Automatic Diagnosis of Attention Deficit Hyperactivity Disorder with Continuous Wavelet Transform and Convolutional Neural Network

Sinan Altun¹, Ahmet Alkan¹, Hatice Altun²

Departments of ¹Electrical and Electronics Engineering, ²Child and Adolescent Psychiatry, Kahramanmaras Sutcu Imam University, Kahramanmaras, Turkey

Objective: The attention deficit hyperactivity disorder has a negative impact on the child’s educational life and relationships with the social environment during childhood and adolescence. The connection between temperament traits and the attention deficit hyperactivity disorder has been proven by various studies. As far as we know, there is no machine learning study to diagnose the attention deficit hyperactivity disorder in a dataset created using temperament characteristics.

Methods: Machine learning-based semi-automatic/fully automatic expert decision support systems are frequently used for the diagnosis of various diseases. In this study, it was aimed to reveal the success of a semi-automatic expert decision support system in the diagnosis of attention deficit hyperactivity disorder by using temperament characteristics. The high classification success achieved is a resource for a potential diagnosis of attention deficit hyperactivity disorder expert decision support system. In this respect, this study includes original qualities and innovations.

Results: Many different deep learning methods were used in the research. Deep learning methods are models that achieve high success by using a large number of images in various image processing competitions. The images of the signals in the data set were first obtained by Continuous Wavelet Transform. The highest classification success in our data set was obtained with the Squeeze Net model with 88.33%.

Conclusion: The model we propose shows that an automatic system based on artificial intelligence can be created, as well as revealing the relationship between temperament characteristics in the diagnosis of attention deficit hyperactivity in the data set we created.

KEY WORDS: ADHD; Deep learning; Diagnostic imaging; Computer assisted diagnosis.

INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental disorder that is common in childhood, directly affects social life, social relations and communication, is characterized by symptoms of hyperactivity, impulsivity and/or inattention, and its symptoms can continue in adolescence and adulthood [1]. The average worldwide rate of ADHD is 8% to 12% [2]. Two recent comprehensive analysis studies found that the mean ADHD values worldwide were 5.29% [3] and 5.9 – 7.1% [4] which was similar among children worldwide. Studies have determined that 30% of children with ADHD are also affected by ADHD in their adulthood, 60% have problems in academic, work and social areas due to persistent ADHD symptoms, and the remaining 10% have severe mental illness [5,6]. In adulthood, symptoms such as forgetfulness, absent-mindedness, inability to listen, being late, difficulty in making decisions, lack of planning skills, lack of time concept, delaying every task, inability to fully commit to one’s duties, inability to complete the required work on time, taking a long time to get ready for work are observed in these individuals due to carelessness during adulthood [7]. Various guidelines [1] recommend non-medication
or drug therapy for children with ADHD [8]. However, it is important to diagnose these children correctly and quickly. Various studies have shown that temperament and personality traits of children and adolescents are associated with the development of psychopathology [9,10]. Since some symptoms of ADHD overlap with normal childhood characteristics, it is thought that knowing the individual’s temperament and personality traits will facilitate the understanding of the heterogeneous structure of ADHD [11,12]. It can be said that image processing techniques have reached advanced levels thanks to deep learning. Very successful deep learning models have been created with State-of-Art competitions using datasets containing a large number of images. These models have been tested for computer aided diagnostic systems that help diagnose different diseases and have been very successful. In this study, experimental studies were carried out for the diagnosis of ADHD by obtaining deep learning models, scalogram images of temperament features. The applied models have achieved the success they have achieved in State-of-Art competitions and different disease images here as well. It is obvious that artificial intelligence will make its presence felt in every field from production to disease diagnosis in the future. The success of this proposed model clearly demonstrates the feasibility of computer aided diagnosis systems in ADHD diagnosis.

**Literature Research**

A study was conducted using machine learning methods to determine the response of children with ADHD treated early with SPN-812. Information from 1,397 children aged 6−17 with ADHD was used. Their classification success was 75% positive predictive power, 75% sensitivity, 74% specificity [13].

A model has been proposed to predict comorbid disorders in individuals with ADHD using machine learning methods [14]. The dataset used consisted of comorbid disorders (substance use, depression, anxiety and obesity) in a sample of 34,009 ADHD patients from Swedish medical records. The success performance of the proposed model reported the area under the receiver operating characteristic (ROC) curve as 0.74.

A study has been conducted to predict the diagnosis of ADHD based on electroencephalogram (EEG) by studying the alpha, beta and theta frequency time spectra separately in deep learning [15]. The study includes information on a patient group, 13 ADHD-C, 12 ADHD-I, and 14 control groups. Classification success between 72% and 93% was achieved in the model in which they determined the deep learning architecture layers themselves. Although high classification success is achieved, to our knowledge, ADHD diagnosis based on EEG is not a common practice and its reliability is open to debate. It is not difficult to predict that there will be difficulties in practice since the EEG recording in question will be performed on the child.

Epilepsy is a neuropathy that many young and old people from all over the world suffer from. EEG signals are mainly used to check the patient’s condition. The most important part of a successful surgery is the discovery of seizures in the brain. For this reason, it can be very useful to automatically detect the seizure area before surgery. In the study, a new method based on continuous wavelet transform (CWT) and two-dimensional convolutional neural network (CNN) is proposed to predict focal and non-focal seizures. AlexNet was used to automatically classify 2D scale map images into local and non-native captures using pre-trained models InceptionV3, Inception-ResNetV2, ResNet50, and VGG16. The performance of five pre-trained models was compared and the detection results of the two-dimensional scale map were checked. Among the other four pre-training models used, the Inception V3 model achieved the best classification accuracy of 92.27%. In conclusion, pre-trained models of focal and non-focal EEG signals and 2D scaled images would be useful for neurologists to predict preoperative seizures quickly and reliably [16].

Since EEG signals carry comprehensive information about cognition skills, it was desired to classify EEG signals using EEG signals while playing games with healthy individuals to detect ADHD [17].

Since 22 children with ADHD and control methods could be applied by extracting various features by principal Component Analysis methods of EEG signals of a total of 47 children [18].

They classified 30 control and 20 ADHD children using EEG signals using multilayer artificial neural networks. At the end of the study, 92−93% success was achieved [19].

**METHODS**

Required permission was given by Kahramanmaraş Sütçü İmam University Non-Interventional Clinical Research
The Ethics Committee with the decision numbered 02 in the session numbered 2021/29 dated 7/9/2021.

The study included 60 ADHD patients and 60 control children. The test scores of these children were collected from Kahramanmaras Sutcu Imam University Faculty of Medicine, Child and Adolescent Psychiatry Clinic, by obtaining the necessary ethics committee permission. Those with ADHD were diagnosed according to the Diagnostic and Statistical Manual of Mental Disorders 5th edition test. Children with ADHD are between the ages of 8–12, 41 boys and 19 girls. The control group consisted of children between the ages of 8–12. Of these, 39 are boys and 21 are girls. The graded tests used in the dataset are as follows:

1. Sociodemographic data form
2. Conners Family Assessment Form
3. Temperament Scale for School Age Children
4. Temps-A Temperament Scale
5. Behavioral Inhibition System/Behavioral Activation System Scale
6. Wechsler Intelligence Scale for Children-Revised (WISC-R) intelligence test

Detailed information about time series signals can be obtained from images obtained by continuous wavelet transform. Obtaining very detailed information from images is carried out with CNN. In this study, it was aimed to investigate the classification success of ADHD diagnosis by combining these two methods. Figure 1 shows the block diagram of the study. As can be seen here, the preprocessing part includes the acquisition of images with CWT. Scientists are working hard to obtain images from the signs and to classify them by extracting various features from these images with CNN. The results obtained are also quite satisfactory. It is possible to use different numbers of convolution, pooling and fully coherent layers in CNN models. Looking at the literature, there are CNN deep learning models that use their own number of layers, and there are many more studies using CNN models such as AlexNet, GoogLeNet and Squeeze Net. In this context, AlexNet, GoogLeNet and Squeeze Net models were preferred, as there is no study that obtains signs from temperament characteristics and classifies ADHD and control classification with machine learning methods, according to our knowledge.

In the second part of the study, there is a summary of the related studies in the literature, and in the third part, a detailed explanation of the dataset and models used. In Chapter 4, the information obtained in the experimental studies and in Chapter 5, the recommendations given according to the information obtained from the study are included.

**Preprocessing**

**Continuous wavelet transform**

The time-frequency diagram can be obtained by analyzing the time-varying signal with CWT. In the pattern recognition method, it is very important to choose the method used for the time-frequency domain conversion. The wavelet transform is very suitable for this type of transformation. Because this conversion is a very effective method for non-stationary signals such as EEG, electrocardiogram, and electromyogram [20-22].

Time-frequency representations of signals are created. These representations are called scalograms. A scalogram is the absolute value of the CWT coefficients of a signal. In order to obtain the scalograms of the signs used in this study, the CWT filter bank was first calculated, and these processes were performed in Matlab 2019b. Precalculating the CWT filter bank in Matlab is the preferred method when obtaining CWT of many signals using the same parameters. The default wavelet used in the filter bank is the analytical Morse wavelet. You can change the time-

![Fig. 1. Block diagram of the proposed model.](image-url)
bandwidth and symmetry parameters for Morse wavelets to tailor the Morse wavelets to your needs. Equation 1 contains the generalized Morse wavelet function, which is given by default in the Matlab CWT filter bank and used in the study [16]. In the equation \( U(\alpha) \) is the Heaviside function. It is seen that the function in the equation is defined depending on two parameters that change as “\( \gamma \)” and “\( \beta \)”, is the normalization constant and is given in Equation 2.

\[
\hat{\psi}_{\beta, \gamma}(\alpha) = U(\alpha) N_{\beta, \gamma}(\alpha)^{\beta} \exp\left[-(\alpha\alpha)^{\gamma}\right] \quad (1)
\]

\[
N_{\beta, \gamma} = 2 \left( \frac{e^{\gamma}}{\beta} \right)^{\beta/\gamma} \quad (2)
\]

You can also use analytical Morlet wavelets or impact wavelets. Morlet wavelet function is given in Equation 3. Here \( c \) is the coefficient and 5 or 6 is taken [16].

\[
m(x) = \prod_1^4 \exp(\text{i}cpx)\exp\left(-\frac{x^2}{2}\right) \quad (3)
\]

For improved computational efficiency when analyzing multiple signals in time frequency, you can precompute the filters once and then pass the filter bank as input to CWT. With the filter bank you can visualize wavelets in time and frequency. You can also create filter banks with specific frequency or period ranges. You can specify the quality factor for wavelets in the filter bank. In Figure 2, there is an example sign in the data set and the scalogram image obtained by CWT transformation.

**Convolutional neural network feature extraction layers**

**Input layer:** This layer accepts three-layer data as input in the form of an image’s size (width \( \times \) height) and has a depth representing three RGB colors. The process is continued by transforming the images into a matrix according to the image size [23, 24].

**Convolutional layer:** The first layer of feature extraction is the convolutional layer. Convolutional layers are similar to multilayer neural networks. The convolutional layer does not use multiple single neurons to become a neuron, but uses a locally connected group of nodes (neurons) to transform the input data of the previous layer. Figure 3 shows a weight map obtained from the images in the data set [23, 24].

**Pooling layer:** This layer extracts many features from the image and creates high-level features. Helps reduce size and control overfitting. Maximum pooling is called subsampling and has no super parameters. This operation is used to spatially adjust the size of the input data, but does not change the depth size [23, 24].

**Fully connected layer:** It is considered as the output of

---

**Fig. 3.** An example of feature maps obtained from images.

**Fig. 2.** Example of continuous wavelet transform. ADHD, attention deficit hyperactivity disorder.
the network. It flattens the properties of the meshes and calculates the class probabilities of the higher-ordered features in the images. In fact, the operation of this layer is the operation of the artificial neural networks method. This number of layers may also vary in different models [23,24].

**AlexNet**

The architecture includes eight layers: five convolutional layers and three fully connected layers [25]. But that’s not what makes AlexNet special. These:

**ReLU:** AlexNet uses the flatness unit (ReLU) instead of the standard tanh function. The advantage of ReLU is the
training time; CNN using ReLU can achieve 25% error on the CIFAR-10 dataset, which is six times the CNN using tanh [25].

**Multiple GPUs**: In the past, the GPU still used about 3GB of memory. This is especially bad because there are 1.2 million images in the training set. AlexNet implements multi-GPU training by placing half of the model’s neurons on one GPU and the other half on another GPU. This means not only that larger models can be trained, but also that the training time can be shortened [26].

**Overlapping pooling**: Traditionally, CNN uses the outputs of adjacent groups of neurons as “pools” without overlapping. However, when the authors started overlapping, they found that the error decreased by about 0.5%, and that models with overlapping pools often found overlapping difficult. Figure 4 shows the AlexNet architecture [26].

**GoogLeNet**

It was suggested by Google researcher Christian Szegedy in 2014 and won the 2014 ILSVRC competition, thereby reducing the error rate of the top five to less than 7% [27]. Although it has more layers than AlexNet, its parameters are 12 times less. There are a total of 22 layers and about 4M parameters. This makes it faster [28,29].

Figure 5 shows the LeNet architecture. As can be seen in the figure, the Initiation Module is an important part of the architecture and the convolution and pooling layers of the module are shown in detail.

**SqueezeNet**

SqueezeNet is a convolutional network that performs better than AlexNet with 50 times fewer parameters. SqueezeNet consists of fifteen layers with five different layers, two convolution layers, three maximum pool layers, eight fire layers, one overall average pool layer and one output layer softmax, Figure 6 shows the SqueezeNet architecture [30].

SqueezeNet uses the initial module idea to design a Fire module with compression layer and expansion layer. The structure of the fire protection module is shown in Figure 7. While 1 × 1 filters are used in the compression stage, 1 × 1 and 3 × 3 dimensional filters are used in the expansion stage. First, the H × W × C input tensor is compressed and the convolution number is equal to the C/4 input tensor.
channels. After the first stage, the data will be expanded and the depth of the data is expanded to C/2 of the depth of the output tensor. Both compression and expansion stages are connected to the ReLu unit. Finally, the expanded output is incrementally summed in the depth dimension of the input tensor [31].

**Performance Calculation Methods**

Figure 8 shows how to calculate the accuracy, sensitivity, specificity, precision, and negative predictive values used to measure the performance of machine learning methods [32-34].

---

**RESULTS**

It is wrong to think that every method will be successful on all datasets, whether it is classical machine learning methods or deep learning methods. An example of this is seen in this study. As can be seen in Table 1, deep learning by image processing has been more successful here. The fact that the proposed model is more suitable for this data set proves that the success in working with deep learning images obtained with CWT is higher than classical machine learning methods.

Almost all of the machine learning studies in the literature compare the classification success of the data sets by using different methods or by changing the hyperparameters. In our study, three different deep learning models called "State of Art" were used to calculate the highest classification success. In Table 1, classification achievements and other information that can provide detailed information are presented. As can be seen here, there are differences in success between the models. Although the AlexNet model has the least success, it cannot be considered unsuccessful according to the value obtained. As can be seen from Table 1, the most suitable model for the data set used in the study was SqueezeNet. The GoogLeNet model has calculated an acceptable success. Looking at the literature, there are many studies with success rates below 75% [14].

ROC curves are often presented, especially in studies in the field of medicine. Thanks to this curve, the success of the classification can be examined in detail. The area under the ideal ROC curve (AUC) should be close to 1. The higher the success, the greater the field will be. Figure 9 shows the ROC curves obtained with 3 different models used in the study. The line shown in blue is the curve obtained as a result of the classification. As can be seen from the X and Y coordinates, the True positive rate and the

---

**Table 1. Performances of methods**

| Method           | Accuracy | False positive rate | Precision |
|------------------|----------|---------------------|-----------|
| Logistic regression | 79%      | 0.208               | 0.796     |
| kNN              | 78%      | 0.199               | 0.783     |
| SVM              | 80%      | 0.191               | 0.81      |
| Naive bayes      | 76%      | 0.23                | 0.76      |
| AlexNet          | 66.7%    | 0.50                | 0.58      |
| GoogLeNet        | 75%      | 0.24                | 0.75      |
| SqueezeNet       | 83.33%   | 0.16                | 0.83      |

kNN, k-nearest neighbor; SVM, support vector machine.
False positive rate are effective in the formation of the curve. As the true positive ratio increases, the curve approaches the ideal value of 1. In this study, the SqueezeNet model achieved the highest success and the area under the curve was calculated as 0.83.

**DISCUSSION**

Studies using various machine learning techniques to classify ADHD are available in the literature. Table 2 shows the success of some of these studies and the success of the proposed model [13,33-36]. In addition, as far as we know, in a data set created using different tests (Sociodemographic data form, Conners Family Assessment Form, Temperament Scale for School Age Children, Temps-A Temperament Scale, Behavioral Inhibition System/Behavioral Activation System Scale, WISC-R intelligence test) that determine temperament characteristics. There is no machine learning study to diagnose ADHD. In this regard, this is another unique aspect of research. Obtaining time-frequency images of the signals obtained from these features and classifying them by extracting various features from these images further deepens the study. When Table 2 is examined, most of the studies are current studies. This shows that there is a need for new and different studies in this field using machine learning methods for the diagnosis of ADHD. We believe that the proposed model shows the value it will add to the literature, as it makes classification with deep learning from images obtained from time series signals and is more successful than current studies, as can be seen in Table 2.

ADHD is a common mental disorder in children. Children with ADHD are difficult to diagnose and treat early, especially when adequate family support is not available or there is no equal opportunity to diagnose and treat ADHD in rural areas. If ADHD is not treated, it can cause many negative behavioral problems such as anger outburst, irritability, harming the environment and yourself, as well as academic, social and work life, both in children and in adulthood.

Computer-aided diagnosis (CAD) systems are very useful in monitoring diagnosis and treatment using various
medical images, biomarkers or graded scale information. To build a CAD system, data on various machine learning methods need to be studied. The classification success obtained from these methods provides important information about the applicability of the CAD system, and the model obtained with this training enables the classification of new data in the future.

With deep learning, high successes can be obtained from biomedical image processing techniques. Here, it is aimed to calculate the classification success of ADHD diagnosis using image processing methods. In order to obtain an image, CWT analysis of the signals in the data set was performed. It is thought that automatic image conversion is more advantageous than different biomedical images (magnetic resonance imaging, EEG) obtained from children. Because while even adults may have difficulties in obtaining the images in question, it is of course more difficult to obtain these images in children.

With the Squeeze Net deep learning model, 88.33% classification success was achieved in images. This success demonstrates that temperament traits can be used to construct a CAD system for ADHD diagnosis. ADHD can be diagnosed quickly with great success with a graded test that is easy to administer and without tiring the child/family. CAD can be made into a mobile phone application. Thus, the difficulty of benefiting from health services for ADHD diagnosis in rural areas can be eliminated and equal medical care can be provided to people.

**Funding**

None.

**Conflicts of Interest**

No potential conflict of interest relevant to this article was reported.

**Author Contributions**

Conceptualization: Sinan Altun, Ahmet Alkan, Hatice Altun. Data acquisition: Sinan Altun, Hatice Altun. Formal analysis: Sinan Altun, Ahmet Alkan, Hatice Altun. Supervision: Ahmet Alkan, Hatice Altun. Writing—original draft: Sinan Altun. Writing—review & editing: Ahmet Alkan, Hatice Altun.

**ORCID**

Sinan Altun https://orcid.org/0000-0002-2356-0460

Ahmet Alkan https://orcid.org/0000-0003-0857-0764

Hatice Altun https://orcid.org/0000-0002-6802-8216

**REFERENCES**

1. NICE. Attention deficit hyperactivity disorder: diagnosis and management [Internet]. NICE; 2018 Mar 14 [cited at 2021 Sep 8]. Available from: https://www.nice.org.uk/guidance/ng87.

2. Guze SB. Book forum. Diagnostic and statistical manual of mental disorders, 4th ed. Am J Psychiatry 1995;152:1228.

3. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. Am J Psychiatry 2007;164:942-948.

4. Willcutt EG. The prevalence of DSM-IV attention-deficit/hyperactivity disorder: a meta-analytic review. Neurotherapeutics 2012;9:490-499.

5. Biederman J, Monuteaux MC, Mick E, Spencer T, Wilens TE, Silva JW, et al. Young adult outcome of attention deficit hyperactivity disorder: a controlled 10-year follow-up study. Psychol Med 2006;36:167-179.

6. Caye A, Rocha TB, Anselmi L, Murray J, Menezes AM, Barros FC, et al. Attention-deficit/hyperactivity disorder trajectories from childhood to young adulthood: evidence from a Birth Cohort supporting a late-onset syndrome. JAMA Psychiatry 2016;73:705-712.

7. Kim S, Kim MS. Deficits in verbal working memory among college students with attention-deficit/hyperactivity disorder traits: an event-related potential study. Clin Psychopharmacol Neurosci 2016:14:64-73.

8. Kim JW, Kim SY, Choi JW, Kim KM, Nam SH, Min KJ, et al. Differences in resting-state quantitative electroencephalography patterns in attention deficit/hyperactivity disorder with or without comorbid symptoms. Clin Psychopharmacol Neurosci 2017;15:138-145.

9. Rettew DC, McKee L. Temperament and its role in developmental psychopathology. Harv Rev Psychiatry 2005;13:14-27.

10. Muris P, Ollendick TH. The role of temperament in the etiology of child psychopathology. Clin Child Fam Psychol Rev 2005;8:271-289.

11. De Pauw ST, Mervielle I. Temperament, personality and developmental psychopathology: a review based on the conceptual dimensions underlying childhood traits. Child Psychiatry Hum Dev 2010;41:313-329.

12. Martel MW, Nigg JT. Child ADHD and personality/temperament traits of reactive and effortful control, resiliency, and emotionality. J Child Psychol Psychiatry 2006;47:1175-1183.

13. Faraone SV, Gomieni R, Hull JT, Busse GD, Melyan Z, O’Neal W, et al. Early response to SPN-812 (viloxazine extended-release) can predict efficacy outcome in pediatric subjects with ADHD: a machine learning post-hoc analysis of four randomized clinical trials. Psychiatry Res 2021;296:113664.

14. Faraone S, James ZY, Chen Q, Larsson H. Predicting comorbid disorders in ADHD using machine learning. Biol Psychiatry 2019;85(S1):56.
15. Ahmadi A, Kashfi M, Shahrokhi H, Nazari MA. Computer-aided diagnosis system using deep convolutional neural networks for ADHD subtypes. Biomed Signal Process Control 2021;63:102227.

16. Narin A. Detection of focal and non-focal epileptic seizure using continuous wavelet transform-based scalogram images and pre-trained deep neural networks. IRBM 2020. doi: 10.1016/j.irbm.2020.11.002. [Epub ahead of print]

17. Alchalabi AE, Shirmohammadi S, Eddin AN, Elsharnouby M. FOCUS: detecting ADHD patients by an EEG-based serious game. IEEE Trans Instrum Meas 2018;67:1512-1520.

18. Mercado-Aguirre IM, Gutierrez-Ruiz KP, Contreras-Ortiz SH. EEG feature selection for ADHD detection in children. In: The 16th International Symposium on Medical Information Processing and Analysis; Oct 3-4, 2020; Lima, Peru.

19. Mohammadi MR, Khaleghi A, Nasrabadi AM, Rafieivand S, Begol M, Zarfshan H. EEG classification of ADHD and normal children using non-linear features and neural network. Biomed Eng Lett 2016;6:66-73.

20. Faust O, Acharya UR, Adeli H, Adeli A. Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. Seizure 2015;26:56-64.

21. Yochum M, Renaud C, Jacquir S. Automatic detection of P, QRS and T patterns in 12 leads ECG signal based on CWT. Biomed Signal Process Control 2016;25:46-52.

22. Yochum M, Bakir T, Lepers R, Binczak S. Estimation of muscular fatigue under electromyostimulation using CWT. IEEE Trans Biomed Eng 2012;59:3372-3378.

23. Garg D, Verma GK. Emotion recognition in valence-arousal space from multi-channel EEG data and wavelet based deep learning framework. Proceed Comput Sci 2020;171:857-867.

24. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM 2017;60:84-90.

25. Jiang B, He J, Yang S, Fu H, Li T, Song H, et al. Fusion of machine vision technology and AlexNet-CNNs deep learning network for the detection of postharvest apple pesticide residues. Artif Intell Agric 2019;1:1-8.

26. Chen J, Wan Z, Zhang J, Li W, Chen Y, Li Y, et al. Medical image segmentation and reconstruction of prostate tumor based on 3D AlexNet. Comput Methods Programs Biomed 2021;200:105878.

27. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov R, et al. Going deeper with convolutions. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Jun 7-12, 2015; Boston, USA.

28. Ritesh, Bhagvati C. Word representations for gender classification using deep learning. Proceed Comput Sci 2018;132:614-622.

29. Jahandad, Sam SM, Kamardin K, Amir Sjarif NN, Mohamed N. Offline signature verification using deep learning convolutional neural network (CNN) architectures GoogLeNet inception-v1 and inception-v3. Proceed Comput Sci 2019;161:475-483.

30. Ucar F, Korkmaz D. COVIDiagnosis-Net: DeepBayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. Med Hypotheses 2020;140:109761.

31. Wang A, Wang M, Jiang K, Cao M, Iwahori Y. A dual neural architecture combined SqueezeNet with OctConv for LiDAR data classification. Sensors (Basel) 2019;19:4927.

32. Chiesa-Estomba CM, Echaniz O, Sistiaga Suarez JA, González-García JA, Larruscaín E, Altuna X, et al. Machine learning models for predicting facial nerve palsy in parotid gland surgery for benign tumors. J Surg Res 2021;262:57-64.

33. Tenev A, Markovska-Simoska S, Kocarev L, Pop-Jordanov J. Application of machine learning techniques for classification of ADHD and control adults. Neurosci Lett 2011;500 Suppl e31-e32.

34. Kautzky A, Vanicek T, Philippe C, Kranz GS, Wadsak W, Mitterhauser M, et al. Machine learning classification of ADHD and HC by multimodal serotonergic data. Transl Psychiatry 2020;10:104.

35. Riaz A, Asad M, Alonso E, Slabaugh G. DeepFMRI: End-to-end deep learning for functional connectivity and classification of ADHD using fMRI. J Neurosci Methods 2020;335:108506.

36. Mao Z, Su Y, Xu G, Wang X, Huang Y, Yue W, et al. Spatio-temporal deep learning method for ADHD fMRI classification. Int Sci 2019;499:1-11.