Autism Spectrum Disorder Analysis and Detection System

Dr Ananthajothi K
Department of CSE
Rajalakshmi Engineering College
Chennai, India
ananthajothi.k@rajalakshmi.edu.in

Mohanapriya E
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701164@rajalakshmi.edu.in

Mukhilan S S
Department of CSE
Rajalakshmi Engineering College
Chennai, India
210701169@rajalakshmi.edu.in

***Abstract*— Autism Spectrum Disorder is a neurodevelopmental disorder characterized by social interaction, communication impairments, and repetitive behaviors. In recent years, there has been an increase in prevalence, and it is important to identify individuals early for appropriate intervention to improve their quality of life. The following study was conducted to deepen the understanding and management of the disorder by employing a comprehensive two-module approach. Module one emphasizes the detection of ASD through sophisticated machine learning techniques for early and accurate diagnosis. Module two is on improving social and communication skills of children with ASD via game-based assessments, offering them an interactive yet enjoyable learning environment. This project, not only in contributing to the scientific community, but also through practical applications contributes and offers easy solutions for caregivers and educators. Integration of technology with both detection and intervention holds huge advancement in this area, promising better outcomes for ASD individuals.**

***Index Terms*—Autism Spectrum Disorder, Exploratory Data Analysis(EDA), Clinical Data, FeatureSelection, Deep Learning, Elderly Population, Early Diagnosis, Prediction Model, AI in Healthcare.**

I. Introduction

Autism Spectrum Disorder is a pervasive complex neurodevelopmental disorder where in an individual has impairment with communication, social interactions, and repetitive behaviors. Aspects of the disorder run a spectrum; therefore, various symptoms may be wide-ranged and different in many subjects. Some individuals will also have intellectual disabilities with extensive impairments, while some show exceptional skills in specific tasks such as mathematics, music, or art. Common signs of ASD include problems with social communication and interaction, such as a failure to maintain eye contact,a failure to understand certain social cues, and a failure to form relationships. Repetitive behaviors, restricted interests, and an intense need for routine and consistency are also characteristics of the disorder.

The cause of ASD remains unknown. It is therefore thought to be an effect of the interplay of genetic and environmental factors. Several genetic mutations increase susceptibility to developing ASD; no one cause has, however, been identified. The known environmental factors include exposure during prenatal periods to specific chemicals or medications, older advanced parental age, and certain complications during birth. Early diagnosis and intervention are crucial in the pursuit of better outcomes in children with ASD. These could rangefrom behavioral therapies to educational support, speech and language therapy, occupational therapy, to even medication that could target associated symptoms such as anxiety or over-activity. The goals of these interventions are to improve the development of life competencies of individuals with ASD, effectively communicate, and engage peers and people around them for the best quality of life

Early and accurate detection would allow timely interventions, which could drastically improve the developmental outcomes and quality of life for the affected. However, there are critical challenges to the existing methods, which limit their effectiveness. The current methods are very subjective and highly variable; they depend on clinical observation and caregiver questionnaires that are greatly influenced by the experience and interpretation of the clinician. Besides, diagnosis usually takes so much time and resources due to several evaluations and protracted observations that precede the start of interven- tions. It also happens to further delay interventions. Such a thing is highly unbearable in the case of the under-served or isolated regions without the accessibility to the professional diagnostic services and therefore leaves great inequalities behind in terms of early discovery and provision of assistance for individuals suffering from ASD.

These challenges call for the need to have an improved, objective, and more accessible diagnostic tool for ASD. The proposed module, therefore, aims at establishing an advanced ASD detection system based on machine learning algorithms and comprehensive behavioral data analysis. By reducing the factor of subjectivity, with the introduction of data-driven approaches, the system shall have improved diagnostic accuracy and precision. The diagnostic process will be more efficient with the help of automated analysis tools. Thus, ASD can

be diagnosed quickly, and interventions can be made earlier. Furthermore, an online or software-based diagnostic tool will reach out to people in many geographical locations and reduce the barriers for diagnosis. With automation of parts of the diagnostic process, the system will also relieve the pressure on healthcare professionals, optimize resource use, and make better use of limited healthcare resources.

II. RELATED WORKS

[1] Self-imposed methods of dietary control in children with autism lead to poor nutrition and food deficiencies. This study considers the eating habits of children who have autism or other similar neurodevelopmental disorders. A survey of 141 children looking, at food usage patterns and the rates of use of diets (e.g gluten free casein) associated with athletes on the spectrum was conducted.

[2] A comprehensive solution to the problem of longer waiting periods for the diagnosis of ASD is presented which is development of a system that utilizes VR for screening as well as classifying autism from non-verbal behavior analysis. Healthy participants performed as normal customers in a VR shopping scenario, where a virtual shop assistant was present. Analysis included gaze and head movements.

[3] A Platform for Autism Home-Based Therapeutic Interven tion highlights the importance of providing cost effective in-home programs for children diagnosed with Autism Spectrum Disorders (ASD) because conventional ABA therapy is quite expensive.This platform includes a mobile application and a desktop application that allow parents to coordinate therapy in the immediate surroundings, offering customized content and instructions.

[4] Autistic spectrum conditions (ASC) present a variety of challenges, one of which is the acquisition of an appropriate emotion and its expression which impedes socialization. To promote both the recognition and the expression of emotions in children diagnosed with high-functioning autism, a serious game with an integrated emotion recognition system based on 3D motion data is developed. For capturing body motion RGB-D sensors are utilized and the recognition of emotions is achieved using linear support vector machines (SVM).

[5] The research addresses issues of diagnosis and treatment of ASD. The study explores the potential of Eye Tracking (ET) and Electroencephalography (EEG) as tools in the search for reliable metrics in the diagnosis of ASD. By analyzing the data collected through eyetracking in combination with the brain activity as measured with EEG, the research discerns the patterns of the brain activity that arise during cognitive engagement that involves vision. This integrated strategy aids in the identification of specific markers that are associated with different behavioral patterns and their corresponding neural activity.

[6] Autistic Children Probability Estimation Using Hidden Markov Models reveals the genetic heritability of Autism Spectrum Disorder (ASD), and seeks to find out if using Hidden Markov Models (HMMs), the probability of children born to autistic parents being autistic can be estimated.

[7] Using statistical information about the heritability of autism and the sister-brother recurrence of autism, the authors establish an HMM that predicts if a child is likely to be autistic based on their parent's features. The model suggests probabilities of = 33 percentage for female children and = 80 percentage for male children regarding the inheritance of the ASD from autistic parents.

[8] Diagnosing Adults with High-Functioning Autism through Eye Tracking and Machine Learning it examines the effi-cacy of eye-tracking information in detecting high-functioning autism (HFA) in grown-up individuals. The goal is to create a computerized screening procedure relying on the application of machine learning visual processing classifiers based on eye tracking while subjects view web pages.The authors gathered eye movement data of adult participants, both autistic and non-autistic while engaging with web pages and trained machine learning classifiers on that data achieving approximately 74 percentage accuracy on autism detection.

[9] Profiling Autism Spectrum Disorder Through Characterization of Regional Cortical Activation and Functional Connectivity as Global Using a Game and Mobile Electroencephalography. This research study aims to assess and diagnose ASD by examining local cortical activations and global brain connectivity in social interactions using a game-based EEG system, as subjective behavioral tests have been criticized for several limitations Participants perform social cognitive tasks under a game-based platform wearing a mobile EEG headset. Collect the relevant data concerning brain activity. Apply machine learning models to classify ASD.

[10] Prediction of Symptom Severity in Autism Spectrum Disorder Using EEG Metrics discusses how EEG metrics may be used to predict the severity of ASD symptoms. The study is an attempt to find objective, reliable, and quantitative biomarkers that can be used to assess the severity of ASD symptoms, thus trying to overcome the subjective nature of clinical assessments. The study applied EEG data from the Autism Biomarkers Consortium for Clinical Trials dataset,with 257 children with ASD and 110 (TD) children. It built EEG brain networks and computed four types of EEG met- rics: network properties, power spectral density (PSD), spatial pattern features, and correlated connectivity weights.

[11] Oral Health and Quality of Life among Autistic Spectrum Disorder Individual will discuss the relationship between oral health and quality of life in children with ASD, pointing out poor oral hygiene and its impact on well-being and behavioral problems. Oral Health Assessment Tool, OHAT, was the tool used to assess the oral health, and EQ-5D-Y for the assessment of quality of life. The participants' number who had ASD

amounted to 163, dental health and its impact on generalquality of life evaluated.

[12] Children with ASD. are characterized by deficits in auditory temporal processing, which influence theirability to process sound durations and inter-stimulus intervals (ISI), and could be responsible for the social and communi- cation deficits.

III. PROPOSED METHODOLOGY

Autism Spectrum Disorder is the significant neurodevelopmental complexity and profoundly alters interaction in society, communication patterns, and repetitive behaviors. Early diagnosis and proper identification are the most critical steps leading to the success of any intervention that can heavily influence the ASD-affected individual's outcome of developmental process and the quality of life. Till date, though, diverse methods used for detection remain problematic. They heavily depend on the subjective clinical observations and caregiver questionnaires, a modification that is susceptible to varying clinician experiences and personal interpretations. The process leading to the diagnosis is resource-intensive as several follow-ups are necessary, and in the process, the intervention may end up being delayed.

In addition to these challenges, there is a disparity in access to specialized diagnostic services, particularly in underserved or remote areas, exacerbating the difficulty in achieving timely identification and support for individuals with ASD. These barriers highlight the urgent need for a more efficient, objective, and accessible diagnostic tool for ASD. Such a tool would aim to streamline the diagnostic process, reduce subjectivity, and ensure that accurate diagnoses can be made early and reliably, thus facilitating prompt and effective intervention to improve outcomes for individuals with ASD.

B. Dataset:

The datasets used in this ASD detection module are comprised of two separate datasets, namely a childdataset and an adult dataset. These datasets contain a range of features related to behavioral and diagnostic criteria used in theidentification of ASD in different age groups. The proposed child dataset aims at filling this gap and comes in handy to give value data for screening ASD. The dataset provided here shows information from a screening tool known as Q-Chat-10,which feature ten behavioral features. Such a dataset records the answer to questions whose answer comes in the form of two binary values, that are 1 or 0, based on the responselikelihood of exhibiting ASD traits. Specifically, in Q1-Q9,it has received a score of 1 when the response has been "Sometimes", "Rarely," or "Never". For question 10, when theresponse has been "Always," "Usually," or "Sometimes," then a score of 1. A score above 3 on Q-Chat-10 means potential ASD features.

This adult dataset of more than 700 is from the responses to survey questions obtained through an app form and contains labels signifying if they have received a diagnosis of ASD. The data can be applied to make predictions about likelihood of ASD in relation to any survey or demographic variables while exploring autism impact across all genders, ages, or other variables.

Dataset Description

The dataset's features are as follows: Demographics: age, gender, ethnicity, jaundice, austim, contryofres, usedappbefore

Behavioral Scores: A1Score, A2Score, A3Score, A4Score, A5Score, A6Score, A7Score, A8Score, A9Score, A10Score

Cognitive/Behavioral Assessments: , result, agedesc, relation, Class/ASD

C. Data Preprocessing and Feature Selection

In this analysis, A combined dataset of children and adults diagnosed with ASD is created in this analysis. The first step entails the addition of a new field that represents age groups in both datasets-'Children' for the children's dataset and 'Adults' for the adults' dataset.

The count of people diagnosed with ASD is produced, showing there are 330 diagnosed and 666 non-diagnosed within the dataset. Visualization techniques include a count plot for the Class/ASD column and heatmap to highlight missing values, which provides insight into characteristics of the dataset. Lastly, a correlation heatmap has been developed to understand relations between various features in the dataset. Then the dataset is shuffled and split for two equal parts,patients diagnosed with ASD and random sample of those without such a disorder in order not to distort data balance.

Pre-training models, this input data is divided into featureset and target label. There are several columns of score, demographic features, as well as some contextual features in the feature set whereas a target label is set to true, if that patient is already diagnosed else false. Applying MinMaxScaler for improving model performances, numerical values of such specific features would be scaled. Using label encoding or one-hot encoding for Categorical features prepares the dataset in front of ML algorithms.

D. System Architecture

The proposed system architecture of the ASD detection system has therefore been carefully crafted to ensure robust and accurate identification of these disorders. This starts from raw datasets that contain information both from children and adults, retrieved from various data repositories or clinical

records meant to give a comprehensive diversified representation of the target population.

Preprocessing forms the initial critical step that follows after raw data gathering. In this process, several steps are involved for cleaning and standardizing data. Handling missing values within the dataset is done with imputation techniques to prevent loss of valuable information. Duplicates are identified and eliminated to ensure that there would be no bias or redundancy in model training.

One - Hot Encoding is a procedure to transform categorical variables into a binary matrix. Therefore, the data becomes prepared for machine learning algorithms which expect numerical inputs. Along with this, min-max scalar normalization is also done so that the data becomes normalized within a given range, usually 0 to 1, to aid in faster convergence of the algorithms used for learning.

After having preprocessed the data, it is further divided into three different subsets, namely, children's dataset, adult's dataset, and a combined dataset where both are merged together. This division will then help make a detailed analysis and understanding of ASD within the different age groups,and also a holistic view if both datasets are considered.

The next step is performing statistical analysis on these data sets. Statistical tests like the Chi-squared test are conducted to determine relevant features with a high value that influence the detection of ASD. These features play a role in determining what the model learns and ultimately its predictability.

The principal characteristics identified would split up the dataset into training and testing sets at a 70-30 ratio. This would ensure that enough data is provided for model training, while a considerable amount is held out to test the modelon unseen data, thus giving a more realistic measure of its accuracy and generalization.

RandomizedSearchCV is an optimization technique, which searches across a grid of hyperparameters for finding the optimal configuration for the model. Fine-tuning the selected models will make use of RandomizedSearchCV.

Once the hyperparameters are fine-tuned, it further trains the models on the complete training set. In doing so, it makes the most of the available data and thereby ensures that the model is well-prepared to face all kinds of scenarios and variations in the dataset. After that, the trained models are tested against the reserved test set.

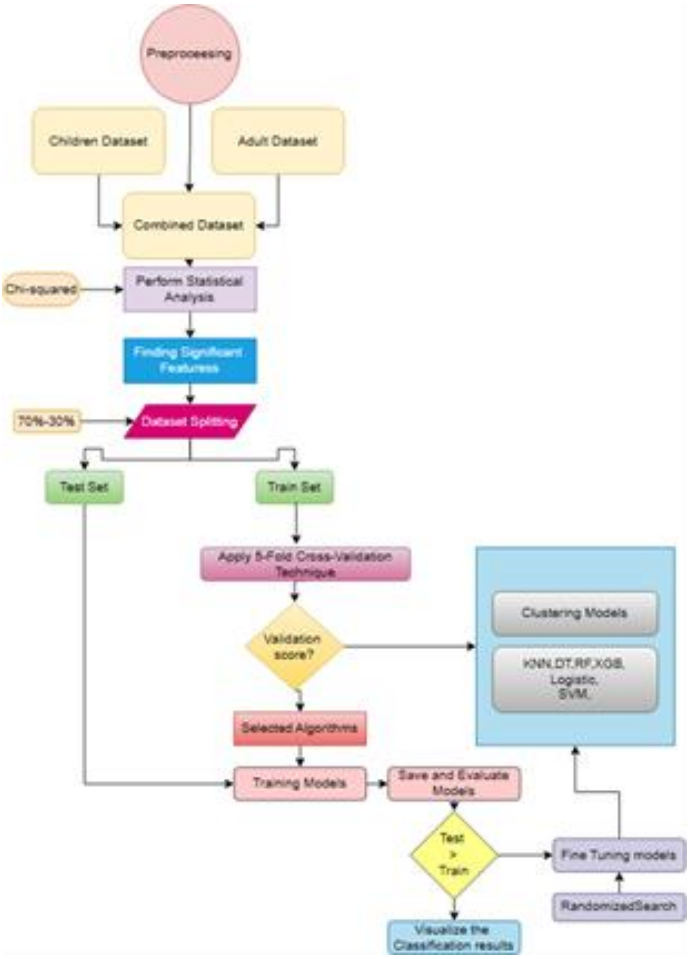

Fig. 1. Working

If the testing accuracy of the model exceeds the training accuracy of the model, there will be generalization of data that does not lead to overfitting. Such models are robust, and they were saved in the course of deployment for further use. The last step would be to display the classification outcomes. Thus,that will enhance the understanding of the model's decision towards the presented case and shed light on which factors influence the detection of ASD.

In summary, this architecture builds the ASD detection system in a methodological manner. Advanced preprocessing techniques, statistical analysis, cross validation, and fine-tuning on machine learning models are used, making this architecture robust and highly accurate in order to detect an ASD disorder. This comprehensive framework allows not only the accurate detection but insight into the patterns and features associated with ASD.

E. Performance Metrics:

Accuracy:

The closeness of a measurement to the true or accepted value is known as accuracy. Accuracy evaluates the model's overall prediction performance by determining the ratio of correct predictions.

Precision:

: Precision refers to how close measurements of the same item are to each other. Precision reflects the accuracy of the model's positive predictions by calculating the proportion of true positives out of all instances labeled as positive.It is computed with the formula:

TP / (TP + FP)

where FP represents false positives and TP stand for True Postives.

Recall (Sensitivity):

Recall, or sensitivity, gauges the model’s effectiveness in identifying positive cases within the dataset. This metric is determined by the formula:

TP / (TP + FN)

where FN represents false negatives.

Confusion Matrix: The model predictions are compiled interms of true positive, False Positive,True Negative, and False Negative results in a confusion matrix for a concise overview of the performance of the model. From these figures, different other measurements can be computed.

IV. Results and Discussion

A. Experimental Setup:

The experimental setup of the project was done on a Windows 11 operating system with an Intel Core i7 processor (3rd generation), 16GB of RAM, and a 512GB SSD-the heart of machines that would process and store all data. The application is developed by the help of Python 3.11 and Jupyter Notebook, and all those keys are there inside Anaconda 3. Most of the main libraries related to exploratory data analysis, namely, Pandas, Seaborn, and Matplotlib helped in carrying out data manipulation and proper visualizations; the most popular Tab-Net algorithm utilized during this is PyTorch and Scikit-learn for

modeling. Before experimentation, the dataset was cleanedup and preprocessed to avoid data duplication, and tuning of hyperparameters was conducted to fine-tune performance to get the best outcome.

Observations:

On observing the dataset: First, preprocess the pipeline-tackle missing values in the dataset. Missing data seriously under-mines the performance of machine learning models, making way for biased or erroneous predictions. Different imputation techniques may include mean, median, or mode imputation, or some more complex techniques that involve k-nearest neighbors (KNN) imputation to fill those gaps without valuable information being lost. Categorical variables often cannot be

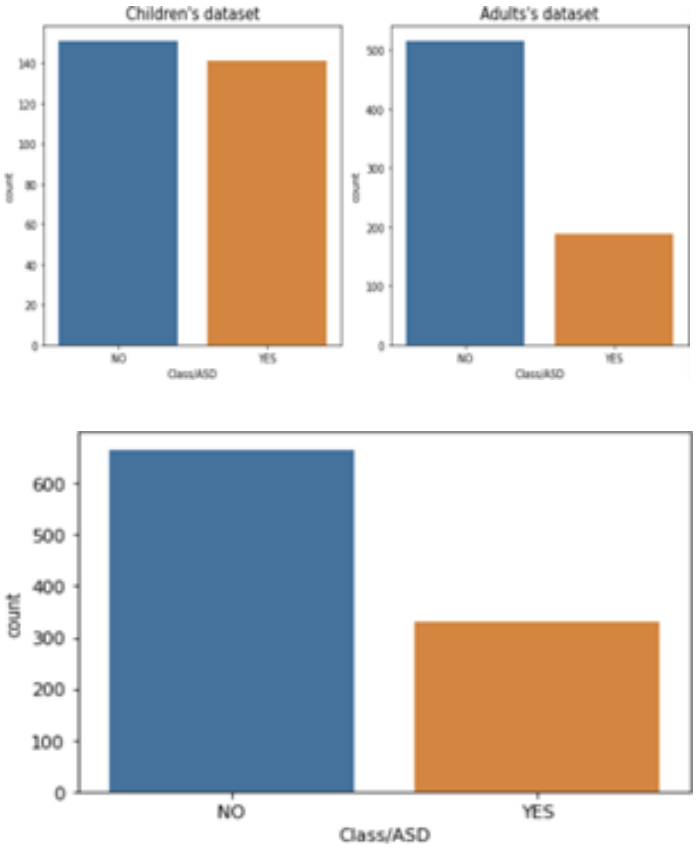

Fig. 2. ASD Diagnosis CHILD vs ADULT vs COMBINED dataset

directly used in machine learning algorithms. It turns each category into a binary vector, and this is inter- pretable by the models numerically. This is particularly im- portant for models requiring numerical input, such as logistic regression and support vector machines.

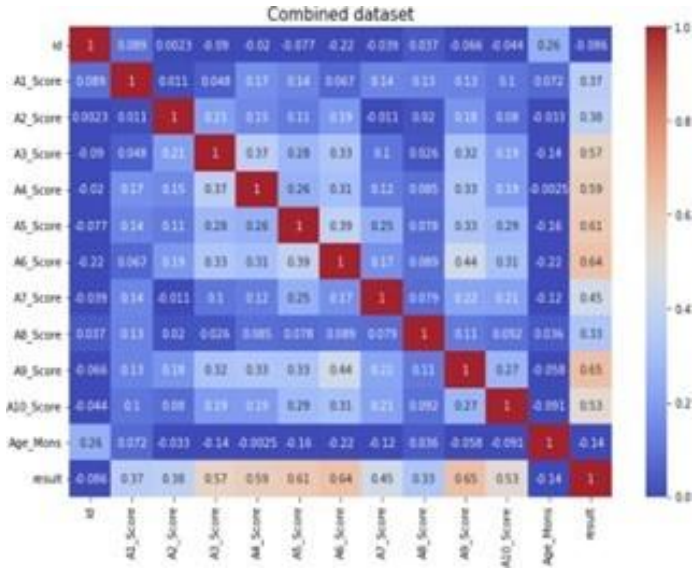

Fig. 3. Correlation Graph of all features

The K-Nearest Neighbors model with the best hyperparameters, namely algorithm='brute', leaf size=112, neighbors=112, showed a increased accuracy of 0.92 and a well-balanced F1 score of 0.922 while having a relatively higher log loss of 1.8997. The SVM model achieved a perfect score for all performance metrics; hence, it may be extremely effective or overfitting. The Random Forest classifier, tuned with hyperparameters criterion='entropy', max depth=6, maxfeatures='sqrt', min- samplesleaf=2, nestimators=100 worked fine. The Decision Tree model obtained perfect scores, so its fit is excellent and therefore it requires independent evaluation on a test set to avoid overfitting. The XGBoost model, optimized hyperparameters: learning rate = 0.1, max depth = 6, n estimators = 100 showed perfect accuracy, precision, recall, and F1 score, meaning it's robust and effective to capture ASD patterns.The same was noticed with Logistic Regression about perfect performance metrics.

Model	Precision	Recall	F1-score	Accuracy
Extreme Gradient Boosting (XGB)	90.01	85.67	88	92.6

Fig. 4. : Performance metrics for XGB

The Confusion Matrix categorizes predictions into four types: correct positives, correct negatives,and two kinds of mistakes: false positives and false negatives.

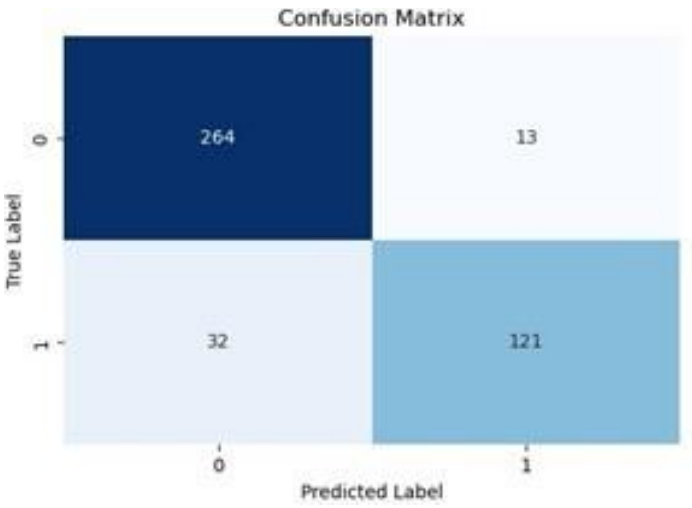

Fig. 5. Confusion Matrix

This table helps you look at a glance where the model gets things right and where it goes wrong, making it easier to improve accuracy. Graphical observation of the performance

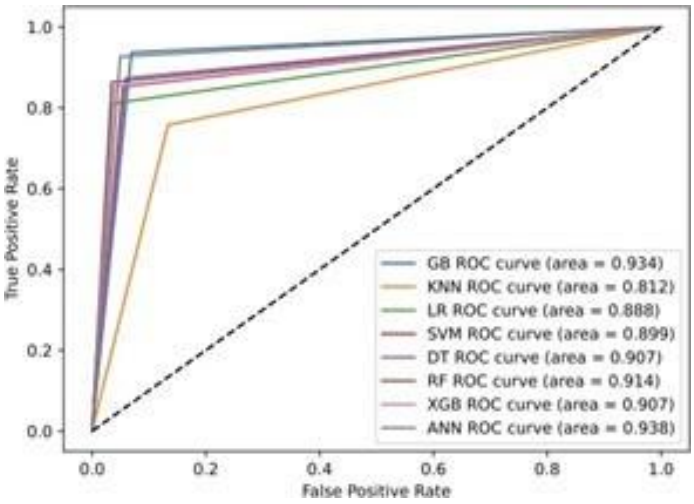

Fig. 6. Training loss over epochs

of different models is provided for the ASD. The consistent performances of SVM, Random Forest, Decision Tree, XGBoost, and Logistic Regressionpoint towards the efficiency of optimizing the hyperparameters but lay emphasis on the validation results of the models on separate datasets. Among these, XGBoost is shown to be an optimal model in respect of complexity vs performance which would mean a good classifying result of ASD.

V. CONCLUSION AND FUTURE WORK

After trying all these models on the dataset, it can be found that all these models do great work with high performances; however, XGBoost is considered robust and highly accurate for detecting ASDs

The current undertaking progresses towards the next level with an intention of improving sustenance and developmental progress of individuals with Autism Spectrum Disorder (ASD) through an assessment-based stage incorporation of game mechanics. This phase of the project will be about development of meaningful and purposeful fun games aimed at assessing and training cognitive social skills and several behavioral patterns of children and adults with autism. The system will encourage learning in a more focused and enthusiastic way by combining Game-Based Approach with therapeutic techniques and giving appropriate intervention and feedback.

The game-based assessment module will utilize adaptive algorithms, allowing to modify the level of challenge and type of tasks according to the record of performance of each user in order to give each user a meaning experience that corresponds to his or her needs and abilities.

VI. REFERENCES

[1] Ahmed, S., Hossain, M.F., Nur, S.B., Shamim Kaiser, M., Mahmud, M., 2022. In-school and Community based Ma-chine Learning oriented Psychological assessment of Autism Spectrum Disorders. Proceedings of Trends in Electronics and Health Informatics: TEHI 2021. Springer, pp. 139–149

[2] P. Kumar, S. Senthil Pandi, T. Kumaragurubaran and V. Rahul Chiranjeevi (2024), "Human Activity Recognitions in Handheld Devices Using Random Forest Algorithm," 2024 International Conference on Automation and Computation (AUTOCOM), Dehradun, India, 2024, pp. 159-163, doi: 10.1109/AUTOCOM60220.2024.10486087.

[3] M.R. Alteneiji, L.M. Alqaydi, M.U. Tariq, 2020. Diagnosis of Autism Spectrum Disorder Using Machine Learning Techniques. International Journal of Advanced Computer Science and Applications, 11.

[4] Bala, M., Ali, M.H., Satu, M.S., Hasan, K.F., Moni, M.A., 2022. Efficient machine learning models for early stage detection of autism spectrum disorder. Algorithms 15, 166

[5] Ahamad, M.M., Aktar, S., Uddin, M.J., Rahman, T., Alyami, S.A., AlAshhab, S., Akhdar, H.F., Azad, A., Moni, M.A, 2022. Machine Learning Approaches for detecting Ovarian Cancer Using Clinical Data. Journal of personalized medicine 12, 1211.

[6] Cavus, N., Lawan, A.A., Ibrahim, Z., Dahiru, A., Tahir, S., Abdulrazak, U.I., Hussaini, A., 2021. A systematic literature review on the application of machine-learning models in behavioral assessment of autism spectrum disorder. Journal of Personalized Medicine 11, 299.

[7] Gaspar A., Oliva D., Hinojosa S., Aranguren, I., Zaldivar D., 2022. In autismspectrum disorder diagnosis from gaze tracking images, kernel extremelearning machine (KELM) with sub KELM classifiers performance analysis.Applied Soft Computing 120, 108654.

[8] Ozsivadjian, A., Knott, F., & Magiati, I. (2012). Parent and child perspectives on the nature of anxiety in children and young people with autism spectrum disorders: a

[9] focus group study. Autism, 16(2), 107–121. https://doi.org/10.1177/1362361311431703

[10] Joudar, S.S., Albahri, A., Hamid, R.A., 2022. A systematicreview of artificial intelligence-based triage as well as priority-based healthcare diagnosis of autism spectrum disorders and its genetic factors. Computers in Biology and Medicine 146, 105553..

[11] Kashef, R., 2022. Diagnostic approach of autism spectrum disorder using ecnn: Enhanced convolutional neural networks.
Cognitive Systems Research 71, 41–49

[12] Mashudi N.A., Ahmad N., Noor N.M., 2021. A Machine Learning Approach for Classifying Adult Autistic Spectrum Disorder. IAES International Journal of Artificial Intelligence 10 (4) 743.

[13] P. Kumar, V. Subathra, Y. Swasthika and V. Vishal, "Disease Diagnosis Using Machine Learning on Electronic Health Records," 2024 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, 2024, pp. 1-6, doi: 10.1109/IC3IoT60841.2024.10550235.

[14] Qureshi, M.S., Qureshi, M.B., Asghar, J., Alam, F., Aljarbouh, A., et al.,2023. Using Machine Learning Techniques to Predict and Assess Autism Spectrum Disorder. Journal of healthcare engineering 2023.

[15] Sharif, H., Khan, R.A., 2022. Implementation of a machine learning based framework for the identification of Autism Spectrum Disorder. Applied Artificial Intelligence 36, 2004655.

[14Sherkatghanad, Z., Akhondzadeh, M., Salari, S., Zomorodi-Moghadam, M.,Abdar, M., Acharya, U.R., Khosrowabadi, R., Salari, V. (2020). Automated detection of autism spectrum disorder using deep learning. Frontiers in Neuroscience, 13, 1325.

[15] The Authors Wei, Q., Xu, X., Xu, X., and Cheng, Q., in the Year 2023. Diverse Instruments Integration towards Early Diagnosis of Autism – A Machine Learning Approach Testing Clinical Applicability. Psychiatry Research 320, 115050.