iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.58939

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

Detection of Autism Spectrum Disorder using Machine Learning: A Review

Vaishnavi Sirigiri[1], Srilekha Katta[2]
Student, Dept. of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology, Telangana, India

Abstract: ***Social interaction and communication impairments are caused by an illness known as Autism Spectrum Disorder (ASD), which has neurological and genetic components. Statistics from the World Health Organization (WHO) show that the number of people with ASD diagnoses is progressively rising. The majority of recent research focuses on data gathering, brain image analysis, and clinical diagnosis; it does not address the diagnosis of ASD using machine learning. Currently, the only techniques available for diagnosing ASD are clinical standardized testing. This results in longer diagnostic times as well as a sharp rise in medical expenses. Machine learning approaches are being used to supplement traditional methods in order to enhance diagnosis precision and time required. The dataset is being subjected to several machine learning approaches, with the aim of constructing predictive models that are contingent on the results. The best accurate machine learning model to identify the disease in its early stages is then found by analyzing each technique according to its assessment metrics.***
Keywords: Autism Spectrum Disorder (ASD), Machine Learning

I. INTRODUCTION

Early-life brain development is linked to autism, a neurodevelopmental disorder that affects social interactions and interaction problems in a person. While this disease is characterized by limited and repetitive behavioural patterns, the term "spectrum" refers to a broad range of symptoms and severity. Understanding Autism Spectrum Disorder is a multifaceted endeavour rooted in the exploration of complex neurological conditions that impact social interaction, communication, and behaviour. The spectrum nature of ASD encompasses a diverse range of presentations, making it a challenge to both comprehend and address effectively. ASD is characterized by a broad spectrum of symptoms, often appearing in early childhood and persisting throughout an individual's life. The diagnostic criteria include confined and repetitive actions in addition to difficulties with social communication and interaction. But each person with ASD experiences the spectrum nature of the disorder differently, leading to a wide range of demands and experiences. ASD's complexity emerges from a confluence of genetic predispositions, environmental influences, and neurological intricacies, making it a challenge to decipher its underlying mechanisms solely through traditional means. In this landscape, machine learning emerges as a transformative force in ASD research and understanding. Machine learning algorithms have the capacity to parse through vast and intricate datasets, amalgamating genetic markers, behavioural patterns, neuroimaging data, and environmental factors. Even so, using conventional behavioural research, the diagnosis and identification of ASD are extremely complex and challenging. According to the Centers for Disease Control's (CDC) most recent study from 2023, one in every 36 children has an autistic diagnosis. From one in 44 children two years ago, this is an increase. Since the report was just published, it is likely that through 2024 the numbers will remain unchanged. Based on its severity, autism can also be identified at a later age. Typically, it is diagnosed around the age of two. There are numerous therapy approaches available to identify autism as soon as feasible. These diagnostic techniques aren't usually applied in clinical settings until there is a significant risk of autism development. In the last several years, a number of research have been carried out using diverse machine learning techniques to quickly detect and assess ASD as well as other conditions like diabetes, stroke, and heart failure. Machine learning algorithms, with their capacity for pattern recognition, offer the potential to analyse diverse datasets to identify subtle markers or behavioural patterns indicative of ASD. This predictive ability not only aids in early diagnosis but also facilitates a more comprehensive understanding of the developmental trajectories of individuals with ASD, empowering clinicians and caregivers to implement tailored intervention strategies.

II. RELATED WORK

A dataset of adult ASD screening results from the UCI machine learning repository is used in the linked research project. Federative Learning approaches are used in the study in con- junction with Support Vector Machine and Logistic Regression algorithms.

©IJRASET: All Rights are Reserved

By training two distinct ML classifiers—logistic regression and support vector machines—locally for the identification of ASD variables and the detection of ASD in both children and adults, the FL approach has been specifically used to the diagnosis of autism. The central server receives the findings of these classifiers due to FL, and a meta classifier is trained there to identify the best accurate method for detecting ASD in both adults and children.

The development and assessment of a machine learning model intended for the early detection of autism spectrum disorder is the main focus of the research that the researchers did. The authors introduce a method that strives for high accuracy in detecting ASD at its nascent stages. Using a variety of feature selection strategies, the study examines a variety of feature subsets related to toddlers, children, adolescents, and adults. Furthermore, various categorization techniques are used on these subsets. Rigorous performance assessments of the classifiers are conducted through non-parametric statistical tests. The identification of optimal feature subsets and significant features of ASD is facilitated by the integration of an explainable AI methodology. The study encompasses a comparative analysis of eight classifiers, including but not limited to Random Forest, Naive Bayes and Support Vector Machine. The objective is to ascertain the most effective classifier for each age group and feature subset. Furthermore, non-parametric statistical methods are deployed to assess the pairwise significance of the classifiers. The authors put forth an explainable AI approach to interpret the results and pinpoint the significant features associated with ASD. The study provides insightful information about how to create an effective machine learning model for ASD early diagnosis [1].

The paper aims to fill the gap in literature reviews concerning recommender systems. The authors critique existing review papers for their narrow focus on aspects like system evaluation, often overlooking the analysis of dataset descriptions and simulation platforms used. This systematic review delves into recent contributions in recommender systems, spanning applications such as books, movies, and products. The analysis centers on the algorithmic facets, constructing a taxonomy that encompasses the diverse components essential for effective recommender system development. Moreover, the paper scrutinizes datasets, simulation platforms, and performance metrics linked to each contribution. It addresses prevalent challenges like scalability, cold-start, and sparsity, advocating for efficient techniques to surmount these issues. The evaluation of recommender system performance is discussed, emphasizing the absence of a standardized measure. The paper outlines various performance metrics, including recall, MAE, precision, F1-measure, accuracy, and RMSE. Furthermore, the authors underscore the need for expanded research in recommender systems, especially in domains such as health, tourism, and education, where the number of identified research papers is relatively limited. They predict a significant surge in recommender system research in the future [2].

This study proposes a novel machine learning architecture that targets four different age groups (toddlers, adolescents, children, and adults) in order to detect autism spectrum disorders (ASD) in their early stages. The framework combines classification algorithms with feature selection (FS) techniques in an effort to identify the best strategies for each dataset. A wide range of statistical metrics, applied to feature-scaled ASD datasets, are used to assess classification results. These metrics include Accuracy, Precision, Recall, F1-score, Mathews Correlation Coefficient (MCC), Kappa score, and Log loss. According to the study, LDA is the best option for datasets pertaining to adolescents and adults, whereas AB is the best classification strategy for datasets pertaining to toddlers and children. Remarkably great accuracy is attained [3].

[4] The goal of this research is to accurately classify ASDs using a variety of machine learning techniques. The dataset includes 703 individuals, both patients and non-patients, and 16 carefully chosen features. The experiments are carried out in a simulation environment and are analysed using the Waikato Environment for Knowledge Analysis (WEKA) platform. The techniques used in the studies include k-Nearest Neighbours (kNN), J48, Linear Support Vector Machine (SVM), Bagging, Stacking, AdaBoost, and Naive Bayes. Three, five, and tenfold cross-validation are used to calculate the predictions of ASD status. The analysis assesses the specificity, sensitivity, and accuracy of each approach. Comparative findings show that J48, Bagging, Stacking, Naive Bayes, linear SVM, and bagging consistently attain the maximum accuracy at 100. This study's method uses machine learning models to categorize each participant as either having autism spectrum disorder (ASD) or not. The method uses a number of factors, such as age, sex, and ethnicity, to classify the data, the technique of the proposed system, which comprises preprocessing the dataset to eliminate noise, encode categorical categories, and eliminate outliers and missing values. Feature engineering is also used to choose the best features from all the features in the dataset. After preprocessing the dataset, classification techniques such Naive Bayes, Support Vector Machine, K-Nearest Neighbours, Random Forest Classifiers, and Logistic Regression are used to predict the output label (ASD or no ASD) [5].

The paper suggests many approaches, such as supervised learning techniques like Support Vector Machines (SVM), decision trees, and logistic regression, for applying machine learning algorithms in ASD identification and analysis. The use of computational intelligence methods for feature selection and dimensionality reduction, such as Variable Analysis (Va), is also mentioned in the paper [6].

The use of machine learning algorithms to forecast autism spectrum disorder (ASD) is investigated in this work. In order to determine how well various Associative Classification (AC) algorithms predict ASD, the authors examine CBA, CMAR, MCAR, FACA, FCBA, ECBA, and WCBA. The article provides experimental data evaluating these algorithms' precision, recall, accuracy, and F-measure. Using a dataset of adults with autism, the authors carried out a thorough analysis to assess the reliability of the chosen algorithms using different values for minimum support and minimum confidence. The so-called "Weighted Classification Based on Association Rules (WCBA)" method is commended for its exceptional performance in 2018 with regard to F-measure, accuracy, recall, and precision. By these statistical measures, the WCBA algorithm performed better than any other AC method. Additionally, the authors highlight the potential utility of the AC technique in supporting important domains and suggest additional research on this topic by suggesting new AC algorithms or altering current ones to obtain high accuracy when applied to relevant domains. The WCBA algorithm outperforms existing AC algorithms in predicting ASD, indicating that the research offers significant insights into the potential of machine learning techniques in diagnosing and comprehending ASD [7].

This specific study looks on the use of machine learning methods to identify autism spectrum disorder. Three ASD datasets from the UCI database were used in the study to evaluate the machine learning models' sensitivity, accuracy, F-measure scores, and area under the curve (AUC). The study used three machine learning algorithms to identify ASD data: Random Forest (RF), Support Vector Machine (SVM), and k-Nearest Neighbours (kNN). For the kNN technique, the Euclidean distance and the three nearest neighbours were chosen. The kernel function of the SVM approach is the radial basis function (RBF). There are sixty trees that make use of the RF method. Because the study only included three ASD datasets from the UCI database, the findings might not hold true for additional datasets. Furthermore, because the study did not compare the effectiveness of the machine learning techniques with traditional diagnostic techniques, it is unknown how well the machine learning models would perform in a clinical setting [8].

In order to better understand autism spectrum disorder (ASD), 45 experiments that used supervised machine learning are reviewed in-depth in this publication. It covers a broad spectrum of text analysis and categorization algorithms. The purpose of the review is to help researchers interested in developing more clinically, computationally, and statistically sound methods for mining ASD data by identifying and characterizing trends in supervised machine learning in the literature on ASD. Five academics with backgrounds in autism spectrum disorder, computer science, and data science assessed the abstracts, methods, and results parts of the publications as part of the review process. A group of people with ASD had to be included, the publication had to be published in a peer reviewed journal, and the main analytical technique had to be a supervised machine learning model. The paper addresses the use of text mining on a broad scale across multiple fields and its ability to address discrepancies in published findings. It also emphasizes the value of cross-validation in enhancing a model's predictive power on future data and the significance of model correctness and generalizability to new datasets beyond those used for training. The study also covers the advantages of certain supervised machine learning algorithms, like as Naive Bayes, that are employed in ASD research, including their simplicity, computational effectiveness, and competitiveness with more sophisticated models in fields like text categorization. The study also discusses the effectiveness of various machine learning models, with ridge logistic regression and linear SVM performing better than the others in some analyses [9].

The study describes how the researchers used terms and expressions found in child developmental evaluations to build a machine learning system for the surveillance of autism spectrum disorder (ASD). In order to predict the status of ASD cases, the study used a random forest classifier to process evaluation text using a "bag-of-words" approach. The limitations of the approach are also discussed in the article, including the need for additional verification and the potential for bias in the data used to train the algorithm. The study also points out that further research is required to see whether the findings can be generalized to different populations or circumstances and that the methodology might not be appropriate in every instance [10].

III. COMPARISON ANALYSIS

The first set of studies focuses on utilizing machine learning approaches for accurate Autism Spectrum Disorder (ASD) classification. Various techniques, such as linear Support Vector Machine, Naive Bayes, Stacking, Bagging, Adaboost, J48 and K-Nearest Neighbours are employed on a dataset comprising 703 individuals. Results consistently show high accuracies, with linear Support Vector Machine, Naive Bayes, Bagging and Stacking achieving very good accuracies in comparative analyses. Another study investigates k-NN, SVM, and Random Forest on three UCI ASD datasets, emphasizing the need for further validation in clinical contexts. A third study proposes supervised learning approaches, computational intelligence techniques, and Federative Learning with logistic regression and SVM for ASD detection, showcasing diverse methodologies.

The second set of studies presents a thorough analysis of 45 studies using supervised learning in the autism diagnosis research. The review identifies trends and guides researchers, emphasizing the importance of model accuracy and generalizability.

Ridge logistic regression and linear SVM outperform other models in certain analyses. Another study explores a machine learning system for ASD surveillance using a random forest classifier, acknowledging the need for further verification. A separate research introduces a machine learning model for early ASD identification, employing feature selection techniques and explaining AI for interpretation. The studies collectively highlight the potential and challenges of applying machine learning in ASD research.

The third set of studies delves into the application of machine learning algorithms for predicting ASD, particularly using Associative Classification (AC) algorithms. A variety of AC algorithms are evaluated, where the Weighted Classification based on Association Rules (WCBA) algorithm stands out for its superior performance in accuracy and other evaluation metrics. The research highlights the potential of AC approaches and recommends additional investigation in related fields. Another research introduces a novel machine learning framework for early-stage ASD detection, achieving high accuracy with tailored methods for different age groups. Overall, these studies contribute valuable insights into the effectiveness of machine learning in ASD diagnosis and understanding. It is also very important to consider the fact that all the accuracies derived are also dependent on the dataset used, its size and other characteristics.

TABLE I
LITERATURE SURVEY

SNo.	Paper Title & Year	Methodologies Used	Observations
1	Detection of Autism Spectrum Disorder in Children and Adults using Machine Learning (2023)	Logistic Regression and Support Vector Machine	Owing to Federative Learning, the output from these classifiers has been sent to a central server, where the meta classifier is trained to identify the most accurate method for identifying ASD in both adults and children.
2	Efficient Machine Learning Models for Early Stage Detection of Autism Spectrum Disorder (2022)	Support Vector Machine, Random Forest and Naive Bayes	Classification techniques through a comparison of their performance, employing nonparametric statistical tests to ascertain their significance. The suggested approach demonstrates promise in accurately identifying Autism Spectrum Disorder (ASD) at an early stage.
3	A systematic review and research perspective on recommender systems (2022)	meta-heuristic-based approaches, content based filtering, collaborative filtering based approaches, and optimization-based approaches	It observes a scarcity of papers addressing recommender systems in health, tourism, and education. Additionally, the paper emphasizes the lack of a uniform measure for evaluating recommender system performance, pointing out the utilization of varied evaluation metrics in the reviewed papers.
4	A Machine Learning Framework for Early Stage Detection of Autism Spectrum Disorders (2022)	Support Vector Machine, Decision Tree, Logistic Regression, Linear Discriminant Analysis and Adaboost	Introduced evaluating multiple classifiers and feature selection techniques. Adaboost and Linear

			Discriminant Analysis emerge as the top-performing classifiers for distinct age groups.
5	Classification of Adult Autistic Spectrum Disorder using Machine Learning approach (2021)	Linear Support Vector Machine, AdaBoost, Naive Bayes, Bagging, Stacking, K-Nearest Neighbours and J48 Decision Tree	Three, five, and ten-fold cross-validation are used to calculate the predictions of ASD status. The analysis assesses the specificity, sensitivity, and accuracy of each approach. According to comparison results, J48, Bagging, Stacking, Naive Bayes, and linear SVM consistently attain the maximum accuracy at 100.
6	Detection of Autism Spectrum Disorder in Children using Machine Learning Techniques (2021)	Logistic Regression, K-Nearest Neighbours, Naive Bayes, Support Vector Machine and Random Forest Classifiers	The classification is made by the system using a variety of characteristics, including age, sex, and ethnicity. The suggested system's methodology, which entails preprocessing the dataset to remove noise, encode categorical categories, and remove missing values and outliers.
7	Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques (2020)	Support Vector Machine, Decision Trees, and Logistic Regression, Variable Analysis	Used Support Vector Machines (SVM), decision trees, and logistic regression, computational intelligence techniques like Variable Analysis (Va) for feature selection and dimensionality reduction.
8	Predicting Autism Spectrum Disorder using Machine Learning Technique (2020)	Classification Based on Association (CBA), Classification Based on Multiple Association Rules (CMAR), Fast Associative Classification Algorithm (FACA), Enhanced Classification Based on Association (ECBA), Weight based Classification Based on Association (WCBA)	Considered various Associative Classification (AC) algorithms. Notably, WCBA algorithm stands out for its exceptional performance in accuracy, recall, precision, and F-measure. The findings of this study have implications for improving our knowledge of ASD and our capacity to diagnose it using machine learning methods.

9	Autism Spectrum Disorder Detection with Machine Learning Methods (2019)	Random Forest, k-Nearest Neighbours and Support Vector Machine	Three nearest neighbours and the Euclidean distance were chosen. There are sixty trees that make use of the Random Forest method. Because the study only included 3 ASD datasets from the UCI database, the findings might not hold true for additional datasets.
10	Applications of Supervised Machine Learning in Autism Spectrum Disorder Research (2019)	Naive Bayes, Logistic Regression and Linear SVM	It underscores the role of cross-validation in enhancing a model's predictive capacity for future data. The study explores the benefits of several supervised machine learning methods, including Naive Bayes, that are used in ASD research.
11	Development of a Machine Learning Algorithm for the Surveillance of Autism Spectrum Disorder (2016)	Bag-of-Words, Random Forest Classifier	"Bag-of-Words" approach was employed to extract words. These linguistic features were utilized to train a random forest classifier. The created machine learning method achieved a sensitivity of 0.83 and a positive predictive value of 0.76, indicating great accuracy in the classification of ASD.

IV. CONCLUSION

In conclusion, the extensive examination of existing methodologies consistently demonstrates their efficacy in handling specific aspects of the Autism Spectrum Disorder at hand. However, a noticeable limitation surfaces when these approaches are confronted with a broader scope of variables. Acknowledging this challenge, we advocate for the development of a novel and dedicated Machine Learning model, designed specifically to tackle the intricacies within a more extensive spectrum of parameters. We propose a strategy that centers around the creation of a specialized ML model, honed to navigate the complexities and enhance accuracy across a diverse array of factors.

By addressing the inherent challenges associated with a broader range of considerations, our approach aims to not only bridge existing gaps but elevate the precision in categorization. The machine learning component of our suggested methodology focuses on predicting and understanding the unique needs and behaviors of individuals on the autism spectrum. By analyzing a diverse range of data, including behavioral patterns, feature scaling techniques, sensory sensitivities, and individual responses to various stimuli, the system aims to provide personalized insights for caregivers and professionals. Apart from the machine learning component, the technique highlights the significance of taking preventive measures that are customized to meet the unique requirements of people with ASD.

This includes creating a user-friendly interface that allows caregivers to input and track potential triggers, enabling a proactive approach to managing challenging behaviors and promoting a supportive environment. Furthermore, we also suggest incorporating a nutrition recommendation module and therapy recommendation system, considering the impact of diet and lifestyle on individuals with ASD.

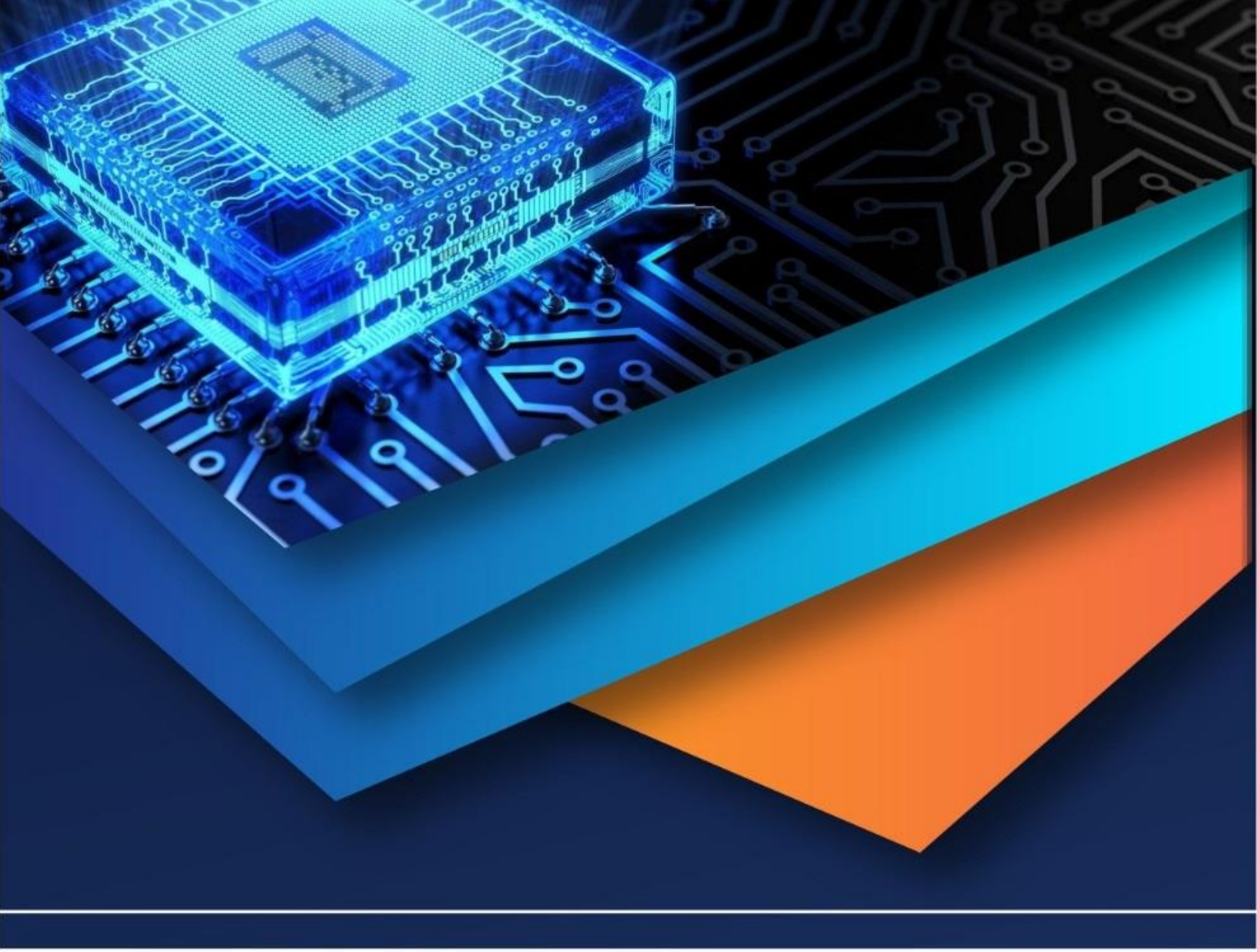
V. ACKNOWLEDGMENT

We are grateful to the Chaitanya Bharathi Institute of Technology's Department of Computer Science and Engineering for their outstanding collaboration and priceless input.We would want to use this occasion to express our gratitude to everyone who has supported us throughout this project. We could not have progressed with the research article without their enthusiastic support, encouragement, and collaboration.

REFERENCES

[1] M. S. M. H. F. M. M. Bala, M.; Ali, “Efficient machine learning models for early stage detection of autism spectrum disorder,” 2022.
[2] D. Roy and M. Dutta, “A systematic review and research perspective on recommender systems,” 2022.
[3] S. M. Mahedy Hasan, M. P. Uddin, M. A. Mamun, M. I. Sharif, A. Ulhaq, and G. Krishnamoorthy, “A machine learning framework for early-stage detection of autism spectrum disorders,” IEEE Access, vol. 11, pp. 15038–15057, 2023.
[4] Nurul Amirah Mashudi, Norulhusna Ahmad, Norliza Mohd Noor, "Classification of adult autistic spectrum disorder using machine learning approach", IAES International Journal of Artificial Intelligence (IJ-AI) Vol. 10, No. 3, September 2021, pp. 743~751 ISSN: 2252-8938, DOI: 10.11591/ijai.v10.i3.pp743-751.
[5] Kaushik Vakadkar, Diya Purkayastha, Deepa Krishnan,"Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques"(2021)
[6] S. M. Suman Raja, “Analysis and detection of autism spectrum disorder using machine learning techniques,” in International Conference on Computational Intelligence and Data Science (ICCIDS), 2019.
[7] R. G. Jaber Alwidian, Ammar Elhassan, “Predicting autism spectrum disorder using machine learning technique,” in International Journal of Recent Technology and Engineering (IJRTE), 2020.
[8] Uğur Erkan, Dang N.H. Thanh, "Autism Spectrum Disorder Detection with Machine Learning Methods",Current Psychiatry Research and Reviews, 2019, 15, 297-308.
[9] K. K. H. . M. N. N. . N. L. . C. P.-P. . R. A. . D. R. D. . E. Linstead, “Applications of supervised machine learning in autism spectrum disorder research,” Review Journal of Autism and Developmental Disorders, pp. 1–5, 2019.
[10] V. N. B. K. C. D. S. L. Maenner MJ, Yeargin-Allsopp M, “Development of a machine learning algorithm for the surveillance of autism spectrum disorder,” 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), 2016.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)