IJAS

A Review on Diagnostic of Autism Spectrum Disorder Based on the Machine Learning Approaches

M.KAVITHA[1], Dr.M.KASTHURI[2]

[1]*Research Scholar, Department of Computer Science, Bishop Heber College, Affiliated to Bharathidasan University, Tiruchirappalli-620017.*
[2]*Assistant Professor, Department of Computer Science, Bishop Heber College, Affiliated to Bharathidasan University, Tiruchirappalli-620017.*

Email: [1]kavitha.ca@bhc.edu.in, [2]kasthuri.ca@bhc.edu.in

Abstract- *Autism Spectrum Disorder is a disorder associated with genetic and neurological component with a lifelong effect on communication and interaction with others. Autism Spectrum Disorders children have some disturbance activities. Understanding their necessities is one of the most challenging tasks for caregivers. The classification algorithms helps to diagnose and improve the children's lives by applying the Machine Learning techniques. This paper provides a systematic review of the Autism Spectrum Disorder approaches in the context of Data Mining. The central goal of this review is to recognize the important research trends in the field of Autism Spectrum Disorder. The analysis classification methods for Autism Spectrum Disorder approaches is provided based on evaluation metrics such as Accuracy and Sensitivity.*

Keywords—ASD, Machine Learning, Supervised Learning, SVM, Naive Bayes, Accuracy.

1. INTRODUCTION

Data mining is a technique that deploys data analysis methods and algorithms to extract meaningful information from data. The part of data mining in support of healthcare varies with the requirements and the available technologies. Data mining emphases on retrieving patterns from a given data set by using many concepts of Machine Learning (ML), database systems and statistics. Data mining which is a field which is the analysis of large data sets to identify irregularities, dependencies, interesting patterns. Data mining techniques can be classified as Supervised and unsupervised. Supervised learning is a type of machine learning algorithm which uses a known dataset.

Supervised learning has two types of algorithms:

　　　1) Classification (for categorical values)
　　　2) Regression (for continuous values)

Classification is one among the data mining tasks. Classification algorithms are employed to construct classification models that are deployed for prediction. The constructed model is first trained with some random data known as training data and then it is deployed for prediction on the unseen testing data.

Autism Spectrum Disorder (ASD) is rising and increasing steadily in the recent years. Data mining based techniques are very much useful in predicting ASD in children and adults. Generally the prediction can be done only with behavioural attributes and without doing any clinical diagnosis. With the help of novel ASD data mining based detection techniques, early detection of autism can be done, the adverse situation can be avoided and finally costs associated with delayed diagnosis can be significantly reduced. Autism Spectrum Disorder is a disease associated with the brain development. ASD may be seen at a very early age. Autism Spectrum Disorders are common in boys than in girls. It distresses a child's behaviour and interaction. It affects a child's understanding and social bonds. Apart from this children with ASD experience various types of symptoms viz.

1) Trouble in interacting with others
2) Monotonous behaviours
3) Struggling to function properly in carrying out day-to-day activities.

ASD screening is normally conducted by a medical specialist. Diagnosing the ASD will have various assessment tools is also known as screening tool which have wide range of intelligence quotient. The intelligence quotients contains of questionnaire used for specific age groups which aims to investigate whether child, adolescent and adult have symptoms of ASD. The ASD process involves collecting a large number of responses of the child or adult behaviour either given by the parents or a caregiver. The result of the analysis can help us in understanding the skills, the health condition and education styles to be adopted for such people. It also can help us in taking precautions against certain practices which may worsen the victim's condition. The major part of the people with autism in India are unlike diagnosed and they did not receive the assistance they need. The Medical Professionals misunderstood the status of patients because of their lack of knowledge for diagnosing autism illnesses.

2. RELATED WORKS

Data mining provides us with various techniques and during this paper we might be focussing on the classification technique.

Logistic regression is a calculation used to expect a binary final results: both something happens, or does not. This can be exhibited as yes/no, pass/fail, alive/useless. Naive Bayes calculates the opportunity of whether or not a information point belongs inside a sure class or does no longer. In text evaluation, it may be used to categorize phrases or terms as belonging to a prese "tag" (class) or no longer. k-nearest neighbours (k-nn) is a sample reputation algorithm that makes use of training datasets to discover the okay Closest spouse and children in destiny examples. A choice tree is a supervised getting to know set of rules that is perfect for type issues, as it's able to order instructions on a particular degree. It really works like a drift chart, keeping apart facts factors into two comparable classes at a time from the "tree trunk" to "branches," to "leaves," wherein the categories become more finitely similar. The random wooded area set of rules is a spread of selection tree, in that, you first assemble a few-axis actual-world selection Timber with schooling facts, then healthy your new records inside one of the bushes as a "random wooded area". A guide vector machine (SVM) makes use of algorithms to educate and classify records within stages of polarity, taking it to a point beyond x/y prediction.

The table-I (Appendix-I) is illustrates the related works in ASD.

3. CONCLUSION

Data mining techniques plays vital role in the decision-making processes in many areas including social media, medical field and text analysis. By using ML techniques the precision, efficiency, specificity, accuracy, recall, and confusion matrix were computed for the dataset acquired from the UCI repository. The SVM, Naïve Bayes, KNN, AC, Random forest algorithms have been used in the classification. Based on that, all of those algorithms showed good performance in serving the autism patients, additionally to reinforce the prediction process that decide if the person has autism spectrum disorder or not. The acquired results indicate that the SVM method produces high classification scores, indicated by sensitivity, accuracy and f-measure. Through experiments we found that the SVM method is more effective than the other methods for ASD data classification.

4. REFERENCES

[1]   R. Geetha Ramani and K. Sivaselvi, "Autism Spectrum Disorder Identification using Data Mining techniques", In: International Journal of Pure and Applied Mathematics, Vol. 117, pp. 427-436, 2017.

[2]   B.Tyagi, R.Mishra and N.Bajpai, "Machine Learning Techniques to Predict Autism Spectrum Disorder", In: IEEE Punecon, Pune, India, pp. 1-5, 2018.

[3]   Fadi Thabtah and David Peebles, "A New Machine Learning Model Based on Induction of Rules for Autism Detection", In: Heal Informatics journal, Vol. 26, pp. 264-283, 2019.

[4]   Gok Murat, "A Novel Machine Learning Model to Predict Autism Spectrum Disorders Risk Gene", In: Neural Computing and Applications, Vol. 31, pp. 6711-6717, 2019.

[5]   Kazi Shahrukh Omar, Prodipta Mondal, Nabila Shahnaz Khan,   Md. Rezaul Karim Rizvi and Md. Nazrul Islam, "A Machine Learning Approach to Predict Autism Spectrum Disorder", In: International Conference on Electrical Computer and Communication Engineering, Cox's Bazar, Bangladesh, pp. 7-9, 2019.

[6]   Richard E. Griffioen, Steffie van der Steen, Theo Verheggen, Marie Jose Enders and Slegers, "Changes in Behavioral Synchrony during Dog-Assisted Therapy for Children with Autism Spectrum Disorder and Children with Down Syndrome", In: Journal of Applied Research in Intellectual Disabilities, pp. 1-11, 2019.

[7]   S.B.Shuvo, J.Ghosh and A.S.Oyshi, "A Data Mining Based Approach to Predict Autism Spectrum Disorder Considering Behavioral Attributes", In: 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, pp. 1-5, 2019.

[8]   Tania Akter, Md. Shahriare Satu, Md. Imran Khan, Mohammad Hanif Ali and Shahadat Uddin, "Machine Learning –Based Models for Early Stage Detection of Autism Spectrum Disorders", In: IEEE, Vol. 7, 2019.

[9]   Ugur Erkan and Dang N. H. Thanh, "Autism Spectrum Disorder Detection with Machine Learning Methods", In: Current   Psychiatry Reviews, Vol. 15, pp. 297-308, 2019.

[10] Pooja Rani, "Emotion Detection of Autistic Children Using Image Processing", In: Fifth International Conference on Image Information Processing (ICIIP), pp. 532-535, 2019.

[11] R. Abitha, and S.Mary Vennila, "A Swarm Based Symmetrical Uncertainty Feature Selection Method for Autism Spectrum Disorders", In: IEEE Xplore International Conference on Inventive Systems and Control (ICISC 2019), pp.665-669, 2019.

[12] S.Sadiq, M.Castellanos, J.Moffitt, M.Shyu, L.Perry and D.Messinger, "Deep Learning Based Multimedia Data Mining for Autism Spectrum Disorder (ASD) Diagnosis", International Conference on Data Mining Workshops (ICDMW), Beijing, China, pp. 847-854, 2019.

[13] Suman Raj and Surfaraz Masood, "Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques ", In: Elsevier, pp. 994-1004, 2020.

[14] Anshu Sharma and Dr. Poonam Tanwar, "Deep Analysis of Autism Spectrum Disorder Detection Techniques", In: IEEE Xplore, pp. 455-459, 2020.

[15] Jaber Alwidian, Ammar Elhassan and Rawan Ghnemat, "Predicting Autism Spectrum Disorder using Machine Learning Technique", In: International Journal of Recent Technology and Engineering, Vol. 8, pp. 2277-3878, 2020.

[16] N. Priya, and C. Radhika, "Effective Implementation of Pre-Processing Techniques in Machine Learning for Autism Spectrum Disorder", In: International Journal of Innovative Technology and Exploring Engineering (IJITEE), Vol-9, Issue-5, pp.2253-2257, 2020.

[17] R. Abitha, and S. Mary Vennila, "A Proficient Optimized Feature Selection Method Germane for Autism Spectrum Disorder Classification", In: International Journal of Scientific & Technology Research, Volume-9, Issue-02, pp.1133-1137, 2020.

[18] Shuaibing Liang, Aznul Qalid Md Sabri, Ady Alnajjar and Chu Kiong Loo, "Autism Spectrum Self-Stimulatory Behaviors Classification Using Explainable Temporal Coherency Deep Features and SVM Classifier", In: IEEE Xplore, Volume-9, pp.34264-34275, 2021.

APPENDIX-I
TABLE-I
RELATED WORKS

Name of the Author	Year	Title	Proposed Methods	Data Set	Accuracy	Advantage	Limitations
R.Geetha Ramani, et.al [1]	2017	Autism Spectrum Disorder Identification using Data Mining Techniques	Supervised data mining techniques are applied, Naive Bayes, SVM, Random Tree, C4.5, CS-CRT	UCLA's Center for Autism Research and Treatment (CART) [15]. Total ASD 60. Males 52. females 8 Total TD 45 Males 38 and 7 female	Random tree classifier with Fisher feature selection achieved a higher accuracy of 88.46%	The methodology can be used to assist the medical practitioners in the identification of ASD	Ensemble classification is applied on these features and no improved results have been obtained.

Bhawana Tyagi, et.al, [2]	2018	Machine Learning Techniques to Predict Autism Spectrum Disorder	KNN, SVM, LR, CART, Naïve Bayes, LDA	UCI Machine learning	LDA Accuracy 72.2024%	Comparative analysis of the performance of different algorithms to diagnose ASD	Only adult people from the age of 16 to 17 years were considered as samples.
Fadi Thabtah, and David Peebles, et.al, [3]	2019	A New Machine Learning Model Based on Induction of Rules for Autism Detection	logistic regression Decision tree (CART), neural network and CHAID algorithms, Random Forest	Real data sets collected using a mobile application called ASDTests and University of California Irvine Repository (UCI)	logistic regression produced Accuracy 95%	Least number of items required for screening. Sensitivity and specificity improved.	Only used for adults. Toddlers are not including.
Gok Murat [4]	2019	A Novel Machine Learning Model to Predict Autism Spectrum Disorders Risk Gene	Naive Bayes, Bayes network, Random forest, Linear SVM, RBF SVM	BrainSpan Atlas of the Developing Human Brain developmental transcriptome dataset, The dataset consists of 366 known ASD genes panning from 2128 disease genes	RBF SVM Produced Accuracy 83.6%	Imaging and biomarkers are tools for diagnostic early ASD	Proposed model achieved 83.6% Ensemble of classification provide better accuracy.
Saad Sadiq, et.al, [12]	2019	Deep Learning Based Multimedi	Linear Regression, Nearest	UCI and real data set	84.7%	The proposed framework	Not applicable to under 4 years

		a Data Mining for Autism Spectrum Disorder (ASD) Diagnosis	Neighbour, Naïve Bayes, Random forest Support Vector machines.			achieved state-of-the- art-predictive diagnostic estimates of ASD severity compared to other leading algorithms.	of age children.
Kazi Shahrukh Omar, et.al, [5]	2019	A Machine Learning Approach to Predict Autism Spectrum Disorder	Merging Random forest, CART, ID3	People from with and without autistic AQ-10 data set and 250 real data set are collected.	Overall accuracy of 97.10%	Usability is high. The proposed system can be enhanced effectively by collecting user experience.	Lack of large data set. Below 3 years of aged kids cannot use this.
Ugur Erkan, et.al, [9]	2019	Autism Spectrum Disorder Detection with Machine Learning Methods	SVM, RF,KNN	UCI data base, AQ-10-Adult, AQ-10-child, AQ-10-Adolescence	RF-99% SVM-90% KNN showed lower than RF and SVM	We detect ASD easily, fast and very high accuracy	Accuracy depends on the data sets.
Richard Eric Griffioen, et.al, [6]	2019	Changes in Behavioral Synchrony During Dog-Assisted Therapy for Children	Dog-assisted therapy	UCI, Raw child Behaviour Checklist(CBCL)	78.2%	Test synchrony hypothesis for ASD child	The CBCL has also been criticized for only measuring children's emotional and behavioural problem, but not the

		with Autism Spectrum Disorder and Children with Down Syndrome					presence or absence of pro-social behaviour. Measures failed to reach significance and unclear
Shaon Bhatta Shuvo, et.al [7]	2019	A Data Mining Based Approach to Predict Autism Spectrum Disorder Considering Behavioural Attributes	Random forest classifier algorithm	UCI, Real time data sets were collected.	Overall accuracy of 96%.	Accuracy, sensitivity, specificity were calculated	Random Forest is difficult because it gives different results in different times
Tania akter, et.al [8]	2019	Machine Learning – Based Models for Early Stage Detection of Autism Spectrum Disorders	SVM log, ZScore, FT methods	Kaggle and the University of California-Irvine (UCI)	Toddler-98.77%, Child-97.20%, Adolescent-93.89% Adult 98.36%	It will useful for physicians to detect ASD in early stage	Need more number of data set for detection of ASD and neurodevelopmental disorder
Pooja Rani [10]	2019	Emotion Detection of Autism children using Image Processing	SVM, Neural Network	Data set downloaded from google 40% used for testing purpose, 60% for training purpose	SVM gave better Accuracy 90%	High performance is achieved	Age and gender attributes are not used for prediction of facial expressions.
Suman Raj and Surfaraz	2019	Analysis and Detection of Autism Spectrum	Naïve Bayes, Support Vector Machine,	Datas collected from UCI Repository	Among all SVM and CNN model	Comparing more than one algorithms helps	This algorithm could be more efficient if

Masood [13]		Disorder Using Machine Learning Techniques	KNN, logistic regression, Convolutional neural network	Three types of dataset used Adult, children, adolescent	gave best result, accuracy 98.30%	in achieving the Highest accuracy	they would have predicted for infant data sets also
R.Abitha, and S. Mary Vennila [11]	2019	A Swarm Based Symmetrical Uncertainty Feature Selection Method for Autism Spectrum Disorders	Feature selection methods like SU, IG, CS and optimization techniques like PSO, GA and ACO were explored	Datas collected from UCI Repository	Naïve Bayes Accuracy 86.747%	Proposed SSU-FS method provided better results than the existing methods.	No clear explanation on the identification of the redundant and irrelevant features was given
N. Priya, C. Radhika [16]	2020	Effective Implementation of Pre-Processing Techniques in Machine Learning for Autism Spectrum Disorder	Random Forest, SVM, Logistic Regression, KNN, Naive Bayes.	Dataset from Kaggle Repository. The data set consists of 1054 observations of 18 features	Preprocessing technique and Random forest method shows 92% accuracy	The proposed technique sstandardized the data efficiently to carry out successful analysis	The minimal subset feature were not identified which were essential to detect the ASD in early stage
Anshu Sharma and Dr. Poonam Tanwar [14]	2020	Deep Analysis of Autism Spectrum Disorder Detection Techniques	SVM, Random forest Scan, decision trees, logistic regression	250 Real dataset collected from different age group with and without autism people.	Among all SVM gave best result, accuracy 98.27%	Accuracy and precisions are improved	Only structured dataset considered for diagnosis

Jaber Alwidian, et.al, [15]	2020	Predicting Autism Spectrum Disorder using Machine Learning Technique	Association Classification	21 attributes to cover 704 instances, where 515 instances classified under no autism class label and 189 instances under autism class label	Overall accuracy of 97%	Good Performance in serving the autism patients. Gave a strong indicator about the potential power of critical domain.	Modifying one of the existing AC algorithms achieve low accuracy
R.Abitha, and S. Mary Vennila [17]	2020	A Proficient Optimized Feature Selection Method Germane For Autism Spectrum Disorder Classification	Symmetrical Uncertainty and Cultural Algorithm based feature selection (ca-fs) method	Dataset was composed of several special schools in and around Chennai. Novel dataset collected of 75 characteristics	89.41%	proposed CA-FS method extracted the least number of features than the existing feature selection techniques	Justification for the selection of robust features was not clearly given
Shuaibing Liang, Aznul Qalid Md Sabri, Chu Kiong Loo [18]	2021	Autism Spectrum Self-Stimulatory Behaviors Classification Using Explainable Temporal Coherency Deep Features and SVM Classifier	Unsupervised Temporal Coherency Deep Networks (TCDN) method with Optimised Supervised learning methods (such as SVM, k-NN,	Real-time detection of the children's behaviour and SSBD [25] was selected as our training test dataset. 75 self-stimulatory behaviour videos	State-of-the-art results by combining the unsupervised and supervised learning methods 73.6%	Many unlabelled autistic self-stimulatory behaviours videos were explored to obtain thorough knowledge	Small data set was deployed so that it cannot be generalized to fit high volume data

			Discriminant)				