



International Journal of Science, Architecture, Technology and Environment

Intelligent Predictive Modelling for Autism Spectrum Disorder Using Machine Learning Approaches

G. Divya^{1*}, Dr. V. Maniraj²

¹Research Scholar, Department Of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur (Dt), Affiliated to Bharathidasan University, Thiruchirappalli, Tamilnadu

²Associate Professor, Research Supervisor, Head of the Department, Department of Computer Science, A.V.V.M Sri Pushpam College (Autonomous), Poondi, Thanjavur (Dt), Affiliated To Bharathidasan University, Thiruchirappalli, Tamilnadu

*Corresponding author, gdivya19.mca@gmail.com

DOI: <https://doi.org/10.63680/ijate1125012.012>

Abstract

Autism spectrum disorder (ASD) is a Neuro complaint that influences an individual's communication, announcement, and knowledge talents. Analysis of Autism can be completed at any age-group level. Autism patients look at diverse kinds of disputes learning disabilities, and complexity with meditation. Mental health problems, motor difficulties, and sensory problems are some of the problems faced by Autism patients. Earlier diagnosis and proper medication at the early stage are essential to control ASD. The ASD prediction framework is built to support a behavioural aspect-based analysis model without any device in this research. The ASD prediction process is focused on the childhood and adolescent analysis model utilized in the system. The behavioural parameters are collected with the support of the Autism Query collections. The decision tree (DT) and Support Vector Machine (SVM) techniques, K-Nearest Neighbours (KNN), and Artificial Neural Network (ANN) are applied for the ASD prediction process. The Correlated Feature selection based Random Forest (CFS-RT) algorithm is applied for the ASD prediction process, giving an accuracy of 93.03%, and ANN produces 97.68% and outperformance other methods.

Keywords: Autism Spectrum Disorder, Decision Tree, Machine Learning, Data Mining, Support Vector Machine.

1. Introduction

Autism spectrum disorder (ASD) is a progressive disability that can benefit important basis community, contact, and behavioural issues. There are frequent tells how a community with ASD looks that sets them together from new persons. Still, individuals with ASD could converse, interrelate, function, perform and hear in the approach part from mainly individual people. The knowledge, thoughts, and difficulty-resolving facility of persons with ASD can provide a choice from talented to more confronted. An analysis of ASD contains a lot of provisions that applied to be analyzed alone as autistic disorder syndrome. These situations in named ASD.

ASD is a state connected to mind growth that forces how a person recognizes and meet peoples, obtaining difficulty in public contact and dealings. The disorder also contains genetic factors, environmental aspects, biological factors, and inadequate and recurring behavior patterns. In the United States, more children's affected by Autism. Each person with Autism has different issues, like anxiety, seizures, and depression, as shown in Figure 1. Autism can be obtained at two or three and diagnosed as early as before 18 months. Early detection gives a good impact on the life of persons. Persons with ASD may perform work, interconnect, interrelate, and acquire behaviors that are changed from person to person. The skills of a person with ASD can differ significantly. The person with ASD are issues with the social announcement and limited behaviors. The children have varying habits of culture and give less attention. Diagnosing ASD is also a problem because there is no proper medical test or blood test to identify the disorder. Doctors see the child's behaviour and can start a diagnosis to increase the quality of life.

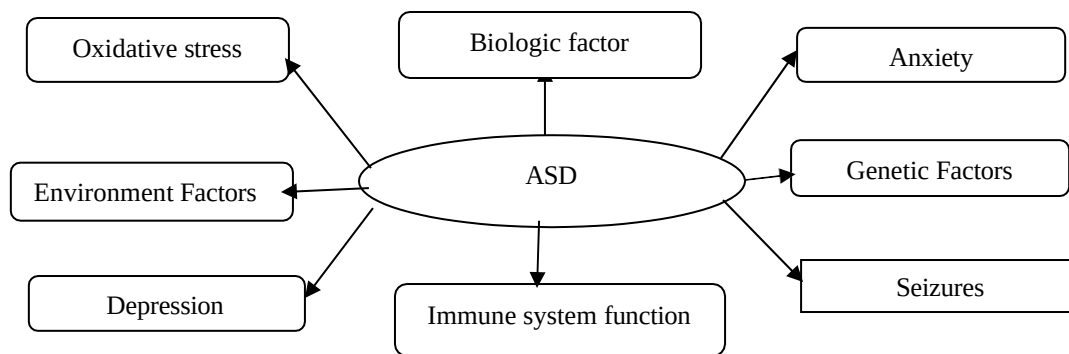


Fig.1 Outline the Factors for obtaining ASD

ASD disturbs how an individual reacts, study, and states themselves. For parents of a child with ASD, significantly which level the child has can help prepare them for issues their child might face in daily life? Figure 2 shows the level of ASD. The first level needs help for every work, and the second level indicates the need for sustainable help. Three-level indicate the need for more sustainable help and very problem to changing focus and actions.

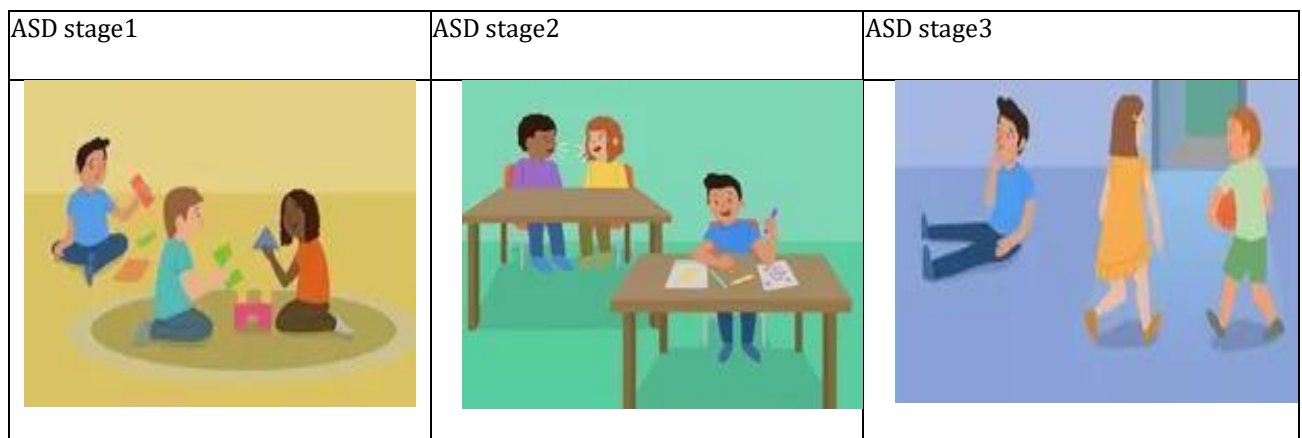


Fig. 2 Type of ASD

1.1 Supervised Machine Learning

Machine Learning (ML) is a division of Artificial Intelligence (AI) that studies and determines useful models in data to formulate diagnoses in the medical field. It is put up on mathematical opinions of likelihood and data and information. This supervised learning is an ML approach that studies by plotting the input features and output mutable for prediction. It is shown in Figure 3 and learning through training and test data set.

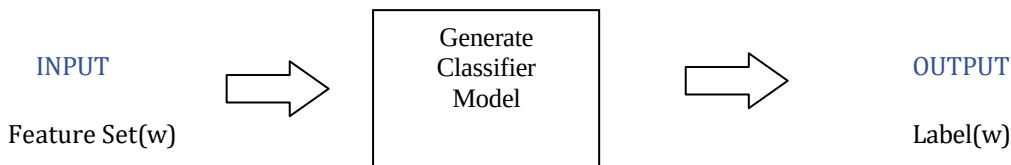


Fig.3 Classifier Model

As shown in Figure 1, a representative controlled learning difficulty includes attribute space W called features in the data set, a label space Z , and a prediction space Z' . Classifier maps each feature w to one of the already defined class labels z .

A training sample set is illustrated as $Y = ((w_1, z_1), \dots, (w_m, z_m)) \in (W \times Z)^m$ that

Consists of prearranged class labels, where each $w_i \in Z$. The output is a model $H_s: W \rightarrow Z'$ that learns from the sample set. These approaches find the class of new instances, $Z' = Z = \{\pm 1\}$ and input samples are a map with two labels as $Z = \{\pm 1\} = \{-1, +1\}$ in binary classification.

1.2 Classification Algorithms

The ID3 algorithm is an approximation supervised learning algorithm through the sample for various labels. The decision tree root and internal nodes hold the feature test uniqueness, whereas the leaf node includes the class label. The decision tree holds numeric and categorical variables from the sample and produces a visual representation. Many open-source tools contain a build-in decision tree algorithm. For example, WEKA produces good data visualization by visualize the tree option.

ID3 identifies features that discriminate one label from another. The features should be continuous or chosen from a group of identified values, and the feature should be well-known in progress. At the same time, this algorithm is more responsive to features when a huge number of data. Limitations in the ID3 are overcome through the C4.5 algorithm with the help of information gain measures.

Naïve Bayes is a supervised learning classifier for data mining and the field of machine learning that solve the classification issues by generating the Bayes rule. This classifier holds any volume of data and also holds missing values effectively. The target is to forecast the class l as the label, f_i as features, and select a significance of l that exploit $P(l|F_1, F_2, \dots, F_n)$ given a sample with attributes $(m_1, m_2, \dots, m_n)$ is shown in equation (1)

$$P(c|F) = \frac{P(m|l)P(l)}{P(m)} \quad (1)$$

K-Nearest Neighbours is an instance-based ML method for classifier, regression approach, and it's lazy learning because there is no explicit training. This algorithm stores the training observations and evaluates against the new opinions with the help of a similarity assessment as Euclidean distance according to real and binary values as shown in equations (2) and (3) and positions the stored data at some point in the test instant to formulate calculations. This algorithm also holds the missing values effectively.

Euclidean Distance:

$$f(w_i, z_j) = \sqrt{\sum_{im=1}^D (w_{im} - w_{jm})^2} \quad (2)$$

Hamming Distance:

$$f(w_i, z_j) = \sum_{im=1}^D II(w_{im} \neq w_{jm}) \quad (3)$$

The sample data j fields to frame a j -dimensional vector, $w = (w_1, w_2, \dots, w_j)$, and the training data is characterized by $F = \{(w_1, z_1), \dots, (w_j, z_j)\}$ where x_i is the content of the input and y_i is the named as the label. The w_i is a vector that contains F attributes, whereas x_{im} represents the m -th attributes of w . That means the result is $z_i \in \{1, \dots, C\}$.

2. Problem Definition

The AI and ML techniques are used in autism detection. The neuro imaging and kinematic data are used to discover the Autism Spectrum ASD levels. The ASD classification is carried out with electroencephalogram (EEG) signal dispensation and knowledge approaches. The EEG is a consistent biomarker for diagnosing ASD abnormalities. The electrodes attached to the scalp are used to detection the EEG signals. The mobile applications collect answers from the people for the Autism Questions. The MLg-based methods are applied to predict ASD at any age-group level. The Random Forest-CART (Classification and Regression Trees) algorithm predicts ASD traits. The Random Forest-ID3 (Iterative Dichotomiser 3) algorithm is applied to perform the ASD prediction process. The Correlation Feature Selection based Random Forest (CFS-RF) technique is built to discover ASD with behavioural features from different age levels.

Specific Objectives

- To design a system for Autism Spectrum Disorder (ASD) discovery with behavioral analysis.
- To perform age group based analysis on the ASD discovery process
- To design the Correlation-based Feature Selection with Random Forest (CFS-RT) technique for the ASD prediction process
- To eliminate device and technician dependency in the ASD detection process

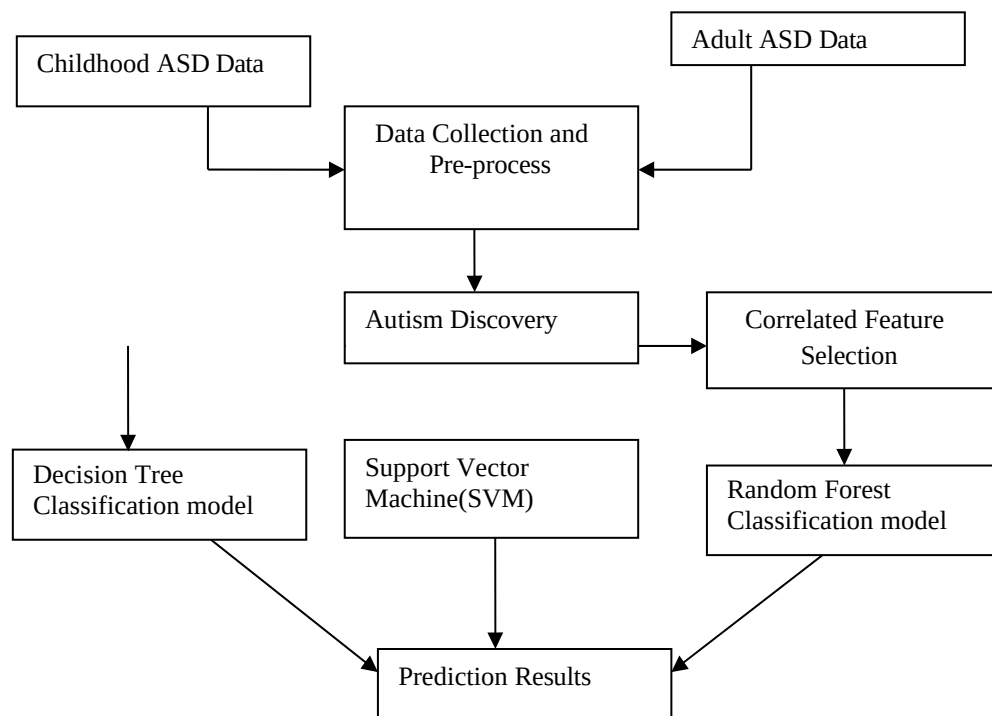
- To minimize cost and time in the ASD prediction operations

3. Proposed System Methodology

3.1 ASD Detection through Behavioural Features using Decision Trees

Adult autism detection is designed to analyze the behavioural attributes of patients above 18 years. The behavioural features are gathered with the help of a questionnaire. The Adult Questioner (AQ10) data set is get from the University of California, Irwin ML repository. Each person is diagnosed with 10 questions. The ASD is identified using the responses given by the individual. The decision tree classification is applied to discover the disease levels. Figure 4 shows the proposed system architecture.

The children dataset and the adolescent dataset is stored in the data repository. It contains missing and incorrect values. This can be pre-processed and get relevant data to build the ML classification models after pre-processing, splitting the operation as testing and training to generate the classifier to detect the Autism of children and adolescents. The proposed system uses DT, SVM, and Random Forest models to generate a good model to predict Autism in the early stage. In classification, 10-fold cross-validation is applied to estimate the model.



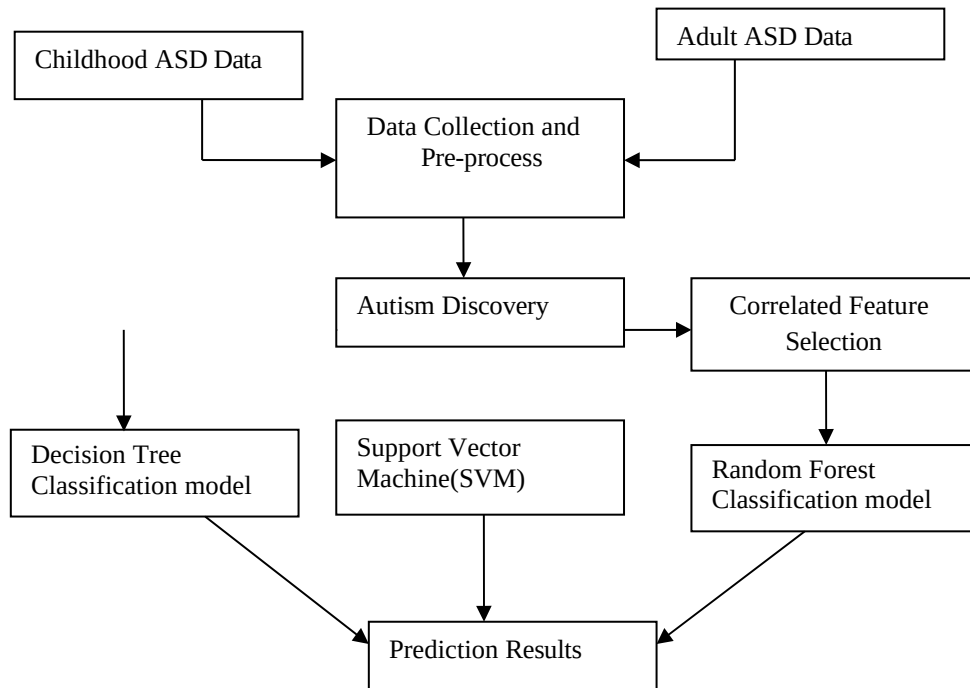


Fig.4 System Architecture Diagram

3.2 Autism Spectrum Disorder Prediction using Random Forest

The Autism Spectrum Disorder (ASD) prediction framework is built to support a behavioral aspect-based analysis model without any device. The behavioral parameters are collected with the support of the Autism Query collections. The ASD prediction process is focused on the age boundary-based analysis process. The childhood and Adolescents analysis model is utilized in the system. Random Forests is an ensemble classification that creates multiple decision trees during the training period and produces individual trees by means of mode or means finding. Random Forest of this machine learning algorithm avoids the overfitting problem in the Decision Tree.

3.3 Other Machine Learning Models

K- Nearest Neighbour (KNN) model used for regression and classification issues. The 'K' denotes the number of kernel theme that is to be nominated to minimize the error. KNN is constructed on the knowledge of similarity, which can relate to Euclidean distance, nearness, or closeness. The artificial Neural Network (ANN) model is a neural network that has a joining with numerous neurons.

4. Experiment Result and Discussion

4.1 Environmental setup

The autism prediction models are implemented and verified using the R language. The prediction process is evaluated with four factors. They are False Positive Rate (FPR), False Negative Rate (FNR), prediction accuracy, and prediction time measures. The FPR rate and FNR rate parameters are used to estimate the fault

level of the ASD detection process. The FPR and the FNR are the findings of error criteria. The error stage should be reduced to increase the detection accuracy levels. The prediction accuracy is applied to estimate the ASD discovery result accuracy levels. The processing time is calculated for the prediction process.

4.2 Data set description

The ASD detection process is implemented to detect the autism stages in childhood, adolescence, and adulthood. The AG10 dataset is taken from the University of California machine learning repository. The children's data set has 292 instances. The adolescent data set has 104 instances. The adult data set is collected for 704 patients. The entire data set has 1100 instances with 21 categorical, continuous, and binary attributes. The ML-based ASD forecast process uses the Decision Tree Classification (DTC), SVM, and CFS-RF techniques. Table 1 and Table 2 show the ASD data set representation.

Table1.Dataset representation

S.No	Name of the dataset	Instances in dataset
1	Adult-ASD screening dataset	704
2	Children-ASD screening dataset	292
3	Adolescent-ASD screening dataset	104

5. Measures used for evaluation

The verified criteria are used to assess the worth of a classifier. For verifying a classifier by contingency matrix, precision and recall, Accuracy, ROC, and time analysis. The contingency matrix contains True Positive (tp), False Positive (fp), False Positive (fp), and False Negative (fn)

- tp – Quantity of optimistic instances in the trial or experimental dataset that are categorized as positive
- fp – Quantity of undesirable instances in the trial or experimental dataset that are categorized as positive
- tn – Quantity of undesirable instances in the trial or experimental dataset that are categorized as negative
- fn – Quantity of optimistic instances in the trial or experimental dataset that are categorized as negative

Accuracy of the classifier: Accuracy is repeatedly used result metric for prediction. Obtaining more accuracy in the model in the ML approach is the highest aspect of education (5).

$$\text{Classification Error} = \frac{fp+fn}{tp+tn+fp+fn} \quad (5)$$

6. Conclusion and Future work

The Correlation Feature Selection based Random Forest model increases the accuracy level in the prediction process. The behavioural and age group parameters are integrated into the prediction process. The data dimensionality is managed with feature selection methods. The system solves the device dependency, technician requirements, and delay problems. The ASD mechanism can be enhanced to analyze the adult autism levels. The ASD prediction system can be enhanced with the following features.

- The system can be improved with automated medicine and a treatment suggestion model.
- The system can be improved to analyze the treatment and its impacts on regular patients' Activities.
- The ASD prediction process can be upgraded by integrating behavioural features and EEG- Based features analysis.
- The ASD prediction process can be implemented as a mobile-based application to Conduct an automated self-assessment test.

Declaration of Conflicting Interests

The authors declare no potential conflicts of interest with respect to the research, authorship and publication of this article.

Funding

The author received no financial support for the research, authorship and publication of this article.

References

- [1] M. Thangamani, R. Vijayalakshmi, M. Ganthimathi, Rajita, P. Malarkodi, S. Nallusamy, Efficient classification of heart disease using K-Means clustering Algorithm, International Journal of Engineering Trends and Technology, 68(12) (2020) 48-53.
- [2] N. Suresh Kumar, M. Thangamani, V. Sasikumar S. Nallusamy, An Improved Machine Learning Approach for Predicting Ischemic Stroke, An Improved Machine Learning Approach for Predicting Ischemic Stroke, 68(1)(2021) 111-115.
- [3] Thangamani M., Ganthimathi M., Sridhar S.R., Akila M., and Keerthana R. Engineering, Detecting Coronavirus Contact using Internet of things., International Journal of Pervasive Computing and Communications, 16(5) (2020) 447-456. <https://doi.org/10.1108/IJPCC-07-2020-0074>.
- [4] Dilantha Haputhanthri, Gunavarana Brihadiswaran, Sahan Gunathilaka, Dulani Meedeniya, and Sampath Jayarathna, An EEG based Channel Optimized Classification Approach for Autism Spectrum Disorder, Moratuwa Engineering Research Conference (Mercon), (2019).
- [5] Surendiran, R., and Alagarsamy,K., 2010. Skin Detection Based Cryptography in Steganography (SDBCS). International Journal of Computer Science and Information Technologies (IJCSIT), 1(4) (2010) 221-225.
- [6] Supriya H S, Sumanth Alva R*, Suprith K P, Vikas Kumar L, and P Kyshan Neheeth, applying supervised learning

- technique to diagnose autism spectrum disorder (asd), *International Journal of Recent Scientific Research* , 11(5b) (2020) 38439-38441. DOI: <http://dx.doi.org/10.24327/ijrsr.2020.1105.5312>.
- [7] Inon Wiratsin, L. Narupiyakul, Feature Selection Technique for Autism Spectrum Disorder, *Proceedings of the 5th International Conference on Control Engineering and Artificial Intelligence*, DOI, 10.1145/3448218.3448241, (2021).
- [8] Surendiran, R., 2017. Development of Multi-Criteria Recommender System, *SSRG International Journal of Economics and Management Studies (IJEMS)* ISSN: 2393 - 9125, 4(1) (2017) 28-33.
- [9] Daniel Bone, Somer Bishop, Matthew P. Black, and Shrikanth S. Narayanan, Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion, *Journal of Child Psychology and Psychiatry*, (2016).
- [10] G. Divya & Dr. V. Maniraj, "Leveraging Machine Learning for a Comprehensive Analysis of Autism Spectrum Disorder: A Synthesis of Scholarly Insights", *International Journal of Computational Research and Development*, Volume 9, Issue 2, July - December, Page Number 91-94, 2024