International Journal Of Public Mental Health And Neurosciences
ISSN: 2394-4668
(An Official publication of Sarvasumana Association)

Deep Learning Approach for Prediction of Autism Spectrum Disorder

Preenon Bagchi [1,2,3], Gauri Hiremath [4], Nandini Ghuge [4], Aman Tiple [3]*, and Shivani V [2]

[1]Dept. of Computer Science, Madhav University, Abu Road, Rajasthan-307026
[2] Karnataka Sanskrit University, Pampa Mahakavi Road, Chamarajapet, Bengaluru – 560018
[3]Vasishth Genomics Research Lab Pvt.Ltd., Bangalore-560085
[4] Rajiv Gandhi Institute of Information Technology & Biotechnology, Bharti Vidyapeeth Deemed University, Pune, Maharashtra 411045
*Corresponding Author: amantiple97@gmail.com

Abstract— Autism Spectrum Disorder (ASD) is a neurodevelopmental state with complex cerebral pattern, observed by patients facing a variety of challenges in social interaction, verbal communication, confined behaviours and speech. The impacts of this disorder and the inflexibility of symptoms vary from person to person. The symptoms appear at the age of 2 to 5 and continue throughout adolescence to the adulthood. While ASD cannot be cured fully, studies have shown that early discovery of this pattern can help in maintaining the behavioural and cerebral development of children. The present literature on ASD is mainly focuses on prediction system based on traditional deep learning algorithm i.e. Sequential model with multiple layers. The proposed models are validated by using performance dimension parameters similar as accuracy, precision and recall. In this exploration, autism spectrum disorder prediction has been delved and compared using common parameters similar as operation type, simulation system, comparison methodology, and input data. This research paper revolves around the use of deep learning algorithm to predict ASD, noticing the significance of early diagnosis for effective intervention.

Index Terms— ***ASD Prediction, Deep Neural Network, Early Diagnosis***

I. INTRODUCTION

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder that affects communication, social engagement, behavioral patterns and cognitive maturation. ASD has a wide range of symptoms, which is why it is called a "spectrum disorder." People with ASD have varying degrees of impairment, making diagnosis and effective treatment difficult [1,2].

The precise cause of ASD is still not completely recognized, but studies indicate that an association of both genetic and environmental factors is significant. Some of the known factors that contribute to the development of ASD include several genes that have been identified as potential contributors to ASD [3]. It is thought that changes in these genes impact the development and operation of the brain, resulting in the defining traits of ASD. Environmental factors can also affect the likelihood of developing ASD in the period before and after birth. Some of these factors consist of maternal infections while pregnant, exposure to specific toxins, and problems during childbirth [4].

Studies utilizing brain imaging techniques have identified variations in brain structure and operation among people with ASD. These variances influence the processing and integration of information, resulting in noticeable behavioral trends [5].Typical symptoms of an introverted individual with a medical condition include lack of eye contact, limited range of interests or intense focus on specific topics, repetitive behaviors like rocking back and forth, repeating words or phrases, or flicking a switch, preferences for textures, smells, or visuals that may seem ordinary to others, significant sensitivity to sounds, and difficulty listening or looking at others.

The current rapid spread of mental illness worldwide is diverse and rapidly increasing. Some people with this disorder can live independently, while a small number of individuals need ongoing care and assistance [6]. Over the last 20 years, there has been a notable rise in the global occurrence of autistic spectrum disorder (ASD), currently affecting around 1 in 100 children, with significant differences seen across different areas. In the year 2020, an estimated 2.8% of American children who were 8 years old were thought to have ASD. Timely detection of ASD is crucial in order to facilitate prompt interventions that can enhance results and enable children to reach their maximum capabilities [7]. There is significance of various factors such as age, gender, ethnicity, and geographic location in predicting ASD, providing insights into potential risk factors.

The early detection of autism through various clinical methods is possible, although these diagnostic processes are rarely performed unless there is a high likelihood of ASD development. Machine learning (ML) allows for training ASD models faster and with greater precision. ML methods are essential for promptly and accurately evaluating ASD risk and simplifying the diagnostic process, helping families access crucial therapies sooner [8,9]. Different ML classification models can be utilized for early autism prediction to avoid its long-term impacts on both adults and children [10]. Machine Learning techniques are becoming a hopeful remedy to tackle this issue in the healthcare sector. By utilizing algorithms like Support Vector Classifier, Logistic Regression, Random Forest, and Voting classifier. Our goal is to greatly speed up the ASD diagnosis process.

We are currently in a time period known as the big data era, where various fields of science and industry produce vast quantities of data. This presents us with new challenges when it comes to analyzing and interpreting them. Therefore, there is a pressing demand for new machine learning and artificial intelligence techniques that can assist in making use of this data [11]. Novel computational methods in machine learning, such as mathematical learning, statistical estimation, and information theories, autonomously uncover valuable patterns in extensive datasets.

This approach offers the benefit of precise and dependable forecasting utilizing data with extensive variables and causal inference within non-experimental datasets. Recent research in the field of psychiatry has demonstrated the successful application of these methods in diagnosing ASD.[12]. Significantly, machine learning has effectively been used in various application challenges.

With the rise of machine learning techniques in different sectors, professionals in the medical field can now receive help in detecting ASD at an early stage. Identifying and focusing on the key features that lead to accurate prediction of ASD is a challenging task for practitioners, highlighting the necessity of implementing an automated method. Moreover, the existence of various symptoms associated with ASD in young children leads to the development of a comprehensive dataset of characteristics [13]. By using this dataset and the deep learning approach, this research creates a model to identify patterns in autism.

The objectives of this study are:

i. Develop a reliable predictive model to identify ASD in individuals at an early stage, enabling timely intervention and support. healthcare professionals in diagnosing ASD more efficiently and consistently.

ii. Analyze the significance of various factors such as age, gender, ethnicity, and geographic location in predicting ASD, providing insights into potential risk factors.

iii. Detect patterns and trends (based on age, gender, ethnicity, and geographic settings) in the dataset that could lead to a better understanding of ASD, potentially uncovering new avenues for research and treatment

II. Methodology

Dataset: The model was built and trained on the data from the .csv file which contained the responses to 10 screening questions, each having binary response either '0' or '1'. It also has certain features like gender, age, if born with Jaundice or not, country of residence.

Data preprocessing: It is a method within data mining which converts raw data into a more practical and effective format. The data is processed to remove impurities before use to train the model. It includes:

- Null Values: Removal of the observations with the missing values to complete the dataset.

- String to integer conversion: To facilitate the model training, categorial string values is converted into the numerical values.

- Normalization: Normalize numerical features (like age) to a consistent scale, typically between 0 and 1, to improve model performance.

Data splitting: Data is split into 80-20 split. The data is divided into training data, 80% and testing data, 20%. The Model Selection and Architecture can be summarized as:

- Deep Learning algorithms were used for training

- The sequential model was built using Keras with a TensorFlow backend

- It consisted of Input layer: matched to the number of input features. First hidden layer: 32 neurons with ReLU activation. Second hidden layer: 16 neurons with ReLU activation. Output layer: 1 neuron with sigmoid activation, suitable for binary classification. The models were trained using the dataset.

Evaluation the model performance by using remaining 20% testing dataset. The model predictions were compared against the actual labels using a confusion matrix and a classification report, which included precision, recall, and F1-score for each class. These metrics gave a more comprehensive understanding of the model's ability to correctly classify both ASD and non-ASD cases. The trained model was used to predict the ASD for the new data points.

The Results of the EDA was Visualized by various plots. The common visualization techniques used to analyze results were Scatter Plots, to visualize relationships between variables (e.g., age vs. ASD prediction); Heat Maps, To show the density of data points across different categories (e.g., country and age) and Histograms: To display frequency distributions of variables (e.g., age distribution within countries).

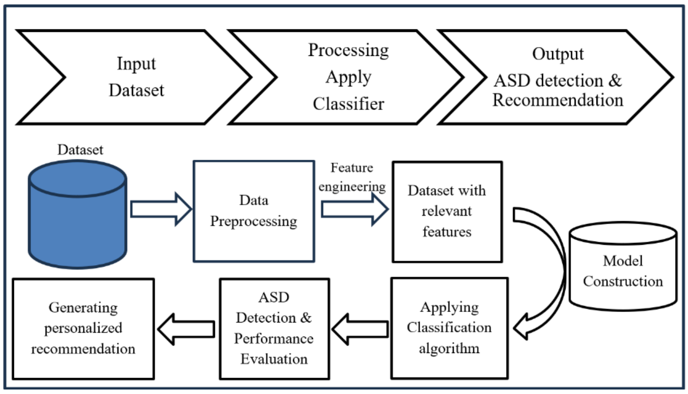

Fig. 1. Workflow of the Study

III. Results

The deep learning model we built for predicting Autism Spectrum Disorder (ASD) showed quite encouraging results. We used a simple, yet efficient Multilayer Perceptron (MLP) architecture. It had two hidden layers: the first with 32 neurons and the second with 16, both activated using the ReLU function. The final layer had just one neuron with a sigmoid activation, which suited our binary classification problem well

We trained the model using the Adam optimizer, and used binary cross entropy as the loss function. The training spanned 100 epochs, and the learning process was generally smooth—the loss decreased steadily while the accuracy improved with each epoch. There weren't any major signs of overfitting, which is always a good sign in deep learning experiments.

After the training was completed, we evaluated the model on the test dataset. The model achieved:

- Training Accuracy: 92.31%
- Testing Accuracy: 89.77%

To further understand the model's prediction capability, we generated a classification report. The report showed strong metrics for both ASD and non-ASD categories. Here's a breakdown of the results:

Table 1. Classification Report for the Model

Class	Precision	Recall	F1-Score
ASD	0.90	0.87	0.88
Non-ASD	0.89	0.92	0.90

These metrics suggest that the model did a good job detecting individuals with ASD (high recall) while also keeping false positives low (high precision). The F1-score, which balances precision and recall, remained strong and consistent across both classes. We also examined the confusion matrix, which gives a more detailed picture of the model's predictions:

Table 2. Confusion Matrix

	Predicted ASD	Predicted non-ASD
Actual ASD (88)	78	12
Actual non-ASD (113)	9	103

As seen above, the model correctly identified 78 out of 88 ASD cases and 103 out of 108 non-ASD cases, which is quite promising. There were only 12 false negatives and 9 false positives, which are tolerable in initial screening models.

Overall, these results suggest that even a relatively simple neural network can be quite effective in identifying ASD based on structured input data. This kind of model could serve as a foundation for building more advanced diagnostic tools, especially in areas where specialist access is limited. With further fine-tuning, more training data, and perhaps ensemble techniques, the performance can likely be improved even more.

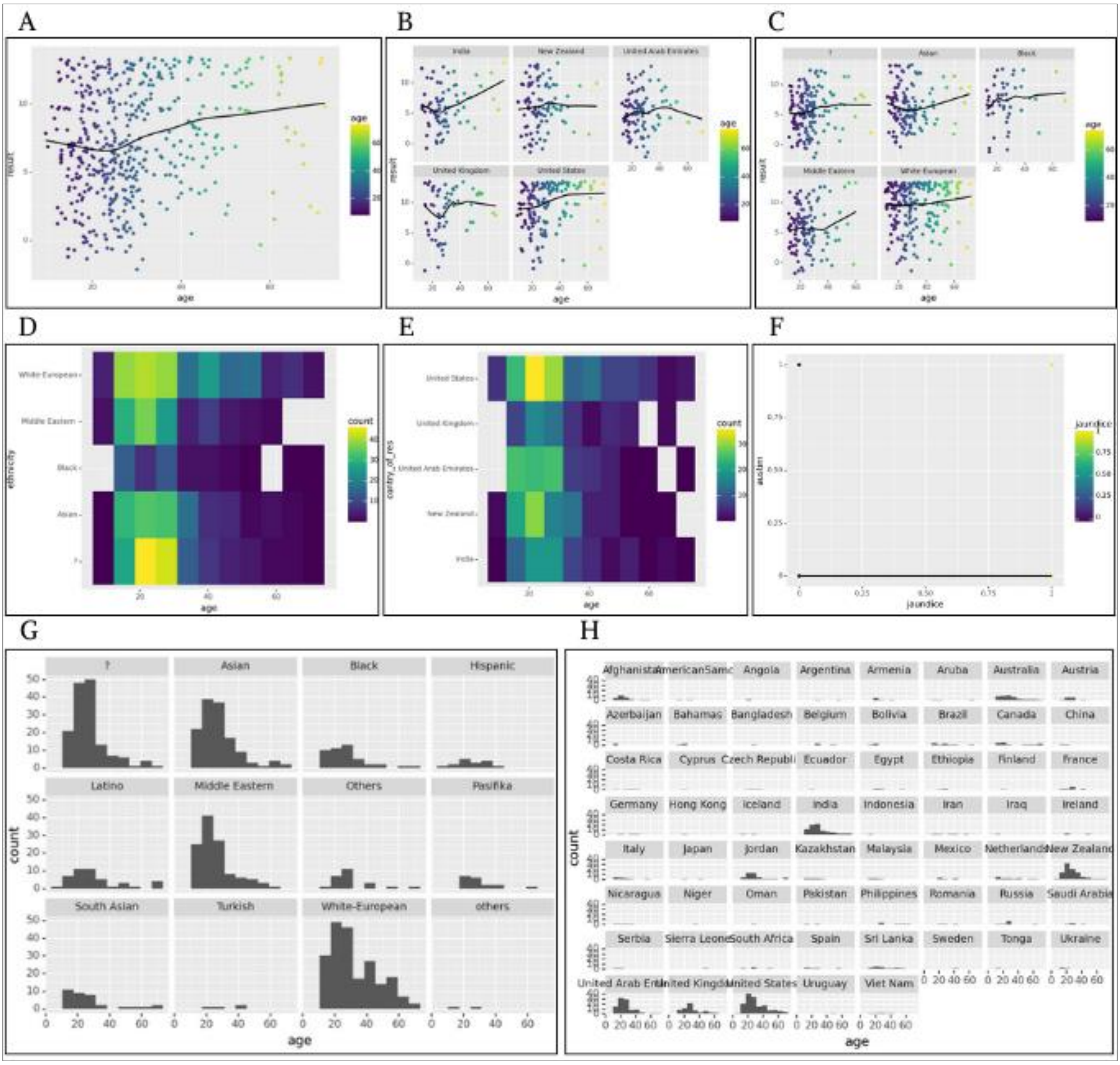

Fig. 2. Results of Exploratory Data Analysis: A. Scatter Plot Result v/s age, B. Scatter Plot Comparison of Different Countries: Result vs. Age, C. Scatter Plot Comparison of Ethnic races: Result vs. Age, D. Heat map between respective country and age, E. Heat map between ethnicity and age, F. Scatter Plot: Autism vs. Jaundice: Organization of the Text, G. Histogram for Ethnic races: Count vs. Age, H. Histogram for Different Countries: Count vs. Age

The scatter graph **(A)** displays how the predicted outcomes of the autism spectrum disorder (ASD) model correlate with the age of the individuals in the dataset. It could reveal trends or patterns in prediction accuracy among various age groups. For instance, it could indicate whether the accuracy of the model is higher for younger kids than for older people, or vice versa. The scatter plot **(B)** shows how the prediction of ASD is compared with age in various countries. The aim of this visualization is to determine whether there are any trends specific to individual countries in terms of prediction accuracy or the distribution of ASD predictions. It can illuminate disparities in the model's performance across various geographical areas or cultural settings. The scatter plot **(C)** chart shows how the prediction of ASD is compared with age within different races. The aim of this visualization is to determine whether there are any trends specific to ethnic races in terms of prediction accuracy or the distribution of ASD predictions. The heatmap **(D)**, color-coded depending on their density or frequency, showing the relationship between country and age. This scenario could reveal the distribution of individuals across various age groups per country in the dataset. The heatmap **(E)** illustrates how data points are spread out according to age and ethnicity. It can indicate whether specific age ranges have a higher concentration of particular ethnic groups. This visualization ensures that the dataset for training and testing the ASD prediction model includes a variety of ethnic groups to be representative. The scatter plot **(F)** analyzes how autism presence is related to a history of jaundice during birth. Determining any correlation between these two variables can be crucial in understanding the cause of ASD. The histogram **(G)** displays the frequency distribution of various age groups within the different races. It can provide insights into the demographic makeup of the data set and show if there are more individuals in a certain age group from a specific race. The histogram **(H)** illustrates how age groups are distributed in each country. This can offer information about the demographics of the dataset and indicate the presence of certain age groups from particular countries. Understanding whether the model has been trained and tested on a balanced dataset is crucial information.

IV. Discussion

Identifying and intervening early can enhance the results for children with ASD. We introduced a model for forecasting ASD using information gathered during standard developmental monitoring. Our models showed strong performance in young children and significant enhancement as they got older. A small model, based mainly on reaching milestones data, is proposed for a regular long-term monitoring situation, like the Israeli developmental surveillance program. The recommended forecasting tool can be integrated into the typical clinical process in various manners. Longitudinal developmental evaluation using age-appropriate milestones can incorporate ASD score calculation through the compact model into the EHR system. Moreover, milestone achievement metrics may consist of continuous DSS scores or basic binary indicators not tied to any particular developmental scale. Establishing a policy for determining the prediction age and thresholds of scores is necessary to accurately predict ASD while maintaining cost-effective results. Utilizing deep learning for identifying autism is important because of its trustworthiness, precision, and speed. Datasets were utilized in the proposed model to locally train Sequential model. The classifiers' results are sent to a central server where a meta classifier is trained to create a global model for detecting autism.

V. Conclusion

The proposed arrangement would determine the mental imbalance range, which may predict chemical imbalance qualities for distinct age groups by examining the appearances. Converting string values into whole integers. Using the Autism Spectrum dataset, the model that we will propose may accurately predict mental unevenness in the case of an event involving children, adolescents, and adults. This finding indicated greater execution, which differed from the current approach of screening mental unevenness. The assessment of ASD has been related with numerous disorders recognised as traits, including behavioural, emotional, structural, and mental disorders, making it difficult to forecast due to the lack of medical testing for all of the aspects required to detect ASD in an individual.

The detection process is time-consuming and complex because the symptoms are not visible. There is now no optimised and extensively established screening approach to particularly detect ASD, nor is there a screening test capable of reliably diagnosing ASD. Deep Learning is the most current research that can help detect autism more correctly, saving a lot of time. Deep Learning can assist diagnose ASD in patients of all ages, including children and adults.

REFERENCES

[1] B. Awaji *et al.*, "Hybrid techniques of facial feature image analysis for early detection of autism spectrum Disorder based on combined CNN features," *Diagnostics*, vol. 13, no. 18, p. 2948, Sep. 2023, doi: 10.3390/diagnostics13182948.

[2] V. Tisato *et al.*, "Genetics and Epigenetics of One-Carbon Metabolism Pathway in Autism Spectrum Disorder: A Sex-Specific Brain Epigenome?," *Genes*, vol. 12, no. 5, p. 782, May 2021, doi: 10.3390/genes12050782.

[3] M. A. Ortega *et al.*, "Microbiota–gut–brain axis mechanisms in the complex network of bipolar disorders: potential clinical implications and translational opportunities," *Molecular Psychiatry*, vol. 28, no. 7, pp. 2645–2673, Jan. 2023, doi: 10.1038/s41380-023-01964-w.

[4] P. Chaste and M. Leboyer, "Autism risk factors: genes, environment, and gene-environment interactions," *Dialogues in Clinical Neuroscience*, vol. 14, no. 3, pp. 281–292, Sep. 2012, doi: 10.31887/dcns.2012.14.3/pchaste.

[5] N. Kohn, J. Szopinska-Tokov, A. L. Arenas, C. F. Beckmann, A. Arias-Vasquez, and E. Aarts, "Multivariate associative patterns between the gut microbiota and large-scale brain network connectivity," *Gut Microbes*, vol. 13, no. 1, Jan. 2021, doi: 10.1080/19490976.2021.2006586.

[6] A. M, "Early Stage Detection of Autism Spectrum Disorder using Machine Learning," *International Journal for Research in Applied Science and Engineering Technology*, vol. 10, no. 6, pp. 3190–3194, Jun. 2022, doi: 10.22214/ijraset.2022.44584.

[7] G. Amit *et al.*, "Early prediction of autistic spectrum disorder using developmental surveillance data," *JAMA Network Open*, vol. 7, no. 1, p. e2351052, Jan. 2024, doi: 10.1001/jamanetworkopen.2023.51052.

[8] Md. M. Rahman, O. L. Usman, R. C. Muniyandi, S. Sahran, S. Mohamed, and R. A. Razak, "A review of Machine learning Methods of feature selection and Classification for autism Spectrum Disorder," *Brain Sciences*, vol. 10, no. 12, p. 949, Dec. 2020, doi: 10.3390/brainsci10120949.

[9] M. S. Farooq, R. Tehseen, M. Sabir, and Z. Atal, "Detection of autism spectrum disorder (ASD) in children and adults using machine learning," *Scientific Reports*, vol. 13, no. 1, Jun. 2023, doi: 10.1038/s41598-023-35910-1.

[10] V. Yaneva, L. A. Ha, S. Eraslan, Y. Yesilada, and R. Mitkov, "Detecting High-Functioning Autism in Adults using eye tracking and machine Learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 6, pp. 1254–1261, Jun. 2020, doi: 10.1109/tnsre.2020.2991675.

[11] F. Emmert-Streib, Z. Yang, H. Feng, S. Tripathi, and M. Dehmer, "An introductory review of deep learning for prediction models with big data," *Frontiers in Artificial Intelligence*, vol. 3, Feb. 2020, doi: 10.3389/frai.2020.00004.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[13] J. P. S *et al.*, "Autism spectrum disorder prediction by an explainable deep learning approach," *Computers, Materials & Continua/Computers, Materials & Continua (Print)*, vol. 71, no. 1, pp. 1459–1471, Jan. 2022, doi: 10.32604/cmc.2022.022170.

[14] "Autism screening on adults," *Kaggle*, Aug. 17, 2020. https://www.kaggle.com/datasets/andrewmvd/autism-screening-on-adults/data

[15] C. J. Newschaffer *et al.*, "The Epidemiology of autism spectrum Disorders," *Annual Review of Public Health*, vol. 28, no. 1, pp. 235–258, Apr. 2007, doi: 10.1146/annurev.publhealth.28.021406.144007.

[16] J. H. Miles, "Autism spectrum disorders—A genetics review," *Genetics in Medicine*, vol. 13, no. 4, pp. 278–294, Apr. 2011, doi: 10.1097/gim.0b013e3181ff67ba.

[17] U. Erkan and D. N. H. Thanh, "Autism Spectrum Disorder Detection with Machine Learning Methods," *Current Psychiatry Research and Reviews*, vol. 15, no. 4, pp. 297–308, Jan. 2020, doi: 10.2174/2666082215666191111121115.

[18] C. Wu *et al.*, "Machine Learning Based Autism Spectrum Disorder Detection from Videos," *IEEE*, Mar. 2021, doi: 10.1109/healthcom49281.2021.9398924.

[19] M. N. Parikh, H. Li, and L. He, "Enhancing diagnosis of autism with optimized machine learning models and personal characteristic data," *Frontiers in Computational Neuroscience*, vol. 13, Feb. 2019, doi: 10.3389/fncom.2019.00009.

[20] J. Elder, C. Kreider, S. Brasher, and M. Ansell, "Clinical impact of early diagnosis of autism on the prognosis and parent-child relationships," *Psychology Research and Behavior Management*, vol. Volume 10, pp. 283–292, Aug. 2017, doi: 10.2147/prbm.s117499.

[21] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset," *NeuroImage Clinical*, vol. 17, pp. 16–23, Jan. 2018, doi: 10.1016/j.nicl.2017.08.017.

[22] H. Alkahtani, T. H. H. Aldhyani, and M. Y. Alzahrani, "Deep learning Algorithms to identify Autism Spectrum Disorder in Children-Based Facial Landmarks," *Applied Sciences*, vol. 13, no. 8, p. 4855, Apr. 2023, doi: 10.3390/app13084855.