IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Machine Learning Based Model for Prediction of Autism Spectrum Disorder

Sandhya A Kulkarni [1]
Asst.Professor
Department of Computer Science
KSSEM
Bangalore, India
sandhya.a.kulkarni@kssem.edu.in

Dhanush Vasudevaraju [2]
Student
Department of Computer Science
KSSEM
Bangalore, India soccerdhanush@gmaill.com

C Ramesh Chandra [3]
Student
Department of Computer Science
KSSEM
Bangalore, India
rameshchalla2001@gmail.com

Dinesh M [4]
Student
Department of Computer Science
KSSEM
Bangalore, India
dinesh96061@gmail.com

C Sai Ranganath [5]
Student
Department of Computer Science
KSSEM
Bangalore, India
challasairanganath@gmail.com

Abstract *— Autism spectrum disorder is a neurodevelopmental disorder that affects a person's interaction, communication and learning skills. Although diagnosis of autism can be done at any age, its symptoms generally appear in the first two years of life and develop through time. Autism patients face different types of challenges such as difficulties with concentration, learning disabilities, mental health problems such as anxiety, depression, motor difficulties, sensory problems, and many others. Diagnosis of autism requires significant amount of time and cost. Earlier detection of autism can come to a great help by prescribing patients with proper medication at an early stage. It can prevent the patient's condition from deteriorating further and would help to reduce long term costs associated with delayed diagnosis. Thus, an efficient, accurate and easy screening test tool is very much required which would predict autism traits in an individual. The main idea behind this project is to detect autism spectrum disorder in an individual (male/female). This project is implemented by making use of a Machine Learning model using parameters such as an individual's age, gender, ethnicity, Autism Quotient Tool. The detection derived from this project will help an individual to get required diagnosis in time to prevent further complications of developing Alzheimer's disease.*

Keywords—machine learning,autism spectrum disorder, parameters,autism quotient tool

I. INTRODUCTION

This paper focuses on proposing a model which would assist in prediction of ASD in an individual, so that daignosis can be don and further treatments may be followed. The dataset used is the Autistic Spectrum Disorder Screening Data [3]. The datasets provide insights into various factors affecting the prediction of the disorder. Machine learning algorithms like decision tree, random forest, logistic regression, support vector classifier and artificial neural networks are used for finding out the optimal modelel for each dataset. Several performance metrics are used in order to analyze and compare each model from every angle possible.

II. RELATED WORK

Machine Learning(ML) is a type of artificial intelligence(AI) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so.Classical machine learning is often categorized by how an algorithm learns to become accurate in its predictions. There are four basic approaches:

A. Supervised learning

In this type of machine learning, data scientisits supply algorithms with labeled training data and define the variables they want the algorithm to assess for correlations. Both the input and the output of the algorithm is used.

B. Unsupervised learning

This type of machine learning involves algorithms that train on unlabeled data. The algorithm scans through data sets looking for any meaningful connection. The data that algorithms train on as well as the predictions or recommendations they output are predetermined.

C. Semi-supervised learning

This approach to machine learning involves a mix of the two preceding types. Data scientists may feed an algorithm mostly labeled training data, but the model is free to explore the data on its own and develop its own understanding of the data set.

D. Reinforcement learning

Data scientists typically use reinforcement learning to teach a machine to complete a multi-step process for which there are clearly defined rules. Data scientists program an algorithm to complete a task and give it positive or negative cues as it works out how to complete a task. But for the most part, the algorithm decides on its own what steps to take along the way.

To further describe machine learning, its life-cycle is described below:-

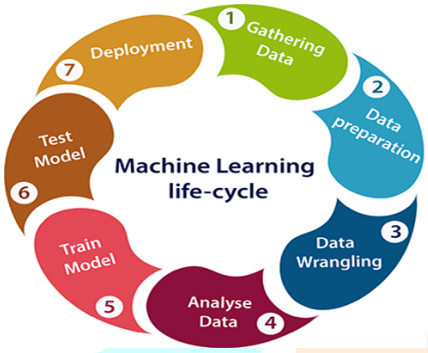

Fig 1 :- Machine Learning Life-Cycle

1. *Gathering Data*

Data Gathering is the first step of the machine learning life cycle. The goal of this step is to identify and obtain all data-related problems.

2. *Data preparation*

After collecting the data, we need to prepare it for further steps. Data preparation is a step where we put our data into a suitable place and prepare it to use in our machine learning training.

3. Data Wrangling

Data wrangling is the process of cleaning and converting raw data into a useable format. It is the process of cleaning the data, selecting the variable to use, and transforming the data in a proper format to make it more suitable for analysis in the next step. It is one of the most important steps of the complete process. Cleaning of data is required to address the quality issues.

4. Data Analysis

Now the cleaned and prepared data is passed on to the analysis step. This step involves:

- Selection of analytical techniques
- Building models
- Review the result

5. *Train Model*

In this step we train our model to improve its performance for better outcome of the problem. We use datasets to train the model using various machine learning algorithms. Training a model is required so that it can understand the various patterns, rules, and, features.

6. Test Model

In this step, we check for the accuracy of our model by providing a test data set to it. Testing the model determines the percentage accuracy of the model as per the requirement of project or problem.

7. Deployment

The last step of machine learning life cycle is deployment, where we deploy the model in the real-world system. If the above-prepared model is producing an accurate result as per our requirement with acceptable speed, then we deploy the model in the real system. But before deploying the project, we will check whether it is improving its performance using available data or not.

IV. The Proposed Method

A. Random Forest Algorithm

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees. The existing base papers make use of this algorithm.

1. **"A Machine Learning Approach to Predict Autism Spectrum Disorder" [1]**

The objective of this work is to propose an autism prediction model using ML techniques and to develop a mobile application that could effectively predict autism traits of an individual of any age. In other words, this work focuses on developing an autism screening application for predicting the ASD traits among people of age groups 4-11 years, 12-17 years and for people of age 18 and more.

2. Autistic Spectrum Disorder Screening: Prediction with Machine Learning Models [3]

This paper uses ASD screening dataset for analysis and prediction of probable cases in adults, children and adolescents. The dataset for each of the age groups are analyzed and inferences are drawn from them. Machine learning algorithms like Random Forest.

B. Support Vector Machines

Support Vector Machines (SVMs, also support vector networks are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. It is formally defined by a separating hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyperplane which categorizes new examples.The existing base papers make use of this algorithm.

1. **Machine Learning-Based Models for Early Stage Detection of Autism Spectrum Disorders [2]**

Various classification techniques were implemented with these transformed ASD datasets and assessed for their performance. SVM showed the best performance. The results of these analytical approaches indicate that, when appropriately optimised, machine learning methods can provide good predictions of ASD status. This suggests that it may possible to apply these models for the detection of ASD in its early stages.

2. **Diagnostic prediction of autism spectrum disorder using complex network measures in a machine learning framework [17]**

In this paper, objective imaging-based biomarker discovery for psychiatric conditions is critical for accurate diagnosis and treatment. Using a machine learning framework, this work investigated the utility of brain's functional network topology extracted from functional magnetic resonance imaging (fMRI) functional connectivity (FC) as viable biomarker of autism spectrum disorder (ASD).

A battery of complex network features were computed from the FC network using graph theoretic techniques. Recursive-Cluster-Elimination Support Vector Machine algorithm was employed to compare the predictive performance of three independent feature sets, (i) FC, (ii) complex network measures, and (iii) both combined. The study found that FC could diagnose ASD with 67.3 % accuracy and graph measures with 64.5 % accuracy, while the combined feature set could diagnose with 70.1 % accuracy (all accuracies were significantly different, $p < 10^{-30}$).

C. K-Nearest Neighbour

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another.The existing base papers make use of this algorithm.

1. Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children [19]

In this paper, the classification method for ASD diagnosis was used in children aged 4-11 years. The K-Nearest Neighbor (KNN) algorithm is used for classification. To test the algorithms, 30 percent of the data set was selected as test data and 70 percent as training data. As a result of the work done, the accuracy of the KNN algorithm is 88.5%. For the For KNN algorithm, these values are calculated as 0.9762 and 0.80. F-measure values are calculated as 0.8913 for the KNN algorithm.

2. Autism Spectrum Disorder Prediction Using Machine Learning Algorithms [44]

The objective of the research is to foresee Autism Spectrum Disorder (ASD) in toddlers with the help of machine learning algorithms. Of late, machine learning algorithms play vital role to improve diagnostic timing and accuracy. This research work precisely compares and highlights the effectiveness of the feature selection algorithms K-Nearest Neighbor (KNN), and to improve the efficiency of Random Tree classification algorithm while modelling ASD prediction in toddlers. and to improve the efficiency of Random Tree classification algorithm while modelling ASD prediction in toddlers. Analysis results uncover that the Random Tree dependent on highlights chosen by Extra Tree calculation beat the individual methodologies.

V. CONCLUSION

This research comes under machine learning which is advanced technique at present. This research provides a comparative view among different ML approach in terms of their performance. The results showed that Random Forest algorithm provides highest accuracy compared to other machine learning algorithms such as Support Vector Machines and K-Nearest Neighbour. Different existing data mining procedures and its application were considered or explored. Utilization of machine learning algorithms was connected in various medical data sets. Machine learning strategies have diverse power in different medical data sets. Previously mentioned conventional machine learning techniques gave less exact outcome and results additionally shifts in light of the procedures has been utilized for the prediction. The existing system **c**annot properly represent the diversity of autism classes with complex intra-class variability and inter-class similarity and it is less effective for autism classification.

REFERENCES

[1] Omar, Kazi Shahrukh, Prodipta Mondal, Nabila Shahnaz Khan, Md Rezaul Karim Rizvi, and Md Nazrul Islam. "A machine learning approach to predict autism spectrum disorder." In 2019 International conference on electrical, computer and communication engineering (ECCE), pp. 1-6. IEEE, 2019.doi:10.1109/ecace.2019.8679454

[2] Akter, Tania, Md Shahriare Satu, Md Imran Khan, Mohammad Hanif Ali, Shahadat Uddin, Pietro Lio, Julian MW Quinn, and Mohammad Ali Moni. "Machine learning-based models for early stage detection of autism spectrum disorders." IEEE Access 7 (2019): 166509-166527.doi:10.1109/access.2019.2952609

[3] Baranwal, Astha, and M. Vanitha. "Autistic spectrum disorder screening: prediction with machine learning models." In 2020 International conference on emerging trends in information technology and engineering (ic-ETITE), pp. 1-7. IEEE, 2020. . doi:10.1109/ic-etite47903.2020.186

[4] Eni, M, Dinstein, I, Ilan, M., Menashe, I, Meiri, G, & Zigel, Y. (2020). Estimating Autism Severity in Young Children From Speech Signals Using a Deep Neural Network. IEEE Access, 8, 139489–139500. doi:10.1109/access.2020.301253

[5] Raj, S, & Masood, S. (2020). Analysis and Detection of Autism Spectrum Disorder Using Machine Learning Techniques. Procedia Computer Science, 167, 994-1004.doi:10.1016/j.procs.2020.03.39

[6] Mamun, K. A. A., Bardhan, S., Ullah, M. A., Anagnostou, E., Brian, J., Akhter, S., & Rabbani, M. G. (2016). Smart autism — A mobile, interactive and integrated framework for screening and confirmation of autism. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi:10.1109/embc.2016.7592093

[7] Bardhan, S., Mridha, G. M. M. M., Ahmed, E., Ullah, M. A., Ahmed, H. U., Akhter, S., … Mamun, K. A. A. (2016). Autism Barta — A smart device based automated autism screening tool for Bangladesh. 2016 5th International Conference on

Informatics, Electronics and Vision (ICIEV). doi:10.1109/iciev.2016.7760073

[8] Min, C.-H. (2017). Automatic detection and labeling of self-stimulatory behavioral patterns in children with Autism Spectrum Disorder. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). doi:10.1109/embc.2017.8036816

[9] Kirana, M. C., Lubis, M. Z., & Amalia, E. D. (2018). The Effect of Sound Manipulation to Know Response Rate in Autism Children Using FFT. 2018 International Conference on Applied Engineering(ICAE).doi:10.1109/incae.2018.8579418

[10] Obafemi-Ajayi, T., Settles, L., Su, Y., Germeroth, C., Olbricht, G. R., Wunsch, D. C., … Miles, J. (2017). Genetic variant analysis of boys with Autism: A pilot study on linking facial phenotype to genotype. 2017 IEEE International Conference on Bioinformatics and Biomedicine(BIBM).doi:10.1109/bibm.2017.8217658

[11] Thabtah, F., Kamalov, F., & Rajab, K. (2018). A new computational intelligence approach to detect autistic features for autism screening. International Journal of Medical Informatics, 117, 112–124. doi:10.1016/j.ijmedinf.2018.06.009

[12] Liaqat, S., Wu, C., Duggirala, P. R., Cheung, S. S., Chuah, C.-N., Ozonoff, S., & Young, G. (2021). Predicting ASD diagnosis in children with synthetic and image-based eye gaze data. Signal Processing: Image Communication, 94, 116198. doi:10.1016/j.image.2021.116198

[13] Wijesinghe, A., Samarasinghe, P., Seneviratne, S., Yogarajah, P., & Pulasinghe, K. (2019). Machine Learning Based Automated Speech Dialog Analysis Of Autistic Children. 2019 11th International Conference on Knowledge and Systems Engineering (KSE). doi:10.1109/kse.2019.8919266

[14] Wang, C., Xiao, Z., Wang, B., & Wu, J. (2019). Identification of Autism Based on SVM-RFE and Stacked Sparse Auto-Encoder. IEEE Access, 7, 118030–118036. doi:10.1109/access.2019.2936639 .

[15] Wei, W., Liu, Z., Huang, L., Wang, Z., Chen, W., Zhang, T., Xu, L. (2021). Identify autism spectrum disorder via dynamic filter and deep spatiotemporal feature extraction. Signal Processing: Image Communication, 94, 116195. doi:10.1016/j.image.2021.116195

[16] Mazumdar, P., Arru, G., & Battisti, F. (2021). Early detection of children with Autism Spectrum Disorder based on visual exploration of images. Signal Processing: Image Communication, 94, 116184. doi:10.1016/j.image.2021.116184

[17] Chaitra, N., Vijaya, P. A., & Deshpande, G. (2020). Diagnostic prediction of autism spectrum disorder using complex network measures in a machine learning framework. Biomedical Signal Processing and Control, 62, 102099. doi:10.1016/j.bspc.2020.102099

[18] Sato, J. R., Calebe Vidal, M., de Siqueira Santos, S., Brauer Massirer, K., & Fujita, A. (2018). Complex Network Measures in Autism Spectrum Disorders. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 15(2), 581–587. doi:10.1109/tcbb.2015.2476787

[19] Altay, O., & Ulas, M. (2018). Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-nearest neighbor in children. 2018 6th International Symposium on Digital Forensic and Security (ISDFS). doi:10.1109/isdfs.2018.8355354

[20] Abraham, A., Milham, M. P., Di Martino, A., Craddock, R. C., Samaras, D., Thirion, B., & Varoquaux, G. (2017). Deriving reproducible biomarkers from multi-site resting-state data: An Autism-based example. NeuroImage, 147, 736–745. doi:10.1016/j.neuroimage.2016.10

[21] Milne, M., Powers, D., & Leibbrandt, R. (2009). Development of a software-based social tutor for children with autism spectrum disorders. Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group on Design: Open 24/7 - OZCHI '09. doi:10.1145/1738826.1738870

[22] Zhang, L., Wade, J., Bian, D., Fan, J., Swanson, A., Weitlauf, A., … Sarkar, N. (2017). Cognitive Load Measurement in a Virtual Reality-Based Driving System for Autism Intervention. IEEE Transactions on Affective Computing, 8(2), 176–189. doi:10.1109/taffc.2016.2582490

[24] [23] Kundu, M. R., & Suranjan Das, M. (2019). Predicting Autism Spectrum Disorder in Infants Using Machine Learning. Journal of Physics: Conference Series, 1362, 012018. doi:10.1088/1742-6596/1362/1/012018

[25] Cannon, J., O'Brien, A. M., Bungert, L., & Sinha, P. (2021). Prediction in Autism Spectrum Disorder: A Systematic Review of Empirical Evidence. Autism Research, 14(4), 604–630. doi:10.1002/aur.2482

[25] Skafidas, E., Testa, R., Zantomio, D., Chana, G., Everall, I. P., & Pantelis, C. (2012). Predicting the diagnosis of autism spectrum disorder using gene pathway analysis. Molecular Psychiatry, 19(4), 504–510. doi:10.1038/mp.2012.126

[26] Howsmon, D. P., Kruger, U., Melnyk, S., James, S. J., & Hahn, J. (2017). Classification and adaptive behavior prediction of children with autism spectrum disorder based upon multivariate data analysis of markers of oxidative stress and DNA methylation. PLOS Computational Biology, 13(3), e1005385. doi:10.1371/journal.pcbi.1005385

[27] Liu, W., Yu, X., Raj, B., Yi, L., Zou, X., & Li, M. (2015). Efficient autism spectrum disorder prediction with eye movement: A machine learning framework. 2015 International Conference on Affective Computing and Intelligent Interaction (ACII). doi:10.1109/acii.2015.7344638

[28] Gök, M. (2018). A novel machine learning model to predict autism spectrum disorders risk gene. Neural Computing and Applications. doi:10.1007/s00521-018-3502-5

[29] Oh, D. H., Kim, I. B., Kim, S. H., & Ahn, D. H. (2017). Predicting Autism Spectrum Disorder Using Blood-based Gene Expression Signatures and Machine Learning. Clinical Psychopharmacology and Neuroscience, 15(1), 47–52. doi:10.9758/cpn.2017.15.1.47

[30] Xu, G., Strathearn, L., Liu, B., & Bao, W. (2018). Prevalence of Autism Spectrum Disorder Among US Children

and Adolescents, 2014-2016. JAMA, 319(1), 81. doi:10.1001/jama.2017.17812

[31] Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. Translational Psychiatry, 5(2), e514–e514. doi:10.1038/tp.2015.7

[32] Lyall, K., Croen, L., Daniels, J., Fallin, M. D., Ladd-Acosta, C., Lee, B. K., … Newschaffer, C. (2017). The Changing Epidemiology of Autism Spectrum Disorders. Annual Review of Public Health, 38(1), 81–102. doi:10.1146/annurev-publhealth-031816-044318

[33] Ejlskov, L., Wulff, J. N., Kalkbrenner, A., Ladd-Acosta, C., Fallin, M. D., Agerbo, E., … Schendel, D. (2021). Prediction of Autism Risk From Family Medical History Data Using Machine Learning: A National Cohort Study From Denmark. Biological Psychiatry Global Open Science, 1(2), 156–164. doi:10.1016/j.bpsgos.2021.04.007

[34] Bar-Haim, Y., Shulman, C., Lamy, D., & Reuveni, A. (2006). Attention to Eyes and Mouth in High-Functioning Children with Autism. Journal of Autism and Developmental Disorders, 36(1), 131–137. doi:10.1007/s10803-005-0046-1

[35] Brown, C. J., Kawahara, J., & Hamarneh, G. (2018). Connectome priors in deep neural networks to predict autism. 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). doi:10.1109/isbi.2018.8363534

[36] Åsberg Johnels, J., Hovey, D., Zürcher, N., Hippolyte, L., Lemonnier, E., Gillberg, C., & Hadjikhani, N. (2016). Autism and emotional face-viewing. Autism Research, 10(5), 901–910. doi:10.1002/aur.1730

[37] Duan, H., Min, X., Fang, Y., Fan, L., Yang, X., & Zhai, G. (2019). Visual Attention Analysis and Prediction on Human Faces for Children with Autism Spectrum Disorder. ACM Transactions on Multimedia Computing, Communications, and Applications, 15(3s), 1–23. doi:10.1145/3337066

[38] Eman, D., & Emanuel, A. W. . (2019). Machine Learning Classifiers for Autism Spectrum Disorder: A Review. 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). doi:10.1109/icitisee48480.2019.9003807

[39] Stevens, E., Dixon, D. R., Novack, M. N., Granpeesheh, D., Smith, T., & Linstead, E. (2019). Identification and analysis of behavioral phenotypes in autism spectrum disorder via unsupervised machine learning. International Journal of Medical Informatics, 129, 29–36. doi:10.1016/j.ijmedinf.2019.05.006

[40] Heinsfeld, A. S., Franco, A. R., Craddock, R. C., Buchweitz, A., & Meneguzzi, F. (2018). Identification of autism spectrum disorder using deep learning and the ABIDE dataset. NeuroImage:Clinical, 17, 16–23. doi:10.1016/j.nicl.2017.08.017

[41] Ghosh, T., Banna, M. H. A., Rahman, M. S., Kaiser, M. S., Mahmud, M., Hosen, A. S. M. S., & Cho, G. H. (2021). Artificial intelligence and internet of things in screening and management of autism spectrum disorder. Sustainable Cities and Society,74,103189. doi:10.1016/j.scs.2021.103189

[42] Bussu, G., Jones, E. J. H., Charman, T., Johnson, M. H., & Buitelaar, J. K. (2018). Prediction of Autism at 3 Years from Behavioural and Developmental Measures in High-Risk Infants: A Longitudinal Cross-Domain Classifier Analysis. Journal of Autism and Developmental Disorders, 48(7), 2418–2433. doi:10.1007/s10803-018-3509-x

[43] Chowdhury, K., & Iraj, M. A. (2020). Predicting Autism Spectrum Disorder Using Machine Learning Classifiers. 2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT). doi:10.1109/rteict49044.2020.9315

[44] Selvaraj, S., Palanisamy, P., Parveen, S., & Monisha. (2020). Autism Spectrum Disorder Prediction Using Machine Learning Algorithms. Advances in Intelligent Systems and Computing, 496–503. doi:10.1007/978-3-030-37218-7_56

[45] Sharma, A., & Tanwar, P. (2020). Deep Analysis of Autism Spectrum Disorder Detection Techniques. 2020 International Conference on Intelligent Engineering and Management (ICIEM). doi:10.1109/iciem48762.2020.9160123 10.1109/ICIEM48762.2020.9160123

[46] Nor, M. N. M., Jailani, R., & Tahir, N. M. (2016). Analysis of EMG signals of TA and GAS muscles during walking of Autism Spectrum Disorder (ASD) children. 2016 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). doi:10.1109/iscaie.2016.7575068

[47] Thabtah, Fadi; Peebles, David (2019). A new machine learning model based on induction of rules for autism detection. Health Informatics Journal, (), 146045821882471–. doi:10.1177/1460458218824711

[48] Thabtah, Fadi; Peebles, David (2019). Early Autism Screening: A Comprehensive Review. International Journal of Environmental Research and Public Health, 16(18), 3502–. doi:10.3390/ijerph16183502