JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EARLY-STAGE DETECTION OF AUTISM SPECTRUM DISORDER USING MACHINE LEARNING FRAMEWORK

[1] S.Vinayagan, [2] Ajila Jaya Reeta C, [3] Shiny Malar .F.R

[1] Student, [2]Assistant Professor, [3]Associate Professor,
[1&2] Department of Computer Science and Engineering, Lord Jegannath College of Engineering and Technology, Ramanathichanputhur
[3] Department of Computer Science and Engineering, Stella Mary's College of Engineering, Aruthenganvilai, Nagercoil, India

Abstract: Autism Spectrum Disorder (ASD) is a developmental disability caused by differences in the brain. People with ASD often have problems with social communication and interaction, and restricted or repetitive behaviors or interests. People with ASD may also have different ways of learning, moving, or paying attention. In this project, we present a novel approach for the detection of ASD using machine learning techniques, implemented in Python. We employed two distinct algorithms, namely the Random Forest Classifier and the Decision Tree Classifier, to analyze a dataset containing some records with 21 features. The dataset includes a diverse range of attributes, such as sensory perception, cognitive abilities, demographics, and medical history, which are potentially indicative of ASD. Our model's performance on this dataset is a testament to the power of machine learning inhealthcare applications.The Random Forest Classifier achieved remarkable results with a training accuracy score of 100% and a testing accuracy score of 99%. This indicates that the model can effectively learn from the training data and generalize well to unseen cases. The Decision Tree Classifier, while achieving a training accuracy score of 100%, maintained a testing accuracy score of 96%, showcasing robust performance. The dataset used in this project encompasses a comprehensive set of attributes, including sensory perception (A1_Score - A6_Score), cognitive abilities (A7_Score - A10_Score), age, gender, ethnicity, parental medical history (jundice), autism diagnosis (austim), country of residence, prior app usage, and various demographic features. The extensive range of attributes ensures that our model takesinto account a multitude of factors when making predictions.

***Index Terms* - Autism Spectrum Disorder, Machine Learning, Forest Classifier and Decision Tree Classifier, Feature Extraction, Classification, Convolutional Neural Networks.**

I. INTRODUCTON

Autism Spectrum Disorder, commonly referred to as ASD, is a complex neurodevelopment condition that affects the way an individual perceives and interacts with the world. ASD is characterized by a wide range of symptoms and behaviors, and it is often described as a "spectrum" because the severity and manifestations of the disorder can vary significantly from person to person. This condition typically emerges in early childhood and persists throughout a person's lifetime.

ASD is distinguished by challenges in three core areas:

Social Interaction: Individuals with ASD often experience difficulties in understanding and responding to social cues, making it challenging to establish and maintain meaningful relationships. They may struggle with interpreting emotions, maintaining eye contact, and engaging in reciprocal conversations.

Communication: Impaired communication is a hallmark of ASD. This can manifest as delayed speech development, limited vocabulary, difficulty with non- verbal communication (such as gestures and facial expressions), and a tendency to engage in repetitive language patterns or echolalia.

Repetitive Behaviors and Interests: Individuals with ASD may exhibitrepetitive or restrictive behaviors, such as intense interests in specific topics, a preference for routines and sameness, and repetitive movements or gestures. These behaviors can serve as a source of comfort or self-regulation for some individuals.

ASD is a highly prevalent condition, with a significant impact on individuals, families, and communities. It is estimated that approximately 1 in 54 children in the United States is diagnosed with ASD, making it one of the most prevalent neurodevelopment disorders. The condition affects individuals of all backgrounds, regardless of race, gender, or socioeconomic status.

Diagnosing ASD involves a comprehensive assessment that considers a person's behavior, developmental history, and communication skills. Early diagnosis and intervention are crucial, as they can lead to improved outcomes and quality of life for individuals with ASD.

The causes of ASD are multifactorial and still the subject of ongoing research. Genetic, environmental, and neurobiological factors are thought to contribute to the development of the disorder. Despite the challenges associated with ASD, many individuals with the condition have unique strengths and abilities that can be harnessed and celebrated.

Understanding ASD and advancing research and interventions in this field are critical in providing support and enhancing the lives of individuals with autism. This introduction provides a broad overview of the condition, setting the stage for a deeper exploration of its characteristics, diagnosis, treatments, and the ongoing efforts to improve the well-being of those living with ASD.

1.1 MACHINE LEARNING

Machine Learning is a system of computer algorithms that can learn from example through self-improvement without being explicitly coded by a programmer. Machine learning is a part of artificial Intelligence which combines data with statistical tools to predict an output which can be used to make actionableinsights.

The breakthrough comes with the idea that a machine can singularlylearn from the data (i.e., example) to produce accurate results. Machine learning is closely related to data mining and Bayesian predictive modeling. The machine receives data as input and uses an algorithm to formulate answers.

A typical machine learning tasks are to provide a recommendation. For those who have a Netflix account, all recommendations of movies or series are based on the user's historical data. Tech companies are using unsupervised learning to improve the user experience with personalizing recommendation.

1.2 ARTIFICIAL INTELLIGENCE

In the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what was then termed "neural networks"; these were mostly perceptions and other models that were later found to be reinventions of the generalized linear models of statistics. Probabilistic reasoning was also employed, especially in automated medical diagnosis.

However, an increasing emphasis on the logical, knowledge-based approach caused a rift between AI and machine learning. Probabilistic systems were plagued by theoretical and practical problems of data acquisition and representation. By 1980, expert systems had come to dominate AI, and statistics was out of favor. Work on symbolic/knowledge-based learning did continue within AI, leading to inductive logic programming, but the more statistical line of research was now outside the field of AI proper, in pattern recognition and information retrieval. Neural networks research had been abandoned by AI and computer science around the same time. This line, too, was continued outside the AI/CS field, as "connectionism", by researchers from other disciplines including Hopfield, Rumelhart and Hinton. Their main success came in the mid-1980s with the reinvention of back propagation. Machine learning (ML), reorganized as a separate field, started to flourish in the 1990s. The field changed its goal from achieving artificial intelligence to tackling solvable problems of a practical nature. It shifted focus away from the symbolic approaches it had inherited from AI, and toward methods and models borrowed from statistics and probability theory.

As of 2020, many sources continue to assert that machine learning remains a subfield of AI. The main disagreement is whether all of ML is part of AI, as this would mean that anyone using ML could claim they are using AI. Others have the view that not all of ML is part of AI where only an 'intelligent' subset of ML is part of AI.

2 METHODOLOGIES:

2.1 Data Collection

In the first module of A Machine Learning Framework for Early-Stage Detection of Autism Spectrum Disorders, we developed the system to get the input dataset. Data collection process is the first real step towards the real development of a machine learning model, collecting data. This is a critical step that will cascade in how good the model will be, the more and better data that we get; the better our model will perform. There are several techniques to collect the data, like web scraping, manual interventions. Our dataset is placed in the project and it's located in the model folder. The dataset is referred from the popular standard dataset repository kaggle where all the researchers refer it. The dataset consists of numerical data. The following is the URL for the dataset referred from kaggle.
Dataset

The dataset consists of 704 individual data. There are 21 columns in the dataset, which are described below.

A1_Score: I often notice small sounds when others do not
A2_Score: I usually concentrate more on the whole picture, rather than the small details
A3_Score: I find it easy to do more than one thing at once
A4_Score: If there is an interruption, I can switch back to what I was doing very quickly
A5_Score: I find it easy to 'read between the lines' when someone is talking to me
A6_Score: I know how to tell if someone listening to me is getting bored
A7_Score: When I'm reading a story I find it difficult to work out the characters' intentions
A8_Score: I like to collect information about categories of things (e.g. types of car, types of bird, types of train, types of plant etc)
A9_Score: I find it easy to work out what someone is thinking or feeling just by looking at their face
A10_Score: I find it difficult to work out people's intentions
Age: Age number
Gender: Female and male
Ethnicity: Types of ethnicity
Jundice: No or yes
Austim: No or yes
Contry_of_res: Types of country of res
Used_app_before: No or yes
Result: Result range
Age_desc: 18 and more
Relation: Types of relation

Class/ASD: No or yes

2.2. Data Preparation:

Wrangle data and prepare it for training. Clean that which may require it (remove duplicates, correct errors, deal with missing values, normalization, data type conversions, etc.). Randomize data, which erases the effects of the particular order in which we collected and/or otherwise prepared our data. Visualize data to help detect relevant relationships between variables or class imbalances (bias alert!), or perform other exploratory analysis. Split into training and evaluation sets

2.3. Random Forest Classifier:

Random Forest Classifier machine learning algorithm, an accuracy of 100% on train set so we implemented this algorithm. Let's understand the algorithm in layman's terms. Suppose you want to go on a trip and you would like to travel to a place which you will enjoy. So what do you do to find a place that you will like? You can search online, read reviews on travel blogs and portals, or you can also ask your friends.

Let's suppose you have decided to ask your friends, and talked with them about their past travel experience to various places. You will get some recommendations from every friend. Now you have to make a list of those recommended places. Then, you ask them to vote (or select one best place for the trip) from the list of recommended places you made. The place with the highest number of votes will be your final choice for the trip. In the above decision process, there are two parts. First, asking your friends about their individual travel experience and getting one recommendation out of multiple places they have visited. This part is like using the decision tree algorithm. Here, each friend makes a selection of the places he or she has visited so far.

The second part, after collecting all the recommendations, is the voting procedure for selecting the best place in the list of recommendations. This whole process of getting recommendations from friends and voting on them to find the best place is known as the random forests algorithm. It technically is an ensemble method (based on the divide-and-conquer approach) of decision trees generated on a randomly split dataset. This collection of decision tree classifiers is also known as the forest. The individual decision trees are generated using an attribute selection indicator such as information gain, gain ratio, and Gini index for each attribute. Each tree depends on an independent random sample. In a classification problem, each tree votes and the most popular class is chosen as the final result. In the case of regression, the average of all the tree outputs is considered as the final result. It is simpler and more powerful compared to the other non-linear classification algorithms.

2.4 Decision Tree Classifier

The decision tree classifier machine learning algorithm, an accuracy of 100% on train set so we implemented this algorithm.

2.4.1. Decision Tree Classification Algorithm

Decision Tree is a supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome.

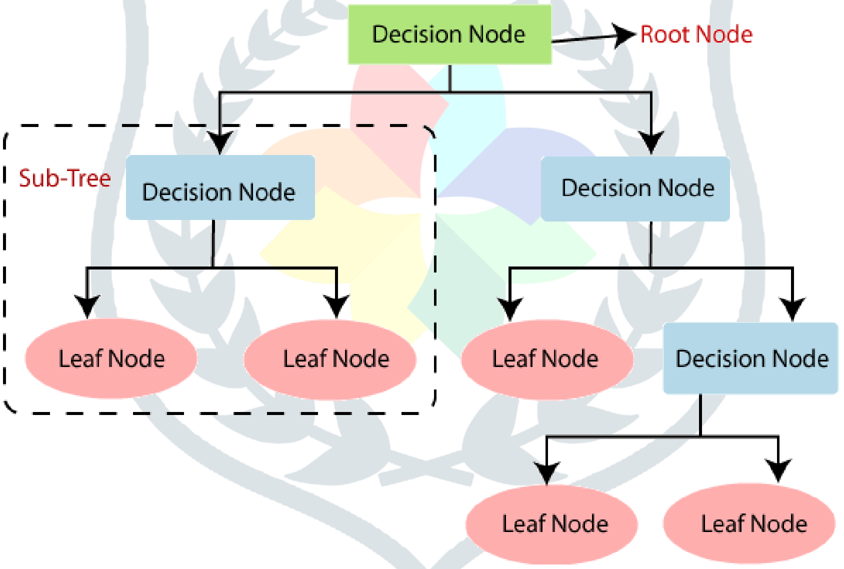


Fig 2.4.1. Decision Tree Classification Algorithm

In the figure 2.4.1illustrated a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset.

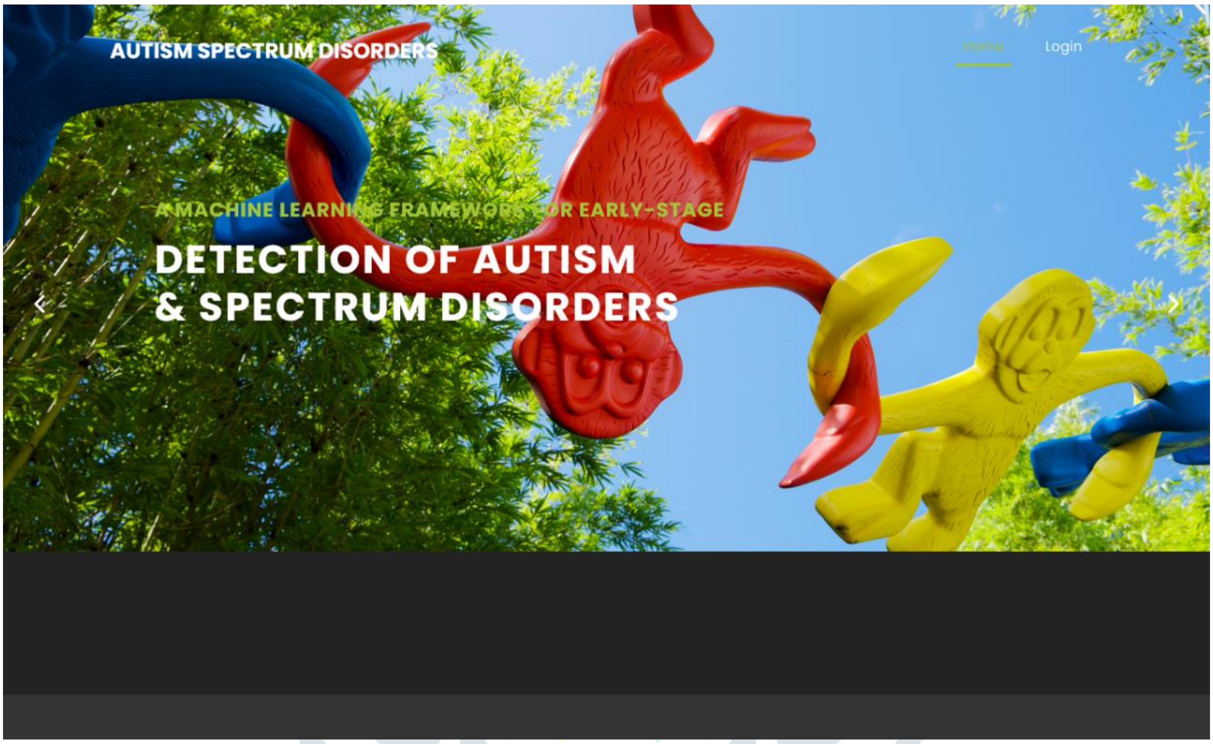
It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure. In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into sub-trees.

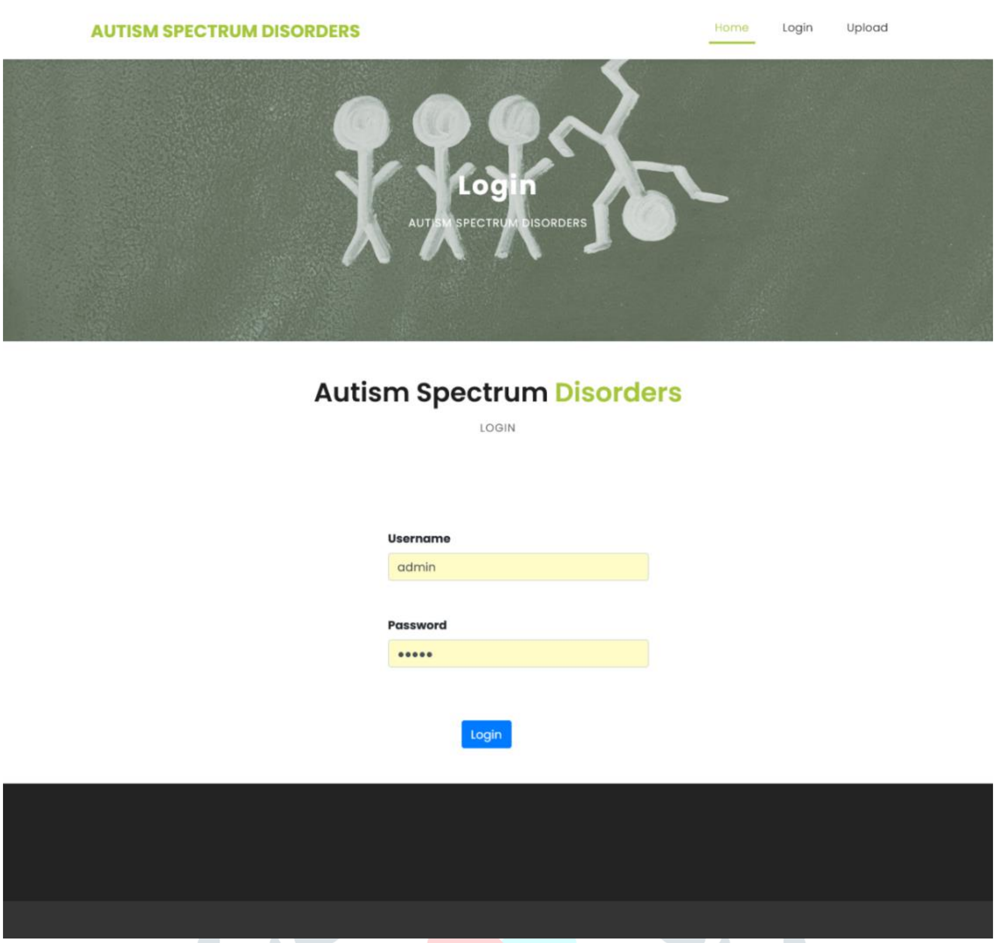
3. EXPERIMENTAL RESULT

Detection of Autism Spectrum Disorder (ASD) using Python-based machine learning models, specifically the Random Forest Classifier and the Decision Tree Classifier. The project builds upon the existing system, which achieved a notable accuracy of 97.95% using the Ada Boost algorithm, by addressing its limitations and enhancing the diagnostic process.

The proposed system exhibits exceptional accuracy, with the Random Forest Classifier achieving a training accuracy score of 100% and a testing accuracy score of 99%, while the Decision Tree Classifier maintains a training accuracy of 100% and a testing accuracy of 96%. These results underscore the system's capability to distinguish between individuals with ASD and those without, with high precision.

A key strength of the proposed system lies in its rich feature set, which encompasses sensory perception, cognitive abilities, demographics, medical history, and other relevant attributes. This comprehensive approach allows for a more holistic assessment of individuals, enabling the model to capture complex relationships between variables. Early diagnoses of ASD are of paramount importance for timely intervention and support. The proposed system facilitates early detection, which can significantly improve the quality of life for affected individuals. Furthermore, the system's consideration of diverse demographic attributes ensures its applicability across different populations and demographic groups. The interpretability of the Decision Tree Classifier provides healthcare professionals with insights into the model's decision-making process, enhancing their understanding of the diagnostic process.





CONCLUSION

In conclusion, the proposed system marks a significant advancement in the field of healthcare and machine learning for ASD detection. Its high accuracy, robust generalization, rich feature set, early diagnosis capabilities, interpretability, and clinical relevance collectively position it as a valuable tool for healthcare professionals working with individuals on the autism spectrum. This project exemplifies the potential of machine learning in improving the diagnosis of complex neurodevelopment conditions and contributes to the broader goal of enhancing the quality of life for individuals with ASD.

REFERENCES

[1]. M. Bala, M. H. Ali, M. S. Satu, K. F. Hasan, and M. A. Moni, ``Efficientmachine learning models for early stage detection of autism spectrumdisorder," Algorithms, vol. 15, no. 5, p. 166, May 2022.

[2]. D. Pietrucci, A. Teofani, M. Milanesi, B. Fosso, L. Putignani, F. Messina,G. Pesole, A. Desideri, and G. Chillemi, ``Machine learning data analysishighlights the role of parasutterella and alloprevotella in autism spectrumdisorders," Biomedicines, vol. 10, no. 8, p. 2028, Aug. 2022.

[3]. R. Sreedasyam, A. Rao, N. Sachidanandan, N. Sampath, andS. K. Vasudevan, ``AaryaA kinesthetic companion for childrenwith autism spectrum disorder," J. Intell. Fuzzy Syst., vol. 32, no. 4,pp. 2971-2976, Mar. 2017.

[4]. J. Amudha and H. Nandakumar, ``A fuzzy based eye gaze pointestimation approach to study the task behavior in autism spectrumdisorder," J. Intell. Fuzzy Syst., vol. 35, no. 2, pp. 1459-1469,Aug. 2018.

[5]. H. ChahkandiNejad, O. Khayat, and J. Razjouyan, ``Software developmentof an intelligent spirography test system for neurological disorder detectionand quantification," J. Intell. Fuzzy Syst., vol. 28, no. 5, pp. 2149- 2157,Jun. 2015.

[6]. F. Z. Subah, K. Deb, P. K. Dhar, and T. Koshiba, ``A deep learning approachto predict autism spectrum disorder using multisite resting-state fMRI,"Appl. Sci., vol. 11, no. 8, p. 3636, Apr. 2021.

[7]. K.-F. Kollias, C. K. Syriopoulou-Delli, P. Sarigiannidis, and G. F. Fragulis,``The contribution of machine learning and eye-tracking technology inautism spectrum disorder research: A systematic review," Electronics,vol. 10, no. 23, p. 2982, Nov. 2021.

[8]. I. A. Ahmed, E. M. Senan, T. H. Rassem, M. A. H. Ali, H. S. A. Shatnawi,S.M. Alwazer, and M. Alshahrani, ``Eye tracking-based diagnosis andearly detection of autism spectrum disorder using machine learning anddeep learning techniques," Electronics, vol. 11, no. 4, p. 530, Feb. 2022.

[9]. P. Sukumaran and K. Govardhanan, ``Towards voice based prediction andanalysis of emotions in ASD children," J. Intell. Fuzzy Syst., vol. 41, no. 5,pp. 5317-5326, 2021.

[10]. S. P. Abirami, G. Kousalya, and R. Karthick, ``Identification and explorationof facial expression in children with ASD in a contact less environment,"J. Intell. Fuzzy Syst., vol. 36, no. 3, pp. 20332042, Mar. 2019.

[11]. M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, ``Detecting autismspectrum disorder using machine learning techniques," Health Inf. Sci.Syst., vol. 9, no. 1, pp. 1-13, Dec. 2021.

[12]. C. Allison, B. Auyeung, and S. Baron-Cohen, ``Toward brief `red flags'for autism screening: The short autism spectrum quotient and the shortquantitative checklist in 1,000 cases and 3,000 controls," J. Amer. Acad.Child Adolescent Psychiatry, vol. 51, no. 2, pp. 202-212, 2012.

[13]. F. Thabtah, F. Kamalov, and K. Rajab, ``A new computational intelligenceapproach to detect autistic features for autism screening," Int. J. Med.Inform., vol. 117, pp. 112-124, Sep. 2018.

[14]. M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, andM. A. Moni, ``Heart disease prediction using supervised machine learningalgorithms: Performance analysis and comparison," Comput. Biol. Med.,vol. 136, Sep. 2021, Art. no. 104672.

[15]. E. Dritsas and M. Trigka, ``Stroke risk prediction with machine learningtechniques," Sensors, vol. 22, no. 13, p. 4670, Jun. 2022.

[16]. V. Chang, J. Bailey, Q. A. Xu, and Z. Sun, ``Pima Indians diabetes mellitusclassification based on machine learning (ML) algorithms," NeuralComput. Appl., early access, pp. 1-17, Mar. 2022.

[17]. F. Thabtah, ``Machine learning in autistic spectrum disorder behavioralresearch:Areviewandways forward," Inform. Health Social Care, vol. 44,no. 3, pp. 278-297, 2018.

[18]. K. S. Omar, P. Mondal, N. S. Khan, M. R. K. Rizvi, and M. N. Islam,``A machine learning approach to predict autism spectrum disorder," inProc. Int. Conf. Electr., Comput.Commun. Eng. (ECCE), Feb. 2019,pp. 1-6.