	Sum of Squares	df	Mean Square	F	Significance
Between Groups	10.720	3	3.573	10.399	<0.001
Within Groups	5.498	16	0.344		
Total	16.218	19			

The results of Levene’s test for homogeneity of variances indicate that the assumption of equal variances is met, as none of the significance levels are below 0.05; specifically, the significance level based on the mean is 0.205, indicating no significant difference in variances across groups.

The ANOVA results reveal that there are statistically significant differences in the AUC scores among the models. The F-value is 10.399 with a significance level of less than 0.001, which is well below the 0.05 threshold. This indicates that the differences in AUC scores between the groups are highly significant. Consequently, we can conclude that the introduction of different attention mechanisms leads to statistically significant variations in model performance. To further investigate these differences, we conducted post hoc tests using Bonferroni’s method (see Table 10 for detailed results). The results indicate that there are significant differences in AUC scores between Model 1 and Model 2 as well as between Model 1 and Model 4. Figure 10 shows the ROC curves of different models in the ablation study.

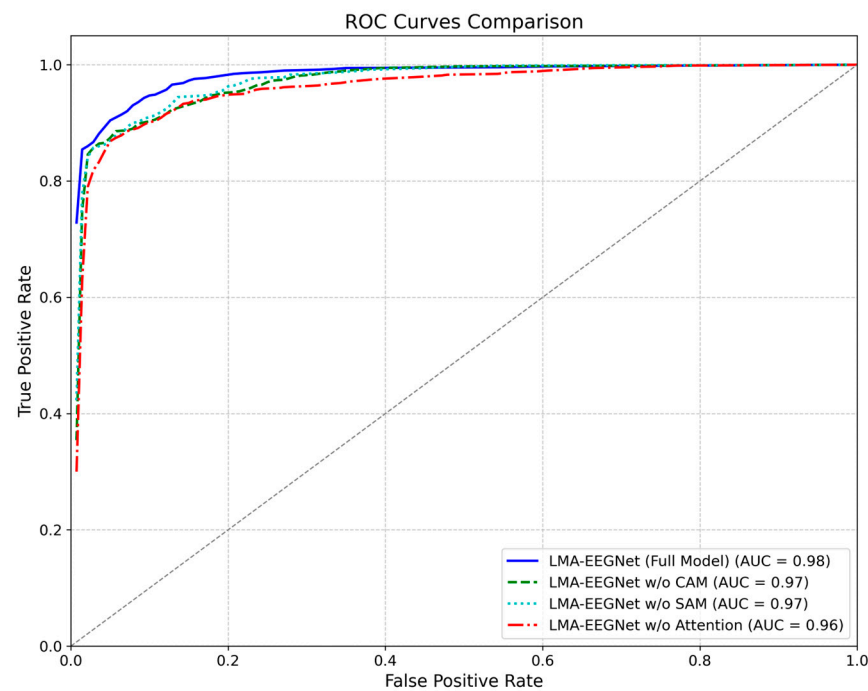
**Table 10.** Post hoc multiple comparisons of AUC for different models (Bonferroni’s method) in the ablation Study.

(I) Model	(J) Model	Mean Difference (I-J)	Std. Error	Significance	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	1.298 *	0.37073	0.018	0.1827	2.4133
	3	1.088	0.37073	0.058	−0.0273	2.2033
	4	2.046 *	0.37073	<0.001	0.9307	3.1613
2	1	−1.298 *	0.37073	0.018	−2.4133	−0.1827
	3	−0.210	0.37073	1.000	−1.3253	0.9053
	4	0.748	0.37073	0.364	−0.3673	1.8633
3	1	−1.088	0.37073	0.058	−2.2033	0.0273
	2	0.210	0.37073	1.000	−0.9053	1.3253
	4	0.958	0.37073	0.120	−0.1573	2.0733
4	1	−2.046 *	0.37073	<0.001	−3.1613	−0.9307
	2	−0.7480	0.37073	0.364	−1.8633	0.3673
	3	−0.958	0.37073	0.120	−2.0733	0.1573

\* The mean difference is significant at the 0.05 level.

The experimental results showed that the complete LMA-EEGNet model (containing all attention mechanisms) not only achieved a mean accuracy of 93.29% on the test set but also exhibited a mean AUC value of 0.9853, indicating the model’s high classification performance and excellent generalization ability.





**Figure 10.** ROC curves of different models in the ablation study.

When the channel attention and spatial attention modules were removed separately, the model's performance declined. After removing the channel attention, the model's mean accuracy dropped to 91.00%, and the mean AUC value decreased to 0.9723. This suggests that channel attention plays an important role in enhancing the model's ability to capture the associations between different channels. When the spatial attention was removed, the model's mean accuracy dropped to 91.36%, and the mean AUC value decreased to 0.9744, reflecting the crucial role of spatial attention in enhancing the model's ability to capture spatial features.

The most significant performance decline occurred in the model where all attention mechanisms were removed simultaneously, with the mean accuracy dropping to 89.86% and the mean AUC value decreasing to 0.9648. This significant performance degradation highlights the importance of attention mechanisms in integrating and enhancing temporal and frequency features, especially when dealing with complex EEG signal data.

#### 4. Conclusions

This study successfully developed a novel neonatal epilepsy detection network based on deep learning. Our network introduces two major innovations in the field of neonatal brain seizure detection: the first application of dilated depthwise separable convolution (DDS Conv), and the initial use of point convolution layers for efficient and accurate classification. These two lightweight design innovations significantly reduce the number of parameters and computational complexity, lowering the demand for computational resources, which makes the model particularly suitable for deployment in resource-constrained environments. Additionally, the performance of the model is enhanced by employing various attention mechanisms and by integrating temporal and spectral features.

In the experimental section, we utilized a publicly available neonatal electroencephalogram (EEG) dataset for validation. The results demonstrate that, compared to existing methods, the model proposed in this study exhibits superior performance on key performance indicators such as accuracy, sensitivity, specificity, and the Area Under the Curve (AUC), while significantly reducing the model size and computational complexity. The model achieved an accuracy of 95.71% on the test set, with a sensitivity of 95.00%, specificity also at 96.43%, and an Area Under the Curve (AUC) of 0.9862. Additionally, we explored the performance of the model under different configurations, including the effects of var-

ious types of attention mechanisms and critical parameters (such as dilation rate). This analysis not only validates the effectiveness of the techniques employed but also provides valuable guidance for future research directions.

Although our model demonstrated outstanding performance in several aspects, there are some limitations. Good generalization capability is crucial for a neonatal seizure detection model [34], and our model's training and validation were performed on a specific dataset. Future research needs to validate the model's generalization ability on a broader range of datasets. We anticipate the availability of datasets with more samples, longer sample durations, and high-quality labels. Moreover, while our model primarily focuses on the detection of neonatal epileptic seizures, the types and specific forms of epilepsy are diverse [35]. It is currently unclear whether the model is equally effective in detecting different types of epileptic seizures. Therefore, future research should take into account the diversity of epilepsy and develop algorithms capable of recognizing and classifying different types of epileptic seizures.

Furthermore, the gap between our testing conditions and real-world application scenarios must be acknowledged. In clinical settings, data quality and conditions can vary significantly from the controlled environments typically used for model training and testing. This discrepancy can affect the model's performance in practice. Real-world applications may involve more noise, variability in signal quality, and differences in patient conditions, which are not fully captured in our current dataset. Addressing these differences will be crucial for the successful deployment of our model in clinical practice.

In future research, we will focus on further optimizing the model structure to accommodate a wider range of application scenarios and data types. Additionally, considering the highly complex, nonlinear, and noise-rich characteristics of EEG signals, we performed certain preprocessing on the data. In real-time detection scenarios, this preprocessing can consume significant computational resources. Therefore, we will also strive to develop lightweight seizure detection models for raw EEG signals. Furthermore, as the labels originate from subjective judgments by experts [36], we aim to leverage various possible methods to enhance the model's interpretability, facilitating a better understanding of the model's decision-making process by medical professionals, thereby increasing its acceptance and trust in clinical applications [37]. Moreover, improving the model's interpretability can also help us identify and address the model's performance shortcomings in specific situations, further enhancing its accuracy and robustness. To this end, we plan to introduce more interpretability mechanisms, such as attention maps and activation mappings, which can intuitively showcase the signal portions the model focuses on the most when making predictions.

**Author Contributions:** Conceptualization, W.Z. (Weicheng Zhou) and W.Z. (Wei Zheng); Funding acquisition, Y.F. and X.L.; Methodology, W.Z. (Weicheng Zhou); Validation, W.Z. (Weicheng Zhou); Writing—original draft, W.Z. (Weicheng Zhou); Writing—review and editing, W.Z. (Weicheng Zhou), W.Z. (Wei Zheng) and Y.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Natural Science Foundation of China, grant number 61601206.

**Data Availability Statement:** The dataset utilized in our study is available for download at <https://zenodo.org/records/1280684> (accessed on 3 January 2024). The relevant ethical approval documents can be viewed at the following link: [[https://zenodo.org/records/4940267/files/Ethics\\_approval.pdf?download=1](https://zenodo.org/records/4940267/files/Ethics_approval.pdf?download=1)] (accessed on 3 January 2024).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Kaminiów, K.; Kozak, S.; Paprocka, J. Neonatal Seizures Revisited. *Children* **2021**, *8*, 155. [[CrossRef](#)] [[PubMed](#)]
2. Ai, D.; Mo, F.; Cheng, J.; Du, L. Deep learning of electromechanical impedance for concrete structural damage identification using 1-D convolutional neural networks. *Constr. Build. Mater.* **2023**, *385*, 131423. [[CrossRef](#)]

3. Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.* **2018**, *100*, 439–453. [\[CrossRef\]](#)
4. Mohammed, B.; Nagy, R.; Ahmed, H.H. Healthcare predictive analytics using machine learning and deep learning techniques: A survey. *J. Electr. Syst. Inf. Technol.* **2023**, *10*, 40.
5. Liu, Y.; Lin, Y.; Jia, Z.; Ma, Y.; Wang, J. Representation based on ordinal patterns for seizure detection in EEG signals. *Comput. Biol. Med.* **2020**, *126*, 104033. [\[CrossRef\]](#) [\[PubMed\]](#)
6. O'Shea, A.; Lightbody, G.; Boylan, G.; Temko, A. Investigating the Impact of CNN Depth on Neonatal Seizure Detection Performance. *IEEE Eng. Med. Biol. Soc.* **2018**, *2018*, 5862–5865.
7. O'Shea, A.; Lightbody, G.; Boylan, G.; Temko, A. Neonatal seizure detection from raw multi-channel EEG using a fully convolutional architecture. *Neural Netw.* **2020**, *123*, 12–25. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Pavel, A.M.; Rennie, J.M.; de Vries, L.S.; Blennow, M.; Foran, A.; Shah, D.K.; Pressler, R.M.; Kapellou, O.; Dempsey, E.M.; Mathieson, S.R.; et al. A machine-learning algorithm for neonatal seizure recognition: A multicentre, randomised, controlled trial. *Lancet Child Adolesc. Health* **2020**, *4*, 740–749. [\[CrossRef\]](#)
9. Striano, P.; Minetti, C. Deep learning for neonatal seizure detection: A friend rather than foe. *Lancet Child Adolesc. Health* **2020**, *4*, 711–712. [\[CrossRef\]](#)
10. Artur, G.; Jarosław, G. A deep learning framework for epileptic seizure detection based on neonatal EEG signals. *Sci. Rep.* **2022**, *12*, 13010.
11. Debelo, B.S.; Thamineni, B.L.; Dasari, H.K.; Dawud, A.A. Detection and Severity Identification of Neonatal Seizure Using Deep Convolutional Neural Networks from Multichannel EEG Signal. *Pediatr. Health Med. Ther.* **2023**, *14*, 405–417. [\[CrossRef\]](#)
12. Visalini, K.; Alagarsamy, S.; Nagarajan, D. Neonatal seizure detection using deep belief networks from multichannel EEG data. *Neural Comput. Appl.* **2023**, *35*, 10637–10647. [\[CrossRef\]](#)
13. Xu, J. Detection methods of Parkinson's Disease based on physiological signals and machine learning methods. *Highlights Sci. Eng. Technol.* **2023**, *36*, 813–822. [\[CrossRef\]](#)
14. Xuegang, H.; Yuanjing, L. Lightweight multi-scale attention-guided network for real-time semantic segmentation. *Image Vis. Comput.* **2023**, *139*, 104823.
15. Feng, X.; Pei, L.; Xiaoyong, L. Multi-scale convolutional attention network for lightweight image super-resolution. *J. Vis. Commun. Image Represent.* **2023**, *95*, 103889.
16. Yazıcı, A.Z.; Öksüz, İ.; Ekenel, K.H. GLIMS: Attention-guided lightweight multi-scale hybrid network for volumetric semantic segmentation. *Image Vis. Comput.* **2024**, *146*, 105055. [\[CrossRef\]](#)
17. Zhu, Y.; Yu, S.; He, H.; Xia, Y.; Zhou, F. A lightweight deep convolutional network with inverted residuals for matching optical and SAR images. *Int. J. Remote Sens.* **2024**, *45*, 3597–3622. [\[CrossRef\]](#)
18. Ryu, S.; Back, S.; Lee, S.; Seo, H.; Park, C.; Lee, K.; Kim, D.S. Pilot study of a single-channel EEG seizure detection algorithm using machine learning. *Child's Nerv. Syst. ChNS Off. J. Int. Soc. Pediatr. Neurosurg.* **2021**, *37*, 2239–2244. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Anchal, Y.; Singh, M.C. A new approach for ocular artifact removal from EEG signal using EEMD and SCICA. *Cogent Eng.* **2020**, *7*, 835146.
20. Gaur, P.; Gupta, H.; Chowdhury, A.; McCreadie, K.; Pachori, R.B.; Wang, H. A Sliding Window Common Spatial Pattern for Enhancing Motor Imagery Classification in EEG-BCI. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 4002709. [\[CrossRef\]](#)
21. Göker, H. Welch Spectral Analysis and Deep Learning Approach for Diagnosing Alzheimer's Disease from Resting-State EEG Recordings. *Trait. Du Signal* **2023**, *40*, 257–264. [\[CrossRef\]](#)
22. Lu, L.; Liu, T.; Jiang, F.; Han, B.; Zhao, P.; Wang, G. DFANet: Denoising Frequency Attention Network for Building Footprint Extraction in Very-High-Resolution Remote Sensing Images. *Electronics* **2023**, *12*, 4592. [\[CrossRef\]](#)
23. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, NV, USA, 27–30 June 2016.
24. Wang, P.; Chen, J.; Wang, Z.; Shao, W. Fault diagnosis for spent fuel shearing machines based on Bayesian optimization and CBAM-ResNet. *Meas. Sci. Technol.* **2024**, *35*, 025901. [\[CrossRef\]](#)
25. Zhao, W.; Wang, G.; Atapattu, S.; He, R.; Liang, Y.C. Channel Estimation for Ambient Backscatter Communication Systems with Massive-Antenna Reader. *IEEE Trans. Veh. Technol.* **2019**, *68*, 8254–8258. [\[CrossRef\]](#)
26. Zhu, F.; Liu, C.; Yang, J.; Wang, S. An Improved MobileNet Network with Wavelet Energy and Global Average Pooling for Rotating Machinery Fault Diagnosis. *Sensors* **2022**, *22*, 4427. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Ali-Gombe, A.; Elyan, E. MFC-GAN: Class-imbalanced dataset classification using Multiple Fake Class Generative Adversarial Network. *Neurocomputing* **2019**, *361*, 212–221. [\[CrossRef\]](#)
28. Tao, Z.; Wanzhong, C. LMD Based Features for the Automatic Seizure Detection of EEG Signals Using SVM. *IEEE Trans. Neural Syst. Rehabil. Eng. A Publ. IEEE Eng. Med. Biol. Soc.* **2017**, *25*, 1100–1108.
29. Khalilpour, S.; Ranjbar, A.; Menhaj, M.B.; Sandooghdar, A. Application of 1-D CNN to predict epileptic seizures using eeg records. In Proceedings of the 2020 6th International Conference on Web Research (ICWR), Tehran, Iran, 22–23 April 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 314–318.
30. Hossain, M.S.; Amin, S.U.; Alsulaiman, M.; Muhammad, G. Applying Deep Learning for Epilepsy Seizure Detection and Brain Mapping Visualization. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2019**, *15*, 1–17. [\[CrossRef\]](#)

31. Tian, X.; Deng, Z.; Ying, W.; Choi, K.-S.; Wu, D.; Qin, B.; Wang, J.; Shen, H.; Wang, S. Deep multi-view feature learning for eeg-based epileptic seizure detection. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2019**, *27*, 1962–1972. [[CrossRef](#)]
32. Liang, W.; Pei, H.; Cai, Q.; Wang, Y. Scalp eeg epileptogenic zone recognition and localization based on long-term recurrent convolutional network. *Neurocomputing* **2020**, *396*, 569–576. [[CrossRef](#)]
33. Wang, Y.; Yuan, S.; Liu, J.X.; Hu, W.; Jia, Q.; Xu, F. Combining EEG Features and Convolutional Autoencoder for Neonatal Seizure Detection. *Int. J. Neural Syst.* **2024**, *14*, 2450040. [[CrossRef](#)] [[PubMed](#)]
34. Zhang, Z.; Xiao, M.; Ji, T.; Jiang, Y.; Lin, T.; Zhou, X.; Lin, Z. Efficient and generalizable cross-patient epileptic seizure detection through a spiking neural network. *Front. Neurosci.* **2024**, *17*, 1303564. [[CrossRef](#)] [[PubMed](#)]
35. Shellhaas, R.A. Neonatal seizures reach the mainstream: The ILAE classification of seizures in the neonate. *Epilepsia* **2021**, *62*, 629–631. [[CrossRef](#)] [[PubMed](#)]
36. Isaev, D.Y.; Tchapyjnikov, D.; Cotten, C.M.; Tanaka, D.; Martinez, N.; Bertran, M.; Sapiro, G.; Carlson, D. Attention-Based Network for Weak Labels in Neonatal Seizure Detection. *Proc. Mach. Learn. Res.* **2020**, *126*, 479–507.
37. Ho, C.Y.; Robert, T.; Naomi, S. Machine learning in medicine: Should the pursuit of enhanced interpretability be abandoned? *J. Med. Ethics* **2021**, *48*, 581–585.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.