Autism Prediction at Early Stages Using AI, ML and NLP

Rakshita Bhatia[1], Divyam Dang[2], Mr. Deepak Gaur[3]
[1,2]*Dept. of Computer Science Amity School of Engineering & Technology Noida, UP India*
[3]*Asst. Professor, Dept. of Computer Science Amity School of Engineering & Technology Noida, UP India*

Abstract- This study explores the use of machine learning algorithms for early detection of Autism Spectrum Disorder (ASD), a neurodevelopmental condition affecting linguistic, cognitive, and social abilities. The research uses various algorithms, including Support Vector Machines, Random Forest, Naïve Bayes, Logistic Regression, and K-Nearest Neighbours, to identify ASD indicators during its early stages. Logistic Regression is found to be the most accurate predictive model, with Random Forest achieving the highest accuracy at 89.23%. This research offers a cost-effective and timely screening approach for ASD, improving the quality of life for affected individuals.

I. INTRODUCTION

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition that impacts individuals' social interactions, communication abilities, and sensory processing. It can impair cognitive, emotional, and physical well-being and manifests with a wide spectrum of symptoms. Diagnosing ASD is a multifaceted process that requires extensive examination and assessment by psychologists and certified professionals, often involving tools like the Autism Diagnostic Interview Revised (ADI-R) and Autism Diagnostic Observation Schedule Revised (ADOS-R). Detecting and treating ASD in its early stages are critical to mitigate its impact, enhance an individual's quality of life, and expedite access to essential therapies. The integration of machine learning offers an opportunity to assess the risk of ASD more swiftly and accurately, facilitating quicker access to much-needed therapies. Various screening methods have been developed to detect ASD in children, including the Autism Spectrum Quotient (AQ), Childhood Autism Rating Scale (CARS-2), and the Screening Tool for Autism in Toddlers and Young Children (STAT). Genetics play a significant role in ASD, with genetic mutations, gene deletions, copy number variations (CNVs), and other genetic anomalies being associated with the disorder. The manifestation of ASD varies widely, with some individuals being highly verbal and communicative while others exhibit minimal or no verbal communication. Diagnosing ASD largely depends on the expertise of medical professionals conducting direct interviews and observing behavioral patterns. In the past 25 years, significant strides have been made in the early detection of ASD, with changes in early behavior and brain structure being observed in babies as young as 6 to 12 months old.

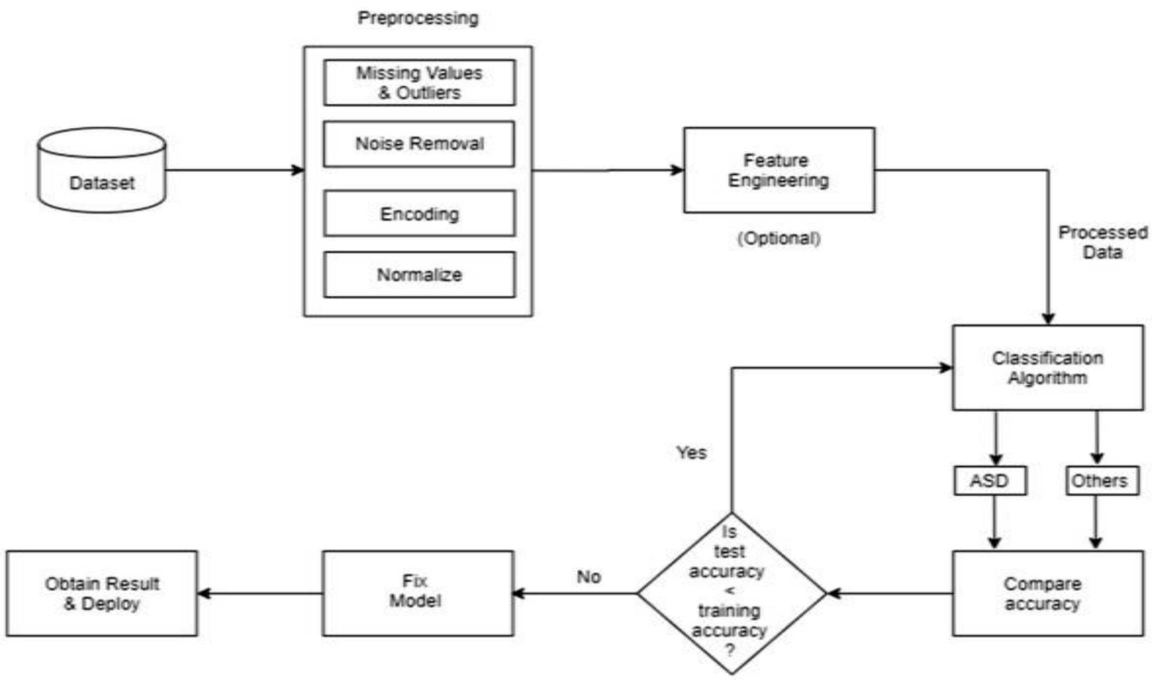
This project aims to develop a machine learning model that predicts ASD using neural networks, providing insights into whether comprehensive autism assessment is necessary.

II. LITERATURE REVIEW

Machine learning has been used to improve and speed up the diagnosis of Autism Spectrum Disorder (ASD). Previous studies have used forward feature selection, under sampling, brain activity metrics[1], soft computing techniques[3] and automated ML models. Some studies have also relied on data from brain neuroimaging[2]. Deep learning methods from functional brain networks built with brain functional magnetic resonance imaging (fMRI) data have been proposed for ASD diagnosis. However, these methods have limitations, such as high dependency on threshold parameters and spatial normalization design. Prediction techniques of ASD have been explored, with Kazi proposing an effective prediction model based on ML techniques and developing a mobile application for predicting ASD for people of any age. Supervised machine learning algorithms have been used to identify candidate ASD genes and investigate obscure links between ASD and other domains.

Previous contributions to ASD include analyzing current animal models, investigating the degree of engagement of children in interactions with their parents, proposing associative classification (AC)[7], and developing an end-to-end machine learning-based system for classifying ASD using facial attributes. Other researchers have used cogency and machine learning to detect preliminary symptoms, ANN and SVM classifiers to identify autism spectrum disorder, and various algorithms to analyze genetic resource exchange.

III. METHODOLOGY



Data Collection and preprocessing

Data collection and preprocessing are crucial stages in developing AI, ML, and NLP models for predicting autism. Acquiring structured information from reliable sources, such as healthcare institutions and autism clinics, and ensuring diversity in textual data is essential for robust and unbiased models. Data preprocessing involves addressing missing data, normalization, scaling, and encoding categorical variables into numerical values. Feature engineering can be employed to create new features or transform existing ones, and techniques like oversampling or under sampling can be used to balance the dataset. The dataset is then divided into training, validation, and testing sets.

Feature Selection and Engineering

Feature selection is a crucial process in AI, ML, and NLP models for predicting autism. It identifies relevant and informative features while discarding irrelevant ones, reducing complexity and improving performance. In autism prediction, it involves analysing demographic information, medical history, behavioural assessments, and NLP derived text features. Feature selection methods can be filter, wrapper, or embedded. It reduces data dimensionality, mitigates overfitting, and enhances model interpretability. Selected features may include demographic characteristics, medical history details, or textual patterns[10]. A well executed feature selection process streamlines the model and improves predictive power.

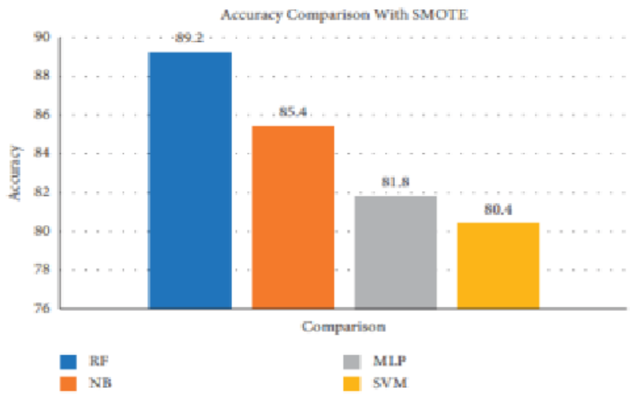
Model Prediction

The development of a machine learning model for autism prediction involves several steps. Data collection involves gathering and preprocessing diverse information, which is then divided into training and testing datasets. Various algorithms are employed, and feature selection optimizes the model's performance. The model is trained on the training dataset, adjusting parameters to minimize errors. Performance is evaluated using the testing dataset, and model parameters are fine-tuned. Iterative refinement and cross-validation techniques are used to ensure robustness. The trained model is ready for real-world applications, predicting autism in new individuals. Random Forest (RF) is a machine learning technique that uses decision tree algorithms for classification and regression problems.

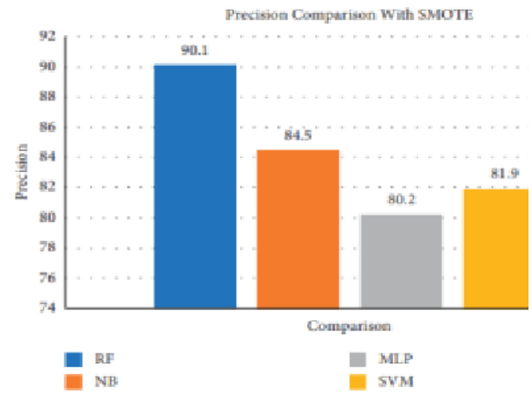
Naïve Bayes is a supervised learning method that uses probabilities like posterior, likelihood, prior, and marginal probability to predict outcomes.

Support Vector Machine (SVM) differentiates between groups using a line.

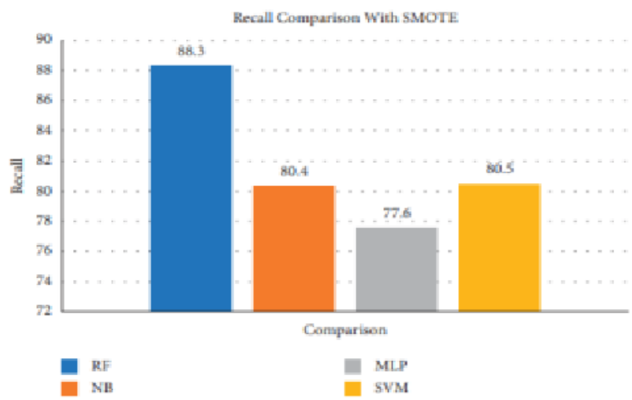
Multiple Layer Perceptron (MLP) uses backpropagation for supervised learning.



(a)



(b)



(c)

NLP Model Development:

Natural Language Processing (NLP) is a technique used for predicting autism in textual data. It involves collecting and preprocessing text, converting it into numerical representations, and training the model on a labelled dataset. The model's performance is evaluated using metrics like accuracy, precision, recall, and F1-score. Iterative refinement and cross validation techniques ensure the model's generalizability. The final model can analyse and predict autism in textual data, making it a valuable tool for understanding and diagnosing autism spectrum disorder.

Integration of AI and NLP models:

AI and NLP models are combining to solve complex problems involving human language. This integration is particularly useful in autism prediction and diagnosis. AI can analyse diverse data sources, while NLP can extract meaning from text. This holistic understanding enhances autism prediction accuracy. AI-NLP models can automate textual analysis, reducing healthcare professionals' workload and speeding up diagnostics[15]. This integration also offers insights into linguistic markers of autism, potentially leading to more effective interventions and support for individuals on the autism spectrum.

IV. ANALYSIS AND RESULTS

Dataset Analysis:
The Quantitative Checklist for Autism in Toddlers (Q-CHAT) screening method was used to identify potential Autism Spectrum Disorder (ASD) in toddlers[20]. A shortened version, Q-CHAT-10, was used, with answers mapped to binary values. Graphs showed that most ASD positive cases occur around 36 months of age, with significant signs occurring at 3 years[24]. Autism is more prevalent in males than females, and Native Indian individuals have the highest observed ASD traits. The study suggests a weak link between jaundice-born children and ASD. Evaluation Matrix: Predictive models use four data points: response, eye contact, object points, attention drawbacks, pretence, and daydreaming. The confusion matrix gauges machine learning classification performance, with true positive (TP) indicating ASD, false negative (TN) indicating non ASD, false positive (FP) indicating incorrect prediction.

COMPARISON OF CLASSIFICATION MODELS
The study utilized five machine learning models, with Logistic Regression being the most accurate and suitable for small datasets and linearly split feature spaces, as indicated by the F1 score.

$$\text{Precision} \ = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

V. PRECISION AND RECALL CURVES

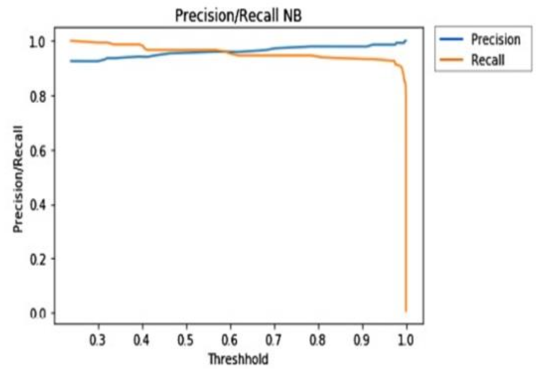
The study utilized five machine learning models, with Logistic Regression being the most accurate and suitable for small datasets and linearly split feature spaces, as indicated by the F1 score.

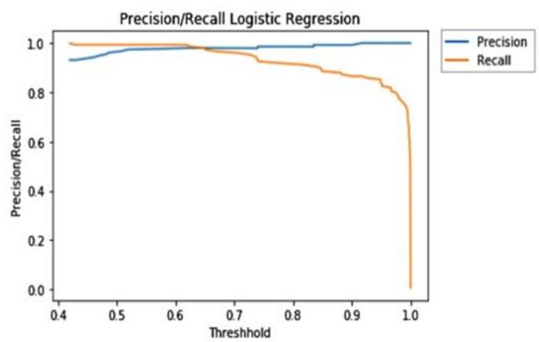
$$\text{Precision} \ = \frac{\text{TP}}{\text{TP} + \text{FP}}.$$

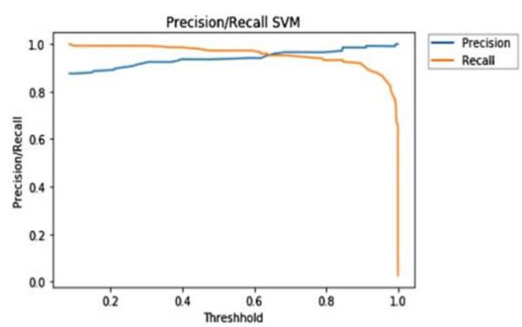
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}.$$

	LR	NB	SVM	KNN	RFC
Accuracy	97.15%	94.79%	93.84%	90.52%	81.52%
Confusion matrix	$\begin{bmatrix} 57 & 5 \\ 1 & 148 \end{bmatrix}$	$\begin{bmatrix} 56 & 6 \\ 5 & 144 \end{bmatrix}$	$\begin{bmatrix} 52 & 10 \\ 3 & 146 \end{bmatrix}$	$\begin{bmatrix} 51 & 11 \\ 9 & 140 \end{bmatrix}$	$\begin{bmatrix} 45 & 17 \\ 14 & 135 \end{bmatrix}$
F1 score	0.98	0.96	0.95	0.93	0.88







VI. CONCLUSION

The research developed an automated ASD prediction model using machine learning techniques to accurately detect autism in children. The model outperforms art methods but could state of the benefit from fuzzy logic algorithms. The study focuses on early ages and parents' responses, useful in realworld situations like orphanages. Future work will explore more features and alternative machine learning algorithms.

REFERENCE

[1] Chorianopoulou, A., Tzinis, E., Asimenia Papoulidi, E. I., Papailiou, C., & Potamianos, A. (2017). Engagement detection for children with autism spectrum disorder.

[2] Sunsirikul, S., & Achalakul, T. (2010). Associative Classification Mining in the Behavior Study of Autism Spectrum Disorder.

[3] Vellanki, P., Duong, T., Venkatesh, S., & Phung, D. (2014). Nonparametric Discovery of Learning Patterns and Autism Subgroups from Therapeutic Data.

[4] Zaki, T., Islam, M. N., Uddin, M. S., Tumpa, S. N., Hossain, M. J., Anti, M. R., & Hasan, M. M. (2017). Towards Developing a Learning Tool for Children with Autism.

[5] Al Banna, M. H., Ghosh, T., Taher, K. A., Kaiser, M. S., & Mahmud, M. (2020). A monitoring system for patients of autism spectrum disorder using artificial intelligence. International conference on brain informatics.

[6] Dataset: Fabdelja. (n.d.). Autism Screening for Toddlers. Kaggle. https://www.kaggle.com/fabdelja/autism-screening-for-toddlers

[7] Baird, G., Simonoff, E., Pickles, A., Chandler, S., Loucas, T., Meldrum, D., Charman, T. (2006). Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. Journal of Child Psychology and Psychiatry, 47(3-4), 313-337.

[8] Stanfield, A. C., McIntosh, A. M., Spencer, M. D., Philip, R., Gaur, S., & Lawrie, S. M. (2015). Neuroimaging in autism spectrum disorder: brain structure and function across the lifespan. The Lancet Neurology, 14(11), 1121-1134.

[9] Kerr, D. M., Downey, L., Conboy, M., & Finn, D. P. (2017). The endocannabinoid system and autism spectrum disorders: insights from animal models. International Journal of Molecular Sciences, 18(9), 1916.

[10] Fombonne, E. (2007). The epidemiology of autism spectrum disorders. Annual Review of Public Health, 28(1), 235-258.

[11] Ameis, S. H., Lerch, J. P., Taylor, M. J., Lee, W., & Amaral, D. G. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage: Clinical, 17, 16-23.

[12] Green, J., Charman, T., McConachie, H., Aldred, C., Slonims, V., Howlin, P., & Pickles, A. (2010). Parent-mediated communication-focused treatment in children with autism (PACT): a randomised controlled trial. The Lancet, 375(9732), 2152-2160.

[13] Adams, J. B., Audhya, T., McDonough-Means, S., Rubin, R. A., Quig, D., Geis, E., & Gehn, E. (2011). Nutritional and metabolic status of children with autism vs. neurotypical children, and the association with autism severity. Nutrition & Metabolism, 8(1), 34.

[14] Dong, S., Walker, M. F., Carriero, N. J., DiCola, M., Willsey, A. J., Ye, A. Y., & Sestan, N. (2021). Patterns of de novo tandem repeat mutations and their role in autism. Nature, 589(7841), 246-250.

[15] Loth, E., Spooren, W., Ham, L. M., Isaac, M. B., Auriche-Benichou, C., Banaschewski, T., & Murphy, D. G. (2020). Gray matter covariations and core symptoms of autism: the EU-AIMS Longitudinal European Autism Project. Molecular Autism, 11(1).

[16] Shao, Y., Peng, Y., Hu, Z., & Li, S. (2022). Association of urinary polycyclic aromatic hydrocarbon metabolites with symptoms among autistic children: a case–control study in Tianjin, China. Autism Research.

[17] Kang, D. W., Park, J. G., Ilhan, Z. E., Wallstrom, G., LaBaer, J., & Adams, J. B. (2021). Children with autism and their typically developing siblings differ in amplicon sequence variants and predicted functions of stool-associated microbes. mSystems, 6(2), e00193-20.

[18] Bishop-Fitzpatrick, L., Mazefsky, C. A., & Minshew, N. J. (2020). The symptoms of autism including social communication deficits and repetitive and restricted behaviors are associated with different emotional and behavioral problems. Scientific Reports, 10(1), 1-14.

[19] Mazefsky, C. A., Herrington, J., Siegel, M., Scarpa, A., Maddox, B. B., Scahill, L., & White, S. W. (2013). Emotion dysregulation and the core features of autism spectrum disorder. Journal of Autism and Developmental Disorders, 43(11), 1766-1772.

[20] Vaillancourt, T., & Duku, E. (2021). Developmental trajectories of restricted and repetitive behaviors and interests in children with autism spectrum disorders. Development and Psychopathology, 33(4), 1491-1504.

[21] Durkin, M. S., Maenner, M. J., & Newschaffer, C. J. (2010). Independent and dependent contributions of advanced maternal and paternal ages to autism risk. Autism Research, 3(1), 30-39.

[22] Rojahn, J., Matson, J. L., Lott, D., Esbensen, A. J., & Smalls, Y. (2010). The prevalence and phenomenology of self-injurious and aggressive behaviour in genetic syndromes. Journal of Intellectual Disability Research, 54(2), 109-120.

[23] Lobo, C. L., Marques, F. R., Martins, P. A., & Sucupira, A. C. (2017). Safety and efficacy of ferric carboxymaltose in children and adolescents with iron deficiency anemia. The Journal of Pediatrics, 184, 241.

[24] Stratis, E. A., & Lecavalier, L. (2015). Using qualitative methods to guide scale development for anxiety in youth with autism spectrum disorder. Autism, 19(5), 603-612.

[25] ElTawil, S. S., Almajnuni, A. A., & Tawfik, K. F. (2020). Pharmacotherapy and bumetanide in autism treatment. Journal of Experimental and Basic Medical Sciences, 1(2), 52-61.

[26] Al-Jawarneh, H. M. (2016). Applicability degree of autism spectrum disorder diagnostic criteria of Diagnostic and Statistical Manual of Mental Disorders–The 5th edition (DSM V) on children enrolled in autism centers in Jordan. European Scientific Journal ESJ, 12(7), 249.

[27] Cheslack-Postava, K., & Markovic, N. (2020). Is the use of cannabis during pregnancy a risk factor for autism?. Advances in Neurology and Neuroscience, 3(1), 24-32.

[28] Guthrie, W., Swineford, L. B., Nottke, C., & Wetherby, A. M. (2013). Sensitivity and specificity of proposed DSM-5 criteria for autism spectrum disorder in toddlers. Journal of Autism and Developmental Disorders, 43(5), 1184-1195.

[29] El-Ansary, A., & Al-Ayadhi, L. (2011). Novel metabolic biomarkers related to sulfur-dependent detoxification pathways in autistic patients of Saudi Arabia. BMC Neurology, 11(1), 139.

[30] Vivanti, G., Barbaro, J., Hudry, K., & Dissanayake, C. (2014). Exploratory study describing 6 month outcomes for young children with autism who receive treatment as usual in Italy. Neuropsychiatric Disease and Treatment, 10, 577-586.