Machine Learning Solutions for Autism Spectrum Disorder Characterization

Varun Chawla, Divesh Chhoda, Atharva Sawant, Sunny Bhatia, Veena Trivedi
CSE Department, VESIT, Chembur, India
veena.trivedi@ves.ac.in
2020.varun.chawla@ves.ac.in
2020.divesh.chhoda@ves.ac.in
2020.sunny.bhatia@ves.ac.in
d2020.athurva.sawant@ves.ac.in

Abstract

Autism Spectrum Disorder (ASD) is a complex neurodevelopmental condition marked by diverse social, communication, and behavioral challenges. This review explores recent strides in applying machine learning and deep learning to autism research. It delves into leveraging various data sources like behavioral observations, neuroimaging, genetic information, and electronic health records to build models for early diagnosis, personalized treatment, and unraveling autism's underlying mechanisms. The paper also addresses the challenges inherent in these approaches, such as data heterogeneity, limited sample sizes, and interpretability issues. By synthesizing current research, this paper underscores the potential of machine learning and deep learning in advancing our understanding of autism and improving targeted interventions for individuals on the spectrum.

Keywords- Autism Spectrum Detection, Machine Learning, Decision Tree, Matlab, Prediction Model.

I. INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that affects a person's ability to learn, communicate, and participate. Symptoms are most common in the first two years of life and change over time.[4] People with ASD face various challenges, including focus issues, learning disabilities, mental health conditions, movement impairments, and sensory sensitivity. The number of cases of ASD has increased worldwide, with about 1 in 160 children receiving an ASD diagnosis.[1]

Diagnosing autism is time-consuming and costly, and rapid identification can transform intervention by enabling timely prescriptions and preventing symptom aggravation. This research aims to develop an autism prediction model using machine learning (ML) and a smartphone app to identify autism symptoms by age groups.

The paper aims to create an autism screening tool for predicting ASD characteristics in individuals aged 4-11, 12-17, and 18 and beyond. Integrating deep and machine learning enhances the ability to discern subtle details defining ASD traits, improving the prediction model's accuracy for precise assessments and facilitating early intervention plans across age groups.

II. BACKGROUND

Autism Spectrum Disorder: Background and Prevalence

Autism Spectrum Disorder (ASD) is a neurological condition involving difficulties in speech, social interaction, and repetitive behaviors.[4] Coined in 1911 to describe symptoms of schizophrenia,[2] modern diagnostic criteria were established in the 1940s by Drs. Leo Kanner and Hans Asperger, identifying specific behavioral patterns linked to autism. The definition of ASD has evolved over time to encompass a wider range of symptoms. Key points about ASD include:

Prevalence:

This estimate represents an average figure, and reported prevalence varies substantially across studies. Some well-controlled studies have, however, reported figures that are substantially higher.[11]

1. Approximately 1 in 100 children (1%) are estimated to have ASD.
2. In The United States 1 in 36 children (2.8%) have ASD, according to the CDC's Autism and

Developmental Disabilities Monitoring (ADDM) Network.
3. In India every 1 in 100 children below 10 years has autism.
4. Prevalence varies across racial and ethnic groups white children have a 2.4% prevalence, while Black children have 2.9%, Hispanic children 3.2%, and Asian or Pacific Islander children 3.3%.
5. Males are nearly 4 times more likely to be diagnosed with ASD than females.

Challenges and Importance of Early Detection of Autism Spectrum Disorder (ASD)

Early detection of ASD is crucial for maximizing a child's potential and improving their long-term outcomes. However, there are several challenges associated with early detection.

Challenges:

1. Lack of awareness: Many parents and caregivers are unaware of the early signs and symptoms of ASD, which can delay diagnosis.
2. Limited access to diagnostic services: There is a shortage of qualified professionals trained to diagnose ASD, particularly in rural areas.
3. Cost of diagnosis: The cost of diagnosis can be prohibitive for some families.
4. Cultural stigma: In some cultures, there is a stigma associated with mental health conditions, which can discourage parents from seeking help.
5. Misdiagnosis: ASD can sometimes be misdiagnosed as other conditions

Importance:

Early detection of autism is important because it can enable early intervention, which can improve the developmental outcomes and quality of life for children with autism and their families.[11] Some of the benefits of early detection are:

- Early intervention: Early intervention has been shown to significantly improve outcomes for children with ASD.
- Improved quality of life: Early diagnosis can help children with ASD develop the skills they need to live happy and fulfilling lives.
- Reduced burden on families: Early intervention can also help reduce the stress and burden on families.
- Improved educational outcomes: Children with ASD who receive early intervention are more likely to succeed in school.
- Better social and emotional development: Early intervention can help children with ASD develop the social and emotional skills they need to form healthy relationships.

III. METHODOLOGY

A. Problem formulation and Data Collection

The project "Machine Learning Solutions for Autism Spectrum Disorder" focuses on leveraging machine learning (ML) algorithms to address the challenges associated with diagnosing and characterizing Autism Spectrum Disorder (ASD). By employing ML models, the research aims to objectively analyze neurological data and extract discriminative features related to ASD, thereby facilitating early diagnosis and personalized interventions. The significance of this endeavor lies in its potential to offer a quantifiable approach to ASD characterization, leading to earlier interventions, enhanced understanding of ASD's neurological aspects, and the development of tailored treatment strategies. This research underscores the advantages of ML over traditional methods, highlighting its ability to detect intricate patterns in complex data and continuously improve its accuracy. Moreover, it contributes to the field by demonstrating the feasibility of ML in ASD diagnosis, identifying potential features and models for clinical applications, and laying the groundwork for future research on personalized interventions for individuals with ASD.

Data collection constitutes the foundational step in our endeavor, "Machine Learning Solutions for Autism Spectrum Disorder." In this phase, we meticulously gathered data from established datasets, leveraging their richness and depth to inform our research. Additionally, we employed tailored questionnaires to solicit valuable input directly from users, ensuring a holistic perspective encompassing diverse viewpoints and experiences. Subsequently, our focus shifted towards data refinement through a comprehensive cleaning and preprocessing regimen. This involved the meticulous identification and removal of outliers, as well as the transformation of raw data into a standardized format conducive to the requirements of our algorithm. By undertaking these preparatory measures, we aimed to enhance the quality and reliability of our dataset, thereby laying a robust foundation for subsequent analysis and modeling.

B. *Choosing the right algorithm*

.

The initial step in our project methodology involves the careful selection and fine-tuning of a machine learning algorithm tailored to the task of detecting Autism Spectrum Disorder (ASD). To accomplish this, we meticulously evaluated several well-established algorithms known for their efficacy in similar diagnostic tasks. These algorithms encompass a variety of methodologies, including Logistic Regression, Decision Trees, Random Forests, and K Nearest Neighbors (KNN).

Each of these algorithms offers distinct advantages and operates on different principles, thereby presenting a diverse range of approaches to ASD detection. However, to ensure optimal performance for our specific application, we engaged in a rigorous process of parameter tuning. Parameter tuning involves adjusting the settings within each algorithm to maximize its accuracy, sensitivity, and specificity in identifying ASD-related patterns within the data.

By meticulously fine-tuning the parameters of these machine learning algorithms, we aim to enhance their ability to discern subtle but significant patterns indicative of ASD. This step is crucial in laying the groundwork for subsequent stages of our research, as it ensures that our chosen algorithm is finely calibrated to the nuances of ASD detection, thereby laying a solid foundation for further analysis and model development.

C. *Model Training and Evaluation*

The initial phase of our project, "Machine Learning Solutions for Autism Spectrum Disorder," focuses on model training, which begins with the curation of a dataset containing labeled data, each instance associated with a known classification or outcome related to autism spectrum disorder (ASD). This dataset forms the foundational input for our machine learning models to discern patterns and relationships within the data. Employing a meticulous cross-validation methodology, we assess the efficacy and robustness of our model by partitioning the dataset into subsets, training the model on a portion, and validating its performance on the remainder. This iterative process provides a comprehensive understanding of the model's ability to generalize and accurately classify or predict ASD-related attributes. Additionally, during model training, we adhere to best practices in feature selection, preprocessing, and algorithm selection to optimize performance while addressing concerns such as overfitting or bias. This rigorous approach ensures that our machine learning solutions are both effective and ethically and scientifically sound.

We prioritized model evaluation as a fundamental step in developing machine learning solutions for Autism Spectrum Disorder (ASD). This involved a thorough assessment of the efficacy and reliability of our model through a comprehensive evaluation process. To achieve this, we partitioned our dataset into two subsets: one for model training and another held-out dataset exclusively for evaluation purposes, ensuring unbiased evaluation metrics. Following model training, we rigorously tested its performance on the held-out dataset, utilizing key performance metrics such as accuracy, precision, and recall. Accuracy provided an overview of the model's correctness across all classes, while precision focused on the accuracy of positive predictions, and recall assessed the model's ability to identify all relevant instances accurately. Leveraging these metrics allowed us to comprehensively evaluate the model's performance, identifying both strengths and areas for improvement. This evaluation process lays a critical foundation for subsequent project phases, guiding further refinement and optimization of our machine learning solution for ASD.

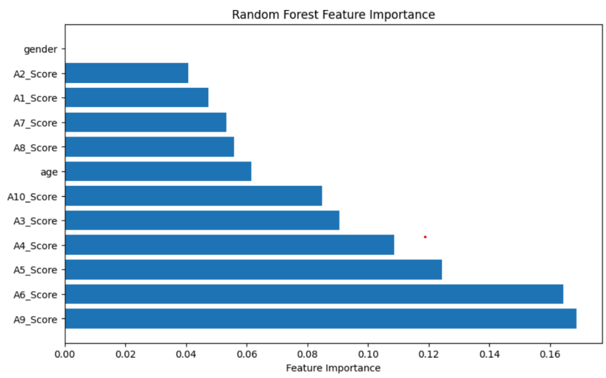

Fig1.1

The above figure illustrates the importance of feature importance in random forest technique. The A1-A10 scores would later on be distinguished as prominent factors of Autism symptoms important in abbreviating the detection analysis.

```
Random Forest:
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.99      0.99       150
           1       0.99      0.98      0.98       133

    accuracy                           0.99       283
   macro avg       0.99      0.99      0.99       283
weighted avg       0.99      0.99      0.99       283
```

Fig 1.2

```
KNN:
Classification Report:
              precision    recall  f1-score   support

           0       0.98      0.79      0.87       150
           1       0.80      0.98      0.88       133

    accuracy                           0.88       283
   macro avg       0.89      0.88      0.88       283
weighted avg       0.89      0.88      0.88       283
```

Fig 1.3

```
Decision Tree:
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.97      0.96       150
           1       0.96      0.95      0.95       133

    accuracy                           0.96       283
   macro avg       0.96      0.96      0.96       283
weighted avg       0.96      0.96      0.96       283
```

Fig 1.4

```
Logistic Regression:
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       150
           1       1.00      1.00      1.00       133

    accuracy                           1.00       283
   macro avg       1.00      1.00      1.00       283
weighted avg       1.00      1.00      1.00       283
```

Fig 1.5

Above are the evaluation metrics observed for all the 4 algorithms used after applying the smote technique on our dataset.

```
Accuracy Scores:
Random Forest: 0.9522727272727273
KNN: 0.8863636363636364
Decision Tree: 0.9
Logistic Regression: 1.0
```

Fig 1.6

Following are the accuracies observed for all 4 algorithms.

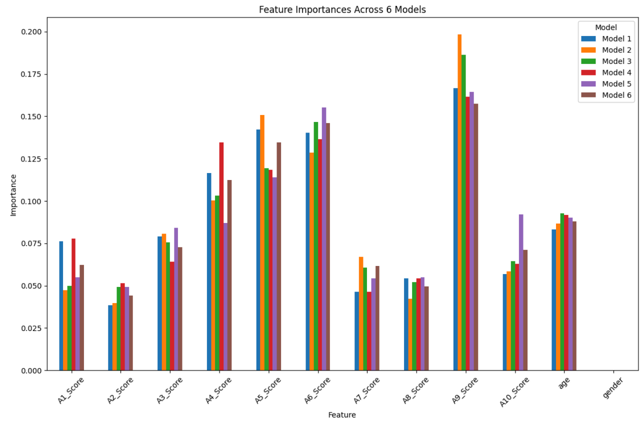

Fig 1.7

Finally, the feature importance across all models.

D. Deployment

The final step of our project methodology involves deploying our machine learning solutions through the development of a web application. This process includes integrating our algorithms and models into a user-friendly interface that can be accessed through standard web browsers. This deployment enables various stakeholders such as caregivers, clinicians, and individuals with autism spectrum disorder (ASD) to seamlessly interact with our tools and services.

Our web application acts as a centralized platform, offering personalized interventions, assessments, and support tailored to the unique needs and characteristics of individuals with ASD. By leveraging machine learning technologies, we aim to provide targeted assistance that is both effective and adaptable to the diverse requirements of the autism community.

By making our solutions accessible online, we prioritize enhancing accessibility, scalability, and usability. This approach empowers stakeholders by providing them with convenient access to cutting-edge tools and resources, ultimately facilitating the utilization of machine learning techniques to tackle the multifaceted challenges associated with ASD.

```
Enter value for A1_Score: 1
Enter value for A2_Score: 1
Enter value for A3_Score: 1
Enter value for A4_Score: 1
Enter value for A5_Score: 1
Enter value for A6_Score: 1
Enter value for A7_Score: 1
Enter value for A8_Score: 1
Enter value for A9_Score: 1
Enter value for A10_Score: 1
Enter value for age: 1
Enter value for gender: M
Severity Classification: High Severity ASD
```

Fig 1.8

The above figure illustrates the metric evaluation of how all the 10 factors contribute towards analyzing the severity of ASD.

IV. LITERATURE REVIEW

The literature review for the survey paper on machine learning-based models for early stage detection of Autism Spectrum Disorders (ASD) encompasses a diverse range of studies, each contributing valuable insights to the field. T. Akter et al. 's work, published in IEEE Access in 2019, delves into the analysis of ASD likelihood in toddlers and adolescents by utilizing various ASD datasets. Another notable contribution, presented at the 2019 International Conference on Electrical, Computer, and Communication Engineering (ECCE) by K. S. Omar et al., introduces a machine learning approach for predicting ASD.

Frith U. and Happé F.'s 2005 study adopts a cognitive approach, exploring symptoms and classifying different types of ASD. A comprehensive review by R. A. Bahathiq et al. focuses on machine learning-based ASD diagnosis, mapping the research landscape and providing a literature-based taxonomy over the past five years. The World Health Organization (WHO) contributes a detailed overview of ASD, including symptoms, analysis, and statistical information on autism patients.

Research by Bone D., Bishop S.L., and others were published in the Journal of Child Psychology and Psychiatry in 2016 on the use of machine learning to improve autism diagnosis and diagnostic tools in terms of quality, efficiency, and many fusion machines. . Multiple sclerosis. Farooq, R. Tehseen, M. Sabir, and Z. Atal investigated ASD in children and adults using various machine learning algorithms. Also S.R. Dutta, S. Datta and M. Roy report the use of a tool algorithm for early stage autism diagnosis.

Together, these studies provide a comprehensive overview of the current landscape, highlighting the significance of machine learning in early-stage ASD detection and addressing various aspects such as prediction techniques, cognitive approaches, and algorithmic applications in different developmental stages.

V. CONCLUSION

Despite extensive research using ASD datasets, there is still much room for improvement in ASD prediction. To examine the key features of ASD, we collected data on early diagnosis of ASD at different stages of life, including infancy, childhood, adolescence, and adulthood. We then analyze the results using different classification methods.
In a randomly distributed sample of test results, we find 100% of the results; this is the best guess for any real test; However, we determined the average value to compare with previous Comparison studies.

REFERENCES:

1. U. Frith and F. Happé, “Autism spectrum disorder,” Current biology, vol. 15, no. 19, pp. R786–R790, 2005.
2. WHO, Autism spectrum disorders, 2017 [Accessed August 22, 2018].
3. Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, Md. Rezaul Karim Rizvi, “A Machine Learning Approach to Predict Autism Spectrum Disorder”,IEEE2019
4. Sushama RAni Dutta, Sujoy Datta, Monideepa Roy, “Using Cogency and Machine Learning for Autism Detection from a Preliminary Symptom”, IEEE 2019
5. Muhammad Shoaib Farooq , RabiaTehseen , Maidah Sabir1 & Zabihullah dAta, “Detection of autism spectrum disorder (ASD) in children and adults using machine learning”, Scientific reports, 2023
6. Reem Ahmed Bahatiq, Haneen Banjar, Ahmed K. Bamaga, Salma Kammoun Jarraya, “Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging”, September 2022
7. Suman Raj, Sarafaraz Masood, “Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques” , ICCIDS 2019
8. D. Bone, S. L. Bishop, M. P. Black, M. S. Goodwin, C. Lord, and S. S.Narayanan, “Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion,” Journal of Child Psychology and Psychiatry, vol. 57, 2016.
9. Jetli Chung and Jason Teo, “Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges”,

Hindawi Applied Computational Intelligence and So Computing Volume 2022, Article ID 9970363, 19 pages

10. Reem Ahmed Bahathiq, Haneen Banjar, Ahmed K. Bamaga and Salma Kammoun Jarraya"Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging" , Frontiers 22 September 2022
11. Kaushik Vakadkar, Diya Purkayastha & Deepa Krishnan "Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques",Springer, 22 July 2021
12. Tania Akter, MD. Shahriare Satu, MD. Imran Khan, Mohammad Hanif,Shahadat Uddin, "Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders"11 November 2019.