Attention Deficit Hyperactivity Disorder(ADHD) Prediction System

[1]Dr. Anil Kumar Dubey,[2] Akshat Saxena,[3] Atul Kumar,[4] Ayush Rai
Computer Science and Engineering Department
ABES Engineering College, Ghaziabad
Uttar Pradesh, India, 201009
[1]anil.dubey@abes.ac.in, [2]akshat.18bcs1159@abes.ac.in, [3]atul.18bcs1082@abes.ac.in,
[4]ayush.18bcs1038@abes.ac.in

***Abstract* - The purpose of this paper is to give a ephemeral impression about the tasks undertaken by us during the project 'ADHD Prediction System'. Initial part of the paper provides an overview of the problem statement, tailed by the work done and approach taken and lastly the role of various algorithms. Due to the limited availability of doctors and its negative impact on the patients we are a/so working simultaneously on identifying suitable prospects for the establishment of systems that bridge the gap between the patients and doctors. The working of Prediction system along with the differentiation between the various approaches is also described within the report. Flowcharts and Use Case Diagrams form a basis of all the further discussions and tasks accomplished by us during the project. To conclude, the paper encompasses of the implementation specifics and outcomes obtained in a processed and progressive manner**

Keywords-ADHD, ADHD-I, ADHD-Common, MRI, fMRI, RELEIF algorithm, PCA, Phenotypic, D2 test-Support vector machine, MRI- Magnetic resonance Imaging, CNN, EEG-electroencephalogram, IQ –Intelligence Quotient.

I. INTRODUCTION

Attention deficit hyperactivity disorder is a mental disorder which creates learning disabilities in the children which is unbearable at the early learning stage of life. Its symptoms are not always well understood. Most children do not get a proper diagnosis and are constantly told by their teacher that the subject is not doing well in aspects, not doing his or her best. Neuro-health disorders are a disorder that is affecting more and more people. ADHD is a very serious and neglected mental illness that is found in children and people of all ages suffering from different ADHD subtypes [1] and [2]. We as a society have found it very problematic and disgraceful to accept it as a disease. The people in the society have a feeling that if a person has mental illness, it is bad for the status of the person. People forget that these disorders are so harmful and impactful that it can be life threatening for an individual. It is absolutely crucial to understand that these are curable only if diagnosed at early stages. If it does not happen then in such a case it brings a lot of difficulties and issues. The symptoms are not self-sufficient to prove that the person is ill so it is important to observe the behaviour of the subject too. The MRI scans of ADHD shows that the brain when compared to the normal brain is reduced in size. This cause the hindered growth of the individual and causes more severe problems. The main problem that is associated with it is the flow of the blood is reduced giving less oxygen and improper growth to the brain. With the constant advancements being made in the various fields of technology, the field of medicine does not lag too far behind the new medical diagnosis techniques being developed every day. Furthermore, there is a rise in the number of people requiring medical attention with the continuously increasing worldwide population numbers. Even though the technologies are improving, the number of doctors still remains a little bit scarce as compared to the number of patients. With the use of efficient, better methods and techniques we want to help in detection of this problem at early state and its types so that it can be beneficial to the people in the real time.

II. LITERATURE SURVEY

After studying all various literature related to our area of project, we come to the point that there is various way to predict and classify ADHD. All those studies which are using machine and deep learning techniques are majorly focusing on effective Dataset collection and after that implementation of most optimum machine learning algorithm. In dataset collection Phenotypic and Image based data is most effective source of collection. Phenotypic data refers to parameter based on age, sex, intelligence, handedness etc and image-based data refers to Magnetic Resonance Imaging known as MRI. In MRI functional MRI which show blood flow in brain during performing some activity is very useful for dataset. Various Image enhancement techniques like PCA are used for optimal image resolution and angle rotation. RELEIF algorithm are used for feature selection in some research. For implementing Machine learning model and Deep learning model SVM is very effectively used in various papers. In some of them SVM is integrated with some other model like CNN for better efficiency. EEG which used to plot brain simulation is also used in some studies. The average results accuracy rate is 72%to 76%

Zou, Liang [3] predicted ADHD using deep learning. fMRI is used as input for CNN model. Different feature of fMRI is used in 3D dataset and 3D CNN model is generated. Multi kernel learning, SVM are compared to 3D CNN used in ADHD-200 data set. This proposed method predicts with highest accuracy of 65.67%.

Brown [4] Machine learning classifier classify the disease into three groups -ADHD-C, ADHD impulsive and ADHD inattentive (ADHD-I) without taking much advantage of fMRI data rather using individuals personal characteristics

such as intelligent level, age, gender, IQ (Verbal IQ, Full Scale)etc. The research developed various set of fMRI dataset. This obtained feature set were served as input to ML classifier. The accuracy was 62.52 % for this model.

Sidhu [5] investigated functional Magnetic Resonance Image (fMRI) scans to come up with an automated methods of diagnosis for ADHD and explored the different fMRI data to compare the diagnostic accuracy by applying various dimensionality reduction methods. Model have two type of data set known as phenotypic and image based FMRI to use for training of model. Phenotypic refers to the age ,IQ, gender, handedness sex etc .fMRI have image of blood flow in brain using magnetic resonance technique. Accuracy is calculated in various combination of dataset types which is stated below.

- Only Phenotypic dataset: 73%
- Only image dataset: 50%
- Both: 76%

Sen B. [6] used both sMRI and fMRI scans to get the accurate disease prediction of ADHD. Two types of scans used are structural MRI and fMRI. Structural MRI provide physical parameter of brain like size, shape, weight , mass whereas Functional MRI shows the image of brain associated with blood flow in different section of brain. More blood flows show high activity in that region. sMRI produces a volumetric representation of the brain anatomy using non-invasive technique. fMRI scans tells about Blood Oxygenation Level fluctuation over the cycle of period. The researchers explored the sequence of three machine learning models namely LEFM*S*, LEFM*F*, LEFM*SF*.

Peng, X. [7] made an automated classification model which helps in detecting ADHD in various region of brain with the help of classification algorithm like linear and radial SVM. In this model HD images of MRI scans are used. The researchers collected wide range of brain measures input. The features that are to be fed as input to the SVM classifier are optimally selected.

TABLE.I. Comparative Study

Research paper	Model used	Average accuracy
Zou,Liang(2017) [1]	CNN deep learning	65.67%
Brown(2012) [2]	Linear SVM	62.52%
Sidhu(2012)[3]	Linear SVM with only Phenotypic dataset	73%
	Linear SVM with only image dataset	50%
	Linear SVM with both datasets combined	76%
Sen B.(2018)[4]	LEFMS	63%
	LEFMF	62%
	LEFMSF	65%
Bledsoe(2016) [6]	SVM	62.3%
	Decision tree	61%

Bledsoe [8] and Miao [9] classify ADHD patients with the help of d2 test. It is a test in which interview is conducted with parent of children who live and grown in a controlled environment. Proper questionnaire for the parents was prepared specially for this test to perform. Accuracy achieved using SVM is 62.3% and with decision tree it is 61%.

III. PROBLEM DEFINITION AND IDENTIFICATION

With the constant advancements being made in the various fields of technology, the field of medicine does not lag too far behind with new medical diagnosis techniques being developed every day. Furthermore, there is a rise in the number of people requiring medical attention with the continuously increasing worldwide population numbers. Even though the technologies are improving, the number of doctors still remains a little bit scarce as compared to the number of patients. The main aim of this project is to cut down on the labour time and cost required to be put down in the laboratories testing the Attention Deficit Hyperactivity disorder(ADHD) samples and thus allow the doctors to be able to treat more patients in the same amount of time as well as cure in beginning evolution of disease. By using established and composition of different method we tried to enhance the existing method and approach in order to help ADHD detection and classification for patient.

IV. PROPOSED METHODOLOGY

After going through the study done earlier, we decided to implement support vector classifier and decision tree classifiers as to do the classification. We implemented support vector classifier because the accuracy achieved with it is significant and less computation power is required.

Decision tree classifier [11] was implemented because of its ability to capture the descriptive decision making knowledge from the supplied data. The data points were not generalized so the data was scaled using StandardScaler() so that the distance between them should be lower. To increase the accuracies and to let the machine learning models easily understand the data it was transformed. The dataset we used is named as ADHD-200 AND parameters are:

1) Site
   - i 1 -Peking University
   - ii 2 -Bradley Hospital/Brown University
   - iii 3 -Kennedy Krieger Institute
   - iv 4 -NeuroIMAGE Sample
   - v 5 -New York University Child Study Center
   - vi 6 -Oregon Health & Science University
   - vii 7 -University of Pittsburgh
   - viii 8 -Washington University in St. Louis
2) Gender
   - i) 0 - Female
   - ii) 1 - Male
3) Diagnosis
   - i 0 - Typically Developing Children
   - ii 1 - ADHD-Combined
   - iii 2 - ADHD-Hyperactive/Impulsive
   - iv 3 - ADHD-Inattentive
4) ADHD Measure
   - i 1 - ADHD Rating Scale IV (ADHD-RS)
   - ii 2 - Conners' Parent Rating Scale-Revised, Long version (CPRS-LV)
   - iii 3 - Connors' Rating Scale-3rd Edition
5) IQ Measure
   - i 1 - Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV)

ii 2 - Wechsler Abbreviated Scale of Intelligence (WASI)
iii 3 - Wechsler Intelligence Scale for Chinese Children-Revised (WISCC-R)
iv 4 - Two subtest WASI
v 5 - Two subtest WISC or WAIS – Block Design and Vocabulary

6) Medication Status
   i 1 - Medication Naïve
   ii 2 - Not Medication Naïve
7) Quality Control
   i 0 Questionable
   ii 1 Pass

Algorithm for Support Vector Machine:

1. Import the Libraries.
2. Choose the kernel.
3. Scale the dataset using StandardScaler().
4. Divide the dataset into train and test set.
5. Build the model.
6. Transform the parameters.
7. Fit the model on the train set.
8. Check the accuracy

Algorithm for the Decision Tree:

1. Import the libraries.
2. Choose the criterion and the random_state.
3. Scale the dataset using StandardScaler().
4. Divide the dataset into train and test set.
5. Build the model.
6. Transform the parameters.
7. Fit the model on the train set.
8. Check the accuracy.

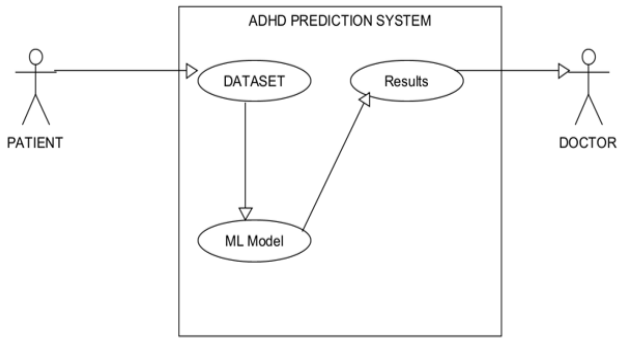

Fig. 1. ADHD use case diagram

The above Figure shows the use-case for the implementation of the project work. The aim is to collect the dataset pre-process it and then apply the machine learning models such as SVM, decision tree , obtain the results and use them. The actors are-Patient and Doctor in the fig.

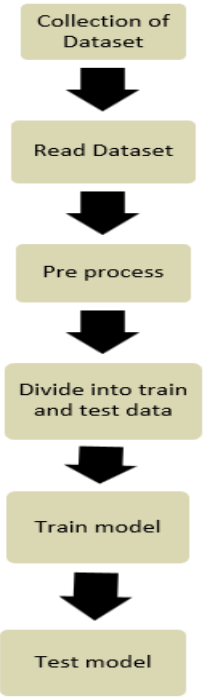

Fig. 2. Flow of steps in a machine learning model

The above figure (Figure 2) shows the basic flow chart of The machine learning model implementation. Firstly, the task is to collect the dataset, read the dataset and pre-processing it. Further dataset is divided into training and Test .With the training data the data is tested and then tested with the test dataset .

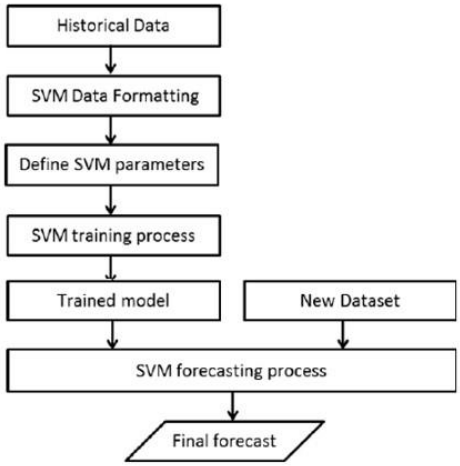

Fig. 3. SVM Model Flow

The above figure is the representation of the methodology taken to implement a basic SVM model. Firstly the data is collected. SVM parameters are selected and then the data is trained the trained model is tested for the final predictions.

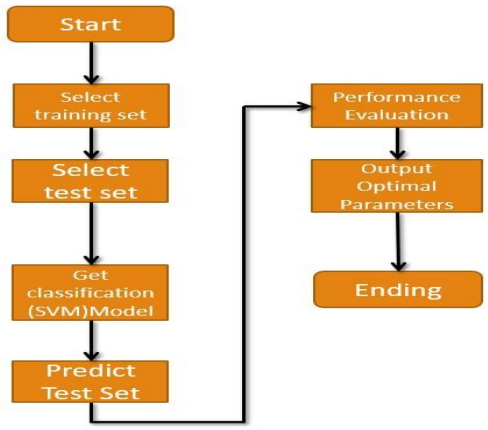

Fig. 4. Support Vector Machine Classifier Model

The above flowchart shows the SVM model that is implemented by us. Dataset was collected , pre-processed and then fitted into the model. We also did the scaling of the data to achieve better results. The values were predicted and tested against the test data . The performance was evaluated and the accuracy achieved was 69% for the support vector machine.

V. SIMULATION RESULT AND DISCUSSION

The dataset consist of various phenotypic characteristics such as age, gender, etc. then data was pre-processed the data and useful features were selected for training the model. the supervised machine learning models like support vector machine and decision tree. The models used are support vector machine with and without scaling the data, decision tree and linear regression. Successful implementation is being done of the SVM and the decision tree algorithm on the data set. The accuracy achieved with SVM is 69% and that with the Decision Tree is nearly 50%. clearly SVM is so far the best model for the prediction. The F1-score for Support vector machine classifier came out to be 0.675 and the Recall was also 0.675. The final results of various models are concluded in the table below.

TABLE.II. Accuracies of various models

Machine learning Model	**Weighted Average precision**	**Weighted Average recall**	**Weighted Average F1-score**	**Accuracy**
Support Vector Machine (without scaling data)	65%	65%	64%	65%
Support Vector Machine (with scaling data)	69%	67.5%	67.5%	69%
Machine learning Model	**Weighted Average precision**	**Weighted Average recall**	**Weighted Average F1-score**	**Accuracy**
Decision Tree Model (without scaling data)	42%	43%	42%	42%
Decision Tree Model (with scaling data)	49%	49.2%	49.2%	49.2%

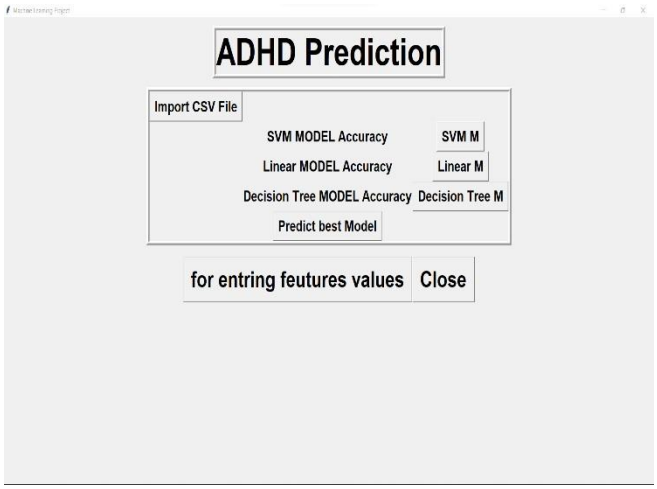

Fig. 5. GUI for ADHD Prediction

The above is the Graphical user interface of the implemented machine learning models. We used tkinter to build it. The first step is to click on the import button on the gui to import the dataset, then click on the buttons representing the various machine learning models to get their accuracies on the dataset. After clicking on the predict best model button we get the model that has the highest accuracy on the dataset. It is then implemented on the dataset. If the user feels the need to close the gui he can use the close button on the mainframe of the gui or else can click on the close button of the gui which will close the current window and exit.

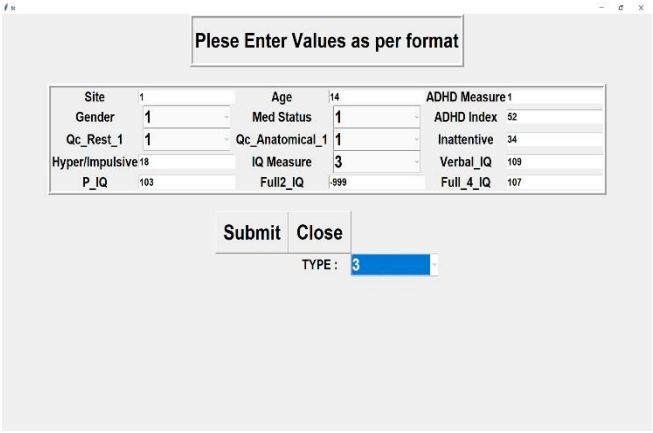

Fig. 6. Result window of GUIs

Using tkinter [12], The next window of the gui is self explainatory as the user has to enter the values in respective fields. To enter the feature values one needs to click on the for entering feature values button. If the user feels to quit the

window, he can use the close button of the window or close button of the gui. The window has different fields to take inputs of the given parameters. After entering the values user needs to click to the submit button. After the submission of the values the machine learning model with highest accuracy is implemented and the type of ADHD is displayed. If the user wants to quit, he/she may click on the close button in the gui.

VI. CONCLUSION

The conclusion of the work done is to be able to contribute to the betterment of the people suffering from ADHD. This disorder is highly ignored and not dealt with the efficient approaches and methods at the early stages. Also when known, it's hard to deal with the type of the disorder and its cure is really time consuming. We want to bridge the gap for the people and the doctors through our work. To be able to scale this work to a platform where we can save the time consumed in the process of detecting the disorder and the treatment. For doing so, we went through a lot of hard work by studying the previous work done and researches published in this domain. Also, the efficiency of the system has to be absolutely at its peak so as to minimize the possibility of an error. We know the importance of life and want to do our every little bit to the betterment of the life and the existing way. To reach to a solid and concluding decision we want to compare the work on the physical data and the work done on image data. The data obtained through the MRI is really time consuming and expensive so we want to achieve a better efficiency in the work through the physical data. The work studied involved a lot of high dimensional images which had to be transformed and checked for the linearity. Various dimensionality reducer algorithms such as the FFT were used to achieve a better efficiency. A lot of pre-processing on the data has to be done to be able to include it or induct it in the machine learning model for better accuracy. We aim to build a model, a system that has optimal efficiency takes less time and bridges the gap providing the exact solution to the existing problem. The process and the work reduce to find a unique way to pre-process the data such that it includes the data that is reliable, the features that have more correlation to the other attributes. The current system gives an accuracy of 69% on the physical data using the SVM algorithm which is to be improved and worked upon by using different kernels and other machine learning techniques and algorithms.

VII. FUTURE SCOPE

The future scope of the work is to increase the accuracy of the applied machine learning models [10], to increase the efficiency of the predictions. We also want to compare the accuracies achieved with the phenotypic data to that of the FMRI scans. Later we want to extend the same machine learning model to other area of medical Study like tumor hypertension etc. At optimum accuracy a web based portal would be developed for ease of use in between general public .Web based application includes all ML based logical concept and work as a interface for user. We want to create a system where a person can know if he/she is suffering from disease and can contact the hospital through the system directly without going to the hospital physically. This will help in reducing the time taken for the overall process and help the doctors and patients greatly.

REFERENCES

[1] Sagi, O., & Rokach, L., "Ensemble learning: A survey" in Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1249, 2018.

[2] Paulus, M. P., Kuplicki, R., & Yeh, H. W., "Machine Learning and Brain Imaging: Opportunities and Challenges" in Trends in neurosciences, 42(10), pp. 659-661, 2019.

[3] Zou, Liang, Jiannan Zheng, and Martin J. McKeown, "Deep learning based automatic diagnoses of attention deficit hyperactive disorder" in 2017 Global Conference on Signal and Information Processing (GlobalSIP).

[4] Brown, M. R., Sidhu, G. S., Greiner, R., Asgarian, N., Bastani, M., Silverstone, P. H., ... & Dursun, S. M., "ADHD-200 Global Competition: diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements" in *Frontiers in systems neuroscience*, *6*, 69, 2012

[5] Sidhu, G. S., Asgarian, N., Greiner, R., & Brown, M. R., "Kernel Principal Component Analysis for dimensionality reduction in fMRI-based diagnosis of ADHD" in *Frontiers in systems neuroscience*, *6*, 74, 2012

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

[7] Peng, X., Lin, P., Zhang, T., & Wang, J., "Extreme learning machine-based classification of ADHD using brain structural MRI data", PloS one, 8(11), e79476, 201

[8] Bledsoe, J. C., Xiao, D., Chaovalitwongse, A., Mehta, S.,T. J., Semrud-Clikeman, M., ... & Breiger, D., "Diagnostic classification of ADHD versus control: support vector machine classification using brief neuropsychological assessment" in Journal of attention disorders, 1087054716649666, 2016

[9] Miao, B., Zhang, L. L., Guan, J. L., Meng, Q. F., & Zhang, Y. L. "Classification of ADHD Individuals and Neurotypicals Using Reliable RELIEF: A Resting-State Study" IEEE Access, 7, pp. 62163-62171, 2019.

[10] Saeed, Fahad. "Towards quantifying psychiatric diagnosis using machine learning algorithms and big fMRI data." in Big Data Analytics 3.1, 7, 2018

[11] Javatpoint Tkinter tutorial (last accessed time 10:30 pm 13-01-2022)

[12] GeeksForGeeks tutorials on machine learning (last accessed time 10:30 pm 13-01-2022