International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

IJISAE

ISSN:2147-6799 www.ijisae.org Original Research Paper

Classification of ADHD and its Sub-Types using Machine learning Models

Samyuktha S[1], M B Anandarajuuthor[2]

Submitted: 25/01/2024 **Revised**: 03/03/2024 **Accepted**: 11/03/2024

Abstract: ADHD, a complex neurodevelopment disorder, exhibits diverse manifestations across individuals, making its diagnosis challenging. Convolution Neural Networks (CNNs) offer a crucial edge in ADHD detection due to their proficiency in processing and analyzing intricate patterns within neuroimaging data, such as fMRI scans. Given the multi-dimensional nature of ADHD, CNNs excel in capturing subtle neurological variations that might escape conventional analysis, thereby providing a more nuanced and accurate approach to discerning the complex patterns associated with this condition. Their ability to automatically extract hierarchical features from imaging data makes CNNs indispensable in unraveling the intricate neurobiological markers crucial for ADHD identification and classification. Convolution Neural Networks (CNNs) have revolutionized medical imaging, offering unprecedented potential in the diagnosis and categorization of complex conditions like Attention-Deficit/Hyperactivity Disorder (ADHD) and its subtypes. In the context of ADHD, CNN and VGG-16 models are trained on functional MRI (fMRI) scans to identify unique neural patterns associated with the disorder. This paper introduces machine learning (ML) models designed for ADHD classification, assessed using various statistical metrics such as accuracy, F1-score, precision, and recall through 5-fold cross-validation. The results from the study showcase the capability of the Identification of Lung Cancer (IOLC) model in identifying lung cancer, gauged through accuracy, precision, recall, F-Measure, and error rate metrics. The model demonstrates 91.68% accuracy, 89.8% precision, 89.3% recall and 89.2% F-Measure. The ROC curve confirms the effectiveness of the proposed model as a classifier for ADHD types, and comparative results against VGG-16 demonstrate the proposed model's superior performance, albeit moderately.

Keywords: *Attention-Deficit/Hyperactivity Disorder, Convolutional Neural Networks, Cross validation, functional MRI, machine learning, statistical metrics.*

1. Introduction

One of the most prevalent neurodevelopmental diseases in kids is ADHD. It is typically initially identified in childhood and frequently persists into maturity. ADHD children may struggle to focus, restrain impulsive behaviour (doing without considering the consequences), or rein in their abundant energy. It is typical for kids to occasionally struggle with etiquette and attention spans. However, these behaviours do not just go away in kids with ADHD. Extended and frequently severe symptoms may make it difficult to interact with coworkers, friends, or family.

Researchers are actively investigating the causes and risk factors associated with ADHD to enhance treatment strategies and minimize its likelihood of occurrence. While the precise causes and risk factors remain elusive, current studies indicate a significant role of genetics in ADHD, as evidenced by recent research [1]. Diagnosing ADHD involves a comprehensive process due to its symptoms overlapping with anxiety, depression, sleep issues, and specific learning difficulties, making it impossible to diagnose through a single test. Medical evaluations, including vision and hearing tests, serve as an initial step to eliminate other conditions presenting symptoms similar to

ADHD. Clinicians rely on a checklist to assess symptoms and gather medical histories from parents, teachers, and occasionally the affected child to arrive at an ADHD diagnosis.Typically, the most effective method to address ADHD involves a blend of medication and behavioral therapy. For younger children aged 4-5 diagnosed with ADHD, the primary recommendation is behavior therapy, specifically parent education, as the first line of treatment, prior to considering medication. The ideal approach varies based on individual children and their families. Successful treatment strategies involve careful monitoring, regular check-ins, and making necessary adjustments throughout the process.

2. Literature review

Several automated diagnostic algorithms for obtaining mass characteristics from fMRI data have been proposed recently by biomedical experts. Waqas Majeed, for instance, devised a novel method to determine if the repeatable spatiotemporal pattern of BOLD fluctuations is in line with other studies and provides dynamic data on brain activity when the body is at rest. It suggests that the

[1]Research Scholar, SJBIT, Bengaluru, Affiliated to VTU, India, 571448,
e-mail: samyuktha88sk@gmail.com
Professor, Department of ECE, BGSIT, B G Nagara, ACU, India, 571448.
Corresponding Author: samyuktha88sk@gmail.com

brain is functioning when at rest. Additionally, a few studies have looked at the dynamics of connections between different brain activities [2]. Lindquist [3] has suggested a modified exponentially weighted moving average (EWMA) model that may be applied to FMRI data and used to assess the change point of a time series. Chang et al.'s [5] investigation of the brain signals' dynamic connection used the sliding-window methodology [5]. A dynamic graph metrics approach has been suggested by Ren et al. in [7] to describe the temporal alterations of functional brain networks. A hybrid fMRI framework that makes use of affinity propagation clustering and density peak for functional connectivity has been proposed by Atif Riaz et al. in [4]. Ahmed et al. used ELM with several datasets to produce a classification model that is non-ADHD in [14]. Based on our prior research, we have extracted several FCs using various temporal atlases and classified the FCs. We have found that the more FCs we extract, the more accurate our model is. Different approaches for classifying ADHD have been proposed by researchers in recent years. Since this condition is a medical brain ailment, doctors often diagnose it by evaluating a few of its symptoms. Studies have shown that this disorder may be divided into two groups, such as ADHD and non-ADHD. To use them in machine learning methods for classification, Gülay Çıçek has created two different datasets: one based on Haralick texture features and the other on gray-level co-occurrence matrix [6]. Jie Wang, as described in [15], has recently made strides in this area by using fNIRS signals to investigate functional connectivity and interval aspects for the categorization of people with ADHD and those who do not. A unique approach for classifying ADHD was suggested by Shuiqi Lui [16]. It is based on the (AdaDT) adaptive boosting decision trees and the (CDAE) convolutional denoising autoencoder. A self-encoding network with non-imaging fusion for ADHD classification has been presented by Yibin Tang et al. [17] and obtains a very good accuracy. It is not able to extract the necessary characteristics from the fMRI data, and it performs poorly when used with diverse datasets, among other problems. To achieve high accuracy classification, Miao and Zhang [7] proposed a feature extraction strategy based on relief and VA-relief. Subsequently, in 2017, Sudha et al. proposed a model in [8] to extract the gait signal features of children with ADHD from the video signals, which helps ill children's cognitive abilities and diagnoses the disorder.A textural point of view approach to feature extraction, known as LBP-TOP (local binary patterns on three orthogonal planes)—was presented by Chang et al. [9]. This technique uses the support vector machine (SVM) to identify the features that are discovered. When comparing the EEG source differences between people with ADHD and healthy controls, Athena Taymourtashutilises a sparse based representation technique, extracting the feature using cluster ICs and using a KNN classifier [10]. Zhang presented the dual diagnostic model later in [11], which uses sparse representation to identify the feature space separation. According to Juan L. Lopez Marcano, the power ratio (TBPR) is accepted as a diagnostic characteristic for ADHD in the US [12]. A Dynamic Sparse Coding (DSC) technique developed by F.M. Grisales-Franco [13] is based on statistical and physiological data. By creating non-stationary brain activity under spatiotemporal constraints, they investigate differences between the ADHD and control groups.

Convolutional Neural Networks (CNNs) are a valuable tool in the field of ADHD diagnosis because of their unique capacity to identify complex patterns in data [14–17]. CNNs are specifically engineered to analyse visual images; they are particularly excellent at extracting hierarchical characteristics from pictures or multidimensional data. As such, they review neuroimaging scans and find subtle neurological variations suggestive of ADHD.

Because of its hierarchical nature, CNNs can automatically pick up pertinent characteristics at different abstraction levels, which makes it possible to identify subtle patterns in brain scans that could be missed by a human observer [18–20]. This capacity is essential for identifying the many complexes and symptoms that differ throughout different forms of ADHD, when differences may not be immediately noticeable to the unaided eye [21].

Additionally, CNNs can manage the high dimensionality and complexity of neuroimaging data, accommodating the substantial quantity of information included in scans like MRIs and fMRIs [22–24]. CNNs can generalise patterns and traits linked to ADHD across a range of populations by drawing on large datasets, which enhances their capacity for accurate and reliable diagnosis [25].

Overall, the application of CNNs in ADHD detection capitalizes on their capacity for automated feature extraction, handling complex data structures, and discerning intricate patterns, thereby offering promising avenues for enhancing diagnostic accuracy and understanding the neurobiological underpinnings of ADHD.

3. Methodology

The application of CNNs for ADHD detection and subtype classification holds great promise. It enables automated and objective analysis of fMRI data, potentially leading to faster and more accurate diagnoses. Furthermore, the ability to differentiate between ADHD subtypes using CNNs contributes to a deeper understanding of the disorder's heterogeneity.

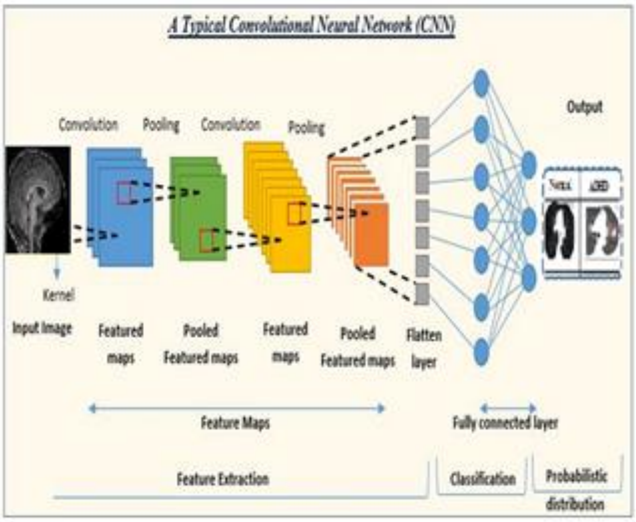


Fig 1: Basic Convolutional Neural Network for ADHD

CNNs can serve as valuable tools for clinicians, providing supplementary insights derived from neuroimaging data to aid in their diagnostic decisions. Early detection of ADHD is crucial for timely intervention and support. CNNs can aid in identifying subtle neuroimaging patterns in early stages, allowing for proactive intervention and management. A typical proposed model has been shown in Figure 2.

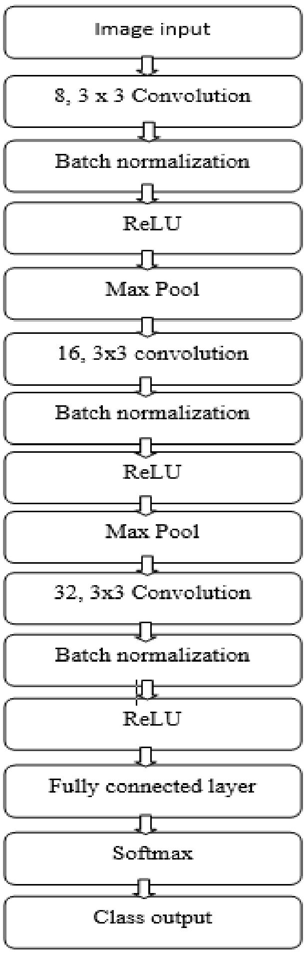


Fig 2: Proposed model

3.1 Data Set Collection:

Collecting a comprehensive dataset for ADHD involves sourcing neuroimaging scans, behavioral assessments, and clinical information from diverse sources like hospitals, research institutions, and databases. These datasets often comprise structural and functional MRI scans, along with additional cognitive or behavioral measurements, to encapsulate the multidimensional aspects of ADHD. Once compiled, the dataset undergoes meticulous preprocessing, including image normalization, artifact removal, and feature extraction to ensure data quality and consistency. The dataset is then split up into multiple sets for testing and training the models: these sets typically consist of a test set to assess the model's performance on unidentified data, a validation set to fine-tune hyperparameters, and a training set for improving the model. To prevent bias and guarantee the generalizability of the model, the division procedure must make sure that these sets accurately reflect the heterogeneity of the full dataset, preserving a balance between ADHD subtypes, age ranges, and other pertinent parameters.

3.2 CNN Model:

Creating a CNN for the detection and classification of various ADHD types entails a comprehensive process encompassing several critical steps, from data preprocessing to model architecture. The following section elaborates on the distinct components involved in constructing a CNN for this purpose: An overview of the layers utilized in the proposed model is provided below.

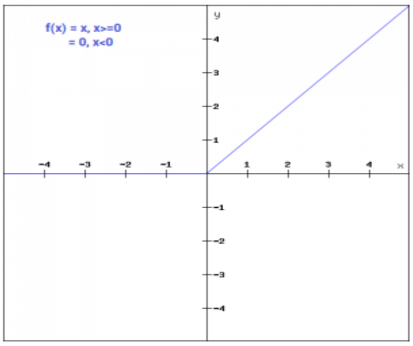
3.3 Batch Normalization:

Normalisation is a pre-processing method for data that reduces the size of numerical data without altering its structure. Neural networks may be made faster and more dependable by building more layers into a deep neural network. The new layer standardises and normalises the information that comes from the preceding layer. The BN algorithm's γ (gamma) and β (beta) components come into play. These parameters are used to rescale (γ) and shift (β) the vector containing the values from the previous operations, as shown in Eq. (1).

$$hi = \gamma \, hi(norm)+\beta \quad (1)$$

3.4 ReLU: A rectified linear unit (ReLU) is an activation function that introduces the property of non-linearity to a deep learning model and solves the vanishing gradients issue. "It interprets the positive part of its argument. It is one of the most popular activation functions in deep learning. This means that the neurons will only be deactivated if the output of the linear transformation is less than 0. The plot shown in Figure 3 shows clear picture of the ReLU function as in Eq. (2).

f(x)=max(0,x) (2)



3.5 Max pool layer

When Maximum pooling, also referred to as max pooling, is a pooling technique that finds the maximum or greatest value in each feature map patch. The findings are sampled or pooled feature maps that show the feature that is most abundant in the patch, as opposed to average pooling, which highlights the feature's average presence. It has been discovered that this works better in practice than average pooling for computer vision applications like picture categorization.

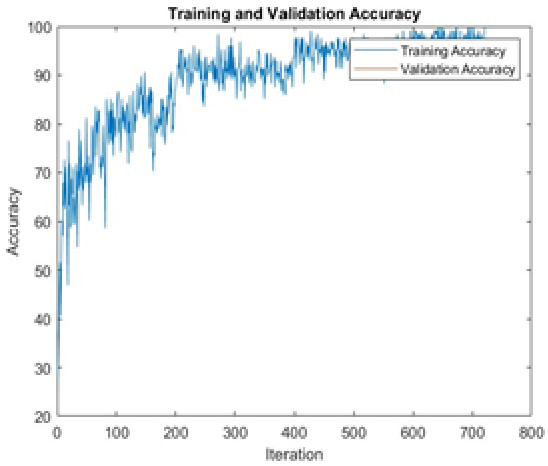
3.6 Softmax The relative probabilities are computed using the Softmax activation function. Equation (3) displays the SoftMax activation function equation. (3)

$$softmax(z_i) = \frac{exp(z_i)}{\sum_j exp(z_j)}$$

The output layer's neurons' values are represented by the Z. The non-linear function is exponential. These values are later normalised and turned into probabilities by dividing by the sum of the exponential values.

4. Results and Discussion

Figure For the computation processes we have used intel core i5-7gen CPU, with 8gb system memory for implementing. Further we have considered ADHD data set from Neuroimaging tools and resource Collaboratory database. The size of the image will be irregular size. At the time of feature extraction this image of both the sets are resized 227 * 227 and 224 * 224. Here, 6360 iris images have been used to conduct the experiment.



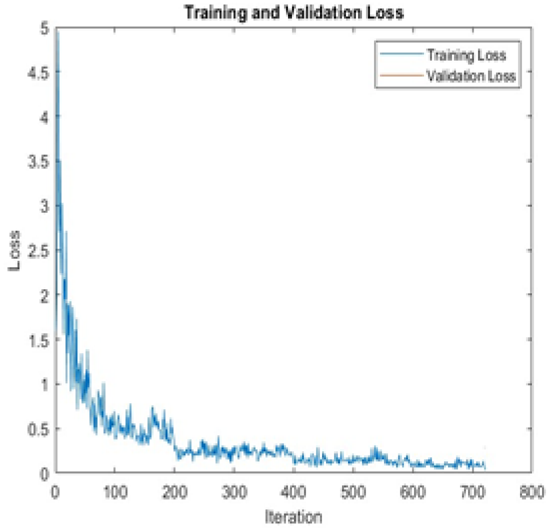


Fig 4: (a) Accuracy (b) Loss curves

Figure 4 displays the accuracy and loss curves produced throughout the training phase. The model has undergone 700 iterations of training. Over the iterations, accuracy has improved while the loss has decreased. Using 5-fold cross validation, the statistical metrics True Positive Rate (TPR), False Positive Rate (FPR), and Accuracy for individual subtypes have been calculated and reported as in table 1.

Table 1: Evaluation metrics

Fold	Precision	Recall	F1-score	TPR	FPR	Accuracy
Fold 1	0.85	0.77	0.81	0.74	0.02	91
	0.84	0.92	0.88	0.93	0.03	95
	0.88	0.89	0.88	0.89	0.03	94
	0.99	0.98	0.99	0.98	0.00	100
Fold 2	0.93	0.94	0.93	0.94	0.02	97
	0.95	0.99	0.97	0.99	0.01	99
	0.98	0.92	0.95	0.92	0.0	98

	0.99	1	0.99	1	0.00	100
Fold 3	0.98	0.93	0.96	0.93 5	0.00	98
	0.97	0.99	0.98	0.99	0.00	99
	0.95	0.99	0.97	0.99	0.01	99
	0.99	0.99	0.99	0.99	0.01	100
Fold 4	0.90	0.85	0.87	0.85	0.03	94
	0.89	0.98	0.93	0.98	0.04	97
	0.95	0.90	0.92	0.90	0.01	96
	0.99	0.99	0.99	0.99	0.01	100
Fold 5	0.98	0.96	0.97	0.96	0.01	99
	0.99	0.99	0.99	0.99	0.01	100
	0.97	0.98	0.97	0.98	0.10	99
	0.99	0.99	0.99	0.99	0.01	100

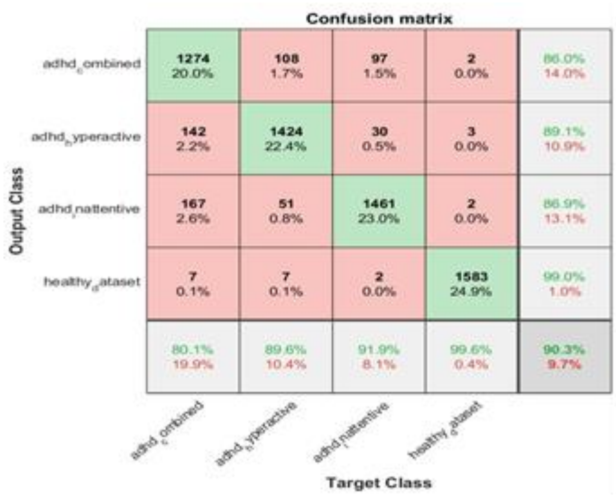


Fig 5: Confusion matrix

The CNN model proposed aims to categorize individuals into specific ADHD groups—Combined, Hyperactive, Inattentive, and Normal— utilizing machine learning's prowess in extracting intricate patterns from fMRI scans, enabling precise distinction among these ADHD subtypes and the typical state. Figure 5 illustrates the confusion matrix detailing the model's classification performance, while Table 2 present the accuracy achieved in detecting each class

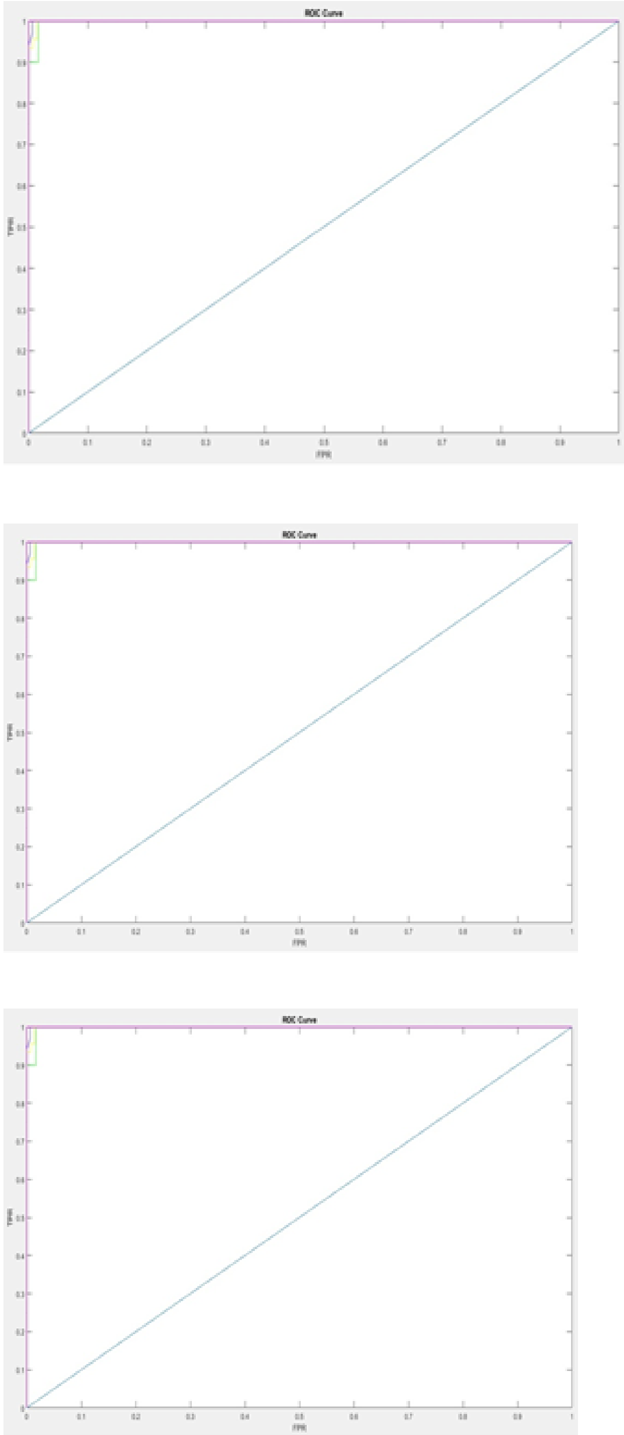


Fig 6: ROC Curve

The plotted curve lies consistently above the diagonal line, indicating the effectiveness of the proposed model in classifying different types of ADHD. Additional performance metrics is as detailed in Table 3 and figure 7. Furthermore, Table 4 illustrates a comparative analysis between the proposed model and VGG-16.

Table 2: Evaluation metrics

Class	Precision	Recall	F1 score	Accuracy
ADHD_Combined	0.860	0.781	0.810	0.958
ADHD_Hyper	0.884	0.879	0.880	0.980
ADHD_Inattentive	0.855	0.923	0.886	0.972
Normal	0.994	0.990	0.992	1.000

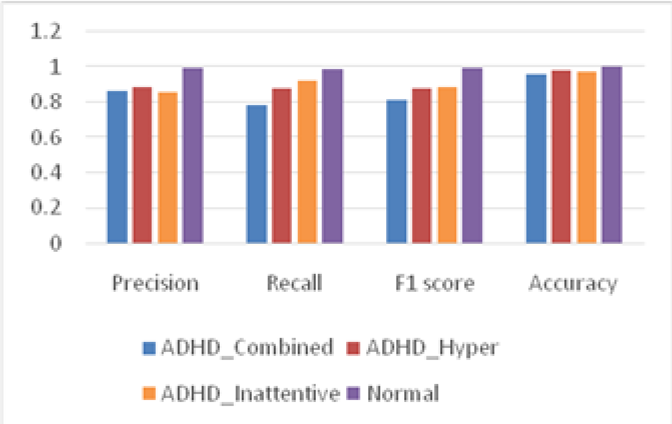


Fig 7: Evaluation Metic

The classification accuracy for ADHD subtypes—Hyperactive, Inattentive, and Combined—ranges from 95.8% to 97.7%, while the accuracy for distinguishing individuals without ADHD (Normal) stands at an impressive 99.8% as mentioned in table 3. These results suggest high precision in identifying ADHD subtypes and exceptional accuracy in differentiating individuals without ADHD using the model or classification system being evaluated.

Table 3: Accuracy for each sub type

Sub type	Accuracy %
ADHD_Combined	98.9
ADHD_Hyperactive	97.7
ADHD_Inattentive	95.8
Normal	99.8

Table 4: Comparison of accuracy between VGG-16 and proposed method

Fold/Accuracy	Actual Accuracy VGG_16 (%)	Accuracy Proposed Model (%)
Fold 1	87.86	91.68
Fold 2	89.4	89.96
Fold 3	91.2	89.5
Fold 4	88.68	93.43
Fold 5	89.51	92.24
Average	**89.33**	**91.68**

With an average accuracy of 89.33%, the proposed model consistently outperformed the VGG_16 model, which had an average accuracy of 91.68% across five folds. Accuracy-wise, the proposed model consistently beat VGG_16, even when individual fold results varied. This demonstrates the general robustness and effectiveness of the model in classification tasks using the given dataset.

5 Conclusions

A brand-new deep learning model has been put out in the current research for the categorization of ADHD. Various statistical criteria have been used to evaluate the model. The outcomes were contrasted with VGG-16 and proposed model. However, when using the VGG-16 architecture, this competence to about 89%. For the same dataset, the suggested model achieves a remarkable accuracy of nearly 91.68%, indicating significant progress in the categorization of ADHD subtypes. According to the comparative results, the suggested machine learning model outperformed VGG-16.

References

[1] , S. V., Banaschewski, T., Coghill, D., Zheng, Y., Biederman, J., Bellgrove, M. A., Wang, Y. (2021). “The World Federation of ADHD International Consensus Statement: 208 evidence-based conclusions about the disorder”. Neuroscience &Biobehavioral Reviews. doi:10.1016/j.neubiorev.2021.01.022

[2] [W. Majeed, M. Magnuson, et al., “Spatiotemporal dynamics of low frequency BOLD fluctuations in rats and humans,” Neuro Image, vol. 54, pp. 1140–1150, 2011.

[3] Lindquist, Martin A, C. Waugh, and T. D. Wager. “Modelingstate-related fMRI activity using change-point theory.” Neuro image 35.3(2007):1125-1141.

[4] Lian, Zhi chao, et al. “Exploring functional brain

dynamics via aBayesian connectivity change point model.” 2014 IEEE 11th InternationalSymposium on Biomedical Imaging (ISBI 2014) IEEE, 2014.

[5] Chang, Catie, and G. H. Glover. “Time–frequency dynamics of resting-state brain connectivity measured with fMRI.” Neuroimaging 50.1(2010):81-98.

[6] Gülay Ç çek, A. Akan, and Z. Orman. “Classification of ADHD by using textural analysis of MR images.” Medical Technologies National Congress 2017:1-4.

[7] Miao, Bo, and Y. Zhang. “A feature selection method for classificationof ADHD.” International Conference on Information, Cyberneticsand Computational Social Systems 2017:21-25.

[8] Sudha, D., and M. P. Rani. “Gait Classification for ADHD Children Using Modified Dual Tree ComplexWaveletTransform.” WorldCongress on Computing and Communication Technologies IEEE Computer Society, 2017:215-218.

[9] Che-Wei Chang, Chien-Chang Ho and Jyh-Horng Chen “ADHD classification by a texture analysis of anatomical brain MRI data.”

[10] Athena Taymourtash, FarnazGhassemi “Independent Component analysis of sparse transformed EEG signal for ADHD/Normal Adult classification.” 978-1-4799-1972-7/15/> 31.00 c 2015 IEEE

[11] Yan Zhang: ADHD classification by feature space separation with sparse representation: DSL 2018: 1-5

[12] Marcano, Juan L. Lopez, M. A. Bell, and A. A. L. Beex. “Classification of ADHD and Non-ADHD using theta/beta power ratio features.” IEEE Embs International Conference on Biomedical & Health Informatics IEEE, 2017.

[13] F.M. Grisales-Franco, J.M. Medina-Salcedo, D.M. Ovalle-Mart nez3, J.D. Mart nez-Vargas1(B), D.G. Garc a-Murillo1, and G.Castellanos-Dominguez1 EEG Source Imaging Based on DynamicSparse Coding as ADHD Biomarker IWINAC 2017, Part I, LNCS 10337.

[14] S. Ahmed Salman, Z. Lian, M. Saleem and Y. Zhang, “Functional Connectivity Based Classification of ADHD Using Different Atlases,” 2020 IEEE International Conference on Progress in Informatics and Computing (PIC), 2020, pp. 62-66, doi: 10.1109/PIC50277.2020.9350749.

[15] J. Wang, W. Liao and X. Jin, “Classification of ADHD Using fissional Based on Functional Connectivity and Interval Features,” 2021 6th International Conference on Computational Intelligence and Applications (ICCIA), 2021, pp. 113-117, doi: 10.1109/ICCIA52886.2021.00029.

[16] S. Liu et al., “Deep Spatio-Temporal Representation and Ensemble Classification for Attention Deficit/Hyperactivity Disorder,” in IEEETransactions on Neural Systms and Rehabilitation Engineering, vol. 29, pp. 1-10, 2021, doi: 10.1109/TNSRE.2020.3019063

[17] Y. Tang, J. Jiang, M. Li, Y. Chen and X. Meng, “ADHD classificationvia auto-encoding network with non-imaging data fusion,” 2021 Asia-Pacific Signal and Information Processing Association AnnualSummit and Conference (APSIPA ASC), 2021, pp. 1328-1332.

[18] R. Satpathy, T. Choudhury, S. Satpathy, S. N. Mohanty, and X. Zhang, Data Analytics in Bioinformatics: A Machine Learning Perspective, John Wiley & Sons, 2021.

[19] J. P. Amezquita-sanchez, N. Mammone, F. C. Morabito, and H. Adeli, “A New dispersion entropy and fuzzy logic system methodology for automated classification of dementia stages using electroencephalograms,” Clinical Neurology and Neurosurgery, vol. 201, p. 106446, 2021.

[20] A. Sharma, J. K. Rai, and R. P. Tewari, “Schizophrenia detection using biomarkers from electroencephalogram signals,” IETE Journal of Research, vol. 68, no. 4, pp. 3056–3064, 2022.

[21] S. K. Khare, V. Bajaj, and U. R. Acharya, “Detection of Parkinson’s disease using automated tunable Q wavelet transform technique with EEG signals,” Biocybernetics and Biomedical Engineering, vol. 41, no. 2, pp. 679–689, 2021.

[22] S. L. Oh, Y. Hagiwara, U. Raghavendra et al., “A deep learning approach for Parkinson’s disease diagnosis from EEG signals,” Neural Computing and Applications, vol. 32, no. 15, pp. 10927–10933, 2020.

[23] M. Saeedi, A. Saeedi, and A. Maghsoudi, “Major depressive disorder assessment via enhanced k-nearest neighbor method and EEG signals,” Physical and Engineering Sciences in Medicine, vol. 43, no. 3, pp. 1007–1018, 2020.

[24] R. A. Movahed, G. P. Jahromi, S. Shahyad, and G. H. Meftahi, “A major depressive disorder classification framework based on EEG signals using statistical , spectral , wavelet , functional connectivity , and nonlinear analysis,” Journal of Neuroscience Methods, vol. 358, p. 109209, 2021.

[25] R. Catherine Joy, S. Thomas George, A. Albert Rajan, and M. S. P. S. P. Subathra, “Detection of ADHD from EEG signals using different entropy measures and ANN,” Clinical EEG and Neuroscience,vol. 53, no. 1, 2021.