Optimized Machine Learning Classification Approaches (Omlc-Asd) For Early Detection of Autism Spectrum Disorder

Asif Mohamed H B[1]*, Dr.Md. Sameeruddin Khan[2], Dr. Mohan K G[3], Dr.Parashuram Baraki[4]

[1*]Research Scholar, School of CSE, Presidency University, Bengaluru, Karnataka, India,

[2] Professor & Dean, School of CSE, Presidency University, Bengaluru, Karnataka, India,

[3] Professor, Department of CSE, GITAM University, Bengaluru, Karnataka, India.

[4] Professor, Department of CSE , Smt.Kamala and Sri Venkappa M Agadi College of Engineering & Technology, Lakshmeshwar Dist:Gadag

*Corresponding Author: Asif Mohamed H B

*Research Scholar, School of CSE, Presidency University, Bengaluru, Karnataka, India,

Abstract

Autism spectrum disorder is a neurological condition that impairs social interaction and communication because it alters how people perceive and relate to one another. ASD screening refers to the process of identifying suspected autistic features in individuals through exams administered by a physician, carer, or parent. This study evaluates the performance of various machine learning algorithms and pre-processing strategies to categorise medical datasets used to identify early signs of autism in children and adults. Many earlier studies in this field use sophisticated pre-processing and machine learning methods to perform accurate categorization. Yet, this experiment shows that a range of classifier algorithms, such as logistic regression, KNN, and Random Forest, can produce results that are comparable to the state-of-the-art.

1. Introduction

Many industries desire process automation to increase productivity and cut costs. One of the most significant industries that could gain from faster processing is healthcare. It is possible to diagnose human health problems rapidly and precisely. One of the many illnesses that are now difficult to identify is autism. Due to the lengthy process, it might take up to 6 months to definitively diagnose a child with autism. A child must consult numerous specialists, beginning with developmental paediatricians, neurologists, psychiatrists, or psychologists, in order to receive an autism diagnosis. With the current, accepted methods, concluding an autism diagnosis requires some time. Hence, using machine learning techniques could significantly speed up the process. It is well-known that children with autism require early intervention in order to improve. Accelerating the diagnostic process is unquestionably more important in cases involving autism. The lengthy and challenging processes of diagnosis and therapy can be predicted and accelerated with the help of big data and machine learning. The enormous amount of health and medical data that is currently available can be used for predictive modelling and analysis using a machine learning system. In this essay, a comparison of several machine learning models and methods will be evaluated. Pre-processed data is used to predict which test respondents would be labelled as autistic based on various categories. Algorithms for categorisation can be utilised in a variety of ways. Each classifier employs a distinct method for data collection, data filtering, and feature extraction before using these methods to generate training data for the model. In this study, the efficacy of several machine learning algorithms for determining the success of a given course of treatment or predicting how those therapies would turn out is examined.

Patients and controls are obtained (dataset) using a screening technique like the A1 Score-A10 Score based on Figure 1. A classification method is used to create a classifier—a predictive model—after the dataset has been processed. The classifier then successfully categorises the target class using fresh test instances (subjects) that were never presented to the model during training (ASD traits or not). We believe that using machine learning to evaluate applicants for the autism application rather than traditional scoring algorithms can enhance the screening's performance in terms of projected accuracy, sensitivity, and specificity. Most crucially, because models are created using historical data and controls, the conclusion will be completely objective as opposed to subjective. Doctors and other medical professionals may be able to diagnose patients more precisely as a result of this.

2. Literature Review

In order to diagnose psychological problems, quantitative tests are typically used less frequently than monitoring the symptoms or characteristics that exist in a human. Hence, the differential diagnosis and grading of a problem, which is more challenging

doi.org/10.31838/hiv23.03.70 Received: 17.10.22, Revised: 20.11.22, Accepted: 13.12.22.

than diagnosing a disease, require a high level of clinical competence. Here is a discussion of how various classification approaches are used in relation to the work. There are a number of ensemble learning models, including boosting and bagging. The user must provide the sample size that will be utilised to train the local classifier in the Bagging Learning model. This sample is typically chosen at random from the training dataset with the final user-specified size. For instance, if the sample size is set to 50%, then 50% of the training dataset will be created along with additional instances that are randomly generated. The user should also specify how many bags (classifiers). This value will frequently grow dynamically until the developed classifiers stop improving. The cost of processing time and computational resources, however, can increase as the number of bags increases.

Crippa, A., Salvatore, C., Perego, P., Forti, S., Nobile, M., Molteni, M., and Castiglioni (2015) proposed an approach utilising machine learning to identify children with autism and related motor difficulties. Researchers have started a proof-of-concept trial to see if a straightforward upper-limb movement can help in the diagnosis of low-functioning children with autism spectrum disorder (ASD) who are between the ages of 2-4. As a result, researchers in the field are attempting to find novel, more effective solutions to the multiple pass problem. In their "Classification Based on Many Class-Association Rules (CMAR)" algorithm, the authors of Pie (2001) established an effective method for frequent rule mining by creating a class distributed-associated FP-tree. Together with further assessments based on correlation rates, confidence, and database coverage, the authors chose a CR-tree to preserve the organisation of mined association rules and to improve the storing and retrieval procedures. As a result, the classification model will be better able to predict new class labels. In comparison to the C4.5 and CBA models, the CMAR model gave more accurate results. The clinical methods used today to diagnose autism demand lengthy waiting times. Researchers are creating new early screening methods utilising machine learning to lessen this and make it easier to detect autism in kids and toddlers. Some scholars have used acoustic analysis of vocal production to identify autistic features. However, other researchers are analysing non-verbal or behavioural data from video recordings.

Bekerom used a variety of machine learning (ML) approaches, including naive Bayes, SVM, and random forest algorithm, to analyse the data and identify characteristics of ASD in children, such as obesity, developmental delay, and decreased physical activity. The functional tree and the AD Tree worked well with high sensitivity, specificity, and accuracy, according to Wallet alwork's study on detecting autism using a quick screening test and validation. Heinsfeld used a large brain imaging dataset from the Autism Imaging Data Exchange along with deep learning techniques and neural networks to identify ASD patients (ABIDE I). His classification accuracy scored between 66% and 71%, on average, with a score of 70%. The SVM classifier had a mean accuracy of 65% compared to the Random Forest classifier's 63%. An ASD screening frequently includes a series of behavioural questions on social interaction, repetitive behaviour, communication, etc. in attempt to identify any autistic symptoms in a person. The child's doctor, carer, parent, or teachers commonly answer these questions on their behalf during the screening procedure. The young person is then given a final score to determine whether he or she may be exhibiting autistic symptoms and whether further medical testing is necessary.

2. Proposed Methodology

Early detection can greatly lower the medical expenses linked to autism spectrum disorder (ASD), a neurodevelopmental illness. But getting an ASD diagnosis requires a long wait and expensive treatments. The rising incidence of ASD cases globally and the financial impact of autism highlight the urgent need for the creation of simple screening techniques. Soon, medical professionals and patients will have access to a quick and easy ASD test that will help them decide whether to pursue an official clinical diagnosis. Datasets on behavioural characteristics are required due to the substantial rise in ASD prevalence worldwide. Due to the dearth of such datasets, it is challenging to perform in-depth research to enhance the efficacy, sensitivity, specificity, and predictive accuracy of the ASD screening technique. There are currently very few datasets available that are connected to clinical or screening-related aspects of autism and most of these datasets are genetic in nature.

2.1 algorithm – omlc-asd

A variety of categorical, continuous, and binary type attributes can be found in the input
data. Toddlers and adults make up the two datasets. a group of classifier models that were built using this dataset as the basis.

Step 1: Import Data Set – autism spectrum disorder
Step 2: Data pre-processing.
Step 3: A variety of examples and attributes are used to construct graph plots.
Step 4: A new dataset is built using only the relevant attributes from the adult and toddler datasets after feature engineering.
Step 5: Using a variety of classifiers and these pertinent attributes, a model is constructed.
Step 6: Efficiency is evaluated and the F1 Score, Recall, and Precision are computed.
Step 7: Plots of graphs are made using the F1 score in order to choose the optimum classification algorithm.

2.2 data analysis

A1 int64
A2 int64
A3 int64

A4 int64
A5 int64
A6 int64
A7 int64
A8 int64
A9 int64
A10 int64
Age_Mons int64
Qchat-10-Score int64
Sex object
Ethnicity object
Jaundice object
Family_mem_with_ASD object
Class/ASD Traits object
dtype: object
Index(['Case_No', 'A1', 'A2', 'A3', 'A4', 'A5', 'A6', 'A7', 'A8', 'A9', 'A10', 'Age_Mons', 'Qchat-10-Score', 'Sex', 'Ethnicity', 'Jaundice', 'Family_mem_with_ASD', 'Who completed the test', 'Class/ASD Traits '],
dtype='object')

Table 1. ASD Statistics

Case_No	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	Age_Mons	Qchat-10-Score	
count	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000	1054.000000
mean	527.500000	0.563567	0.448767	0.401328	0.512334	0.524668	0.576850	0.649905	0.459203	0.489564	0.586338	27.867173	5.212524
std	304.407895	0.496178	0.497604	0.490400	0.500085	0.499628	0.494293	0.477226	0.498569	0.500128	0.492723	7.980354	2.907304
min	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	12.000000	0.000000
25%	264.250000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	23.000000	3.000000
50%	527.500000	1.000000	0.000000	0.000000	1.000000	1.000000	1.000000	1.000000	0.000000	0.000000	1.000000	30.000000	5.000000
75%	790.750000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	36.000000	8.000000
max	1054.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	36.000000	10.000000

Figure 1. Heat Map – ASD

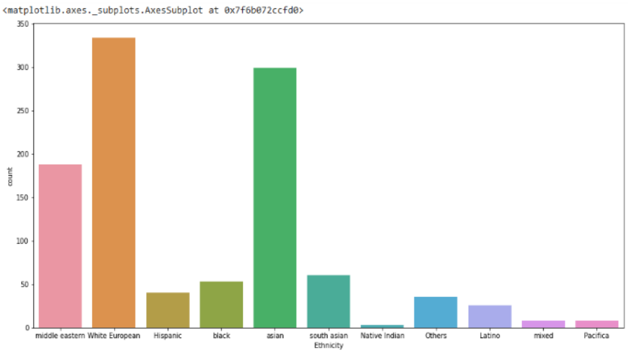

Figure 2 – ASD Statistics

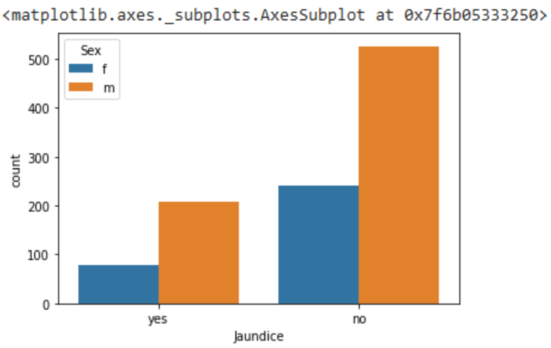

Figure 3 . Jaundice Count - ASD

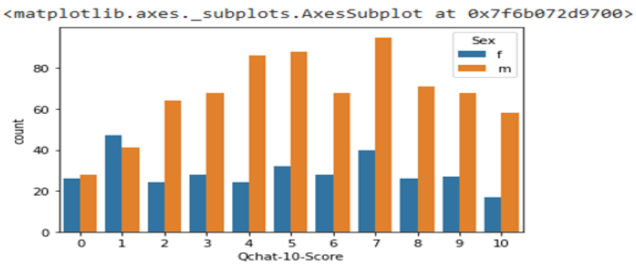

Figure 4. QChat-10 Score - ASD

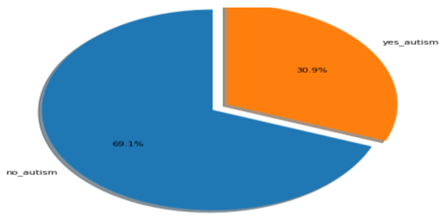

Figure 5 – Yes_Autism / No_Autism – ASD

3. Performance Measure

A thorough study of the experimental data was done in order to compare the three well-known AC methods—F-measure, recall, and precision—as statistical metrics. Several minimal support and minimum confidence levels were utilized to assess the efficacy of the chosen algorithms for the dataset of autistic adults and toddlers. The evaluation procedure uses three commonly used statistical indicators to assess how each strategy performed overall on the Autism Adult UCI Dataset (F1, Precision, and Recall).

3.1 Comparisons of Algorithms

Table 2 ML Model Comparison

0	LogisticRegression	1.0
1	LinearDiscriminantAnalysis	0.96
2	KneighborsClassifier	0.91
3	DecisionTree Classifier	0.91
4	GaussianNB	0.94
5	SVC_beforegrid	0.78
6	RandomForest_beforegrid	0.64
7	XGBClassifier	0.99
8	GradientBoosting	0.64
9	AdaBoosting	0.49
10	SVC_aftergrid	1.0
11	RandomForest_aftergrid	0.96
12	Neuralnetwork_SKLearn Accuracy	0.99
13	Neuralnetwork_Keras	0.95

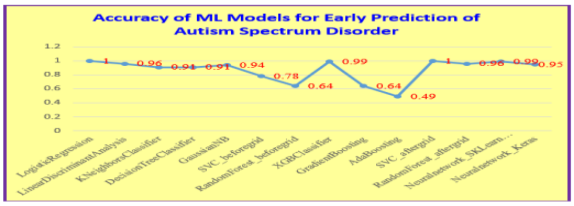

Figure 6. Accuracy

The accuracy score for several classifiers is displayed. Overall, the performance of Logistic Regression and SVC- classifier is 1 which is better Accuracy when compared to other Machine Learning Approaches.

4. Conclusion

This work focused on building classification models using machine learning techniques such the Random Forest, Logistic Regression, and K NN , DT ,NB, SVC , Adaboost, NN Classifiers. The accuracy score for several classifiers is displayed. Overall, the performance of Logistic Regression and SVC- classifier is 1 which is better Accuracy when compared to other Machine Learning Approaches.

References

Constantino, John N., Patricia D. Lavesser, Y. I. Zhang, Anna M. Abbacchi, Teddi Gray, and Richard D. Todd. (2007) "Rapid quantitative assessment of autistic social impairment by classroom teachers." Journal of the American Academy of Child & Adolescent Psychiatry 46(12): 1668-1676.

Daniel Bone, Chi-Chun Lee, Matthew P. Black, Marian E. Williams, Sungbok Lee, Pat Levitt, and Shrikanth Narayanan .(2014). "The psychologist as an interlocutor in autism spectrum disorder assessment: Insights from a study of spontaneous prosody." Journal of Speech, Language, and Hearing Research, 57(4): pp.1162-1177.

Daniel Bone, Matthew S. Goodwin, Matthew P. Black, Chi-Chun Lee, Kartik Audhkhasi, and Shrikanth Narayanan. (2015) "Applying machine learning to facilitate autism diagnostics: pitfalls and promises." Journal of autism and developmental disorders 45(5): 1121-1136.

Dennis P. Wall, Rebecca Dally, Rhiannon Luyster, Jae-Yoon Jung, and Todd F. DeLuca. (2012) "Use of artificial intelligence to shorten the behavioral diagnosis of autism." PloS one, 7(8): e43855.

Dennis Paul Wall, J. Kosmicki, T. F. Deluca, E. Harstad, and Vincent Alfred Fusaro. (2012) "Use of machine learning to shorten observationbased screening and diagnosis of autism." Translational psychiatry, 2(4): e100.

Dr.R.Chinnaiyan, Sabarmathi.G (2017)," Investigations on Big Data Features , Research Challenges and Applications", IEEE International Conference on Intelligent Computing and Control Systems, ICICCS 2017, 782 – 786

Dr.S.Nirmala,R.Chinnaiyan.et.al. (2021). Blockchain based Secured Framework for Road Traffic Management using Fog Computing . International Journal of Computational Intelligence in Control 13(2)

Fadi Thabtah. (2017) "ASD Tests. A mobile app for ASD screening." www.asdtests.com [accessed December 20th, 2017]. [11] Baihua Li, Arjun Sharma, James Meng, Senthil Purushwalkam, and Emma Gowen. (2017) "Applying machine learning to identify autistic adults using imitation: An exploratory study." PloS one, 12(8): e0182652.

Fadi Thabtah. (2017). "Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment." In Proceedings of the 1st International Conference on Medical and Health Informatics, pp. 1-6. ACM.

G. Sabarmathi and Dr.R. Chinnaiyan, "Reliable Machine Learning Approach to Predict Patient Satisfaction for Optimal Decision Making and Quality Health Care," 2019 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2019, pp. 1489-1493.

G. Sabarmathi and Dr.R. Chinnaiyan, "Big Data Analytics Framework for Opinion Mining of Patient Health Care Experience" International Conference on Computing Methodologies and Communication (ICCMC 2020), IEEE Xplore Digital Library

G. Sabarmathi and R. Chinnaiyan, "Reliable feature selection model for evaluating patient home health care services opinion mining systems," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-4, doi: 10.1109/ICAECA52838.2021.9675485.

G.Sabarmathi , Dr.R.Chinnaiyan (2018), "Envisagation and Analysis of Mosquito Borne

Fevers – A Health Monitoring System by Envisagative Computing using Big Data Analytics" in ICCBI 2018 – Springer on 19.12.2018 to 20.12.2018 (Recommended for Scopus Indexed Publication IEEE Xplore digital library)

G.Sabarmathi , Dr.R.Chinnaiyan, Reliable Data Mining Tasks and Techniques for Industrial Applications, IAETSD JOURNAL FOR ADVANCED RESEARCH IN APPLIED SCIENCES, VOLUME 4, ISSUE 7, DEC/2017,PP- 138-142, ISSN NO: 2394-8442

G.Sabarmathi, Dr.R.Chinnaiyan "Mining Patient Health Care Service Opinions for Hospital Recommendations" International Journal of Engineering Trends and Technology 69.9(2021):161-167

Hari Pranav A; M. Latha; Ashwin. M. S; R. Chinnaiyan, "BlockchainAs a Service (BaaS) Framework for Government Funded Projects e-Tendering Process Administration and Quality Assurance using Smart Contracts," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-4, doi: 10.1109 /ICCCI5 0826. 2021.9402348.

Hari Pranav A;M. Senthilmurugan; Pradyumna Rahul K;R. Chinnaiyan , "IoT and Machine Learning based Peer to Peer Platform for Crop Growth and Disease Monitoring System using Blockchain," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-5, doi: 10.1109/ICCCI50826.2021.9402435.

M. Latha, M. Senthilmurugan and R. Chinnaiyan, "Brain Tumor Detection and Classification using Convolution Neural Network Models," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-5, doi: 10.1109/ICAECA 52838.2021.9675585.

M. Senthilmurugan, M. Latha and R. Chinnaiyan, "Analysis and Prediction of Tuberculosis using Machine Learning Classifiers," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), 2021, pp. 1-4, doi: 10.1109/ICA ECA52838 .2021.9675482.

Manoj Challa , Dr.R.Chinnaiyan (2019) "Optimized Machine Learning Approach for the Prediction Of Diabetes-Mellitus (Recommended for Scopus Indexed Publication in Springer - Advances in Intelligent Systems and Computing Series)

Preetika B,; M. Latha; M. Senthilmurugan; R. Chinnaiyan, , "MRI Image based Brain Tumour Segmentation using Machine Learning Classifiers," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-9, doi: 10.1109/ICCCI50826.2021.9402508

R. Chinnaiyan and D. Stalin Alex, "Early Analysis and Prediction of Fetal Abnormalities Using Machine Learning Classifiers," 2021 2nd International Conference on Smart Electronics and Communication(ICOSEC),2021,pp.1764-1767,doi: 10.1109/ICOSEC51865.2021.9591828. https://ieeexplore.ieee.org/abstract/document/9591828)

R. Chinnaiyan and S. Alex, "Machine Learning Approaches for Early Diagnosis and Prediction of Fetal Abnormalities," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, pp. 1-3, doi: 10.1109/ICCCI50826.2021.9402317. https://ieeexplore.ieee.org/abstract/document/9402317)

R.Chinnaiyan,Dr.StalinAlex (2021). Optimized Machine Learning Classifier for Early Prediction of Fetal Abnor-malities . International Journal of Computational Intelligence in Control 13(2)

Vaishali, R., and R. Sasikala. "A machine learning based approach to classify Autism with optimum behaviour sets. (2018)" International Journal of Engineering & Technology 7(4): 18.