

Prediction of Autism Spectrum Disorder Using Supervised Machine Learning Algorithms

T. Lakshmi Praveena¹ and N. V. Muthu Lakshmi²

¹Research Scholar, ²Assistant Professor

^{1&2}Sri Padmavati Mahila Visva Vidyalayam, Tirupati, Andhra Pradesh, India

E-Mail: praveenalaxmi1@gmail.com, nvmthulakshmi@gmail.com

(Received 3 July 2019; Accepted 28 August 2019; Available online 9 September 2019)

Abstract - Autism appears to be a neuro developmental disorder that is visible in the early years. It is a wide-spectrum disorder that indicates that the severity and symptoms can vary from person to person. The Centre for Disease Control found that one in 68 was diagnosed with autism spectrum disorder with increasing numbers in every year. Detection of autism in adults is a cumbersome procedure because in adults, many symptoms can blend with some other mental health, motor impairment disorders so misinterpretation of actual diseases can in turn lead to a terrible life without proper diagnosis and effective treatment mechanisms. Machine learning is a powerful computer tool that supports different application domains Learning complex relationships or patterns from large datasets to draw accurate conclusions. Disease assessment can be done with predictive health data analysis and more appropriate treatment mechanisms that are now a hot area of research. Supervised learning is an important step of Machine learning which uses a rule-based approach by examining empirical data sets to build accurate predictive models. In this paper, decision tree, random forest, SVM, neural networks algorithms are applied on autism spectrum data which have been collected from UCI repository. The results of decision tree, random forest, SVM, neural networks algorithms on autism dataset are presented in this paper in an efficient manner. Analysis performed over these accurate results which will be useful to make right decisions in predicting autism spectrum disorder (ASD) at early stages. Thus, early autism intervention using machine learning techniques opens up a new way for autistic individuals to develop the potential to lead a better life by improving their behavioural and emotional skills.

Keywords: Autism Spectrum Disorder, Supervised Learning, Decision Trees, Random Forest, SVM, Neural Networks

I. INTRODUCTION

Autism spectrum disorder (ASD) is a neuro-developmental disorder. In general, neuro developmental disorders symptoms are hyperactivity, deficiency in social communication, deficiency in learning and language. America's children and the environment (ACE) published an article on neuro developmental disorders and also detailed review of ASD and other disorders [1]. ASD can also be caused by genetic, chemicals used in different food items or environmental factors. ASD affected children's ratio is increasing day by day with ratio of 1:68. ASD can be detected for a person by diagnosis, based on the children and the level of ASD it may require 2 or 3 days. Diagnosis process is manual process conducted by multiple streams of doctors by observing the behaviour of a child. Actually,

ASD is detected at the age of 3 years or above 3 years [2] with manual diagnosis. Treatments started at this level takes long time to show benefits. For the past few years, many research works have done to speed up the diagnosis process so that ASD detection can be done at an early age which improves the curing mechanism. Researchers have worked with many techniques to address this issue among many; machine learning is one, which is efficient and reliable to detect ASD with less processing time. Machine learning can train the system with past data and then machine will predict ASD within short period of time.

Machine learning can be broadly categorized into two supervised and unsupervised. Supervised learning is more suitable and accurate in predicting ASD which uses a rule-based approach by examining empirical data sets to build accurate predictive models. There are many algorithms that exist in supervised learning, In this paper, the efficient algorithms are identified and are considered in detecting ASD which are decision tree, random forest, Support Vector Machine(SVM), neural networks algorithms and are applied over autism spectrum data which have been collected from UCI repository. The results obtained from these algorithms decision tree, random forest, SVM, neural networks on autism dataset have presented in this paper. Analysis performed over these accurate results which will be useful to make right decisions in predicting autism spectrum disorder (ASD) at early stages. Thus, early autism intervention using machine learning techniques opens up a new way for autistic individuals to develop the potential to lead a better life by improving their behavioural and emotional skills. Following sections presents ASD, Machine learning, supervised learning, Analysis process and result analysis.

II. AUTISM SPECTRUM DISORDER (ASD)

ASD can be known as psychological and neurological disorder. ASD is caused by unequal development of some regions of brain. Autism spectrums disorders (ASD) are typically are diagnosed by manual process. Even the symptoms dissent for different ASD persons and individuals [3]. The general similarities between ASD individuals are difficulties in social interaction, communication and restricted unimaginitive behaviors. ASD has increased dramatically from the mid of 20th century. Today this increase reached to 1 out of 68 individuals [4, 5]. Need of

research in diagnosis methods, prediction methods and detection methods are increasing [6]. ASD is also called as pervasive disorder. ASD include sub categories Autistic Disorder and Asperger disorder [12]. The symptoms of ASD are difficult to observe at the early age less than 3 years. Early diagnosis of ASD helps to promote nonmedical treatment like physiotherapy, communication therapy, and sensory integration therapy. So that it is very important to find efficient prediction methods to detect ASD at early stage of less than 3 years. In recent years technological revolutions are taken place in ASD research. Machine learning is evolved from this research. The main objective of using technical methods in ASD prediction is to speed up diagnosis process and detect ASD at early stage. Machine learning algorithms are best suited for this purpose. Next section discuss about machine learning.

III. MACHINELEARNING ALGORITHMS

Machine learning is known as training an algorithm to apply on different test data [7].The main aim of machine learning algorithms is to train system and work automatically without human intervention [13]. The learning procedure starts by pre-processing the data and removing noisy data from dataset like missing values, outliers. The pre-processed data is used for generating model using machine learning algorithms. Data set is partitioned into two halves such as training data and test data.

The machine is trained by learning algorithm from training dataset and it generates a new model and this model tested with test dataset. Once the model is generated then the model can e used for any test data without algorithm. As machine learning is classified into two supervised, unsupervised which are further classified and are shown in fig.1. Supervised machine learning algorithms are SVM, Decision trees, Random Forest, Naive Bayes and neural networks [8] which are based on classification. For prediction of value supervised learning algorithms are used. The present paper presents prediction analysis of ASD using

supervised learning algorithms and comparative analysis of algorithms.

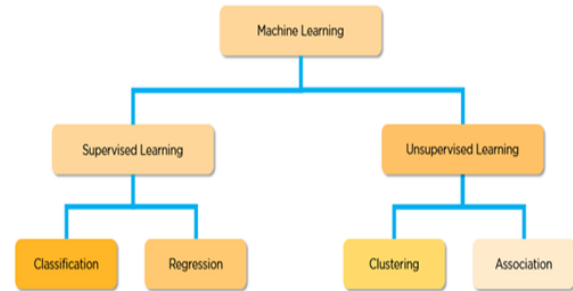


Fig. 1 Machine learning classification

Next section explains supervised learning algorithms in brief.

IV. SUPERVISED LEARNING

Supervised learning is one of the parts of machine learning and Artificial Intelligence (AI). The real time systems that uses supervised learning and AI are self-driving cars, facial recognition systems etc [10]. ASD prediction is performed by generating a model with existing dataset which is used for future prediction of ASD with less processing time [11]. Supervised learning uses existing input, output data to generate a model for future data processing or prediction [9]. The process of prediction of ASD using supervised learning algorithms includes two steps. In the first step training data is used to generate a model classifier. In the second step generated classifier model is applied with test data [14]. Supervised learning algorithms selected in present paper for ASD prediction are Decision Trees, SVM, Random Forest, and Neural Networks. These algorithms predict ASD with optimal accuracy. This statement is proved in by comparing prediction error rates and accuracy of prediction. The following section describes chosen ASD data set to train the machine using supervised learning algorithms.

TABLE I ASD ADOLESCENT DATASET ATTRIBUTE DESCRIPTION

S. No.	Attribute Name	Description	Possible Values
1	A1 to A10	Diagnosis Questionnaire result.	Yes-1,No-0
2	Age	Age of the diagnostic patient. Continuous values	minimum value: 1year and maximum value : 80 years
3	Sex	Gender of the patient.	M-Male,F-Female
4	Ethnicity	This is a category of people who identify with each other based on similarities such as common ancestry, language, history, society.	White, Black, Hispanic, Aboriginal, Latino Middle Eastern, Asian, South Asia, Others.
5	Jaundice	Jaundice effected immediate after delivery or not.	Yes-1,No-0
6	Family ASD	ASD existence in family hierarchy	Yes-1,No-0
7	Residence	Country or state of residence	Different states and countries in Asia, South Asia and others.
8	Used App Before	Whether patient assessment is first time or not.	Yes-1,No-0
9	Score	Sum of values of A1 to A10 attributes	Minimum value: 0 and Maximum Value:10
10	Screening Type	Age group of patients selected for screening	1-3,4-11,12-16,17 and above
11	Language	Language used for communication	English, Russian, Farsi, Spanish, Arabic, French.
12	User	Person who bought patient for screening.	Self, Parents, Others, Relative
13	ASD	Class variable. Classified under ASD or not	Yes and No

V. ASD DATASET AND ATTRIBUTES ANALYSIS

ASD adolescent dataset has 23 attributes with combination of 10 binary attributes, 2 continuous attributes, 10 categorical attributes and one binary class variable. Table I shows attribute information of adolescent dataset. The dataset size in terms of number of records in adolescent dataset is 249.

VI. COMPARITIVE ANALYSIS OF SUPERVISED LEARNING ALGORITHMS

Comparison of supervised learning algorithms applied on ASD dataset. Error matrix shown in below table II gives percentage of correct classification and count of records that are classified with and without error.

TABLE II ERROR RATE ANALYSIS OF SUPERVISED MACHINE LEARNING ALGORITHMS FOR ASD PREDICTION

Supervised Machine Learning Algorithm	Test Data (20%)					Training Data (80%)				
	Proportions		Count of Records		Error Rate	Proportions		Count of Records		Error Rate
Decision Tree(J48)	No	Yes	No	Yes		No	Yes	No	Yes	
No	55.3	0	21	0	0	49.9	0	85	0	0
Yes	0	44.7	0	17	0	0	50.9	0	88	0
Random Forest										
No	55.3	0	21	0	0	49.9	0	85	0	0
Yes	0	44.7	0	17	0	0	50.9	0	88	0
SVM Model										
No	52.6	2.6	20	1	4.8	49.9	0	85	0	0
Yes	2.6	42.1	1	16	5.9	0	50.9	0	88	0
Neural Networks										
No	52.6	2.6	20	1	4.8	49.9	0	85	0	0
Yes	5.3	39.5	2	15	12	0	50.9	0	88	0

TABLE III ACCURACY ANALYSIS OF SUPERVISED LEARNING ALGORITHMS FOR ASD PREDICTION

Supervised Machine Learning Algorithm	Training Data Prediction Accuracy	Test Data Prediction Accuracy	Complete Dataset Prediction Accuracy
Decision Tree(J48)	100%	100%	100%
Random Forest	100%	100%	100%
SVM Model	100%	87.30%	95.90%
Neural Networks	100%	83.20%	95.20%

The comparison of supervised learning algorithms is used in selection of classification algorithm for efficient prediction of ASD. As per the results generated decision tree and random forest algorithms are efficient algorithms for better prediction of ASD.

VII. CONCLUSION AND FUTURE WORK

ASD is a neuro-developmental disorder and is a lifelong disorder. The prediction process of ASD is a long process. For better treatment and improvement in behaviour, detection process must be fast, efficient and accurate. Machine learning is one of the emerging methodology which takes less time to predict new results. This paper presents comparative analysis of supervised learning algorithms which helps in predicting ASD and selection of supervised learning algorithm. This work can extended to work with different types of ASD data like MRI scan data, EEG data and gene sequences data.

REFERENCES

- [1] A.M. Daniels, R.E. Rosenberg, J.K. Law, C. Lord, W.E. Kaufmann, and P.A. Law, "Stability of initial autism spectrum disorder diagnoses in community settings," *Journal of Autism and Developmental Disorders*, Vol. 41, No. 1, pp. 110–121, 2011.
- [2] Tracy A Becerra, Maria L Massillon, Vincent M You, Ashley A Owen-Smith, Frances L Lynch, Phillip M Crawford, MS, Kathryn A Pearson, Magdalena E Pomichowski, Virginia P Quinn, Cathleen K Yoshida, and Lisa A Crone, *A Survey of Parents with Children on the*

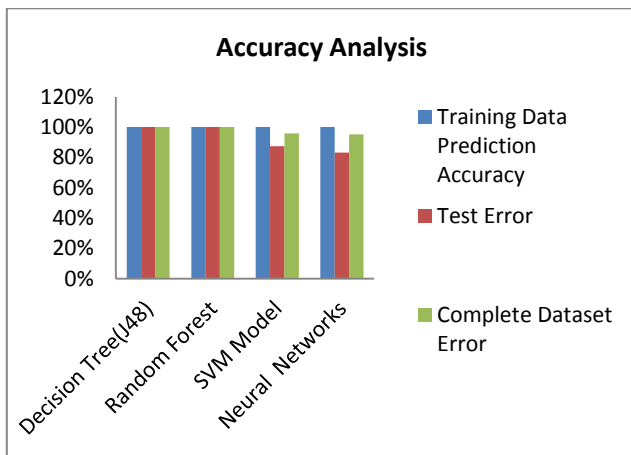


Fig. 2 Accuracy Analysis Graph

- Autism Spectrum: Experience with Services and Treatments*, [Online] Available at: <https://doi.org/10.7812/TPP/16-009>
- [3] Caroline P. Whyatt and Elizabeth B. Torres, *Autism Research: An Objective Quantitative Review of Progress and Focus Between 1994 and 2015*, Doi: 10.3389/fpsyg.2018.01526.
- [4] F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward," *Informatics for Health and Social Care*, pp. 1–20, 2018.
- [5] Alessandro Crippa, Christian Salvatore, Paolo Perego, Sara Forti, Maria Nobile, Massimo Molteni, and Isabella Castiglioni, *Use of Machine Learning to Identify Children with Autism and Their Motor Abnormalities*, DOI 10.1007/s10803-015-2379-8.
- [6] B. Auyeung, S. Baron-Cohen, S. Wheelwright, and C. Allison, "The autism spectrum quotient: Childrens version (aq-child)," *Journal of autism and developmental disorders*, Vol. 38, No. 7, pp. 1230–1240, 2008.
- [7] D. Bone, S.L. Bishop, M.P. Black, M.S. Goodwin, C. Lord, and S.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*, Vol. 57, No. 8, pp. 927–937, 2016.
- [8] R. GeethaRamani and K. Sivaselvi, "Autism Spectrum Disorder Identification Using Data Mining Techniques", *International Journal Of Pure And Applied Mathematics*, Vol. 117, No. 16, pp. 427-436, 2017.
- [9] Sheena Angra and SachinAhuja, "Machine learning and its applications: A review", *In proc. International Conference on Big Data Analytics and Computational Intelligence (ICBDAC-2017)*, DOI: 10.1109/ICBDACI.2017.8070809
- [10] Murat Gök, "A novel machine learning model to predict autism spectrum disorders risk gene", *Neural Computing and Applications*, pp.1–7, Springer.
- [11] Matthew J. Maenner, Marshalyn Yeargin-Allsopp, Kim Van Naarden Braun, Deborah L. Christensen, and Laura A. Schieve, *Development of a Machine Learning Algorithm for the Surveillance of Autism Spectrum Disorder*, published: December 21, 2016, [Online] Available at: <https://doi.org/10.1371/journal.pone.0168224>
- [12] Kiyoharu Takara, and Tsuyoshi Kondo, "Autism spectrum disorder among first-visit depressed adult patients: diagnostic clues from backgrounds and past history", *General Hospital Psychiatry*, Elsevier, Vol. 36, pp. 737–742, 2014.
- [13] Hana S. Alarifi and G. S. Young, "Using Multiple Machine Learning Algorithms to Predict Autism in Children", *In proc. Int'l Conf. Artificial Intelligence / ICAI'18, CSREA Press*.
- [14] M.S. Mythili, and A.R. Mohamed Shanavas, "A Study on Autism Spectrum Disorders using Classification Techniques", *International Journal of Soft Computing and Engineering (IJSCE)*, Vol. 4, No.5, Nov.2014.