International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 114-119

Autism Neuro Developmental Disorder Analysis Using Machine Learning Models

Dr. Madhu B G[1], Dr. Sunitha M R[2], Nischal B J[3], Sambram Gowda S M[4], Romiya Taranum[5], Ayesha Shabnam[6]

[1,] *(Dept. of CS&E, Adichunchanagiri Institute of Technology, Chikkamagaluru, India)*
[2,] *(Dept. of CS&E, Adichunchanagiri Institute of Technology, Chikkamagaluru, India)*
[3,] *(Dept. of CS&E, Adichunchanagiri Institute of Technology, Chikkamagaluru, India)*
[4,] *(Dept. of CS&E, Adichunchanagiri Institute of Technology, Chikkamagaluru, India)*
[5,] *(Dept. of CS&E, Adichunchanagiri Institute of Technology, Chikkamagaluru, India)*
[6,] *(Dept. of CS&E, Adichunchanagiri Institute of Technology, Chikkamagaluru, India)*

Abstract: *Autism disorder is Spectrum of neuro developmental disorder which affect the behavior and mental process of gaining knowledge and comprehension. Autism is observed in all the age groups of children, adults and adolescents. Hence to draw the inference this paper uses the Autism dataset of all the age groups. Machine learning algorithms like SVM, KNN and Naïve Bayes is used in classifying and comparison of disorder.*

Keywords *- Neurodevelopmental Disorder, Screening, Autism, adolescents, behavior.*

I. Introduction

The autism disorder is also called as Asperger. Diagnosing this disorder can be done at any time after first two years. Autism [2]victims might confront many kinds of difficulties, for example, trouble in centring, capacity forlearning is low, mental shortcoming issue like tension, sadness, tangible issues, etc. At the present pace, mental imbalance all over the planet is detonating variously and it is expanding at quicker rate. By WHO review it is reasoned that out of 160 kids 1 will experience the ill effects of autism. Individuals experiencing mental imbalancemight live without reliance, but few might require reliance.

The conventional conclusion is tedious and exorbitant as well. Early identification of confusion can assist withdiagnosing patients with legitimate solution at a beginning phase as it were. It assists the patient with improving with their wellbeing and decrease the expense of post late analysis. So natural, exact and helpful screening test instrument is particularly deprived by which would foresee autism characteristics in an individual and distinguishwhether they require a complete mental imbalance treatment further. The primary subject behind this work is to propose ML model that foresee medically introverted individual precisely for any matured individual. Subsequently, this work upgrades the genuine strategy and further set forth the progression to work on the productivity in distinguishing proof.

II. Releted Work

The pre-handled [1] fMRI of ASD and ID are downloaded from an ABIDE (Autism Brain Imaging Data Exchange). Withstand is a multisite stage that accumulated practical and underlying cerebrum imaging information gathered from 17 unique labs all over the planet. Datasets has underlying and pre-handled resting state fMRI with phenotypic portrayal. The rs-fMRI are cut time and movement amended and standardized. For this assignment, all rs-fMRI are preprocessing pipeline and band pass sifted (0.01-0.1Hz).

In spite [4] we have 1112 datasets; one cannot use them all to train any model. Recent literatures point out thatthe configuration of instruments used in different labs are subjected to be different. Moreover, the time-series length is different for different sources. So as to remove this difficulty, I have used the datasets from NYU only for this task owing to the fact that it has the largest number of fMRI samples. For this task, I have used Automated Anatomical Labelling (AAL) brain parcellation which contains 116 regions of interest (ROIs). After obtaining the feature vectors and class labels, state-of-the-art techniques like SVM, KNN, etc. can be used for classification.For each classifier, its hyperparameters are tuned via a grid search using 5-fold cross-validation.

For SVM, various kernel functions can be used and it also has tunable hyperparameters C and γ. For this task, I used linear, quadratic, radial and sigmoid basis functions as kernels and varied C from 2^{-5} to 2^{15}, and γ from 2^{-15} to 2^{3}. By using linear basis function as kernel, the accuracy observed is same for all the values of hyperparameters.The observed accuracy was 59.92%. For radial basis function is quite high for lower values of γ and higher valuesof C. The best accuracy was observed to be 67.97% for $\gamma = 2^{-13}$ and $C \geq 2^{1}$. The model with sigmoid kernel function gives best accuracy of 61.04% for $\gamma = 2^{-13}$ and $C = 2^{1}$. For quadratic kernel, the best accuracy is 63.41%for values of C and γ satisfying,

$$log2\,(\gamma) + \frac{1}{2}\,log2\,(C) + 25^2 = 0$$

Thus, the best kernel to be used in SVM for this task is radial with corresponding hyperparameters. In KNN, the value of parameter k and distance function defined as Minkowski distance with parameter p. p=1 corresponds to Manhattan distance the best accuracy is 0.6455 for k=29 and p=2 corresponds to Euclidean distance the best accuracy score is 0.6399 for k=23. I varied the parameter k from 1 to 50 for both values of p.

From the previous section, it is observed that a standalone finely-tuned SVM classifier performs the best for the given task. In this section, various neural net architectures are used to perform the same task. It can be concluded that a standalone finely-tuned SVM model is best fitted for this task and can provide accuracy as highas 68%.

III. Methodology

A. The Dataset

The dataset [2] for the train and test dataset of the model contains the Autism dataset with all age group from children, adults and adolescents. The target feature is the class_asd which determines whether the individual is autistic or not. Then this feature set is used to build model using SVM, Naïve Bayes algorithms for classification.

TABLE 1.DESCRIPTION OF DATASET

SerialNo.	Attribute	Brief explanation	Data type
1to10	a1_scr to a10_scr	Answers for the asked questions	binary
11.	age_prs	Person’s age	integer
12.	gender_prs	Person’s gender	String (M, F)
13.	indvl_ethnicity	Individual belong to which ethnicitygroup	String
14.	jaundice	Jaundice of person at birth	String (Y, N)
15.	autsm_score	Assuming family member or family case history of the individual wasdetermined to have the disorder	String (N, Y)
16.	contry_res	Residence country	String
17.	app_befr_usd	Person used screening app before	String (Y, N)
18.	result_scr	Screening test score out of 10.	integer
19.	age_desc	Description of age range	String
20.	relation	Screening question answered bywhom.	String
21.	class_asd	Individual diagnosed as ASD fromapp	String (Y, N)

B. Data Cleaning

The properties like gender, ethnicity, jaundice, and so on, are switched over completely to twofold qualities (0, 1) so to facilitate the demonstrating system. The age attribute null is imputed with mean. Hence data cleaning is done by using attribute mean, attribute construction and label encoding method using label encoder isused for categorical values.

C. Exploratory Data Analysis

From exploratory data analysis the following summary is drawn that majority of people who took the survey were from USA then from UK, India and so on. Then majority of people who took the test are White-European ethnicity followed by Asian, Middle Eastern, black and so on. Person born without jaundice has high chance of being ASD positive and ASD is more common among male than female irrespective of being born with jaundice.Most of the white European people have ASD, followed by black, Asian and so on.

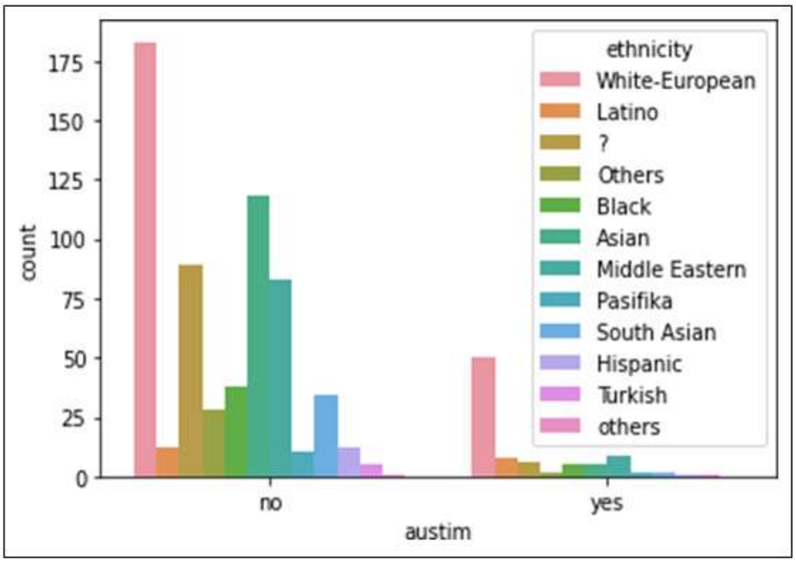

Fig. 1 Data analysis for ethnicity.

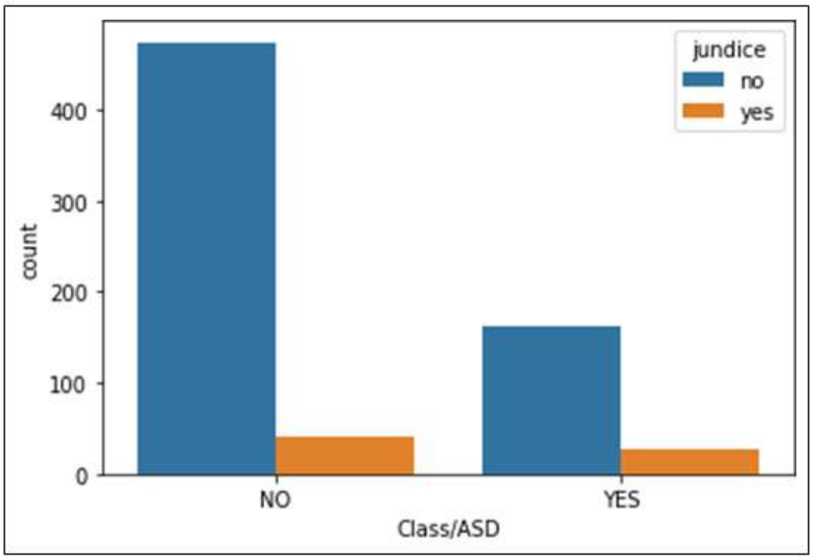

Fig. 2 Data analysis for jaundice.

***D.* Classification**

The dataset of autism is then classified using SVM, KNN and Random Forest algorithms. Before classification the dataset is split for testing and training of the model in fashion of 80:20 respectively. Then the algorithms weretuned and the model is fitted. The fitted model is test against the test dataset for accuracy. The accuracy is calculated by comparing the predicted value and original test value.

***E.* Data Analysis**

A score for the Question and the Answering of the screening test, if it results more than 7 then the classificationwill be positive else negative can be done at user interface. There is no likely connection between the individual brought into the world with jaundice and analyzed as Autistic and furthermore there is no plausible connection between the individual who was determined to have the Autistic and individual himself lying with Autistic. By using further more dataset the relation of the above may be improved.

IV. Performance Metrics

A. True Positive (TP)
Positive class anticipated by model accurately.

B. True Negative (TN)
The negative class anticipated by model mistakenly.

C. False Positive (FP)
Here model predicts the positive class inaccurately.

D. False Negative (FN)
The negative class predicted by model incorrectly.

E. F1 Score
The F1 Score consolidates accuracy and review of a classifier into a solitary measurement with thethought of harmonic mean.

F. Accuracy Score (AS)
The accuracy score defined as performance measure. It is a ratio of correctly predicted observation tothe total observation.

$$AS = \frac{TruePositive + TrueNegative}{TruePositive + TrueNegative + FalseNegative + FalsePositive}$$

G. ROC curve.
An ROC curve which shows the performance of a classification model at all classification thresholds.
It plots two parameters:

i. True-Positive Rate (TPR)
It is a measure of how many positive cases in a set of data are correctly classified as such:TPR = $\frac{TP}{TP+FN}$

ii. False Positive Rate (FPR)
The absolute of negative cases wrongly detailed as certain cases to the all outnumber of negative cases is portrayed in this formula:

$$FPR = \frac{FP}{FP+TN}$$

H. Area Under Curve
AUC estimates the whole 2-D region beneath the whole ROC bend. AUC gives a significant proportion of execution across all conceivable order edges.

V. Result and Evaluation

The five classifying model used are: SVM, Naïve Bayes, Logistic Regression and KNN classifiers. Naïve Bayes classifier model is tested with accuracy of 0.95 is also an optimal for the classification. The F1 Score of this model is 93%. The AUC and ROC is also good for this model.

The model developed with the SVM classifier have the test dataset accuracy of 0.98 compared to Naïve Bayes which has the test dataset accuracy of 0.93. The F1 Score of this model is 95% which is also excellent compared to previous model. Thus, the model has optimal ROC curve with a high AUC of 98% compared to Naïve Bayes classifier.

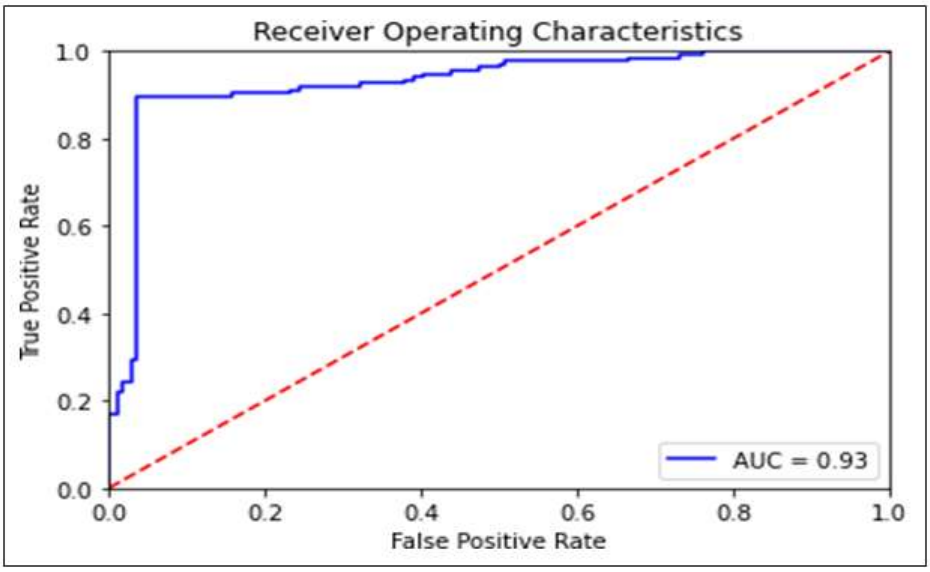

Fig. 3 ROC graph for Naïve Bayes for autism screening dataset.

The model of Logistic Regression has the F1 Score of 70% which the least and test data accuracy of the 0.71which also not a good accuracy. The ROC curve with low AUC of 0.73 which not pose the optimal model. The KNN model is out fitted model for this dataset with test dataset accuracy 0.68 and have the F1 Score of 66% which is too least to be the optimal model. The ROC curve has a low AUC which is far to be the optimal model.

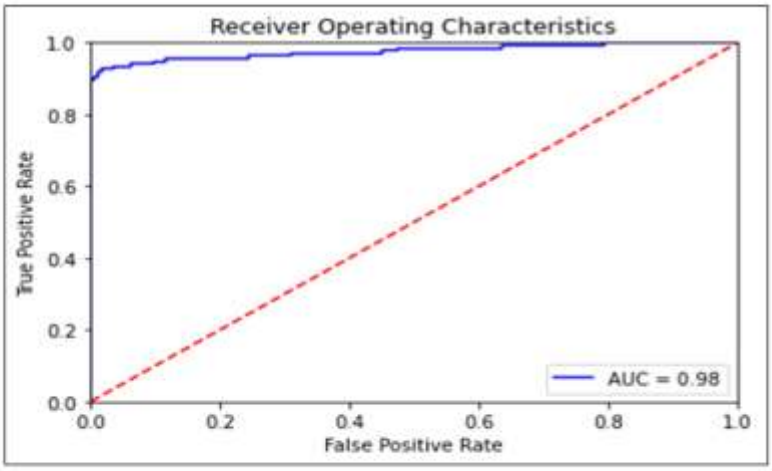

Fig. 4 ROC graph for SVM for autism screening dataset.

The Naïve Bayes gives the high specificity but fails to give the actual result that is required. The Logistic Regression and KNN do not perform that much excellent compared to the SVM and Naïve Bayes models.

TABLE 2.COMPARISON OF ALGORITHMS FOR AUTISM SPECTRUM DISORDER DATASET.

Algorithms	**F1_score**	**Accuracy**	**AUC**
SVM	0.95	0.98	0.98
Naïve Bayes	0.92	0.95	0.92
Logistic Regression	0.70	0.71	0.73
KNN	0.66	0.68	0.67

The SVM perform well compared to Naïve Bayes, Logistic Regression and KNN with respect to all theperformance metrics of the models.
SVM perform well by establishing the best F1 Score, Accuracy Score, AUC, specificity and sensitivity rates.

VI. Conclusion

The person with Autism is diagnosed as Autistic by many attributes. The early detection of disorder leads toearly treatment and recovery of disorder. The diagnose of this disorder is based on many disorders hence require larger dataset size, so that an inference can be drawn through it that the model was built with algorithms like SVM, KNN, Logistic Regression etc. Through performance metrics these models F1 score, Accuracy Score, AUC and ROC graphs are plotted. Thus indicates the coverage of the dataset for the performance of the models.

Thus, by using the larger dataset the result with highest accuracy can be drawn. But certain conclusions like whether the jaundice attribute or any disease of family case history of person may vary the accuracy of the resultcan be drawn by further evaluation of the model with more complex dataset.

References

[1] Anibal Sólon Heinsfeld, Alexandre Rosa Franco, R. Cameron Craddock, Augusto Buchweitz and Felipe Meneguzzi, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset," NeuroImage: Clinical, vol. 17, pp. 16-23, 2018.

[2] Astha Baranwal, Vanitha M, " Autistic Spectrum Disorder Screening: Prediction with Machine Learning Models", 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE).

[3] Suman Raj, Sarfaraz Masood, "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques", International Conference on Computational Intelligence and Data Science (ICCIDS 2019), Procedia Computer Science 167 (2020) 994–1004.

[4] Utkarsh Patel, "Identification of Autism from Functional Connectivity Analysis using Support Vector Machine and K Nearest Neighbour Classifier", IIT Kharagpur.

[5] Mariam M. Hassan and Hoda M. O. Mokhtar, "Investigating autism etiology and heterogeneity by decision tree algorithm," Informatics in Medicine Unlocked, vol. 16, 100215, 2019.

[6] Fadi Thabtah, Firuz Kamalov and Khairan Rajab, "A new computational intelligence approach to detect autistic features for autism screening," International Journal of Medical Informatics, vol. 117, pp. 112-124, sep 2018.

[7] Sanjeevani Choudhery, Chuan Huang and Daifeng Wang, "T253. A Machine Learning Approach to Predict the Changes of Brain Functional Connectivity in Autism Spectrum Disorder From the Gene Expression Data," Biological Psychiatry, vol. 83, issue 9, supplement, pp. S227-S228, 1 May 2018.

[8] Alex M. Pagnozzi, Eugenia Conti, Sara Calderoni, Jurgen Fripp and Stephen E. Rose, "A systematic review of structural MRI biomarkers in autism spectrum disorder: A machine learning perspective," International Journal of Developmental Neuroscience, vol. 71, pp. 68-82, dec 2018.

[9] Elizabeth Stevens, Dennis R.Dixon, Marlena N. Novack, Doreen Granpeesheh, Tristram Smith, Erik Linstead, "Identification and analysis of behavioral phenotypes in autism spectrum disorder via unsupervisedmachine learning," International Journal of Medical Informatics, vol. 129, pp. 29-36, sep 2019.

[10] Haishuai Wang, Li LiLianhua Chi, Ziping Zhao, "Autism Screening Using Deep Embedding Representation,"International Conference on Computational Science, Lecture Notes in Computer Science, vol 11537, pp. 160-173, jun 2019. Muhammad Nazrul Islam, Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan, "A Machine Learning Approach to Predict Autism Spectrum Disorder," International Conference on Electrical, Computer and Commmunication Engineering, feb 2019.

[11] Kayleigh K. Hyde, Marlena N. Novack, Nicholas LaHaye, Chelsea Parlett-Pelleriti, Raymond Anden, DennisR. Dixon, Erik Linstead, "Applications of Supervised Machine Learning in Autism Spectrum Disorder Research: a Review," Review Journal of Autism and Developmental Disorders, vol. 6, pp. 128–146, feb 2019.

[12] Fadi Thabtah, "Autism Spectrum Disorder Screening: Machine Learning Adaptation and DSM-5 Fulfillment," Proceedings of the 1st International Conference on Medical and Health, pp.1-6, may 2017.