Exploring the Potential of Machine Learning in Healthcare Accuracy Improvement

SINDHU VEERAMANI[1,*], S. M. RAMESH[2], B. GOMATHY[3]
[1]Dept. of Computer Science and Engineering,
Dr. NGP Institute of Technology,
Coimbatore, Tamil Nadu,
INDIA

[2]Dept. of Electronics and Communication Engineering,
KPR Institute of Engineering and Technology,
Coimbatore, Tamil Nadu,
INDIA

[3]Dept. of Computer Science and Business Systems,
Dr. NGP Institute of Technology,
Coimbatore, Tamil Nadu,
INDIA

Abstract: - Machine learning techniques have shown great potential in the medical industry, particularly in the field of neuroimaging and the identification of neurological illnesses such as Autism Spectrum Disorder (ASD). By utilizing machine learning algorithms, researchers aim to predict the type of disability and analyze the predicted variations using different types of predictive models. These predictive models can be trained on neuroimaging data to identify patterns and markers that are indicative of ASD. By analyzing these patterns, machine learning algorithms can help in accurately predicting the presence and type of ASD in individuals. This can be immensely valuable in early diagnosis and intervention, leading to better outcomes for individuals with ASD. Furthermore, the applications of machine learning in the healthcare industry extend beyond just prediction. Machine learning algorithms can also be used to analyze large amounts of medical data, identify trends, and assist in decision-making processes. This can help healthcare professionals in providing more accurate diagnoses, personalized treatment plans, and improved patient care. It is important to note that the success and accuracy of machine learning models in the healthcare industry depend on various factors, including the quality and quantity of data available, the choice of algorithms, and the expertise of the researchers. Ongoing research and advancements in machine learning techniques hold great promise for improving the accuracy and effectiveness of medical diagnoses and treatments.

Key-Words: - Autism Diagnostic Observation Schedule (ADOS), Autism Spectrum Disorder (ASD), Autism Diagnostic Interview-Revised (ADI-R), Classification Algorithms, Machine Learning(ML), Support Vector Machine (SVM), Decision Tree (DT), K Nearest Neighbors (KNN), Naive Bayes (NB), Logistic Regression (LR), Random Forest Classifier (RFC).

Received: August 28, 2023. Revised: November 22, 2023. Accepted: December 19, 2023. Published: December 31, 2023.

1 Introduction

Autism Spectrum Disorder (a neurological illness) is an enlightening inability caused by Variances in the brain. There are specific difficulties with verbal and gesture communication, social interaction, and recurring behavior among persons with autism. Autism spectrum disorder will be acknowledged by Medical Consultants by analyzing the variation in people's behavior with the condition based on their age and level of competence. [1], [2], [3], ASD has a long-term impact on individuals, their families, and the provision of educational and therapeutic services. DSM-5 describes human interaction abilities as (a) deficiencies in teamwork, (b) irregularities in communication skills, and (c) problems establishing, sustaining, and comprehending relationships. In light of problems with effective interaction and repeating and restricted behavior, DSM-5 classified ASD into three severity levels. Level one - Requiring Support, Lack of social communication causes obvious impairments if there are no supports in place. Level two - Requiring Substantial Support, Social impairments are noticeable even with assistance, and Significant weaknesses in verbal and nonverbal social communication abilities level three - Requiring Very Substantial Support - Ability to initiate social contacts is severely hampered by severe inadequacies and there is little reaction to social advances made by others.

The ADI-R and ADOS are two clinical instruments and diagnostic approaches that are unfortunately time-consuming and necessary for the medical identification of people with autism (ADOS), [4]. Parent interviews are a part of the rigorous, organized, and structured ADI-R test, which looks at the child's early life, [5]. ADOS advises the parent about the child's current social, communication, and play abilities during a prearranged play session.

2 Literature Survey

Several kinds of research with the help of machine learning was done to speed up and enhance the diagnosis of ASD. Utilizing machine learning, which had an accuracy of 84%, to categorize the retinal obsessions of kids with ASD and TD. These investigations showed that when compared to established diagnostic measures, machine learning is more efficient and objective, [6]. The integrating multidimensional features gives the best degree of accuracy for depicting the correlation coefficients of the brain when compared to examining individual heterogeneous variables, [7]. With an accuracy of 97.6%, The Support Vector Machine technique to screen for ASD. The limitation of this paper is the small sample containing 612 autistic cases and 11 non autistic cases, [8]. Sixty-five Social Impartiality individuals from two thousand nine hundred and twenty-five individuals with ASD or ADHD with the help of six machine learning models. They employed forward feature selection with minimum sampling, [9]. Five out of the sixty-five tests, with an efficiency of 96.4%, were sufficient to differentiate ASD from ADHD. There was a considerable imbalance favoring the ASD class in the dataset, which was mostly based on collections about autism, [10]. There were 367 variables in the 95,577 children's data, 256 of which were deemed sufficient, and the accuracy was 87.1% in SVM, RFC, and NB. The innovative intelligence-sharing structure was designed to hide responsive and simultaneously unstable individual data. The study also recommended linear variance analysis as a simple way for keyword separation (LVA), [11]. SVM was the technique applied in this filtration procedure. The study's findings offer hope for employing text mining to safeguard private health information transmitted over the Internet, [12]. ASD is predicted using parameters based on brain activity, 95.9% accuracy was achieved using SVM with 2 groups and 19 features. The amount of data was relatively small. It used a cross-validation approach to extract 6 personality variables from the data of 851 individuals to train and evaluate their ML models. This was utilized to categorise patients into those with and those without ASD, [13]. Facial expressions from photo to identify psychological stress is not possible. Using haar cascade algorithm is used to evaluate the stress via logarithmic regression, [14], [15].

3 Working Principle

Data processing converts raw data into a more comprehensible format from a preliminary step. Figure 1 illustrates the workflow of our proposed system. The dataset will be preprocessed for missing data, duplicate data, and noisy data. Pre-processing of the data can be done by dimensionality reduction technique. Autism can be predicted using classification approaches after the dataset is preprocessed. The efficiency of each classification may contrasted. The training efficiency will be greater than the test efficiency if the classification performs properly. Then, this classification algorithm may be used for the next training and classification as the best model.

3.1 Dataset Pre-processing

One of the most essential components in Machine Learning is the dataset. Once the dataset is collected, researchers should employ various dataset types that may depend on their prediction model. Preprocessing can convert the data into a pattern that is more easily and productively refined in data mining, and machine learning through dimensionality reduction techniques using selection and extraction.

3.2 Feature Selection

It is the process of contracting the proposal to your model by using only proper data so the noise in data can be cleared. The most essential task for building an efficient classification system. There are three methods of feature selection are available.

3.2.1 Filter Method
Features are processed based on universal aspects of the dataset such as interaction with dependent variables. A faster and better approach when the sum of features is enormous.

3.2.2 Wrapper Methods
It is the process of removing the weakest feature and applying a model until a desired feature is selected.

3.2.3 Embedded Methods
It is the process of selecting a feature based on its weight. It combines the qualities of filter and wrapper methods.

3.3 Classification Algorithm
It utilizes proposed training data to predict the expectation that the data will fall into one of the fixed grades. The training set and testing set were separated from the entire dataset. 80% of the data used to train the classification algorithm will come from the training set, and 20% will come from the testing set. This sample data partitioning into training sets and testing sets enables us to evaluate the overfitting or underfitting of our model. Later the collection of preprocessed data, the grade models are enforced to detect ASD. Each classification algorithm's performance is assessed by the level of accuracy it achieves, including NB, SVM, LR, RFC and KNN

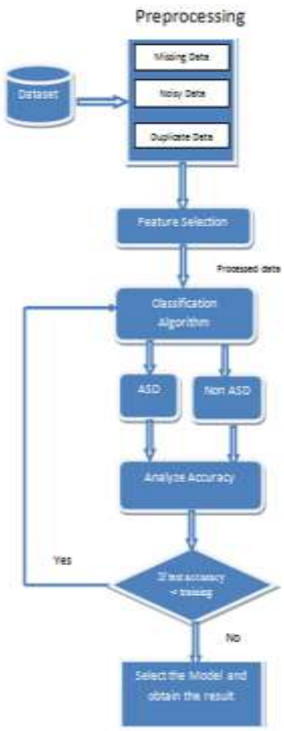

Fig. 1: Workflow of Proposed System

3.3.1 Decision Tree (DT)
It is the most influential and prominent tool for allocation and prognosis. To forecast the output class, DT builds the supervised learning model using a collection of IF-THEN rules from the training set. Based on characteristics in the dataset, a hierarchical tree is built. A root node, an internal node, and a leaf node are the three different types of nodes that make up a decision tree, [16].

3.3.2 K Nearest Neighbors (KNN)
In this strategy, we compare the new sample with the training dataset to identify the k examples that are close to the new sample in terms of class. We refer to them as neighbors. The classification of this sample of data is then established via a majority vote among neighbors, [17].

3.3.3 Support Vector Machine (SVM)
Regression or classification problems are used in SVM. Every data is plotted as a spot in n-dimension using the support vector machine, where n = number of functionalities that possess and each functionality is the value of a certain coordinate, [18].

3.3.4 Random Forest Classifier (RFC)
It is a classification algorithm, which has many decisions trees. Consists of a large number of personal decision trees that operate as an ensemble. Each personal tree in the forest spits out a class prediction. Both continuous variables, as in regression, and predictor variables, as in classification, can be included in data sets that RFC can handle. It produces better results in terms of classification problems

3.3.5 Naive Bayes (NB)
It makes the process easier to build machine learning models that can produce accurate predictions. Learners and classifiers are immensely quick, when related to refined forms. This helps to ease issues from the curse of dimensionality, [19].

3.3.6 Logistic Regression (LR)
Logistic Regression is used when the dependent variable (target) is categorical. It is equal to the Linear Regression except for how they are used. It can be used to classify the perception using various types of data and it can easily resolve the most efficient variables used for the classification, [20].

4 Evaluation

Evaluation is based on the performance of classification how much the correctness of the result of the training system and the testing system. The scale are listed below:

Table 1. Performance of Classification algorithm

Classification algorithm / Matrices	Accuracy %	Specificity %	Sensitivity %
SVM	89	91	88
KNN	88	92	90
DT	88	86	90
RF	91	93	89
LR	90	91	88
NB	88	88	89

True Positive (TP): The amount of outcomes to be true in both predicted and diagnosed cases
True Negative (TN): The amount of outcomes to be false in both predicted and diagnose cases.
False Positive (FP): The amount of outcomes that are predicted to be false but are really true.
False Negative (FN): The amount of outcomes that are predicted to be true but are really false.

Accuracy: It is a very important metric in Machine Learning. It is calculated as the percentage of total right predictions overall predictions. Figure 2 represent the evaluation graph of accuracy.

$$\text{Accuracy} = \frac{\mathbf{TP+TN}}{\mathbf{FN+TP+TN+FP}}$$

Specificity: Another name for specificity is the True Negative Rate (TNR). It is a capability to estimate a true negative for each accessible category. Figure 3 represent the evaluation graph of specificity.

$$\text{Specificity (TNR)} = \frac{\mathbf{TN}}{\mathbf{TN+FP}}$$

Sensitivity: Other names for sensitivity include the True Positive Rate (TPR) or recall. Since it allows us to see how many situations the model was able to correctly identify, it is used to evaluate model performance. Figure 4 represent the evaluation graph of sensitivity.

$$\text{Sensitivity (TPR)} = \frac{\mathbf{TP}}{\mathbf{TP+FN}}$$

Based on the analysis of classification the following Table 1. Indication of the matrices of accuracy, specificity, and sensitivity in various classification algorithms.

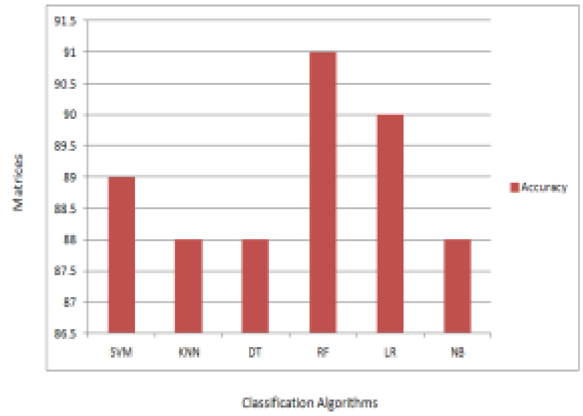

Fig. 2: Evaluation graph of Accuracy

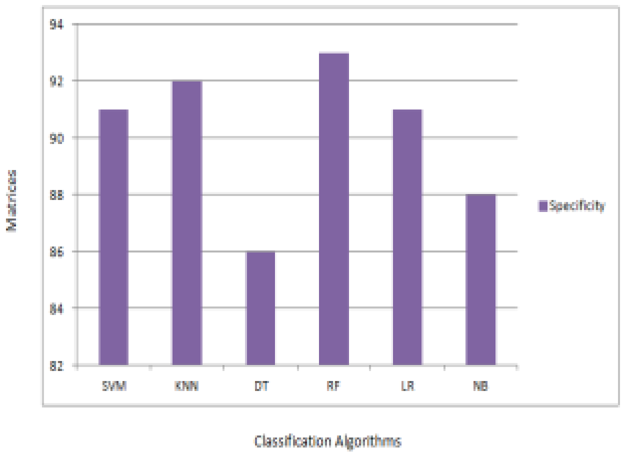

Fig. 3: Evaluation graph of Specificity

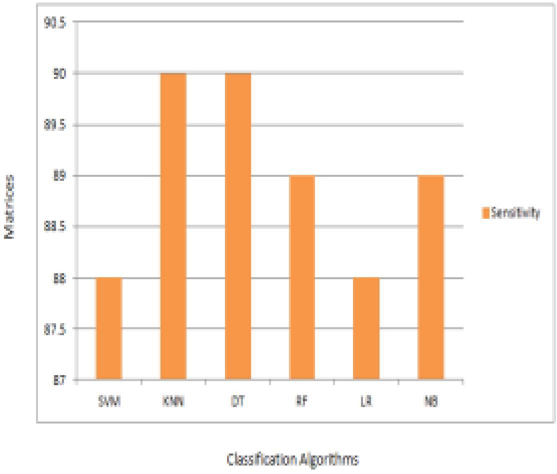

Fig. 4: Evaluation graph of Sensitivity

From this graph analysis, we can conclude Random forest classifier and Logistic Regression have more accuracy when compared with another classification algorithm.

5 Conclusion

This analysis of the study used several ML techniques to identify autism spectrum disorder. Using several performance evaluation measures, including specificity, accuracy, and sensitivity for detecting ASD. Having greater accuracy rates of 91% and 90%, respectively, were the classification algorithms Logistic Regression(LR) and Random Forest(RF) Classifier. Future autism spectrum disorder diagnoses will be more sensitive, specific, and accurate thanks to a range of machine-learning techniques.

References:

[1] Guangyao Shen, Jia Jia, Liqiang Nie, Fuli Feng, Cunjun Zhang, Tianrui Hu, "Depression Detection via Harvesting Social Media: A Multimodal Dictionary Learning Solution", *international Joint Conferences on Artificial Intelligence* vol.4, pp. 3838-3844, 2017.

[2] N. Rajaraman and A. P. R. Bhuja, "G Depression Detection of Tweets and A Comparative Test", *International Research Journal of Engineering and Technology (IRJET),* vol. 09, no. 03, March 2020, ISSN: 2278-0181.

[3] L. Squarcina, F. M. Villa, M. Nobile, E. Grisan and P. Brambilla, "Deep learning for the prediction of treatment response in depression", *Journal of Affective Disorders,* vol. 281, pp. 618-622, 2021.

[4] T. Zhang, A. M. Schoene, S. Ji and S. Ananiadou, "Natural language processing applied to mental illness detection: a narrative review", *NPJ digital medicine*, vol. 5, no. 1, pp. 1-13, 2022.

[5] Thabtah, Fadi. An Accessible and Efficient Autism Screening Method for Behavioural Data and Predictive Analyses. *Health informatics*, 25, 1739-1755 (2018).

[6] Vaishali, R and Sasikala, R, A Machine Learning Based Approach to Classify Autism with Optimum Behaviour Sets, *International Journal of Engineering & Technology*. 7 (2018).

[7] Sen B, Borle NC, Greiner R, Brown MR. A General Prediction Model for the Detection of ADHD and Autism Using Structural and Functional MRI. *PLoS ONE*. 13: e0194856. (2018).

[8] Thabtah, Fadi and Peebles, David.: A New Machine Learning Model Based on İnduction of Rules for Autism Detection, *Health Informatics Journal*. 26(1) 264-286, (2020).

[9] J. Chen, G. Wang, K. Zhang, G. Wang, and L. Liu,: A Pilotstudy on Evaluating Children with Autism Spectrum Disorder, *Journal Of Healthcare Engineering Using Computer Games. Computers in Human Behavior.* Vol.90, pp.204–214, (2019).

[10] Duda M, Ma R, Haber N, Wall DP.: Use of Machine Learning for Behavioral Distinction of Autism and ADHD, *Transl Psychiatry*. 6: e732. Doi. 10.1038/tp.2015.221, (2016).

[11] Suman Raj and Sarfaraz Masood,: Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques, *Procedia Computer Science, Elsevier,* vol.167 pp.994-1004 (2020).

[12] Vakadkar K, Purkayastha D, Krishnan D. Detection of Autism Spectrum Disorder in Children Using Machine Learning Techniques. *SN Comput Sci*. 2021; 2(5):386. doi: 10.1007/s42979-021-00776-5.

[13] Deshpande G, Libero LE, Sreenivasan KR, Deshpande HD, Kana RK.: Identifcation of Neural Connectivity Signatures of Autism using Machine Learning, *Frontiers in Human Neuroscience,* vol. 7, doi. 10.3389/fnhum.2013.00670, (2013).

[14] Andrew Yates, Arman Cohan and Nazli Goharian, "*Depression and selfharm risk assessment in online forums*", arXiv preprint, 2017, doi.10.48550/arXiv.1709.01848.

[15] Minsu Park, David W McDonald and Meeyoung Cha, Perception differences between the depressed and non-depressed users in twitter, *Seventh International AAAI Conference on Weblogs and Social Media*, 2013, 7(1), pp.476-485, doi.10.1609/icwsm.v7i1.14425

[16] I. Salehin, I. M. Talha, N. Nessa Moon, M. Saifuzzaman, F. N. Nur and M. Akter, "Predicting the Depression Level of Excessive Use of Mobile Phone: Decision Tree and Linear Regression Algorithm," *2020 IEEE International Conference on Sustainable Engineering and Creative Computing (ICSECC)*, Indonesia, 2020, pp. 113-118, doi: 10.1109/ICSECC51444.2020.9557394.

[17] M. R. Islam, A. R. M. Kamal, N. Sultana, R. Islam, M. A. Moni and A. ulhaq, "Detecting Depression Using K-Nearest Neighbors (KNN) Classification Technique," *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)*, Rajshahi, Bangladesh, 2018, pp. 1-4, doi: 10.1109/IC4ME2.2018.8465641.

[18] A. Ahmed, R. Sultana, M. T. R. Ullas, M. Begom, M. M. I. Rahi and M. A. Alam, "A Machine Learning Approach to detect Depression and Anxiety using Supervised Learning," *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*, Gold Coast, Australia, 2020, pp. 1-6, doi: 10.1109/CSDE50874.2020.9411642.

[19] Islam MR, Sohan M, Daria S, "Evaluation of inflammatory cytokines in drug-naïve major depressive disorder: A systematic review and meta-analysis"2023 *International Journal of Immunopathology and Pharmacology*, vol. 37, pp. 1- 15.

[20] S. Jayawardena, J. Epps and E. Ambikairajah, "Ordinal Logistic Regression With Partial Proportional Odds for Depression Prediction," in *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 563-577, 1 Jan.-March 2023,

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This research received no specific grant from any Funding agency in the public, commercial, or not for-profit sectors.

Conflict of Interest

The authors have no conflicts of interest to declare.