

Article

Effects of the Hyperparameters on CNNs for MDD Classification Using Resting-State EEG

Chia-Yen Yang ^{1,*} and Hsin-Min Lee ² ¹ Department of Biomedical Engineering, Ming-Chuan University, Taoyuan 333321, Taiwan² Department of Physical Therapy, I-Shou University, Kaohsiung 840301, Taiwan

* Correspondence: cyyang@mail.mcu.edu.tw

Abstract: To monitor patients with depression, objective diagnostic tools that apply biosignals and exhibit high repeatability and efficiency should be developed. Although different models can help automatically learn discriminative features, inappropriate adoption of input forms and network structures may cause performance degradation. Accordingly, the aim of this study was to systematically evaluate the effects of convolutional neural network (CNN) architectures when using two common electroencephalography (EEG) inputs on the classification of major depressive disorder (MDD). EEG data for 21 patients with MDD and 21 healthy controls were obtained from an open-source database. Five hyperparameters (i.e., number of convolutional layers, filter size, pooling type, hidden size, and batch size) were then evaluated. Finally, Grad-CAM and saliency map were applied to visualize the trained models. When raw EEG signals were employed, optimal performance and efficiency were achieved as more convolutional layers and max pooling were used. Furthermore, when mixed features were employed, a larger hidden layer and smaller batch size were optimal. Compared with other complex networks, this configuration involves a relatively small number of layers and less training time but a relatively high accuracy. Thus, high accuracy (>99%) can be achieved in MDD classification by using an appropriate combination in a simple model.

Keywords: depression; resting-state electroencephalography (EEG); convolutional neural network (CNN); input; hyperparameter



Citation: Yang, C.-Y.; Lee, H.-M. Effects of the Hyperparameters on CNNs for MDD Classification Using Resting-State EEG. *Electronics* **2024**, *13*, 186. <https://doi.org/10.3390/electronics13010186>

Academic Editor: Stanley C. Ahalt

Received: 5 December 2023

Revised: 28 December 2023

Accepted: 30 December 2023

Published: 31 December 2023



Copyright: © 2023 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/>).

1. Introduction

Depression is a common and severe mental disorder. This multifaceted disorder is caused by diverse factors, including genetics, major events, certain medications, medical conditions, and changes in the brain [1,2]. The World Health Organization used depression as the main theme of the 2017 World Health Day, placing special emphasis on the increasing problems of depression [3]. Furthermore, the impact of the COVID-19 pandemic has rapidly aggravated these problems, resulting in a 27.6% increase in the number of major depressive disorder (MDD) cases [4]. Depression is expected to be the largest contributor to mental disorders by 2030 [5].

Although the pattern of depression varies from one person to another, depression has some common symptoms, including feelings of hopelessness, anger or irritability, a lack of energy, and sleep disturbances [6]. The clinical diagnosis of depression is usually based on a detailed history of symptoms assessed by a physician or mental health professional according to certain criteria, such as those set by the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [7]. To address questionnaire limitations, many studies have explored physiological distinctions between healthy people and patients with depression, using these distinctions as identification markers. For example, Wu et al. [8] extracted relative electroencephalography (EEG) power features from five frequency bands to train a novel conformal kernel support vector machine. Their findings showed the best performance (accuracy = 83.64%, sensitivity = 87.50%, specificity = 80.65%) in delta power

at FT8-T4, theta power at Fz-FCz, FP1-FP2, and FT8-T4, alpha power at FP1-FP2, and gamma power at CP3-CP4 and FT8-CP4. Furthermore, Schwartzmann et al. [9] explored whether resting-state EEG power could assess the outcomes of cognitive behavioral therapy. They found increased delta power (0.5–4 Hz) and decreased alpha power (8–12 Hz) in responders, suggesting the potential utility of resting EEG in predicting optimal treatment for patients with MDD.

With advances in computer science, researchers have begun using deep learning technology for diagnosing diseases. For example, Acharya et al. [10] proposed a model using 13-layer convolutional neural networks (CNNs) to test EEG signals obtained from 15 patients with depression and 15 healthy controls (HCs); this model obviates the necessity of feeding a semimanually selected set of features into a classifier. They reported that the proposed model achieved classification accuracy rates of 93.5% and 96.0% for EEG signals recorded from the left (FP1-T3) and right (FP2-T4) hemispheres, respectively. Li et al. [11] separately applied CNNs to a two-dimensional (2D) data form of functional connectivity matrices from five EEG bands to identify mild depression. Their graph theory analysis results revealed that the brain functional network observed for patients with mild depression had a greater characteristic path length and a smaller clustering coefficient than that observed for HCs. The accuracy of their proposed model was 80.74%. Moreover, Uyulan et al. [12] compared three different deep CNN structures, namely ResNet-50, MobileNet, and Inception-v3, to dichotomize 46 patients with MDD and 46 HCs. Data were collected from 19 electrodes for delta, theta, alpha, and beta bands. Their results indicated that the MobileNet architecture exhibited the highest spatially dependent classification accuracies, which were 89.33% and 92.66% for the left and right hemispheres, respectively; in addition, the ResNet-50 and MobileNet architectures had the highest frequency-dependent accuracies, which were 90.22% (in the delta band) and 80.25% (in the theta band), respectively.

Those neural networks can automatically learn discriminative features from diverse input data, leading to a rapid increase in interest for applications. However, many studies have utilized well-known models without fully grasping the impact of hyperparameters on those. Adopting inappropriate input forms and network structures may degrade the performance and efficiency of deep learning algorithms [13]. Therefore, network structures in different input modalities should be tested before clinical implementation. EEG is valued for depression diagnosis due to its high temporal resolution, noninvasiveness, and cost-effectiveness [14]. Accordingly, the aim of the present study was to systematically evaluate the effects of various CNN model architectures using two different EEG inputs on the MDD classification performance. In line with prior research (e.g., [15]), two common input modalities, such as one-dimensional (1D) mixed features (identified in our previous study by using a support vector machine), and 2D representations of raw EEG signals were investigated in the present study. Here, a simple CNN structure adjusted by five hyperparameters, namely the number of convolutional layers, filter size, pooling type, hidden size, and batch size, was selected to determine the most suitable forms for the classifiers. Accuracy and training time were evaluated to investigate how those hyperparameters affect the characteristics of depression classifiers. Finally, Grad-CAM and saliency map were applied to visualize the learned features from the trained models, to confirm the performance of optimal selections, and to verify the findings associated with MDD. Our combination exhibits fewer layers and quicker training times compared to other complex networks yet maintains high accuracy. This emphasizes the potential for achieving high classification accuracy by employing a suitable combination within a simpler model. Furthermore, using raw resting-state EEG data directly streamlines data processing and benefits future applications.

The rest of this paper is organized as follows. Section 2 provides the details of the data sets, analysis, and classification method of the CNN model. Section 3 describes the test results as well as the performance and saliency maps of the model. Section 4 analyzes the

effects of various hyperparameters and compares the accuracy with other studies. Section 5 summarizes and concludes the paper.

2. Materials and Methods

The following describes the data sets, methodology for analysis and classification, and the evaluation metrics used in our study.

2.1. Data Sets

EEG data were downloaded from the Patient Repository of EEG Data + Computational Tools developed by Cavanagh et al. [16]. A total of 21 patients with MDD and 21 HCs aged 18 to 25 years were included. None of the participants had a history of head trauma or seizures, and none of them were currently on any psychoactive medication. The Beck Depression Inventory (BDI) is a 21-item self-report questionnaire assessing depression levels via characteristic attitudes and symptoms [17]. In contrast to the HCs, who had a BDI score of <7 points, the patients with MDD had a score of >13 points and were diagnosed according to the Structured Clinical Interview for Depression (Table 1).

Table 1. Parameters of the CNN model for seizure prediction.

Item	MDD	HC	Statistic (<i>t</i> -Test)
Gender	8 M: 13 F	8 M: 13 F	-
Age	18.86 ± 1.35	18.67 ± 0.73	<i>p</i> = 0.573
BDI	21.52 ± 5.66	1.00 ± 1.05	<i>p</i> < 0.001

2.2. Data Acquisition

Throughout their experiments, all participants were instructed to minimize movements and maintain a state of thoughtfulness but without falling asleep. The resting state was divided into two conditions: one with eyes closed and the other with eyes open. EEG was recorded for 5 min using whole-head 64 Ag–AgCl scalp electrodes on a SynAmps2 system (Neuroscan, Charlotte, NC, USA). The electrical responses were sampled at 500 Hz and bandpass filtered between 0.5 and 100 Hz. Data were then selected from 58 channels according to the International 10–20 System. The online reference was a single channel placed between Cz and CPz, and the electrode impedance was set to <10 kΩ.

2.3. Data Analysis

The collected data were downsampled to 250 Hz and then divided into two types used as inputs for classification. The first type comprised raw EEG signals with a matrix size of 58 (channels) × 500 (time points) (Figure 1a). The signals were preprocessed in two steps: the first step involved detrending all truncated signals to remove any offsets and slow linear drifts over the time course, and the second step entailed filtering all detrended signals by using a 0.5–50-Hz bandpass filter. Furthermore, signals collected in the eyes-open condition were decomposed using the FastICA algorithm to manually remove components containing eye movements and blink artifacts. Finally, each input was truncated by a half-overlapping window of 2 s.

The second type comprised mixed features with a matrix size of 1 × 4441 (Figure 1b). After the signals were preprocessed, they were decomposed into five frequency bands by using discrete Daubechies wavelet decomposition (order = 4): delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (30–50 Hz). Seven features identified using Student's *t*-test in our previous study were then calculated:

1. The first feature was the frequency power, which was derived by dividing the power calculated in each band by the total power across the frequency spectrum.
2. The second feature was the alpha interhemispheric asymmetry, which was derived by calculating a total of 25 pairs of indices by the formula $(R - L)/(R + L)$, where R and L represent the right- and left-hemisphere power, respectively.

3. The third feature was the left–right coherence, which was derived by evaluating a total of 625 (25 left \times 25 right) electrode pairs for interhemispheric coherence.
4. The fourth feature was the structural properties of the network, which were evaluated by binarizing the coherence values obtained between electrodes into a network using a threshold and then calculating the strength, clustering coefficient, and path length of the network.
5. The fifth feature was the sample entropy, which was derived by modifying the approximate entropy used for assessing complexity.
6. The sixth feature was the multiscale entropy, which was derived by extending the sample entropy calculated at multiple time scales and then summing all calculated values.
7. Finally, the seventh feature was detrended fluctuation, which was derived by dividing the cumulative sum calculated from time-series data into time windows of length n and then linearly fitting them using least-squares errors; the slope of the log of the root-mean-square deviation from the trend against $\log n$ within each time window was calculated using a least-squares method.

Each input was a combination of all features: 290 (58 channels \times 5 bands) frequency power, 25 alpha asymmetry, 3125 (625 pairs \times 5 bands) coherence, 15 (3 \times 5 bands) network characteristic, 348 (58 channels \times 5 bands + 58 channels in the whole band) sample entropy, 348 (58 channels \times 5 bands + 58 channels in the whole band) multiscale entropy, and 290 (58 channels \times 5 bands) detrended fluctuation, totaling 4441.

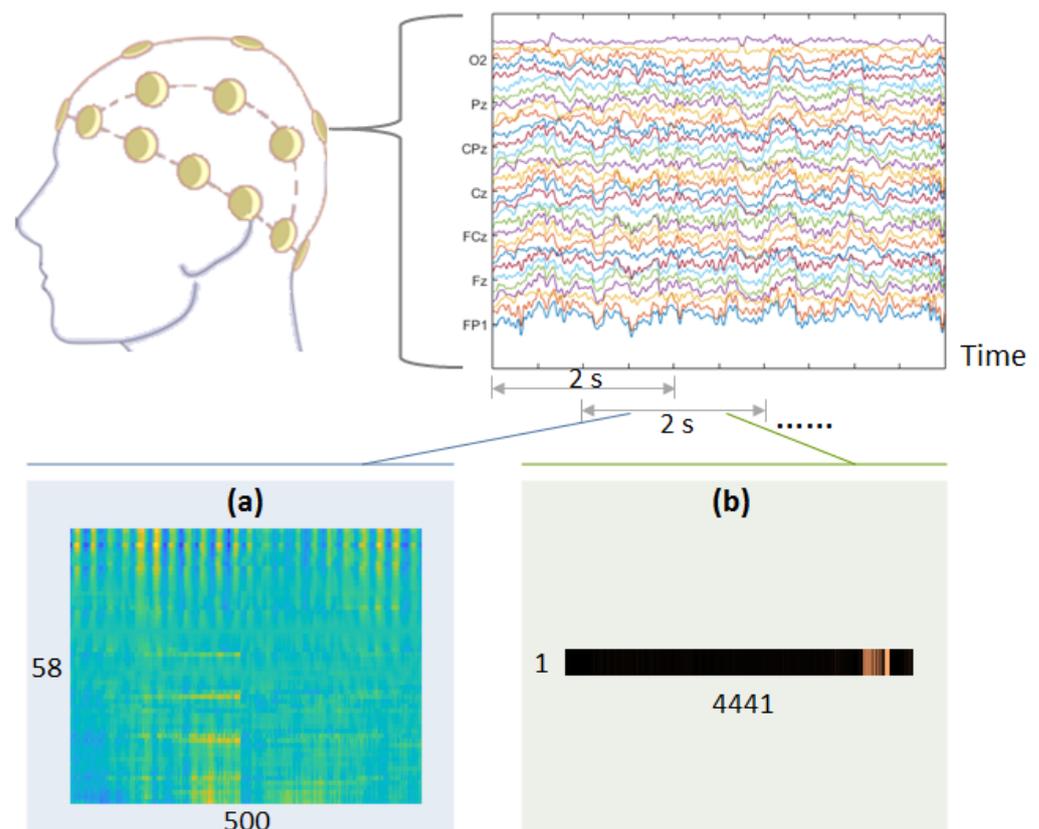


Figure 1. Illustration of the two types of input data for CNN: (a) raw EEG signals and (b) mixed features. Distinct colors represent different input values.

2.4. Data Classification

The architecture of the 2D CNN model used in this study for the classification of MDD is presented in Figure 2 and Table 2. Several hyperparameters selected from the literature, including the number of convolutional layers (i.e., 3, 4, or 5), filter size (i.e., 3 or 5), pooling type (i.e., average and max), hidden size (i.e., 128, 256, or 512), and batch size (i.e., 16, 32,

64, or 128) were tested. Adam optimization was then applied for parameter learning with a learning rate of 0.001. The input data were randomly divided into two sets (with 80% of the data used for training and 20% used for testing) by using a record-wise method. Subsequently, fivefold cross-validation was used to evaluate the performance of the trained model. An early stop strategy was implemented with a patience of 10 to terminate the training process. Finally, the optimized model was tested using the testing data set. These steps were performed ten times (Figure 3).

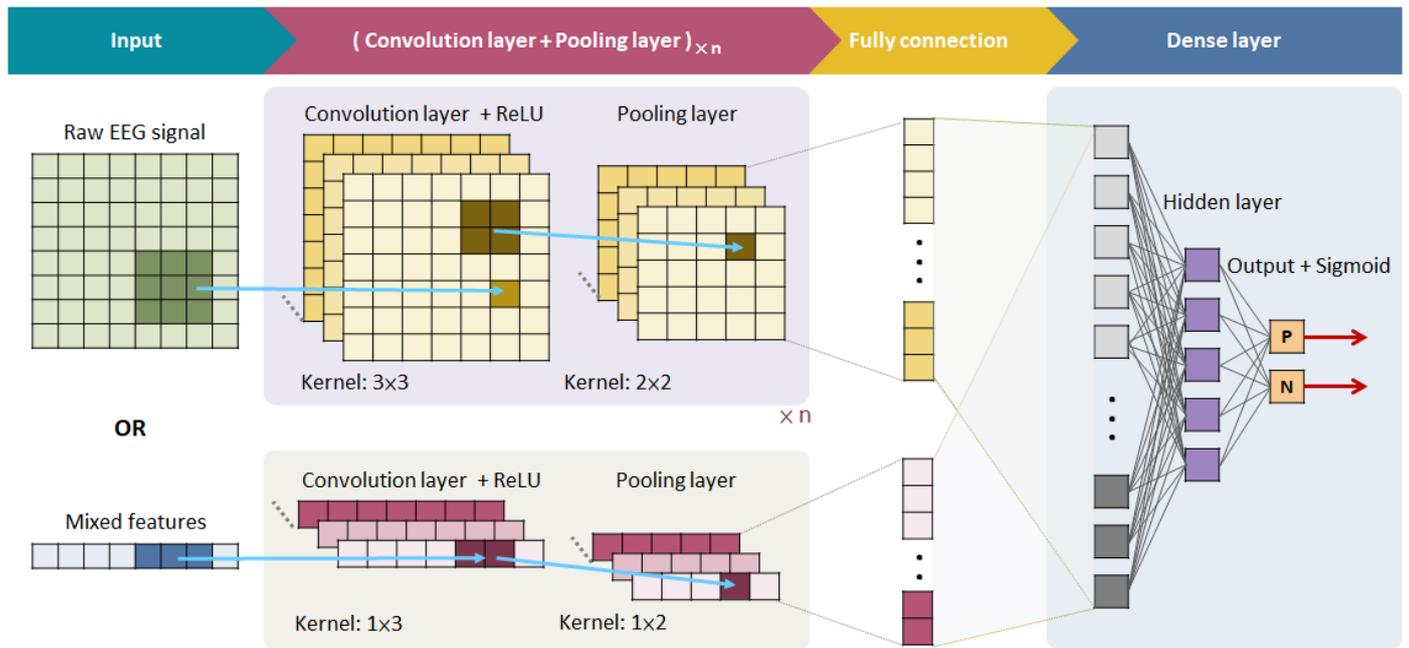


Figure 2. Illustration of two types of inputs in the CNN architecture for MDD classification, spanning from the input layer to the output layer. n indicates the number of convolutional layers.

Table 2. Parameters of the CNN model for MDD classification.

Layer	Type	Filter Size	# Filter	Stride	Output
conv_1	Conv	3/5	16	1	$58 \times 500 \times 16/1 \times 4441 \times 16$
pooling_1	Max/Avg	2	1	2	$29 \times 250 \times 16/1 \times 2221 \times 16$
conv_2	Conv	3/5	32	1	$29 \times 250 \times 32/1 \times 2221 \times 32$
pooling_2	Max/Avg	2	1	2	$15 \times 125 \times 32/1 \times 1111 \times 32$
conv_3	Conv	3/5	64	1	$15 \times 125 \times 64/1 \times 1111 \times 64$
pooling_3	Max/Avg	2	1	2	$8 \times 63 \times 64/1 \times 556 \times 64$
conv_4	Conv	3/5	128	1	$8 \times 63 \times 128/1 \times 556 \times 128$
pooling_4	Max/Avg	2	1	2	$4 \times 32 \times 128/1 \times 278 \times 128$
conv_5	Conv	3/5	256	1	$4 \times 32 \times 256/1 \times 278 \times 256$
pooling_5	Max/Avg	2	1	2	$2 \times 16 \times 256/1 \times 139 \times 256$
flatten_1	Flatten	-	-	-	8192/35584
dense_1	Dense	-	-	-	512/256/128
dropout_1	Dropout (0.5)	-	-	-	512/256/128
dense_2	Dense	-	-	-	2

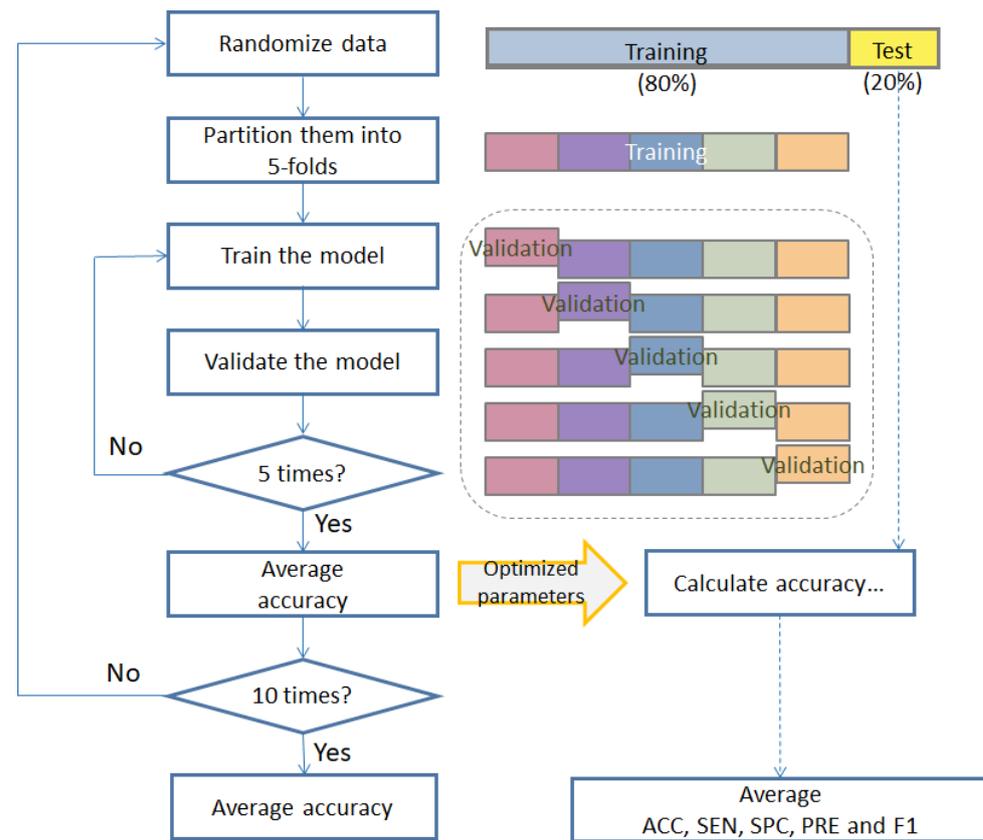


Figure 3. CNN architecture for MDD classification. n indicates the number of convolutional layers.

2.5. Performance Evaluation

To evaluate the effects of the various model architecture combinations on classification performance, a confusion matrix and calculated key performance metrics including accuracy, specificity, sensitivity, precision, and F1 score were generated. Then, one-way analysis of variance was used to test for significant differences in accuracy ($p < 0.05$), with the Tukey–Kramer correction employed for pairwise multiple comparisons.

The classification performance of the model was evaluated in terms of these metrics by using a graphics processing unit (GPU) server with an Intel Xeon W-3225 processor, Nvidia Titan RTX GPU, 192 GB of RAM, and running Ubuntu 18.04 LTS with CUDA 10.1.

2.6. Feature Visualization

Saliency maps were utilized to visualize the learned features and confirm the intended use of the trained model for MDD classification. The mean saliency values across all subjects were aggregated for a comprehensive overview. Furthermore, class-specific gradients were generated for each feature map in every convolutional layer using the Grad-CAM technique [18]. Gradient variances from sample trials were then calculated for further comparative analysis.

3. Results

3.1. Test Results Obtained with Various Configuration

The raw EEG and mixed features contributed 4116 inputs each. First, the CNN model was trained and tested using the raw EEG signals. When utilizing signals under the eyes-closed condition, the classification accuracy of the model significantly increased from 84% to 96% as the number of layers increased and from 90% to 93% as the pooling type changed to max (Figure 4a). The training time increased from 19.85 to 30.28 s as the number of layers increased, and from 24.59 to 25.72 s as the pooling type changed to max. When utilizing signals under the eyes-open condition, the classification accuracy of the model significantly

increased from 91% to 98% as the number of layers increased and from 95% to 96% as the pooling type changed to max (Figure 4b). The training time increased from 20.97 to 30.74 s as the number of layers increased and from 26.79 to 25.86 s as the pooling type changed to max.

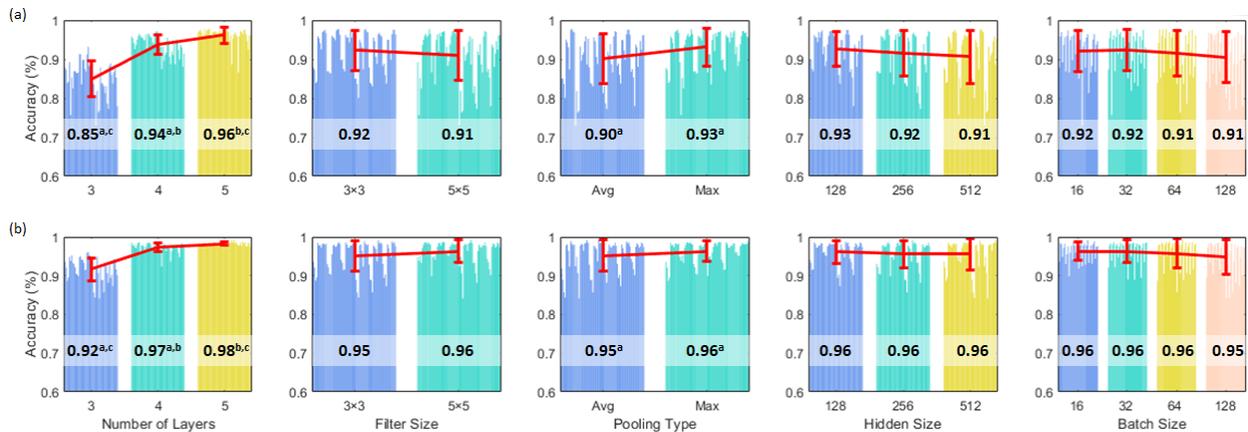


Figure 4. Mean accuracy rates derived under various network settings with raw EEG signals under the (a) eyes-closed and (b) eyes-open conditions. Different colors denote distinct hyperparameter configurations. The red lines indicate global averaged values in each hyperparameter test. The superscripts a, b, and c indicate statistical significance.

Then, the CNN model was trained and tested using the mixed features. When signals were utilized under the eyes-closed condition, the classification accuracy of the model significantly increased from 89% to 90% and from 92% to 93% as the filter size and hidden size increased, respectively, and from 88% to 90% as the pooling type changed to max, but it significantly decreased from 93% to 92% as the batch size increased (Figure 5a). The training time increased from 21.41 to 22.17 s and from 20.33 to 24.19 s as the filter size and hidden size increased, respectively, but it decreased from 25.02 to 18.56 s as the pooling type changed to max and from 34.29 to 14.15 s as the batch size increased. When utilizing signals under the eyes-open condition, the classification accuracy of the model significantly increased from 92% to 93% as the number of layers and hidden size increased, but it significantly decreased from 93% to 92% as the batch size increased (Figure 5b). The training time increased from 19.44 to 26.48 s and from 21.72 to 25.26 s as the number of layers and hidden size increased, respectively, but it significantly decreased from 35.63 to 15.50 s as the batch size increased.

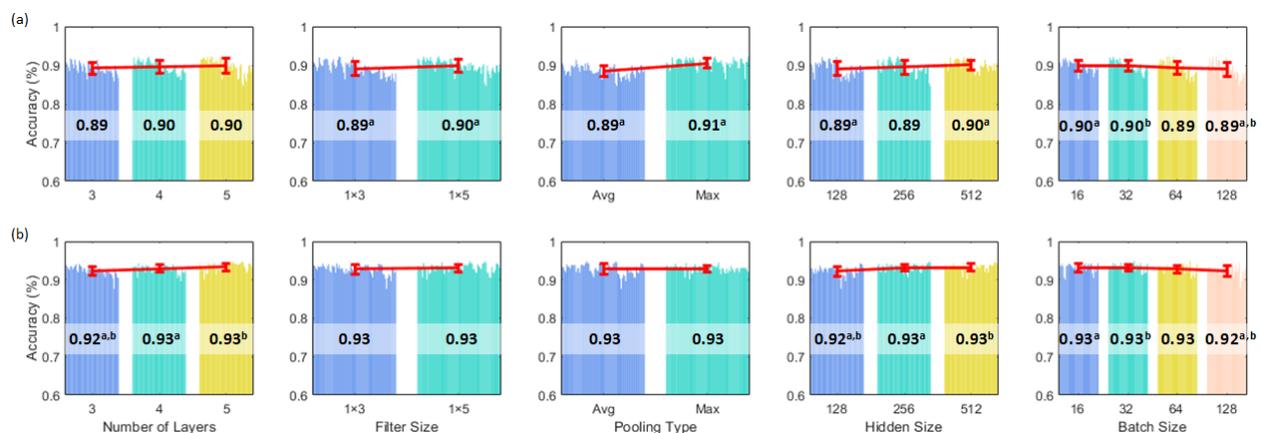


Figure 5. Mean accuracy rates derived under various network settings using mixed EEG features under the (a) eyes-closed and (b) eyes-open conditions. Different colors denote distinct hyperparameter configurations. The red lines indicate global averaged values in each hyperparameter test. The superscripts a and b indicate statistical significance.

3.2. Classification Performance

To evaluate the effects of the input and model architecture combinations on CNN classification performance, we identified hyperparameters associated with the two highest mean accuracy rates of the trained model for each input data type under each condition (Table 3). Under the eyes-closed condition, when raw EEG signals were used as the input, the highest mean accuracy (97.8%) was obtained when the number of layers was set to 5, the filter size was set as 3×3 , the pooling type was set as avg, the hidden size was set as 512, and the batch size was set as 32. In this case, the training time (i.e., fivefold validation) and testing time were 150.0 and 0.17 s, respectively. Furthermore, when mixed features were used as the input, the highest mean accuracy (92.3%) was obtained when the number of layers was set as 4, the filter size was set as 1×5 , the pooling type was set as max, the hidden size was set as 256, and the batch size was set as 64. The training time and testing time were 73.6 and 0.08 s, respectively.

Table 3. Hyperparameters associated with the top two CNN model classification performances for different input types under the eyes-closed and eyes-open conditions.

Eyes-Closed Condition											
# of Layer	Filter Size	Pooling Type	Hidden Size	Batch Size	ACC (%)	SPEC (%)	SEN (%)	PRE (%)	F1 (%)	Training Time (s)	Test Time (s)
• Raw EEG:											
5	3×3	Avg	512	32	97.8 ± 1.13	98.2 ± 1.97	97.4 ± 1.64	98.2 ± 1.87	97.7 ± 1.13	150.0 ± 15.33	0.17 ± 0.004
		Max	256	64	97.7 ± 0.96	97.3 ± 1.88	98.2 ± 1.22	97.3 ± 1.72	97.7 ± 0.95	111.8 ± 14.24	0.17 ± 0.003
• Mixed Features:											
4	1×5	Max	256	64	92.3 ± 0.83	91.9 ± 1.85	92.8 ± 1.79	90.3 ± 1.87	91.5 ± 0.88	73.6 ± 2.87	0.08 ± 0.002
5	1×3				92.2 ± 1.25	93.0 ± 1.59	91.2 ± 2.81	91.3 ± 1.69	91.2 ± 1.50	80.4 ± 2.85	0.08 ± 0.001
Eyes-Opened Condition											
# of Layer	Filter Size	Pooling Type	Hidden Size	Batch Size	ACC (%)	SPEC (%)	SEN (%)	PRE (%)	F1 (%)	Training Time (s)	Test Time (s)
• Raw EEG:											
5	5×5	Max	512	64	99.4 ± 0.25	99.3 ± 0.38	99.5 ± 0.39	99.3 ± 0.38	99.4 ± 0.24	127.8 ± 8.63	0.20 ± 0.005
		Avg			99.3 ± 0.50	99.2 ± 0.64	99.4 ± 0.43	99.2 ± 0.63	99.3 ± 0.49	170.7 ± 15.84	0.20 ± 0.005
• Mixed Features:											
5	1×5	Avg	256	64	94.9 ± 0.95	95.2 ± 2.36	94.3 ± 2.92	94.2 ± 2.61	94.2 ± 1.07	125.2 ± 13.58	0.09 ± 0.002
			512	32	94.8 ± 0.71	94.9 ± 2.12	94.7 ± 2.61	93.9 ± 2.29	94.2 ± 0.82	180.8 ± 17.73	0.09 ± 0.002

Under the eyes-open condition, when raw EEG signals were used as the input, the highest mean accuracy (99.4%) was obtained when the number of layers was set as 5, the filter size was set as 5×5 , the pooling type was set as max, the hidden size was set as 512, and the batch size was set as 64. In this case, the training time and testing time were 127.8 and 0.20 s, respectively. Moreover, when mixed features were used as the input, the highest mean accuracy (94.9%) was obtained when the number of layers was set as 5, the filter size was set as 1×5 , the pooling type was set as avg, the hidden size was set as 256, and the batch size was set as 64. In this case, the training time and testing time were 125.2 and 0.09 s, respectively. To sum up, superior classification accuracy could be attained using EEG data under the eyes-open condition as inputs (see Figure 6 for the trend of accuracy and loss functions).

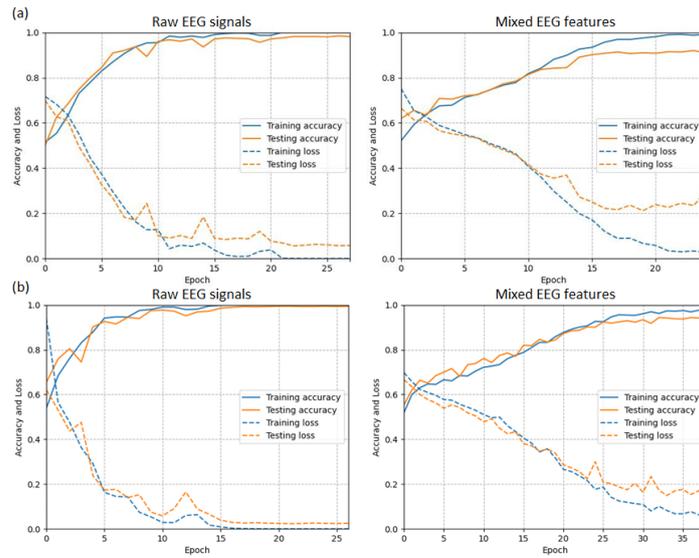


Figure 6. Accuracy and loss functions of the best-performing model when using raw EEG signals and mixed EEG features collected under the (a) eyes-closed and (b) eyes-open conditions, respectively.

3.3. Saliency Maps

Following the training of the model, we examined the averaged saliency distributions across all trials for each subject. Figure 7 shows the grand averaged saliency maps from the top-performing model among the 21 patients with MDD during the eyes-open condition. When using raw EEG signals, we observed high saliency values primarily in the right parieto-occipital region, with some in the frontal region. When using the mixed features, we noted high saliency values to some features, including the delta power around the fronto-central area (i.e., F1, Fz, F2, F4, F6, F8, FT7, FC5, FC3, FC1, FCz, FC2, FC4, FC6, FT8, T7, C5, and C1), the gamma power around the parieto-occipital area (i.e., POz, PO8, Oz, and O2), the alpha asymmetry around the frontal (i.e., FP2-FP1, AF4-AF3, F8-F7, F6-F5, and F4-F3), frontal-temporal (i.e., FT8-FT7), fronto-central (i.e., FC6-FC5 and FC4-FC3), temporal (i.e., T8-T7), central (i.e., C4-C3), temporal-parietal (i.e., TP8-TP7), and parietal (i.e., P8-P7) areas, and the delta coherence between the frontal and central areas (i.e., FP1-FC2, FP1-C2, and FP1-CP2).

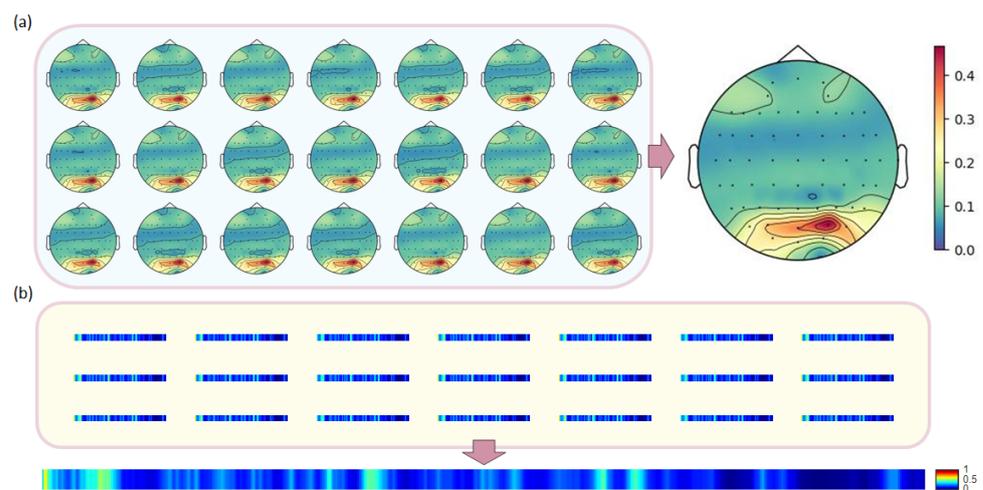


Figure 7. Averaged saliency maps of the model using (a) raw EEG signals and (b) mixed EEG features from 21 MDD patients under the eyes-open condition.

4. Discussion

CNNs are widely employed for image recognition and classification due to their ability to learn visual features [19]. Therefore, in health applications involving physiological information, 2D inputs are typically used. In this study, we chose 2D representations of raw EEG signals, the most common type of input, because they are easy to prepare after data collection. Furthermore, in machine learning, feature selection is a critical part of the classification process and often requires considerable manual procedures for identifying appropriate features among candidates [20]. Therefore, we chose 1D mixed features as an alternative input to evaluate the efficacy of automatic feature selection in the deep learning algorithm.

We observed that when using mixed features as model input, the classification accuracy increased slightly as the number of layers increased, whereas when using raw EEG signals as input, the accuracy significantly increased, highest with five convolutional layers, which agrees with the results of Khan et al. [21]. Because the number of layers used in a CNN model is related to the data complexity, more layers might achieve better classification performance [22]. However, the trade-off between accuracy and computational time should be considered during model construction. In that case, we utilized the Grad-CAM technique to compare gradient variances across different layers (Figure 8). The fourth stage exhibited the highest variances, signifying more complex information that warrants further analysis [18]. Hence, our results demonstrated that data with prior feature extraction could be classified using a shallower convolutional layer model, achieving good classification performance, whereas when dealing with raw data, a deeper layer model is preferable for effective classification.

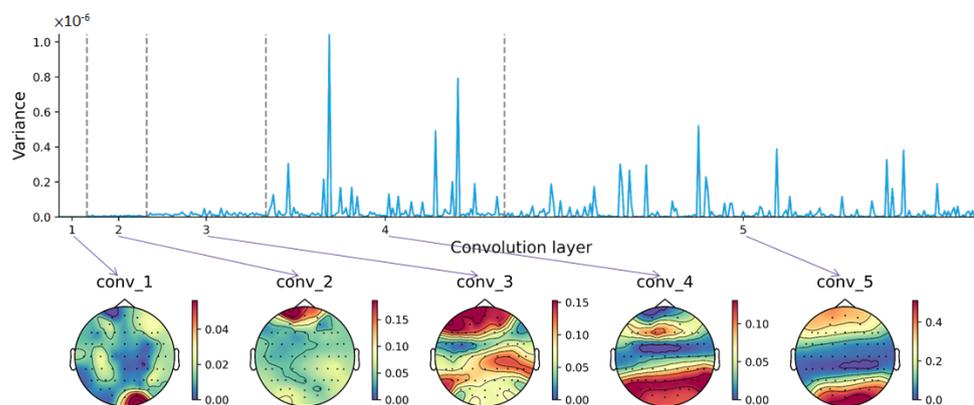


Figure 8. Variances of the gradients for each feature map and Grad-CAM topographic maps at different stages of CNN using raw EEG signals from 21 MDD patients under the eyes-open condition.

When using mixed features as model input, we observed that the accuracy increased slightly as the filter size increased. Generally, kernel size is the receptive field in a CNN [23]. A large kernel size may help extract coarse-grained feature information, whereas a small kernel size may help extract fine-grained feature information. Sharma et al. [24] proposed a novel EEG-based hybrid neural network, called the depression hybrid neural network, for screening depression. They tested different filter sizes (i.e., 4, 5, 6, and 7) and observed that the classification performance improved when the filter size increased. However, they also noted that filters with sizes larger than five did not further increase the accuracy due to the excessive parameters or over-smoothing. Since the difference in the filter sizes in our CNN model using raw EEG signals was only 8 ms, it had no significant effect on the classification of depression; however, when using extracted features, a larger filter size is recommended.

In this study, the classification accuracy increased when the pooling type changed from avg to max. In general, average pooling enables the extraction of smooth features, whereas max pooling enables the extraction of more pronounced features; therefore, most researchers adopt max pooling. For example, Bera and Shrivastava [25] analyzed the performance of five pooling strategies in CNN models: max pooling, average pooling,

stochastic pooling, rank-based average pooling, and rank-based weighted pooling. Their results indicated that the classification accuracy of models with max pooling was higher than those of other models. However, the major disadvantage of max pooling is that with the exception of the largest feature, other features are completely ignored, especially when the majority of features have large magnitudes [26]. Therefore, the choice of the pooling type should be based on the input type, and our results indicated that max pooling is more effective in a CNN model using raw EEG signals than in a model employing mixed features.

When mixed features were used as the model input, we observed that the accuracy increased as the hidden size increased. Kamhi et al. [27] optimized the robustness of CNNs for classifying motor imagery EEG data by varying the number of nodes from 32 to 1024 in the first three dense layers. They found that the use of more than 300 dense units improved the classification performance. In addition, Bakhtyari and Mirzaie [28] examined three dense unit settings (i.e., 64, 128, and 256) in their second dense layer to investigate the effect of the hidden size on the classification of attention deficit hyperactivity disorder (ADHD). Their results indicated that applying 128 dense units may help reach the best accuracy in diagnosing ADHD. These differences are primarily attributable to different model structures. In this study, since the increment of the hidden size did not apparently increase the computational time, using a hidden size greater than 256 in a CNN may be an option for effective classification through EEG.

In this study, we observed that the accuracy marginally decreased as the batch size increased. Batch size is the number of samples simultaneously entered into a model. A large batch size may cause an overfitting problem, whereas a small batch size may increase the time to convergence [29]. According to the literature, the performance of a model declines as the batch size increases, and small sizes (e.g., below 256) are recommended [24,30,31]. However, when the batch size decreases below 32, the computational time increases. Therefore, the trade-off between accuracy and computational time should be considered during model construction.

Overall, in this study, using EEG under the eyes-open condition produced better classification accuracy. Variables decreased from eyes-closed to eyes-open conditions may be a possible reason [32]. Thus, optimal results for MDD classification were obtained when using raw EEG signals collected in the eyes-open condition in a CNN with five convolutional layers, a 5×5 filter, max-pooling layers, a hidden size of 512, and a batch size of 64. An accuracy of 99.4%, a training time of 127 s, and a testing time of 0.20 s were achieved under the combination. After visualizing the extracted features exploited by the trained model, we observed that the informative brain regions were around the parieto-occipital and frontal areas, which were echoed in others' results (e.g., [33,34]). Frontal asymmetry in EEG is frequently used as a risk predictor of depression, associated with emotion. Marcu et al. [35] explored the correlation between resting-state EEG alpha asymmetry across various sites (frontal, frontolateral, and parietal) and diverse severities of depressive disorders to assess the biomarker's validity, reliability, and predictive value. Their findings suggested frontal asymmetry as a reliable marker, while parietal asymmetry was considered a potential marker for expanding the comprehension of depression's physiological foundations. Furthermore, Kaushik et al. [36] analyzed data from 40 non-clinically depressed individuals with varying depression scale levels. They noted increased EEG amplitude in the left frontal channel and decreased amplitude in the right frontal and occipital channels among individuals more susceptible to depression in resting-state EEG. Their suggestion emphasized the need to identify the most informative subset of these biomarkers to determine which ones are more effective in detecting MDD. In that case, the learned features from the trained model might provide valuable insights, prompting further avenues for exploration [37].

We ultimately compared the performance of our model with those of models reported by other previous similar studies on depression detection using EEG data (Table 4). Acharya et al. [10] used a 13-layer CNN model with a filter size of 5 to evaluate signals

recorded from the left (FP1-T3 channel) and right (FP2-T4 channel) hemispheres of the brain for 5 min with the eyes open and closed. They trained their model using 90% of an EEG data set through 10-fold cross-validation, achieving the highest accuracy of 93.5%. Ay et al. [38] evaluated the same data set by using an 11-layer CNN–long short-term memory (LSTM) model with specific filters. They trained their model by performing random splitting tests (i.e., 70%, 15%, and 15% of the EEG data set were used for training, validation, and testing, respectively), achieving the highest accuracy of 97.7%. On average, their training time was 52.13 s. Moreover, Sharma et al. [24] evaluated the same data set as that evaluated by Cavanagh et al. [16] by using 4–6-layer CNN–LSTM models with specific filters. They trained their model by using random splitting tests (i.e., 70%, 20%, and 10% of the EEG data were used for training, validation, and testing, respectively), achieving the highest accuracy of 99.1%. Their training time for a single epoch was 112 s. Among those major results in the literature, our combination demonstrated excellent performance, achieving an accuracy of 99.4% with a training time of 3.03 s for an epoch. The confusion matrix obtained during the model testing for each participant is shown in Figure 9. Overall, 1.21% of normal EEG signals were incorrectly classified as depressive, while 0.24% of depressive signals were mistakenly categorized as normal. Due to the greater risks associated with false negatives, most applications were designed to minimize their occurrence.

Table 4. Accuracy rates of different depression detection systems based on CNNs with EEG signals.

Authors	Year	# of Subjects	Model	Accuracy	TN FN	FP TP
Acharya et al. [10]	2018	15 Normal, 15 Depressed	CNN	93.5%	2055 175	104 1984
Ay et al. [38]	2019	15 Normal, 15 Depressed	CNN-LSTM	97.7%	2066 38	64 2382
Sharma et al. [24]	2021	24 Normal, 21 Depressed	CNN-LSTM	99.1%	452 4	5 539
This study	2023	21 Normal, 21 Depressed	CNN	99.4%	407 1	5 411

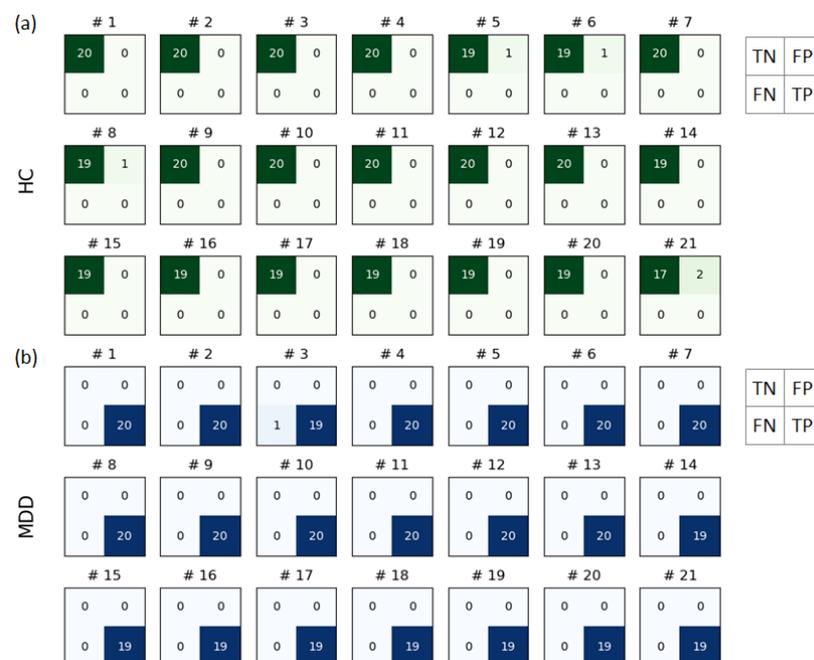


Figure 9. Confusion matrix of each participant in the (a) HC and (b) patient groups during model testing for EEG signals. The four quadrants are TP (True Positive), FP (False Positive), FN (False Negative), and TN (True Negative).

This study has some limitations. First, we examined our classifier by using one database, which included 42 participants. Hence, our sample size may not be adequate for broader clinical applications, and the generalization of varying data sources should be considered further. Second, the five hyperparameters that we selected from the literature were tuned in a specific range, and manual tuning may not typically lead to the best-performing model. Finally, our classifier was based on a standard CNN model. Therefore, further research, e.g., incorporating time information, is warranted to evaluate a more sophisticated approach.

5. Conclusions

Many studies have employed various classification models to categorize data pertaining to a wide range of topics. However, the types of input data and hyperparameter settings have often been determined based on previous experience or studies. In addition, the effects of combinations have yet to be systematically evaluated. In this study, we evaluated the effects of various EEG input and CNN model architecture combinations on the MDD classification performance of our CNN-based classifier. When raw EEG signals were employed, optimal performance and efficiency were achieved as more convolutional layers and max pooling were used. When mixed EEG features were employed, a larger hidden layer and smaller batch size were optimal. Compared with other complex networks, our combination has a relatively small number of layers and training time but relatively high accuracy rates. Thus, high classification accuracy can be achieved using an appropriate combination in a simple model. Furthermore, direct use of raw resting-state EEG data can facilitate the processes of data collection and processing and benefit future applications. This can be useful for patients who cannot perform some examination tasks, as well as for diagnosticians unfamiliar with the extraction or selection of features. In conclusion, our findings provide crucial information for other researchers regarding the design of their models for applications.

Author Contributions: All authors contributed to the study conception and design. Material preparation and data collection were conducted by H.-M.L., while data analysis was performed by C.-Y.Y. The first draft of the manuscript was written by C.-Y.Y. and all authors commented on previous versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this study are openly available in the database at <http://predict.cs.unm.edu/downloads.php> (accessed on 14 November 2021).

Acknowledgments: This study was supported in part by the National Science and Technology Council (MOST 110-2221-E-130-002), Taiwan.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Brigitta, B. Pathophysiology of depression and mechanisms of treatment. *Dialogues Clin. Neurosci.* **2002**, *4*, 7–20. [[CrossRef](#)] [[PubMed](#)]
2. Lang, U.E.; Borgwardt, S. Molecular mechanisms of depression: Perspectives on new treatment strategies. *Cell Physiol. Biochem.* **2013**, *31*, 761–777. [[CrossRef](#)] [[PubMed](#)]
3. Sau, A.; Bhakta, I. Predicting anxiety and depression in elderly patients using machine learning technology. *Healthc. Technol. Lett.* **2017**, *4*, 238–243. [[CrossRef](#)]
4. COVID-19 Mental Disorders Collaborators. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet* **2021**, *398*, 1700–1712. [[CrossRef](#)] [[PubMed](#)]
5. Khosla, A.; Khandnor, P.; Chand, T. Automated diagnosis of depression from EEG signals using traditional and deep learning approaches: A comparative analysis. *Biocybern. Biomed. Eng.* **2022**, *42*, 108–142. [[CrossRef](#)]
6. Buchwald, A.M.; Rudick-Davis, D. The symptoms of major depression. *J. Abnorm. Psychol.* **1993**, *102*, 197–205. [[CrossRef](#)] [[PubMed](#)]
7. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, 5th ed.; American Psychiatric Association: Arlington, VA, USA, 2013.

8. Wu, C.T.; Dillon, D.G.; Hsu, H.C.; Huang, S.; Barrick, E.; Liu, Y.H. Depression detection using relative EEG power induced by emotionally positive images and a conformal kernel support vector machine. *Appl. Sci.* **2018**, *8*, 1244. [[CrossRef](#)]
9. Schwartzmann, B.; Quilty, L.C.; Dhami, P.; Uher, R.; Allen, T.A.; Kloiber, S.; Lam, R.W.; Frey, B.N.; Milev, R.; Müller, D.J.; et al. Resting-state EEG delta and alpha power predict response to cognitive behavioral therapy in depression: A Canadian biomarker integration network for depression study. *Sci. Rep.* **2023**, *13*, 8418. [[CrossRef](#)]
10. Acharya, U.R.; Oh, S.L.; Hagiwara, Y.; Tan, J.H.; Adeli, H.; Subha, D.P. Automated EEG-based screening of depression using deep convolutional neural network. *Comput. Methods Programs Biomed.* **2018**, *161*, 103–113. [[CrossRef](#)]
11. Li, X.W.; La, R.; Wang, Y.; Hu, B.; Zhang, X. A deep learning approach for mild depression recognition based on functional connectivity using electroencephalography. *Front. Neurosci.* **2020**, *14*, 192. [[CrossRef](#)]
12. Uyulan, C.; Ergüzel, T.T.; Unubol, H.; Cebi, M.; Sayar, G.H.; Asad, M.N.; Tarhan, N. Major depressive disorder classification based on different convolutional neural network models: Deep learning approach. *Clin. EEG Neurosci.* **2021**, *52*, 38–51. [[CrossRef](#)] [[PubMed](#)]
13. Cho, K.O.; Jang, H.J. Comparison of different input modalities and network structures for deep learning-based seizure detection. *Sci. Rep.* **2020**, *10*, 122. [[CrossRef](#)]
14. Deslandes, A.; Veiga, H.; Cagy, M.; Fiszman, A.; Piedade, R.; Ribeiro, P. Quantitative electroencephalography (qEEG) to discriminate primary degenerative dementia from major depressive disorder (depression). *Arq. Neuropsiquiatr.* **2004**, *62*, 44–50. [[CrossRef](#)] [[PubMed](#)]
15. Craik, A.; He, Y.; Contreras-Vidal, J.L. Deep learning for electroencephalogram (EEG) classification tasks: A review. *J. Neural Eng.* **2019**, *16*, 031001. [[CrossRef](#)] [[PubMed](#)]
16. Cavanagh, J.F.; Bismark, A.W.; Frank, M.J.; Allen, J.J.B. Multiple dissociations between comorbid depression and anxiety on reward and punishment processing: Evidence from computationally informed EEG. *Comput. Psychiatr.* **2019**, *3*, 1–17. [[CrossRef](#)] [[PubMed](#)]
17. Beck, A.T.; Ward, C.H.; Mendelson, M.; Mock, J.; Erbaugh, J. An inventory for measuring depression. *Arch. Gen. Psychiatry* **1961**, *4*, 561–571. [[CrossRef](#)] [[PubMed](#)]
18. Jiang, P.T.; Zhang, C.B.; Hou, Q.; Cheng, M.M.; Wei, Y.Y. LayerCAM: Exploring hierarchical class activation maps for localization. *IEEE Trans. Image Process* **2021**, *30*, 5875–5888. [[CrossRef](#)]
19. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaria, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [[CrossRef](#)]
20. Remeseiro, B.; Bolon-Canedo, V. A review of feature selection methods in medical applications. *Comput. Biol. Med.* **2019**, *112*, 103375. [[CrossRef](#)]
21. Khan, D.M.; Masroor, K.; Jailani, M.F.M.; Yahya, N.; Yusoff, M.Z.; Khan, S.M. Development of wavelet coherence EEG as a biomarker for diagnosis of major depressive disorder. *IEEE Sens. J.* **2022**, *22*, 4315–4325. [[CrossRef](#)]
22. Ke, H.; Cai, C.; Wang, F.; Hu, F.; Tang, J.; Shi, Y. Interpretation of frequency channel-based CNN on depression identification. *Front. Comput. Neurosci.* **2021**, *15*, 773147. [[CrossRef](#)] [[PubMed](#)]
23. Lei, Y.; Belkacem, A.N.; Wang, X.; Sha, S.; Wang, C.; Chen, C. A convolutional neural network-based diagnostic method using resting-state electroencephalograph signals for major depressive and bipolar disorders. *Biomed. Signal Process Control* **2022**, *72*, 103370. [[CrossRef](#)]
24. Sharma, G.; Parashar, A.; Joshi, A.M. DepHNN: A novel hybrid neural network for electroencephalogram (EEG)-based screening of depression. *Biomed. Signal Process Control* **2021**, *66*, 2393. [[CrossRef](#)]
25. Bera, S.; Shrivastava, V.K. Effect of pooling strategy on convolutional neural network for classification of hyperspectral remote sensing images. *IET Image Process.* **2020**, *14*, 480–486. [[CrossRef](#)]
26. Zafar, A.; Aamir, M.; Nawari, N.M.; Arshad, A.; Riaz, S.; Alruban, A.; Dutta, A.K.; Almotairi, S. A comparison of pooling methods for convolutional neural networks. *Appl. Sci.* **2022**, *12*, 8643. [[CrossRef](#)]
27. Kamhi, S.; Zhang, S.; Amou, M.A.; Mouhafid, M.; Javaid, I.; Ahmad, I.S.; Kader, I.A.E.; Kulsum, U. Multi-classification of motor imagery EEG signals using bayesian optimization-based average ensemble approach. *Appl. Sci.* **2022**, *12*, 5807. [[CrossRef](#)]
28. Bakhtyari, M.; Mirzaie, S. ADHD detection using dynamic connectivity patterns of EEG data and ConvLSTM with attention framework. *Biomed. Signal Process Control* **2022**, *6*, 103708. [[CrossRef](#)]
29. Chen, H.; Jin, M.; Li, Z.; Fan, C.; Li, J.; He, H. MS-MDA: Multisource marginal distribution adaptation for cross-subject and cross-session EEG emotion recognition. *Front. Neurosci.* **2021**, *15*, 778488. [[CrossRef](#)]
30. Thoduparambil, P.P.; Dominic, A.; Varghese, S.M. EEG-based deep learning model for the automatic detection of clinical depression. *Phys. Eng. Sci. Med.* **2020**, *43*, 1349–1360. [[CrossRef](#)]
31. Seal, A.; Bajpai, R.; Agnihotri, J.; Yazidi, A.; Herrera-Viedma, E.; Krejcar, O. DeprNet: A deep convolution neural network framework for detecting depression using EEG. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–13. [[CrossRef](#)]
32. Barry, R.; Clarke, A.R.; Johnstone, S.J.; Magee, C.A.; Rushby, J.A. EEG differences between eyes-closed and eyes-open resting conditions. *Clin. Neurophysiol.* **2007**, *118*, 2765–2773. [[CrossRef](#)] [[PubMed](#)]
33. Stewart, J.L.; Coan, J.A.; Towers, D.N.; Allen, J.J. Resting and task-elicited prefrontal EEG alpha asymmetry in depression: Support for the capability model. *Psychophysiology* **2014**, *51*, 446–455. [[CrossRef](#)] [[PubMed](#)]

34. Yang, J.; Zhang, Z.; Xiong, P.; Liu, X. Depression detection based on analysis of EEG signals in multi brain regions. *J. Integr. Neurosci.* **2023**, *22*, 93. [[CrossRef](#)] [[PubMed](#)]
35. Marcu, G.M.; Szekely-Copindean, R.D.; Radu, A.M.; Bucuța, M.D.; Fleacă, R.S.; Tănăsescu, C.; Roman, M.D.; Boicean, A.; Băcilă, C.I. Resting-state frontal, frontolateral, and parietal alpha asymmetry: A pilot study examining relations with depressive disorder type and severity. *Front. Psychol.* **2023**, *14*, 1087081. [[CrossRef](#)]
36. Kaushik, P.; Yang, H.; Roy, P.P.; van Vugt, M. Comparing resting state and task-based EEG using machine learning to predict vulnerability to depression in a non-clinical population. *Sci. Rep.* **2023**, *13*, 7467. [[CrossRef](#)]
37. Cerquitelli, T.; Meo, M.; Curado, M.; Skorin-Kapov, L.; Tsiropoulou, E.E. Machine learning empowered computer networks. *Comput. Netw.* **2023**, *230*, 109807. [[CrossRef](#)]
38. Ay, B.; Yildirim, O.; Talo, M.; Baloglu, U.B.; Aydin, G.; Puthankattil, S.D.; Acharya, U.R. Automated depression detection using deep representation and sequence learning with EEG signals. *J. Med. Syst.* **2019**, *43*, 205. [[CrossRef](#)]

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.