AUTISM SPECTRUM DISORDER USING IMAGEPROCESSING AND MACHINE LEARNING

Dr. Vidhya KAssoc prof
Dept Of ISE
EWIT,Bangalore
vidya.k2591@gmail.com

Darshan G B Student
(BE) Dept. Of ISE
EWIT, Bangalore
Darshandruvan1@gmail.com

Giridhar S T Student (BE)
Dept. of ISE
EWIT,Bangalore
Giridhar.st2001@gmail.com

Mohan Katti Student
(BE) Dept. of ISE
EWIT,Bangalore
Kattimohanc206@gmail.com

S V Keertan Nursimha
Student (BE)
Dept. of ISE
EWIT,Bangalore
Keertan222@gmail.com

Abstract— In present day autism spectrum disorder (ASD) is gaining its momentum faster than ever. Applying Image processing technique Detecting autism traits through screening tests is very expensive and time consuming. With the advancement of artificial intelligence and machine learning (ML), autism can be predicted at quite an early stage. Though a number of studies have been carried out using different techniques, these studies didn't provide any definitive conclusion about predicting autism traits in terms of different age groups. Therefore, this paper aims to propose an effective prediction model based on ML technique and to develop a user interface forpredicting ASD for people of any age. As outcomes of this research, an autism prediction model was developed by merging Random Forest-CART (Classification and Regression Trees) and Random Forest-ID3 (Iterative Dichotomiser 3) and cnn algorithm also a user interface was developed based on the proposed prediction model. The proposed model was evaluated with AQ10 dataset and 250 real datasets collected from people with and without autistic traits.city, sensitivity, precision and false positive rate (FPR) for thThe evaluation results showed that the proposed prediction model provides better results in terms of accuracy, specifie data.

Keywords- Image Processing Technique, Random Forest CART, Random Forest- ID3, AQ10 dataset, False Positive Rate.

I. INTRODUCTION

Autism spectrum disorder (ASD) is a cognitive disability that can cause serious social, communication and behavioural challenges and its traits generally seen in the first two years of life and gradually develop through time. People with autism face different kinds of struggles with concentration, learning disabilities, mental health issues like depression, anxiety and so on, motor difficulties, sensory issues and others. Experimentation suggests that both the genes and environment play significant roles.

Present day explosion rate of autism around the world is plentiful and it is growing at a very high rate. As stated by WHO about 1 out of every 160 individuals has autism spectrum disorder. Few people with this disorder can live without depending on anyone while others reqiure lifelong support and care.

Autism spectrum disorder is a neurodevelopmental disorder that affects a person's interaction, communication and learning skills. Although diagnosis of autism can be done at any age, its symptoms generally appear in the first two years of life and develop through time. Autism patients face different types of challenges such as difficulties with concentration, learning disabilities, mental health problems such as anxiety, depression etc, motor difficulties, sensory problems and many others Image processing technique can be used to analyze data related to ASD ,such as brain imaging,facial expression recognition,, and eye tracking these technique can help identify patterns and features that may be indicative of ASD as well as track changes inn behavior and brain function over time.

Autism spectrum disorder can be controlled if found at an early stage by advising people with proper medication. This could prevent the patient's condition from getting worse and would decrease long term costs related to delayed diagnosis. Thus an effective, accurate and easy screening check tool is highly required which would detect the traits in a person and recognize whether the person requires thorough autism syndrome assessment or not. In this paper we use machine learning to find out a set of conditions that are put together to be predictive of autism spectrum disorder. This can be immensely useful to physicians, assisting them to detect autism spectrum disorder at a very early stage.

Current explosion rate of autism around the world is numerous and it is increasing at a very high rate. According to WHO, about 1 out of every 160 children has ASD. Some

people with this disorder can live independently, while others require life-long care and support. Diagnosis of autism requires a significant amount of time and cost. Earlier detection of autism can come to a great help by prescribing patients with proper medication at an early stage. It can prevent the patient's condition from deteriorating further and would help to reduce long term costs associated with delayed diagnosis.

II. LITERATURE REVIEW

A. Use of Artificial Intelligence to Shorten the Behavioral Diagnosis of Autism by Dennis P. Wall1, Rebecca Dally1, Rhiannon Luyster3, Jae-Yoon Jung1, Todd F. DeLuca

They used machine learning techniques to study the complete sets of answers to the ADI-R available at the Autism Genetic Research Exchange (AGRE) for 891 individuals diagnosed with autism and 75 individuals who did not meet the criteria for an autism diagnosis.Their analysis showed that 7 of the 93 items contained in the ADI-R were sufficient to classify autism with 99.9% statistical accuracy.They further tested the accuracy of this 7- question classifier against complete sets of answers from two independent sources, a collection of 1654 individuals with autism from the Simons Foundation and a collection of 322 individuals with autism from the Boston Autism Consortium.

B. Image Machine Learning for Autism Diagnostics: Applying Support Vector Classification by F. Hauck1 and N. Kliewer1

In this paper, they present an approach for developing a support vector machine (SVM) that can facilitate autism diagnostics. Models are trained and tested on a dataset of about 2,500 records with autism diagnostic observation schedule (ADOS) and autism diagnostic interview revised (ADI-R) items. The results show that a small combination of selected ADOS and ADI R items is sufficient to achieve a good performance in diagnosing ASD. Their models reach between 85.6% and 94.3% sensitivity and between 80.9% and 89.3% specificity with 10 features at the most. Feature selection is performed by using a greedy backward-elimination process and SVM is trained using a linear kernel as well as a radial basis function (RBF) kernel. Cross-validation (CV) is applied to ensure a high generalization performance.

C. Autism Screening using Deep Embedding Representation by Haishuai Wangl, Li Li2, Lianhua Chi3 and Ziping Zhao

In this paper, they apply novel feature engineering and feature encoding techniques, along with a deep learning classifier for ASD screening. Algorithms were created via a robust deep learning classifier and deep embedding representation for categorical variables to diagnose ASD based on behavioral features and individual characteristics. The proposed algorithm is effective compared with baselines, achieving 99% sensitivity and 99% specificity. The results suggest that deep embedding representation learning is a reliable method for ASD screening.

D. Engagement Detection with Autism Spectrum Disorder using Machine Learning by Aishwarya J, Akshatha N, Anusha H, Shishira J, Deepa Mahadev

This paper aims to propose an autism prediction model using ML techniques and to develop a web application that could effectively predict autism traits of an individual. In other words, this work focuses on developing an autism screening application for predicting the ASD traits among people of age group 3 years and below. The proposed model was evaluated with AQ-10 datasets (1054 datasets) and 50 real dataset collected from people with and without autistic traits. The evaluation results showed that the proposed prediction model provides better results in terms of accuracy, specificity, sensitivity, precision and f1- score forboth kinds of datasets. The primary limitation of the study is lack of sufficiently large data to train the prediction model. Another limitation is that the screening application is not designed for age group above 3 years.

III. EXISTING SYSTEM

The multiclass problem is implemented using the machine learning approaches like Naive Bayesian probabilistic approach, Decision Tree Classifier, K Nearest Neighbour Classifier and Support Vector Machine Classifier which are well suitable for classification problems. Used only Jupyter Notebook and not developed.User interface for Prediction.No Extra Modules were present in existing system.Used three different datasets of each age group.

IV. PROPOSED ARCHITECTURE

Data Collection:

The data required for the autism prediction is the heterogeneousgenomes which vary from each individual. Autism is a heterogeneous neurodevelopment syndrome. It involves complex genetics, etiology, DNA and genes. It has a large dataset with complex genetic structure which has to be handled to remove the noisy and inconsistent data. The dataset will be in the form of AQ 10 dataset developed using the Autism spectrum tool. The AQ 10 dataset is divided into three types based on the age. Child AQ dataset (4- 8 years), Adolescent AQ 10 dataset (teen12-18 years) Adult AQ 10 (Adults 40-65 years).

Data Pre-Processing:

The Preprocessing of genetic information includes the following:

Data Transformation:

Normalization: scaling the values to a selected range.

Aggregation: distribution probabilistic values to the genes.

Construction: replacement or adding new genes inferred by the prevailing genes.

Data Reduction:

Searching for a lower dimensional house which will best represent the data. Removing the irrelevant information from the order dataset. Sampling can be accustomed to alter the method of classification using small dataset.

Applying algorithm:

At first the Decision Tree-CART algorithm was implemented to predict autism traits in an individual. For further improvement Random Forest-CART was implemented and better results were obtained. Finally, the Random Forest-CART classifier was modified to get improved results by merging it along with the Random Forest-ID3 classifier.

The three algorithms consecutively used to implement the system are discussed below:

Prediction model based on Decision Tree-CART:

Initially, the Decision Tree-CART classifier was selected to create the prediction model. At the beginning, the tree root consists of the whole dataset. Then data would be split using the best feature. The splitting process will continue recursively until a node consists of data of a unique label class. Sequential attribute selection method is resolved by Gini Impurity and Information Gain (IG) as shown in equation 1 and 2. Attributes with maximum IG will be chosen first to split data.

$$Gini(data) = 1 - X\, i \in unique\_classes\ P(i)\ 2 \quad \text{---------- (1)}$$

$$InfoGain\ (data, featureX) = Gini(data) - X\, i \in f\ eatureX\ AvgGini(i) \quad \text{---------- (2)}$$

Prediction model based on Random Forest:

CART In a random forest, each node is split using the best among a subset of predictors randomly chosen. This somewhat counter intuitive strategy turns out to perform very well compared to many other classifiers, including discriminant analysis, support vector machines and neural networks, and is robust against over-fitting. To make the predictive model more accurate, Random Forest-CART classifier was implemented. Here alsothe algorithm can be split into two phases: generating random forest and classifying test data.

Prediction model based on merging Random Forest-CART and Random Forest-ID3:

In order to improve the performance, a prediction model is proposed that merges the concept of random forest- CART with the concept of random forest - ID3 [Algorithm 3]. The algorithm for the proposed prediction model can be split into two phases like before: generating the merged random forest and classifying test data. Difference of it from 2 is that here randomnessisincreased more by generating and adding ID3 decision trees to therandom forest [In line 3-13]. Algorithm 3 tends to work better than Algorithm 2 because addition of ID3 decision trees limits overfitting and thus further reduces error compared to Algorithm 2.

Creating User Interface:

We have used a flask framework to design the front end. Syntax such apps. route is used to connect html pages. Separate html pages contain design layout of the webpage. Designing web pages is done using normal HTML tags. The session kept active using a secret key. After running the algorithm we have come to the conclusion that random forest base 99% accuracy.

DEVELOPING A CHATBOT:

A bot is a program that automatically completes an action based on specific triggers and algorithms. A chatbot is a computer program that's designed to simulate human conversation. Users communicate with these tools using a chat interface or via voice, just like they would converse with another person. Chatbots interpret the words given to them

by a person and provide a pre-set answer. Chatbots, like regular applications, have application layers, databases, conversational user interfaces (CUIs) and APIs.

V. REQUIREMENT AND SPECIFICATIONS

A System Requirements Specification (SRS) is a document that specifies the requirements for a system. It is used to describe the characteristics, capabilities, and constraints of a system, and to define the interfaces between the system and its stakeholders. It typically includes a description of the system's functions, performance, and behavior, as well as the requirements for hardware, software, and other resources which helps to ensure that the final system meets the needs of the users and other stakeholders.

Hardware Requirements :

Hardware refers to the physical components of a computer or other electronic system. These may include the processor, memory, storage devices, input/output devices, and other components that are necessary for the system to function. The specific hardware components that are used can have a significant impact on the performance and capabilities of the system.

1. Processor: Intel® Core™ - i5 or higher.
2. Memory RAM: 8GB or higher
3. Hard Disk Space: 512GB or more.
4. Architecture: 64-bit (x64) windows.

Processor: Intel® Core™ - i5 or higher:

Intel Core i5 is a family of mid-range processors produced by Intel, which offers a balance between performance and cost. The i5 processors are suitable for general-purpose computing tasks, such as web browsing, office applications, and light gaming, among others. They typically feature four to six cores, with clock speeds ranging from 2.4 to 4.2 GHz, and are compatible with a wide range of motherboards. However, for more demanding tasks such as video editing, 3D modeling, or heavy gaming, you may want to consider upgrading to an Intel Core i7 or i9 processor, which offer higher clock speeds, more cores, and better multi-threading capabilities. Ultimately, the choice of processor depends on your specific needs, budget, and the applications you intend to run on your computer.

Memory RAM: 8GB or higher:

RAM (Random Access Memory) is a type of computer memorythat temporarily stores data that the computer needs to access quickly. It is different from a hard drive in that it is much faster but has a much smaller storage capacity. When a computer runsa program, it loads the necessary data into the RAM. Having 8GB or higher of RAM means that the computer can store more data in its temporary memory, allowing it to run more applications simultaneously and improve overall performance. With 8 GB of RAM, you will have enough memory to run several programs at once. You can open lots of browser tabs at once, use photo or video editing programs, stream content, and play mid-to-high-end games. Many Windows 10 and macOS computers or laptops come with 8 GB of memory installed thesedays. For example, if you're using a computer for video editing or gaming, having 8GB or more of RAM can significantly improve the speed at which the computer processes data and reduces the likelihood of the computer slowing down or freezing. Therefore, having 8GB or higher of RAM is essential for running demanding applications and multitasking on a computer.

Autism spectrum disorder can be controlled if found at an early stage by advising people with proper medication. . This could prevent the patient’s condition from getting worse and would decrease long term costs related to delayed diagnosis. Thus an effective, accurate and easy screening check tool is highly required which would detect the traits in a person and recognize whether the person requires thorough autism syndrome assessment or not. In this paper we use machine learning to find out a set of conditions that are put together to be predictive of autism spectrum disorder. This can be immensely useful to physicians, assisting them to detect autism spectrum disorder at a very early stage.

Current explosion rate of autism around the world is numerous and it is increasing at a very high rate. According to WHO, about 1 out of every 160 children has ASD. Some people with this disorder can live independently, while others require life-long care and support. Diagnosis of autism requires a significant amount of time and cost. Earlier detection of autism can come to a great help by prescribing patients with proper medication at an early stage. It can

prevent the patient's condition from deteriorating further and would help to reduce long term costs associated with delayed diagnosis.

Software Requirements :

Software refers to the computer programs and related data that provide the instructions for telling a computer what to do and how to do it. Software is an essential part of any computer system, and it is responsible for the programs and applications that run on the system. Software is an integral part of any computer system, and it is responsible for enabling the system to perform a wide range of tasks and functions.

1. Microsoft Windows 10 or Higher
2. Jupyter Notebook
3. Anaconda Navigator
4. Visual Studio: IDE and Code Editor

Microsoft Windows 10 or Higher:

Microsoft Windows 10 is a computer operating system that helps you run software and perform tasks on your computer. It has a user-friendly interface that allows you to easily navigate and customize your settings. Windows 10 also comes with built-in security features to help protect your computer from viruses and malware. With Windows 10, you can use a variety of applications such as web browsers, email clients, and productivity tools to help you get your work done efficiently. Microsoft Windows 10 is the latest version of the Windows operating system. It features a user-friendly interface, improved security and productivity, virtual desktops, Cortana virtual assistant, and improved security and compatibility with a wide range of devices. Subsequent updates and versions have added additional features and enhancements. Additionally, Windows 10 has many accessibility features that make it easier for peoplewith disabilities to use computers.

Jupyter Notebook:

Jupyter Notebook is an open-source web application that allows users to create and share documents that contain live code, equations, visualizations, and narrative text. It supports multiple programming languages, including Python, R, and Julia, and allows users to create interactive data science workflows and reproducible research. Jupyter Notebook runs on a web browser and provides an intuitive interface for code editing, debugging, and visualization. It also supports the creation of rich media content and interactive widgets. Jupyter Notebook has become a popular tool for data scientists and researchers due to its flexibility, interactivity, and ability to promote reproducible research practices.

Anaconda Navigator:

Anaconda Navigator is a graphical user interface (GUI) that comes with the Anaconda distribution for Python programming. It provides an easy-to-use interface for managing packages, environments, and Jupyter Notebooks. With Anaconda Navigator, you can install, update, and remove packages, as well as create and manage virtual environments for your Pythonprojects. It also includes a suite of popular data science packages, such as NumPy, Pandas, Matplotlib, and SciPy. In addition, Anaconda Navigator allows you to launch Jupyter Notebooks, a web-based interactive computing environment that is widely used in data science and scientific research. Overall, Anaconda Navigator is a valuable tool for Python developers and data scientists.

Visual Studio: IDE and Code Editor:

Visual Studio is an integrated development environment (IDE) and code editor that provides a comprehensive set of tools and features for building and debugging software applications. The IDE includes a code editor with syntax highlighting, codecompletion, and code navigation features to help developers write code efficiently. It also provides project management tools, version control integration, and a debugging environmentto help developers identify and fix errors in their code. The IDE supports multiple programming languages, including C++, C#, and Visual Basic, and enables developers to build applications for a variety of platforms, including Windows, iOS, and Android. Overall, Visual Studio is a powerful tool for software development that streamlines the development process and improves code quality.

VI. IMPLEMENTATION

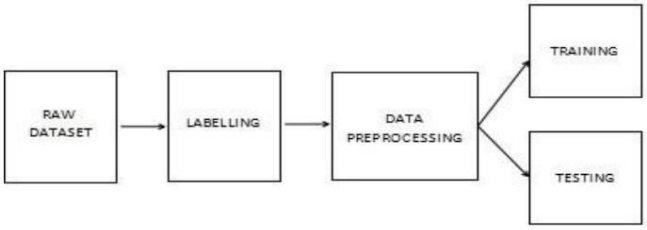

Fig 1: Data Pre-processing

Raw Dataset:

Raw data, also known as primary data, are data (e.g., numbers, instrument readings, figures, etc.) collected from a source. In the context of examinations, the raw data might be described as a raw score.

Labelling:

In machine learning, data labeling is the process of identifying raw data (images, text files, videos, etc.) and adding one or more meaningful and informative labels to provide context so that a machine learning model can learn from it.

Data Preprocessing:

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviours or trends, and is likely to contain many errors.

Training:

The model is initially fit on a training dataset, which is a set of examples used to fit the parameters of the model. The model is trained on the training dataset usinga supervised learning method, for example using optimization methods such as gradient descent or stochastic gradient descent. In practice, the training dataset often consists of pairsof an input vector (or scalar) and the corresponding output vector (or scalar), where the answer key is commonly denoted as the target (or label). The current model is run with the training dataset and produces a result, which is then compared with the target, for each input vector in the training dataset. Based on the result of the comparison and the specific learning algorithm being used, the parameters of the model are adjusted. The model fitting can include both variable selection and parameter estimation.

Testing:

A test dataset is a dataset that is independent of the training dataset, but that follows the same probability distribution as the training dataset. If a model fit to the training dataset also fits the test dataset well, minimal overfitting has taken place. A better fitting of the training dataset as opposed to the test. dataset usually points to overfitting. A test set is therefore a set of examples used only to assess the performance (i.e. generalization) of a fully specified classifier. To do this, the final model is used to predict classifications of examplesin the test set. Those predictions are compared to the examples' true classifications to assess the model's accuracy.

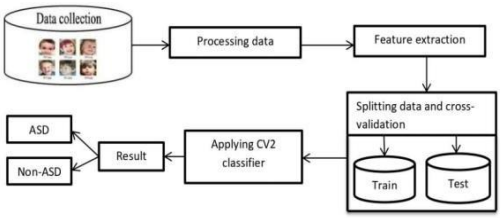

Fig 2: Implementation

In the process of prediction initially, the patient/user will input the data for AQ10 dataset which has 20 attributes in total where 10 general questions and other 10 attributes are personal questions like age, country etc. This is provided by the doctor or any medical assistant. After filling up the data given by the user, several required machine learningalgorithms were used to predict whether the patient/user had ASD traits or not.

VII. RESULT AND DISCUSSION

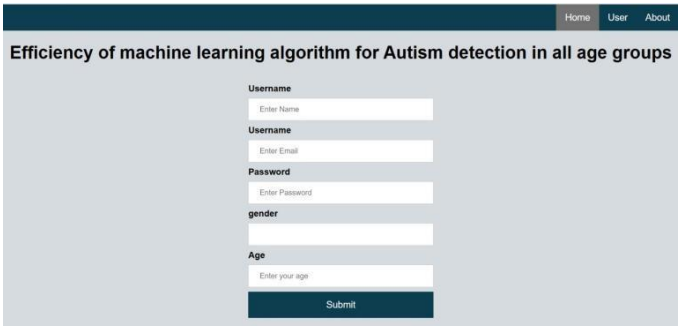

Fig 3: Register

A GUI application as been developed in which the user has to register using username, email-id, password, gender, age.

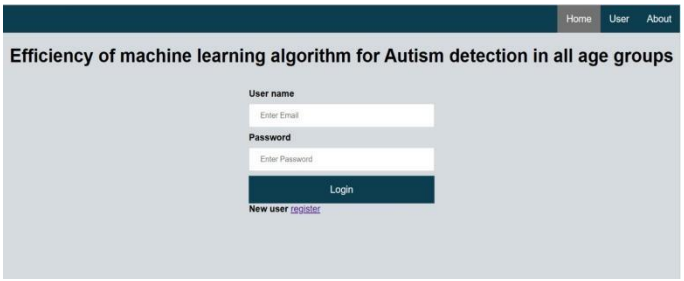

Fig 4: Login

A GUI application as been developed in which the user has to login using username and password. After login it will take direct to main page.

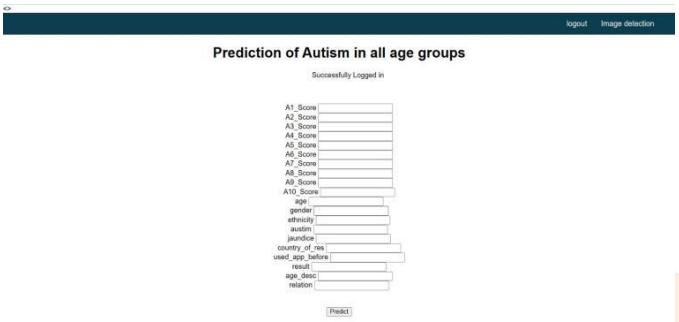

Fig 5: Main Page

The above page represents the main page where the user can choose the questions or image processing.

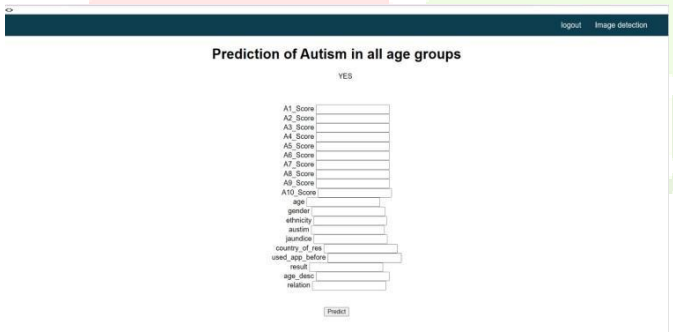

Fig 6: Predicting Autism using set of questions

In this above figure we give answers to the set of questions. Based on the answers the Chat-Bot predicts wether the user has Autism or not.

Fig 7: Choosing file to predict Autism

In the above figure we can choose the image to predict wether the user has Autism or not.

Fig 8: Predicting Autism using image.

In the above figuer after the user choose the image, the image is processed and prediction is given.

VII. CONCLUSION

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that affects social interaction, communication, and behavior. It is a spectrum disorder, which means that the severity of symptoms and challenges can vary greatly among individuals with ASD.Research has shown that there is no single cause of ASD, but rather a combination of genetic and environmental factors. While there is no known cure for ASD, early intervention and treatment can help individuals with ASD lead fulfilling and productive lives.Individuals with ASD may face challenges in social interaction, communication, and behavior, but they also have unique strengths and abilities. It is important to recognize and appreciate these strengths, and to work to create a more inclusive and supportive society for individuals with ASD.Overall, ASD is a complex and varied disorder that requires individualized support and understanding. With the right resources and support, individuals with ASD can thrive and make valuable contributions to their communities.

REFERENCES

[1] U. Frith and F. Happé, “Autism spectrum disorder,” Current biology, vol. 15, no. 19, pp. R786–R790, 2005.

[2] WHO, Autism spectrum disorders, 2017 [Accessed August 22, 2018]. Available: http://www.who.int/news-room/fact-sheets/detail/ autism-spectrum-disorders.

[3] J. A. Cruz and D. S. Wishart, “Applications of machine learning in cancer prediction and prognosis,” Cancer informatics, vol. 2, 2006.

[4] N. S. Khan, M. H. Muaz, A. Kabir, and M. N. Islam, “Diabetes predicting mhealth application using machine learning,” in 2017 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE). IEEE, 2017, pp. 237–240.

[5] D. P. Wall, R. Dally, R. Luyster, J.-Y. Jung, and T. F. DeLuca, "Use of artifificial intelligence to shorten the behavioral diagnosis of autism," PloS one, vol. 7, no. 8, p. e43855, 2012.

[6] D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S. Narayanan, "Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, effificiency, and multi-instrument fusion," Journal of Child Psychology and Psychiatry, vol. 57, 2016.

[7] C. Allison, B. Auyeung, and S. Baron-Cohen, "Toward brief "red flags" for autism screening: the short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls," Journal of the American Academy of Child Adolescent Psychiatry, vol. 51, 2012.

[8] F. Thabtah, "autism spectrum disorder screening: machine learning adaptation and dsm5 fulfillment," in Proceedings of the 1st International Conference on Medical and Health Informatics 2017. ACM, 2017.

[9] F. Hauck and N. Kliewer, "Machine learning for autism diagnostics: Applying support vector classification."

[10] B. van den Bekerom, "Using machine learning for detection of autism spectrum disorder," 2017.

[11] D. Wall, J. Kosmicki, T. Deluca, E. Harstad, and V. Fusaro, "Use of machine learning to shorten observation-based screening and diagnosis of autism," Translational psychiatry, vol. 2, no. 4, p. e100, 2012.

[12] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identifification of autism spectrum disorder using deep learning and the abide dataset," NeuroImage: Clinical, vol. 17, 2018.

[13] W. Liu, M. Li, and L. Yi, "Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework," Autism Research, vol. 9, no. 8, pp. 888–898, 2016.

[14] D. Bone, M. S. Goodwin, M. P. Black, C.-C. Lee, K.Audhkhasi, and S. Narayanan, "Applying machine learning to facilitate autism diagnostics: pitfalls and promises," Journalof autism and developmental disorders, vol. 45, no. 5, pp. 1121–1136, 2015.

[15] J. Kosmicki, V. Sochat, M. Duda, and D. Wall, "Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning," Translational psychiatry, vol. 5, no. 2, p. e514, 2015.

[16] T. Booth, A. L. Murray, K. McKenzie, R. Kuenssberg, M. O'Donnell, and H. Burnett, "An evaluation of the aq-10 as a brief screening instrument for asd in adults."

[17] A. Liaw, M. Wiener et al., "Classification and regression by random forest," R news, vol. 2, no. 3, pp. 18–22, 2002.