Volume 13, Issue 1, Page 19-25, 2024 https://doi.org/10.46810/tdfd.1388893 Research Article

BİNGÖL UNIVERSITY

Türk Doğa ve Fen Dergisi
Turkish Journal of Nature and Science
www.dergipark.gov.tr/tdfd

# Advanced EEG-Based Analysis for ADHD Identification Utilizing Convmixer and Continuous Wavelet Transform

**Buğra KARAKAŞ [1*], Salih Taha Alperen ÖZÇELIK[2], Hakan UYANIK[3], Hüseyin ÜZEN[4], Abdulkadir ŞENGÜR[5]**

[1] National Education Directorate Department, Elazig, Türkiye
[2] Bingol University, Engineering Faculty, Electrical-Electronics Engineering Department, Bingöl, Türkiye
[3] Munzur University, Engineering Faculty, Electrical-Electronics Engineering Department, Tunceli, Türkiye
[4] Bingol University, Engineering Faculty, Computer Engineering Department, Bingöl, Türkiye
[5] Firat University, Technology Faculty, Electrical-Electronics Engineering Department, Elazig, Türkiye

Buğra KARAKAŞ ORCID No: 0000-0002-8319-7480
Salih Taha Alperen ÖZÇELIK ORCID No: 0000-0002-7929-7542
Hakan UYANIK ORCID No: 0000-0002-6870-7569
Hüseyin ÜZEN ORCID No: 0000-0002-0998-2130
Abdulkadir ŞENGÜR ORCID No: 0000-0003-1614-2639

**Corresponding author: bugrakarakas@gmail.com*

(Received: 10.11.2023, Accepted: 7.01.2024, Online Publication: 26.03.2024)

**Keywords**
EEG,
Deep learning,
ADHD,
Continious wavelet transform

**Abstract:** Children with ADHD may experience challenges such as attention deficits, behavioral problems, educational problems, and low self-confidence. This study summarizes research aiming to evaluate the diagnosis of attention deficit hyperactivity disorder (ADHD) with electroencephalography (EEG) signals. The research used EEG data from 30 children diagnosed with ADHD and 30 healthy control groups. EEG data was first processed for noise reduction purposes and then classified using deep learning models such as ConvMixer, ResNet50, and ResNet18. The findings show that ConvMixer demonstrates high accuracy in classification. while requiring low computational resources. Additionally, the effects of different channels on the usability of EEG signals in the diagnosis of ADHD were examined, and the T8 channel was found to be particularly effective. In conclusion, the study emphasizes the effectiveness of lightweight models and underscores the significance of specific EEG channels in diagnosing ADHD using EEG signals.



# Convmixer ve SDD Kullanılarak DEHB Hastalığının EEG Sinyalleri ile Otomatik Olarak Tespit Edilmesi

**Anahtar Kelimeler**
EEG,
Derin Öğrenme,
DEHB,
Sürekli dalgacık dönüşümü

**Öz:** DEHB, çocuklarda dikkat eksikliği, davranış problemleri, eğitimle ilgili sorunlar ve düşük özgüven gibi problemler oluşturabilir. Bu çalışma, Dikkat Eksikliği Hiperaktivite Bozukluğu (DEHB) teşhisini elektroensefalografi (EEG) sinyalleriyle değerlendirmeyi hedefleyen bir araştırmayı özetlemektedir. Araştırma, 30 DEHB tanısı almış çocuk ve 30 sağlıklı kontrol grubunun EEG verilerini kullanmıştır. EEG verileri öncelikle gürültü azaltma amacıyla işlenmiş ve ardından ConvMixer, ResNet50 ve ResNet18 gibi derin öğrenme modelleri kullanılarak sınıflandırılmıştır. Bulgular, ConvMixer'in düşük hesaplama kaynaklarına ihtiyaç duyarak yüksek sınıflandırma başarısı elde ettiğini göstermektedir. Ayrıca, EEG sinyallerinin DEHB teşhisinde kullanılabilirliği konusunda farklı kanalların etkileri incelenmiş ve T8 kanalının özellikle etkili olduğu tespit edilmiştir. Bu çalışma, EEG tabanlı DEHB teşhisi için daha hafif modellerin kullanılabilirliğini ve EEG kanallarının önemini vurgulamaktadır.

## 1. INTRODUCTION

ADHD is a neurodevelopmental disorder that typically manifests in childhood, affecting at least 5 out of every 100 children today [1]. ADHD can cause problems in children such as attention deficits, behavioral problems, educational problems, and low self-confidence. Early diagnosis of ADHD helps to create a treatment plan appropriate to children's needs, improve school performance, and support social and emotional development. ADHD symptoms cause some changes in brain activity. For this reason, an electroencephalogram (EEG) may reveal some findings that show symptoms of ADHD. Since deep learning methods can automatically analyze EEG data, they can be used as an effective tool to distinguish between individuals with ADHD and healthy individuals.

Numerous studies have been conducted in the literature using artificial intelligence techniques to classify people with ADHD and healthy individuals. Tosun [2] investigated how different frequency light stimuli affected the diagnosis of ADHD. This data set was obtained by the researcher using the power spectral densities and spectral entropy values that were generated from the EEGs of the patients. The researcher created his own data set. The researcher used support vector machines (SVM) and long-short-term memory (LSTM) as classifiers. With the eyes closed, the results showed 88.88% classification accuracy in the Fp1 and F7 channels and 92.15% at rest.

Five Azure Kinect units and depth sensors were used to gather the skeletal data of children as they played a game that Lee et al. [3] designed for the purpose of screening and diagnosing ADHD in children. The child is required to follow a robot that follows a predetermined path in a game meant to screen diagnoses. The skeletal data utilized in this study were separated into two categories: "play" data, which was collected while the child was playing, and "waiting" data, which was collected when the child was waiting while the robot guided. The RNN series, bidirectional layer, and weighted cross-entropy loss function of the GRU, RNN, and LSTM algorithms were used to classify the resultant data. Out of all of these techniques, the LSTM algorithm with a weighted cross-entropy loss function and a bidirectional layer achieved 97.82% classification accuracy.

Two novel deep learning techniques for the classification of ADHD based on functional magnetic resonance imaging (fMRI) were presented by Wang et al. [4]. Convolutional neural networks and independent component analysis were employed in the first. The correlation autoencoder method was applied in the second one. Both approaches outperformed traditional approaches in terms of performance.

The gradient-weighted class activation mapping technique was used by Chen et al. [5] to visualize EEG signals in their investigation. 50 children with attention deficit hyperactivity disorder were studied; 9 girls and 41 boys provided EEG signals for the data collection. From spatial frequency anomalies in the EEGs, they were able to obtain the power spectral density. The researchers' classification accuracy was 90.29% when they used this feature as the convolutional neural network's (CNN) input.

Lee et al. [6] used deep learning and skeletal data to classify ADHD in children. Data from engaging games were accurately classified into three groups: ADHD, ADHD-RISK, and Normal. 98.15% classification accuracy was attained with the use of bi-directional LSTM and channel attention model. A major contribution to the differentiation of the ADHD-RISK class was made by the study.

Saurabh et al. utilized functional magnetic resonance imaging (fMRI) data during resting-state to diagnose ADHD. They classified ADHD using voxel data in RSN active regions, utilizing a modified BLSTM model [7]. For classification accuracy, the model scored 87.50%. An evaluation was conducted in comparison to alternative approaches.

Tang et al.[8] used the ADHD-200 database as a data set. The suggested network architecture consisted of a modified autocoding network and a binary hypothesis testing framework. The study employed binary hypothesis testing as a means of handling incomplete data, and during feature selection, the brain functional connections from the test and training data sets were combined. The purpose of the modified autocoding network was to capture more useful features. The ADHD-200 database was used for experiments, and the method's average accuracy was 99.6%.



EEG signals were used as the data set by Ahmadi [9], who proposed a computer-aided diagnosis system that can accurately diagnose ADHD. Deep convolutional neural network (CNN) architecture is used in the suggested technique. With the combination of β1, β2, and γ bands, the highest classification accuracy was obtained. It was discovered that the success rate attained was 99.46%. In this study, deep learning and an EEG signal were used to classify children with ADHD and healthy children.

A total of 121 children and 19 channels, aged 7 to 12, comprised 61 children with ADHD and 60 healthy children [10] were the data set utilized by Maniruzzzaman et al. [11]. In the study by et al., key features were chosen using the LASSO logistic regression model after optimal channels were chosen as the network architecture using two different techniques based on SVM and t-test. Consequently, the following six machine learning-based classifiers were employed: logistic regression, multilayer perceptron, k-nearest neighbor, random forest, and Gaussian process classification (GPC). By using these techniques, it was possible to identify children with ADHD from healthy children with an accuracy rate of 97.53%.

Park et al. [12] objectively detected physical aggression in children by utilizing machine learning and physical activity data from wearable sensors. An activity monitor worn three times a week by 39 individuals with and without ADHD served as the data set for the study conducted by. The random forest method was used to perform machine learning, and patterns describing times of physical aggression were examined. With an 89.3% area under the curve, 82.4% F1 score, 85.0% recall, 82.2% sensitivity, and 82.0% accuracy, the model was able to distinguish between episodes of physical aggression. The sensor's vector magnitude feature was crucial to the model's operation. This research may offer a useful method for remotely identifying and controlling child aggression.

Ghasemi et al. used machine learning to improve the accuracy of ADHD diagnosis. Event-Related Potentials (ERP) data from ADHD patients and healthy control groups were used as the data set in the study by [13]. Features for frequency bands were calculated by processing ERP signals. Seven distinct machine learning algorithms were used for the classification process. Selected features have been combined with care. Deep Learning, Logistic Regression, and Generalized Linear Modeling techniques produced the best classification results. The AUC value is greater than 0.999 and the average accuracy rate is 99.85%. The ability to differentiate ADHD from the control group was better demonstrated by high and low frequencies (Beta, Delta). A machine learning expert system that reduces ADHD misdiagnosis and aids in treatment efficacy evaluation was created in this study.

Mikolas [14] trained a linear SVM classifier with anonymized data from clinical records to identify participants with ADHD from a population presenting a variety of psychiatric conditions. With 66.1% accuracy, children and adolescents diagnosed with ADHD were distinguished from those without the disorder. SVM with single features produced accuracies with slightly different and overlapping standard deviations. Their developed method, which combined 19 features in an automatic feature selection process, produced the best results.

In this study, a deep learning-based method was developed to automatically diagnose ADHD from EEG signals. The open-access data set [10], created with the participation of boys and girls between the ages of 7 and 12, consists of the EEG signals of 61 children with ADHD and 60 healthy children. EEG signals consisting of 19 channels were divided into segments by a multi-part signal segmentation procedure, with each channel becoming a separate vector. Each EEG signal is divided into 4-second segments. A 50Hz notch filter was applied to reduce noise and fluctuation in each EEG segment. Using a continuous wavelet transform, each EEG segment was converted into an image in the frequency-time domain. Each of the 19 channels is classified separately into two classes: ADHD and healthy. In Part 2, the dataset and materials used will be given. In the third part, the findings will be mentioned, and in the fourth part, discussion and conclusions will be given.

## 2. MATERIAL AND METHOD

### 2.1. Dataset

In this research, the open-access data set obtained from [10] was used. This data set includes 30 children diagnosed with Attention Deficit Hyperactivity Disorder (ADHD) and 30 healthy controls for whom an experienced child and adolescent psychiatrist confirmed this diagnosis according to DSM-IV criteria. The group diagnosed with ADHD consists of 22 boys and 8 girls, and their average age is $9.62 \pm 1.75$ years. The healthy control group consists of 25 boys and 5 girls, and their average age is $9.85 \pm 1.77$ years. Among the children diagnosed with ADHD, 25 had combination subtypes, 3 had attention deficit subtypes, and 2 had hyperactivity subtypes.

Children diagnosed with ADHD were referred to the child and adolescent psychiatry clinic at Roozbeh Hospital, and these children had never used medication before. The control group was selected from two different sources: the first was selected from a primary school with 25 male students; the other 5 female students were selected from an all-female primary school. Children in the control group were evaluated by a child and adolescent psychiatrist to determine possible disorders. As a result of this evaluation, it was determined that none of the children in the control group had psychiatric problems.

Exclusion criteria for children diagnosed with ADHD and healthy controls included history of significant neurological disorder, brain injury (including epilepsy), history of serious medical illness, learning or verbal disability, other psychiatric disorders, and use of benzodiazepine and barbiturate medications. Additionally, after the Raven Progressive Matrices Test was administered to children, participants who showed above-average success were included in the study.

### 2.2. Preprocessing

EEG signals in the data set were recorded according to the 10-20 standard with a sampling frequency of 128 Hz and were recorded with 19 channels (Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2). Each channel was converted into a separate signal vector. 19 signal vectors were obtained for each sample. Then, a 50 Hz Notch Filter was applied to these signal vectors to minimize noise. Then, each signal vector was divided into 4-second segments without overlapping. For each channel, a total of 79287 segments were obtained, including 2330 ADHD segments and 1843 healthy segments. The pre-processing procedure performed is given in Figure 1.

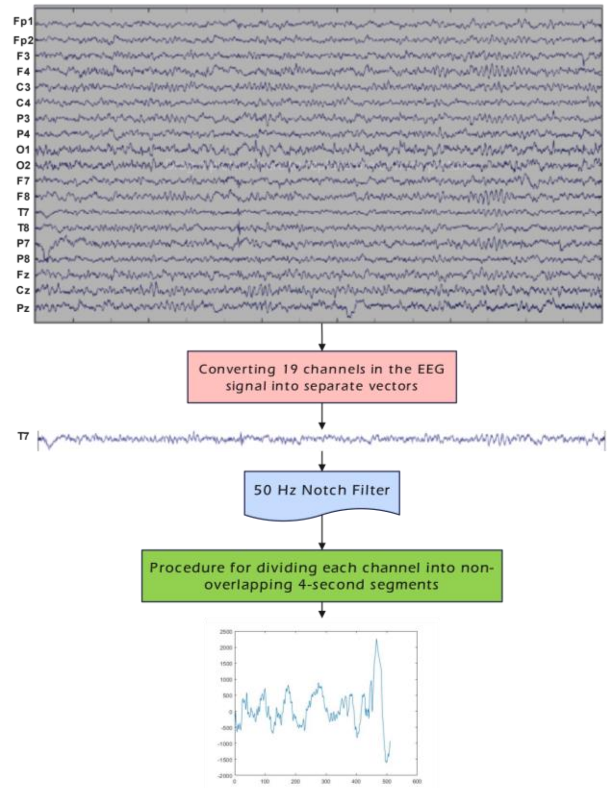


**Figure 1.** Preprocessing procedure performed

### 2.3. Continuous Wavelet Transform

Each of the 4-second segments obtained in the pre-processing section was converted into heat map images in the frequency-time domain using the continuous wavelet transform (CWT). The resulting images were set to 224x224, which is the input size of the deep learning networks to be used. CWT is a mathematical process or a spectral analysis method applied to analyze a signal in time-frequency space [15]. EEG signals examine changes in brain activity over time. Frequency components of the EEG are also of high importance. Because oscillations at different frequency levels represent different brain activities [16] . CWT analyzes EEG signals in the time-frequency domain and shows which frequency components are effective in which time interval. The basic equation of CWT is as follows:

$$CWT(a,b) = \int_{-\infty}^{\infty} x(t) \times \psi^{*}\left(\frac{t-b}{a}\right) dt \qquad (1)$$

Here a is the scale factor and determines the time scale of the signal. b is the position parameter and determines the time position of the signal. The result of CWT(a,b) is indexed by time (a) and scale (b). This represents the analysis of the signal at different time scales. x(t) represents the signal under consideration. $\psi^{*}$ is a wavelet function and is used to measure the frequency components and time positions of the signal. By performing the CWT transformation, wavelet coefficients were obtained as follows:

$$COEFS = CWT(x(t), f_s) \qquad (2)$$

Here, $COEFS$ are the wavelet coefficients. $f_s$ is the sampling frequency of the EEG device. Since the sampling frequency is given as 128 Hz in the data set, $f_s$was chosen as 128 Hz.

The obtained wavelet coefficients were multiplied by a scalar, their logarithm was taken, and then their absolute value was taken.

$$img = log|COEFS.* \, COEFS| \qquad (3)$$

Here *img* is the resulting image matrix. This matrix was visualized to scale, and a heat map image (HMI) was obtained. For each channel, a total of 79287 HMI was obtained, including 2330 ADHD HMI and 1843 healthy HMI. The stages of obtaining an HMI and an example of the obtained HMI are given in Figure 2.

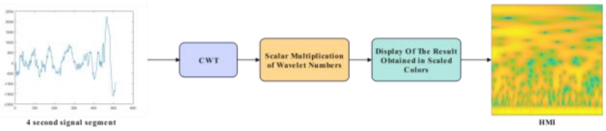


**Figure 2.** Stages of obtaining HMI

### 2.4. Classification Method

The resulting 224x224 size images were classified using Convolutional Neural Network (CNN). The flow diagram of the classification is given in Figure 3.

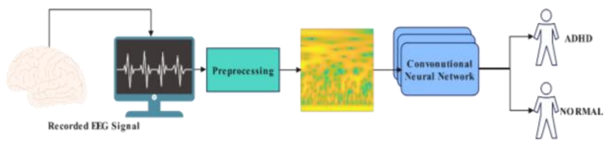


**Figure 3.** Flow diagram of the applied system

ConvMixer, ResNet50 and ResNet18 were used as classifiers. 70% of the data is reserved for training and 30% for testing. Each channel is classified separately.

#### 2.4.1. ConvMixer

ConvMixer[17] is a deep learning model very similar to MLP-Mixer[18]. In MLP-Mixer, a multilayer perceptron is used to process data in the spatial dimension and mix the channel size, while in ConvMixer, depth convolution called DepthWise is used for spatial mixing. This structure first involves a patch placement layer and is then built by multiple iterations of a simple fully convolutional block. Patch embeddings defined by p and embedding size h can be implemented as a convolution operation with input channels $c_{in}$, output channels h, kernel size p and step size p:

$$z_0 = BN(\sigma\{conv_{c_{in}} \rightarrow (X, stride = p, kernel_{size} = p)\}) \qquad (4)$$

ConvMixer combines several components to process data:

Patch Embedding Layer: Input data is processed with this layer. In this layer, the data is divided into small patch regions and each patch region is converted into a vector by embedding. The stage of converting the input data into lower-dimensional vector representatives is achieved with this layer.

Convolutional Block: ConvMixer consists of fully connected layers repeated consecutively. These blocks use the basic convolution operation. Each block consists of two steps. The first of these steps is depth convolution, where a convolution is applied where the number of groups is equal to the number of channels. It changes the overall structure of the feature map by comparing channel-level information. The second step is the point convolution step. This step allows further processing of pixel and position features.

Normalization and Activation: Normalization and activation are applied after each convolution. In this way, the model is ensured to stabilize education and learning.

Global Pooling: After a series of repetitions of ConvMixer blocks, global pooling is applied to create a general summary of the entire feature map, and a feature vector is obtained as a result of this process.

Classification Layer: In the final stage, this feature vector is passed to a softmax classification layer and classification of objects or patterns is performed through this layer.

The network architecture used is given in Figure 4.

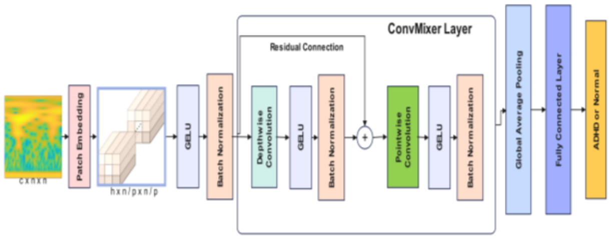


**Figure 4.** ConvMixer Network Architecture

### 2.4.2. ResNet50

ResNet [19], developed by Kaiming He and other researchers in 2015, is a family of network architectures. The main feature of ResNet is the use of so-called "skip connections" or "shortcut connections". These connections allow the network to be made deeper, making it easier to train the network and more resistant to overfitting. In this study, ResNet models were also used to classify ADHD and healthy and compare them with the ConvMixer architecture.

Input Layer: ResNet-50 model takes an image of 224x224 pixels as input. This image is usually represented as a tensor consisting of three-color channels (RGB).

Convolution Layers: ResNet-50 contains several convolution layers. Convolution layers are used to capture and extract different features of the image. Each convolution layer implements a convolution operation that comes with weight matrices (W) and bias terms (b). The mathematical expression of this process is as follows:

$$H_i = f(W_i * H_{i-1} + b_i) \quad (5)$$

Here $H_i$ represents the output of the layer, $W_i$ represents the convolution kernel, $H_{i-1}$ represents the output of the previous layer, $b_i$ represents the bias term and $f$ represents the activation function.

$$Y = f(WX + b) \quad (6)$$

Here Y is the classification result; W, weight matrix; X, feature vector; b is the bias term and f is the activation represents the function.

Skip Connections: ResNet-50 uses connections called "skip connections" or "redundant connections" that specifically help train deeper networks. These connections add the output of one layer to the input of another layer, making the flow of information smoother and allowing the network to become deeper.

### 2.4.3. ResNet18

Basically, ResNet18 has a similar structure to ResNet50. ResNet18 consists of 18 layers in total, while ResNet50 consists of 50 layers. For this reason, the depth of ResNet50 is greater than ResNet18. This will directly increase the number of parameters of ResNet50 compared to ResNet18. In terms of generalization ability, ResNet50 can generalize better. For this reason, it will work better than ResNet18 in complex situations. Being a model with a higher number of parameters indicates that ResNet50 requires more computational power and memory. In short, while ResNet50 can give better results in larger and more complex data sets, ResNet18 can be described as a simpler model that achieves better results with less data. In the study, classification was performed with both ResNet18 and ResNet50 and these two models were compared in terms of classification success.

Deep learning techniques were applied in a MATLAB environment to automate the classification of Attention Deficit Hyperactivity Disorder (ADHD) using EEG signals. The EEG data, collected from 121 children 60 diagnosed with ADHD and 61 healthy controls were processed into individual vectors for each channel to facilitate detailed analysis. A 50 Hz Notch filter was applied to clean each signal from noise. Each vector, cleared of noise after the Notch filter, was divided into 4-second segments. 4173 segments were obtained for each channel, 2330 of which were ADHD and 1843 were healthy. The resulting 4173 segments were converted into heat map images in the time-frequency domain using the SDD method. The resulting 4173 images were set to 224x224 size. Of these 4173 images, 80% were randomly divided as training data and 20% as test data. For each channel, 3378 training and 835 test images were obtained. A total of 4173*19=79287 images were obtained. In total, 64182 images were allocated to training and 15105 images were allocated to testing.

Some parameters of pre-trained models are given in Table 1.

## 3. RESULTS OF EXPERIMENTS

Analyzes were carried out in MATLAB environment to automatically classify ADHD disease with deep learning using EEG signals. EEG signals received from a total of 121 children, 60 of whom had ADHD and 61 of whom were healthy, were turned into a separate vector for each channel. A 50 Hz Notch filter was applied to clean each signal from noise. Each vector, cleared of noise after the Notch filter, was divided into 4-second segments. 4173 segments were obtained for each channel, 2330 of which were ADHD and 1843 were healthy. The resulting 4173 segments were converted into heat map images in the time-frequency domain using the SDD method. The resulting 4173 images were set to 224x224 size. Of these 4173 images, 80% were randomly divided as training data and 20% as test data. For each channel, 3378 training and 835 test images were obtained. A total of 4173*19=79287 images were obtained. In total, 64182 images were allocated to training and 15105 images were allocated to testing. Some par ameters of pre-trained models are given in Table 1.

**Table 1** Parameters for Pre-Trained Models

| Model Parameters | ConvMixer | ResNet50 | ResNet18 |
|---|---|---|---|
| **Input Image Size** | 224x224x3 | 224x224x3 | 224x224x3 |
| **Mini-batch size** | 64 | 64 | 64 |
| **Number of epochs** | 10 | 10 | 10 |
| **Initial Learning rate** | 0.001 | 0.001 | 0.001 |
| **Optimizer** | Adam | Adam | Adam |
| **Activation Function** | Softmax | Softmax | Softmax |
| **Verification Frequency** | 3 | 3 | 3 |
| **Depth** | 5 | - | - |
| **Patch Size** | 9 | - | - |

An example classification process is given in Figure 5.

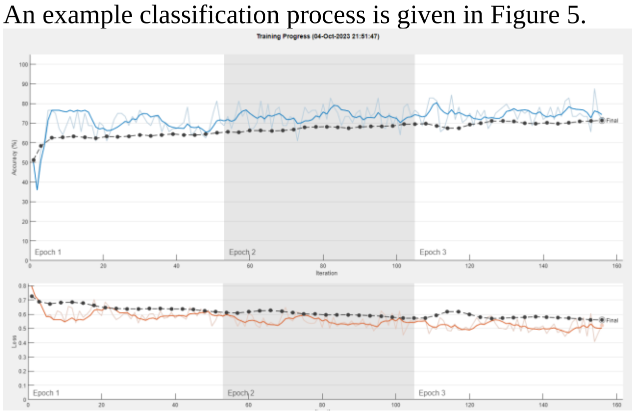

**Figure 1** An example for training process

The accuracy rate and average accuracy rate obtained for each channel using Convmixer in detecting ADHD from EEG signals are as in Table 2.

**Table 2** Achieved Classification Accuracy Rates

| | ConvMixer | ResNet50 | ResNet18 | Mean |
|---|---|---|---|---|
| **Fp1** | 73.16 | 75.33 | 72.18 | 73.5567 |
| **Fp2** | 74.37 | 73.89 | 68.82 | 72.3600 |
| **F3** | 70.9 | 75.81 | 76.02 | 74.2433 |
| **F4** | 70.3 | 68.74 | 71.46 | 70.1667 |
| **C3** | 72.22 | 77.25 | 72.18 | 73.8833 |
| **C4** | 70.18 | 68.86 | 69.3 | 69,44 |
| **P3** | 76.29 | 76.29 | 76.26 | 76.28 |
| **P4** | 65.39 | 67.19 | 65.71 | 66.0967 |
| **O1** | 71.86 | 71.5 | 72.66 | 72.0067 |
| **O2** | 65.15 | 66.23 | 67.63 | 66.3367 |
| **F7** | 73.53 | 70.9 | 73.86 | 72.7633 |
| **F8** | 73.65 | 73.53 | **76.74** | 74.64 |
| **T8** | 75.81 | **78.92** | 77.7 | **77.4767** |
| **P7** | 70.9 | 74.01 | 74.82 | 73.2433 |
| **P8** | 70.9 | 71.98 | 75.3 | 72.7267 |
| **Fz** | **78.68** | 63.47 | 64.51 | 68.8867 |
| **Cz** | 74.13 | 73.65 | 76.98 | 74.92 |
| **Pz** | 77.13 | 78.8 | 69.78 | 75.2367 |
| **Mean** | **72,475** | **72.575** | **72.383** | |

## 4. DISCUSSION AND CONCLUSION

Table 2 presents the classification results achieved for each channel using pre-trained ResNet and ConvMixer architectures. Classification performance varies for each EEG channel. ConvMixer achieved the highest classification accuracy among all channels, with 78.68% for the Fz channel, while ResNet50 attained the highest accuracy at 78.92% for the T8 channel. ResNet18 achieved the highest classification accuracy among the channels, reaching 76.74% for the F8 channel. Upon examining the classification success obtained by averaging results across 19 channels, ResNet50 exhibited the highest accuracy at 72.575%, followed closely by ConvMixer with 72.475%. ResNet18 gave the lowest classification success with 72.383%. However, all three classifiers gave an average classification accuracy of around 72%.

When their classification success was evaluated, almost three classifiers gave approximately the same classification success. However, when ResNet and ConvMixer are compared, ResNet has many more parameters than ConvMixer architectures. This means that ResNet requires more computational resources than Convmixer. ConvMixer is a lighter model and consumes less resources. Again, ResNet uses skip connections and cut connections, while ConvMixer focuses on planar comparison of features. In this study, it was revealed that ConvMixer provides features almost as good as ResNet50 with less workload and cost. In addition, the fact that it gives better results than ResNet18 shows that ConvMixer can give better results at some points, even though it has lower depth compared to ResNet architectures.

Another point of focus was which of the 19 channels arranged according to the 10-20 system while recording EEG signals could be more effective in recognizing ADHD with EEG. As stated above, ConvMixer gave high classification success in the Fz probe, ResNet50 gave high classification success in the T8 probe, and ResNet18 gave high classification success in the F8 probe. In this way, it is difficult to deduce which channel is better for ADHD detection with EEG signal. For this reason, it was estimated which probe would give the best results for automatic ADHD detection by taking the average of the 3 classifiers. As can be seen, the T8 probe was the probe that gave the highest classification success with an average of 77.4767%. This means that the region of the brain where the T8 probe is inserted may be more effective in diagnosing ADHD, and this may be a topic of discussion that may generate new ideas for expert neurologists on this subject.

The biggest disadvantage of this study is its classification accuracy, which is not high. This may be due to the complex and noisy structure of EEG signals.

**REFERENCES**

[1] Willcutt, E. G. . The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. Neurotherapeutics, 2012; 9(3), 490-499.

[2] Tosun, M. Effects of spectral features of EEG signals recorded with different channels and recording statuses on ADHD classification with deep learning. Physical and Engineering Sciences in Medicine, 2021 44(3), 693-702.

[3] Lee, W., Lee, D., Lee, S., Jun, K., & Kim, M. S. . Deep-Learning-Based ADHD Classification Using Children's Skeleton Data Acquired through the ADHD Screening Game. Sensors, 2022; 23(1), 246.

[4] Wang, D., Hong, D., & Wu, Q.. Attention deficit hyperactivity disorder classification based on deep learning. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2022; 20(2), 1581-1586.

[5] Chen, H., Song, Y., & Li, X. . Use of deep learning to detect personalized spatial-frequency abnormalities in EEGs of children with ADHD. Journal of neural engineering, 2019; 16(6), 066046.

[6] Lee, W., Lee, S., Lee, D., Jun, K., Ahn, D. H., & Kim, M. S. . Deep Learning-Based ADHD and ADHD-RISK Classification Technology through the Recognition of Children's Abnormal Behaviors during the Robot-Led ADHD Screening Game. Sensors, 2023; 23(1), 278.

[7] Saurabh, S., & Gupta, P. K.. Deep Learning-Based Modified Bidirectional LSTM Network for Classification of ADHD Disorder. Arabian Journal for Science and Engineering, 2023; 1-18.

[8] Tang, Y., Sun, J., Wang, C., Zhong, Y., Jiang, A., Liu, G., & Liu, X. . ADHD classification using auto-encoding neural network and binary hypothesis testing. *Artificial Intelligence in Medicine*, 2022; 123.

[9] Ahmadi, A., Kashefi, M., Shahrokhi, H., & Nazari, M. A. Computer aided diagnosis system using deep convolutional neural networks for ADHD subtypes. Biomedical Signal Processing and Control, 2021; 63, 102227.

[10] Ali Motie Nasrabadi, Armin Allahverdy, Mehdi Samavati, Mohammad Reza Mohammadi, June 10, 2020, "EEG data for ADHD / Control children", IEEE Dataport, doi: https://dx.doi.org/10.21227/rzfh-zn36.

[11] Maniruzzaman, M., Hasan, M. A. M., Asai, N., & Shin, J. . Optimal Channels and Features Selection Based ADHD Detection From EEG Signal Using Statistical and Machine Learning Techniques. IEEE Access, 2023;11, 33570-33583.

[12] Park, C., Rouzi, M. D., Atique, M. M. U., Finco, M. G., Mishra, R. K., Barba-Villalobos, G., ... & Najafi, B. Machine Learning-Based Aggression Detection in Children with ADHD Using Sensor-Based Physical Activity Monitoring. Sensors, 2023; 23(10), 4949.

[13] Ghasemi, E., Ebrahimi, M., & Ebrahimie, E. Machine learning models effectively distinguish attention-deficit/hyperactivity disorder using event-related potentials. Cognitive Neurodynamics, 2022; 16(6), 1335-1349.

[14] Mikolas, P., Vahid, A., Bernardoni, F., Süß, M., Martini, J., Beste, C., & Bluschke, A. Training a machine learning classifier to identify ADHD based on real-world clinical data from medical records. Scientific Reports, 2022; 12(1), 12934.

[15] Rioul, O., & Duhamel, P. Fast algorithms for discrete and continuous wavelet transforms. IEEE transactions on information theory, 1992; 38(2), 569-586.

[16] Uyanık, H., Ozcelik, S. T. A., Duranay, Z. B., Sengur, A., & Acharya, U. R. Use of differential entropy for automated emotion recognition in a virtual reality environment with EEG signals. Diagnostics, 2022; 12(10), 2508.

[17] Trockman, A., & Kolter, J. Z. Patches are all you need?. arXiv preprint arXiv:2201.09792; 2022.

[18] Tolstikhin, I. O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., ... & Dosovitskiy, A. Mlp-mixer: An all-mlp architecture for vision. Advances in neural information processing systems, 2021; 34, 24261-24272.

[19] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016; (pp. 770-778).