

Intelligent Autism Spectrum Disorder Detection Using Machine Learning Models

¹G. Divya, ²Dr. V. Maniraj

¹Research Scholar, Department Of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur(Dt),
Affiliated to Bharathidasan University, Thiruchirappalli, Tamilnadu
(Mail Id-gdivya19.mca@gmail.com)

²Associate Professor, Research Supervisor, Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous),
Poondi, Thanjavur(Dt), Affiliated To Bharathidasan University, Thiruchirappalli, Tamilnadu,
(Mail Id-manirajv61@gmail.com)

Abstract:

This project pioneers a revolutionary approach to predict Autism Spectrum Disorder (ASD) using advanced machine learning techniques. Implemented in Python through platforms like Jupyter Notebook or Google Colab, the project aims to craft an efficient and precise predictive model for early ASD detection. The primary objective is to develop a tool that facilitates early identification of ASD, enabling timely interventions and support for individuals on the autism spectrum. By harnessing machine learning methodologies, this project aspires to significantly enhance the quality of life for individuals with ASD through early recognition and intervention.

Index terms: Autism Spectrum Disorder, Jupyter Notebook, quality of life, recognition and Intervention.

Introduction

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition characterized by challenges in social interaction, communication, and repetitive behaviours. Early diagnosis and intervention are crucial for improving the long-term outcomes of individuals with ASD. Traditional diagnostic methods, which often rely on behavioural assessments and expert evaluations, can be time-consuming, subjective, and may lead to delays in diagnosis. The advent of machine learning (ML) offers a promising alternative by enabling more efficient, accurate, and objective prediction of ASD. This research explores the innovative application of machine learning techniques to predict ASD, aiming to enhance early detection and intervention strategies. Machine learning, a subset of artificial intelligence, has revolutionized various fields by enabling computers to learn from data and make predictions. In the context of ASD prediction, ML algorithms can analyse large datasets, identify patterns, and predict the likelihood of an individual having ASD based on diverse features such as genetic, behavioural, and environmental factors. By leveraging advanced ML techniques, it is possible to develop models

that provide a more comprehensive understanding of ASD, facilitating earlier and more accurate diagnoses compared to traditional methods.

Objective of the Project:

The primary objective of this research is to develop and validate an innovative machine learning-based model for the early prediction of Autism Spectrum Disorder (ASD). This model aims to leverage advanced machine learning algorithms to analyse a wide range of data, including genetic, behavioural, and environmental factors, to accurately identify individuals at risk of ASD. By improving the accuracy and efficiency of ASD diagnosis, this research seeks to facilitate early intervention, which is crucial for improving developmental outcomes in affected individuals. Additionally, the study aims to contribute to the existing body of knowledge by identifying key predictive features and patterns associated with ASD, thereby enhancing our understanding of the disorder and informing future research and clinical practices.

Problem Definition:

The challenge lies in the timely identification of Autism Spectrum Disorder, considering the diverse manifestations of symptoms and the need for early interventions. Traditional diagnostic processes often lead to delays in providing support, impacting the overall well-being of individuals on the autism spectrum.

Existing System:

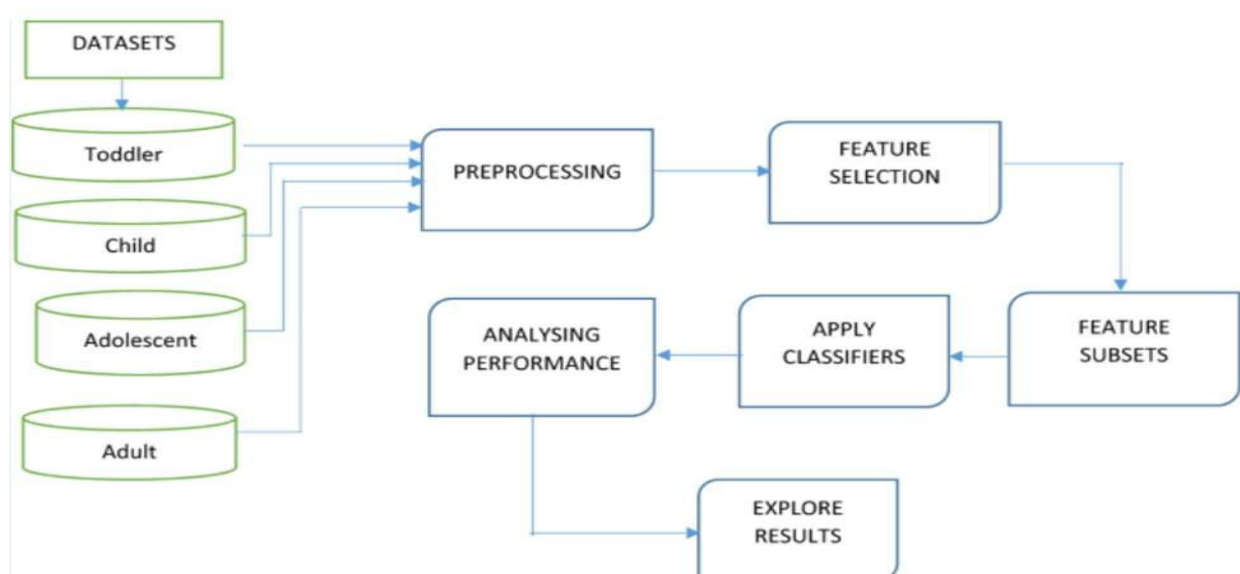
Existing studies in the field of ASD prediction predominantly focus on a variety of machine learning techniques. However, there remains a gap in the literature concerning the comprehensive use of specific algorithms like Support Vector Machines (SVM) and K-Nearest Neighbor (KNN) for early ASD detection. While some studies explore broader aspects of machine learning in healthcare, the application of these algorithms to the intricacies of ASD prediction is underrepresented. This research builds upon existing knowledge by emphasizing the unique advantages of SVM and KNN for ASD prediction and their potential integration into a user-friendly tool for healthcare professionals.

Proposed System:

The proposed approach for predicting Autism Spectrum Disorder (ASD) using machine

learning focuses on integrating various data sources and employing advanced predictive models to enhance diagnostic accuracy and early detection. This approach involves the collection and analysis of diverse data types, including behavioral assessments, genetic information, and neuroimaging data. By leveraging these multi-modal data sources, the proposed method aims to capture the complex and heterogeneous nature of ASD. To achieve this, the approach utilizes a combination of machine learning algorithms such as Support

Vector Machines (SVM), Random Forests, and Convolutional Neural Networks (CNNs) to develop predictive models. Feature extraction and selection techniques are employed to identify the most relevant predictors of ASD, improving model performance and interpretability. Additionally, ensemble methods are used to combine predictions from multiple models, enhancing the overall accuracy and robustness of the prediction system. The integration of these advanced techniques is expected to facilitate more accurate and earlier detection of ASD, ultimately leading to improved outcomes for individuals and more effective intervention strategies.

System Architecture Diagram**Fig1: System Architecture**

A System Architecture Diagram visually depicts the structure and components of a system, illustrating how they work together to accomplish a specific function or goal. It offers a high-level perspective of the system, highlighting the primary components, data flow, and the interactions among various parts. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and

the output data is generated by this system. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

Algorithms and Module Implementation

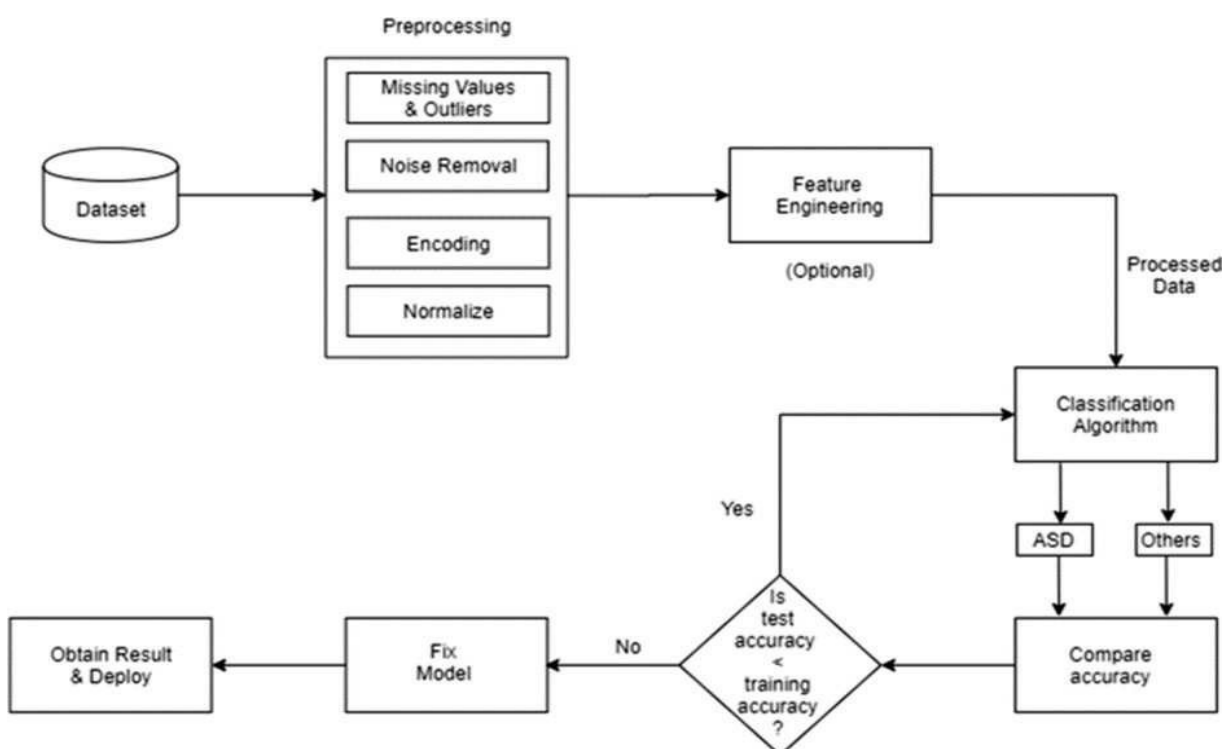


Fig2: Block Diagram

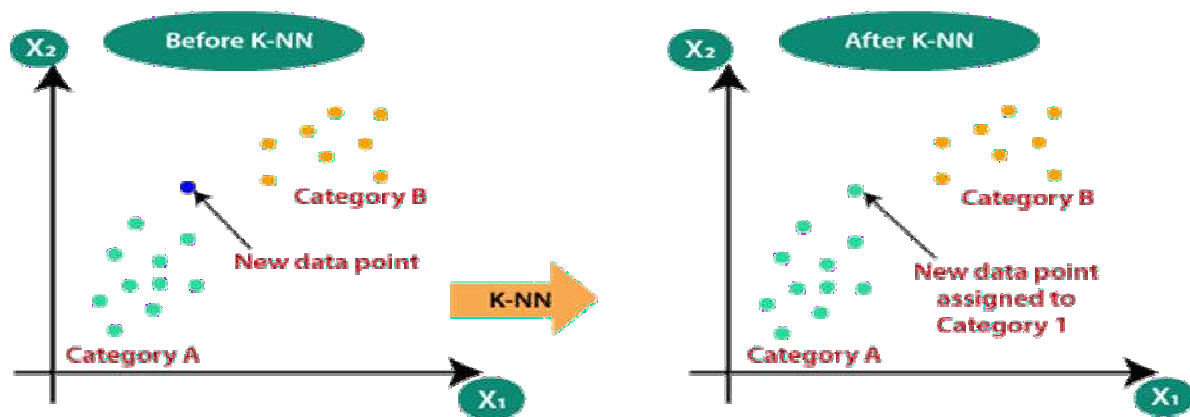
Data Collection: Gather datasets containing features relevant to predicting Autism Spectrum Disorder.

Data Pre-processing: Handle missing values, perform feature selection.

Model Development: Implement supervised learning algorithms like Support Vector Machines (SVM) and K-Nearest Neighbor (KNN).

Training and Evaluation: Split the data into training and testing sets. Train the models and evaluate their performance.

Tool Development: Develop a user-friendly tool for early ASD prediction. This architecture ensures a coherent and effective workflow, culminating in the development of a practical tool for early ASD detection.

k-Nearest Neighbour (kNN):**Fig3: K-NN Algorithm**

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.

- K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset.

- KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

- Suppose there are two categories, i.e., Category A and Category B, and we have a new data point x_1 , so this data point will lie in which of these categories. To solve this type of problem, we need a KNN algorithm. With the help of K-NN,

we can easily identify the category or class of a particular dataset.

Module Implementation

- **Data Collection Module:** Collaborate with medical institutions and research centers to acquire comprehensive datasets. Use standardized protocols for data collection to ensure consistency and accuracy. Anonymize and preprocess the data to maintain patient confidentiality and prepare it for analysis.

- **Feature Selection and Extraction Module:** Implement feature selection techniques such as correlation analysis, mutual information, and principal component analysis (PCA). Extract key features related to genetics, behavior, and environment using domain knowledge and statistical

methods. Regularly update and refine the feature set based on ongoing research findings and model performance.

- **Model Training and Development Module:** Split the dataset into training, validation, and test sets to ensure robust model evaluation. Train multiple machine learning models, including support vector machines (SVM), random forests, and neural networks, to identify the best-performing algorithm. Use techniques like cross-validation and hyperparameter tuning to optimize model performance. Implement ensemble learning methods to combine the strengths of different models for improved accuracy.

- **Model Evaluation and Validation Module:**

Use metrics such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC) to assess model performance. Validate the models using independent test datasets to ensure generalizability and robustness. Perform sensitivity analysis to understand the impact of different features on the model's predictions. Conduct external validation with data from different sources to confirm the model's applicability across various populations.

- **Deployment and Integration Module:**

Develop a user-friendly interface for clinicians and healthcare professionals to input patient data and receive ASD risk predictions. Ensure the deployment platform adheres to data privacy and security standards. Provide training sessions and documentation to healthcare providers on how to use the tool effectively. Monitor the system's performance in realworld settings and collect feedback for continuous improvement.

- **Continuous Learning and Improvement Module:**

Implement mechanisms for continuous data collection and integration of new information. Regularly retrain the model with updated datasets to maintain its predictive accuracy. Stay informed about the latest developments in ASD research and incorporate new findings into the model. Engage with the research community for collaborative efforts and validation studies

Conclusion

The integration of machine learning into Autism Spectrum Disorder (ASD) prediction represents a significant advancement in the field of diagnostic and therapeutic interventions. By leveraging sophisticated algorithms and diverse data sources, such as behavioural assessments, genetic information, and neuroimaging data, machine learning models can enhance the accuracy and timeliness of ASD detection. These predictive models not only offer the potential for earlier diagnosis, which is crucial for effective intervention, but also provide a more personalized

approach to understanding the complexities of ASD. The use of advanced techniques like deep learning and ensemble methods further improves model performance, offering promising improvements over traditional diagnostic methods.

Future Work

Looking ahead, the future of ASD prediction using machine learning will likely be marked by continued advancements in technology and data integration. The focus will shift towards incorporating real-time monitoring and longitudinal data to capture developmental changes more effectively. Additionally, enhancing model interpretability and incorporating explainable AI will be crucial for ensuring that predictions are both reliable and actionable. By continually refining these models and integrating them into clinical practice, there is potential for transformative improvements in early detection and personalized care, ultimately leading to better outcomes for individuals with ASD and their families.

References

1. Duda, M., Ma, R., Haber, N., & Wall, D. P. (2016). Use of machine learning for behavioural distinction of autism and ADHD. *Translational Psychiatry*, 6(5), e732.
2. Thabtah, F. (2017). Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment. *Proceedings of the 1st International Conference on Medical and Health Informatics 2017*, 1-6.
3. Bone, D., Bishop, S. L., Black, M. P., Goodwin, M. S., Lord, C., & Narayanan, S. S. (2016). Use of machine learning to improve autism screening and diagnostic instruments: A systematic review. *Autism Research*, 9(11), 1274-1293.
4. Sadiq, M. A., Jeon, M., & Imran, M. A. (2019). Machine Learning based Diagnosis of Autism Spectrum Disorder using Structural and

Functional MRI. IEEE Access, 7, 115453-115462.

5. Thabtah, F., Peebles, D., Retzler, J., & Peebles, D. (2019). A machine learning autism classification based on logistic regression analysis. Health Information Science and Systems, 7(1), 1-11.

6. Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a minimal set of behaviours for autism detection through feature selection-based machine learning. Translational Psychiatry, 5(2), e514.

7. Zhao, X., Li, X., Jiao, Y., Yu, L., Li, Y., & Li, S. (2018). Support vector machine based classification of autism and control using structural MRI. Biomedical Research International, 2018, 1-10.

8. Sakai, K., & Yamashita, Y. (2015). Deep learning-based classification of autism spectrum disorder using brain network analysis. Journal of Biomedical Science and Engineering, 8(10), 683-695.

9. Iwabuchi, T., Takao, H., & Fukuda, H. (2019). Multivariate pattern analysis of resting-state functional connectivity predicts the severity of autism spectrum disorder. Journal of Autism and Developmental Disorders, 49(3), 792-805.

10. Wei, L., Zhang, Z., & Guo, X. (2019). A hybrid convolutional and recurrent neural network for ADHD diagnosis. Frontiers in Neuroscience, 13, 178.