BRAIN DATA VISUALIZATION AND PREDICTION OF ADHD DISEASE

Himani Bansal[1], Shruti Jaiswal[*1], Sarika Agarwal[2]
[1] Department of CSE & IT, Jaypee Institute of Information Technology, Noida
[2] Department of CSE, Greater Noida Institute Of Technology, Greater Noida

Abstract

ADHD (Attention Deficit Hyperactivity Disorder) is a mental illness that impinges on 5% of children and 3% of adults in the United States. It can induce learning problems in children, which is intolerable during their growing age. This condition necessitates a thorough investigation because it is one of the most difficult to identify. Most of the time, the subtle symptoms of this condition are misunderstood. Most children are misdiagnosed, and teachers frequently comment that they are not functioning well and are not putting forth their best efforts. We as a culture find it extremely difficult and shameful to accept that someone may have a mental illness; this is the primary reason this sickness is never identified. Because of its subtle symptoms, diagnosing this sickness can be difficult. This paper is an endeavour to attain some confidence in identifying ADHD disease. A technical model is developed for the problem by understanding brain spatial arrangement using FMRI scans. ADHD prediction is made by learning and applying the features extracted through FMRI scans using unsupervised learning. The results look promising, with a decent accuracy percentage of ADHD disease prediction.

Keywords: ADHD, Deep Learning, Mental disorder, Unsupervised Learning, FMRI Scans

1. Introduction

The Central Nervous System comprises the human brain, spinal cord, and nervous system, all housed in same head. A human brain weighs approximately 2% of the total weight of a human being. The brain is the most sensitive and precious part of our body and several layers protect it. These protective layers are:

- Skull:- The hardest and the uppermost layer is the skull. The skull is made up of bones contributing to the facial structure of the vertebrates. Along with the brain cavity protection, the skull structure also decides the distance between the two eyes, nose, ears, and mouths.
- Cerebrospinal fluid is the transparent fluid present in the brain that protects the brain from external sources' shocks.
- Blood-brain barrier: This semi-permeable membrane acts as a separator between blood and other fluid.

The cerebrum, Brainstem, and cerebellum are the three primary sections of the human brain, as depicted in Figure 1. The cerebrum is the significant and front most of the brain. The primary purpose of the cerebrum includes body control, logical thinking, planning, vision, etc. The cerebrum is connected to the Brainstem.

The Brainstem is the central part of the brain, also known as the midbrain. The Brainstem is connected to the cerebellum.

Cerebellum: The cerebellum is the smallest part of the brain present at the lower backside of the brain connected to the spinal cord.

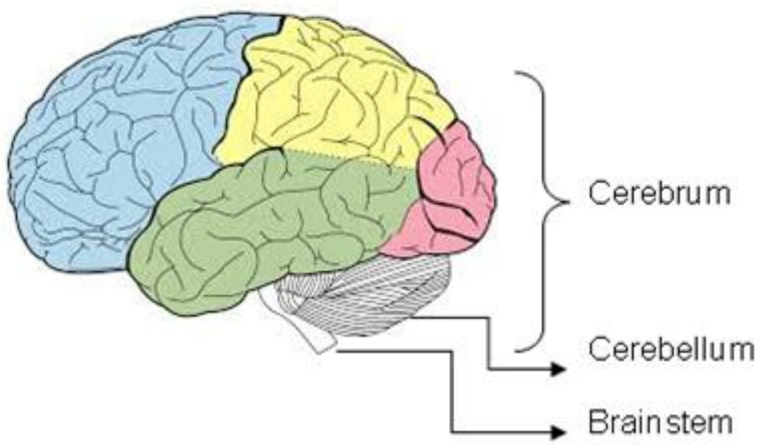


Figure 1. The Brain Structure

ADHD is a multifaceted neurological condition. Compared to the average person, a person with ADHD has a slower and distinct brain growth process. The patient's brain drops self-control and acquires a delayed reaction and comprehension. Symptoms of ADHD in humans include:

1. The individual with ADHD is unable to respond to others effectively.
2. ADHD people develop self-focused conduct, which causes them to barge in other people's activities to get noticed by them.
3. Children with ADHD lose control and have difficulty waiting for their turn in the class, games, and daily activities.
4. ADHD persons also can't adapt their behavior to the different conditions.
5. They cannot sit or stand still in a place.
6. They cannot do things quietly and in a calm environment.
7. They show interest in initiating an activity but cannot successfully finish the same task.
8. Such people cannot focus on learning, writing, playing, conversations, etc.
9. Does not want to add their own mental efforts to complete a task.
10. Shows carelessness and start visualizing thing that does not exist.

According to the findings, this illness affects 5% of children and 3% of adults. It can cause learning disabilities in youngsters, which is not acceptable at this stage of development. Because it is one of the most challenging disorders to identify, it necessitates a thorough investigation. Most of the time, the subtle symptoms of this condition are misunderstood. Most children are not correctly diagnosed, and teachers frequently comment that they are not functioning well and are not putting out their best effort. (Polanczyk et al., 2014) (Wender, 1995) (Polanczyk et al., 2007)

Magnetic Resonance Imaging (MRI) creates detailed images of various organs and tissues using radio waves and magnetic fields. Most MRI scanners take advantage of hydrogen nuclei because of their unique physical property, i.e., it only consists of one proton in its nucleus, and our tissues are also made of protons. These protons generate the signal that is further processed to create detailed images of the tissues, known as MRI scans. The signal produced measures how densely the waves are echoed back to the scanner.

The person who has to be scanned is positioned within the scanners. When the scanners are switched on, it creates a powerful magnetic field in the entire room, and then the radio waves are applied to the organs to be scanned. These radio waves taking advantage of protons, create the image representing the density of protons in that area. The image's darker part means more robust bones and hollow cavities (air) since that does not reflect the radio waves, as shown in Figure 2.

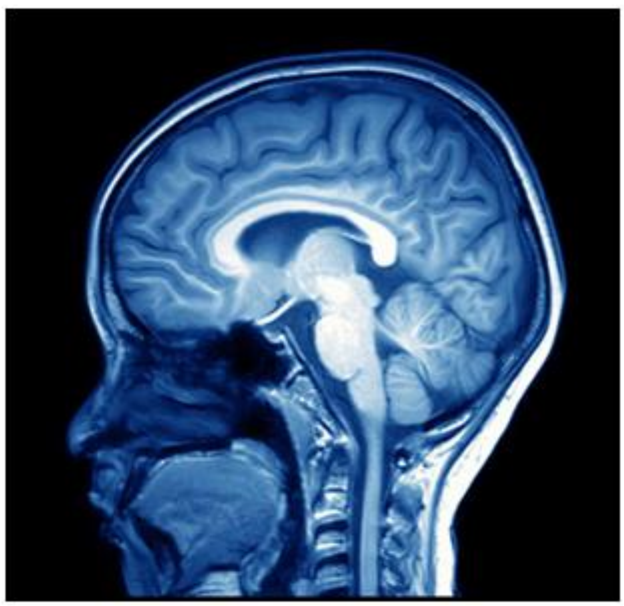


Figure 2. An Example of Brain Scan (side view)

With the ongoing breakthroughs in numerous technology domains, medicine is surely on the way. With the growing population, there is also an increase in the populace who require medical consideration. Despite advances in technology, the ratio of doctors available to patients' remains limited. The primary goal is to reduce the amount of time and money spent in laboratories evaluating ADHD samples, allowing doctors to treat additional amount of patients in the same time by catching the problem early on. We hope to collaborate to advance ADHD diagnosis and classifications via cutting-edge techniques and approaches, thereby benefiting those suffering from mental disorders.

2. Background Study

With the goal of contributing to the growth of society, it is necessary to research the brain disorder known as ADHD. Several forms of study have been conducted to assist doctors in detecting and curing this neurological condition.

The most common global datasets accessible, notably ADHD-200 and ABIDE, are used in most previous research efforts conducted by various academics. Both datasets contain high-resolution structural and functional MRI scans of people with ADHD and healthy people. This imaging dataset is coupled with phenotypic data in several research investigations, including age, gender, handedness, cognitive level, and brain scanning locations. Phenotypic data was shown to be more beneficial in developing models that could accurately detect the condition.

Because the imaging data contains high-resolution images, it is critical to reducing the dataset's resolutions, dimensionality, and linearity. To do this, practically all studies used the most common dimensionality reduction techniques. Specific photos were rotated at different angles, blurred, and sharpened to achieve higher precision. Several tactics were used to enhance or minimize noise depending on the need and training. The feature maps were then submitted to a variety of different processes, such as dataset scaling, k-fold cross-validation, averaging, and so on, depending on the model's requirements.

At this step, the feature map is loaded into machine learning models. The Support Vector Machine (SVM) model is mentioned as important models for distinguishing healthy vs. ADHD patients. Several studies have successfully fused diverse models, such as Convolutional Neural Network (CNN) + SVM (linear/kernel), employing SVM.

Another model employed was feedforward neural networks, also known as Extreme Learning Machines (ELM). The accuracy of the outcomes of various study contributions ranges from 50% to 91%. 72-76% is the average accuracy rate. With ELM mode, the greatest accuracy of around 91 percent was achieved.

Research (Qureshi et al., 2017) was carried out by neuroimaging community and is totally based on fMRI image scans. The primary method used in the study was the global connectivity feature. Maps of global connectivity were calculated from fMRI image scans; then their average was obtained for each class which was fed as the classifier's input. The classifier which they have used is a hierarchical extreme learning machine. AFNI software was used to pre-process the fMRI image data. For feature extraction of the classifier, mainly two methodologies (region of interest, and voxel-based feature extraction) are used. Both of the methodologies mentioned above are used because images of global connectivity from all the voxels were taken at first. Then, using atlas-based parcellation, the region of interest was extracted from all the images of global connectivity. Now the classifier used is HELM. It has two significant components. First, it has a multi-layer sparse autoencoder and an extreme learning machine supervised classifier. Encoding input features in sparse representation is a hierarchical and extreme learning machine's advantage over conventional classifiers. They have achieved an accuracy of 71.28% using HELM, an accuracy of 66.7% using linear support vector machine, an accuracy of 58.9% using KNN, and an accuracy of 57.1% using linear discrimination analysis.

Saeed (Saeed, 2018) developed J – Eros, a model selection strategy on training data that selects an optimal value for k for k– closest neighbors. This model demonstrated a 20% boost in accuracy over conventional techniques. High-Performance Computing (HPC) approaches will erase the computational stress from large amounts of fMRI data. With a high degree of parallelism, Graphics Processing Unit (GPU) is less expensive than other parallel computing techniques. The primary objective is to develop parallel computing methods based on GPUs that can analyse vast volumes of fMRI data.

According to the study conducted and presented in the paper **(**De Silva **et al., 2019),** generally, children who have ADHD are born with learning difficulties. In these children, the symptoms of hyperactivity can be easily observed in enthusiasm for knowledge, creativity, and curiosity. These behaviors and learning difficulties, if not handled properly, can develop into other severe disorders or a combination of the two. Some of the data types utilised to diagnose ADHD include fMRI, EEG, clinical data, and eye movement data. A mixture of these data types, on the other hand, would give more precision in the identification process than any single data type. Eye movements offer unique qualities compared to different data types, such as being less intricate and having a smaller dimensionality. But at the same time, it is not generally a go-to method because of its problems in the data collection process. As a result, researchers in this study chose to use eye movement data to solve the existing challenges in ADHD. According to their research, many studies have found a link between ADHD and eye movements. However, there is a disadvantage: eye movement can be influenced by various factors or disorders. This research aims to see if algorithms like decision trees and classification rules can be used to generate a rule-based component. For feature selection and rule development, optimized data mining methods are utilized.

Eye movement measures were incorporated in the algorithms based on the decision tree and classification rule, with an accuracy of 84 percent and 82 percent, respectively. This means that the rule-based component of both of these algorithms is very accurate.

In the research by Miao et al. (Miao et al. 2019) on ADHD, two widely used names will be there, i.e., fMRI and Fractional amplitude of low–frequency fluctuation (fALFF). fMRI can quantitatively evaluate brain activity as it analyses changes in brain activity of each voxel indirectly based on the BOLD signal. fALFF is a reliable tool for neuroimaging research in the resting state and accurately depicts spontaneous neural activity.

For accurate analysis and classification, age and gender are significant barriers. When the signal-to-noise ratio is poor, typical analytic methods cannot discover anomalies in the human brain. Machine learning and deep learning may be utilised to understand the multidimensional nature of fMRI data completely. Multi-voxel pattern analysis (MVPA), which enables decoding cognitive states from brain signals, was used to analyze the fMRI data.

Multiple linear regression is a suitable methodology for domineering variables in fMRI image classification. In the feature selection algorithm, RELIEF is an effective filter. If the sample size is small, then the feat of the RELIEF algorithm remains unstable. This issue can be resolved by promoting classification effects using the traditional RELIEF algorithm.

In this study, researchers focus on improving the accuracy of ADHD diagnosis. The performance of R–RELIEF is evaluated by comparing classification results with the RELIEF algorithm and the min. Redundancy Maximum Relevance (mRMR) method. Finally, several classifiers were used. The findings were compared to those of other contemporary approaches.

Results of the proposed method were compared to the average results of competition teams (ADHD – 200 competition, Kuang 2014, Zou 2017, and Riaz 2018) and additional cutting-edge techniques. For the KKI dataset, the accuracy of the proposed method was almost equal to that of Riaz 2018, i.e., 81.82%. For the NI dataset, the accuracy was 76%. For the NYU dataset, the accuracy was 70.73%, and for the Peking database, the accuracy was 68.63%.

In the paper by Wanf et al. (Wang et al., 2019) the previous studies, the information for feature selection is carried out at first, and then some vector-based algorithm is applied, due to which the local to the global connectivity of networks is lost. Hence this paper comes with the approach of a graph kernel-based solution to select the features from structural images. The basic idea behind this technique is to compute the similarities between networks and hence take the essential benefit of structural information. The linearly separable samples are being successfully mapped from original data to high dimensional data in many tasks. To normalize the whole topology of the network, the thresholding of the network is necessary, and this is done using a given value of t and then applying the binary function. After the thresholding is done, we move to feature learning, in which the coefficient of clustering is calculated from each region of interest. Then all are integrated to represent each subject. Based on this feature selection method and achieved vector, they applied classification. In classification, SVM is used with parameter C=1 for just patient identification, and then talking more vividly for SVM-based brain disease classification LIBSVM toolbox was provided. This proposed solution improved the performance of ADHD disease prediction and helped diagnose the brain with a more accurate understanding of brain images.

In the paper (Yao et al., 2018), the risk of other psychiatric disorders is more likely to increase in patients with ADHD. As far as the current diagnosis is concerned for ADHD patients, it depends on the clinical symptoms, which is why the misdiagnosis rate is high. fMRI is a frequently used technique in quantitative analysis of a healthy person's brain and the person with psychiatric disorders to better understand the human brain's functional networks. Functional Connectivities (FCs) have been linked to variations in resting-state fMRI between multiple brain regions. According to this study, FS RIEL is a new feature selection method based on Relative Importance and Ensemble Learning that may condense large feature spaces into a more focused subspace. Decision trees are used to calculate the relative importance of features. A forward-backward selection approach is utilised to combine features in order to improve the diversity of the new feature space while keeping the low dimensionality of FCs feature space. When they compared FS RIEL to traditional feature selection methods, they found that the former enhanced ADHD categorization by around 15% in both adults and children, with an accuracy of around 80% to 86%.

In their research (Zou et al. 2017), Zou et al. developed a method for diagnosing neurological and psychiatric illnesses such as ADHD that can be done automatically using a large number of fMRI-based characteristics. These features are then separated into voxel- and region-level features. Because of the large dimensionality of features, feature selection is usually required before classification, even if it is intuitive and simple to extract. fMRI and MRI features form a multi-modality 3D CNN model to combine features of both and are complementary. Currently, it gives 69.15% accuracy on ADHD – 200 global competition testing data.

The **accuracies** of the above-proposed method on other datasets are: -

- PekingU – 62.95%
- KKI – 72.82%
- NYU – 70.5%

In this study (Sidhu et al., 2012), the researchers looked at fMRI scans to develop an automatic way of diagnosing ADHD. They compared the diagnostic accuracy of different functional MRI data using various dimensionality reduction approaches. The all-over model is divided into two stages, i.e., training and performance. While training a model, the functional MRI scans were used to build a classifier. This classifier is used during the performance stage to perform the diagnosis on the test dataset. They have divided the dataset for ADHD into two components phenotypic data and imaging data. Age, sex, handedness, Intelligence level, and brain scanning sites are all examples of phenotypic data, while imaging data comprises functional MRI scans. The dimensionality of the dataset is reduced using feature selection/reduction techniques i.e., PCA and FFT. The dataset undergoes a 7-step dimensionality selection process.

In the first step of pre-processing, the researchers have opted for basic approaches for dimensionality reduction such as PCA and FFT. The advantages of selecting the naïve approaches are that every bit of useful information is preserved for the feature maps. The fMRI pre-processing involves six rigid body motion correction parameters followed by anatomical scans of fMRI scans. These anatomical scans are Non-linear spatial warping which is reduced to template space at 1x1x1 mm resolution. This reduction results in equal spatiotemporal dimensions. The seven-step dimensionality reduction method involves the first three steps of PCA: PCA over time dimensionality, PCA over time and space dimensionality, and Kernelized PCA and the next four steps of FFT. The dataset is converted to a matrix of the waveform of the scans as a row, from which each row is converted to a row vector. After the dimensionality reduction, the functional MRI data is normalized using a technique such as standard scaling. At this point, the dataset is diminished to spatiotemporal dimensions (57x67x50). Firstly, the functional MRI scans are used to produce a diagnostic classifier at the training stage. Then at performance time, machine learning classifiers are used to generate diagnoses that are not used to build the model at the first stage. Different filters are applied to remove noise from various sources. Finally, the data is fed as input to a linear SVM Classifier.

The following are the results for the Healthy vs. ADHA class:

- Phenotypic dataset only – 73%
- Image Dataset Only - 50 %
- 76 percent for both

In this work (Sen et al., 2018), the researchers explored the naive biological ways to utilize the combinations of the structural connectivity to determine if a person has ADHD. Structural MRI scans and functional aspects of fMRI scans were used.

Structural MRI and functional MRI are the two types of scans employed. MRI is a non-invasive method that creates a volumetric image of the brain anatomy. fMRI scans are used to track changes in blood oxygen levels due to brain activity. They employ 3-D structural MRI and 4-D resting-state fMRI to extract structural texture and functional connectivity. The sequence of three machine learning models, LeFMS, LeFMF, and LeFMSF, was examined by the researchers.

LeFM$_S$Model: The structural filters such as spares autoencoders are used to extract structural features from the structural magnetic resonance images. The unsupervised convolution network models are used to merge the features from the structural MRI with original MRIs images. These fused outputs are then served to a linear SVM classifier.

LeFM$_F$ Model: The independent non-stationary spatial components of the functional MRI images are used to generate models to diagnose ADHD. These diagnostic models are used to extract useful features automatically, and then the extracted feature set is fed as input to the SVM-Linear classifier.

LeFM$_{SF}$ Model: This model combines the features extracted from LeFM$_S$ LeFM$_F$ models, and the merged results are fed to the SVM classifier as input.

LeFM$_S$ Model: The Structural MRI scans are used to extract the valuable features, and are fed as input to the machine learning models. In this model features extracted are converted to 3D patches by removing features from 5 x 5 x 5 cubes from the scans. Then the filter encoders, such as sparse auto-encoder, perform the reductions on the extracted features to make the features more precise before serving as input to the convolutional neural network. During the CNN, the max-pooling is applied to reduce the feature set further.

LeFM$_F$Model: In this model, the blind sources separation techniques such as PCA and ICA are used to generate efficient feature maps from the functional MRI scans, which are then served as input to the SVM for further predictions among healthy and ADHD. These source separation techniques are applied to the raw data from the functional MRI scans, which are further used to generate spatial maps. Features from the functional MRI scans are extracted with the help of these spatial maps, which are finally fed to the SVM classifier.

LeFM$_{SF}$ Model: In this model, the features from the LeFM$_S$ and LeFM$_F$ are merged and serve as input to SVM linear classifier.

Results of the above models are: -

- LeFM$_S$– 0.635
- LeFM$_F$– *0.6225*
- LeFM$_{SF}$ – *0.653*

Work presented in the paper (Peng et al., 2013) proposed a method to propose an automated classification model that efficiently uses extreme learning algorithms to help detect the ADHA and the specific brain segment that suffers from ADHA. Magnetic Resonance pictures in HD format are used in this model. The researchers gathered a variety of brain measurements. The features to be supplied as input to the SVM classifier and Extreme learning algorithms are correctly selected using F-score and Sequential Forward Selection (SFS).

Movement correction, average T1-weighted imaging, Talairach space for registering the volume in the hollow cavities, and deformable template model are performed on the original MRI to pre-process the image dataset. The corpus was created by encoding the scans in different spatial dimensions and adjusting the intensity gradients. Different Surface areas and the curvature areas are calculated based on the folding method. To effectively deal with the problem of Over-fitting number of features is reduced to 340 with the help of feature selection techniques such as F-Score and Sequential Forward Selection (SFS). A non-linear vector function that is linearly separable is generated by passing the extracted features to the SVM classifier. The other classifier used was the training model is ELM algorithms such as feedforward neural networks with a hidden layer or backpropagation.

The accuracies are as follows: -

- SVM – Linear: 84.73%
- SVM – RBF: 86.55%
- ELM: 90.18%

According to the research, clinical interviews, behavioral questionnaires, and neuropsychological evaluations are popular clinical diagnosis procedures (Bledsoe et al., 2020). The reliability, sensitivity, and specificity of these procedures are all dependent on clinical interpretation. This study aimed to see how helpful machine learning was in predicting and classifying, using short neuropsychological assessments, children with ADHD–Combined presentation (ADHD-C) (d2 Test of Attention). Parents completed symptom severity questionnaires, and children with ADHD–C and normally developing control children participated in semi-structured clinician interviews and attention/concentration tests. Forward feature selection approach and a decision tree model were used to discover the most informative neuropsychological characteristics to build a rule-based model. Individual youngsters with ADHD and those that do not have it, were classified with remarkable accuracy (100%) using the SVM model (1.0). Individuals with ADHD-C and those that do not have it, were diagnosed with 100% sensitivity and specificity using decision tree algorithms.

This study revealed very precise statistical diagnostic categorization at the personage level in a sample of youngsters with ADHD-C. According to the research, data-driven behavioral algorithms based on short neuropsychological data might give clinicians a rapid and accurate diagnostic tool.

The paper focused on machine learning algorithms, such as classifiers developed for identifying adult ADHD subtypes based on EEG power spectra (Tenev et al., 2014). They looked at 117 persons, with 67 having ADHD and 50 being healthy controls. Data were collected using two resting circumstances (eyes open and closed) and two neuropsychological tests (visual continuous performance test and emotional continuous performance test). The sample is split into four different datasets, one for each of the four different situations. Each dataset is used to train four distinct SVM classifiers, with the output concatenated using a logical expression produced from the Karnaugh map. This strategy makes it easier to tell the difference between ADHD and other conditions.

Neurobehavioral disorder with a wide range of symptoms ADHD is often identified in children and teenagers, although it has lately been discovered to persist into adulthood. As per DSM-IV, the disorder defines inattention, hyperactivity, and impulsivity symptoms. Despite having a distinct name, hyperkinetic disorder(HD), the ICD-10 provides similar criteria for the illness. The disease incidence in children is between

5% and 9%. Although ADHD symptoms may improve with age, more than 50% of children diagnosed with ADHD persist to have clinically significant signs as adults. It indicates that roughly 5% of the adult population in the world is afflicted.

Most EEG studies use a combination of indicators to distinguish ADHD children from healthy controls, such as smaller alpha and beta bands and higher theta and delta bands. The few EEG studies conducted on ADHD patients were wildly divergent. The environment of the quantitative EEG parameters that are evaluated and the developmental element of the illness itself could be one of the answers. In the adult population, the disease is still diagnosed based on the doctor's abilities and knowledge.

Machine learning algorithms are used in only a few research projects to differentiate ADHD patients from control groups. Non-linear classifiers such as SVM and ANN can discover non-linear correlations in data. Mueller et al. built a machine learning system that employs an SVM classifier to identify ADHD people from control groups based on EEG readings' event-related potentials.

This research presents a model for classifying ADHD individuals and control groups based on EEG power spectra acquired under various measurement settings. The EEG power spectrum displays the distribution of the signal's squared amplitude and all frequency bands, and it is derived from EEG data collected from scalp electrodes. The data set we utilized for analysis consisted of four independent EEG readings taken in four different scenarios. The conditions included a visual continuous performance exam and an emotional continuous performance test, both of which give information on the organizational system of the patient's brain and open and closed eyes circumstances.

In their study, the researchers primarily explored how to classify ADHD using a pre-processed and feature-selected fractional amplitude of low-frequency fluctuation (fALFF) in resting-state feature subset fMRI (rs – fMRI) data (Miao & Zhang, 2017). They created a feature selection technique called the relief algorithm and verification accuracy (VA – Relief). Currently, this disease is diagnosed by the experience of a doctor and clinical diagnostic criteria. There is a high rate of misdiagnosis, which makes the children not get proper treatment in the early phase of this disease and hinders their average growth.

The paper **(**Shao et al., 2019) represents the solution to ADHD disease classification using image data with SVM integration. The basic idea was to analyze the differences between the normal controls and ADHD subjects. Two types of data were retrieved using image scans. First was 1-D data of functional connectivity, and another was 3-D data of fewer amplitude frequencies. To combine these two datasets, they have used the gcForest technique, which provides a whole new multi-functional structure using two features—using which a unique feature vector can be formed from each sample. After this, they observed some overlapping data handled using the nearest neighbor edited version. Cascade forest is used at the last stage to provide the classification input, consisting of all features combined concatenated vector.

Accuracy table obtained in different datasets in the ADHD prediction competition by different universities.

Peking dataset gave the best accuracy of 78.05% when both FC and ALFF were combined, KKI gave the best accuracy of 90.91% with ALFF alone, and NYU gave the best accuracy of 78.05% when both were combined, and NI gave the best accuracy of 80% with ALFF alone.

Calculation of all the frequencies in the brain was done by average time series. They have not used techniques like PCA's functional version strategy for fMRI scan data. Their work can also be used to detect and classify other diseases like autism, Alzheimer Ars, etc.

In the paper (Kaur et al., 2018) ADHD has a occurrence rate of 5% in children and 2.5% - 4.3% in adults diagnosed during childhood. Anxiety disorder, depression, and substance misuse are all linked with ADHD. For ADHD adults, the diagnosis involves a subjective assessment of their childhood history that entirely relies on the memory of the patient and his family members. Thus the accurate information that is needed is affected by such misreporting. The symptoms of ADHD are very similar to that of depression; that's why usually, in such cases, ADHD is unidentified.

Electroencephalography (EEG) is a special technique for diagnosing neurological problems and teaches us about the central brain nervous system because of its capacity to represent quick transient cortical processes. The Multiparadigm technique for children with ADHD using synchronization likelihood features among all electrodes and electrode pairs, by Ahmadlou and Adeli. With an accuracy of 95.6 percent, the Radial Basis Function (RBF) neural network was employed to categorize eyes-closed EEG data. From visual task-based EEG data of ADHD and control children, Allahverdy extracted non-linear metrics such as the Lyapunov exponent, Higuchi, Katz, and Sevcik fractal dimensions. Multi-layer perceptron (MLP) accuracy was 68.6 percent, 86 percent, 61 percent, 62 percent, and 55.6 percent, respectively, for all frontal, parietal, central, and occipital regions. When Ghassemi et al. used K-Nearest Neighbour (KNN) on parameters of cognitive task-based EEG in ADHD and normal people, such as correlation dimension and wavelet entropy, and Lyapunov exponent, they got a 96 percent accuracy. Many additional types of research utilized various methodologies and obtained varying degrees of accuracy.

The eigenvector approach and autoregressive (AR) modeling are used in this paper to identify ADHD adults using EEG data and a decision support system. The parameters of the AR model were extracted from the rest state EEG data of 30 ADHD and 30 control individuals using the burg, covariance, and yule-walker techniques.

The results were as follows:-

- When applied to eyes–open state data, the covariance method gave an 85% accuracy in discriminating the two groups.
- Burg and covariance approach on eyes-closed state gave an accuracy of 80%.
- In an eyes-open condition, the relative alpha power provided the maximum accuracy of 80%.

Compared to spectral features, AR model parameters have shown to be better in discriminating between ADHD and control adults.

In the paper (Zhang et al., 2018), the ADHD-200 dataset is used. The number of subjects from the ADHD – 200 dataset is as follows: -

- **NI –** Total subjects: 48 (23 Control Subjects, 25 ADHD Subjects)
- **KKI –** Total subjects: 83 (61 Control Subjects, 22 ADHD Subjects)
- **Peking –** Total subjects: 86 (62 Control Subjects, 24 ADHD Subjects)
- **NYU –** Total subjects: 216 (98 Control Subjects, 118 ADHD Subjects)

Typically, ADHD is diagnosed based on a doctor's judgment and a lengthy procedure. Still, BOLD - fMRI (Blood Oxygen Level Dependent fMRI) has developed as a new technique to learn more about ADHD. Thankfully, a Functional Connectivity (FC) network of the brain can be generated from fMRI data utilizing the temporal and spatial coherence of the BOLD signal, successfully identifying additional elementary dissimilarities between ADHD and control subjects. Consequently, FC analysis has emerged as an essential method of diagnosing and characterizing ADHD.

FC variations are classified as characteristics by incorporating various ML approaches into multiple automatic diagnosis methods for ADHD proposed in recent decades. If another perspective is used and different graph-based methods for ADHD classification are attempted, FC might be considered as a topographic map. In recent years, deep learning has made its way into the spotlight, demonstrating that it can calculate the disrupted connections associated with ADHD. To describe the FCs of ADHD participants, a fully connected network is employed to compute the similarities between the collected features and Siamese architecture using a Convolutional Neural Network (CNN).

As far as sparse representation is concerned, with the help of a self-learned dictionary, it reconstructs the data and mainly consists of sparse coding and dictionary learning. As Least Absolute Shrinkage and Selection Operator (LASSO) and Elastic Net (EN) are used in feature selection for FC, similarly, in the classification of ADHD, sparse coding is implicitly used with its varied expressions.

Inspired by the process mentioned above, they designed a feature space separation via a sparse representation classification framework. The methods were tested on the ADHD-200 database, which shows that their approach can compete with other current methods.

The best accuracy of the NI dataset is 75%, given by their proposed method. KKIdataset gives the best accuracy of 75.6% with Graph MRI method, Peking dataset gives 85.3% with Fusion MRI method, and NYU dataset provides 73% with Deep MRI method.

The technology of deep learning was used to diagnose ADHD (Zou et al., 2017). This study uses a dataset of three typical types of fMRI characteristics as an input dataset to augment the performance of the CNN algorithm. The three features were provincial homogeneousness, voxel connection, and low amplitude of low-frequency variation. These features were integrated, and a 3D dataset was generated, which was then used to create a 3D CNN model. This approach has aided in the accurate diagnosis of ADHD condition. 10-fold cross-validation was also incorporated, allowing them to choose distinct validation sets for each occurrence. They also compared the accuracies of several classifiers used in state-of-the-art approaches.

On the ADHD–200 dataset, the performance of state-of-the-art approaches employing classifiers such as multi-kernel learning and SVM was compared to the suggested method using 3D CNN classifier. Their recommended strategy was the most accurate, with a 65.67 percent accuracy rate.

In work (Hao, He & Yin, 2015), the researchers aimed to use the relationship of the features stored in fMRI scans to enhance the model's prediction accuracy. The researchers merged the Naïve Bayesian network with a belief network to achieve the desired relationship model to detect ADHD. The different normalization and dimensionality reduction techniques are being used to make the fMRI data suitable for the model.

Deep Belief Network: It is one of the most famous unsupervised multi-layer machine learning models that use the concept of a hidden layer of the deep learning models primarily used in image recognition, speech recognition, etc. A deep belief network uses greedy algorithms to generate and update the model's weights by learning in a top-to-down fashion. The stochastic gradient descent and backward propagation models update the weights in several runs from the top to hidden layers until the bottom layer. To predict the outcomes probability method such as maximum likelihood is used.

Bayesian Network: Bayesian Networks are the probability-based graphical model that uses the relationship between the features and their dependability on each other. This relationship-building minimizes the probability calculation to a great extent. The functional features from the MRI scans are visualized in the form of a directed graph. The probability of all the features is evaluated based on the probability of related features. An accuracy of about 66% is achieved.

In the research presented in **(**Brown et al., 2012**),** the researcher aims to build an intelligent model that effectively classifies the homo sapiens fMRI data into three different categories: healthy, ADHD combined (ADHD-c), ADHD inattentive (ADHD-I) uses individuals personal characteristics such as intellectual level, age, sex, etc. Before performing the diagnosis, feature extraction methods include PCA, Fourier transform, and functional connectivity.

The raw dataset of functional MRI data or the individual characteristics data is pre-processed at several stages. The researchers generated the variants of functional MRI data vector that are as follows:

- Averaging functional MRI signal intensity over time dimension.

- Averaging functional MRI signal intensity over spatial dimensions.
- Fourier frequency components.
- Derivatives of feature maps.

This acquired feature set was fed as input to ML classifiers which involve support vector machine linear, quadratic, cubic, and kernel, multidimensional logistic classifier, radial basis function support vector classifier. Different combinations of the pre-processed data were also fed as input to the classifiers. K-fold cross-validation with k=10 is also carried out to make the classifiers learn more effectively and deeply, increasing the accuracy of the test set.

The classifier performs the diagnosis with personal characteristic data eliminating the Imaging data or functional MRI scans. The training is carried out on several different feature sets. The first feature set involves seven features - site from where the data is collected, handedness (left-handed, right-handed), age, gender, intelligent oral level, intelligent physical level, and complete intelligent level. The second set is the same as the first set with two add-on features intelligence level measure, the site from which respective neuroimage is selected. The third feature set only updates the second set replacing all positive values by 1 and negative values with 2. The NULL values in the dataset are filled by the mean values of the specific columns followed by normalizations using standard scaling, bringing all data values between the range of 0-1.

These high dimensional images, if directly served as input to the dataset, the machine learning model will overfit the data. To overcome the problem of overfitting, the single-point precision setting is reduced to one-point precision. To carry out a down-sampling method of averaging the window is used. Finally, the dataset obtained is fed to the classifier to build and predict the ADHD.

For the data used throughout the training period, the researchers set a chance accuracy of 64.2 percent. On all of the feature set combinations, the logistic regression classifier, linear, quadratic, cubic, and kernel support vector machine fared better. Kernel-rbf SVM demonstrated the best performance. The linear SVM achieved a maximum accuracy of 75%.

2.1Integrated summary of the literature studied

The most common global datasets accessible, notably ADHD-200 and ABIDE, are used in most previous research efforts conducted by various academics. Both datasets contain high-resolution structural and functional MRI images of people with ADHD and healthy people. In several of the research studies we've come across, this imaging dataset has been combined with phenotypic data like age, gender, handedness, cognitive level, and brain scanning sites. Phenotypic data was shown to be more beneficial in developing models that could accurately detect the condition.

Because the imaging data contains high-resolution images, it is critical to reducing the dataset's resolutions, dimensionality, and linearity. To achieve this, practically all studies have oftenly used dimensionality reduction techniques, such as PCA and LDA, with some studies including FFT. In order to achieve more accuracy, several images were rotated at different angles, blurred, and sharpened. Various strategies were utilized to enhance or minimize the noise depending on the necessity and training. The feature maps were then subjected to multiple processes depending on the model's requirements, such as dataset scaling, k-fold cross-validation, and averaging.

The feature map created at this point is finally served to machine learning models. SVM models are included in all of the studies as one of the essential models for classifying healthy vs. ADHD patients.

3. Methodology

This section is divided into two parts. In the first part, the methodology followed for machine learning models is discussed, and in the next part of the section, the deep learning methodology is discussed.

1) **For Machine Learning Model:** The methodology followed for the machine learning process is shown in Figure 3. The steps involved are discussed below in detail.

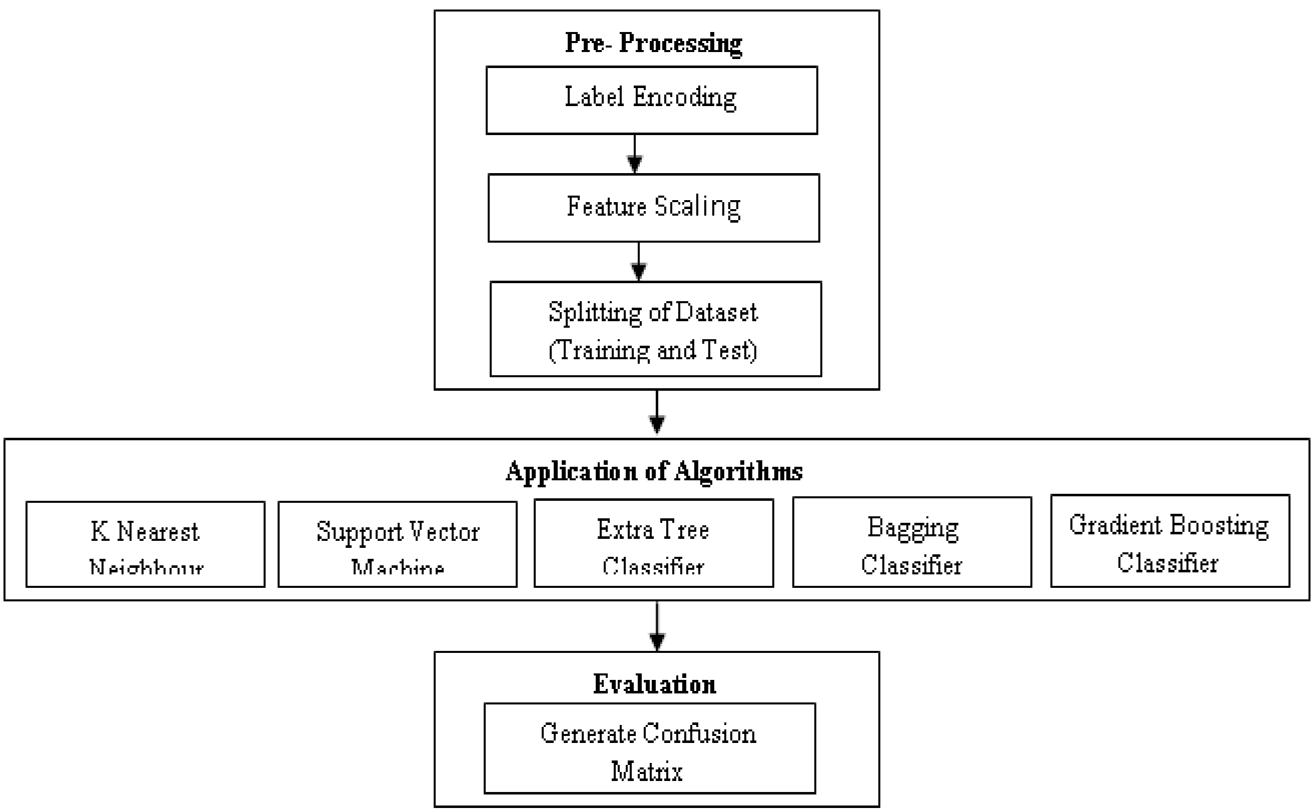


Figure 3. Methodology for Machine Learning

a. Label Encoding: Entails translating labels to numeric form to transform them into integer values, as machine learning models work with integer datasets. Then, using machine learning algorithms, more intelligent judgments may be made about labelling. It is a decisive pre-processing step in supervised learning for the structured dataset. Gender, medical, and aggression were the three columns in our dataset with values in string format. As a result, we used Label Encoding to solve this issue.
b. Feature Scaling: Feature scaling condenses the properties of the data into a small range. It is used in data pre-processing to handle wildly varying magnitudes, values, or units.
c. K-Nearest Neighbour (KNN): KNN is the simplest supervised learning algorithm. In KNN, the K closest train data instances are founded using the 21 optimized distance method for each test data instance. Then the probability of all possible classes is calculated using the formula number of instances of each class in K closest instances divided by K. Then, the class with maximum probability is returned.
d. Support Vector Machine (SVM): SVM is the most famous machine learning algorithm used for regression and classification tasks. We have used SVM for classification to classify ADHD persons and non ADHD. The SVM finds the N-dimensional hyperplane that best distinguishes the classes. The main aim of the SVM model is to find the decision boundary such that it is as wide as possible with the minimum number of violations.
e. Extra Tree Classifier: It is a decision tree ensemble machine learning algorithm. In this approach, many decision tree estimators are built using random subsets of the dataset. The trees are typically created differently, as they are based on a random subsample of each dataset. While testing, each instance is assessed against all of the trees that have been created, and the class with the most votes is returned.
f. Bagging Classifier: In bagging, several weak learners are trained parallel to each other independently on a random subsample of the dataset. During testing, each instance is assessed using all of the estimators that have been created. The class with the most votes is then returned. We have used a decision tree classifier as the learner and created five estimators using a parameter selection algorithm. The five estimators vote and return the maximum probability class for each test case.
g. Gradient Boosting Classifier: In a gradient boosting classifier, several weak learners are trained in a sequence where each new estimator aims to reduce the loss done by the previous learner. Then, during testing, each instance is evaluated on all of the estimators produced, and the class with the most votes is returned. We have used a decision tree classifier as the learner and created ten estimators using a parameter selection algorithm. The ten estimators vote and return the maximum probability class for each test case.
h. Confusion matrix: The confusion Matrix is used to check the model's accuracy on the test dataset. This matrix can be used for any number of classes.

2) **For Deep Learning Model:** The methodology followed for the deep learning model is shown in Figure 4. The steps involved are discussed below in detail.

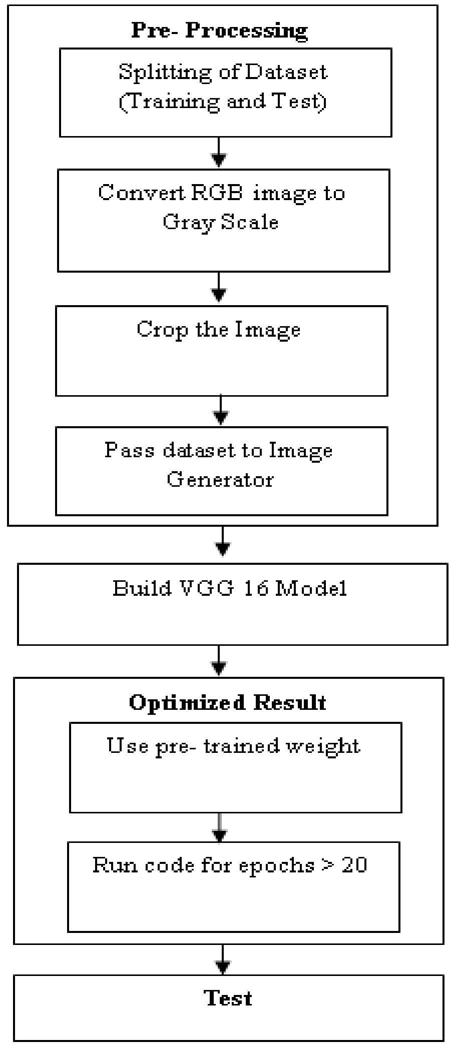


Figure 4. Methodology for Deep Learning

a. CNN VGG-16: It is a 16-layered CNN-based model prioritizes convolutional layers over other hyperparameters. Each convolution layer in VGG-16 employs 3x3 filters with one stride, while the max-pooling layer uses 2x2 filters with two strides. As seen in Figure 5, the output has two completely linked layers followed by a soft-max layer.

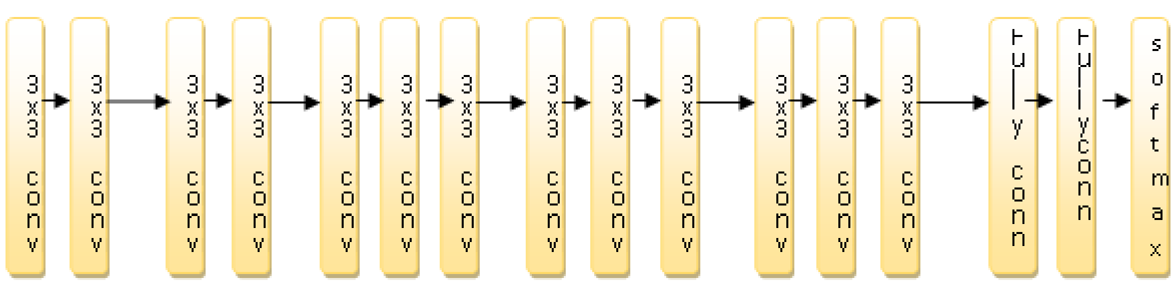


Figure 5. Basic Structure of VGG-16

b. Image data generator(): The image data generator automatically labels all the images in the dataset. Images in the 'yes' folder are labeled as yes, and those in the 'NO' folder are labeled as no.

4. Implementation Details and Issues

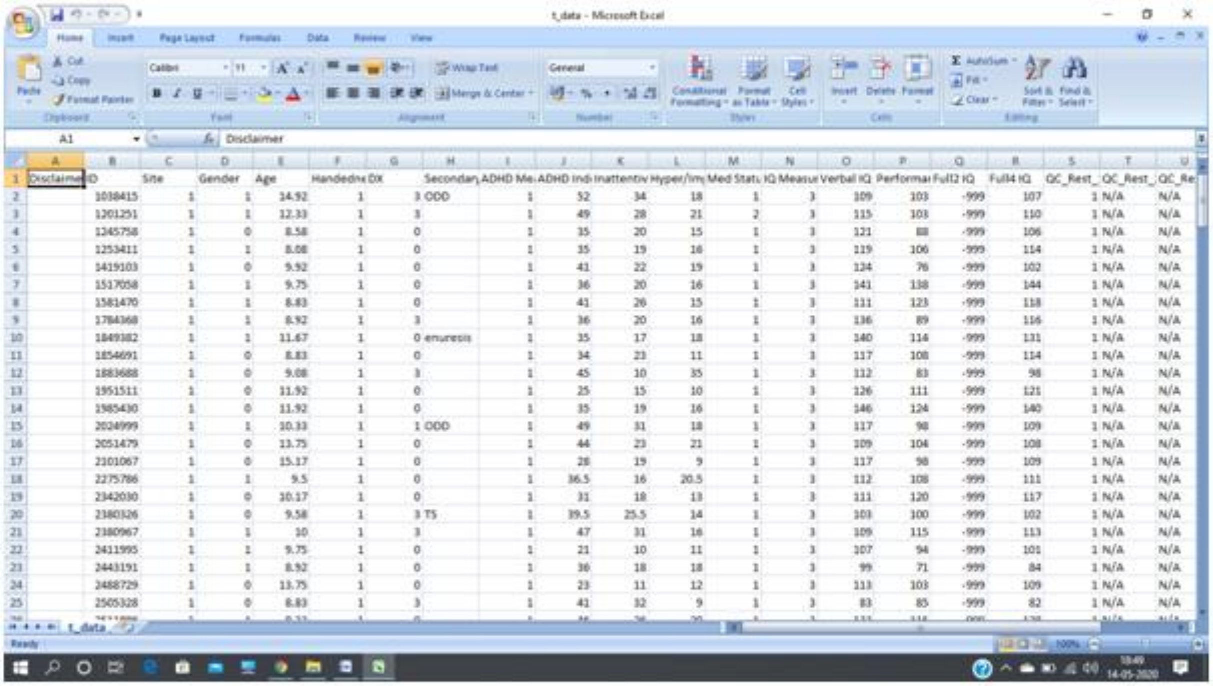
ML and deep learning models are used to predict ADHD disease. The supervised-ML models are first applied to the Phenotypic dataset, followed by the deep learning CNN VGG16 model on the brain scans images dataset.

4.1Dataset

The government launched a competition to forecast ADHD disease using two types of datasets: phenotypic datasets, which are large in number, and brain MRI scans of ADHD and non-ADHD patients, which are small in number. Various universities contributed the datasets. The majority of the datasets were gathered by the general population, who released 776 fMRI images in which anatomical datasets were combined utilizing several imaging locations. The source of the dataset is http://fcon_1000.projects.nitrc.org/indi/adhd200/. Hence, two types of datasets are used.

a. Phenotypic dataset: It contains the phenotypic data of 221 ADHD and non-ADHD individuals. As illustrated in Figure 6, the phenotypic information encompasses eight features: gender, HI, med, aggressiveness, perfIQ, VerbIQ, fullIQ, and

AD.



	Disclaimer	ID	Site	Gender	Age	Handedness	DX	Secondary	ADHD Me	ADHD Ind	Inattentiv	Hyper/Im	Med Statu	IQ Measur	Verbal IQ	Performan	Full2 IQ	Full4 IQ	QC_Rest_	QC_Rest_	QC_Re
2		1038415	1	1	14.92	1	3	ODD	1	52	34	18	1	3	109	103	-999	107	1	N/A	N/A
3		1201251	1	1	12.33	1	3		1	49	28	21	2	3	115	103	-999	110	1	N/A	N/A
4		1245758	1	0	8.58	1	0		1	35	20	15	1	3	121	88	-999	106	1	N/A	N/A
5		1253411	1	1	8.08	1	0		1	35	19	16	1	3	119	106	-999	114	1	N/A	N/A
6		1419103	1	0	9.92	1	0		1	41	22	19	1	3	124	76	-999	102	1	N/A	N/A
7		1517058	1	1	9.75	1	0		1	36	20	16	1	3	141	138	-999	144	1	N/A	N/A
8		1581470	1	1	8.83	1	0		1	41	26	15	1	3	111	123	-999	118	1	N/A	N/A
9		1784368	1	1	8.92	1	3		1	36	20	16	1	3	136	89	-999	116	1	N/A	N/A
10		1849382	1	1	11.67	1	0	enuresis	1	35	17	18	1	3	140	114	-999	131	1	N/A	N/A
11		1854691	1	0	8.83	1	0		1	34	23	11	1	3	117	108	-999	114	1	N/A	N/A
12		1883688	1	0	9.08	1	3		1	45	10	35	1	3	112	83	-999	98	1	N/A	N/A
13		1951511	1	0	11.92	1	0		1	25	15	10	1	3	126	111	-999	121	1	N/A	N/A
14		1985430	1	0	11.92	1	0		1	35	19	16	1	3	146	124	-999	140	1	N/A	N/A
15		2024999	1	1	10.33	1	1	ODD	1	49	31	18	1	3	117	98	-999	109	1	N/A	N/A
16		2051479	1	0	13.75	1	0		1	44	23	21	1	3	109	104	-999	108	1	N/A	N/A
17		2101067	1	0	15.17	1	0		1	28	19	9	1	3	117	98	-999	109	1	N/A	N/A
18		2275786	1	1	9.5	1	0		1	36.5	16	20.5	1	3	112	108	-999	111	1	N/A	N/A
19		2342030	1	0	10.17	1	0		1	31	18	13	1	3	111	120	-999	117	1	N/A	N/A
20		2380326	1	0	9.58	1	3	TS	1	39.5	25.5	14	1	3	103	100	-999	102	1	N/A	N/A
21		2380967	1	1	10	1	3		1	47	31	16	1	3	109	115	-999	113	1	N/A	N/A
22		2411995	1	1	9.75	1	0		1	21	10	11	1	3	107	94	-999	101	1	N/A	N/A
23		2443191	1	1	8.92	1	0		1	36	18	18	1	3	99	71	-999	84	1	N/A	N/A
24		2488729	1	0	13.75	1	0		1	23	11	12	1	3	113	103	-999	109	1	N/A	N/A
25		2505328	1	0	8.83	1	3		1	41	32	9	1	3	83	85	-999	82	1	N/A	N/A

Figure 6. A snippet of Phenotypic Dataset

b. Brain MRI Scans: It includes 98 brain scans of non-ADHD people and 155 brain scans of ADHD people. An example of a brain scan is illustrated in Figure 7.

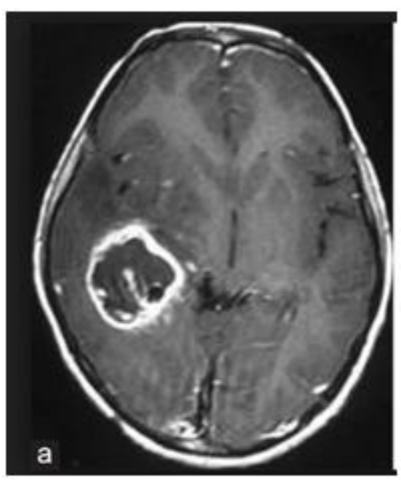


Figure 7. Brain Image Scan (top view)

4.2Machine Learning Model on Phenotypic Dataset

A phenotypic dataset is collected to train our machine learning models. Gender, HI, med, aggressiveness, perfIQ, VerbIQ, fullIQ, and AD are the eight features that make up the phenotypic data. The first step for building Machine Learning models is data pre-processing. Pre-processing the dataset on characteristics such as skewness, dimensionality, null values, categorical columns, complexity, outliers, and maximum and minimum values of each column is done. Null values in the dataset are replaced with their respective column average using the imputer class from Scikit-learn. The categorical columns are handled using the Pandas get_dummies() function that generates binary columns for the categorical columns (gender, med, and aggression). Since the range of all the columns is different, it can decrease the model efficiency; therefore, the dataset is scaled using feature scaling so that all the column values lie in the same range. After that, the correlation analysis is carried out to drop the features with the least correlation. The dataset obtained is divided into a train set, test set, and validation set. Finally, the dataset is trained on five models: KNN, SVM, Extra tree classifier, Bagging classifier, and Gradient Boosting classifier. The models are tested using a test dataset and evaluated on precision, accuracy, and recall parameters. The accuracy recorded varies from 70% to 89%.

4.3 Deep learning model on Brain MRI Scans

We employed the VGG16 model and a collection of brain scan images for deep learning. There are 155 photographs in the ADHD category and 98 images in the non-ADHD category.

The dataset was divided into a training set, a testing set, and a validation set. The process began with image scaling, and RGB photos were transformed to grayscale. The images are then cropped to remove background noise, which reduces the size of the image pixel array. The data set is saved and handed to the image generator. The classes will be assigned to the dataset by the image generator. Now finally VGG16 model is built, and pre-trained weights are used to perform fast training. VGG-16 model summary is shown in Figure8. Then the model is trained for 30 epochs using train and validate sets. Finally, the model is tested on a test dataset, and an accuracy of 90% is achieved.

```
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
vgg16 (Model)                (None, 7, 7, 512)         14714688
_________________________________________________________________
flatten_1 (Flatten)          (None, 25088)             0
_________________________________________________________________
dropout_1 (Dropout)          (None, 25088)             0
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 25089
=================================================================
Total params: 14,739,777
Trainable params: 25,089
Non-trainable params: 14,714,688
_________________________________________________________________
```

Figure 8. CNN VGG-16 Model Summary

4.4 Implementation Results

Figure 9 shows the precision, recall, f1-score, and support value achieved after applying different Machine Learning algorithms, namely KNN, SVM, ETC, Bagging classifier, Gradient Boosting Classification on the phenotypic dataset. Figure 10 and Figure 11 show the model loss and model accuracy of VGG-16 on brain MRI scans, respectively.

	precision	recall	f1-score	support
0.0	0.85	0.88	0.86	32
1.0	0.67	0.62	0.64	13
accuracy			0.80	45
macro avg	0.76	0.75	0.75	45
weighted avg	0.80	0.80	0.80	45

(a) KNN Classification Report

	precision	recall	f1-score	support
0.0	0.88	0.91	0.89	32
1.0	0.75	0.69	0.72	13
accuracy			0.84	45
macro avg	0.81	0.80	0.81	45
weighted avg	0.84	0.84	0.84	45

(b) SVM Classification Report

	precision	recall	f1-score	support
class1	0.87	0.94	0.90	48
class2	0.80	0.63	0.71	19
accuracy			0.85	67
macro avg	0.83	0.78	0.80	67
weighted avg	0.85	0.85	0.84	67

(c) ETC Classification Report

	precision	recall	f1-score	support
0.0	0.90	0.84	0.87	32
1.0	0.67	0.77	0.71	13
accuracy			0.82	45
macro avg	0.78	0.81	0.79	45
weighted avg	0.83	0.82	0.83	45

(d)Bagging Classification Report

	precision	recall	f1-score	support
0.0	0.92	0.92	0.92	26
1.0	0.80	0.80	0.80	10
accuracy			0.89	36
macro avg	0.86	0.86	0.86	36
weighted avg	0.89	0.89	0.89	36

(e) Gradient Boosting Classification Report

Figure 9: Classification Reports

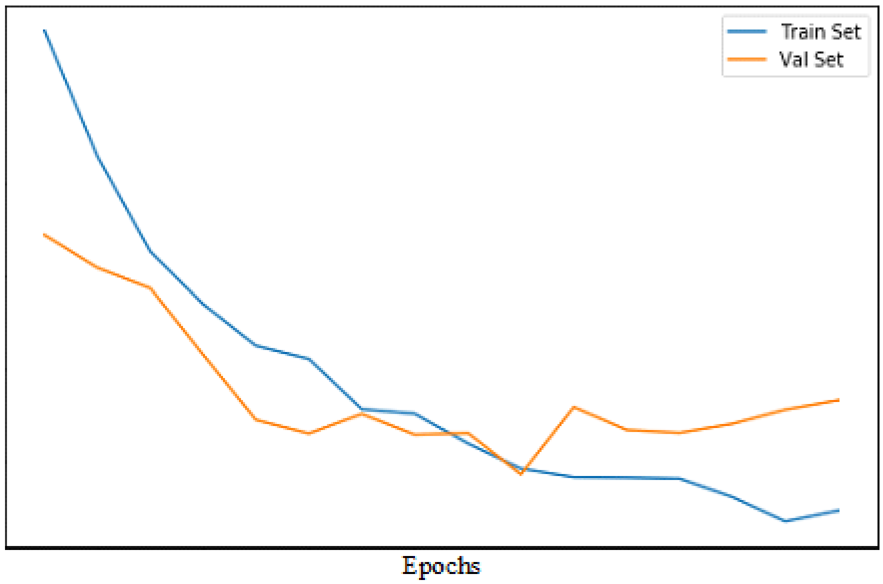


Figure 10. CNN VGG-16 Model Loss

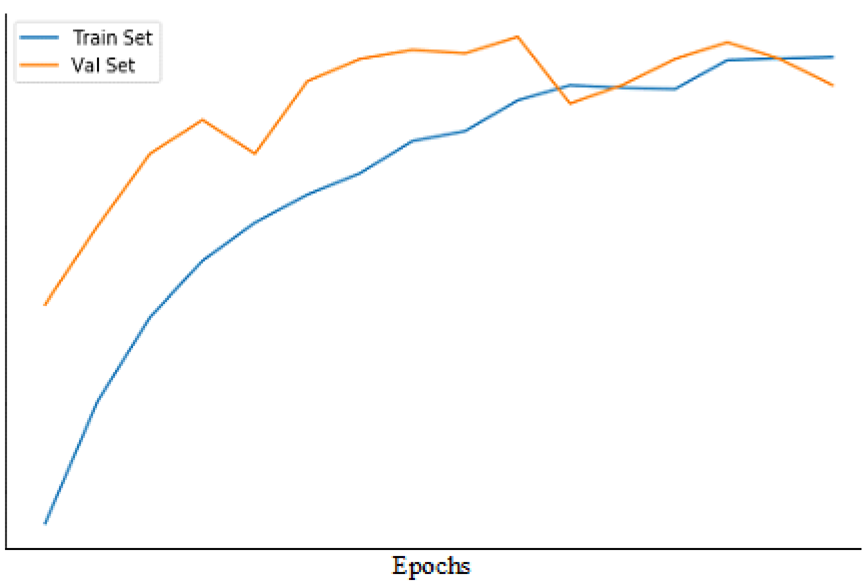


Figure 11. CNN VGG-16 Model Accuracy

Studies employing the DSM-III or DSM-III-R classification system states that the occurrence of ADHD in the general, unscreened, school-age U.S. populace ranges from 4 percent to 12 percent. According to a multiple logistic regression study with random effects, sex, diagnostic tool, and environment are all critical predictors in the occurrence of ADHD, but not age. When impairment is necessary for diagnosis, the event of ADHD is significantly lower than when impairment is not included, according to a single study employing the DSM-IV categorization scheme (7 percent as compared to 16 percent). Boys exhibited greater rates of ADHD than girls across the board, with inattentive ADHD being the most frequent.

According to the statistics, ADHD often co-occurs with the oppositional defiant disorder, conduct disorder, anxiety disorder, depressive disorder, and learning impairment in the general, unscreened school-age population. Around 35% of children diagnosed with ADHD also had an oppositional defiant disorder diagnosis, 28% had a conduct disorder diagnosis, 26% had an anxiety disorder diagnosis, and 18% had a depressive disorder diagnosis. The frequency of learning difficulties in children with ADHD is around 12%. The prevalence of ADHD was estimated to be at 4% in the screened school-age population.

5. Conclusion and Future Work

We chose to research the neurological condition known as ADHD to contribute to society's growth. Several studies have been conducted to aid doctors in detecting and curing this neurological condition.

We employed several ensemble learning models such as Extra tree classifier, bagging, and Ada boosting to enhance illness prediction and classification. Then we used the deep learning model CNN VGG-16 on brain MRI images to predict ADHD disease and reached a 90% accuracy rate. Different models are integrated with SVM as one of the main aspects in many studies to effectively create the findings, such as CNN + SVM (linear/kernel). ELM, or feedforward neural networks, was the other model employed. The accuracy of various study contributions ranges from 50 percent to 91 percent. The accuracy rate ranges from 72 to 76 percent on average. The ELM model reached the greatest accuracy of around 91 percent.

We hope to have a larger dataset in the future so that this project can be more dependable. We're thinking about expanding the dataset, making it more comprehensive, and integrating new characteristics affecting ADHD patients. First and foremost, we are attempting to make it live on a website, followed by completing additional research articles on this topic. We are currently aiming for a complete software system that doctors in hospitals can operate.

References

1. Bledsoe, J. C., Xiao, C., Chaovalitwongse, A., Mehta, S., Grabowski, T. J., Semrud-Clikeman, M., ... & Breiger, D. (2020). Diagnostic classification of ADHD versus control: support vector machine

classification using brief neuropsychological assessment. *Journal of Attention Disorders*, *24*(11), 1547-1556.

2. Brown, M. R., Sidhu, G. S., Greiner, R., Asgarian, N., Bastani, M., Silverstone, P. H., ... & Dursun, S. M. (2012). ADHD-200 Global Competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements. *Frontiers in Systems Neuroscience*, *6*, 69.

3. De Silva, S., Dayarathna, S., Ariyarathne, G., Meedeniya, D., Jayarathna, S., Michalek, A. M., & Jayawardena, G. (2019, July). A rule-based system for ADHD identification using eye movement data. In *2019 Moratuwa Engineering Research Conference (MERCon)* (pp. 538-543). IEEE.

4. Hao, A. J., He, B. L., & Yin, C. H. (2015, November). Discrimination of ADHD children based on Deep Bayesian Network. In *2015 IET International Conference on Biomedical Image and Signal Processing (ICBISP 2015)* (pp. 1-6). IET.

5. Kaur, S., Arun, P., Singh, S., & Kaur, D. (2018, December). EEG Based Decision Support System to Diagnose Adults with ADHD. In *2018 IEEE Applied Signal Processing Conference (ASPCON)* (pp. 87-91). IEEE.

6. Miao, B., & Zhang, Y. (2017, July). A feature selection method for classification of ADHD. In *2017 4th International Conference on Information, Cybernetics and Computational Social Systems (ICCSS)* (pp. 21-25). IEEE.

7. Miao, B., Zhang, L. L., Guan, J. L., Meng, Q. F., & Zhang, Y. L. (2019). Classification of ADHD individuals and neurotypicals using reliable RELIEF: A resting-state study. *IEEE Access*, *7*, 62163-62171.

8. Peng, X., Lin, P., Zhang, T., & Wang, J. (2013). Extreme learning machine-based classification of ADHD using brain structural MRI data. *PloS One*, *8*(11), e79476.

9. Polanczyk, G. V., Willcutt, E. G., Salum, G. A., Kieling, C., & Rohde, L. A. (2014). ADHD prevalence estimates across three decades: an updated systematic review and meta-regression analysis. *International Journal of Epidemiology*, *43*(2), 434-442.

10. Polanczyk, G., de Lima, M.S., Horta, B.L., Biederman, J., Rohde, L.A., 2007. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. The American Journal of Psychiatry 164, 942–948.

11. Qureshi, M. N. I., Jo, H. J., & Lee, B. (2017, April). ADHD subgroup discrimination with global connectivity features using hierarchical extreme learning machine: resting-state FMRI study. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (pp. 529-532). IEEE.

12. Saeed, F. (2018). Towards quantifying psychiatric diagnosis using machine learning algorithms and big fMRI data. *Big Data Analytics*, *3*(1), 1-3.

13. Sen, B., Borle, N. C., Greiner, R., & Brown, M. R. (2018). A general prediction model for the detection of ADHD and Autism using structural and functional MRI. *PloS One*, *13*(4), e0194856.

14. Shao, L., Zhang, D., Du, H., & Fu, D. (2019). Deep forest in ADHD data classification. *IEEE Access*, *7*, 137913-137919.

15. Sidhu, G. S., Asgarian, N., Greiner, R., & Brown, M. R. (2012). Kernel Principal Component Analysis for dimensionality reduction in fMRI-based diagnosis of ADHD. *Frontiers in Systems Neuroscience*, *6*, 74.

16. Tenev, A., Markovska-Simoska, S., Kocarev, L., Pop-Jordanov, J., Müller, A., & Candrian, G. (2014). Machine learning approach for classification of ADHD adults. *International Journal of Psychophysiology*, *93*(1), 162-166.

17. Wang, M., Jie, B., Bian, W., Ding, X., Zhou, W., Wang, Z., & Liu, M. (2019). Graph-kernel based structured feature selection for brain disease classification using functional connectivity networks. *IEEE Access*, *7*, 35001-35011.

18. Wender, P., 1995. Attention-deficit Hyperactivity Disorder in Adults. Oxford University Press, New York.

19. Yao, D., Guo, X., Zhao, Q., Liu, L., Cao, Q., Wang, Y., ... & Sui, J. (2018, July). Discriminating ADHD from healthy controls using a novel feature selection method based on relative importance and ensemble learning. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4632-4635). IEEE.

20. Zhang, Y., Tang, Y., Chen, Y., Zhou, L., & Wang, C. (2018, November). ADHD classification by feature space separation with sparse representation. In *2018 IEEE 23rd International Conference on Digital Signal Processing (DSP)* (pp. 1-5). IEEE.

21. Zou, L., Zheng, J., & McKeown, M. J. (2017, November). Deep learning based automatic diagnoses of attention deficit hyperactive disorder. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (pp. 962-966). IEEE.

22. Zou, L., Zheng, J., Miao, C., Mckeown, M. J., & Wang, Z. J. (2017). 3D CNN based automatic diagnosis of attention deficit hyperactivity disorder using functional and structural MRI. *IEEE Access*, *5*, 23626-23636.