Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.10 : 2024
ISSN :**1906-9685**

A Comprehensive Study of Machine Learning Approaches for ASD Prediction

[#1]LokeshSirela,AssistantProfessor, RaghuEngineering College(A),Dakamarri,Visakhapatnam,India

lokesh.sirela@raghuenggcollege.in

[#2]Venkata Sai Sampath Kumar Gudivada, Student, Raghu Engineering College(A) Dakamarri,Visakhapatnam,India

sampathgudivada21@gmail.com

[#3]Veluturi Sai Prashanth ,Student, Raghu Engineering College(A) Dakamarri, Visakhapatnam, India

saiprashanth506@gmail.com

[#4]Vanapalli Mounika Varalakshmi, Student, Raghu Engineering College(A) Dakamarri,Visakhapatnam, India

mounikavanapalli22@gmail.com

[#5]Omkar Pudi, Raghu Engineering College(A) Dakamarri,Visakhapatnam, India

omkarpudi88@gmail.com

Abstract - Autism Spectrum Disorder (ASD) is a complex neurological condition that affects individuals of all ages, impacting their mental, social, and physical well-being. Traditional diagnostic methods for ASD are time-consuming and expensive. Machine learning techniques offer a promising approach to automate and improve the efficiency of ASD diagnosis. This study aims to identify key characteristics for automating the diagnostic process and evaluates several machine learning algorithms, including Logistic Regression, K-Nearest Neighbor (KNN), Support Vector Machine (SVM), and Naïve Bayes, for predicting ASD occurrence. Experimental results show that the Naïve Bayes algorithm achieves the highest accuracy, reaching 99.6% compared to other algorithms.

Keywords: Machine Learning, Autism Spectrum Disorder, KNN, Logistic Regression, Naïve Bayes, SVM.

1.INTRODUCTION

The human brain, serving as the central organ in the body, coordinates various functions. Autism arises from neural disconnection and disturbances in brain maturation. Although it can appear at any age, autism typically emerges in childhood as a developmental condition. People with Autism Spectrum Disorder (ASD) face challenges in social interaction and communication, often showing restricted interests and repetitive behaviors that affect their daily lives. Accordingto the WHO, autism affects

around one in every hundred and sixty children. Some individuals with autism can live independently, while others may need lifelong care and support.Detecting autism can be time-consuming, but early detection offers significant benefits. It enables timely intervention and appropriate treatment, preventing further health deterioration and reducing the need for unnecessary treatments, thus lowering costs.

This study aims to use classification methods to determine if an individual has Autism Spectrum Disorder (ASD). It also aims to identify the optimal classification algorithm for predicting ASD, using performance criteria such as error rate and accuracy. The dataset on autism is preprocessed, converting textual data into numerical format using techniques like One Hot Encoder and Label Encoder. Data purification is then performed using the mean method. Features in the dataset are identified using training data, while testing data is used for result validation. Significant attributes are selected through feature selection techniques.The efficiency of the decision-making algorithm is evaluated based on criteria such as correctness and inaccuracy frequency. The classifier demonstrating superior effectiveness is considered the most suitable for ASD prediction. After implementing various classification models for autism prediction, it is observed that Naïve Bayes consistently outperforms other algorithms, providing significantly better results.

2. LITERATURE REVIEW

Within the research by Sunsirikul they detail their endeavors to create a data analytics instrument aimed at aiding future medical practitioners in diagnostic procedures. Their investigation entailed extracting behavioral trends from data and establishing a behavioral guideline for patients, with the objective of identifying potential associations between specific behaviors and symptoms of autism, relying on an ample amount of patient records. The paper delves into various data mining techniques, aiming to furnish physicians with a comprehensive set of tools for the intelligent evaluation of patient data. One notable outcome of this study is its revelation of the association between behavioral patterns in autistic children and PDD NOS, potentially leading to improvements in self-esteem, mitigating disabilities, and refining disorder classifications. Moreover, the research highlights the absence of medical information regarding typical children throughout the training period.

Osman et al. developed a methodology that utilizes data analysis methods to diagnose Autism Spectrum Disorder (ASD) in patients. ASD significantly impacts an individual's well-being, characterized by social and interpersonal challenges, repetitive actions, and intense interests or hobbies. This study focuses on using classification algorithms to identify ASD in children. The classification process results determine whether a child is diagnosed with ASD or not. The LDA algorithm achieved an accuracy of 90.8%, while the KNN algorithm reached 88.5% accuracy.

Cincy Raju et al. highlighted heart disease as one of the deadliest illnesses leading to mortality, causing profound long-term damage. This study aims to use data analysis techniques to provide an effective solution for therapeutic scenarios. Various classification algorithms were used, with the Support Vector Machine (SVM) proving to be the most superior among them.

Erik et al. introduced a model for ASD screening that incorporates machine learning adaptation and DSM-5 criteria. Their framework aims to predict autism and introduce a comprehensive healthcare system for gathering, assessing, and managing data related to the evaluation and therapy of children with ASD. The Autism Management Platform (AMP) serves as an intelligent web interface and statistical platform, facilitating real-time collection and extraction of patient data by healthcare professionals and specialists. It also provides automated feedback to adjust data filtering preferences. Similar efforts in predictive analysis have been undertaken.

Canon et al. have provided empirical evidence regarding the prognosis of autism spectrum disorder. This manuscript identifies and presents supplementary evidence needed for further investigation. However, it underscores the importance of examining the functional impacts on individuals.Karunakaran et al. have introduced a method that integrates an adaptive functioning classifier with early learning techniques. This method effectively addresses the challenge of handling less noisy data, a notable limitation. Additionally, alternative approaches for predicting autism spectrum disorder include machine learning analysis and pathway analysis.

The current methods for predicting autism spectrum disorder lack promise, primarily due to the omission of critical parameters in the analysis. Therefore, it is imperative to incorporate all essential parameters to enhance the effectiveness of the proposed algorithm**.**

3.PROPOSED METHODODLOGY

This section outlines the proposed scheme for analyzing autism disorder. The process involves the following steps:

(a) Data Gathering

(b) Data Preprocessing

(c) Model Development

(d) Training and Validation

(a) Data Gathering We collected a dataset comprising 1054 unique cases for predicting autism spectrum disorder. Essential attributes from this dataset were used to train our model. The dataset includes various fields such as indications, age groups, familial backgrounds, living locations, and more.

(b) Data Preprocessing Data often contains missing components, discrepancies, inaccuracies, and non-numeric entries. Data preprocessing offers a robust solution to address these challenges. In this study, we preprocess the data using One Hot Encoding and Label Encoding methodologies. This strategy effectively converts qualitative data into numerical formats.

In the Autism Assessment dataset, data preprocessing involves data cleaning techniques followed by feature selection to filter out irrelevant or redundant features. Subsequently, classification is performed using algorithms such as LDM, SVM, and Naive Bayes. This is followed by forecasting outcomes and assessing for superior precision.

(c) Model Development Various methodologies are employed in constructing the model, aiding in the selection of the most suitable model for achieving the desired outcome. The algorithms used include LDM, SVM, and Naive Bayes.

(i)LDMLinear Discriminant Analysis (LDA) is a statistical method used for predicting qualitative outcome measures using predictor variables. LDA aims to maximize the separability between classes while minimizing the variance within each class. This makes LDA a powerful tool for pattern recognition and classification tasks.

(ii)SVM :When dealing with unlabeled data, supervised learning methods become impractical, leading to the necessity of using unsupervised techniques. In unsupervised learning, data is naturally grouped into clusters, and new data points are then mapped to these clusters. Support Vector Machines (SVM).

The effectiveness of Support Vector Machines (SVM) depends on the choice of kernel, its configurations, and the soft margin setting. The Gaussian kernel, which has just one parameter, is a commonly preferred option. However, SVM has some limitations:

•Complete labeling of data inputs is required.

•SVM's calibration of probabilities for classes may be inaccurate, as it is based on Vapnik's theory, which eliminates the need for probability estimation from finite data.

•SVM is primarily suitable for two-class tasks, so techniques are needed to transform multi-class tasks into a series of binary problems.

•The interpretation of parameters in a solved model can be challenging.

(iii)NaïveBayes

The Naive Bayes algorithm is a well-known method for building predictive models. It uses an "event-based model" based on hypotheses about feature distributions, often employing Multinomial and Bernoulli distributions for categorical attributes encountered in document sorting. This leads to the development of two distinct yet often misinterpreted models.

(iv)Training and Validation: After selecting different models, the dataset undergoes training using various construction models. Subsequently, the data is tested across all models, and the one demonstrating superior accuracy is chosen for further use.

4.TESTED RESULTS AND ANALYSIS

Inthispart,testoutcomesareshowcasedanddeliberated. The format of the dataset is visually illustrated,and the efficiency of the algorithms in forecasting autismdisorder ismeticulouslyexamined.

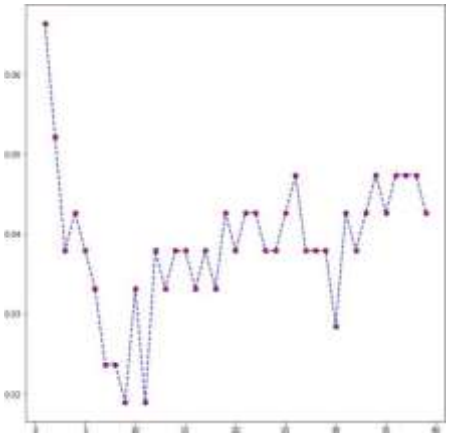
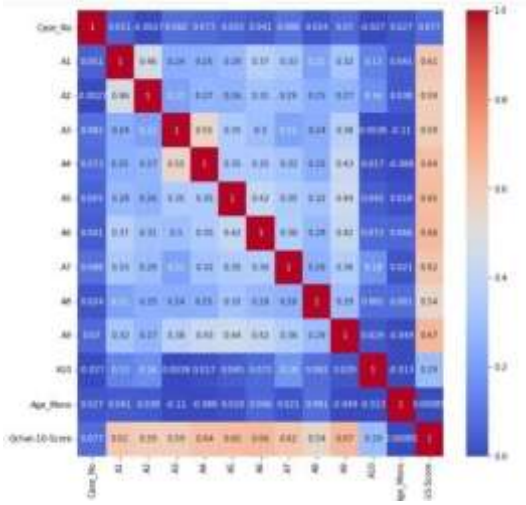


Figure1illustratesthefeaturespresentindatapool

The data pool utilized for identifying autism spectrum condition comprises 1054 observations and includes 19 distinct attributes such as symptoms ,gender, age, ethnicity, nationality, etc.

Thedatasetundergoespreprocessing,andappropriateattributes are chosen for establishing the framework. Prior toframework development and construction, the data's patternsare comprehended through visualization. Figure 3 displaysthe distribution of traits among individuals with and withoutautism,indicatingthecountofaffectedandunaffectedindividualsacrossdifferentcharacteristics.
This following figure illustrates the association amongthe features withindata pool.
Fig 2 heat map



5.Principle Component Analysis

Principal Component Analysis (PCA) is a technique used for dimensionality reduction in data. It helps in transforming a large set of variables into a smaller set that still contains most of the information in the original set.In the context of your project, PCA can be applied to reduce the dimensionality of your dataset, which has 19 distinct attributes. By doing so, you can potentially improve the efficiency of your algorithms in forecasting autism disorder.

Key Aspects of PCA:

Dimensionality Reduction: When conducting principle component analysis (PCA), the goal is to minimize the number of variables as much as possible while retaining as much of the initial information as is practically practicable. A smaller set of variables that are not related to one another is used to accomplish this goal. The original variables are reduced to this smaller set of variables. Principal components are the variables that are reduced to this smaller group. These variables are referred to as the ones that are reduced to this smaller group. It is structured in a descending sequence according to the amount of variation that each component is able to capture, with the first few components preserving the greatest degree of variability in the dataset. This indicates that the components are organized in a manner that is declining. The purpose of this design is to provide the greatest amount of information that can be preserved.

Identifying Patterns and Relationships: Principal component analysis (PCA) is a helpful approach for discovering underlying patterns or structures in the dataset. PCA is an acronym that stands for principal component analysis. It is via this that the variables that tend to change together as well as the ways in which they contribute to the overall variability that is detected in the data are brought to light. This is the case because of the fact that this approach is utilized. This makes it easier to identify the key variables

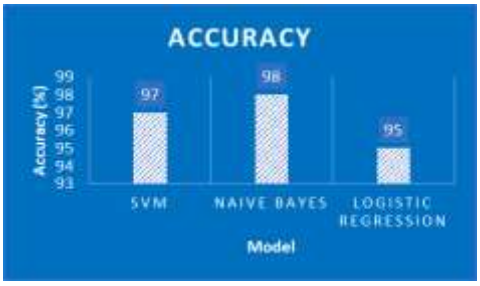
that lead to differences in water quality measurements, as well as likely sources of pollution or variations in regional distribution. Additionally, it makes it easier to identify the components that contributed to the variations in water quality data.

Visualization: Through the utilization of principal component analysis (PCA), which permits the presentation of high-dimensional data in lower dimensions, the data are simplified in terms of their complexity for the purpose of analysis. There are a variety of plots and graphs that can be applied to highlight the relationships that exist between variables, clusters, or groups. Some examples of these plots and graphs are principal component analysis biplots and scree plots. The application of statistical analysis is one method that can be utilized to achieve this success. It is likely that this will be helpful in identifying trends or outliers in the data and information.

Data Compression: Principal component analysis, also known as PCA, is a method that enables the compression of data while still preserving the fundamental components of the dataset. PCA is an acronym that stands for an analysis approach. This decrease in dimensionality also makes it feasible to present complex data findings in a form that is easier to understand. In addition, subsequent studies are simplified as a result of this reduction in dimensionality.

Approach	Accuracy(%)	Error Rate(%)
K-NearestNeighbour	98.2	0.06
LogisticRegression	97.2	0.07
SupportVector Machine	98.4	0.05
NaïveBayes	99.6	0.02

Table1presentsanoverviewofthecontrastbetweenexisting and suggestedmethodologies.**Figure-3** Accuracy



6.CONCLUSION

Autism is one of the vital issues which cannot be prevented but can be treated and it's a big challenge for any family to have a child with autistic disorder. Therefore, it's importance to diagnose it, as early as possible. One of the major issues in today's research is to have improved diagnostic tools to have faster, effective and accurate result. Machine learning techniques have shown favorable results. Machine learning with its leverage methods can be used to diagnose ASD. Use of machine Learning are to create algorithms that are vigorous and have contrive instruments. By using different ML algorithms researchers were able to build such model which shows improved results in terms of accuracy and precisions. In this analysis various classifiers have been used like SVM, Random Forest, decision tree, LMT, KNN, Naïve Bayes', LDA and so on. Among all SVM have shown the best result with an accuracy of 98.27%.

REFERENCES

[1] Karunakaran, P., & Hamdan, Y. B. "Early prediction of autism spectrum disorder by computational approaches to fmri analysis with early learning technique." Journal of Artificial Intelligence, 2(04), 207-216, 2020.

[2] Cannon, J., O'Brien, A. M., Bungert, L., & Sinha, P. "Prediction in autism spectrum disorder: a systematic review of empirical evidence." Autism Research, 14(4), 604-630, 2016.

[3] Madhukeerthana, Y. BevishJinila, "A review on rough set theory in medical images", Research Journal of Pharmaceutical, Biological and Chemical Sciences, Vol.7, Issue 1, pp. 815-822, ISSN : 0975-8585, 2016.

[4] Lord, C., Elsabbagh, M., Baird, G., & Veenstra-Vanderweele, J.(2018). Autism spectrum disorder. The Lancet, 392(10146), 508- 520.

[5] Chaitra, N., Vijaya, P. A., & Deshpande, G. "Diagnostic prediction of autism spectrum disorder using complex network measures in a machine learning framework." Biomedical Signal Processing and Control, 62, 102099, 2020.

[6] Kojovic, N., Natraj, S., Mohanty, S. P., Maillart, T., & Schaer, M. "Using 2D video-based pose estimation for automated prediction of autism spectrum disorders in young children." Scientific Reports, 11(1), 1-10, 2021