Comparison of Different Machine Learning Algorithm for Early Detection of Autism Spectrum Disorder

Tintu Varghese[1] , R.Sri Devi* [2]

Abstract

Rate of kids affected with extreme Autism Spectrum Disorder is increasing rapidly. The exceptionally autistic children challenge problems in conception thinking, intellectual capacity, learning and related skills. The problems that exceptionally autistic children expertise is various. The Autistic Children encounter problems in learning aptitudes, interaction with others and coping with social skills. If autistic children are recognised early, they will have a high quality life by providing thorough care and medical aid. The objective is to conduct a comparative analysis of different machine learning algorithms for the purpose of predicting Autism Spectrum Disorder at an early stage.

Keywords— Autism Spectrum Disorder, Machine Learning, Classification

I. INTRODUCTION

A developmental disorder Autism often known as Autism Spectrum disorder is characterized by difficulties with behaviour, social interaction and communication. The condition lasts lifetime, and each person will experience it differently in terms of symptoms.

ASD can affect individuals of all ages, socioeconomic backgrounds, and various racial communities. ASD diagnosis occurs within the first two year of life, though it can sometimes happen in later. The diagnosis of this condition in adults might come with very severe symptoms or it can come with only minor difficulties.

Although those with Autism Spectrum condition may not exhibit all the sign, they frequently do:

- Making eye contact difficult
- Having trouble listening to and participating in talks.
- Lack of pretend play.
- Inconsistent facial expressions.

Nearest Neighbor (KNN), Linear Regression (LR), and Support VectAn individual diagnosed with ASD might face an increased likelihood of wandering, potentially posing risks to their own safety as well as that of their family and society because he has the potential to damage both himself and other people in act of hostility. So it becomes crucial to identify any indicator that could indicate severe ASD as soon as feasible.

In order to determine whether a person has any characteristics of ASD or not, we used variety of

Machine Learning approaches like LDA (Linear Discriminant Analysis), Naïve Bayes (NB), Classification and Regression Trees (CART), K-or Machine (SVM) are utilized to classify data.

In this instance, we have opted for 70% allocation of data for training purposes and reserved 30% for testing purposes. In our findings, accuracy achieved through KNN is 68.02 %, LR 72.04%, LDA is 73.1%, CART is 68.5%, Naïve Bayes is 69.8% and SVM is 71.2%.

II. LITERATURE REVIEW

Many investigations and test can be carried out in the field of medical science to forecast the disease as soon as possible to potentially initiate appropriate treatment without delay.

One method employed for diagnosing diabetic retinopathy involves the utilization of various

Machine learning techniques like the Naïve Bayes Classifier,

Department of Computer Science,
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
* Corresponding Author

Decision Tree, KNN and SVM are considered in the analysis. Among these algorithms, the SVM demonstrates the highest level of accuracy, as highlighted by Ridam Pal[1].

An alternative method involves the utilization of KNN and LDA for the prediction of ASD disease symptoms [2].

In another study, the K-Means clustering analysis method was employed to classify difficult behaviours in ASD[3].In a separate work, Sami S Alwakeel introduced a Wireless Sensor Network system incorporating machine learning techniques to detect activities associated with autism[4].This system combines eye gaze patterns with demographic features descriptors for improved accuracy.

Shan Canavan conducted autism classification in their study[5].Data mining techniques were employed by Salha M Alzahani to both determine the presence of and forecast heart disease[6].For cardiovascular disease diagnosis, Shauman and AkhilJabbar utilized the K-Nearest Neighbours (KNN) algorithm [7][8]. Nassar H Sweilam applied the Support Vector machine (SVM) approach to cancer diagnosis[9].In a similar way, Suruchi Pimple and Anuja Kumari employed Support Vector Machine methods to diagnose diabetes[10][11].The exploration of functional patterns for the identification of various diseases and the determination of appropriate solutions represents a significant and prevalent application of machine learning techniques[12]

III. ALGORITHMS AND METHODS

A. K Nearest Neighbour

KNN stands as the most frequently employed classification algorithm due to its quick computation time and straightforward analysis of results. The foundation of KNN rests upon a similarity measure, the distance measure from the trained available instances is calculated in order to categorise the new cases. The formula for calculating the distance between points A and B using the normalised Euclidean metric is as follows:

Distance(x,xi)=sqrt(sum(xj-xij)2) (1)

KNN entails four steps.

In an alternative formulation, the process involves calculating the Euclidean distance between the new dataset and the training dataset, followed by sorting the resulting distances. Subsequently, the smallest distance is selected, ultimately determining the class.

B. Linear Regression

It stands as the most widely used statistical leaning technique and ranks among the most straightforward strategies for supervised learning.

. This foresees that the predictor variable A will result in a quantitative reaction B. But it makes the assumption that variables A and B are linearly connected.

$B=\beta 0+\beta 1A1+\beta 2A2+\cdots+\beta nAN$ (2)

Top of Form

In Equation 2, the response is represented as B,the values of β, referred to as model coefficients, undergo a “learning” process during model fitting or training.The coefficient β0 corresponds to the intercept, whileβ1 pertains to the coefficient associated with A1 denotes the first feature, and βn signifies the coefficient associated with an, the nth feature.

C. Linear Discriminant Analysis

Models based on LDA furnish a predictor(s) set X, which forecasts the probability. This algorithm individually models the arrangment of predictor variable X within every response class (given Y; s) and applies Bayes’ theorem to transform them into estimates. If these distributions exhibit normality, this model begins to bear a strong resemblance to logistic regression.

D. Support Vector Machine

Support Vector Machines create decision boundaries by following the concept of decision planes, utilizing this concept to delineate boundaries. The assortment of entities is partitioned through the utilization of decision planes, contingent upon their memberships to different classes. This technique segregates instances of diverse class labels by

arranging the task into hyper planes within a multi dimensional space.SVM excels in managing categorical data as well as multiple continuous variables, and it accommodates both classification and regression assignments. In order to manage categorical variables, a dummy variable is generated; assigning values that can take on either 0 or 1

E. Naive Bayes

Bayes' Theorem operates based on the concept of conditional probability. Conditional Probability pertains to the probability of an event happening when another event has already occurred. The mathematical representation of Bayes' Theorem can be expressed through the equation.

P (B)=P(B|A)P(A)/P(B) where events A and B are considered, and it's important that P(B)$\neq$0.

In a boarder context, the focus is on determining the likelihood of event A occurring when event B has taken place. P(A) represents the antecedent of event A, which is essentially the probability before or the probability of event occurrence A before any supporting evidence is observed. In this context, the evidence is represented by the occurrence of event B.

F. CART

Classification And Regression Tree (CART) model illustrates how the values of a target variable can be forecasted using the values of additional variables. CART generates a decision tree that provides predictions for the outcome variable at each leaf node and splits predictor variables at each branching point. The CART algorithm uses a series of questions, and depending on the answers, the next question is chosen. As a consequence of these questions, a tree-like structure is created, with the terminal nodes acting as leaf nodes when there are no more questions to be answered.

IV. RESULT AND DISCUSSION

Although the ASD can manifest at various stages of life, it is crucial to get a diagnosis as soon as possible. The methods include KNN, LDA,SVM, CART, Naive Bayes, and linear regression have been chosen in this instance to find a person with ASD.To assess this application, we employed 30% of the data for testing purposes and allocated 70% for training. The data selection was randomized.

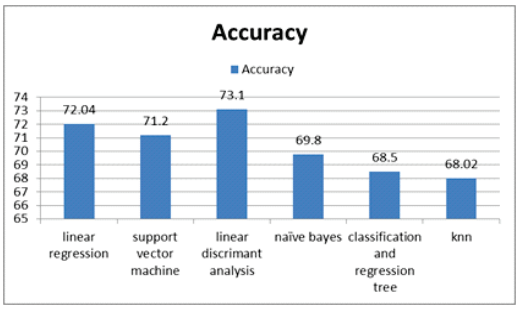

Figure1.Accuracy Chart

This article examined six distinct studies focused on predicting autism spectrum disorder through various machine learning algorithms. The analysis of comparative results for different classification algorithms is showcased in figure 1.Notably,the findings indicate that Linear Discriminant Analysis stands out as the most efficacious algorithm, achieving a peak accuracy of 73.1%

V. CONCLUSION

Many people can avoid destroying their own and other people's lives by getting an early diagnosis of ASD.The K-Nearest Neighbours(KNN), Support Vector Machine(SVM),LogisticRegression(LR),Classification and Regression Trees(CART),Naïve Bayes and Linear Discriminant Analysis(LDA) algorithms were utilized for the classification task..Here, we turn the data for a few attributes into numbers.

Our implementation demonstrates that Linear Discriminant Analysis algorithm produces the most favourable outcome, reaching 73.1% accuracy, and outperforms alternative techniques in terms of precision.

REFERENCES

[1] Ridam Pal, Dr.JayantaPoray and MainakSen, "Application of Machine Learning Algorithms on Diabetic Retinopathy" May 2017 in 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT).

[2] Osman Altay and Mustafa Ulas, " Prediction of the Austim Spectrum Disorder Diagnosis with Linear Discriminant Analysis Classifier and KNearest Neighbour." 2018 in 6th International Symposium on Digital Forensic and Security (ISDFS)

[3] Elizabeth Stevens, Abigail Atchison, Laura Stevens, Eshter Hong, Doreen Granpeesheh, Dennis Dixon and Erik Linstead, " A cluster Analysis of challenging Behaviors in Autism Spectrum Disorder ", 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)

[4] Sami S. Alwakeel, BassemAlhalabi, HadiAggoune and Mohammad Alwakeel, " A machine Learning Based WSN System for Autism Activity Recognition" 2015. IEEE 14th International Conference on Machine Learning and Applications (ICMLA)

[5] Shaun Canavan, Melanie Chen, Song Chen , Robert Valdez , Miles yaeger , Huiyi Lin and Lijun Yin, " Combining Gaze and Demographic Feature Descriptors for Autism Classification", 2017 in IEEE International Conference on Image Processing (ICIP)

[6] Salha M. Alzahani, AfnanAthopity, AshwagAlghamdi, BoushraAlshehri and SuheerAljuaid, " An overview of Data Mining Techniques Applied for Heart Disease Diagnosis and Prediction", December 2014.

[7]Mai shouman , Tim Turner , Rob Stocker , " Applying K Nearest Neighbor in Diagnosing Heart Disease Patients", International Journal Of Information and Education Technology, Vol. 2 No. 3,2012.

[8] M. AkhilJabbar , B.L Deekshatulua , Priti Chandra, " Classification of Heart Disease usung K- Nearest Neighbour and Genetic Algorithm ", CoRR, Volume abs/1508.02061, arXiv:1508.02061.2015.

[9] Nasser H. Sweilam, A.A Tharwat, N.K Abdel Moniem," Support Vector Machine for Diagnosis of Cancer disease: A comparative Study" Egyptian Informatics journal, volume 11, issue 2, PP, 81-92,2010.
22,2013.

[10] Thabtah F.(2017 Dec] UCI Machine Learning Repository [http:// archieve. ics.uci.edu/ml], Irvine, CA: University of California, School of information and Computer Science.

[11] Anuja Kumari, R. Chitra, " Classification of diabetes Disease using Support Vector Machine " International Journel of Engineering Research and application vol.3, pp 1797-1801,ISSN:2248-96.

[12]http://refhub.elsevier.com/S2665-9174(23)00110-12

[13]http://refhub.elsevier.com/S2665-9174(23)00110-1/sref1