International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Comparative Analysis of Autism Spectrum Disorder Using Machine Learning Techniques

U. Chaitanya[1], P. Ashok [2], Samreen[3], S. G. V. Rathnam[4]

[1]Assistant Professor, Mahatma Gandhi Institute of Technology, Hyderabad-500075, Telangana, India
[2,3,4] UG Student, Mahatma Gandhi Institute of Technology, Hyderabad-500075, Telangana, India

A B S T R A C T

A neuro-disease known as autism spectrum disorder or in short can be denoted as "ASD" hampers a person's qualities and abilities to interact and communicate with others throughout their life. Autism can be detected in a person at any point in his life. Generally based on the study on ASD theory the illness persists throughout our lifetime. Attempts to employ Naive Bayes, Support Vector Machine, Logistic Regression, KNN, and MLP for predicting and analysing autism spectrum disorder difficulties in adults are made in this work in light of the growing usage of machine learning techniques in the research aspects of medical diagnosis. On a non-clinical ASD dataset that is available to the public, the proposed strategies are assessed. There are a total of 701 occurrences and 21 attributes included in the dataset of the adults. The results of using several machine learning methods and handling values that are missing strongly indicate that the Naive Bayes algorithm can be used on the data to achieve efficiency.

Keywords: SVM, Naïve Bayes, Logistic Regression, KNN, MLP.

1. Introduction

Autism spectrum disorder (ASD) is being increasingly detected in people throughout the world. It kills the progress of the brain from the infant stage itself. Grown-up adults who are diagnosed with this disorder behave like a child and are restricted to their own world. Detecting Autism spectrum disorder can be helped by analyzing its symptoms being common among other related cases. This disorder develops due to severely underdeveloped parts of a person's neurological system. A person diagnosed with autism cannot communicate and engage with normal people. They are difficult to approach and are allergic to various kinds of things. People who are diagnosed with autism exhibit repetitive behaviors which is uncommon among normal people. It cannot be completely cured but understanding their thoughts, treatments, and therapy can be applied to them thereby helping them lead a meaningful and healthy life. It is important to recognize that every person with ASD is unique, with their own strengths, challenges, and capabilities. Changes in routine can be particularly distressing for some individuals with ASD, leading to increased anxiety and emotional difficulties.Autism can be identified at any stage and upon identification better treatment is required for them to sustain a normal life. Diagnosing a patient with ASD requires in depth knowledge about the behavior of ASD and its symptoms.

2. Background

Vaishali, et al. proposed the need for creating precise prediction models to identify the risk of autism quicker than with conventional diagnostic techniques is highlighted by machine learning-based behavioural analytics. When using the best feature subsets, the results of our technique provided a mean accuracy in a spectrum with a certain higher range of equivalent with a mean accuracy obtained from the all-out information we have provided in this model. This proves the complete hypothesis [2]. Thabtah, et al. proposed that according to their approach, physicians will be better able to diagnose and predict ASD with accuracy. Studies on the usage of machine learning for autism discovery and cure, however, are hindered by theoretical, implementation, and data problems, including the way codes for diagnosis are utilized, the type of selection of features used, the assessment criteria selected, and sort disparities in particulars. The creation of a new machine learning-based approach for diagnosing ASDs is a more serious assertion made in recent research [4]. Mythili, et al. proposed that the examination of ASD is of classification Approaches. The goal was to figure out the autism trouble and the tiers of autism. In this SVM ,Neural Network, and Fuzzy strategies with WEKA tools are employed to investigate pupils' manners and colonial relations [6]. Constantino et al, proposed that Educators day-to-day watch kids in the realistic colonial factors of their classrooms and nourish surveys particularly significant intake in the analysis of considerable psychiatric signs. Their preciseness in demonstrating and estimating autistic symptomatology retains not yet been inducted. This examination corresponded with educators' corresponding to autistic signs those emanated from the elderly [1]. Kosmickil et al. proposed the assumed exploring strategy for the last collection of features for autistic evaluation. In this, the writers utilized a machine learning strategy to assess the clinical review of ASD. It was accomplished on the subset of conducts of kids established on the autism scope. It has four modules. It consists of, 8 distinct machine learning algorithms that were operated, applying the step-wise rearward component tag on score sheets from a diverse range of humans [7].

Dennis P.Wall et al. proposed that in their thesis they made 93 different questions in order to diagnose a certain set of children in the span of 2.5 hours. Upon further analysis, they found out that 7 specific questions were statistically enough to predict the outcome with an accuracy of 99%[3]. Baihua Li et al. proposed that autistic features such as repetitive behavior are the most significant attribute for detecting ASD. It was one of the first studies to be conducted where Machine Learning algorithms are applied to kinematic movements in detecting ASD[5].

3. Method

A. Data Preprocessing

Data Preprocessing is required to be performed for removing the noisy values and null values etc. Values while being entered from the real data involve a lot of mistakes and for thorough processing of the data into further steps, cleaning of the data is required which is removing inconsistent and incomplete values.

B. Loading Pre-processed data

After the cleaning of data such as missing values, and noisy values, we can load the preprocessed data.

C. Apply ML Algorithms

Machine Learning algorithms such as SVM,MLP , Logistic Regression ,Naïve Bayes and KNN are applied to the data, and models are created.

D. Implementation of machine learning

The train-test split process is used to assess the effectiveness of models. These models are required to be tested on the test dataset after being evaluated on the training data. The chosen features from the feature selection module are used to generate the metrics used to evaluate the performance, including accuracy, precision, recall, and f1-score.

E. Deployment

It entails developing a user interface where users can enter values for attributes about the conditions of the person. The input is processed by the system, which then determines whether or not a person has ASD. Users can view the outcome on a result page.

4. Algorithms

A. Support Vector Machine

The prior function of SVM in an ASD observing prototype stands to comprehend and construct a determination border that diverges the two categories founded on the segments yanked from the information. The SVM algorithm strives to locate an optimal hyperplane that maximally separates the two categories, inevitably facilitating reasonable sort. In the matter of ASD recognition, it can be equipped by employing a tagged dataset that possesses characteristics described as ASD characteristics and details regarding even if someone carries has been examined with ASD or not. The SVM standard comprehends for this data to design a category border that can distinguish between ASD and non-ASD someone grounded on the furnished segments.

B. Naïve Bayes.

Naive Bayes part in an ASD learning representative is to figure out the likelihood of someone's affinity to a distinct type (ASD or non-ASD) founded on the monitored segments. It supposes that the components are conditionally self-reliant given the class, learned as the "naive" assumption. The algorithm assesses the dependent likelihoods per element given the type tags. These likelihoods stand then fused utilizing Bayes' theorem to figure the posterior probabilities of per class shown the monitored elements.

C. Logistic Regression

Logistic regression holds definite edges in ASD recognition representatives. The algorithm furnishes decipherable outcomes by assessing the mark per component on the odds of ASD. Further, it can operate both continuous and categorical intake features. Logistic regression can be open to endure multiclass classification undertakings as well. Nevertheless, logistic regression deems an unbent association between the intake elements and the log probabilities of the prey class. In some instances, where it depicts this hypothesis may not retain, additional intricate representatives or trait engineering approaches may be vital.

D. K-Nearest Neighbour (KNN)

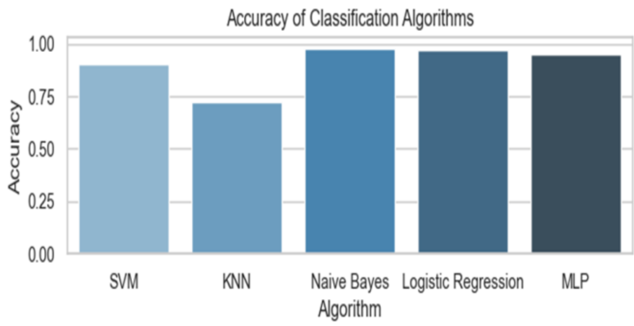
KNN holds indubitable forces in ASD detection standards. It can address problematic and non-linear affinities between elements and class tags, driving it critical when the judgment limit is not well-defined. Further, it is moderately uncomplicated to enforce and can have diverse sort of features.

E. MLP

Several layers of neurons that are artificially coupled and arranged in a feedforward fashion make up the MLP algorithm. A layer for input, a number of layers that are hidden, and an outputting layer are among the layers. Every single neuron in the neural network takes input signals, transforms them mathematically, and then sends the output to the neuronal cells in the layer below.

5. Comparison of algorithms with performance metrics

Upon training models applying different algorithms such as Support Vector Machine, Naïve Bayes, Logistic Regression, K-Nearest Neighbour, and Multi-Level Perceptron, we get different accuracy, precision, recall, and f1 score from different algorithms. Among the trained models, the Naïve Bayes algorithm gives the highest accuracy of them all. Below are the comparisons of different performance metrics vs algorithms.



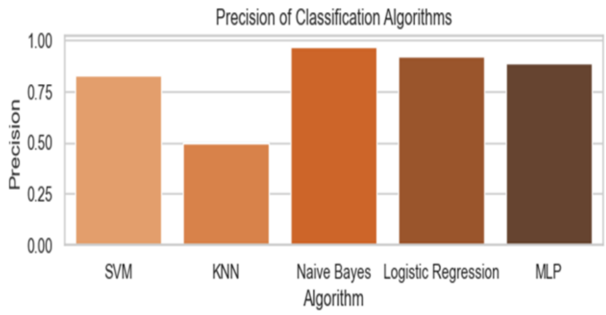
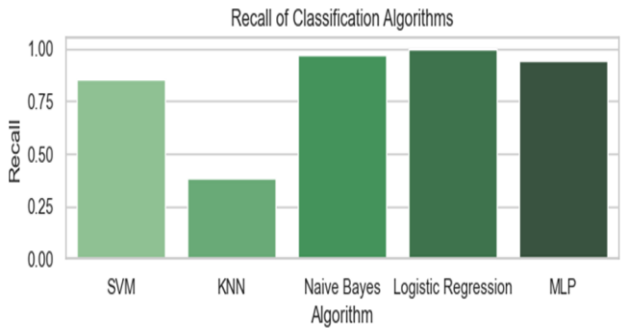


Fig.1.Comparison of Accuracies of different Machine Learning algorithms

Fig.2.Comparison of Precision of different Machine Learning algorithms

The above Fig.1. shows the comparison between SVM, KNN, Naïve Bayes, Logistic Regression, and MLP algorithms with their Accuracies

The above Fig.2. shows the comparison between SVM, KNN, Naïve Bayes, Logistic Regression, and MLP algorithms with their Precision.



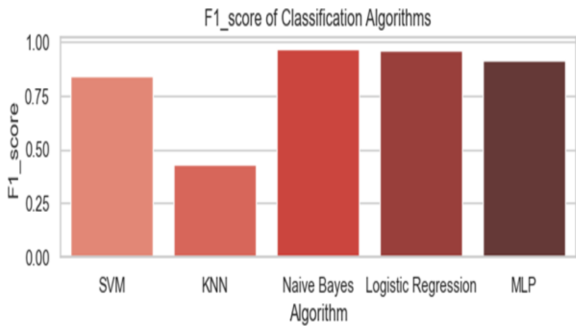


Fig.3.Comparison of Recall of different Machine Learning algorithms

Fig.4.Comparison of F1-Score of different Machine Learning algorithms

The above Fig.3. shows the comparison between SVM, KNN, Naïve Bayes, Logistic Regression, and MLP algorithms with their Recall Score.

The above Fig.4. shows the comparison between SVM, KNN, Naïve Bayes, Logistic Regression, and MLP algorithms with their F1-Score.

Table.1. Comparison of algorithms with performance metrics

	Accuracy	Precision	Recall	F1-Score
SVM	90.9836	82.85	85.29	84.05
Naïve Bayes	**98.36**	**97.05**	97.05	**97.05**
Logistic Regression	97.54	91.89	**100**	95.77
KNN	72.13	50	38.2	83.33
MLP	95.08	88.88	94.11	91.42

6. RESULTS

By comparing the performance metrics, we got the Naïve Bayes algorithm to be the most accurate algorithm for detecting autism spectrum disorder. Below is the whole comparison chart of all the performance metrics vs All the Machine learning Algorithms

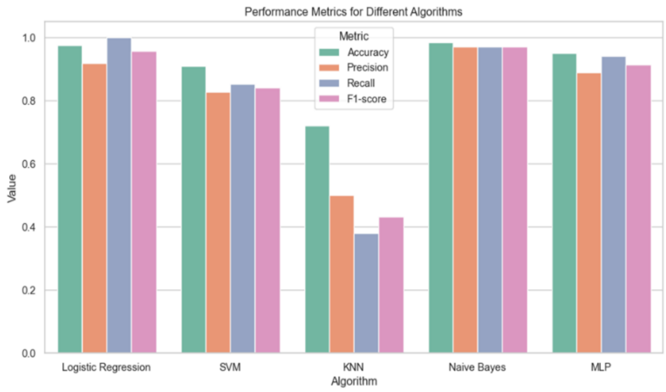


Fig.5. Comparison of all the four-performance metrics vs five Machine Learning Algorithms.

We compare all the four-performance metrics with the ML algorithms and Fig.5. shows the comparison of them. It is useful to identify the suitable algorithm for the dataset and which gives us results with high performance.

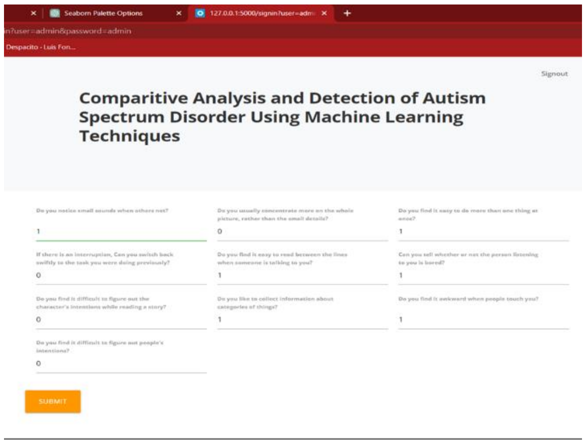


Fig 6. User-friendly interface showcasing an attribute form box

From the 21 Attributes we took initially to train the model, after further processing of data we took only 10 attributes which gave high similarity with the result. The selected 10 attributes amounted to the most impact on the result. The attributes are the specific questions asked to the user which can relate the user’s symptoms with autism spectrum disorder’s symptoms to predict whether the user is having ASD or not. The attributes specified in Fig.6. on the User Interface are:

i) Do you notice small sounds when others don’t?

ii) Do you usually concentrate more on the whole picture, rather than the small details?

iii) Do you find it easy to do more than one thing at once?

iv) If there is an interruption, can you switch back swiftly to the task you were doing previously?

v) Do you find it easy to read between the lines when someone is talking to you?

vi) Can you tell whether or not the person listening to you is bored?

vii) Do you find it difficult to figure out the character's intentions while reading a story?

viii) Do you like to collect information about categories of things?

ix) Do you find it awkward when people touch you?

x) Do you find it difficult to figure out people's intentions?

The Fig.6. shows that the input is taken from the user on this page and upon entering the input values it loads into next page to show the result

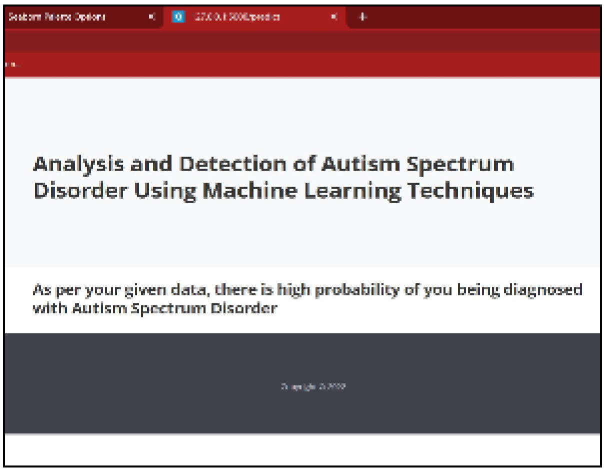


Fig.7. Web page interface showcasing machine learning model predictions

The Fig.7. displays the output to the user confirming the possibility for the user to be diagnosed with autism spectrum disorder.

7. Conclusion

Multiple machine learning techniques and deep learning approaches were used to try to diagnose autism spectrum disorder. Machine learning techniques, each of which is popular and has unique properties, have been employed for this data, and a variety of performance assessment criteria have been used to examine the effectiveness of the models constructed for ASD identification on non-clinical datasets. Naive Bayes provides the best accuracy of the five algorithms used, with a score of 98.36%. In the future, we can develop a more accurate model by combining two or more algorithms so that it gives more accuracy. Patterns can be found to near perfection upon intertwining two or more algorithms to create a model. It covers the loopholes of one algorithm with another.

References

[1] 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.

[2] Thabtah, Fadi. "Machine learning in autistic spectrum disorder behavioural research: A review and ways forward. (2018) " Informatics for Health and Social Care: 1-20.

[3] Wall, D. P., Dally, R., Luyster, R., Jung, Y., & DeLuca, T. F. (2012). Use of Artificial Intelligence to Shorten the Behavioral Diagnosis of Autism. *PLOS ONE*, *7*(8), e43855. https://doi.org/10.1371/journal.pone.0043855.

[4] Thabtah, Fadi, Firuz Kamalov, and Khairan Rajab. (2018) "A new computational intelligence approach to detect autistic features for autism screening." International Journal of medical informatics 117: 112-124.

[5] Li, B., Sharma, A., Meng, J., Purushwalkam, S., & Gowen, E. (2017). Applying machine learning to identify autistic adults using imitation: An exploratory study. *PLOS ONE*, *12*(8), e0182652. https://doi.org/10.1371/journal.pone.0182652.

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

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